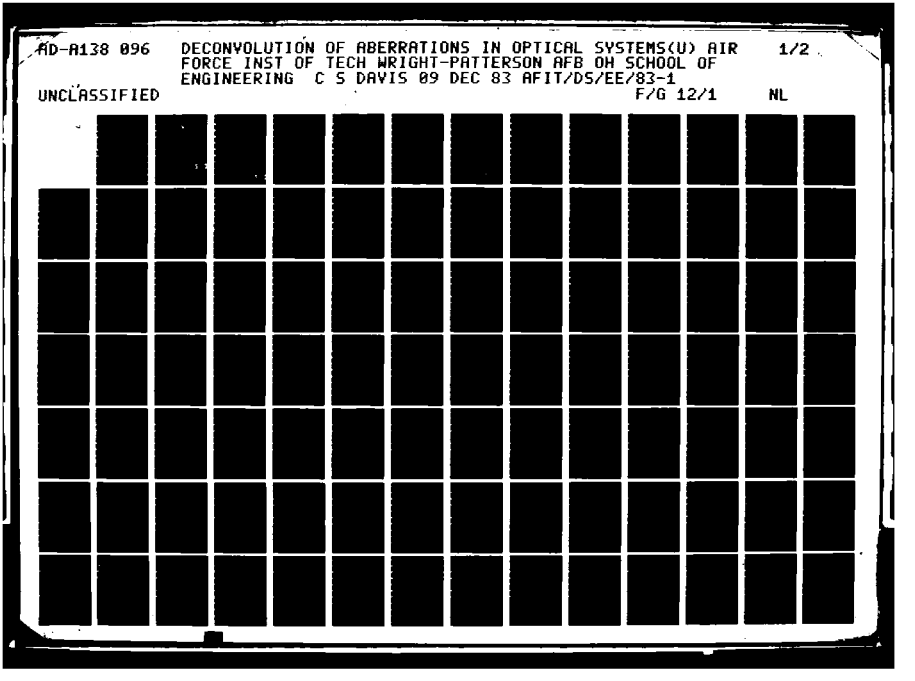
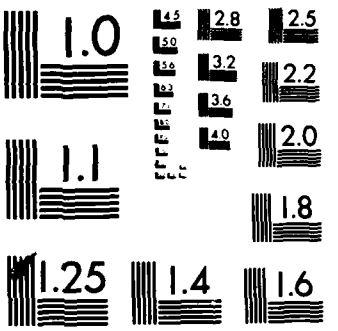
AD-A138 096 DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING C S DAVIS 09 DEC 83 AFIT/DS/EE/83-1 1/2

UNCLASSIFIED F/G 12/1 NL

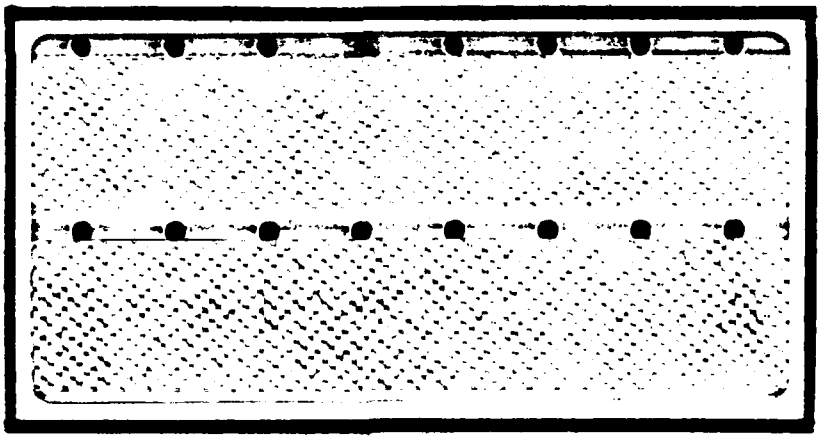




DD2

(1)

AD A138096



DEPARTMENT OF THE AIR FORCE

AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DS/EE/83-1

DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS

DISSERTATION

AFIT/DS/EE/83-1 Clair S. Davis
Capt USAF

AFIT/DS/EE/83-1

DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS

DISSERTATION

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

by

Clair S. Davis, B.S., M.E.
Captain USAF

AFIT/DS/EE/83-1

DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS

by

Clair S. Davis, B.S., M.E.
Captain USAF

Approved:

Chairman	6 Dec 83
Norman C. Shankland	9 Dec 83
Vaqar A. Syed.	Dec 12, 1983
David A. Lee	12 Dec 83

Accepted:

J S Przemieniecki 12 Dec. 1983
Dean, School of Engineering

Preface

The requirement to address the deconvolution problem came from a technology effort to develop very large, adaptive, space-borne optical systems. In optical systems, the convolution integral can be "deconvolved" quite readily by the method of stationary phase. The remaining problem, which much of this research addresses, is to solve a system of bilinear equations.

In the initial phases of this research, my goal was to develop an algorithm that could be applied to any multi-element optical system to solve for the aberrations on the elements. While an algorithm was developed, its capabilities are much more modest than originally envisioned. However, as the research progressed, it became apparent that what was needed even more than a general purpose deconvolution algorithm was a mathematical analysis to provide insight into the problem. I view the insight provided by this analysis, rather than a specific algorithm, to be the major contribution of this research effort.

Thanks are due to many people who provided technical assistance or encouragement. I would especially like to thank my advisor, Dr. Stanley Robinson, who provided both. I wish also to thank the other members of my committee for their insight and recommendations which have significantly enhanced this research. The encouragement and support

given by my parents and by my friends Diane Gourdin and Leo Silvernagel are especially appreciated. But most of all I thank my wife Ada, whose unwavering faith and support kept this effort alive and brought it to a successful conclusion.

Clair S. Davis

Contents

List of Figures

List of Tables

NOTATION

The variables and operators listed below are the ones used most often in this paper. Less frequently used variables are not listed below, but are defined in the text.

Operators

*	Convolution or conjugate (depending on context)
∂	Partial differential
†	Conjugate transpose
E	Expected value
^	Designates a random variable or process

Constants, variables, and functions

a, b	Aperture dimensions
$a_J(n,m)$	Elements of the $\underline{A}$ matrix
$\underline{A}$	Matrix (see Eq (2-29))
c,d,i,l,m,n,p,q	Integer variables
C,D,L,M,N,P,Q	Integer constants
d_i	Represents either F_i or G_i
F_m, G_n	Fourier coefficients
f_x	Spatial frequency in the x direction
h(x)	Propagation impulse response function
$H(f_x)$	Propagation transfer function
$H_J(p)$	Discrete Fourier transform derived from the J^{th} measurement of $U_5(x)$
j	$\sqrt{-1}$

J	As a subscript, designates the J^{th} measurement of the output field. Otherwise, it represents the Fisher information matrix
k	$2\pi/\lambda$
$\hat{N}(x)$	Noise (spatial random process)
Pmax, Qmax	Maximum spatial frequencies of aberrations
$P_{\hat{H}\|H}(\hat{H}\|H)$	Conditional probability density function of $\hat{H}$ given H
A'	Amplitude of input plane wave
R, I	Subscripts used to designate real and imaginary parts of complex numbers
U(x)	Measured field multiplied by known constants
$U_n(x)$	Scalar field
$W_A(x)$, $W_B(x)$	Aberration functions
Y_{mn}	$F_m G_n$
Z_1, Z_2	Distances between optical elements
λ	Wavelength of light
γ, β, ξ, η	Dummy variable of integration
θ, ϕ	Angles with respect to the x and y axes, respectively
σ	Standard deviation
Λ	Covariance matrix (Eq (3-9))

Abstract

Methods are developed for estimating aberrations on the elements of a two-element optical system based on knowledge of the input and output fields. The equation is written for the propagation of a scalar, quasi-monochromatic field through the system, and the output field is assumed measurable at a given plane. The resulting convolution integral contains the unknown aberration functions in the integrand. The integral equation is a Fredholm equation of the first kind, and the integrand is a nonlinear function of the aberrations. The integral is completed by the method of stationary phase. Methods for estimating the aberration functions are developed, and the effects of noisy measurements are considered. A computerized algorithm that will estimate the Fourier coefficients of the aberration functions is demonstrated for a simple deconvolution problem.

DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS

I. Introduction

Problem Statement

The effort to study the deconvolution problem is motivated by a technology effort called HALO (High Altitude Large Optics) which is managed by Rome Air Development Center (RADC/OCS) at Griffiss Air Force Base, New York. HALO is an attempt to develop the technology necessary to build a large spaceborne telescope. The telescope would be large enough that adaptive optical elements would be required to keep the optical surfaces aligned to within a fraction of a wavelength (Refs 12, 17, 24). Figure 1-1 is an example of such a telescope. A Cassegrain telescope is illustrated, but other configurations are possible. Light enters the telescope from the left, is reflected by the primary mirror, propagates to the secondary mirror where it is reflected and propagates to the detector. As the angle of the orbiting telescope changes with respect to the sun, changes in temperature on the mirrors will cause their position to change slightly. The difference between where the surface of the mirror is and where it should be is called the aberration of the mirror. Mechanical accuators are placed on the back side of the mirrors that can posi-

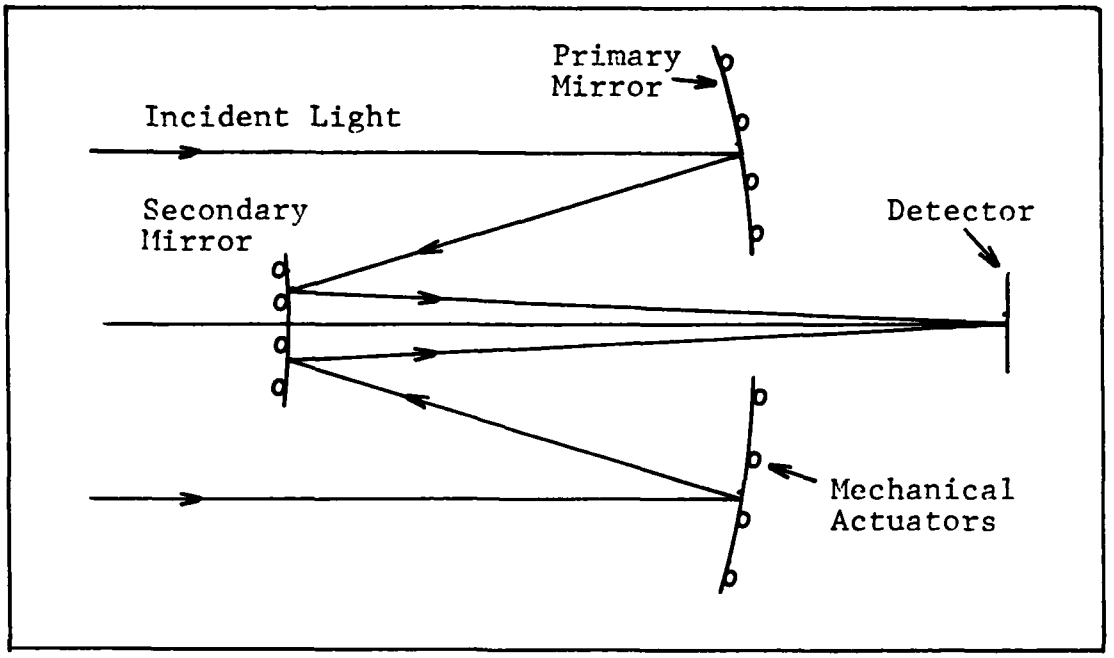


Fig 1-1. Cassegrain Telescope

tion the mirror surfaces to within a fraction of a wavelength. The deconvolution problem may be stated as follows: Given a known input field and measurements of the amplitude and phase at the plane of the detector, can we determine the aberrations on each mirror so that we can correct them?

Figure 1-2 is a block diagram of an optical system such as the telescope of Figure 1-1. The input signal $U_1(x)$ is multiplied by aberration functions $e^{jkW_A(x)}$ and $e^{jkW_B(x)}$, convolved with impulse response functions $h_1(x)$ and $h_2(x)$, and corrupted by additive noise $\hat{N}(x)$. A typical problem from detection and estimation theory is to detect or estimate $U_1(x)$ given $W_A(x)$, $W_B(x)$, $h_1(x)$, $h_2(x)$,

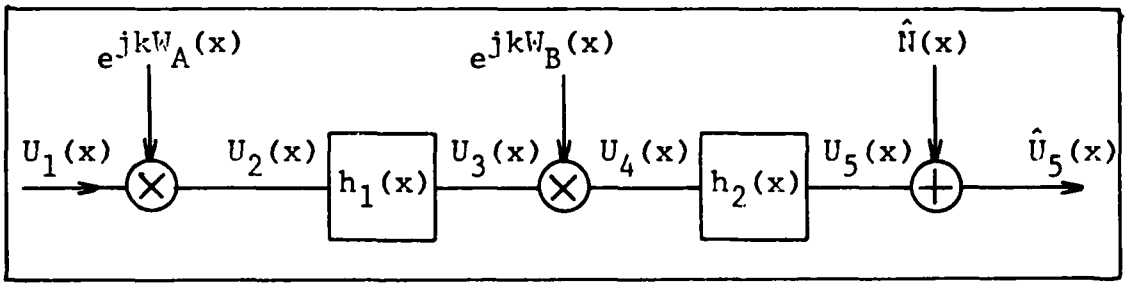


Fig 1-2. Optical System Block Diagram

measurements of $\hat{U}_5(x)$, and statistical information about $\hat{N}(x)$. Detection and estimation problems have been studied extensively (Ref 23). The problem addressed in this paper, which has been studied very little, is the same as the problem stated above except that $W_A(x)$ and $W_B(x)$ are the unknowns and $U_1(x)$ is known. While the deconvolution problem as defined above could arise in other types of applications, the scope of this paper will be limited to optical systems. Limitations and assumptions from optics will be used to define the problem.

The major problem, then, is to determine the exact location of the optical surfaces without adding a lot of additional complex sensors to the telescope. One would like to be able to deduce the location of the mirror surfaces from intensity measurements at the image plane of the telescope. This implies being able to solve two subproblems: first, the phase at some plane in the output of the telescope, hereafter called the measurement plane, must be deduced from intensity measurements because the phase and amplitude of an optical frequency field cannot be measured directly. Second, given a known input field and

the amplitude and phase at the measurement plane, some way must be found to determine the aberrations on the optical surfaces. This second problem is the deconvolution problem.

When a propagating field which is described in the spatial domain encounters an optical element such as a lens, the function representing the field is multiplied by the lens function. When a spatial field propagates, the function representing the field is convolved with the propagation impulse response function as mentioned previously. It would be just as valid to describe a field in the spatial frequency domain (the Fourier transform of the spatial domain) in which case the function representing the field would be convolved with the Fourier transformed lens function and multiplied by the Fourier transformed propagation impulse response function. Throughout this paper, the optical field is described in the spatial domain. Also, scalar field theory is assumed to adequately describe the fields.

In this paper, the deconvolution problem is divorced from any specific optical system such as HALO. Instead, the simplest imaginable system is studied; namely, an optical system consisting of two planar aberration functions or "sheets of glass" followed by a measurement plane. The input field is assumed to be a plane wave arriving at a known angle such as would be the case if a telescope were

imaging a star. The optical system is described in Chapter 2. Despite the fact that the system under consideration is very simple, the extension of the solution to more complex systems is straightforward.

A one-dimensional optical system is treated throughout this paper except in Chapter 5 where the 1-D results are extended to two dimensions. This is for convenience, since the 2-D equations are much longer and more cumbersome than the 1-D equations. As shown in Chapter 5, there is no difficulty in doing this because the extension to two dimensions is straightforward, and the 2-D equations parallel the 1-D equations very closely.

Previous Research

Apparently the deconvolution problem as posed in this paper had never been addressed in the open literature before the HALO effort was begun. During 1976-1977, RADC contracted with the Perkin-Elmer Corporation to develop a deconvolution algorithm. The algorithm Perkin-Elmer produced failed to consider the effects of propagation between optical elements, so it was of little practical value either to solve the deconvolution problem or as a starting point for this research (Ref 17).

Approach

Three approaches were considered in addressing the deconvolution problem. The first was to develop a ray tracing algorithm to deconvolve the aberrations. Such an

algorithm was begun, but it was abandoned in favor of a more mathematically sound approach. The ray tracing approach appears to have the advantage of not generating large systems of equations to be solved. However, it appears to be inaccurate. It might be useful for indicating which direction to move an optical surface to correct an aberration, but not for accurately determining the aberration. The ray tracing algorithm appears to have enough merit to warrant inclusion in this paper as an appendix (Appendix C). The information there may serve as a starting point for future research.

Another approach is to develop a parameter search algorithm. In this approach, the unknown aberration functions would be represented by a set of basis functions with unknown coefficients. An algorithm would try a number of combinations of coefficients and select the combination that yields a calculated system output field which most nearly matches the measured output (Refs 5, 8, 18). A parameter search is a brute force approach often requiring considerable computer time. The number of basis functions needed to describe the aberration could be very large (Ref 6). For these reasons, the parameter search approach was not pursued as a solution to the deconvolution problem.

The third approach, and the one presented in this paper, is to use scalar diffraction theory to write the measured output field of the optical system as a function

of the known input field and the unknown aberration functions, represent the aberration functions by a basis with unknown coefficients, and then try to solve for the coefficients. Specifically, in Chapter 2 the system represented by Figure 1-2 is presented as an optical system with measurable output field $U_5(x)$, known input field $U_1(x)$, and known system impulse response functions $h_1(x)$ and $h_2(x)$. Initially, noise $N(x)$ is assumed to be zero. The aberration functions $W_A(x)$ and $W_B(x)$ are unknown. Appropriate limitations and assumptions from optics are used in defining the problem. The output $U_5(x)$ is then written as a function of the input $U_1(x)$ and the known and unknown system parameters. The resulting equation is a Fredholm equation of the first kind, the solution of which is often an ill-posed problem. However, because the Fredholm integral can be completed by the method of stationary phase, the ill-posed nature of the Fredholm equation causes no difficulty.

The unknown aberration functions are represented by Fourier series with unknown coefficients and, with appropriate precautions, the problem to be solved is represented by a system of bilinear equations.

Chapter 3 draws heavily on estimation theory as presented in Ref 23. Noise with known statistics is added to the system and bounds are established on the variance of the Fourier coefficients used in representing the aberrations. This chapter accomplishes two things: first, it

gives some assurance that the solution to the deconvolution problem is not overly sensitive to input noise under the appropriate conditions; and second, it shows that the variance in the solution is a function of the arrival angles of the input fields ($U_1(x)$). An equation is developed which enables one to select suitable input plane wave arrival angles.

In Chapter 4, a unique approach is developed to solve the bilinear system of equations from Chapter 2. The approach makes use of the fact that the least-square-error solution to the bilinear system of equations is also the maximum likelihood estimate. This guarantees that when the measurements $\hat{U}_5(x)$ are corrupted by Gaussian noise, the least-square-error solution of the bilinear equations is also most likely to be the correct solution. In other words, estimation theory proves that there is no solution with a higher probability of being the correct solution.

Chapter 5 presents the two-dimensional forms of the key one-dimensional equations developed in Chapter 2. The results are straightforward, there being nothing unusual or unexpected in the extension. A simple 2-D problem is presented in the last part of Chapter 5 and the deconvolution algorithm is applied to it to yield the solution. The purpose of this exercise is to show that the algorithm developed in Chapter 4 does in fact work. The sample problem is unrealistically simple because of the excessive

amount of computer time and memory required to solve more typical problems, but it still demonstrates that the algorithm works in principle.

II. One-Dimensional Two-Surface Deconvolution Problem

Overview

In this chapter, limiting assumptions will be stated and discussed, the problem to be solved will be defined mathematically, and the equations representing the problem will be developed into a system of bilinear equations that can be programmed on a computer to get an approximate solution. The problem will be shown to be nonlinear and ill-posed, and the approach to handling these difficulties will be discussed. In the development of a solution to the problem, detours will occasionally be taken to examine such issues as the uniqueness of the proposed solution, the dimensionality of the system of equations that must be solved, and the effects of edge diffraction and vignetting. Throughout this chapter, only the noiseless, one-dimensional problem will be addressed. The effects of measurement noise will be addressed in Chapter 3. In Chapter 5, the more realistic two-dimensional problem will be shown to be a straightforward extension of the one-dimensional problem.

Statement of Problem and Assumptions

Figure 2-1 illustrates the optical system to be considered in this chapter. It consists of two planar aberration functions, $W_A(x)$ and $W_B(x)$, which are parallel to each other and to a measurement plane located some distance to the right of $W_B(x)$. Light propagates from left

to right through the system and is measured at the measurement plane.

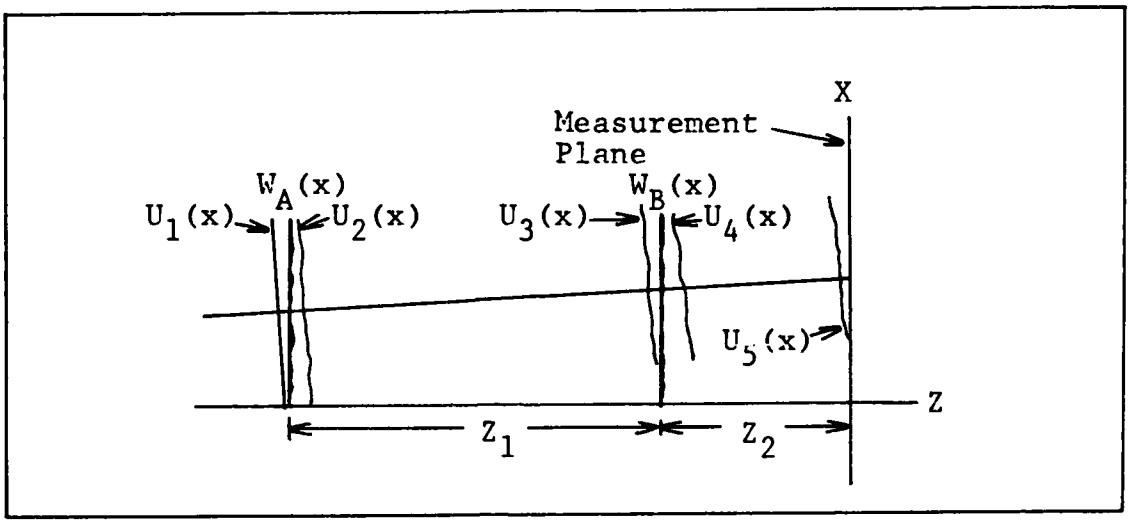


Fig 2-1. One-Dimensional Deconvolution Problem

The following definitions and assumptions pertain to the problem:

1. Scalar diffraction theory may be used to describe the propagation of light in the system. Scalar diffraction theory accurately describes light propagation when the apertures in the system are large compared to the wavelength of the light, when measurements are made at least several wavelengths from the aperture, and when polarization effects are not important. These conditions are all met in this problem.

2. $U_1(x)$ through $U_5(x)$ are complex quantities representing the amplitude and phase of the scalar field at a given plane perpendicular to the Z axis. $U_1(x)$, $W_A(x)$, and

$U_2(x)$ are all considered to be at plane A, and $U_3(x)$, $W_B(x)$, and $U_4(x)$ are all considered to be at plane B. This condition will be met as long as the distance between the true aberration surface and the plane (A or B) is much less than the distance between the various elements of the optical system. This assumption allows us to consider Z_1 and Z_2 as constants. If Z_1 and Z_2 were functions of x, the mathematical description of the system would be much more complex. Note that this approximation is similar to the "thin lens" approximation of geometrical optics.

3. $U_1(x)$ is a known input field. The form of the input field is unimportant except for mathematical simplicity.

4. $W_A(x)$ and $W_B(x)$ have units of length and represent the distance that light passing through the aberration is delayed compared to light propagating in free space. Also, $W_A(x)$ and $W_B(x)$ are continuous with finite first derivatives. This assumption insures that $e^{jkW_A(x)}$ and $e^{jkW_B(x)}$ can be represented by a Fourier series (Eqs (2-21) and (2-22)). For example, if $W_B(x)$ was a step function with step size $n\lambda$, then $e^{jkW_B(x)}$ would have the same value on both sides of the step and would be undefined at the step. A Fourier series could not represent such a function. The aberration functions are assumed to be continuous with finite first derivatives to avoid such complications. In practical optical systems with segmented optics, step

functions do occur. As long as the step size is known to be less than one wavelength, the step should cause no difficulty. Otherwise, there would be an $n\lambda$ ambiguity in the Fourier series representation of the aberration function.

5. The amplitude and phase of $U_5(x)$ can be determined at an arbitrary number of points along the x-axis. This is by no means a trivial assumption. At optical frequencies, detectors only measure intensity, so phase must either be measured interferometrically or it must be deduced from intensity measurements. Others are actively pursuing the problem of deducing phase from intensity measurements (Refs 7, 9, 10, 20) so that problem has not been addressed, but it has been assumed solved in order to narrow the scope of the deconvolution problem.

6. The light will be treated as monochromatic. Actually the light could be quasimonochromatic. The definition of quasimonochromatic is that $\Delta\lambda/\bar{\lambda} << 1$, where $\Delta\lambda$ is the wavelength spread and $\bar{\lambda}$ is the average wavelength. All that is required in the deconvolution problem is that the amplitude and relative phase of the scalar field at the measurement plane be defined at a given wavelength. Two input fields that differ only slightly in wavelength (so that the quasimonochromatic condition is met) will not result in fields at the measurement plane which differ appreciably in relative phase and amplitude.

Therefore, the input field does not need to be as nearly monochromatic as it would for such applications as coherent detection. For ease of notation, λ will be used in this paper, but $\bar{\lambda}$ could be used just as well.

7. The highest spatial frequencies present in either the aberrations or fields are small compared to the inverse of the wavelength of the light. In other words, $f_x < 1/\lambda$, where f_x is spatial frequency in cycles per unit length and λ is wavelength. This assumption is necessary to simplify the propagation transfer function (Ref 11:54)

$$H(f_x) = \begin{cases} \exp\left(j\frac{2\pi Z}{\lambda}\sqrt{1-(\lambda f_x)^2}\right), & f_x < \frac{1}{\lambda} \\ 0 \text{ else} \end{cases} \qquad (2\text{-}1)$$

Since $(\lambda f_x)^2 << 1$, $\sqrt{1-(\lambda f_x)^2} \cong 1-\frac{(\lambda f_x)^2}{2}$, so

$$H(f_x) \cong e^{jkz}\, e^{-j\pi\lambda z f_x^2} \qquad (2\text{-}2)$$

where $k = 2\pi/\lambda$.

System Transfer Function

To describe $U_5(x)$ in terms of $U_1(x)$, $W_A(x)$, and $W_B(x)$, the inverse Fourier transform of H(fx) is needed which can be shown to be

$$h(x) = \frac{e^{-j\frac{\pi}{4}}\, e^{jkz}\, e^{\frac{j\pi x^2}{\lambda Z}}}{\sqrt{\lambda Z}} \qquad (2\text{-}3)$$

Each field in Figure 2-1 can be written in terms of the field to the left of it as follows:

$$U_5(x) = U_4(x) * h_2(x) \tag{2-4}$$

$$U_4(x) = U_3(x)e^{jkW_B(x)}P_B(x) \tag{2-5}$$

$$U_3(x) = U_2(x) * h_1(x) \tag{2-6}$$

$$U_2(x) = U_1(x)e^{jkW_A(x)}P_A(x) \tag{2-7}$$

where $P_A(x)$ and $P_B(x)$ are the pupil functions at planes A and B, respectively, and where "*" is the convolution operator. Combining Eqs (2-4) through (2-7),

$$U_5(x) = \left[[(U_1(x)e^{jkW_A(x)}P_A(x)*h_1(x))]e^{jkW_B(x)}P_B(x)\right]* h_2(x) \tag{2-8}$$

or

$$U_5(x) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_1(\gamma)P_A(\gamma)P_B(\beta)h_1(\beta-\gamma)h_2(x-\beta) e^{jk(W_A(\gamma)+W_B(\beta))}d\gamma d\beta. \tag{2-9}$$

If it is assumed that $U_1(x)$ is a plane wave with amplitude A' propagating at angle θ relative to the x-axis, Eq (2-3) for h_1 and h_2 is substituted in Eq (2-9), and the pupil functions are combined with the limits of integration, then

$$U_5(x,\theta) = \frac{-jA'e^{jk(Z_1+Z_2)}}{\lambda\sqrt{Z_1Z_2}} \int_{\sum P_A}\int_{\sum P_B} \exp\Big[jk(\gamma\sin\theta+$$

$$\frac{(\beta-\gamma)^2}{2Z_1} + \frac{(x-\beta)^2}{2Z_2})\Big] \exp\Big[jk(W_A(\gamma)+W_B(\beta))\Big] d\gamma d\beta \qquad (2\text{-}10)$$

The first exponential term in the above integral is a known kernel, and the second exponential contains the unknown aberration functions. Even though the optical system of Figure 2-1 is very simple, an approach to solving for the aberrations in that system can be easily applied to much more general systems. For example, if the optical elements at planes A and B were lenses with aberrations instead of planar surfaces with aberrations, there would be additional known phase terms inside the integral of the form $e^{\frac{\pm jkx^2}{2fl}}$ where "fl" is the focal length of the lens. These terms could be combined with the known kernel and would not complicate or change the problem. The same can be said for the known input field, and for more general optical elements than lenses.

Completing the Integral of Equation (2-10)

Equation (2-10) is a Fredholm equation of the first kind (Ref 15:905), the solution of which is usually an ill-posed problem (Ref 14, 21, and 22). In the case of Eq (2-10) however, the integrand contributes to the integral at only a few discrete points (usually only one point) so the integral may be approximated by means of an asympto-

tic expansion, and the usual ill-posedness of Fredholm equations of the first kind is not a problem.

The method for approximating the integral in Eq (2-10) is called the method of stationary phase. See Ref 2:747-754 for a complete discussion of this method. Basically, the integral such as the one in Eq (2-10) is approximated by an asymptotic expansion. When the constant $k=2\pi/\lambda$ is large, only the first term of the expansion is significant. The integrand may be approximated by the first term and the integral completed. When the limits of integration are $\pm\infty$, contributions to the integral come only from critical points of the first kind (stationary points) defined by

$$\frac{\partial f(\gamma,\beta)}{\partial \gamma} = \frac{\partial f(\gamma,\beta)}{\partial \beta} = 0$$

where

$$f(\gamma,\beta) = \gamma\sin\theta + \frac{(\beta-\gamma)^2}{2Z_1} + \frac{(x-\beta)^2}{2Z_2} \qquad (2\text{-}11)$$

As explained near the end of this chapter in the section "Dimensionality of the $\underline{A}$ Matrix", the aberration functions $W_A(\gamma)$ and $W_B(\beta)$ are considered to be on the order of a few wavelengths so that $e^{jkW_A(\gamma)}$ and $e^{jkW_B(\beta)}$ are slowly varying functions of γ and β as compared to $e^{jkf(\gamma,\beta)}$. Therefore, the locations of the stationary points are determined by $f(\gamma,\beta)$ as follows:

$$\frac{\partial f(\gamma,\beta)}{\partial \gamma} = \sin\theta - \frac{\beta-\gamma}{Z_1} = 0 \tag{2-12}$$

$$\frac{\partial f(\gamma,\beta)}{\partial \beta} = \frac{\beta-\gamma}{Z_1} - \frac{x-\beta}{Z_2} = 0 \tag{2-13}$$

Solving for γ and β in Eqs (2-12) and (2-13) yields the coordinates of the critical points of the first kind, γ_0 and β_0.

$$\gamma_0 = x-(Z_1+Z_2)\sin\theta \tag{2-14}$$

$$\beta_0 = x-Z_2\sin\theta \tag{2-15}$$

Substituting Eqs (2-14) and (2-15) into (2-11) gives

$$f(\gamma_0,\beta_0) = x\sin\theta - (Z_1+Z_2)\frac{\sin^2\theta}{2} \tag{2-16}$$

The following second partial derivatives are also needed to evaluate the asymptotic expansion:

$$\frac{\partial^2 f(\gamma,\beta)}{\partial \gamma^2} = \alpha' = \frac{1}{Z_1} \qquad \frac{\partial^2 f(\gamma,\beta)}{\partial \beta^2} = \beta' = \frac{1}{Z_1} + \frac{1}{Z_2}$$

$$\frac{\partial^2 f(\gamma,\beta)}{\partial \gamma\, \partial \beta} = \gamma' = -\frac{1}{Z_1} \tag{2-17}$$

From Born and Wolf (Ref 2:754), the results of a two-dimensional asymptotic expansion may be stated as follows:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\gamma,\beta)e^{jkf(\gamma,\beta)}d\gamma d\beta = \frac{2\pi j\sigma}{k\sqrt{|\alpha' \ \beta' \ -\gamma'|^2}}$$

$$g(\gamma_0,\beta_0)e^{jkf(\gamma_0,\beta_0)} \qquad (2\text{-}18)$$

where $\sigma = +1$ for $\alpha'\beta' > \gamma'^2$ and $\alpha' > 0$ as it does in this problem. Substituting Eqs (2-10), (2-16), and (2-17) into Eq (2-18), and noting that $g(\gamma,\beta) = e^{jk(W_A(\gamma)+W_A(\beta))}$,

$$U_5(x,\theta) = \frac{-jA'e^{jk(Z_1+Z_2)}}{\lambda\sqrt{Z_1Z_2}} \frac{2\pi j\sqrt{Z_1Z_2}}{k} e^{jkx\sin\theta}$$

$$e^{-jk(Z_1+Z_2)(\frac{\sin^2\theta}{2})} e^{jkW_A(x-(Z_1+Z_2)\sin\theta)}e^{jkW_B(x-Z_2\sin\theta)}$$

$$= A'e^{jk(Z_1+Z_2)}e^{-j\frac{k}{2}(\sin^2\theta)(Z_1+Z_2)}e^{jkx\sin\theta}$$

$$e^{jkW_A(x-(Z_1+Z_2)\sin\theta)}e^{jkW_B(x-Z_2\sin\theta)} \qquad (2\text{-}19)$$

Since the first three phase terms in Eq (2-19) are all known, it is easy to find values of $U(x,\theta)$ from observations, where

$$U(x,\theta) = \frac{U_5(x,\theta)}{A'}e^{-jk(Z_1+Z_2)}e^{-jkx\sin\theta} e^{jk(Z_1+Z_2)(\frac{\sin^2\theta}{2})}$$

$$= e^{jkW_A(x-(Z_1+Z_2)\sin\theta)}e^{jkW_B(x-Z_2\sin\theta)} \qquad (2\text{-}20)$$

An examination of Eq (2-20) reveals that if $U_5(x,\theta)$ can be measured as θ is continuously varied, it would be

possible to determine the aberration functions exactly to within a complex constant. For example, let $x-Z_2\sin\theta = x_0$, where x_0 is an arbitrary constant. As θ is varied, $e^{jkW_B(x-Z_2\sin\theta)} = e^{jkW_B(x_0)}$ will be an unknown constant while $e^{jkW_A(\gamma)}$ varies as $\gamma = Z_1\sin\theta+x_0$. Therefore, $e^{jkW_A(\gamma)}$, and consequently $W_A(\gamma)$, can be determined exactly over some range to within an unknown constant. Likewise, if $x-(Z_1+Z_2)\sin\theta=x_0$, then as θ is varied, $e^{jkW_A(x_0)}$ will be an unknown constant while $e^{jkW_B(\beta)}$ varies as $\beta=Z_1\sin\theta+x_0$, and $e^{jkW_B(\beta)}$ can be determined over some range to within a constant. The unknown constant is unimportant, because it is only necessary to know the relative magnitude of the aberrations on the mirrors. It isn't necessary to know the absolute position of the mirror surfaces to correct the aberrations.

The method of determining the aberrations described above has an obvious physical interpretation illustrated by Figure 2-2 below. This is the same optical system shown in Figure 2-1. Let $a=b=1$, $Z_1=2$, $Z_2=1$, and $x_0=.5$. Suppose an incident plane wave propagates in the direction of ray 1 in Figure 2-2, and suppose an observer observes the plane wave at the point where ray 1 intersects the measurement plane. The observer is looking through $W_B(.5)$ at $W_A(0)$, and the angle of propagation is about 14°. Now suppose the angle of propagation varies continuously from 14° to -14° (from ray 1 to ray 2) while the observer at the measurement

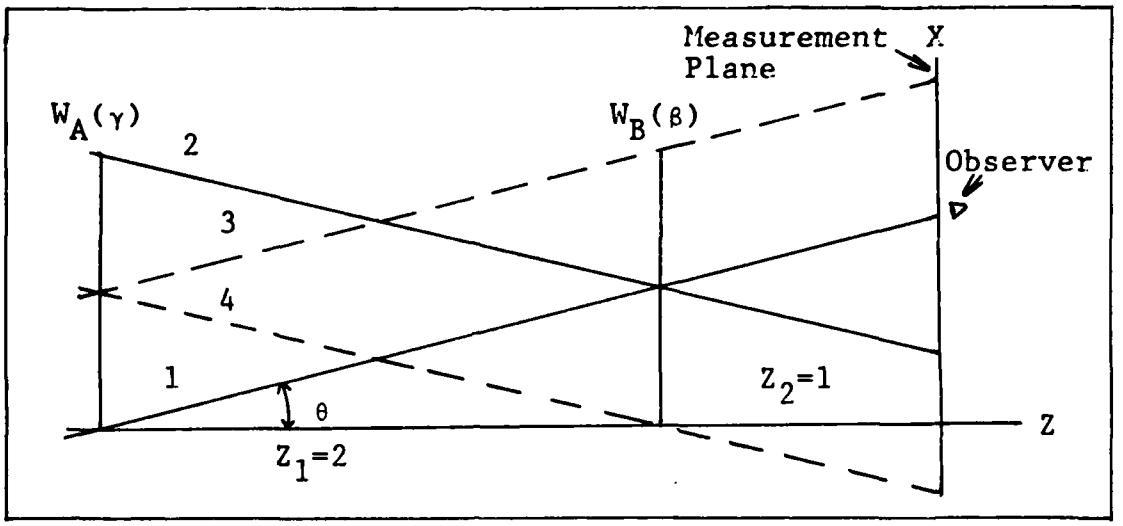


Fig 2-2. Deconvolution With Variable Θ

plane moves in the negative x direction so that he always looks through $W_B(.5)$ to observe the plane wave. Variations in the observed plane wave are due solely to $W_A(\gamma)$, so that $W_A(\gamma)$ can be determined exactly except for a constant phase term which arises because $W_B(.5)$ is not known. Likewise, rays 3 and 4 (dashed lines) show the beginning and ending line of sight for an observer looking through the same point on aberration $W_A(\gamma)$ as the angle of the incident plane wave is varied from +14° to -14°. This time, all observed variations in the plane wave are due to $W_B(\beta)$, so it can be determined exactly except for a constant phase term.

The preceding discussion establishes that in the asymptotic limit of large k, the deconvolution problem can be solved quite easily and directly. Unfortunately, the assumption that measurements of $U_5(x,\theta)$ may be made at

arbitrary θ is generally unrealistic. In most optical systems of interest, measurements of $U_5(x,\theta)$ can only be made at discrete points in the measurement plane, and only for a few discrete plane wave arrival angles. The deconvolution problem to be addressed in the remainder of this paper may be restated as follows: Given measurements of $U_5(x,\theta)$ for discrete points in the measurement plane and for a few discrete plane wave arrival angles, approximate the aberration functions.

The ray tracing deconvolution algorithm documented in Appendix B is similar to the deconvolution method described above except that θ is not continuously variable. The aberrations can only be found at discrete points and their value must be inferred between those points. The ray tracing algorithm was abandoned in favor of the present more mathematically sound approach.

That approach is to approximate the aberration functions by a suitable set of basis functions with unknown coefficients and solve for the coefficients. A finite subset of the basis functions must be selected based on some knowledge of the functions being represented because the mathematics alone may not indicate the best subset. The set of approximating functions should span the function space of the unknown function, and it should closely approximate the unknown function, in some norm, with a minimum number of terms. For example, if the aberrations

are defined over a circular aperture, the circle polynomials of Zernike might closely approximate the aberration functions with a minimum number of unknown coefficients (Ref 2:464). Because the development of the solution to the deconvolution problem in this paper is not geared to any specific system, a Fourier series with a finite number of terms will be used to represent the aberration functions. Fourier series have desirable properties such as orthogonality that make them easy to use in an analysis such as this. A procedure for determining the number of terms to be used in the Fourier series will be discussed later. Let

$$e^{jkW_A(x)} = \sum_m F_m e^{\frac{j2\pi mx}{a}} \tag{2-21}$$

and

$$e^{jkW_B(x)} = \sum_n G_n e^{\frac{j2\pi nx}{b}} \tag{2-22}$$

where F_m and G_n are the unknown coefficients and "a" and "b" are the pupil dimensions corresponding to $W_A(x)$ and $W_B(x)$, respectively. If Eqs (2-21) and (2-22) are substituted into (2-18), where now

$$g(\gamma,\beta) = \sum_m \sum_n F_m G_n e^{\frac{j2\pi m\gamma}{a}} e^{\frac{j2\pi n\beta}{b}}$$

and the constant phase terms are combined with $U_5(x)$ as was done in Eq (2-20), then

$$U(x) = \sum_m \sum_n F_m G_n e^{\frac{j2\pi m}{a}(x-(Z_1+Z_2)\sin\theta)} e^{\frac{j2\pi n}{b}(x-Z_2\sin\theta)} \tag{2-23}$$

Note that U(x) is no longer shown as a function of θ since measurements are only made at a few discrete values of θ.

It is reasonable to let the apertures P_A and P_B in Figure 2-1 be equal in size since a subaperture can always be defined in one of the two apertures such that $a=b$. Vignetting would cause only part of apertures A and B to be used at any plane wave arrival angle except $\theta = 0$, so the aberrations generally can be found only on subapertures. Applying these conditions to Eq (2-23),

$$U(x) = \sum_m \sum_n F_m G_n \exp\left[j\frac{2\pi}{a}[x(m+n)-m\sin\theta(Z_1+Z_2)-nZ_2\sin\theta)]\right] \tag{2-24}$$

Bounding the Fourier Series

Note that Eq (2-24) is a bilinear equation. If the summations are bounded such that $-M \le m \le M$ and $-N \le n \le N$, and we let

$$V(m,n) = \exp\left[j\frac{2\pi}{a}[x(m+n)-m\sin\theta(Z_1+Z_2)-nZ_2\sin\theta]\right],$$

then Eq (2-24) can be written as

$$U(x) = \left[F_{-M}, F_{-M+1}, \cdots F_{M-1}, F_M\right] \left[V_{m,n}\right] \begin{bmatrix} G_{-N} \\ G_{-N+1} \\ \vdots \\ G_{N-1} \\ G_N \end{bmatrix} \quad (2\text{-}25)$$

The equation contains 2M+2N+2 unknowns, so at least that many equations must be generated in order to solve for the Fourier coefficients. However, the bounds M and N must first be found.

In Figure 2-1, $U_5(x)$ is nonzero only over an aperture of width "a" (neglecting diffraction at the edge). Even though $U_5(x)$ (and therefore U(x)) is bounded by the aperture, it can be represented by a Fourier series if the assumption is made that U(x) is one period of a periodic function (Ref 16:99). The resulting Fourier series representation of the function will be unique. Let

$$U(x) = \sum_{p=-\infty}^{\infty} H(p) e^{\frac{j2\pi xp}{a}} \quad (2\text{-}26)$$

where

$$H(p) = \frac{1}{a}\int_0^a U(x)e^{\frac{-j2\pi xp}{a}}dx = \frac{1}{a}\int_0^a \sum\sum_{m\ n} F_m G_n$$

$$e^{\frac{j2\pi x(m+n)}{a}} e^{\frac{-j2\pi xp}{a}} \exp\left[-j\frac{2\pi}{a}[m\sin\theta(Z_1+Z_2) + Z_2 n\sin\theta]\right] dx$$

$$= \sum_{\substack{m \\ m+n=p}} \sum_{n} F_m G_n \exp\left[-j\frac{2\pi}{a}[m\sin\theta(Z_1+Z_2) + Z_2 n\sin\theta]\right] \qquad (2\text{-}27)$$

If U(x) is known only at L equally spaced points rather than on a continuum (see assumption 5, page 12), the Fourier coefficients H(p) can be found using a discrete Fourier transform (DFT) as follows:

$$H(p) = \frac{1}{L}\sum_{\ell=0}^{L-1} U((\ell+1)\Delta x)e^{\frac{-j2\pi\ell p}{L}} \qquad (2\text{-}28)$$

where $\Delta x = \frac{a}{L}$. Of course, there must be enough samples of $U_5(x)$ to meet the Nyquist sampling criteria, which requires that the number of samples of a function be greater than twice the highest spatial frequency of the function. Failure to meet the Nyquist criteria causes aliasing in the frequency domain description of the function, so that the function has not been uniquely and completely described (Ref 16:29).

Equation (2-27) by itself indicates that for a U(x) with finite spatial frequency content (p finite), the magnitude of m and n may grow large without bound as long as $m+n=p$. The measurements taken with only one input

plane wave arrival angle cannot be used to mathematically establish bounds on m and n because the effect of a high spatial frequency aberration at plane A can always be cancelled at the measurement plane by a properly phased aberration of the same frequency at plane B. But when measurements are taken at more than one θ, the presence of frequency components in the aberrations which are of higher order than the order of the frequency components in $U_5(x)$ will generally be revealed. However, there is still a possibility that for a given difference in plane wave arrival angles, higher spatial frequency components of the aberration functions may not be observed at the measurement plane.

As an example, suppose that measurements of $U_5(x)$ are taken at equally spaced points across an aperture "a" in the measurement plane, and that the measurements are taken at two different arrival angles, $\theta_1 = 0$ and $\theta_2 = \arcsin(a/4Z_1)$. Further suppose that all $U(x) = 1+j0$ in both sets of measurements. Examining Eq (2-24), if

$$F_m=G_n=1 \begin{cases} m=n=0 \\ \\ =0 \end{cases} \begin{cases} \\ \\ \text{else,} \end{cases}$$

then all $U(x) = 1+j0$. But if

$$F_m=G_n=1 \begin{cases} m=4,\ n=-4 \\ \\ =0 \end{cases} \begin{cases} \\ \\ \text{else,} \end{cases}$$

then $U(x) = 1+j0$ also. In the first case, $W_A(x)=W_B(x)=0$, but in the second case, $W_A(x) = \frac{4\lambda x}{a}$ and $W_B(x) = \frac{-4\lambda x}{a}$. The second set of aberrations are optical wedges that exactly compensate for each other, and the difference between θ_1 and θ_2 is such that the optical path lengths differ by exactly one wavelength between the measurements. This is illustrated in Figure 2-3 below. Note that the path length difference discussed here is not the one resulting from the term $e^{jk(Z_1+Z_2)\frac{\sin^2\theta}{2}}$ in Eq (2-19) since that path length difference was accounted for in converting from $U_5(x)$ to $U(x)$.

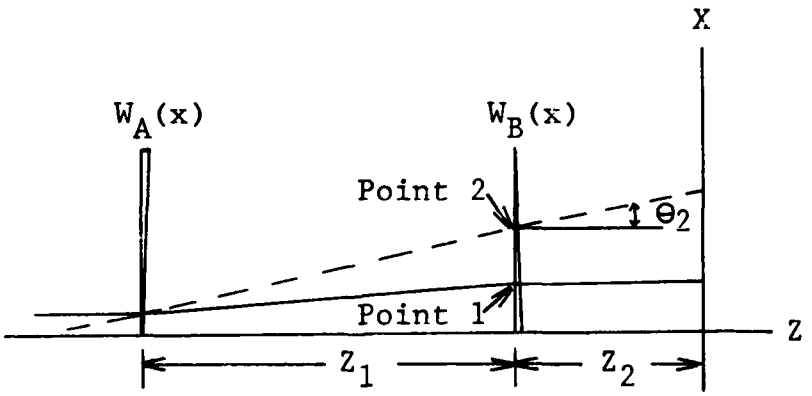


Fig 2-3. Example of a Nonunique Deconvolution Problem

The path length difference is the one resulting from a ray passing through point 2 versus point 1 in the wedge in Figure 2-3. If that difference is a multiple of one wavelength, then it is impossible to distinguish between

the system of Figure 2-3 and a system with zero aberrations. However, if U(x) is measured for a third arrival angle θ_3 such that the path length difference is not a multiple of one wavelength, then it is possible to distinguish between the system of Figure 2-3 and an aberration free system. If the angle of the incident plane wave can be controlled, angles can always be chosen such that the aberrations may be uniquely determined. But if m=p-n increases without bound, there will be an infinite number of angles such that the aberrations cannot be uniquely determined. Therefore, m and n must be limited by making the very reasonable assumption that the highest spatial frequency of either $e^{jkW_A(x)}$ or $e^{jkW_B(x)}$ is less than or equal to the highest spatial frequency of U(x). As long as this assumption holds, the spatial frequency content of U(x) can be determined for a measurement with $\theta = \theta_1$, m and n can be bounded, an optimum θ_2 can be found, and a second measurement of $U_5(x)$ can be made for $\theta = \theta_2$. This yields the same number of equations as unknowns. The procedure for selecting an optimum θ_2 is discussed in Chapter 3.

Choice of a Two-Angle Measurement Scheme

Note that in the preceding paragraph, two independent sets of measurements of $U_5(x)$ were made with two separate input plane waves arriving at different angles to provide as many equations as unknowns. If the spatial frequency content of U(x) is such that the number of Fourier coeffi-

cients H(p) is bounded by $-Pmax \leq p \leq Pmax$, then measuring U(x) at one angle yields 2Pmax+1 equations (Eq (2-27)), but since $-Pmax \leq m(\text{or } n) \leq Pmax$, there are 4Pmax+2 unknowns. That is why at least two independent measurements of $U_5(x)$ are required. The reason for varying the plane wave arrival angle rather than one of the other parameters is now justified.

The parameter that is varied must result in a second set of equations (Eq (2-27)) that are independent of the first set. The parameters that are candidates to be varied between measurements are θ, Z_1, and Z_2. Let

$$a_{sub}(m,n) = e^{-j\frac{2\pi}{a}[m\sin\theta(Z_1+Z_2)+n\sin\theta Z_2]} \qquad (2\text{-}29)$$

where a subscript of 1 on a(m,n) and H(p) represents the first set of measurements, a subscript of 2 represents the second set, etc. Then Eq (2-27) can be written in matrix form as follows:

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} F_{-Pmax}G_0 \\ F_{-Pmax}G_1 \\ \vdots \\ F_{-Pmax}G_{Pmax} \\ F_{-Pmax+1}G_{-1} \\ F_{-Pmax+1}G_0 \\ \vdots \\ F_{-Pmax+1}G_{Pmax} \\ \vdots \\ F_{Pmax}G_{-Pmax} \\ \vdots \\ F_{Pmax}G_0 \end{bmatrix} = \begin{bmatrix} H_1(-Pmax) \\ H_1(-Pmax+1) \\ \vdots \\ H_1(Pmax) \\ H_2(-Pmax) \\ \vdots \\ H_2(Pmax) \end{bmatrix} \qquad (2\text{-}30)$$

where the m, n, and subscript of the elements in the $\underline{A}$ matrix must correspond to the appropriate m, n, and subscript of the FG and H vectors. The only nonzero elements of $\underline{A}$ are those where $m+n=p$, so $\underline{A}$ is sparse. An example will better illustrate the structure of Eq (2-30). Let $Pmax=1$. If two indepedent measurements of $U_5(x)$ are made, then Eq (2-30) would be

$$\begin{bmatrix} a_1(-1,0) & 0 & a_1(0,-1) & 0 & 0 & 0 & 0 \\ 0 & a_1(-1,1) & 0 & a_1(0,0) & 0 & a_1(1,-1) & 0 \\ 0 & 0 & 0 & 0 & a_1(0,1) & 0 & a_1(1,0) \\ a_2(-1,0) & 0 & a_2(0,-1) & 0 & 0 & 0 & 0 \\ 0 & a_2(-1,1) & 0 & a_2(0,0) & 0 & a_2(1,-1) & 0 \\ 0 & 0 & 0 & 0 & a_2(0,1) & 0 & a_2(1,0) \end{bmatrix}$$

$$\begin{bmatrix} F_{-1}G_0 \\ F_{-1}G_1 \\ F_0G_{-1} \\ F_0G_0 \\ F_0G_1 \\ F_1G_{-1} \\ F_1G_0 \end{bmatrix} = \begin{bmatrix} H_1(-1) \\ H_1(0) \\ H_1(1) \\ H_2(-1) \\ H_2(0) \\ H_2(1) \end{bmatrix} \qquad (2\text{-}31)$$

The only rows that could be linear combinations of each other are the rows corresponding to the same "p" (rows 1 and 4, 2 and 5, etc.). The elements of $\underline{A}$ are given by Eq (2-29). If Z_1 is varied between measurements and $Z_1=Z_1'$ in the first measurement and $Z_1=Z_1''$ in the second measurement, then

$$a_2(m,n) = a_1(m,n)\, e^{\frac{-j2\pi}{a}m\sin\theta(Z_1''-Z_1')} \qquad (2\text{-}32)$$

Since m is always different for the different elements in the same row, no two row vectors corresponding to the same "p" will be linearly related, and again the $\underline{A}$ matrix will

be full rank. However, there probably are not many optical systems where deconvolution is of interest and one has the option of varying the distance between the optical elements.

If Z_2 is the parameter that is varied between measurements and $Z_2=Z_2'$ in the first measurement and $Z_2=Z_2''$ in the second measurement, then

$$a_2(m,n) = a_1(m,n)\, e^{\frac{-j2\pi}{a}(m+n)\sin\theta(Z_2''-Z_2')} \qquad (2\text{-}33)$$

Since $m+n=p=$constant for any row in $\underline{A}$, each a(m,n) in a given row would be multiplied by the same constant resulting in the rows in the lower half of $\underline{A}$ being linearly related to the corresponding rows in the upper half of $\underline{A}$. The $\underline{A}$ matrix would only be half the required rank.

If θ is varied between measurements and $\theta=\theta_1$ in the first measurement and $\theta=\theta_2$ in the second measurement, then

$$a_2(m,n) = a_1(m,n)\, e^{\frac{-j2\pi}{a}(\sin\theta_2-\sin\theta_1)[m(Z_1+Z_2)+nZ_2]} \qquad (2\text{-}34)$$

The situation is the same as when Z_1 was the parameter varied between measurements, so the $\underline{A}$ matrix is full rank. Since it is reasonable to vary θ (the input plane wave arrival angle) between measurements in a practical optical system, θ is the parameter that is considered to be variable.

The remaining problem is to solve the bilinear system of equations (Eq (2-30)). Since that is not a trivial problem, it will be addressed in Chapter 4. First, some of the properties of the solution developed in this chapter and some of the assumptions made will be examined.

Diffraction at the Edge

In solving for the integral in Eq (2-10), only critical points of the first kind were considered. This is equivalent to letting the limits of integration go to $\pm\infty$ so that diffraction at the edge is ignored. Suppose the actual limits of integration are 0 to "a" on both integrals in Eq (2-10). Let $g(x,\gamma,\beta) = \exp\left[jk(\gamma\sin\theta + \frac{(\beta-\gamma)^2}{2Z_1} + \frac{(x-\beta)^2}{2Z_2} + W_A(\gamma) + W_B(\beta))\right]$. Then Eq(2-10) can be written

$$U_5(x) = \frac{-jA'e^{jk(Z_1+Z_2)}}{\lambda\sqrt{Z_1Z_2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,\gamma,\beta)d\gamma d\beta$$

$$- \int_{-\infty}^{\infty}\int_{-\infty}^{0} g(x,\gamma,\beta)d\gamma d\beta - \int_{-\infty}^{\infty}\int_{a}^{\infty} g(x,\gamma,\beta)d\gamma d\beta$$

$$- \int_{-\infty}^{0}\int_{0}^{a} g(x,\gamma,\beta)d\gamma d\beta - \int_{a}^{\infty}\int_{0}^{a} g(x,\gamma,\beta)d\gamma d\beta \qquad (2\text{-}35)$$

The first integral in Eq (2-35) gives $U_5(x)$ ignoring diffraction. The last four integrals represent the contributions to $U_5(x)$ that come from edge diffraction. Usually,

not more than one of the last four integrals will be significant for a given x and θ. Specifically, the only significant integral will be one corresponding to the aperture edge which, when projected, is closest to the point of interest in the measurement plane. For example, if $U_5(x_0)$ were of interest, where the diffraction pattern near x_0 was due primarily to diffraction from the lower edge of aperture "A" in Figure 2-1, then

$$U_5(x_0) = -\frac{jA'e^{jk(Z_1+Z_2)}}{\lambda\sqrt{Z_1Z_2}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,\gamma,\beta)d\gamma d\beta - \int_{-\infty}^{\infty}\int_{-\infty}^{0} g(x,\gamma,\beta)d\gamma d\beta \qquad (2\text{-}36)$$

If for a particular optical system it is known that aberrations near the edge of the aperture do not significantly affect the diffraction pattern, then the second integral in Eq (2-31) would be known and it could be taken to the other side of the equation and added to $U_5(x_0)$ to compensate for diffraction. If it cannot be assumed that the diffraction pattern is unaffected by aberrations, then the aberrations can only be found on a subaperature where $U_5(x)$ is not measured in the diffraction pattern.

Vignetting

When the exponentiated aberration functions were represented by Fourier series in Eqs (2-21) and (2-22), the

assumption was implicitly made that the aberration functions were periodic with spatial periods "a" and "b" for $W_A(x)$ and $W_B(x)$, respectively. In Figure 2-1, when measurements are made for two different plane wave arrival angles, θ_1 and θ_2 for example, different portions of each aperture are projected to the measurement plane. The portion of each aperture gained in going from θ_1 to θ_2 is, in fact, not a periodic extension of the portion lost, so error is introduced in the deconvolution algorithm. The error may be minimized by keeping the difference between θ_1 and θ_2 as small as possible without making the bilinear system of equations (Eq (2-30)) ill-conditioned. The optimum value for θ_1-θ_2 is discussed in Chapter 3.

Space Bandwidth Product

When a discrete Fourier transform (DFT) is applied to a set of equally spaced samples of U(x), the resulting Fourier coefficients, H(p), represent the spatial frequency content of U(x). If U(x) is sampled at more than twice the highest spatial frequency, as it must be to meet the Nyquist sampling criteria, then some Fourier coefficients will be insignificantly small. The largest p for which H(p) is significant is the space-bandwidth product of U(x), and was previously called Pmax, a dimensionless number. The spatial bandwidth is just Pmax/a where "a" is the aperture dimension in the measurement plane over which $U_5(x)$ is measured. Care must be taken to properly inter-

pret the results of the DFT. An example will illustrate the point. Suppose "I" samples of U(x) are taken and I=8. Let the samples be called U(0) through U(7). Applying the DFT as in Eq (2-28), H(p) can be found for any given p. H(p) is periodic with period "I", so H(p)=H(p+nI) where "n" is any integer. Most available DFT programs would print values for H(0) through H(7), and if one does not recognize that H(7)=H(-1), one might let Pmax=7. It is a good idea to write the Fourier coefficients centered on zero, such as H(-4) through H(3). Now p represents the real space-bandwidth product and not a periodic extension of it. Note that the Nyquist criterion requires that a periodic function be sampled at a rate greater than twice the highest frequency component of the function. Therefore, if a periodic function is described by eight samples as in the example above, Pmax=3.

Dimensionality of the A Matrix

Once Pmax is determined, the dimensionality of the A matrix in Eq (2-30) can be found. There will be a column for each (m,n) pair subject to the conditions $|m| \leq Pmax$, $|n| \leq Pmax$, and $|m+n| \leq Pmax$. For a given m, there will be $2Pmax+1-|m|$ columns, so summing over all n yields

$$\sum_{m=-Pmax}^{Pmax} (2Pmax+1-|m|) = 3Pmax^2+3Pmax+1 \qquad (2\text{-}37)$$

columns. The row rank is (2Pmax+1) times the number of

independent measurements of $U_5(x)$ that are made. It can be seen that the dimensionality of the $\underline{A}$ matrix grows rapidly with increasing Pmax, thus placing a practical limit on the complexity of the aberrations that can be deconvolved.

Another property of the aberrations that will affect the dimensionality of the $\underline{A}$ matrix can be seen by noting the similarity between the Fourier series representation of the aberration functions in Eqs (2-21) and (2-22), and the representation of frequency modulated (FM) signals (Ref 25:117). If a sinusoidal signal with frequency ω_n is modulated onto a carrier with frequency ω_m, then the signal is represented by the Fourier series

$$e^{j\beta\sin\omega_n t} = \sum_{n=-\infty}^{\infty} J_n(\beta)e^{jn\omega_m t} \qquad (2\text{-}38)$$

Eq (2-38) is exactly the same form as Eq (2-21) if $W_A(x)$ is a sinusoid. The Fourier coefficients in Eq (2-38) are Bessel functions, and the number of significant coefficients is a function of β, the modulation index. For example, if β=10, over 30 coefficients are required to adequately represent

$$e^{j\beta\sin\omega_n t} \quad (\cong \sum_{n=-15}^{15} J_n(\beta)e^{jn\omega_m t}).$$

Obviously, the magnitude of the aberration functions must not be more than a few wavelengths, or Pmax will be extremely large.

One final note on the dimensionality of the $\underline{A}$ matrix: If from some other source of information it is known that certain Fourier coefficients (F_m or G_n) are zero where m or n $\leq$ Pmax, the columns in $\underline{A}$ and rows in the FG matrix of Eq (2-30) which contain those coefficients can be eliminated, thus reducing the dimensionality of the system of equations.

Summary

In this chapter, the simplest possible one-dimensional, two-element optical system was described. The deconvolution problem was then stated mathematically in terms of the simple optical system, and a solution was sought for the aberration functions W_A and W_B. The deconvolution problem was shown to be nonlinear. The nonlinear nature of the problem, as seen in Eq (2-30), is a major difficulty that is addressed in Chapter 4. Some important side issues such as the effects of diffraction, vignetting, and the dimensionality of the bilinear system of equations are also addressed in this chapter.

III. Noise Analysis for the One-Dimensional Problem

Approach

The reason for performing a noise analysis on the deconvolution problem is two-fold: first, to find a suitable procedure for estimating the Fourier coefficients of the aberration functions, and second, to determine how sensitive those estimates are to measurement noise. The approach taken is the one presented by Van Trees (Ref 23) wherein the measurements are assumed to be corrupted by Gaussian noise, the conditional probability density is written for the measured parameters assuming the unknown parameters are given, and either the maximum likelihood (ML) or the maximum a posteriori (MAP) estimate is found. The choice of ML or MAP estimate depends on whether the unknown parameters being sought are treated as real but unknown variables or as random variables. Finally, bounds on the variance of the unknown parameters are sought to give some indication of the sensitivity of the estimates to noise.

Probability Density Function for $\hat{H}$

Assume that the measurements of the complex field $U_5(x)$ are corrupted by noise, so that

$$\hat{U}_5(x) = A'U_5(x) + \hat{N}(x) \qquad (3\text{-}1)$$

where $U_5(x)$ is nonrandom but unknown (Ref 23:63), and A' is the amplitude of the input plane wave. In an actual optical system, $\hat{U}_5(x)$ would probably be found by integrating the optical signal falling on a detector over a finite time period. The temporal variations in the signal would be integrated out, so $\hat{U}_5(x)$ and $\hat{N}(x)$ are considered to be spatial rather than temporal random processes. That is, they are random processes as a function of x, not t. Further assume that the real and imaginary parts of $\hat{N}(x)$ are zero-mean, spatially white, and each have a variance of $\frac{1}{2}\sigma_n^2$.

$$\hat{H}(p,\theta) = \frac{1}{a} e^{-jk(Z_1+Z_2)}\left(1 - \frac{\sin^2\theta}{2}\right) \int_0^a [A'U_5(x)+\hat{N}(x)]$$

$$e^{-jkx\sin\theta}\, e^{-j\frac{2\pi xp}{a}}\, dx = H(p,\theta) + \hat{N}_H(p,\theta) \qquad (3\text{-}2)$$

The first and second moments of $\hat{N}_H(p,\theta)$ are

$$E\,[\hat{N}_H(p,\theta)] = 0 \qquad (3\text{-}3)$$

$$E\,[|\hat{N}_H(p,\theta)|^2] = \frac{1}{a^2}\int_0^a\int_0^a E\,[\hat{N}(x)\,\hat{N}^*(x')]\, e^{-jk(x-x')\sin\theta}$$

$$e^{-j\frac{2\pi(x-x')p}{a}}\, dxdx' \qquad (3\text{-}4)$$

A problem arises because

$$E\left[\hat{N}(x)\,\hat{N}^*(x')\right]\begin{cases} = 0 & \text{when } x \neq x' \\ = \sigma^2 & \text{when } x = x' \end{cases}$$

so the integral in Eq (3-4) is zero. The problem occurs because $\hat{N}(x)$ has been treated as a continuous function when in reality $\hat{N}(x)$ is sampled over the interval 0 to "a". If Eq (3-4) is written as a double summation instead of an integral, then

$$E\left[\left|\hat{N}_H(p,\theta)\right|^2\right] = \frac{1}{M^2}\sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} E\left[\hat{N}(m)\hat{N}^*(m')\right]$$

$$e^{jk(m-m')\sin\theta}\; e^{\frac{-j\pi(m-m')p}{M}} = \frac{1}{M^2}\sum_{m=0}^{M-1}\sigma_n^2 = \frac{\sigma_n^2}{M} \qquad (3\text{-}5)$$

where M is the number of samples of $\hat{N}(x)$ and $\hat{N}(m)$ is the m^{th} sample. If the number of samples goes to infinity as it does in Eq (3-4) where $\hat{N}(x)$ is written as a continuous function, then the variance of $\hat{N}_H(p,\theta)$ goes to zero. The variance of $\hat{N}_H(p,\theta)$ will be called σ henceforth and is equal to ${\sigma_n}^2/M$.

The cross-correlation of $\hat{N}_H(p,\theta)$ is

$$E\left[\hat{N}_H(p,\theta)\hat{N}_H^*(q,\theta)\right] = \frac{1}{a^2}\int_0^a\int_0^a E\left[\hat{N}(x)\hat{N}^*(x')\right]$$

$$e^{-jk(x-x')\sin\theta}\; e^{-j\frac{2\pi}{a}(xp-x'q)}\; dxdx' \qquad (3\text{-}6)$$

Again the double integral must be written as a double summation.

$$E\left[\hat{N}_H(p,\theta)\hat{N}_H^*(q,\theta)\right] = \frac{1}{M^2}\sum_{m=0}^{M-1}\sum_{m'=0}^{M-1} E\left[\hat{N}(m)\hat{N}^*(m')\right]$$

$$e^{jk(m-m')\sin\theta}\, e^{\frac{-j2\pi(mp-m'q)}{M}} = \frac{\sigma_n^2}{M^2}\sum_{m=0}^{M-1} e^{\frac{-j2\pi m(p-q)}{M}}$$

$$\begin{cases} = 0 & \text{if } p \neq q \\ = \sigma^2 & \text{if } p = q \end{cases} \qquad (3\text{-}7)$$

This result is as expected since the measurement noise $\hat{N}(x)$ was projected onto a set of orthogonal coordinates when $\hat{H}(p)$ was found. In summary, given measurement noise $\hat{N}(x)$ which is a zero-mean, spatially white random process with real and imaginary parts each having variance $\frac{1}{2}\sigma_n^2$, then $\hat{N}_H$ is a zero-mean random vector with orthogonal elements each having a variance of σ^2.

Assume now that $\hat{N}(x)$ is Gaussian. Since $\hat{N}_H$ is a linear function of $\hat{N}(x)$, $\hat{N}_H$ is also Gaussian, and the conditional probability density function for $\hat{H}$ given H can be written as

$$P_{\hat{H}|H}(\hat{H}|H) = \frac{1}{(2\pi)^{\frac{K}{2}}\,|\Lambda|^{\frac{1}{2}}}\exp\left[-\tfrac{1}{2}(\hat{H}-H)^{\dagger}\Lambda^{-1}(\hat{H}-H)\right] \qquad (3\text{-}8)$$

where K is the dimensionality of the H vector. From Eq (3-2), $\hat{H}-H=\hat{N}_H$, and from Eq (3-7)

$$\Lambda = E\,[\hat{N}_H^{\dagger}\hat{N}_H] = \sigma^2\,\underline{I} \qquad (3\text{-}9)$$

where "†" is the conjugate transpose and $\underline{I}$ is the identity matrix. Therefore,

$$P_{\hat{H}|H}(\hat{H}|\hat{H}) = \frac{1}{(2\pi)^{\frac{K}{2}}\sigma^K} \exp\left[-\frac{1}{2\sigma^2}(\hat{H}-H)^{\dagger}(\hat{H}-H)\right] \qquad (3\text{-}10)$$

Referring to Eq (2-30), $H=\underline{A}Y$, where Y represents the matrix with elements F_mG_n ($Y_{mn}=F_mG_n$). So Eq (3-10) can be written

$$P_{\hat{H}|Y}(\hat{H}|\hat{Y}) = \frac{1}{(2\pi)^{\frac{K}{2}}\sigma^K} \exp\left[-\frac{1}{2\sigma^2}(\hat{H}-\underline{A}Y)^{\dagger}(\hat{H}-\underline{A}Y)\right] \qquad (3\text{-}11)$$

meaning that the Gaussian density function of $\hat{H}$ is known if the Fourier coefficients F_m and G_n are given. For a given measurement vector $\hat{H}$, the maximum likelihood estimate of F_m and G_n will be those values which minimize $(\hat{H}-\underline{A}Y)^{\dagger}(\hat{H}-\underline{A}Y) = |\hat{H}-\underline{A}Y|^2$ in Eq (3-11). From Eq (177), page 65 of Van Trees (Ref 23), the maximum likelihood equation is

$$\left.\frac{\partial \ln P_{\hat{H}|F,G}(\hat{H}|F,G)}{\partial\, F \text{ (or } G)}\right|_{F=\hat{F}(\hat{H}),\ G=\hat{G}(\hat{H})} =$$

$$-\frac{1}{2\sigma^2}\left.\frac{\partial |\hat{H}-AY|^2}{\partial F \text{ (or } G)}\right|_{F=\hat{F}(\hat{H}),\ G=\hat{G}(\hat{H})} = 0 \qquad (3\text{-}12)$$

$\hat{F}$ and $\hat{G}$ are the maximum likelihood estimates of F and G, respectively, for a given measurement vector $\hat{H}$. The partial derivatives must be taken with respect to the real and imaginary parts of each F_m and G_n separately, because the derivative of a complex number is not defined. Equation (3-12) is necessary but not sufficient for estimating F and G unless the first derivative of the density function has only one zero. Since the density function is Gaussian in this case, the first derivative has only one zero and Eq (3-12) is both necessary and sufficient for finding the ML estimate of F and G.

Equation (3-12) is not very useful for finding F and G, because the system of equations generated by taking the partial derivatives of the real and imaginary parts of each F_m and G_n is still a nonlinear system in terms of F and G. Note, however, that finding the maximum likelihood estimate of F and G by minimizing $|\hat{H}-\underline{A}Y|^2$ is exactly the same as finding the least-square-error solution of Eq (2-30). The values of F and G that minimize $|\hat{H}-H|^2$ form the least-square-error solution to Eq (2-30) where $\hat{H}$ is computed from the measured values of $\hat{U}_5(x)$ (Eq (2-27)) and H is computed from the estimated values of F and G (Eq (2-30)). This point will be exploited in Chapter 4, but for now, bounds will be found for the estimates $\hat{F}$ and $\hat{G}$ rather than the actual solution for F and G.

Fisher Information Matrix

Let F_{mR} and G_{nR} be the real parts of F_m and G_n, and F_{mI} and G_{nI} be the imaginary parts. From Van Trees (Ref 23:79),

$$\sigma_i^2 \triangleq \text{Var}\,[\hat{d}_i(\hat{H})-d_i] \geq (J^{-1})_{ii} \tag{3-13}$$

where di represents F_{mR}, F_{mI}, G_{nR}, or G_{nI}, and where $(J^{-1})_{ii}$ is the ii^{th} element of the square matrix $\underline{J}^{-1}$. The matrix $\underline{J}$ is called the Fisher information matrix (FIM) and has elements

$$J_{ij} = -E\left[\frac{\partial^2 \ln P_{\hat{H}|F,G}(\hat{H}|F,G)}{\partial d_i\ \partial d_j}\right] \tag{3-14}$$

if F and G are nonrandom. If they are random,

$$J_{ij} = -E\left[\frac{\partial^2 \ln P_{\hat{H}|F,G}(\hat{H}|F,G)}{\partial d_i\ \partial d_j}\right] - E\left[\frac{\partial^2 \ln P_{\hat{F},\hat{G}}(\hat{F},\hat{G})}{\partial d_i\ \partial d_j}\right] \tag{3-15}$$

Often the decision whether to treat unknown variables such as F and G as nonrandom but unknown or as random is dictated by the mathematics. Both assumptions must be tried to see which one yields useful results. This is done in the following paragraphs.

Nonrandom Fourier Coefficients

Assuming first that F and G are nonrandom, Eqs (3-14) and (3-11) give

$$J_{ij} = \frac{1}{2\sigma^2} E\left[\frac{\partial^2 |\hat{H}-AY|^2}{\partial d_i \partial d_j}\right] \tag{3-15}$$

where the term

$$\ln \frac{1}{\frac{K}{(2\pi)^2}\sigma^K}$$

is neglected because it disappears when the partial derivatives are taken. The elements of vector $\hat{H}-\underline{A}Y$ are given by Eq (2-30) as

$$\hat{H}_J(p)-\underline{A}Y = \hat{H}_J(p) - \sum_{\substack{m \\ m+n=p}} \sum_{n} F_m G_n a_J(m,n) \tag{3-17}$$

where $a_J(m,n)$ is given by Eq (2-29).

$$|\hat{H}-\underline{A}Y|^2 = \sum_J \sum_p \left| \hat{H}_J(p) - \sum_{\substack{m \\ m+n=p}} \sum_{n} F_m G_n a_J(m,n) \right|^2 \tag{3-18}$$

where the summation over J means that the summation is repeated for each set of equations generated with a different plane wave arrival angle.

After the summations and squaring operation are carried out in Eq (3-18), all terms will contain elements of F and G. Even after $\partial^2/\partial di \partial dj$ is taken, most terms will contain elements of F or G. If F and G are nonrandom but unknown, the expected value operator in Eq (3-16) will not eliminate them and J_{ij} would be defined in terms of the

unknown coefficients. This would not be very useful, so the elements of F and G will be assumed random.

Random Fourier Coefficients

Assume the elements of F and G are random variables with known Gaussian densities. Let the variance and expected values of the real and imaginary parts of the Fs and Gs be given as

$$\text{VAR}\,[\hat{d}_i] = \sigma_{di}^2$$

$$E\,[\hat{d}_i] = 0 \text{ for } d_i \neq F_{0R} \text{ or } G_{0R}$$

$$E\,[\hat{d}_i] = 1 \text{ for } d_i = F_{0R} \text{ or } G_{0R} \qquad (3\text{-}19)$$

The reason for letting $E\,[\hat{F}_{0R}] = E\,[\hat{G}_{0R}] = 1$ is as follows. From Eq (2-21),

$$\frac{1}{a}\int_0^a |e^{jkW_A(x)}|^2 dx = 1 = \frac{1}{a}\int_0^a \left|\sum_m F_m e^{\frac{j2\pi mx}{a}}\right|^2 dx =$$

$$\sum_m |F_m|^2 \qquad (3\text{-}20)$$

Likewise, from Eq (2-22),

$$\frac{1}{a}\int_0^a |e^{jkW_B(x)}|^2 dx = 1 = \frac{1}{a}\int_0^a \left|\sum_n G_n e^{\frac{j2\pi nx}{a}}\right|^2 dx =$$

$$\sum_n |G_n|^2 \qquad (3\text{-}21)$$

Equations (3-19) are made consistent with Eqs (3-20) and (3-21) by letting $E[\hat{F}_{0R}] = E[\hat{G}_{0R}] = 1$.

Assuming independent elements of $\hat{F}$ and $\hat{G}$,

$$P_{\hat{F},\hat{G}}(\hat{F},\hat{G}) = \frac{1}{(2\pi)^{\frac{K}{2}} \prod_i di} \prod_{di \neq F_{0R} \text{ or } G_{0R}} \exp\left[-\frac{di^2}{2\sigma_{di}^2}\right]$$

$$\exp\left[-\frac{(F_{0R}-1)^2}{2\sigma_{F0R}^2} - \frac{(G_{0R}-1)^2}{2\sigma_{G0R}^2}\right] \qquad (3\text{-}22)$$

Combining Eqs (3-15), (3-16), and (3-22), and dropping the constants at the first of the probability density functions which get eliminated when the partials are taken, the elements of $\underline{J}$ can be written as

$$J_{ij} = \frac{1}{2\sigma^2} E\left[\frac{\partial^2 |\hat{H}-AY|^2}{\partial di\ \partial dj}\right] + E\left[\frac{\partial^2}{\partial di\ \partial dj}\right.$$

$$\left.\left(\sum_{\substack{i \\ di \neq F_{0R} \text{ or } G_{0R}}} \frac{di^2}{2\sigma_{di}^2} + \frac{(F_{0R}-1)^2}{2\sigma_{F0R}^2} + \frac{(G_{0R}-1)^2}{2\sigma_{G0R}^2}\right)\right] \qquad (3\text{-}23)$$

The process of finding the elements of and inverting $\underline{J}$ to get $(\underline{J}^{-1})_{ii}$ is straightforward but tedious. The calculations are documented in Appendix A, and only the result is given in this chapter. In Appendix A it is shown that the lower bounds on VAR $[\hat{d}i]$ can be found in the general case by computing the values for the entries in $\underline{J}$, then using the computer to invert the matrix. If the simplifying assumption is made that $H(p,\theta) << H(0,\theta)$ where $p \neq 0$, and $\sigma_{diR}^2 = \sigma_{diI}^2$, $\underline{J}$ reduces to several 4X4 matrices that can be inverted in general. The inverted 4X4 matrices yield an explicit equation for the lower bounds on VAR $[\hat{d}i]$. This is given by Eq (A-12), which is repeated below.

$$VAR[\hat{F}_{iR}] = VAR[\hat{F}_{iI}] \geq$$

$$\frac{\sigma^2 \sum_J (2 \sum_k \sigma_{Fk}^2 + 1) + \frac{\sigma^4}{\sigma_{Gi}^2}}{\left[\left[\sum_J (2 \sum_k \sigma_{Gk}^2 + 1) + \frac{\sigma^2}{\sigma_{Fi}^2} \right] \left[\sum_J (2 \sum_k \sigma_{Fk}^2 + 1) + \frac{\sigma^2}{\sigma_{Gi}^2} \right] - \left[\sum_J \cos\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right) \right]^2 - \left[\sum_J \sin\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right) \right]^2 \right]} \qquad \text{(A-12)}$$

where

$$-P_{max} + i \leq k \leq P_{max} \text{ for } i \geq 0$$

and

$$-P_{max} \leq k \leq P_{max}+i \text{ for } i \leq 0$$

The simplifying assumptions used to derive Eq (A-12) correspond to the case where the aberrations are almost zero. While the assumptions are restrictive, a study of the near-zero aberration optical system yields useful insights into the deconvolution problem. If the exact lower bounds are needed or if the aberrations are not near zero, the lower bounds can always be found with a computer. The near-zero aberration case will be studied exclusively for the remainder of this chapter.

Analysis of Lower Bounds When Aberrations Are Near Zero

A test problem will now be formulated to gain some insight into the behavior of Eq (A-12). Assume that measurements are taken at the output of the optical system described in Chapter 2 for two different plane wave arrival angles, θ_1 and θ_2. Then Eq (A-12) can be written

$$VAR[\hat{F}_{iR}] = VAR[\hat{F}_{iI}] \geq$$

$$\frac{\sigma^2 \left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right)}{\left[\left(4 \sum_k \sigma_{Ck}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)\left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right) - \right.}$$

$$\left. 2 - 2\cos\left(\frac{2\pi}{a} Z_1 i \sin\theta_1\right)\cos\left(\frac{2\pi}{a} Z_1 i \sin\theta_2\right) - \right.$$

$$\left. 2\sin\left(\frac{2\pi}{a} Z_1 i \sin\theta_1\right)\sin\left(\frac{2\pi}{a} Z_1 i \sin\theta_2\right) \right] \qquad (3\text{-}24)$$

Equation (3-24) can be simplified by use of the angle difference relationship $\cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\alpha-\beta)$ and the power relationship $\cos^2\alpha = \frac{1}{2}(1 + \cos 2\alpha)$.

$$\mathrm{VAR}[\hat{F}_{iR}] = \mathrm{VAR}[\hat{F}_{iI}] \geq \frac{\sigma^2 \left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right)}{\left[\left(4 \sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)\left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right) - 4 \cos^2\left[\frac{\pi}{a} Z_1 i(\sin\theta_1 - \sin\theta_2)\right]\right]} \quad (3\text{-}25)$$

The bounds on $\mathrm{VAR}[\hat{G}_{iR}] = \mathrm{VAR}[\hat{G}_{iI}]$ are the same as in inequality (3-24) except that the numerator is

$$\sigma^2 \left(4 \sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)$$

Let the right side of inequality (3-25) be called ε. The limits on ε as a function of σ^2 and σ_{di}^2 are as follows. If $\sigma^2 \to 0$, $\varepsilon \to 0$. This says that if the measurements of $U_5(x)$ are noiseless, the VAR $[\hat{d}i]$ may approach zero as expected. If $\sigma^2 \to \infty$,

$$\varepsilon = \frac{\sigma^4/\sigma_{Gi}^2}{\sigma^4/\sigma_{Fi}^2\,\sigma_{Gi}^2} = \sigma_{Fi}^2 ,$$

which is as expected, since if we have no measurement information, the VAR $[\hat{F}_i]$ is bounded by the variance from the prior statistics. If $\sigma_{Fi}^2 \to 0$, $\varepsilon = \sigma_{Fi}^2/\sigma^2 = 0$, as expected. A limit for $\sigma_{Fi}^2 \to \infty$ would be meaningless since we know that $|di| \leq 1$ (Eq (3-20)).

Uniqueness Versus Input Plane Wave Arrival Angle

Note that for any given di, σ^2 and σ_{di}^2, ε is a maximum when $\cos^2[\frac{\pi}{a}Z_1 i(\sin\theta_1-\sin\theta_2)] = 1$. In this case, the quantity $\frac{\pi}{a}Z_1 i(\sin\theta_1-\sin\theta_2)$ equals 0 or $\pm n\pi$ resulting in $\sin\theta_1-\sin\theta_2=0$ or $\pm na/Z_1 i$. This has a useful physical interpretation which can be illustrated with an example. Suppose $\theta_1=0$, $Z_1=2$, a=1, n=1, and i=2. This is shown in Figure 3-1.

For this case, $\sin\theta_2=0$ or $\pm 1/4$ to maximize ε. The aberration corresponding to i=2 is an optical wedge as can be seen from Eq (2-21):

$$e^{jkW_A(x)} = F_2 e^{\frac{j4\pi x}{a}} = e^{jk2\lambda x},$$

so

$$W_A(x) = 2\lambda x.$$

Now suppose that measurements of the field incident at some point on $W_B(x)$ are taken first for the plane wave arriving at $\theta_1=0$, then for the plane wave arriving at $\theta_2=\arcsin(.25) \cong \arctan(.25)$. In Figure 3-1, the normals of these plane waves are labeled as ray 1 and ray 2,

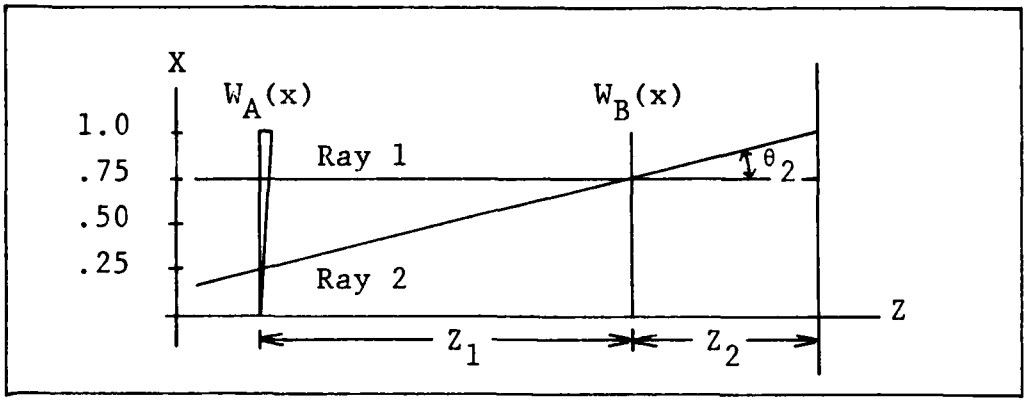


Fig 3-1. Example Nonunique Deconvolution Problem

respectively. The phase difference between plane waves 1 and 2 at $W_B(x)$ is the same whether the aberration $W_A(x)=2\lambda x$ is present or not. This is because ray 1 is delayed exactly one wavelength when the aberration is present as compared to the case when no aberration is present. This is true no matter where the plane wave phase is measured along $W_B(x)$. Therefore, a deconvolution scheme such as the one described in this paper which relies on the relative phase between plane waves arriving at different angles, would not be able to detect some components of the aberration functions for some specific arrival angles. Equation (3-25) may be used to select the optimum difference between arrival angles for a given Fourier component of the aberration funtions.

When the aberration functions consist of many Fourier components instead of just one as in the previous example, some method must be used to determine a "good" difference angle $(\theta_2-\theta_1)$. The difference angle must be "good" in the

sense that ϵ_i, the lower bound on VAR [$\hat{di}$] (right side of inequality (3-24)), is relatively small for every Fourier component d_i. A simple way to select a difference angle is to sum the ϵ_i corresponding to each d_i and select the difference angle that minimizes that sum. Note that the sum $\sum\epsilon_i$ is the trace of the inverted FIM, where the trace of a matrix is defined as the sum of the diagonal elements of a matrix. Let the trace be denoted by T. Then

$$T = \sum_i \epsilon_i =$$

$$\sum_i \frac{2\sigma^2 \left(4\sum_k \sigma_{Fk}^2 + 4 + \frac{\sigma^2}{\sigma_{Gi}^2} + 4\sum_k \sigma_{Gk}^2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)}{\left[\left(4\sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)\left(4\sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right) - 4\cos^2\left[\frac{\pi}{a} Z_1 i(\sin\theta_1 - \sin\theta_2)\right]\right]} \tag{3-26}$$

For given prior statistics, measurement statistics, and number of Fourier coefficients (-Pmax ≤i≤ Pmax), the optimum difference $\sin\theta_1$-$\sin\theta_2$ may be found graphically. This is done for three examples in Figure 3-2. In all three examples, all prior variances σ_{di}^2=.001; SIGMAH=σ^2, DIFF=$\sin\theta_1$-$\sin\theta_2$, Z_1 are as defined in Figure 3-1; and Pmax is the highest Fourier component (bounds "i" in the summation of Eq (3-25)). For all graphs, Z_1=2 and a=1.

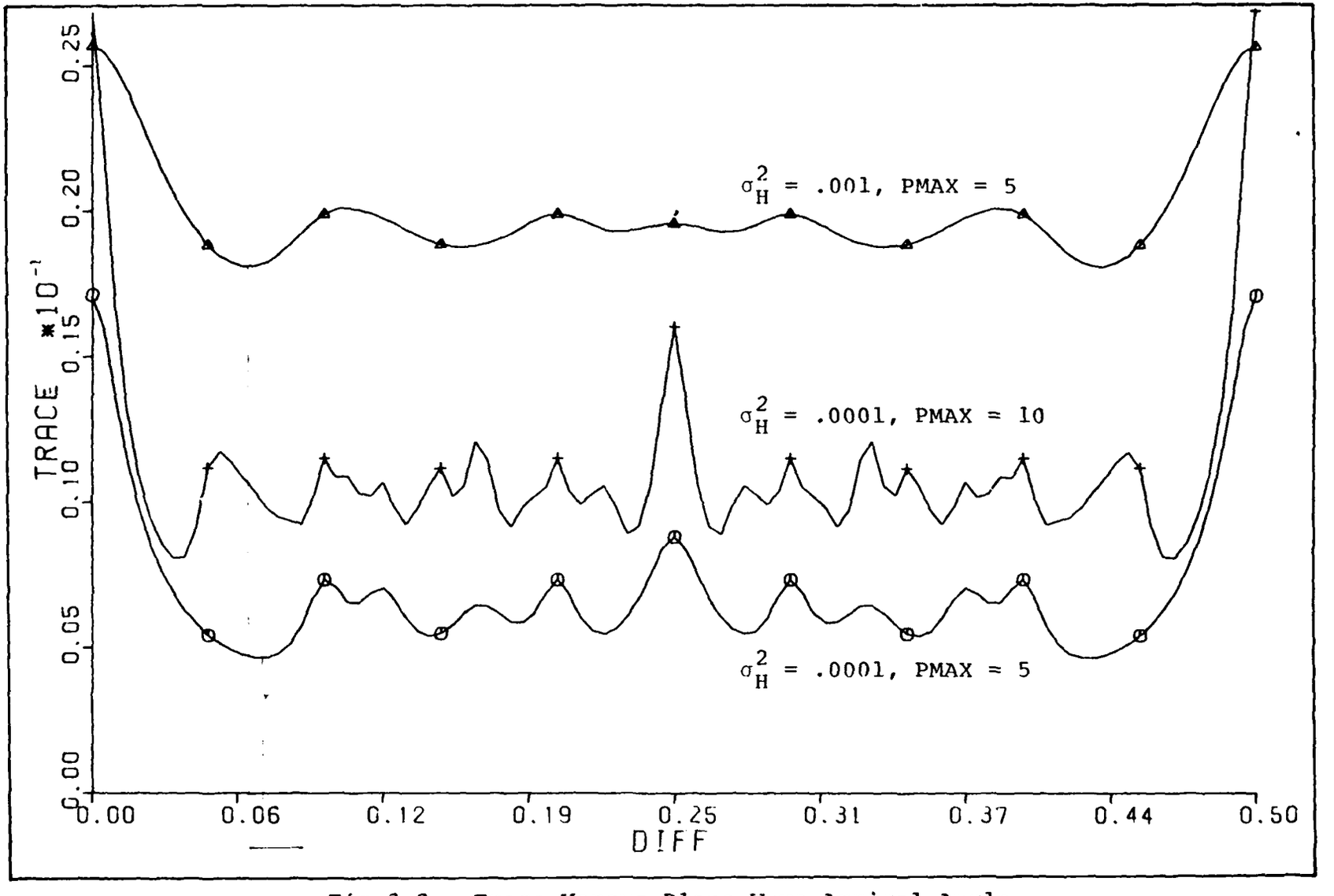


Fig 3-2. Trace Versus Plane Wave Arrival Angle

In Figure 3-2, a comparison of the top and bottom graphs shows that for a given Pmax, the trace is greater when the measurement noise (σ^2) is greater, as expected. The minimum value of the trace occurs at $\sin\theta_1-\sin\theta_2=.065$ for Pmax=5 and $\sigma^2=.001$, and at .07 for Pmax=5 and $\sigma^2=.0001$, and at .034 for Pmax=10 and $\sigma^2=.0001$. If $\theta_1=0$, then the optimum θ_2 for the three cases above is 3.7°, 4.0°, and 1.9°, respectively. There does not appear to be any way to determine the optimum $\sin\theta_1-\sin\theta_2$ except numerically. However, this can be done a priori for given Pmax and σ_{di}^2. The measurement noise σ^2 does not seem to have much effect on the optimum plane wave arrival angles.

Linking TRACE to $W(x)-\hat{W}(x)$

Equation (3-26) establishes T as a lower bound on $\sum$ VAR $[\hat{di}]$. The final step in the noise analysis is to relate T to the error functions $\hat{W}_A(x)-W_A(x)$ and $\hat{W}_B(x)-W_B(x)$. First T must be split into two smaller summations, one corresponding to F_m and the other corresponding to G_n so that (see Eq (3-26))

$$T_F = \sum_{i=-Pmax}^{Pmax} \frac{2\sigma^2 \left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right)}{\left[\left(4 \sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)\left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right) - 4 \cos^2\left[\frac{\pi}{a} Z_1 i(\sin\theta_1 - \sin\theta_2)\right]\right]} \quad (3\text{-}27)$$

and

$$T_G = \sum_{i=-Pmax}^{Pmax} \frac{2\sigma^2 \left(4 \sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)}{\left[\left(4 \sum_k \sigma_{Gk}^2 + 2 + \frac{\sigma^2}{\sigma_{Fi}^2}\right)\left(4 \sum_k \sigma_{Fk}^2 + 2 + \frac{\sigma^2}{\sigma_{Gi}^2}\right) - 4 \cos^2\left[\frac{\pi}{a} Z_1 i(\sin\theta_1 - \sin\theta_2)\right]\right]} \quad (3\text{-}28)$$

Noting that

$$e^{jkW_A(x)} - e^{jk\hat{W}_A(x)} = \sum_n (F_n - \hat{F}_n) e^{\frac{j2\pi nx}{a}} ,$$

T can be related to the error in the aberration functions as follows:

$$E\left[\frac{1}{a}\int_0^a \left|e^{jkW_A(x)}-e^{jk\hat{W}_A(x)}\right|^2 dx\right] =$$

$$E\left[\frac{1}{a}\int_0^a \left|\sum_n (F_n-\hat{F}_n)e^{\frac{j2\pi nx}{a}}\right|^2 dx\right] = E\left[\sum_n |F_n-\hat{F}_n|^2\right] =$$

$$E\sum_n \left[|F_{nR}-\hat{F}_{nR}|^2 + |F_{nI}-\hat{F}_{nI}|^2\right] \qquad (3\text{-}29)$$

If $\hat{F}_{nR}$ and $\hat{F}_{nI}$ are unbiased estimates, then

$$E\sum_n \left[|F_{nR}-\hat{F}_{nR}|^2 + |F_{nI}-\hat{F}_{nI}|^2\right] =$$

$$\sum_n \left(VAR[\hat{F}_{nR}] + VAR[\hat{F}_{nI}]\right) \geq T_F \qquad (3\text{-}30)$$

Equations (3-29) and (3-30) are the same for $W_B(x)$ with G and T_G replacing F and T_F, respectively. The final result of the noise analysis then is that a lower bound can be found for the mean squared error of the exponential of the aberration functions. This result is summarized in the following equations:

$$E\left[\frac{1}{a}\int_0^a \left|e^{jkW_A(x)}-e^{jk\hat{W}_A(x)}\right|^2 dx\right] \geq T_F \qquad (3\text{-}31)$$

$$E\left[\frac{1}{a}\int_0^a \left|e^{jkW_B(x)}-e^{jk\hat{W}_B(x)}\right|^2 dx\right] \geq T_G \qquad (3\text{-}32)$$

Summary

The noise analysis of this chapter is based on the assumption that the variable elements of Eq (2-30) ($\hat{F}$, $\hat{G}$, and $\hat{H}$) are random variables with Gaussian probability densities. This assumption is routinely used when no information on the actual densities is available. While the densities may deviate somewhat from Gaussian in a real optical system, the Gaussian assumption generally yields results that match reality quite well.

Two significant results have been derived in this chapter. First, Eq (3-26) enables one to select optimum plane wave arrival angles θ_1 and θ_2 for a given number of Fourier coefficients. This result has a satisfying intuitive interpretation that was illustrated by the example of Figure 3-1. Second, lower bounds on the difference between the calculated and the actual aberration functions were given by Eqs (3-31) and (3-32). While one would really like to have upper bounds, lower bounds are often very close to the actual error and give some assurance that the problem is not overly sensitive to noise if the proper input plane wave arrival angles θ_1 and θ_2 are chosen.

IV. Solution to a System of Bilinear Equations

Introduction

This chapter will explore a method of solving the nonlinear system of equations (Eq (2-30)). With the aid of a computer, a number of methods could be applied such as the Newton-Raphson method (Ref 4:249). Instead of using one of the standard methods, a novel method is presented that takes advantage of the fact that the nonlinear system of Eq (2-30) is bilinear. This method possesses the advantages of always iterating "downhill" and of being relatively easy to program on the computer. However, like other schemes for solving nonlinear systems of equations, it uses a lot of computer time when the dimensionality of the equations is large, and it will iterate to local minima under certain conditions. No attempt is made to compare the solution method of this chapter with standard methods. However, the computer solutions to some example problems are presented to demonstrate some of the characteristics of the solution method.

Solution Method

The individual equations in Eq (2-30) are given by Eq (2-27). For convenience, both equations are repeated below.

$$H_J(p) = \sum_{\substack{m\ n \\ m+n=p}} \sum F_m G_n \exp\left[-j\frac{2\pi}{a}[m\sin\theta_J(Z_1+Z_2) + Z_2 n\sin\theta_J]\right] \tag{2-27}$$

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} F_{-Pmax}G_0 \\ F_{-Pmax}G_1 \\ \vdots \\ F_{-Pmax}G_{Pmax} \\ F_{-Pmax+1}G_{-1} \\ F_{-Pmax+1}G_0 \\ \vdots \\ F_{-Pmax+1}G_{Pmax} \\ \vdots \\ F_{Pmax}G_{-Pmax} \\ \vdots \\ F_{Pmax}G_0 \end{bmatrix} = \begin{bmatrix} H_1(-Pmax) \\ H_1(-Pmax+1) \\ \vdots \\ H_1(Pmax) \\ H_2(-Pmax) \\ \vdots \\ H_2(Pmax) \end{bmatrix} \tag{2-30}$$

Since $m+n=p$ in Eq (2-27), it can be rewritten as a single summation.

$$H_J(p) = \sum_m F_m G_{p-m} e^{\frac{-j2\pi}{a}[(Z_1+Z_2)m\sin\theta_J + Z_2(p-m)\sin\theta_J]} \tag{4-1}$$

In Eq (4-1), $|m| \leq Pmax$ and $|p-m| \leq Pmax$. The elements of the $\underline{A}$ matrix are given by the exponential factor in Eq (4-1). For a given element in the $\underline{A}$ matrix, $m+n$ must equal p in $H_J(p)$ for the row of the element, and m and $p-m$ must be equal to the m and n, respectively, in the

$F_m G_n$ corresponding to the column of the element. Otherwise, the element is zero.

There is no mathematical guarantee that a consistent solution to Eq (2-30) even exists, so we must rely on the fact that if the equations accurately represent a real-world optical system, they will be consistent. Also, it can be seen by inspection that the solution to Eq (2-30) is not unique unless the magnitude and phase of either the F or G vector is given. This can be seen by noting that if vectors F and G are solutions, so are CF and G/C, where C is any complex number. From Eqs (3-20) and (3-21), the magnitude of F and G are each unity. The phase is entirely arbitrary, so it will be set to $F_{0I}=0$.

One obvious approach to solving Eq (2-30) would be to solve it as a linear system for the $F_m G_n$ pairs, then devise a scheme for dividing different pairs into each other to form ratios such as $F_2G_0/F_0G_0 = F_2/F_0$. The fact that $|F|^2 = 1$ and $|G|^2 = 1$ and the assumption that the phase of F_0 is 0 could then be used to solve for the Fourier coefficients. There are two major drawbacks to this approach. First, if only two measurements of the field at the measurement plane are taken, Eq (2-30) will always be an underdetermined linear system with the $F_m G_n$ pairs as the unknowns. Of course additional measurements of the optical field could be taken, but the number of required measurements becomes large as the dimensionality of Eq (2-30)

increases. For example, the row dimension of $\underline{A}$ per measurement is $2Pmax+1$ and the column dimension is $3(Pmax^2+Pmax)+1$. If $Pmax=1$, it would require three measurements to make Eq (2-30) overdetermined; but if $Pmax=5$, it would require nine measurements.

An even more severe drawback to the solution scheme described in the paragraph above is that if the measurements are corrupted by noise, there is no guarantee that the errors produced by the noise will not be propagated by the scheme in such a way that very small amounts of measurement noise produce very large errors in the solution for the Fourier coefficients. For these reasons, a more robust approach to solving Eq (2-30) is pursued.

A clue to a robust method for solving Eq (2-30) was given in Chapter 3 where it was noted that the maximum likelihood estimate of the unknown coefficients was the least-squared-error solution of Eq (2-30). Therefore, a solution will be sought which minimizes the least squared error ϵ_R, where

$$\epsilon_R = \sum_J \sum_{p=-Pmax}^{Pmax} \left| \hat{H}_J(p) - \sum_m F_m G_{p-m} \right.$$

$$\left. e^{\frac{-j2\pi}{a}[m\sin\theta_J(Z_1+Z_2)+(p-m)\sin\theta_J Z_2]} \right|^2 \qquad (4-2)$$

In Eq (2-30), there are the same number of equations as unknowns. If one set of coefficients (F or G) were

known, Eq (2-30) would be an overdetermined linear system in the remaining unknown coefficients. One approach to finding the coefficients is to assume a solution for one set of coefficients (say F), find the least-squared-error solution for G, substitute the solution for G back into Eq (2-30), find the least-squared-error solution for F, and continue iterating in this fashion until an overall solution is reached. The only known way to study the convergence properties of this algorithm is through simulation. An exhaustive simulation study will not be attempted, but a few example problems will be presented to demonstrate some of the properties of the algorithm.

Real Bilinear Equation Solution

The first example problem is a real bilinear system with six unknown coefficients, three Fs and three Gs. For simplicity, no attempt is made at this point to model a real optical system. The example problem is arbitrary except that the A matrix must be full rank and the system of equations must be consistent. The example problem is given by Eq (4-3) below.

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} F_1G_1 \\ F_1G_2 \\ F_1G_3 \\ F_2G_1 \\ F_2G_2 \\ F_2G_3 \\ F_3G_1 \\ F_3G_2 \\ F_3G_3 \end{bmatrix} = \begin{bmatrix} -.12 \\ .85 \\ .45 \\ .13 \\ -.05 \\ -.25 \end{bmatrix} \qquad (4\text{-}3)$$

$\underline{A}$ is a 6x9 matrix with elements a(i,j) = cos(1.3*I*J). Consistency was assured by assuming values for F and G vectors and solving for H. The actual values for F and G are F=[.447, -.837, .316] and G=[.837, -.316, .447]. However, in solving Eq (4-3), it is assumed that F and G are unknown except that they each have a length of one.

The first step in the solution is to assume a value for G. Assume G=[.447, -.447, .775]. Then G can be combined with $\underline{A}$ as follows:

$$\begin{bmatrix} (a_{11}G_1+a_{12}G_2+a_{13}G_3) & (a_{14}G_1+a_{15}G_2+a_{16}G_3) & (a_{17}G_1+a_{18}G_2+a_{19}G_3) \\ (a_{21}G_1+a_{22}G_2+a_{23}G_3) & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ (a_{61}G_1+a_{62}G_2+a_{63}G_3) & \cdot & (a_{67}G_1+a_{68}G_2+a_{69}G_3) \end{bmatrix}$$

$$X \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} H \end{bmatrix} \qquad (4\text{-}4)$$

Equation (4-4) is now an overdetermined linear system. The least-squared-error solution for F is found by use of the commercially available International Mathematics and Statistical Library (IMSL) subroutine called LLSQF. This subroutine uses a Q-R decompostion of matrix $\underline{A}$ as described in Ref 4:216 and Ref 13. The least-squared-error solution of F is then substituted into Eq (4-3), yielding an over-determined linear system with unknown G. The least-squared-error solution for G is then found the same as for F. This procedure is repeated until the sum of the squared errors stops decreasing. Figure 4-6 is the flow diagram of BILIN2, a computer program which implements the algorithm described above. Note that the above scheme always iterates "downhill" (least-square error always decreases) because if the algorithm yields an F or G vector at some iteration that causes the squared errors to increase, then by definition it is not the least-squared-error solution.

Figures 4-1 and 4-2 show the computer output for BILIN2, each figure corresponding to a different initial guess for G. In the computer output, GSQ is simply $[g_1^2, g_2^2, g_3^2]$. The correct value for G is [.837, -.316, .447]. The algorithm uses G rather than GSQ, but GSQ is computed for use in graphing the iterative solution, as will be shown later. R ($=\varepsilon_R$) is the squared error at each iteration defined by

```
  G=   .7746 0.0000   .6325
GSQ=   .6000 0.0000   .4001

F=   .0668 -.5618 -.0234                                    R=.2977E+00
G=   .8296 -.3361   .4458 GSQ=   .6883   .1130   .1987     R=.1578E+00   TOL=.5433E+00   2
F=   .4258 -.8201   .3149                                   R=.1364E-02
G=   .8362 -.3184   .4465 GSQ=   .6993   .1014   .1993     R=.1799E-03   TOL=.3650E+00   4
F=   .4449 -.8349   .3161                                   R=.2030E-04
G=   .8366 -.3164   .4471 GSQ=   .7000   .1001   .1999     R=.2205E-05   TOL=.3604E+00   6
F=   .4470 -.8365   .3162                                   R=.2313E-06
G=   .8367 -.3162   .4472 GSQ=   .7000   .1000   .2000     R=.2418E-07   TOL=.3599E+00   8
F=   .4472 -.8367   .3162                                   R=.2519E-08
G=   .8367 -.3162   .4472 GSQ=   .7000   .1000   .2000     R=.2624E-09   TOL=.3599E+00  10
F=   .4472 -.8367   .3162                                   R=.2731E-10
G=   .8367 -.3162   .4472 GSQ=   .7000   .1000   .2000     R=.2843E-11   TOL=.3599E+00  12
```

Fig 4-1. BILIN2 Output, Initial G=(.7746, 0, .6325)

```
 H=   -.11680     .85365     .44975     .13038    -.05485    -.25241


RUN 1
  G=   .4472 -.4472  .7746
GSQ=   .2000  .2000  .6000

F= -.4452 -.3456 -.1465                                    R=.2262E+00
G=  .3238 -.7519  .5742 GSQ=  .1049  .5654  .3297  R=.7573E-01  TOL=.3344E+00  2
F= -.7064 -.3378 -.1263                                    R=.1690E-01
G=  .2799 -.7365  .6158 GSQ=  .0783  .5424  .3792  R=.1227E-01  TOL=.2661E+00  4
F= -.7989 -.3371 -.1787                                    R=.7803E-02
G=  .2451 -.7228  .6461 GSQ=  .0601  .5225  .4175  R=.4232E-02  TOL=.2517E+00  6
F= -.8673 -.3416 -.2197                                    R=.2112E-02
G=  .2282 -.7150  .6608 GSQ=  .0521  .5113  .4366  R=.1056E-02  TOL=.2461E+00  8
F= -.8986 -.3453 -.2398                                    R=.6227E-03
G=  .2219 -.7120  .6662 GSQ=  .0492  .5069  .4439  R=.4593E-03  TOL=.2446E+00 10
F= -.9097 -.3470 -.2473                                    R=.4025E-03
G=  .2197 -.7109  .6681 GSQ=  .0483  .5054  .4463  R=.3833E-03  TOL=.2441E+00 12
F= -.9134 -.3476 -.2498                                    R=.3770E-03
G=  .2190 -.7106  .6687 GSQ=  .0480  .5049  .4471  R=.3749E-03  TOL=.2440E+00 14
F= -.9146 -.3478 -.2506                                    R=.3743E-03
G=  .2188 -.7105  .6689 GSQ=  .0479  .5048  .4474  R=.3741E-03  TOL=.2440E+00 16
F= -.9150 -.3478 -.2509                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4474  R=.3740E-03  TOL=.2440E+00 18
F= -.9151 -.3479 -.2510                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 20
F= -.9151 -.3479 -.2510                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 22
F= -.9152 -.3479 -.2510                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 24
F= -.9152 -.3479 -.2510                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 26
F= -.9152 -.3479 -.2510                                    R=.3740E-03
G=  .2187 -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 28
F= -.9152 -.3479 -.2510                                    R=.3740E-03
```

(Continued)

(Continued)

```
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 30
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 32
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 34
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 36
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 38
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 40
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 42
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 44
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 46
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 48
F= -.9152  -.3479  -.2510
G=  .2187  -.7104  .6689 GSQ=  .0478  .5047  .4475  R=.3740E-03  TOL=.2440E+00 50
```

Fig 4-2. BILIN2 Output, Initial G=(.4472, -.4472, .7746)

$$R = |[\underline{A}][FG] - \hat{[H]}|^2.$$

Note that in Figure 4-1, with an initial guess of G=(.7746, 0., .6325), the algorithm converges quickly to the correct solution. But in Figure 4-2 with an initial guess of G=(.4472, -.4472, .7746), the algorithm converges to a local minimum which is not the correct solution.

Since $G_1^2=1-G_2^2-G_3^2$, R is a function of only two independent variables in this problem. This makes it easy to show R as a function of the two variables, which is done in Figure 4-3. The independent variables are G_2 and G_3. GRID1 is the computer program used to generate Figure 4-4. G_1 is assumed positive for all values of G_2 and G_3. The other half of the solution space where G_1 is negative is just the mirror image of the solution space shown in Figure 4-3. For convenience, a vector GS is defined as $GS=[\pm G_1^2, \ldots \pm G_n^2]$, where the sign is chosen as the sign of the corresponding element of G. The values of $\pm G_2^2$ and $\pm G_3^2$ are indicated along the horizontal and vertical axes, respectively, instead of the values of G_2 and G_3. The solid lines on the figure are contour lines lying along a constant value of R. The small Xs are initial guesses for G which were input to BILIN2, and the dashed lines from point to point show the successive values of G computed by BILIN2. The ovals show the position of local minima.

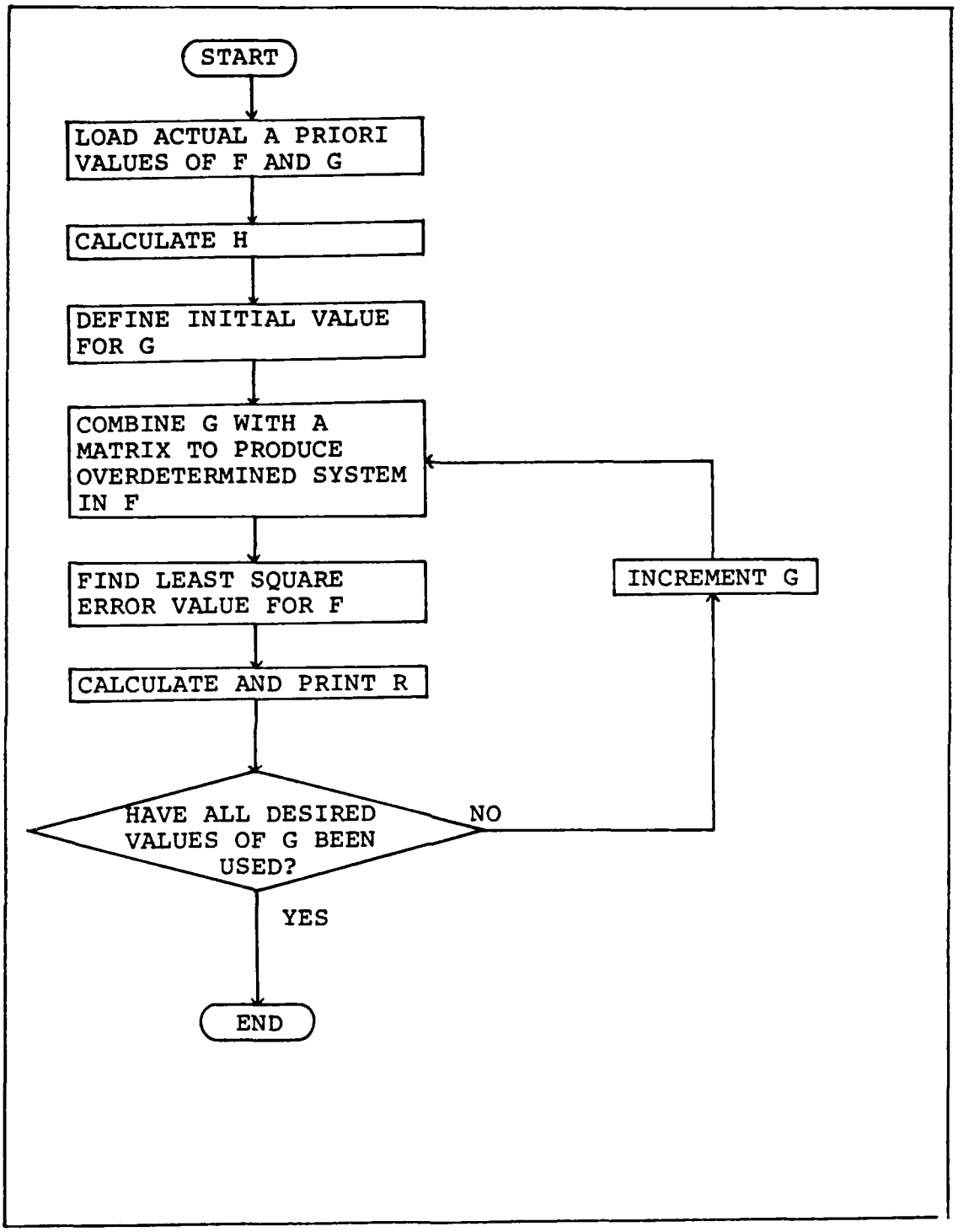


Fig 4-4. Flowchart for GRID1

An examination of Figure 4-3 reveals a number of characteristics of the algorithm under study as well as some basic characteristics of the solution space. First, there are four local minima, only one of which is the correct solution. Second, the algorithm will iterate to the minimum in the same "valley" that contains the initial guess because the algorithm always iterates downhill. Therefore, the initial guess must be reasonably close to the solution. If the algorithm iterates to a local minimum, it is usually obvious. The squared error will be greater than expected, and often $|F|^2$ and $|G|^2$ will not both equal 1. Third, opposite edges of the solution space are adjacent to each other. Note that opposite edges of Figure 4-3 have the same values of R. Because opposite edges are adjacent, the downhill iterations may cross over an edge and reappear on the adjacent edge.

Figure 4-5 is a cross section of the solution space shown in Figure 4-3 with varying amounts of noise added to the H vector. The noise vector was generated by a Gaussian random number generator. The cross section shows the values of R between and on either side of the correct minimum and the local minimum located near $G(2) = -.5$ and $G(3) = +.4$. The four graphs in Figure 4-5 correspond to no noise, 52 dB, 31 dB, and 16 dB signal-to-noise ratio. The location of the minima change very little as a function of noise, but it can be seen from Figure 4-5 that the depth

of the true minimum changes considerably as the bilinear equations become less consistent. Noise, then, may make it more difficult to identify the true minimum, but if the algorithm iterates to the true minimum, the error in the resulting coefficients does not appear to be unusually large. This point is amplified by Figure 4-8 and the associated discussion.

Complex Bilinear Equation Solution

The next step in solving the deconvolution problem is the creation of a program that will solve a complex bilinear system of equations of arbitrary size. Program DECON1 accomplishes this for the one-dimensional problem by using the following relationships: If A=B+jC, F=X+jY, and H=D+jE, and if AF=H, then

$$\begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} D \\ E \end{bmatrix} \qquad (4\text{-}5)$$

The flowchart for DECON1 is basically the same as for BILIN2 in Figure 4-6 except that before processing, DECON1 converts all complex matrices and vectors to real matrices and vectors by use of Eq (4-5). This is done as a matter of convenience because the IMSL subroutine LLSQF which finds the least squares solution only operates on real matrices.

Figures 4-7 through 4-10 show the output of DECON1 for four different test problems. These test problems still do

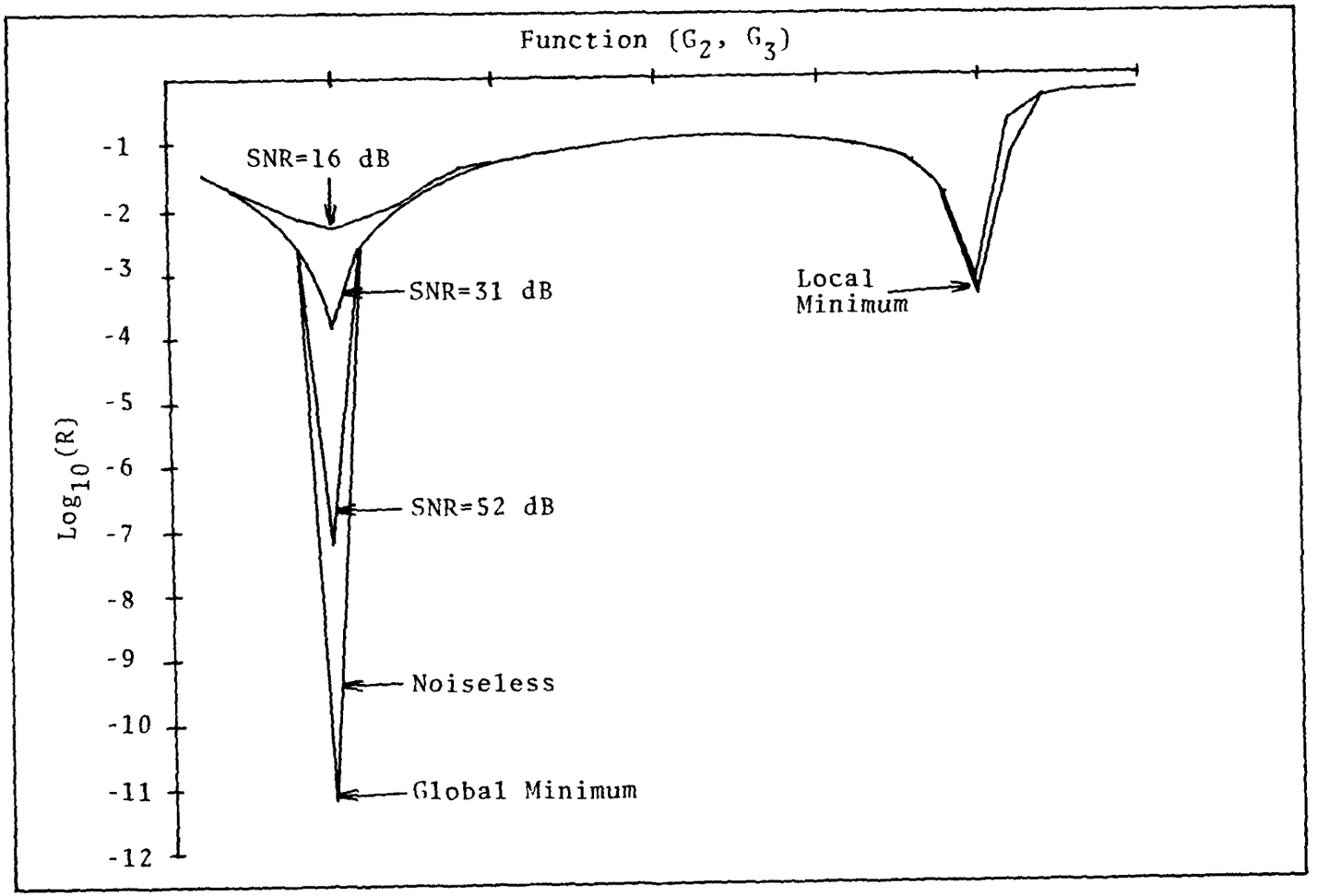


Fig 4-5. Cross Section of Figure 4-3 with Noise

not represent an optical system, but are used to further study the properties of the bilinear equation solver.

The problem associated with Figure 4-7 assumes a maximum spatial frequency component of ±1 for either of the aberrations (Pmax=1). When the complex matrices are converted to real matrices by Eq (4-5), matrix A will be 12X6 after it is combined with an initial value of G, F and G will each be 6X1, and H will be 12X1. This compares with 6X3, 3X1, and 6X1 for A, F, and H, respectively, when all matrices were real. Because all matrices are larger, convergence to the correct solution is much slower as can be seen by comparing Figure 4-7 with BILIN2 in Figure 4-1. Note that in DECON1, the output is only printed at every fourth pair of iterations. LL is the iteration number.

The correct values for G and F associated with Figure 4-7 are G=[(.316, 0.), (.8944, -.316), (0., 0.)] and F=[(0., 0.), (.7746, 0.), (-.4470, .4470)]. The initial guess for G is G=[(.2, 0.), (.9, -.4), (0., 0.)]. Note that after 71 iterations, the sum of the squared errors has dropped to $.946X10^{-6}$, and the algorithm has converged to an answer very close to the correct answer.

Effect of Noisy Measurements

In Figure 4-8, the problem is the same as in Figure 4-6 except that a noise vector has been added to H to see what effect noisy measurements would have on the algorithm. The noise vector N was generated by a zero mean, Gaussian

random number generator with a standard deviation of σ=.01, which is about 3% of the average magnitude of the elements of the H vector. The value of N is N = [(-.0158, -.0058), (.0039, -.0033), (.0092, -.0088), (-.0012, -.0075), (-.0008, -.0170), (.0052, .0010)]. $|N|^2 = .847X10^{-3}$ and $\sum|F'-F|^2 + \sum|G'-G|^2 = 1.264X10^{-3}$, where F' and G' are the values computed by DECON1 and F and G are the actual values. The squared error in H is $|N|^2$. This value is close to the squared error of F' and G', giving some assurance that the algorithm is not overly sensitive to measurement noise. Note that R appears to stop decreasing at a value of $.124X10^{-3}$ in the case of noisy H as compared to $.946X10^{-6}$ in the noiseless case. This is because the equations with noisy H are no longer entirely consistent.

Effect of Increased Dimensionality

Figures 4-9 and 4-10 show the computer output of DECON1 when Pmax=2 and Pmax=4, respectively. H is noiseless in both cases. These figures are included to show that as the dimensionality of the problem increases, the number of iterations required to achieve a squared error of $R \approx 10^{-6}$ increases somewhat. More importantly, the computer execution time went from one second to four seconds to sixteen seconds for Pmax=1, 2, and 4, respectively, indicating that execution time increases as the square of Pmax. Of course, there are other factors that affect the number of iterations and the execution time required to achieve a

given R such as how close the initial guess is for G, so an accurate comparison of these factors in Figures 4-7, 4-9, and 4-10 is not possible. Still, the three computer runs give considerable insight into the behavior of the algorithm as the dimensionality of the problem increases.

In Figure 4-9, the correct values are

G=[(0,0),(.316,0),(.8944,-.316),(0,0),(0,0)]

F=[(0,0),(0,0),(.7746,0),(-.447,.447),(0,0)]

In Figure 4-10, the correct values are

G=[(0,0),(0,0),(0,0),(.3162,0),(-.7071,.3162),
(0,-.3162),(0,.3162),(0,0),(.3162,0)]

F=[(0,0),(0,-.3162),(0,0),(-.4472,.3162),(.7746,0),
(0,0),(0,0),(0,0),(0,0)]

and the initial guess was

G=[(0,0),(0,0),(0,0),(.4,0)(-.6,.3),(0,-.1),(0,.4),
(0,0),(.2,0)].

Summary

The purpose of this chapter has been to present a novel method for solving Eq (2-30) and to exercise the solution to examine its properties. It was shown that the solution scheme is a downhill iterative method that converges to the least-squared-error solution. However, it may converge to a local minimum instead of the global minimum. The scheme converges to a minimum determined by the initial guess for one of the Fourier coefficient vectors F or G. The closer the initial guess is to the

actual value, the more probable it is that the iterative algorithm will converge to the global minimum.

It was shown that noisy measurements cause the least squared error to be larger than in the noiseless case because the noise causes the equations to become less consistant. The iterative scheme does not appear to be especially sensitive to noise.

Finally, it was shown that increasing the dimensionality of the bilinear system of Eq (2-30) drastically slows the convergence of the iterative algorithm. This places limitations on the complexity of the aberrations that can be found.

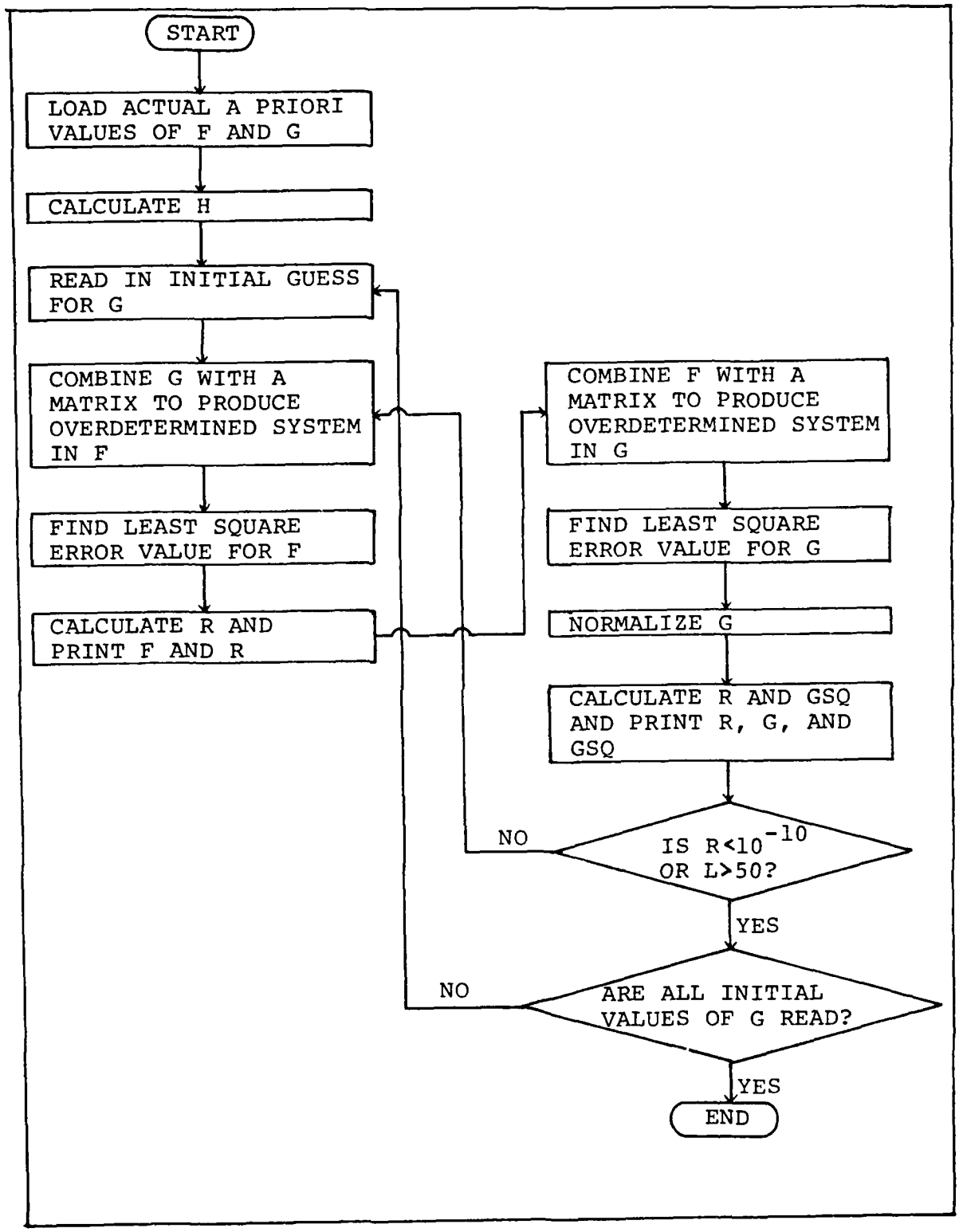


Fig 4-6. Flowchart for BILIN2

```
          -1                  0                  1
H=(  .2448,0.0000)  (  .5516,-.1035)  (-.2585,  .5410)
H=(  .2328,  .0756)  (  .6615,-.0475)  (  .2855,  .5273)
F=(  .0664,-.0180)  (  .7304,0.0000)  (-.4144,  .4756)
                                        LL=  7 R=.140E-02 TOL=.421E+00
G=(  .2522,  .0212)  (  .9249,-.2826)  (-.0178,-.0192)
                                        LL=  8 R=.123E-02 TOL=.183E+00
F=(  .0392,-.0106)  (  .7488,0.0000)  (-.4285,  .4721)
                                        LL= 15 R=.516E-03 TOL=.387E+00
G=(  .2792,  .0134)  (  .9119,-.2994)  (-.0178,-.0188)
                                        LL= 16 R=.459E-03 TOL=.153E+00
F=(  .0239,-.0065)  (  .7589,0.0000)  (-.4362,  .4653)
                                        LL= 23 R=.207E-03 TOL=.377E+00
G=(  .2939,  .0085)  (  .9049,-.3072)  (-.0135,-.0143)
                                        LL= 24 R=.185E-03 TOL=.147E+00
F=(  .0149,-.0041)  (  .7648,0.0000)  (-.4405,  .4594)
                                        LL= 31 R=.848E-04 TOL=.373E+00
G=(  .3023,  .0054)  (  .9009,-.3111)  (-.0092,-.0099)
                                        LL= 32 R=.759E-04 TOL=.145E+00
```

(Continued)

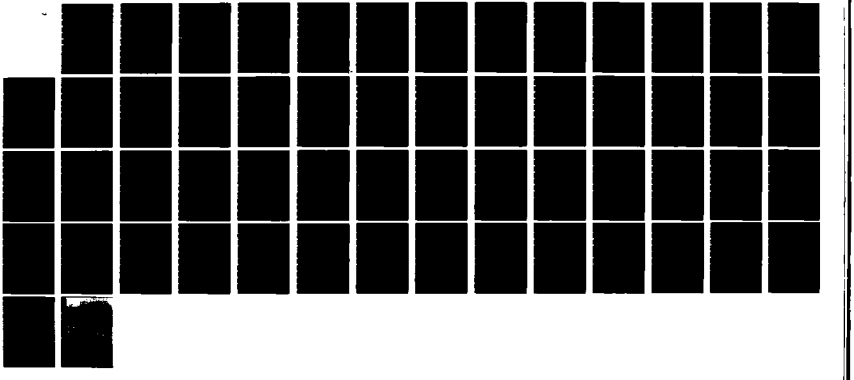
(Continued)

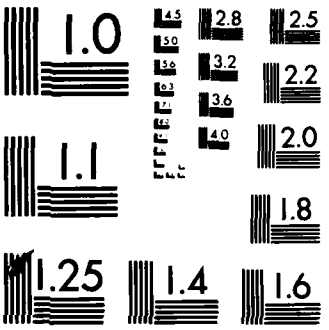
```
F=( .0094,-.0026) ( .7684,-.0000) (-.4430, .4552)
                                          LL= 39 R=.347E-04 TOL=.372E+00
G=( .3075, .0035) ( .8985,-.3131) (-.0061,-.0066)
                                          LL= 40 R=.310E-04 TOL=.145E+00
F=( .0059,-.0017) ( .7706,-.0000) (-.4445, .4523)
                                          LL= 47 R=.142E-04 TOL=.372E+00
G=( .3106, .0022) ( .8970,-.3143) (-.0039,-.0043)
                                          LL= 48 R=.127E-04 TOL=.144E+00
F=( .0038,-.0011) ( .7720,0.0000) (-.4454, .4504)
                                          LL= 55 R=.576E-05 TOL=.372E+00
G=( .3126, .0014) ( .8961,-.3150) (-.0025,-.0028)
                                          LL= 56 R=.514E-05 TOL=.144E+00
F=( .0024,-.0007) ( .7729,-.0000) (-.4460, .4491)
                                          LL= 63 R=.234E-05 TOL=.372E+00
G=( .3139, .0009) ( .8956,-.3154) (-.0016,-.0018)
                                          LL= 64 R=.209E-05 TOL=.144E+00
F=( .0015,-.0004) ( .7735,0.0000) (-.4463, .4483)
                                          LL= 71 R=.946E-06 TOL=.372E+00
```

Fig 4-7. DECON1 Output, P_{MAX} = 1, Noiseless

AD-A138 096 DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING C S DAVIS 09 DEC 83 AFIT/DS/EE/83-1 2/2

UNCLASSIFIED F/G 12/1 NL





	-1	0	1	
H=	(.2290,-.0058)	(.5555,-.1068)	(-.2493, .5322)	
H=	(.2316, .0681)	(.6607,-.0645)	(.2907, .5283)	
F=	(.0619,-.0248)	(.7294,0.0000)	(-.4150, .4638)	
				LL= 7 R=.132E-02 TOL=.431E+00
G=	(.2436, .0203)	(.9262,-.2868)	(-.0021,-.0136)	
				LL= 8 R=.118E-02 TOL=.181E+00
F=	(.0387,-.0185)	(.7455,-.0000)	(-.4299, .4570)	
				LL= 15 R=.589E-03 TOL=.402E+00
G=	(.2673, .0131)	(.9154,-.3004)	(.0040,-.0119)	
				LL= 16 R=.539E-03 TOL=.174E+00
F=	(.0252,-.0146)	(.7547,0.0000)	(-.4379, .4492)	
				LL= 23 R=.314E-03 TOL=.394E+00
G=	(.2805, .0083)	(.9093,-.3070)	(.0105,-.0073)	
				LL= 24 R=.294E-03 TOL=.150E+00
F=	(.0171,-.0122)	(.7602,0.0000)	(-.4425, .4430)	
				LL= 31 R=.203E-03 TOL=.391E+00
G=	(.2883, .0051)	(.9056,-.3105)	(.0155,-.0031)	
				LL= 32 R=.194E-03 TOL=.149E+00
F=	(.0120,-.0106)	(.7636,0.0000)	(-.4451, .4387)	
				LL= 39 R=.156E-03 TOL=.388E+00
G=	(.2931, .0031)	(.9034,-.3123)	(.0189, .0001)	
				LL= 40 R=.153E-03 TOL=.149E+00
F=	(.0088,-.0096)	(.7658,0.0000)	(-.4466, .4359)	
				LL= 47 R=.137E-03 TOL=.381E+00
G=	(.2961, .0018)	(.9020,-.3134)	(.0212, .0023)	
				LL= 48 R=.136E-03 TOL=.148E+00

(Continued)

(Continued)

```
F=( .0068,-.0089) ( .7671,0.0000) (-.4476, .4340)
                                   LL= 55 R=.130E-03 TOL=.376E+00
G=( .2980, .0010) ( .9011,-.3141) ( .0226, .0038)
                                   LL= 56 R=.129E-03 TOL=.144E+00
F=( .0055,-.0085) ( .7680, .0000) (-.4481, .4327)
                                   LL= 63 R=.126E-03 TOL=.373E+00
G=( .2992, .0004) ( .9006,-.3145) ( .0235, .0047)
                                   LL= 64 R=.126E-03 TOL=.144E+00
F=( .0047,-.0082) ( .7685,0.0000) (-.4485, .4320)
                                   LL= 71 R=.125E-03 TOL=.371E+00
G=( .3000, .0001) ( .9002,-.3147) ( .0241, .0054)
                                   LL= 72 R=.125E-03 TOL=.144E+00
F=( .0042,-.0081) ( .7689,0.0000) (-.4487, .4315)
                                   LL= 79 R=.125E-03 TOL=.370E+00
G=( .3005,-.0001) ( .9000,-.3149) ( .0245, .0058)
                                   LL= 80 R=.125E-03 TOL=.144E+00
F=( .0038,-.0080) ( .7691,0.0000) (-.4488, .4311)
                                   LL= 87 R=.124E-03 TOL=.369E+00
G=( .3008,-.0003) ( .8998,-.3150) ( .0247, .0060)
                                   LL= 88 R=.124E-03 TOL=.144E+00
F=( .0036,-.0079) ( .7693,0.0000) (-.4489, .4309)
                                   LL= 95 R=.124E-03 TOL=.369E+00
G=( .3010,-.0004) ( .8997,-.3150) ( .0248, .0062)
                                   LL= 96 R=.124E-03 TOL=.144E+00
```

Fig 4-8. DECON1 Output, P_{MAX} = 1, With Noise

-2	-1	0	1	2
H=(0.0000,0.0000)	(.2448,0.0000)	(.5516,-.1035)	(-.2585, .5410)	(0.0000,0.0000)
H=(0.0000,0.0000)	(.2328, .0756)	(.6615,-.0475)	(.2855, .5273)	(0.0000,0.0000)
F=(-.0035, .0048)	(.0662,-.0186)	(.7273,-.0000)	(-.4124, .4758)	(-.0007,-.0047)
		LL= 7 R=.125E-02 TOL=.348E+00		
G=(-.0179,-.0103)	(.2405, .0306)	(.9306,-.2724)	(-.0081,-.0180)	(-.0118, .0032)
		LL= 8 R=.110E-02 TOL=.885E-01		
F=(.0001, .0042)	(.0415,-.0119)	(.7445,-.0000)	(-.4256, .4756)	(-.0033,-.0078)
		LL= 15 R=.494E-03 TOL=.312E+00		
G=(-.0146,-.0080)	(.2676, .0222)	(.9178,-.2906)	(-.0117,-.0212)	(-.0110, .0039)
		LL= 16 R=.446E-03 TOL=.934E-01		
F=(.0003, .0032)	(.0275,-.0081)	(.7547,0.0000)	(-.4334, .4719)	(-.0044,-.0083)
		LL= 23 R=.231E-03 TOL=.299E+00		
G=(-.0105,-.0056)	(.2839, .0157)	(.9098,-.3012)	(-.0118,-.0197)	(-.0088, .0034)
		LL= 24 R=.212E-03 TOL=.931E-01		
F=(.0003, .0024)	(.0188,-.0057)	(.7610,0.0000)	(-.4381, .4675)	(-.0044,-.0075)
		LL= 31 R=.119E-03 TOL=.294E+00		
G=(-.0075,-.0039)	(.2941, .0112)	(.9048,-.3071)	(-.0104,-.0167)	(-.0068, .0027)
		LL= 32 R=.110E-03 TOL=.966E-01		
F=(.0002, .0017)	(.0132,-.0041)	(.7650,0.0000)	(-.4409, .4633)	(-.0039,-.0064)
		LL= 39 R=.645E-04 TOL=.292E+00		
G=(-.0054,-.0027)	(.3006, .0081)	(.9016,-.3105)	(-.0087,-.0136)	(-.0052, .0021)
		LL= 40 R=.599E-04 TOL=.966E-01		
F=(.0002, .0013)	(.0095,-.0030)	(.7677, .0000)	(-.4427, .4597)	(-.0032,-.0051)
		LL= 47 R=.360E-04 TOL=.292E+00		
G=(-.0040,-.0020)	(.3050, .0059)	(.8995,-.3125)	(-.0070,-.0107)	(-.0040, .0016)
		LL= 48 R=.335E-04 TOL=.967E-01		

(Continued)

(Continued)

```
F=( .0001, .0010) ( .0069,-.0022) ( .7695,0.0000) (-.4439, .4568) (-.0025,-.0041)
                                                  LL= 55 R=.204E-04 TOL=.293E+00
G=(-.0029,-.0015) ( .3080, .0043) ( .8981,-.3137) (-.0055,-.0084) (-.0030, .0012)
                                                  LL= 56 R=.190E-04 TOL=.969E-01
F=( .0001, .0007) ( .0051,-.0016) ( .7708,0.0000) (-.4447, .4545) (-.0020,-.0032)
                                                  LL= 63 R=.116E-04 TOL=.293E+00
G=(-.0022,-.0011) ( .3101, .0032) ( .8972,-.3144) (-.0043,-.0064) (-.0023, .0009)
                                                  LL= 64 R=.108E-04 TOL=.970E-01
F=( .0001, .0005) ( .0038,-.0012) ( .7718,0.0000) (-.4453, .4527) (-.0015,-.0024)
                                                  LL= 71 R=.662E-05 TOL=.294E+00
G=(-.0016,-.0008) ( .3116, .0024) ( .8965,-.3149) (-.0033,-.0049) (-.0017, .0007)
                                                  LL= 72 R=.618E-05 TOL=.972E-01
F=( .0001, .0004) ( .0028,-.0009) ( .7725,0.0000) (-.4457, .4513) (-.0012,-.0019)
                                                  LL= 79 R=.379E-05 TOL=.295E+00
G=(-.0012,-.0006) ( .3128, .0018) ( .8960,-.3152) (-.0025,-.0038) (-.0013, .0005)
                                                  LL= 80 R=.353E-05 TOL=.973E-01
F=( .0001, .0003) ( .0021,-.0007) ( .7730, .0000) (-.4460, .4503) (-.0009,-.0014)
                                                  LL= 87 R=.217E-05 TOL=.295E+00
G=(-.0009,-.0004) ( .3136, .0014) ( .8956,-.3155) (-.0019,-.0029) (-.0010, .0004)
                                                  LL= 88 R=.202E-05 TOL=.974E-01
F=( .0000, .0002) ( .0016,-.0005) ( .7733,0.0000) (-.4463, .4495) (-.0007,-.0011)
                                                  LL= 95 R=.124E-05 TOL=.295E+00
G=(-.0007,-.0003) ( .3142, .0010) ( .8954,-.3156) (-.0015,-.0022) (-.0008, .0003)
                                                  LL= 96 R=.115E-05 TOL=.975E-01
```

Fig 4-9. DECON1 Output, P_{MAX} = 2, Noiseless

```
          -4                -3                -2
          -1                 0                 1
           2                 3                 4

H=(0.0000,-.1000) ( .1000, .2236) (-.2414, .1000)
  ( .5611,-.3650) (-.4477, .3863) (-.1000,-.4863)
  (0.0000, .2449) (-.1414, .1000) ( .2449,0.0000)

H=( .0001, .1000) (-.1644,-.1816) (-.0580,-.1624)
  ( .5966, .1168) (-.5499, .4181) (-.0271,-.3984)
  ( .1439, .1982) (-.1036, .1388) ( .0758,-.2329)

F=( .0147, .0081) ( .0233,-.3125) ( .0162, .0298)
  (-.4529, .2867) ( .7442,0.0000) (-.0873, .0575)
  (-.0071, .0251) (-.0050,-.0155) (-.0073,-.0106)

                                  LL=  7 R=.237E-02 TOL=.871E-01
G=( .0029,-.0024) ( .0082,-.0057) ( .0035, .0108)
  ( .3265,-.0075) (-.6937, .3543) (-.0201,-.2307)
  ( .0221, .3379) ( .0287, .0387) ( .3401,-.0065)

                                  LL=  8 R=.192E-02 TOL=.830E-01
F=( .0019, .0098) ( .0144,-.3092) ( .0190, .0125)
  (-.4601, .2861) ( .7543,0.0000) (-.0487, .0285)
  (-.0024, .0172) (-.0010,-.0032) (-.0030,-.0046)

                                  LL= 15 R=.572E-03 TOL=.881E-01
G=( .0029,-.0004) ( .0025,-.0000) ( .0024, .0076)
  ( .3074,-.0130) (-.7060, .3366) (-.0227,-.2702)
  ( .0109, .3299) ( .0197, .0272) ( .3315,-.0045)

                                  LL= 16 R=.495E-03 TOL=.766E-01
F=(-.0023, .0091) ( .0102,-.3080) ( .0172, .0054)
  (-.4608, .2899) ( .7595,0.0000) (-.0306, .0157)
  (-.0030, .0104) (-.0007, .0002) (-.0007,-.0024)

                                  LL= 23 R=.214E-03 TOL=.895E-01
G=( .0024, .0001) ( .0004, .0015) ( .0022, .0044)
  ( .3042,-.0134) (-.7086, .3283) (-.0193,-.2881)
  ( .0053, .3268) ( .0141, .0203) ( .3266,-.0026)

                                  LL= 24 R=.194E-03 TOL=.772E-01
F=(-.0033, .0077) ( .0075,-.3084) ( .0145, .0024)
  (-.4593, .2945) ( .7629,0.0000) (-.0209, .0096)
  (-.0031, .0067) (-.0007, .0013) ( .0001,-.0013)

                                  LL= 31 R=.107E-03 TOL=.972E-01
G=( .0019, .0002) (-.0003, .0018) ( .0020, .0028)
  ( .3050,-.0118) (-.7090, .3241) (-.0155,-.2974)
  ( .0028, .3245) ( .0105, .0154) ( .3237,-.0016)

                                  LL= 32 R=.996E-04 TOL=.824E-01
```

(Continued)

(Continued)

LL= 40 R=.575E-04 TOL=.897E-01

F=(-.0028, .0050) (.0044,-.3107) (.0094, .0006)
(-.4552, .3023) (.7673,0.0000) (-.0113, .0046)
(-.0023, .0034) (-.0006, .0013) (.0004,-.0005)

LL= 47 R=.367E-04 TOL=.911E-01

G=(.0012, .0002) (-.0005, .0014) (.0013, .0014)
(.3086,-.0078) (-.7086, .3203) (-.0096,-.3064)
(.0010, .3213) (.0061, .0092) (.3205,-.0007)

LL= 48 R=.345E-04 TOL=.726E-01

F=(-.0023, .0040) (.0034,-.3118) (.0074, .0003)
(-.4535, .3052) (.7689,0.0000) (-.0086, .0034)
(-.0019, .0025) (-.0005, .0011) (.0003,-.0003)

LL= 55 R=.222E-04 TOL=.944E-01

G=(.0009, .0002) (-.0005, .0011) (.0011, .0011)
(.3102,-.0062) (-.7083, .3193) (-.0074,-.3088)
(.0007, .3201) (.0047, .0071) (.3195,-.0005)

LL= 56 R=.209E-04 TOL=.694E-01

F=(-.0018, .0031) (.0026,-.3127) (.0058, .0002)
(-.4522, .3076) (.7701,0.0000) (-.0067, .0026)
(-.0015, .0019) (-.0004, .0009) (.0003,-.0002)

LL= 63 R=.135E-04 TOL=.930E-01

G=(.0007, .0001) (-.0004, .0009) (.0008, .0008)
(.3115,-.0049) (-.7081, .3186) (-.0058,-.3105)
(.0005, .3193) (.0037, .0055) (.3188,-.0004)

LL= 64 R=.127E-04 TOL=.792E-01

F=(-.0014, .0024) (.0021,-.3134) (.0046, .0001)
(-.4511, .3094) (.7710,0.0000) (-.0052, .0020)
(-.0012, .0015) (-.0003, .0007) (.0002,-.0002)

LL= 71 R=.821E-05 TOL=.926E-01

G=(.0006, .0001) (-.0003, .0007) (.0007, .0007)
(.3125,-.0038) (-.7079, .3181) (-.0045,-.3118)
(.0004, .3186) (.0028, .0043) (.3182,-.0003)

LL= 72 R=.771E-05 TOL=.781E-01

F=(-.0011, .0019) (.0016,-.3140) (.0036, .0001)
(-.4502, .3109) (.7718, .0000) (-.0040, .0016)
(-.0009, .0012) (-.0002, .0006) (.0002,-.0001)

F=(-.0032, .0063) (.0057,-.3095) (.0117, .0011)
(-.4572, .2987) (.7654,0.0000) (-.0151, .0065)
(-.0028, .0046) (-.0007, .0014) (.0003,-.0008)

LL= 39 R=.614E-04 TOL=.867E-01

G=(.0015, .0002) (-.0005, .0016) (.0016, .0019)
(.3068,-.0097) (-.7089, .3217) (-.0122,-.3029)
(.0016, .3227) (.0080, .0119) (.3218,-.0010)

(Continued)

(Continued)

```
                                        LL= 79 R=.499E-05 TOL=.922E-01
G=( .0004, .0001) (-.0002, .0005) ( .0005, .0005)
  ( .3133,-.0030) (-.7077, .3177) (-.0035,-.3128)
  ( .0003, .3180) ( .0022, .0033) ( .3178,-.0003)

                                        LL= 80 R=.469E-05 TOL=.673E-01
F=(-.0009, .0015) ( .0013,-.3145) ( .0028, .0001)
  (-.4496, .3121) ( .7724,0.0000) (-.0031, .0012)
  (-.0007, .0009) (-.0002, .0004) ( .0001,-.0001)

                                        LL= 87 R=.303E-05 TOL=.919E-01
G=( .0003, .0001) (-.0002, .0004) ( .0004, .0004)
  ( .3140,-.0023) (-.7076, .3173) (-.0027,-.3136)
  ( .0002, .3176) ( .0017, .0026) ( .3174,-.0002)

                                        LL= 88 R=.285E-05 TOL=.672E-01
F=(-.0007, .0012) ( .0010,-.3149) ( .0022, .0000)
  (-.4490, .3130) ( .7729, .0000) (-.0024, .0010)
  (-.0006, .0007) (-.0001, .0004) ( .0001,-.0001)

                                        LL= 95 R=.194E-05 TOL=.916E-01
G=( .0003, .0001) (-.0002, .0003) ( .0003, .0003)
  ( .3145,-.0018) (-.7075, .3171) (-.0021,-.3142)
  ( .0002, .3173) ( .0013, .0020) ( .3172,-.0002)

                                        LL= 96 R=.173E-05 TOL=.671E-01
F=(-.0005, .0009) ( .0008,-.3152) ( .0017, .0000)
  (-.4486, .3137) ( .7732, .0000) (-.0019, .0008)
  (-.0004, .0006) (-.0001, .0003) ( .0001,-.0001)

                                        LL=103 R=.112E-05 TOL=.915E-01
G=( .0002, .0000) (-.0001, .0003) ( .0002, .0003)
  ( .3149,-.0014) (-.7074, .3169) (-.0017,-.3146)
  ( .0001, .3171) ( .0010, .0016) ( .3170,-.0001)

                                        LL=104 R=.105E-05 TOL=.671E-01
```

Fig 4-10. DECON1 Output, P_{MAX}=4, Noiseless

V. Two-Dimensional, Two-Surface Problem

Introduction

Chapters II through IV have considered only the 1-D deconvolution problem. In this chapter it will be shown that when a second dimension is added, the problem does not change in any substantial way. The 1-D development in Chapters II through IV was for convenience, since the 2-D equations are longer and more cumbersome.

In the first part of this chapter, the 2-D equations are developed, with the approach paralleling the 1-D approach. In the last part of the chapter, a 2-D optical system with simple wedges for aberrations is described, and the 2-D deconvolution algorithm is exercised to solve for the aberrations. The example problem is included to show that the deconvolution algorithm works in principle, but because of the computer time and memory required to solve for realistic aberrations, no attempt is made to exercise the algorithm on more general problems.

The 2-D problem is illustrated in Figure 5-1 below. A plane wave enters the optical system from the left with an angle of propagation θ in the x direction and ϕ in the y direction with respect to the Z axis. The plane wave passes through aberrations $W_A(x,y)$ and $W_B(x,y)$, and the amplitude and phase of the resulting field, $U_5(x,y)$, are measured at the measurement plane. All of the assumptions

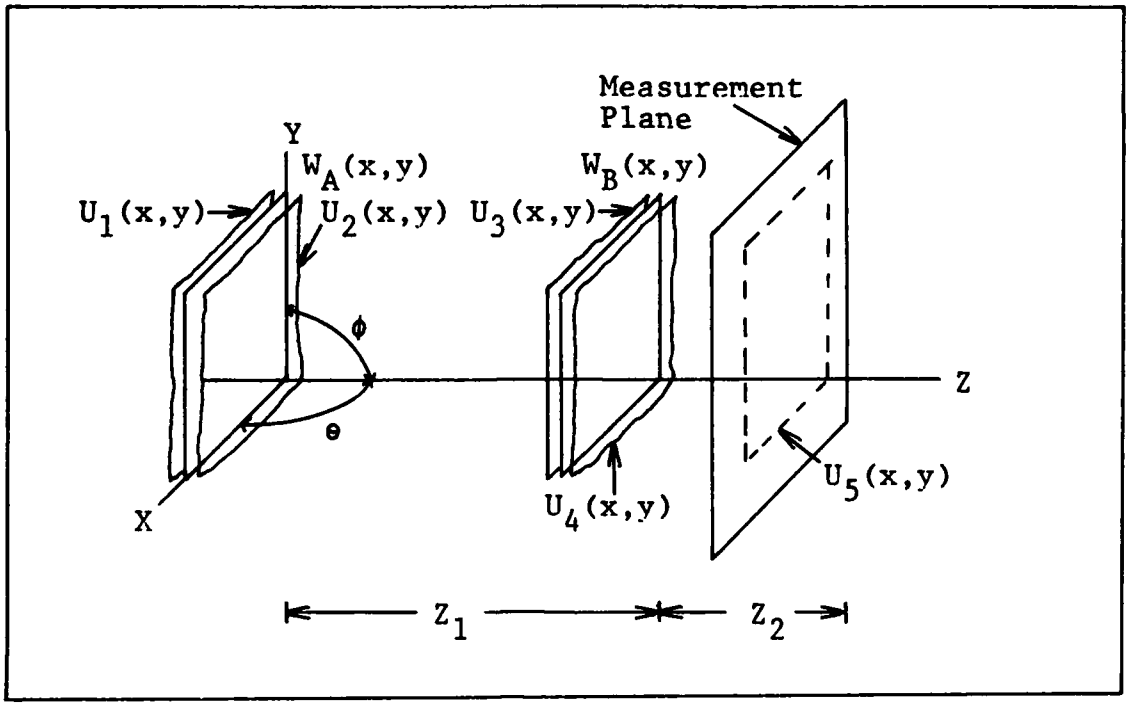


Fig 5-1. Two-Dimensional Deconvolution Problem

listed on the first page of Chapter 2 apply if they are extended to two dimensions.

Some of the key 2-D extensions of 1-D equations listed in Chapter 2 follow. The parallel 1-D equation from Chapters 2 and 3 are listed below each equation number for reference.

$$H(f_x,f_y) = \begin{cases} \exp[jkZ\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}], & f_x^2+f_y^2 < \frac{1}{\lambda^2} \\ 0 \text{ else} \end{cases} \quad \begin{matrix}(5\text{-}1)\\(2\text{-}1)\end{matrix}$$

$$h(x,y) = \frac{e^{jkZ}\, e^{\frac{j\pi(x^2+y^2)}{Z}}}{j\lambda Z} \quad \begin{matrix}(5\text{-}2)\\(2\text{-}3)\end{matrix}$$

$$U_5(x,y) = -\frac{A'e^{jk(Z_1+Z_2)}}{\lambda^2 Z_1 Z_2} \int\int_{\sum P_A}\int\int_{\sum P_B} U(\gamma,\xi) e^{jkW_A(\gamma,\xi)}$$

$$e^{jkW_B(\beta,\eta)}\ e^{j\frac{\pi}{\lambda Z_1}[(\beta-\gamma)^2+(\xi-\eta)^2]}$$

$$e^{j\frac{\pi}{\lambda Z_2}[(x-\beta)^2+(y-\eta)^2]}\ d\gamma\ d\xi\ d\beta\ d\eta \qquad \begin{matrix}(5\text{-}3)\\(2\text{-}10)\end{matrix}$$

$$e^{jkW_A(x,y)} = \sum_{i=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} F_{im} e^{j2\pi(\frac{ix}{a_x}+\frac{my}{a_y})} \qquad \begin{matrix}(5\text{-}4)\\(2\text{-}21)\end{matrix}$$

$$e^{jkW_B(x,y)} = \sum_{l=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} G_{ln} e^{j2\pi(\frac{lx}{b_x}+\frac{ny}{b_y})} \qquad \begin{matrix}(5\text{-}5)\\(2\text{-}22)\end{matrix}$$

$$U_5(x,y) = \frac{-A'e^{jk(Z_1+Z_2)}}{\lambda^2 Z_1 Z_2} \sum_i\sum_m\sum_l\sum_n F_{im}\ G_{ln}$$

$$\int_0^{a_x}\int_0^{b_x} e^{jk\sin\theta\gamma}\ e^{j2\pi(\frac{i\gamma}{a_x}+\frac{l\beta}{b_x})}\ e^{j\frac{\pi}{\lambda Z_1}(\gamma-\beta)^2}$$

$$e^{\frac{j\pi}{\lambda Z_2}(\beta-x)^2}\ d\gamma\ d\beta \int_0^{a_y}\int_0^{b_y} e^{jk\sin\phi\xi}\ e^{j2\pi(\frac{m\xi}{a_y}+\frac{n\eta}{b_y})}$$

$$e^{\frac{j\pi}{\lambda Z_1}(\xi-\eta)^2}\ e^{\frac{j\pi}{\lambda Z_2}(\eta-y)^2}\ d\xi\ d\eta \qquad (5\text{-}6)$$

Since the quadruple integral in Eq (5-3) separates into two double integrals in Eq (5-6), each of the double integrals can be solved using stationary phase. Assuming $a_x=b_x$ and $a_y=b_y$,

$$U_5(x,y) = A' e^{jk(Z_1+Z_2)}\ e^{jk(x\sin\theta+y\sin\phi)}$$

$$e^{-j\frac{k}{2}(Z_1+Z_2)(\sin^2\theta+\sin^2\phi)} \sum_i\sum_m\sum_l\sum_n F_{im}\ G_{ln}$$

$$\exp\left[j2\pi[x(\frac{i+1}{a_x})+y(\frac{m+n}{a_y})]\right]\ \exp\left[-j2\pi[(Z_1+Z_2)\right.$$

$$\left.(\frac{i}{a_x}\sin\theta + \frac{m}{a_y}\sin\phi) + Z_2(\frac{l}{a_x}\sin\theta + \frac{n}{a_y}\sin\phi)]\right] \qquad (5\text{-}7)$$

Combining the three known exponential terms with $U_5(x,y)$,

$$U(x,y) = \sum_i\sum_m\sum_l\sum_n F_{im}\ G_{ln}\ \exp\left[j2\pi[x(\frac{i+1}{a_x}) + y(\frac{m+n}{a_y})]\right]$$

$$\exp\left[-j2\pi[(Z_1+Z_2)(\frac{i}{a_x}\sin\theta + \frac{m}{a_y}\sin\phi) + \right.$$

$$\left. Z_2(\frac{l}{a_x}\sin\theta + \frac{n}{a_y}\sin\phi)]\right] \qquad \begin{matrix}(5\text{-}8)\\(2\text{-}24)\end{matrix}$$

If U(x,y) is known only at a finite number of points on an equidistant C by D grid rather than on a continuum, the DFT of U(x,y) can be taken. C, D, c, and d are all

integers. Let $x=c\Delta x$ and $y=d\Delta y$ where Δx and Δy are the distances between grid points. Then the measurements are taken over an area at the measurement plane with dimensions a_x by a_y. Note that $\Delta x/a_x = 1/C$ and $\Delta y/a_y = 1/D$. The 2-D DFT of U(x,y) is

$$H(p,q) = \frac{1}{CD} \sum_{c=0}^{C-1} \sum_{d=0}^{D-1} U(c\Delta x, d\Delta y) \exp[-j2\pi(\frac{cp}{C} + \frac{dq}{D})]$$

$$= \frac{1}{CD} \sum_{c=0}^{C-1} \sum_{d=0}^{D-1} \sum_{i}\sum_{m}\sum_{l}\sum_{n} F_{im}\, G_{ln} \exp\Big[j2\pi[\frac{c}{C}(i+l) +$$

$$\frac{d}{D}(m+n) - \frac{cp}{C} - \frac{dq}{D}]\Big] A_J(i,m,l,n) \qquad (5\text{-}9)$$

where

$$A_J(i,m,l,n) = \exp\Big[-j2\pi[(Z_1+Z_2)(\frac{i}{a_x}\sin\theta_J + \frac{m}{a_y}\sin\phi_J) +$$

$$Z_2(\frac{l}{a_x}\sin\theta_J + \frac{n}{a_y}\sin\phi_J)]\Big] \qquad (5\text{-}10)$$

The order of summation in Eq(5-9) can be changed so that

$$H_J(p,q) = \sum_{i}\sum_{m}\sum_{l}\sum_{n} F_{im}\, G_{ln}\, A_J(i,m,l,n)\, \frac{1}{CD}$$

$$\sum_{c=0}^{C-1} \sum_{d=0}^{D-1} \exp\Big[j2\pi[\frac{c}{C}(i+l-p) + \frac{d}{D}(m+n-q)]\Big] \qquad (5\text{-}11)$$

The double summation over c and d in Eq (5-11) is zero unless i+l=p and m+n=q, in which case the double summation equals CD. Therefore,

$$H_J(p,q) = \sum_{\substack{i \\ i+l=p}} \sum_{m} \sum_{l} \sum_{\substack{n \\ m+n=q}} F_{im}\, G_{ln}\, A_J(i,m,l,n) \qquad \begin{matrix}(5\text{-}12)\\(2\text{-}26)\end{matrix}$$

Other bounds on the summation indices are $|i| \leq Pmax$, $|l| \leq Pmax$, $|m| \leq Qmax$, and $|n| \leq Qmax$.

Equation (5-12) specifies the equations in the 2-D bilinear system. There is no basic difference between the 2-D and the 1-D bilinear systems except that the 2-D system is larger. For $-Pmax \leq p \leq Pmax$ and $-Qmax \leq q \leq Qmax$, there are $(2Pmax+1)(2Qmax+1)$ equations for a given plane wave arrival angle. There will be twice that many unknown coefficients, so measurements must be taken for at least two different plane wave arrival angles to yield as many equations as unknowns.

Two-Dimensional Noise Analysis

The 2-D noise analysis is a straight forward extension of the 1-D analysis, so only key equations will be shown. The comparable 1-D equation number will be listed beneath each 2-D equation number as was done previously.

$$\hat{U}_5(x,y) = A'U_5(x,y) + \hat{N}(x,y) \qquad \begin{matrix}(5\text{-}13)\\(3\text{-}1)\end{matrix}$$

where the real and imaginary parts of $\hat{N}(x,y)$ are zero-mean, spatially white, and each has a variance of $\frac{1}{2}\sigma_n^2$.

$$\hat{H}_J(p,q) = e^{-jk(Z_1+Z_2)}\ e^{j\frac{k}{2}(Z_1+Z_2)(\sin^2\theta_J+\sin^2\phi_J)}$$

$$\frac{1}{a_x a_y}\int_0^{a_x}\int_0^{a_y} [A'U_5(x,y) + \hat{N}(x,y))]\ e^{-jk(x\sin\theta_J+y\sin\phi_J)}$$

$$e^{-j2\pi(\frac{px}{a_x}+\frac{qy}{a_y})}\ dy\ dx = H_J(p,q) + \hat{N}_{HJ}(p,q) \qquad \begin{matrix}(5\text{-}14)\\(3\text{-}22)\end{matrix}$$

$$E\ [\hat{N}_{HJ}(p,q)] = 0 \qquad \begin{matrix}(5\text{-}15)\\(3\text{-}3)\end{matrix}$$

$$E\ [\hat{N}_{HJ}(p,q)\hat{N}^*_{HJ}(p'',q'')] = \begin{cases}\sigma^2 \text{ if } p=p'' \text{ and } q=q''\\ 0 \text{ else}\end{cases} \qquad \begin{matrix}(5\text{-}16)\\(3\text{-}7)\end{matrix}$$

$$P_{\hat{H}|Y}(\hat{H}|Y) = \frac{1}{(2\pi)^{\frac{K}{2}}\sigma^K}\ \exp[-\frac{1}{2\sigma^2}(\hat{H}-AY)^\dagger(\hat{H}-AY)] \qquad \begin{matrix}(5\text{-}17)\\(3\text{-}11)\end{matrix}$$

Equations (5-17) and (3-11) are identical, meaning that the conditional probability density function is the same for the 2-D problem and the 1-D problem. Equations (3-11) through (3-16) and the associated discussion apply equally to the 2-D and 1-D noise analysis. The form of Eq (3-17) is slightly different for the 2-D analysis as shown below.

$$\hat{H}_J(p,q) - AY = \hat{H}_J(p,q) - \sum_i \sum_m F_{im}\ G_{p-i,q-m}$$

$$a_J(i,m,p-i,q-m) \qquad \begin{matrix}(5\text{-}18)\\(3\text{-}17)\end{matrix}$$

where "a" is given by equation (5-10).

Equations (3-19) through (3-22) apply to the 2-D problem if F_{0R} and G_{0R} are replaced by F_{00R} and G_{00R}, respectively.

The development of the 2-D FIM follows the same procedure as for the 1-D FIM developed in Appendix A. Therefore, only the result will be given.

$$\mathrm{VAR}[F_{imR}] = \mathrm{VAR}[F_{imI}] \geq$$

$$\frac{\sigma^2 \sum_J \left(2 \sum_l \sum_n \sigma_{Fln}^2 + 1\right) + \frac{\sigma^4}{\sigma_{Gim}^2}}{\left[\left[\sum_J \left(2 \sum_l \sum_n \sigma_{Gln}^2 + 1\right) + \frac{\sigma^2}{\sigma_{Fim}^2}\right]\left[\sum_J \left(2 \sum_l \sum_n \sigma_{Fln}^2 + 1\right) + \frac{\sigma^2}{\sigma_{Gim}^2}\right] - \left[\sum_J \cos 2\pi Z_1 \left(\frac{i}{a_x}\sin\theta_J + \frac{m}{a_y}\sin\phi_J\right)\right]^2 - \left[\sum_J \sin 2\pi Z_1 \left(\frac{i}{a_x}\sin\theta_J + \frac{m}{a_y}\sin\phi_J\right)\right]^2\right]} \qquad \begin{matrix}(5\text{-}19)\\(A\text{-}12)\end{matrix}$$

where $-Pmax+i \leq l \leq Pmax$ for $i \geq 0$,

$-Pmax \leq l \leq Pmax+i$ for $i < 0$,

$-Pmax+m \leq n \leq Pmax$ for $m \geq 0$,

$-Pmax \leq n \leq Pmax+m$ for $m < 0$.

Equation (5-19) has the same form as Eq (A-12), so the bounds for each F and G as given by Eq (5-19) can be summed to yield a 2-D trace as was done for the 1-D problem. For any given optical system the trace may be plotted as a

function of θ and ϕ to yield the optimum difference between measurement angles.

Example Two-Dimensional Deconvolution Problem

The computer program in Appendix C called DECON2 will solve for the coefficients of the Fourier series representation of $e^{jkW_A(x,y)}$ and $e^{jkW_B(x,y)}$. The input information required by DECON2 consists of the amplitude and phase measurements of U(x,y) over a uniform grid of points. The number of measurement points must exceed the product of twice the highest space-bandwidth product in each coordinate direction in order to satisfy the Nyquist criterion. Other input information includes the distance between planes A and B (Z_1) and between plane B and the measurement plane (Z_2), plane wave arrival angles for each set of measurements (θ and ϕ), the dimensions of the aberration surfaces (a_x and a_y) which are assumed to be equal for both planes A and B, and the wavelength (λ). Some required initializing parameters are MA, ND, and LMAX. $NU=2^{MA}$ is the square root of the number of samples to be taken of U(x,y), ND^2 is the number of complex Fourier coefficients used to describe each aberration, and LMAX is the number of different plane wave arrival angles for which U(x,y) is measured (normally 2).

Figure 5-2 is the flowchart for DECON2. It is nearly the same as the flowchart for BILIN2 shown in Figure 4-5 except that the values for U(x,y) are loaded and Fourier

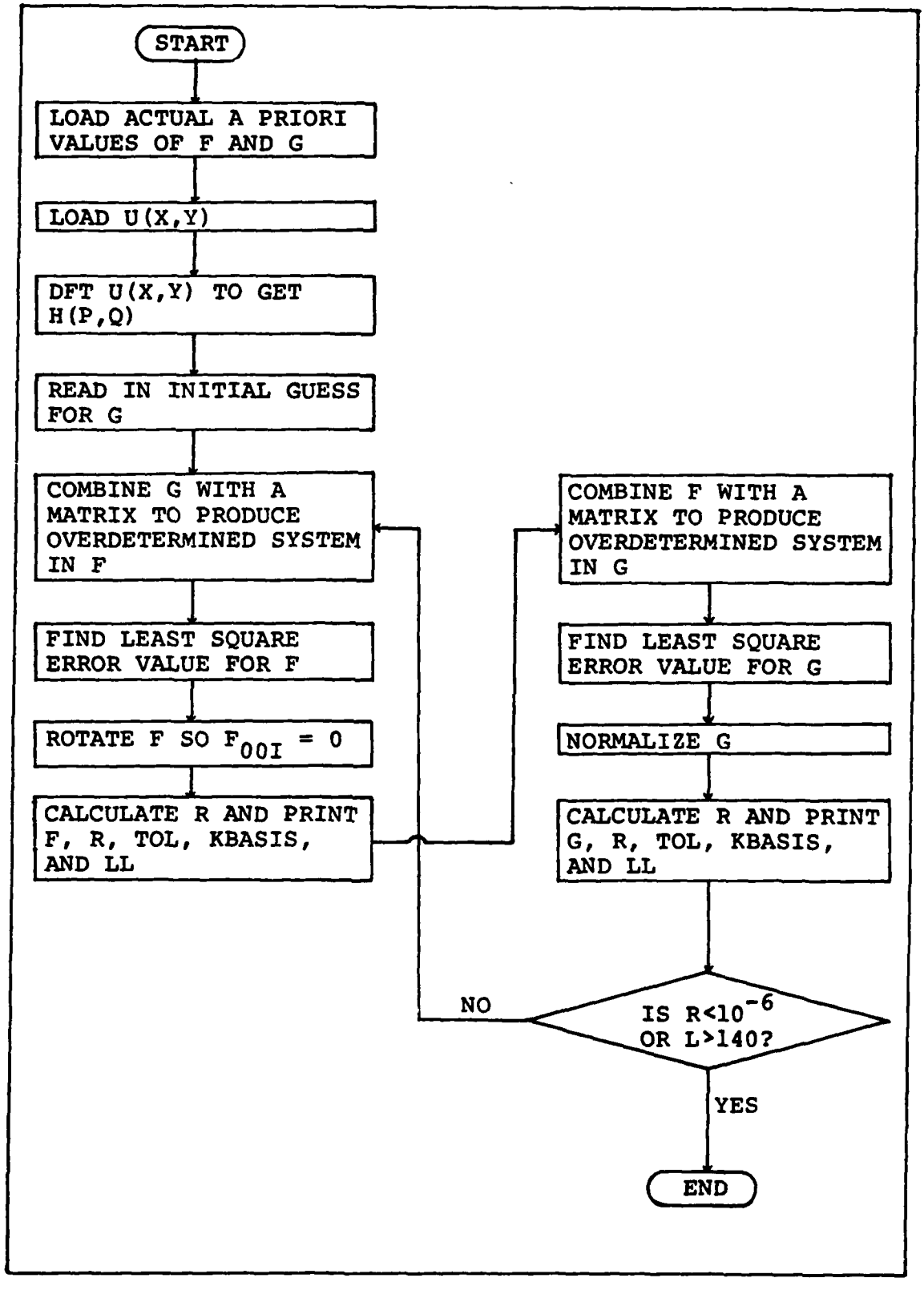


Fig 5-2. Flowchart for DECON2

transformed to get H instead of calculating H from known Fourier coefficients as was done in BILIN2. An added convenience feature is that at every other iteration, the F vector is rotated so that the imaginary part of F_{00} is zero.

In general, an infinite series is required to completely describe an aberration function (see Eq (5-4)). A decision is required as to how many coefficients are sufficient to adequately describe the aberration function. To keep the example 2-D deconvolution problem simple, aberrations were chosen which can be exactly described with a single coefficient for each aberration.

Let $W_A(x,y)=\lambda x$, $W_B(x,y)=\lambda y$, $a_x=a_y=1$, $\lambda=10^{-6}$, $Z_1=2$, $Z_2=1$, $\theta_1=\phi_1=0$, $\theta_2=\phi_2=.05$. An illustration of the problem is shown in Figure 5-3. If refraction at the aberration surfaces is ignored, which is reasonable in this problem, it can be shown that

$$U(x,y) = \exp j2\pi[x-(Z_1+Z_2)\sin\theta + y-Z_2\sin\phi] \qquad (5\text{-}20)$$

This can be seen by picking a point (x,y) at the measurement plane and backward ray tracing through the aberrations for a ray arriving at angle (θ,ϕ). The ray will encounter a phase delay of $2\pi(y-Z_2\sin\phi)$ at plane B and $2\pi[x-(Z_1+Z_2)\sin\theta]$ at plane A. Note that propagation phase terms do not appear in Eq (5-20) because they are

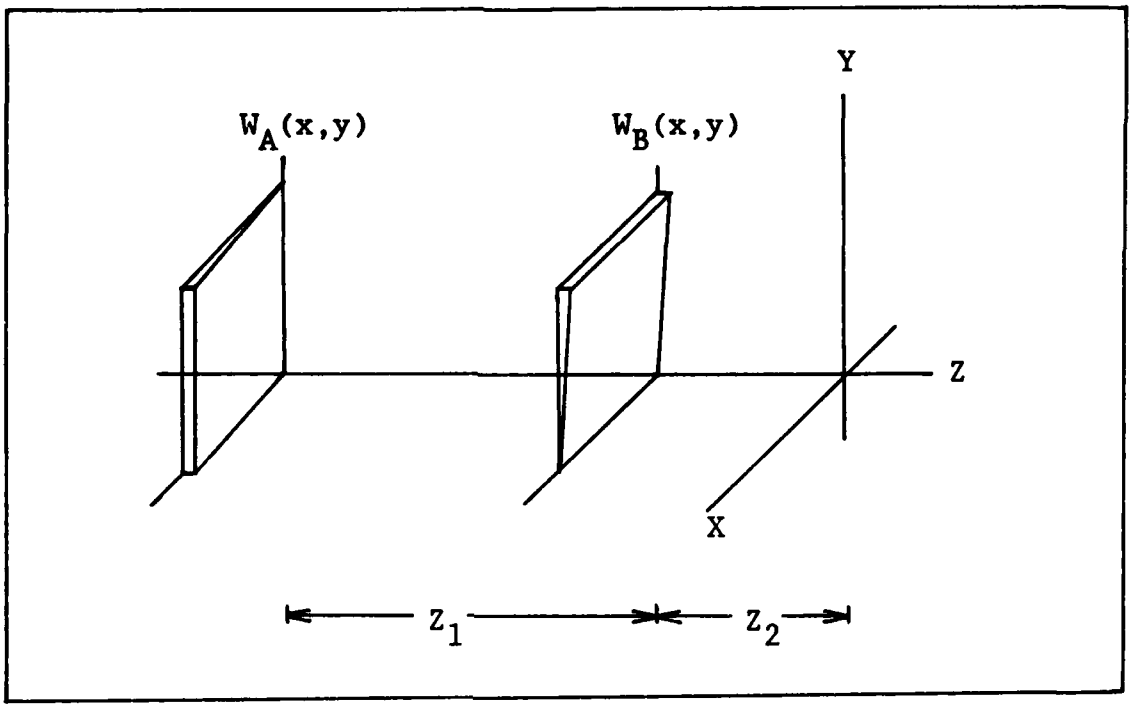


Fig 5-3. Example 2-D Deconvolution Problem

known phase terms which are combined with $U_5(x,y)$ to yield $U(x,y)$.

Equation (5-20) is used by DECON2 to generate $U(x,y)$ over an 8X8 grid for each of the two plane wave arrival angles. These values are shown in Figure 5-4. The values of $U(x,y)$ are then Fourier transformed to yield the values of the H matrix as shown at the beginning of Figure 5-5. An initial guess for G is loaded as follows:

$G_{0,0}$=(.3162, 0), $G_{0,1}$=(.9487, 0)

and all other values of $G_{1,n}$ are zero.

The program now begins iterating. It will iterate a maximum of 140 times or until the sum of the squared errors, R, drops below 10^{-6}. At every fifth pair of

$\theta_1=\phi_2=0.0$

X=0.000	REAL U	1.000	.707	-.000	-.707	-1.000	-.707	.000	.707
	IMAG U=	0.000	.707	1.000	.707	-.000	-.707	-1.000	-.707
X= .125	REAL U	.707	-.000	-.707	-1.000	-.707	.000	.707	1.000
	IMAG U=	.707	1.000	.707	-.000	-.707	-1.000	-.707	.000
X= .250	REAL U	-.000	-.707	-1.000	-.707	.000	.707	1.000	.707
	IMAG U=	1.000	.707	-.000	-.707	-1.000	-.707	.000	.707
X= .375	REAL U	-.707	-1.000	-.707	.000	.707	1.000	.707	-.000
	IMAG U=	.707	-.000	-.707	-1.000	-.707	.000	.707	1.000
X= .500	REAL U	-1.000	-.707	.000	.707	1.000	.707	-.000	-.707
	IMAG U=	-.000	-.707	-1.000	-.707	.000	.707	1.000	.707
X= .625	REAL U	-.707	.000	.707	1.000	.707	-.000	-.707	-1.000
	IMAG U=	-.707	-1.000	-.707	.000	.707	1.000	.707	-.000
X= .750	REAL U	.000	.707	1.000	.707	-.000	-.707	-1.000	-.707
	IMAG U=	-1.000	-.707	.000	.707	1.000	.707	-.000	-.707
X= .875	REAL U	.707	1.000	.707	-.000	-.707	-1.000	-.707	.000
	IMAG U=	-.707	.000	.707	1.000	.707	-.000	-.707	-1.000

$\theta_2=\phi_2=.05$

X=0.000	REAL U	.310	.891	.951	.454	-.310	-.891	-.951	-.454
	IMAG U=	-.951	-.454	.310	.891	.951	.454	-.310	-.891
X= .125	REAL U	.891	.951	.454	-.310	-.891	-.951	-.454	.310
	IMAG U=	-.454	.310	.891	.951	.454	-.310	-.891	-.951
X= .250	REAL U	.951	.454	-.310	-.891	-.951	-.454	.310	.891
	IMAG U=	.310	.891	.951	.454	-.310	-.891	-.951	-.454
X= .375	REAL U	.454	-.310	-.891	-.951	-.454	.310	.891	.951
	IMAG U=	.891	.951	.454	-.310	-.891	-.951	-.454	.310
X= .500	REAL U	-.310	-.891	-.951	-.454	.310	.891	.951	.454
	IMAG U=	.951	.454	-.310	-.891	-.951	-.454	.310	.891
X= .625	REAL U	-.891	-.951	-.454	.310	.891	.951	.454	-.310
	IMAG U=	.454	-.310	-.891	-.951	-.454	.310	.891	.951
X= .750	REAL U	-.951	-.454	.310	.891	.951	.454	-.310	-.891
	IMAG U=	-.310	-.891	-.951	-.454	.310	.891	.951	.454
X= .875	REAL U	-.454	.310	.891	.951	.454	-.310	-.891	-.951
	IMAG U=	-.891	-.951	-.454	.310	.891	.951	.454	-.310

Fig 5-4. Derived Values for U(x,y) in 2-D Example Problem

```
      REAL H= -.000  .000  .000  .000 1.000  .000  .000  .000 -.000
      IMAG H=  .000 -.000 -.000 -.000  .000 -.000 -.000 -.000  .000

      REAL H=  .000 -.000 -.000 -.000  .310 -.000 -.000 -.000 -.000
      IMAG H= -.000 -.000 -.000 -.000 -.951 -.000 -.000 -.000  .000

REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.1558 1.0293 -.1611
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0506  .0000 -.0523
R= .5611E-02     TOL= .2057E-01     KBASIS= 18     LL= 9
REAL G= 0.0000 0.0000 0.0000  .0254  .1527  .9879 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .5063E-02     TOL= .5376E+00     KBASIS=  3     LL= 10
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0949 1.0289 -.1103
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0308  .0000 -.0358
R= .2022E-02     TOL= .1201E-01     KBASIS= 18     LL= 19
REAL G= 0.0000 0.0000 0.0000  .0096  .0927  .9957 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .1828E-02     TOL= .6701E+00     KBASIS=  3     LL= 20
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0566 1.0108 -.0694
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0184  .0000 -.0225
R= .7378E-03     TOL= .8104E-02     KBASIS= 18     LL= 29
REAL G= 0.0000 0.0000 0.0000  .0035  .0561  .9984 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .6672E-03     TOL= .7790E+00     KBASIS=  3     LL= 30
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0341 1.0040 -.0427
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0111  .0000 -.0139
R= .2701E-03     TOL= .4999E-02     KBASIS= 18     LL= 39
REAL G= 0.0000 0.0000 0.0000  .0013  .0340  .9994 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .2443E-03     TOL= .8580E+00     KBASIS=  3     LL= 40
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0206 1.0015 -.0260
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0067  .0000 -.0084
R= .9903E-04     TOL= .3063E-02     KBASIS= 18     LL= 49
```

(Continued)

(Continued)

```
REAL G= 0.0000 0.0000 0.0000  .0005  .0206  .9998 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .8958E-04      TOL= .9110E+00      KBASIS=  3      LL= 50
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0125 1.0005 -.0158
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0041  .0000 -.0051
R= .3632E-04      TOL= .1869E-02      KBASIS= 18      LL= 59
REAL G= 0.0000 0.0000 0.0000  .0002  .0125  .9999 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .3286E-04      TOL= .9450E+00      KBASIS=  3      LL= 60
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0076 1.0002 -.0096
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0025  .0000 -.0031
R= .1332E-04      TOL= .1137E-02      KBASIS= 18      LL= 69
REAL G= 0.0000 0.0000 0.0000  .0001  .0076 1.0000 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .1205E-04      TOL= .9663E+00      KBASIS=  3      LL= 70
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0046 1.0001 -.0058
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0015  .0000 -.0019
R= .4888E-05      TOL= .6904E-03      KBASIS= 18      LL= 79
REAL G= 0.0000 0.0000 0.0000  .0000  .0046 1.0000 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .4421E-05      TOL= .9795E+00      KBASIS=  3      LL= 80
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0028 1.0000 -.0035
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0009  .0000 -.0011
R= .1793E-05      TOL= .4189E-03      KBASIS= 18      LL= 89
REAL G= 0.0000 0.0000 0.0000  .0000  .0028 1.0000 0.0000 0.0000 0.0000
IMAG G= 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
R= .1622E-05      TOL= .9875E+00      KBASIS=  3      LL= 90
REAL F=  .0000  .0000  .0000 -.0000 -.0000  .0000 -.0017 1.0000 -.0021
IMAG F=  .0000  .0000 -.0000  .0000  .0000 -.0000  .0005  .0000 -.0007
R= .6578E-06      TOL= .2540E-03      KBASIS= 18      LL= 99
```

Fig 5-5. DECON2 Output for 2-D Example Problem

iterations, F, G, R, TOL, KBASIS, and the iteration number LL are printed as shown in Figure 5-5. TOL is the inverse of the condition number, which is defined as $||\underline{A}||\ ||\underline{A}^{-1}||$ (Ref 4:176). The closer TOL is to 1.0, the better conditioned the $\underline{A}$ matrix is. KBASIS shows how many columns of the $\underline{A}$ matrix are used to find the least-square-error value for F or G at each iteration. When fewer than 2(2Pmax+1)(2Qmax+1) columns are used, it is because some of the columns are linear combinations of each other. This occurs in the example problem because so many of the Fourier coefficients are near zero. The IMSL subroutine LLSQF throws out columns that cause TOL to drop below some predefined value, 10^{-4} in this case. As noted in Chapter 4, the smaller the system of equations, the more rapidly the deconvolution algorithm will converge. The fact that KBASIS=3 on every other iteration in the example problem undoubtedly speeds the convergence considerably.

After 99 iterations, R has dropped below 10^{-6} and the algorithm has converged to $F_{1,0}$=(1.0, 0.0) and $G_{0,1}$=(1.0, 0.0), with all other coefficients being near zero. From Eqs (5-4) and (5-5),

$$e^{jkW_A(x,y)} = e^{j2\pi x} \tag{5-21}$$

and

$$e^{jkW_B(x,y)} = e^{j2\pi y} \tag{5-22}$$

Comparing the left and right sides of Eqs (5-21) and (5-22), $W_A(x,y) = \lambda x$ and $W_B(x,y) = \lambda y$, which are the correct solutions.

Summary

In this chapter it was shown that the extension of the 1-D deconvolution problem to two dimensions is straightforward, with all 2-D equations having the same form as the 1-D equations. A very simple 2-D problem was presented and the aberrations were found using the 2-D deconvolution algorithm, thus showing that the algorithm works at least on simple problems. The dimensionality of the matrices in more general 2-D problems tends to be large enough that considerable computer time and memory is required for their solution. The 2-D deconvolution algorithm is not exercised on these general problems since the scope of this paper is limited to presenting an approach to solving the deconvolution problem, but does not include an extensive exercise of the algorithm.

VI. Conclusions and Recommendations

Conclusions

The fundamental difficulty with the deconvolution problem is that it is nonlinear. As with most nonlinear problems, it does not lend itself well to solution or analysis. Despite these drawbacks, estimation theory was applied in Chapter 3 to characterize the noise performance. It was shown that the bilinear system of equations representing the deconvolution problem is not unduly sensitive to measurement noise except at certain input plane wave arrival angles.

In Chapter 2 it was shown that as the aberrations become more severe, the dimensionality of the system of nonlinear equations increases dramatically. This is perhaps the most significant limitation in applying the deconvolution algorithm to a practical problem. Computer memory and time requirements become prohibitive even for moderately severe aberrations. It was also shown in Chapter 2 that vignetting and edge diffraction contribute to error in the solution, but these errors can be minimized and do not impose significant limitations on the problem.

In Chapter 4 it was shown that the solution space of the deconvolution problem contains local minima that the algorithm may iterate to. To get the algorithm to iterate to the global minimum, the algorithm must start at a point

in solution space that is fairly close to the global minimum. The implication is that the solution must be approximately known or a time consuming grid search must be conducted over the solution space to find the global minimum. A local minimum can usually be recognized as such because the minimum will not be sufficiently small, and the Fourier coefficient vectors F and G will not both have unit length.

Finally, in Chapter 5 it was shown that the extension of the 1-D problem to two dimensions is straightforward. The application of the deconvolution algorithm to a very simple 2-D problem in Chapter 5 shows that it does work in principle, even though there are a number of practical limitations on its use.

Recommendations

There are undoubtedly a number of alternate approaches to the deconvolution problem that could be explored. The approach presented herein is rather straightforward and is based on the equations of propagation through the system. The intent in this approach has been not only to develop a deconvolution algorithm, but also to explore the basic nature of the problem.

The deconvolution algorithm presented in this paper attempts to define completely the aberration functions with two sets of measurements. In a practical application, it would probably not be necessary to know exactly what the

aberrations are as long as it is known which direction to move the optical surfaces to correct for the aberrations. The optical surfaces could be moved, another measurement taken, the optical surfaces moved again, etc. In this way the optical system itself would become part of an iterative algorithm which would converge to a properly aligned optical system. Algorithms based on geometrical ray tracing such as the one briefly introduced in Chapter 1 might be employed in such a scheme.

Whether or not there is ever any practical application for a deconvolution algorithm depends heavily upon the success of those who are developing methods for deducing amplitude and phase of optical frequency fields from intensity measurements. It may be that since both of these problems are nonlinear, they could profitably be combined into one problem: namely, determining aberrations on optical surfaces directly from intensity measurements at the output of the optical system. This should certainly be explored since it may not be any more difficult to determine aberrations from intensity measurements than it is to determine either amplitude and phase from intensity measurements or to determine aberrations from amplitude and phase measurements.

Bibliography

1. Ahlberg, J.H. et al., The Theory of Splines and Their Applications, New York: Academic Press, 1967.
2. Born, Max and Emil Wolf, Principles of Optics, New York: Pergamon Press, 1975.
3. Chi, Changhwi, "Curvilinear Bicubic Spline Fit Interpolation Scheme," Optica Acta, 20: 979-993 (1973).
4. Dahlquist, Germund and Ake Bjorck, Numerical Methods, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1974.
5. Davidon, William C., "Variance Algorithm for Minimization," Computer Journal, 10: 406-411 (1967).
6. DeSantis, P. and C. Palma, "Degree of Freedom of Aberrated Images," Optica Acta, 23: 743-752 (Sep 1976).
7. Eikonix Corporation, "Interim Report: Analytical Studies of Phase Estimation." Unpublished report prepared for Rome Air Development Center, Griffiss AFB NY (Proprietary), September 14, 1978.
8. Fletcher, R. and M.J.D. Powell, "A Rapidly Convergent Descent Method for Minimization," Computer Journal, 6: 163-168 (1963).
9. Foley, John T. and R. Russell Butts, "Uniqueness of phase retrieval from intensity measurements," JOSA, 71: 1008-1014 (1981).

10. Gerchberg, R.W. and W.O. Saxton, "Phase Determination from Image and Diffraction Plane Pictures in the Electron Microscope," Optik, 34(3): 275-284 (1971).

11. Goodman, Joseph W., Introduction to Fourier Optics, San Francisco: McGraw-Hill, Inc., 1968.

12. Hardy, John W., "Active Optics: A New Technology for the Control of Light," Proceedings of the IEEE, 66(6): 651-697 (June 1978).

13. Lawson, C.L. and R.J. Hanson, Solving the Least Squares Problem, Englewood Cliffs, NJ: Prentice-Hall, Inc., 1974.

14. Lee, D.A. et al., "Some Practical Aspects of the Treatment of Ill-Posed Problems by Regularization." Unpublished interim report prepared for Aerospace Research Laboratories, Wright-Patterson AFB OH (February 1975).

15. Morse, Philip M. and Herman Feshbach, Methods of Theoretical Physics, New York: McGraw-Hill, Inc., 1953.

16. Oppenheim, Alan V. and Ronald W. Schafer, Digital Signal Processing, London: Prentice-Hall, Inc., 1975.

17. Perkin-Elmer Corporation, "HALO Deconvolution." Unpublished report for RADC/OCSE, Griffiss AFB NY.

18. Ortega, J.M. and W.C. Rheinbolt, Iterative Solutions of Nonlinear Equations in Several Variables, New York: Academic Press, 1970.

19. Peters, W.N. et al., "Stellar Interferometer for Figure Sensing of Orbiting Astronomical Telescopes," Applied Optics, 14(11): 2622-2626 (November 1975).

20. Robinson, Stanley R., "On the Problem of Phase From Intensity Measurements," JOSA, 68(1): 87-92, (January 1978).

21. Tikhonov, Andrey N., "Regularization of Incorrectly Posed Problems," (Russian) English Translation, Soviet Mathematics, 4: 1624-1627 (1963).

22. Tikhonov, Andrey N. and Vasiliz Y. Arsenin, Solution of Ill-Posed Problems, New York: John Wiley & Sons, Inc., 1968.

23. Van Trees, Harry L., Detection, Estimation, and Modulation Theory, New York: John Wiley & Sons, Inc., 1968.

24. Wissinger, A. and T. Facey, "On-Orbit Optical Control of the Space Telescope," Optical Spectra: 30-38 (March 1977).

25. Ziemer, R.E. and W.H. Tranter, Principles of Communications, Boston: Houghton Mifflin, 1976.

Appendix A

Development of the 1-D Fisher Information Matrix

This appendix documents the mathematical calculations required by Eq (3-23) to find the elements of the Fisher Information Matrix (FIM) (Ref 23:80). The FIM is then separated and inverted so that the lower bounds on VAR[$\hat{di}(H)$-di] can be found, where di is used to represent the real or imaginary parts of any element of F or G. F and G are the Fourier coefficient vectors which describe the aberrations.

The equations in Chapter 3 which are required are repeated here for convenience.

$$J_{ij} = \frac{1}{2\sigma^2} E\left[\frac{\partial^2 |\hat{H}-AY|^2}{\partial di\ \partial dj}\right] + E\left[\frac{\partial^2}{\partial di\ \partial dj}\left(\sum_{\substack{i \\ di \neq F_{0R} \text{ or } G_{0R}}} \frac{di^2}{2\sigma^2_{di}} + \frac{(F_{0R}-1)^2}{2\sigma^2_{F0R}} + \frac{(G_{0R}-1)^2}{2\sigma^2_{G0R}}\right)\right] \quad (3\text{-}23)$$

$$|\hat{H}-AY|^2 = \sum_J \sum_p |\hat{H}_J(p) - \sum_{\substack{m \\ m+n=p}} \sum_n F_m G_n a_J(m,n)|^2 \quad (3\text{-}18)$$

$$E[\hat{d}_i] = 0 \text{ for } d_i \neq F_{0R} \text{ or } G_{0R}$$

$$E[\hat{F}_{0R}] = E[\hat{G}_{0R}] = 1$$

$$VAR[\hat{d}_i] = \sigma^2_{di} \quad (3\text{-}19)$$

$$\sigma_{\varepsilon i}^2 \triangleq \text{Var}\,[\hat{d}_i(\hat{H})-d_i] \geq (J^{-1})_{ii} \tag{3-13}$$

where $(J^{-1})_{ii}$ is the iith element of matrix $\underline{J}^{-1}$, the inverse of the FIM. For convenience, Eq (3-23) will be divided into two parts corresponding to the two expected values, so the FIM is now J_T, and

$$J_T = J_D + J_P \tag{A-1}$$

By inspection, J_P will be a diagonal matrix because all terms within the summation are of the form

$$\frac{di^2}{2\sigma_{di}^2}$$

and

$$\frac{\partial^2}{\partial di\,\partial dj}\, di^2 = 0$$

$$\frac{\partial^2}{\partial di^2}\left(\frac{di^2}{2\sigma_{di}^2}\right) = \frac{1}{\sigma_{di}^2} \tag{A-2}$$

so

$$J_P = \begin{bmatrix} \frac{1}{\sigma_{d_1}^2} & & 0 & \cdots & 0 \\ 0 & & \frac{1}{\sigma_{d_2}^2} & \cdots & \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & & & \frac{1}{\sigma_{d_k}^2} \end{bmatrix} \qquad \text{(A-3)}$$

Finding J_D is not so simple. The squaring and summing operations in Eq (3-18) may yield hundreds of terms for even an average size matrix. Of course, after the partials and expected values are taken, most terms go to zero. Carrying out the squaring operation in Eq (3-18) yields

$$|\hat{H}-AY|^2 = \sum_J \sum_p \hat{H}_J(p)\ \hat{H}_J^*(p) - 2\mathrm{Re}\Big[\hat{H}_J(P) \sum_{\substack{m \\ m+n=p}} \sum_n F_m^*\ G_n^*\ a_J^*(m,n)\Big] + \sum_{\substack{m \\ m+n=p}} \sum_n F_m\ G_n\ a_J(m,n)$$

$$\sum_{\substack{m' \\ m'+n'=p}} \sum_{n'} F_{m'}^*\ G_{n'}^*\ a_J^*(m',n') \qquad \text{(A-4)}$$

Note that because $m+n=p$ in Eq (A-4), the double summation and quadruple summations can be reduced to single and double summations, respectively. Now the first part of

Eq (3-23) can be written as

$$J_{Dij} = \frac{1}{2\sigma^2} E\left\{\frac{\partial^2}{\partial di\ \partial dj}\left[\sum_J \sum_p \left(|\hat{H}_J(p)|^2 - 2Re[\hat{H}_J(p) \sum_{\substack{m \\ |p-m|\le Pmax}} F_m^* G_{p-m}^* a_J^*(m,p-m)] + \sum_m \sum_{\substack{m' \\ |p-m|\le Pmax \\ |p-m'|\le Pmax}} F_m G_{p-m} F_{m'}^* G_{p-m'}^* a_J(m,p-m)\ a_J^*(m',p-m')\right)\right]\right\} \quad \text{(A-5)}$$

The term $|\hat{H}_J(p)|^2$ in Eq (A-5) goes to zero when the partials are taken. Noting that $E[\hat{H}_J(p)]=H_J(p)$, the contributions to J_{Dij} from

$$E\left[\frac{\partial^2}{\partial di\ \partial dj}\left(-2Re[\hat{H}_J(p) \sum_{\substack{m \\ |p-m|\le Pmax}} F_m^* G_{p-m}^* a_J(m,p-m)]\right)\right]$$

are summarized below for four combinations of di and dj:

TABLE A-1

Terms from the Second Term in Eq (B-5)

di	dj	Term
F_{iR}	G_{jR}	$-2Re[H_J(p)\ a_J^*(i,j)]$
F_{iR}	G_{jI}	$-2Im[H_J(p)\ a_J^*(i,j)]$
F_{iI}	G_{jR}	$-2Im[H_J(p)\ a_J^*(i,j)]$
F_{iI}	G_{jR}	$+2Re[H_J(p)\ a_J^*(i,j)]$

For any other combinations of di and dj, the second term in Eq (A-5) does not contribute to J_{Dij}.

In the third part of Eq (A-5), the second partial differential operates on two of the coefficients and the expected value operator operates on the remaining coefficients. The remaining coefficients must be the zero order coefficients (F_0, G_0) or they must be of the form $|di|^2$ or the expected value will be zero. There are eight different cases where terms from

$$\sum_{\substack{m \\ |p-m|\leq Pmax \\ |p-m'|\leq Pmax}} \sum_{m'} F_m\, G_{p-m}\, F^*_{m'},\, G^*_{p-m'},\, a_J(m,p-m)\, a^*_J(m',p-m')$$

may yield nonzero values when operated on by the partial and expected value operators. The eight cases arise from taking two of the coefficients at a time, setting them equal to di and dj, respectively, and letting the remaining two coefficients be the zero order coefficients. These eight cases are summarized in the table below. Use is made of the fact the $a(0,0)=1$ and $a(m,0)a^*(0,m)=a(m,-m)$ (see Eq (2-29)).

Table A-2 can best be explained with an example. In case 1, the term specified in the table is from the last part of Eq (A-5),

$$E\left[\frac{\partial^2}{\partial di\ \partial dj}\ (F_i\ G_i\ F^*_0\ G^*_0\ a_1(i,j)\ a^*_1(0,0))\right] .$$

The constraints arise from simple algebra performed on the values assigned to m, p-m, m', and p-m'. The entries in columns RR through II are the multipliers that the result

TABLE A-2

Terms From the Third Term in Eq (A-5)

Case	m	p-m	m'	p-m'	Constraint	RR	RI	IR	II	Result
1	i	j	0	0	p=0, i=-j	1	j	j	-1	a(i,-i)
2	0	0	i	j	p=0, i=-j	1	-j	-j	-1	$a^*(i,-i)$
3	i	0	j	0	i=j=p	1	0	0	1	$4\sigma^2_{G0}+2$
4	i	k	j	k	i=j, k=0	1	0	0	1	$4\sigma^2_{Gk}$
5	0	i	0	j	i=j=p	1	0	0	1	$4\sigma^2_{F0}+2$
6	k	i	k	j	i=j, k=0	1	0	0	1	$4\sigma^2_{Fk}$
7	i	0	0	j	i=j=p	1	-j	j	1	a(i,-i)
8	0	i	j	0	i=j=p	1	-j	j	1	$a^*(i,-i)$

of the E $[\partial^2/\partial di \partial dj]$ operator must be multiplied by depending on whether di and dj correspond to real or imaginary components of F_m or G_n. For example, if the third part of Eq (A-5) is operated on by E $[\partial^2/\partial F_{iR} \partial G_{iI}]$, only two nonzero terms result. The first term corresponds to case 1 in Table A-2, and that term, a(i,-i), must be multiplied by j because it corresponds to the real part of F_i and dj corresponds the the imaginary part of G_{-i}. Therefore, the multiplier in the RI column is used. Likewise, the second term, which corresponds to case 2, must be multiplied by -j.

The results of the E $[\partial^2/\partial di \partial dj]$ operator listed in Table A-2 are for a single measurement "θ". When $\hat{H}$-AY includes terms corresponding to more than one θ, then those terms are simply added, and for each case, the "Result" column in Table A-2 will contain one term for each θ.

Now we are in a position to summarize the results of this appendix by writing a generalized expression for J_{ij}, the entries in the FIM. This is done most easily by listing all possible cases of J_{ij}, where J_{ij} is given by Eq (3-23).

TABLE A-3

Fisher Information Matrix Entries

di	dj	J_{ij} $(=J_{di,dj})$	Constraints
F_{iR}	G_{-iR}	$\frac{1}{\sigma^2}[\mathrm{Re}[a(i,-i)]-\mathrm{Re}[a(i,-i)H_J^*(0)]$	$i\neq 0$
F_{iR}	G_{-iI}	$\frac{1}{\sigma^2}[-\mathrm{Im}[a(i,-i)]+\mathrm{Im}[a(i,-i)H_J^*(0)]$	$i\neq 0$
F_{iI}	G_{-iR}	$\frac{1}{\sigma^2}[-\mathrm{Im}[a(i,-i)]+\mathrm{Im}[a(i,-i)H_J^*(0)]$	$i\neq 0$
F_{iI}	G_{-iI}	$\frac{1}{\sigma^2}[-\mathrm{Re}[a(i,-i)]+\mathrm{Re}[a(i,-i)H_J^*(0)]$	$i\neq 0$
F_{iR}	F_{iR}	$\frac{1}{\sigma^2}(2\sum_{k=-Pmax+i}^{Pmax}\sigma_{Gk}^2+1)+\frac{1}{\sigma_{FiR}^2}$	$i\geq 0$
F_{iI}	F_{iI}	$\frac{1}{\sigma^2}(2\sum_{k=-Pmax+n}^{Pmax}\sigma_{Gk}^2+1)+\frac{1}{\sigma_{FiR}^2}$	$i\geq 0$

F_{iR}	F_{iR}	$\frac{1}{\sigma^2}(2\sum_{k=-Pmax}^{Pmax}\sigma_{Gk}^2+1)+\frac{1}{2}\sigma_{FiR}$	$i<0$
F_{iI}	F_{iI}	$\frac{1}{\sigma^2}(2\sum_{k=-Pmax}^{Pmax+i}\sigma_{Gk}^2+1)+\frac{1}{2}\sigma_{FiI}$	$i<0$
G_{iR}	G_{iR}		
or		Same as F_i,F_i with Fs and Gs exchanged	
G_{iI}	G_{iI}		
F_{0R}	G_{0R}	$\frac{1}{\sigma^2}(2-Re[H_J(0)])$	
F_{0R}	G_{0I}	$-\frac{1}{\sigma^2}Im[H_J(0)]$	
F_{0I}	G_{0R}	$-\frac{1}{\sigma^2}Im[H_J(0)]$	
F_{0I}	G_{0I}	$\frac{1}{\sigma^2}Re[H_J(0)]$	
F_{iR}	G_{iR}	$\frac{1}{\sigma^2}[Re[a(i,-i)]-Re[a(i,i)H_J^*(2i)]$	$i\neq0$
F_{iR}	G_{iI}	$\frac{1}{\sigma^2}[Im[a(i,-i)]-Im[a^*(i,i)H_J(2i)]$	$i\neq0$
F_{iI}	G_{iR}	$\frac{1}{\sigma^2}[Im[a(-i,i)]-Im[a^*(i,i)H_J(2i)]$	$i\neq0$
F_{iI}	G_{iI}	$\frac{1}{\sigma^2}[Re[a(i,-i)]+Re[a^*(i,i)H_J(2i)]$	$i\neq0$
F_{iR}	G_{iR}	$-\frac{1}{\sigma^2}Re[H_J(p)a^*(i,j)]$	$i\neq\pm j$, $\lvert i+j\rvert\leq Pmax$
F_{iR}	G_{iI}	$-\frac{1}{\sigma^2}Im[H_J(p)a^*(i,j)]$	"

F_{iI}	G_{iR}	$-\frac{1}{\sigma^2}\mathrm{Im}[H_J(p)a^*(i,j)]$	"
F_{iI}	G_{iI}	$\frac{1}{\sigma^2}\mathrm{Re}[H_J(p)a^*(i,j)]$	"

The E $[\partial^2/\partial di \partial dj]$ operator yields zero for all combinations of di and dj not listed in Table A-3.

In Table A-3, $H_J(p)$ is not known exactly, but if σ_H^2 is small, then $\hat{H}_J(p)$, the measured value of $H_J(p)$, can be used instead of $H_J(p)$. Then for a given measurement of $U_5(x,\theta)$, all the entries in Table A-2 can be calculated to yield the total FIM. The FIM will be square with dimensions of $4(2P_{max}+1)$ on each side. According to Eq (3-13), the FIM may be inverted to yield lower bounds on the variance of the real and imaginary parts of the unknown coefficients, F_m and G_n. A very large variance for any di would indicate that the problem of finding the coefficients is ill-conditioned.

A simple example will be used to further explore the properties of the FIM. Suppose $P_{max}=1$. The Xs in Figure A-1 show the locations of the nonzero entries in the FIM for this case, and the Ys show the location of entries that may approach zero. (The values for these entries can be found in Table A-3, but are not written out in Figure A-1 for brevity.)

	F_{-iR}	F_{0R}	F_{iR}	G_{-iR}	G_{0R}	G_{iR}	F_{-iI}	F_{0I}	F_{iI}	G_{-iI}	G_{0I}	G_{iI}
F_{-iR}	X	-	-	X	Y	Y	-	-	-	X	Y	Y
F_{0R}	-	X	-	Y	X	Y	-	-	-	Y	X	Y
F_{iR}	-	-	X	Y	Y	X	-	-	-	Y	Y	X
G_{-iR}	X	Y	Y	X	-	-	X	Y	Y	-	-	-
G_{0R}	Y	X	Y	-	X	-	Y	X	Y	-	-	-
G_{iR}	Y	Y	X	-	-	X	Y	Y	X	-	-	-
F_{-iI}	-	-	-	X	Y	Y	X	-	-	X	Y	Y
F_{0I}	-	-	-	Y	X	Y	-	X	-	Y	X	Y
F_{iI}	-	-	-	Y	Y	X	-	-	X	Y	Y	X
G_{-iI}	X	Y	Y	-	-	-	X	Y	Y	X	-	-
G_{0I}	Y	X	Y	-	-	-	Y	X	Y	-	X	-
G_{iI}	Y	Y	X	-	-	-	Y	Y	X	-	-	X

Fig A-1. Location of the Nonzero Entries in the FIM.

With the use of a computer, the inverse of the matrix in Figure A-1 could be taken for a number of values of any of the variables in the matrix elements, and the effect of the variations on the lower bound of the variance of F_m or G_n could be studied. This would be tedious. An easier way to gain some insight into the behavior of the lower bounds is to assume that the aberrations are small so that $H_J(0) \cong 1$ and $H_J(p) \cong 0$ where $p \neq 0$. Then all the "Y" entries in Figure A-1 go to zero and the 12 X 12 matrix can be separated into three 4X4 matrices as shown below.

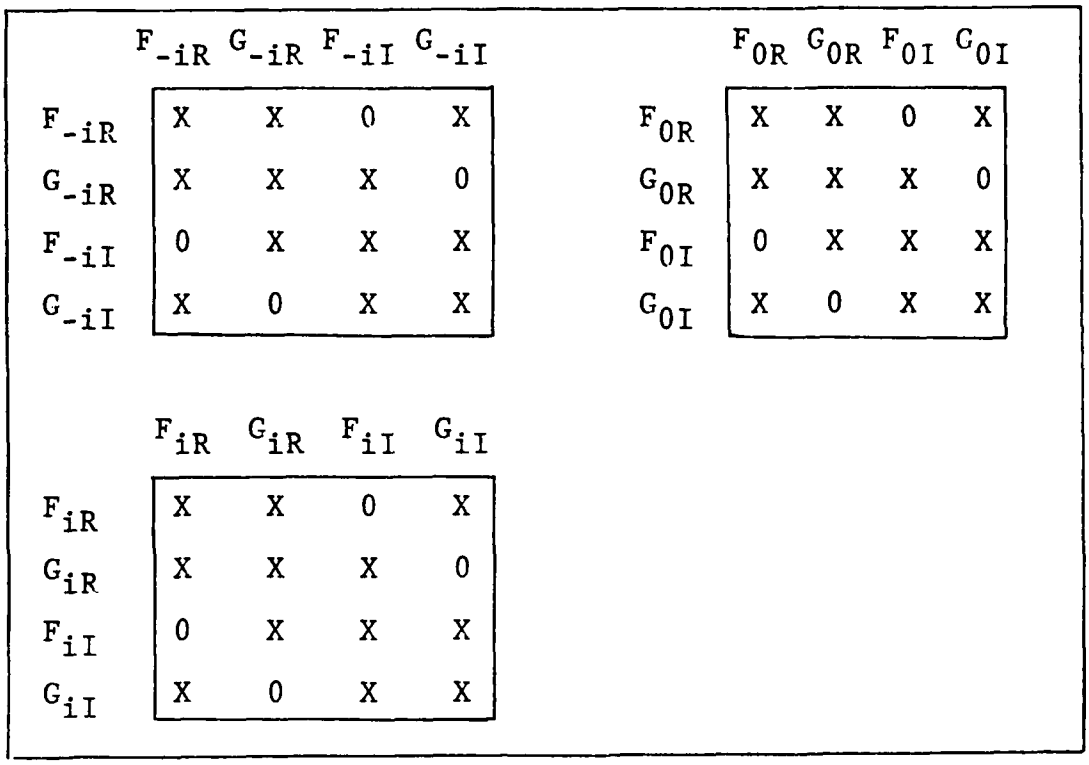


Fig A-2. Reduced FIM.

The overall FIM of Figure A-1 will always collapse into a group of 4 X 4 matrices when the aberrations approach zero, regardless of the dimensionality of the FIM. Each 4 X 4 matrix corresponds to one component of the Fourier series used to represent the aberrations. Assuming the variance of the real and imaginary parts of any F_m or G_n are equal, the 4 X 4 matrices of Figure A-2 can be represented as shown in Figure A-3. Note that many of the elements of the matrix are equal. This matrix can easily be inverted in general.

	F_{iR}	G_{iR}	F_{iI}	G_{iI}
F_{iR}	c	a	0	b
G_{iR}	a	d	-b	0
F_{iI}	0	-b	c	a
G_{iI}	b	0	a	d

Fig A-3. Reduced FIM Structure.

In this case, only the diagonal values of the inverse are of interest, and they can be shown to be

$$(FIM^{-1})_{11} = (FIM^{-1})_{33} = \frac{d}{cd-a^2-b^2} \qquad \text{(A-6)}$$

and

$$(FIM^{-1})_{22} = (FIM^{-1})_{44} = \frac{c}{cd-a^2-b^2} \qquad \text{(A-7)}$$

Substituting in the values of a, b, c, and d from Table A-2, the lower bounds on the variance of the Fourier coefficients may now be written. The values are given with the summation over different values of θ (summation over J) included.

$$c = \frac{1}{\sigma^2} \sum_J \left(2 \sum_k \sigma_{Gk}^2 + 1\right) + \frac{1}{\sigma_{Fi}^2} \qquad \text{(A-8)}$$

$$d = \frac{1}{\sigma^2} \sum_J \left(2 \sum_k \sigma_{Fk}^2 + 1\right) + \frac{1}{\sigma_{Gi}^2} \qquad \text{(A-9)}$$

$$a = \frac{1}{\sigma^2} \sum_J \mathrm{Re}[a(i,-i)] = \frac{1}{\sigma^2} \sum_J \cos\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right) \qquad \text{(A-10)}$$

$$b = \frac{1}{\sigma^2} \sum_J \mathrm{Im}[a(i,-i)] = -\frac{1}{\sigma^2} \sum_J \sin\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right) \qquad \text{(A-11)}$$

Substituting Eqs (A-8) through (A-11) into (A-6) and (A-7) yields

$$\mathrm{VAR}[F_{iR}] = \mathrm{VAR}[F_{iI}] \geq$$

$$\frac{\sigma^2 \sum_J \left(2 \sum_k \sigma_{Fk}^2 + 1\right) + \frac{\sigma^4}{\sigma_{Gi}^2}}{\left[\left[\sum_J \left(2 \sum_k \sigma_{Gk}^2 + 1\right) + \frac{\sigma^2}{\sigma_{Fi}^2}\right]\left[\sum_J \left(2 \sum_k \sigma_{Fk}^2 + 1\right) + \frac{\sigma^2}{\sigma_{Gi}^2}\right] - \right.}$$
$$\left.\left[\sum_J \cos\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right)\right]^2 - \left[\sum_J \sin\left(\frac{2\pi}{a} Z_1 i \sin\theta_J\right)\right]^2\right]$$

(A-12)

where

$$-P_{max}+i \leq k \leq P_{max} \text{ for } i \geq 0$$

and

$$-P_{max} \leq k \leq P_{max}+i \text{ for } i < 0$$

The bounds on $\mathrm{VAR}[G_{iR}]=\mathrm{VAR}[G_{iI}]$ are the same as Eq (A-12) except that the numerator is

$$\sigma^2 \sum_J (2 \sum_k \sigma_{Gk}^2 + 1) + \frac{\sigma^4}{\sigma_{Fi}^2}$$

Appendix B: Ray Trace Deconvolution Algorithm

The ray trace deconvolution algorithm presented in this appendix was "invented" rather than derived. Not much is known about its accuracy, stability, or uniqueness, but because it shows promise as a suitable alternative to the algorithm developed in the text, at least in some types of optical systems, it is documented here.

A simple version of the ray tracing algorithm is illustrated in Figure B-1. $W_A(x)$ and $W_B(x)$ are one-dimensional aberration functions separated by a distance d. Phase maps are obtained immediately behind $W_B(x)$ for plane waves arriving at two different angles, say 0 and θ. From these phase maps the optical distance traveled by a given ray can be found. The rays in Figure B-1 with an α subscript correspond to the plane wave arriving at angle θ and the β rays correspond to the plane wave arriving at angle 0. The assumption is made that $W_A(0)=W_A'(0)=W_B(0)=W_B'(0)=0$. In other words, an optical axis is selected where, based on physical consideration, it can be assumed that there are no aberrations. The rest of the aberration function will be calculated relative to this reference axis. Now the following steps are taken:

1. Measure 1α (read from phase map α).

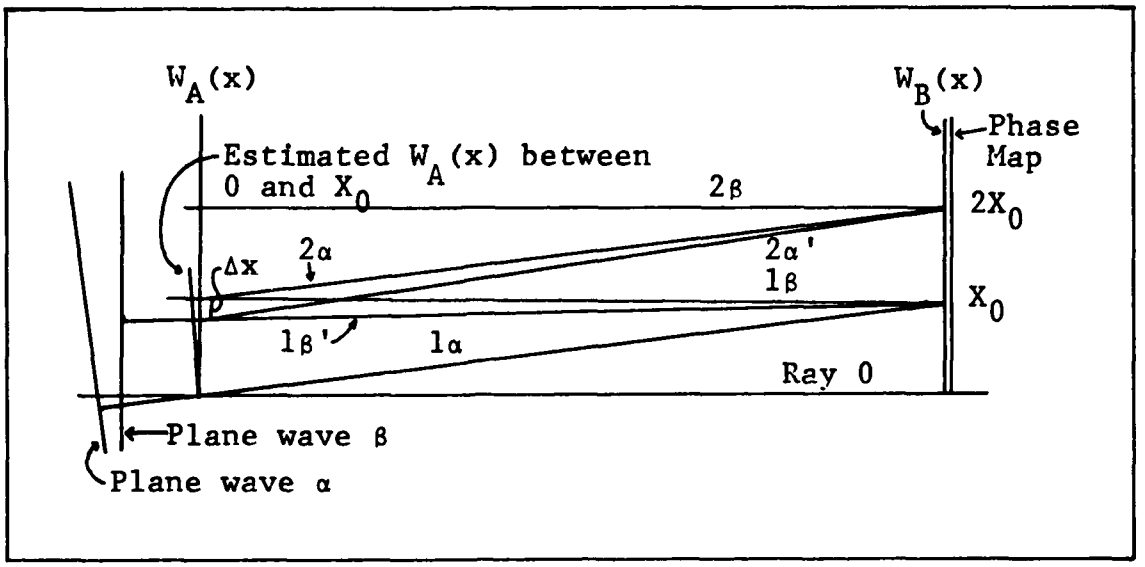


Fig B-1. Ray Trace Deconvolution Algorithm

2. Calculate what 1α would be with no aberrations $(=(x_o^2+d^2)^{\frac{1}{2}})$.

3. Measured 1α - calculated $1\alpha = W_B(x_o)$ because $W_A(0)=W_A'(0)=0$. The optical delay at $W_B(x_o)$ is known perfectly.

4. Measure 1β (read from phase map β).

5. Calculate $1\beta = d + W_B(x_o)$.

6. Measured 1β - calculated 1β = error due to W_A near x_o.

With only a finite number of phase maps, two in this case, one can't tell exactly what the form of $W_A(x)$ is in the region near x_o, and since the form of $W_A(x)$ is not known, one can't tell where ray 1β really came from. Suppose the actual ray passing through $W_B(x_o)$ is $1\beta'$ and came through $W_A(x)$ at $x_o + \Delta x$ (where Δx is negative as

drawn in Figure B-1). Δx can be determined only if a form for $W_A(x)$ is assumed. Suppose it is assumed that $W_A(x)$ is piecewise linear between all points nx_o and $(n+1)x_o$ where n is an integer. Then there is enough information to uniquely determine Δx, $W_A(x+\Delta x)$, and $W_A(x_o)$. Now using rays 2α and 2β, steps 1 through 6 can be repeated to find $W_A(2x_o)$ and $W_B(2x_o)$. This procedure is repeated until the edge of the aperture is reached.

A problem with the algorithm is that the errors that occur as a result of assuming a slightly incorrect form for $W_A(x)$ accumulate so that the error between the calculated aberrations and the actual aberrations near the edge of the aperture may be quite large. If, from design considerations, the aberrations could be assumed to be zero along some other optical paths besides the reference axis, the aperture could be divided into subapertures, one for each aberration-free path. This would help to solve the problem of cumulative errors, but there are probably not many adaptive imaging systems where more than one reference path can be assumed.

Table B-1 shows how the algorithm worked for a test problem where $W_A(x)$ and $W_B(x)$ were rather severe misfocus aberrations represented by thin lenses. The parameters were as follows: x_o=.1, d=5, focal length of lens A was 100, and focal length of lens B was -150 (the negative sign indicates the lens is concave). The numbers under the

"algorithm" column are the values of the aberration functions predicted by the algorithm, and the numbers under the "actual" column are the actual values of the aberrations.

TABLE B-1

Algorithm Performance for Focusing Aberrations

	$W_A(x) \times 10^{-6}$		$W_B(x) \times 10^{-6}$	
x	Algorithm	Actual	Algorithm	Actual
.1	-51.9	-50	33.3	33.3
.2	-154.6	-200	80	133
.3	-306.7	-450	138	300
.4	-507	-800	208	533
.5	-755	-1250	288	833
.6	-1049	-1800	376	1200

The errors in Table B-1 are quite large, but this test problem is a severe case in some respects. All the errors that accumulate have the same sign resulting in large errors near the edge of the aperture. In a real-world aberration, many errors would have opposite signs and would cancel each other. Also, a linear form for $W_A(x)$ was assumed when it was actually quadratic. The better the assumed aberration function approximates the real function, the smaller the errors will be.

In spite of the errors shown in Table B-1, this algorithm would be sufficient for some adaptive systems if for

each pass of the algorithm the system was corrected in the right direction. The overall system would then converge to the correct configuration. However, several things could be done to improve the accuracy of the algorithm. Cubic spline functions would better approximate elastic surfaces than piecewise linear functions and would yield better results (Refs 1, 3). Also, additional wavefront measurements could be incorporated into an algorithm to improve its accuracy. For example, if a third phase map were available which corresponded to an arrival angle halfway between the arrival angles for the first two phase maps, then the values of the aberration functions could be estimated at $nx_o/2$. This information could be incorporated into the interpolation scheme to increase its accuracy. As discussed in Chapter 2, if the phase maps were known for a continuum of arrival angles, enough information would be available to solve the problem exactly. Finally, if an imaging system was designed so that the entire region within $+\frac{1}{2}x_o$ of the reference axis (see Figure B-1) could be considered aberration free, then a number of reference rays could be used as starting points all at the same time, resulting in an algorithm which would provide much more detailed coverage of the aberrations and yield greatly improved results.

Appendix C: Program DECON2 Listing

```
      PROGRAM DECON2(INPUT,OUTPUT)
C  THIS PROGRAM SOLVES FOR THE FOURIER COEFFICIENTS OF TWO
C  ABERRATION FUNCTIONS IN A 2-D OPTICAL SYSTEM.  THE INPUT
C  TO THE PROGRAM, U, IS A NORMALIZED SAMPLING OF THE
C  AMPLITUDE AND PHASE AT THE OUTPUT OF THE SYSTEM FOR TWO
C  OR MORE INPUT PLANE WAVE ARRIVAL ANGLES.
C  U(NU,NU) IS THE COMPLEX SET OF SAMPLE POINTS WHERE THE
C  POINTS ARE TAKEN ON AN EQUIDISTANT NU X NU GRID.
C  ND**2 IS THE NUMBER OF COMPLEX FOURIER COEFFICIENTS WHICH
C  WILL BE USED TO DESCRIBE EACH ABERRATION.
C  ND MUST BE LESS THAN NU/2.
C  DIMENSIONS ARE AS FOLLOWS:  U(NU,NU), A(KK,NE),
C  G(NE), F(NE), H(KK), HW(NE), B(KK), AND IP(NE).
      COMPLEX U(8,8),ZZ,ZW
      REAL A(36,18),G(18),F(18),H(36),HW(18),B(36),LMDA
      INTEGER IP(18),P,Q
      COMMON U
C     LOAD PARAMETERS
C  Z1 IS THE DISTANCE BETWEEN ABERRATIONS AND Z2 IS THE
C  DISTANCE FROM THE SECOND ABERRATION TO THE OUTPUT
C  MEASUREMENT PLANE.
C  LMDA IS WAVELENGTH IN METERS.  THETA1 AND THETA2 ARE PLANE
WAVE ARRIVAL ANGLES IN RADIANS WITH RESPECT TO THE X-AXIS.
C  AND PHI1 AND PHI2 ARE WITH RESPECT TO THE Y-AXIS.
      Z1=2.$Z2=1.$LMDA=1.0E-6$MA=3$THETA1=0.$THETA2=.05$ND=3$LMAX=2
      PHI1=0.$PHI2=.05$AX=1.0$AY=1.0$PI=3.141592654
      SINT1=SIN(THETA1)$SINT2=SIN(THETA2)
      SINP1=SIN(PHI1)$SINP2=SIN(PHI2)$NU=2**MA$NC=(ND+1)/2
      NA=ND**2$NE=2*NA$KK=NA*LMAX$IA=2*KK
      PRINT 906
      DO 20 L=1,LMAX
      DO 10 I=1,NU
      DO 15 J=1,NU
      X=(I-1)*AX/NU
      Y=(J-1)*AY/NU
      IF (L.EQ.1)SINT=SINT1
      IF (L.EQ.1)SINP=SINP1
      IF (L.EQ.2)SINT=SINT2
      IF (L.EQ.2)SINP=SINP2
      U(I,J)=CONJG(CEXP(CMPLX(0.,2.*PI*(X-3.*SINT+Y-SINP))))
15    CONTINUE
      PRINT 901,X,(REAL(U(I,J)),J=1,8)
      PRINT 903,(=AIMAG(U(I,J)),J=1,8)
10    CONTINUE
      PRINT 906
C
C     TAKE THE 2-D FFT OF U TO GET H
      CALL DFT(MA)
C     LOAD U INTO H MATRIX
      MM=NA*(L-1)
      DO 20 I=1,ND
      DO 20 J=1,ND
```

```
      MM=MM+1
      NN=MM+KK
      II=I$ JJ=J
      IF(I.GT.NC) II=NU-ND+I
      IF(J.GT.NC) JJ=NU-ND+J
      H(MM)=REAL(U(II,JJ))
      H(NN)=0.-AIMAG(U(II,JJ))
20    CONTINUE
      PRINT 904,(H(I),I=1,9)
      PRINT 905,(H(I),I=19,27)
      PRINT 904,(H(I),I=10,18)
      PRINT 905,(H(I),I=28,36)
C
C     LOAD A ASSUMING G IS GIVEN
      DO 60 I=1,NE
      G(I)=0.
      F(I)=0.
60    CONTINUE
      KA=0
      G(5)=.3162 $ G(6)=.9487
      DO 140 LL=1,140
      LLFLAG=(-1)**LL
      DO 70 L=1,LMAX
      IF(L.EQ.1)SINT=SINT1
      IF(L.EQ.1)SINP=SINP1
      IF(L.EQ.2)SINT=SINT2
      IF(L.EQ.2)SINP=SINP2
      MM=NA*(L-1)
      DO 80 J=1,ND
      DO 80 K=1,ND
      MM=MM+1
      NN=MM+KK
C     DETERMINE Q AND P FOR ROW J,K
C     ROW NUMBER=(J-1)*ND+K+(L-1)*KK
      Q=K-1
      IF(K.GT.NC)Q=K-ND-1
      P=J-1
      IF(J.GT.NC)P=J-ND-1
      NL=0
C     DETERMINE I,M,L,ANDN FOR COLUMN IB,MB
C     COLUMN NUMBER=(IB-1)*ND+MB FOR REAL PART
      DO 90 IB=1,ND
      DO 90 MB=1,ND
      NL=NL+1
      NR=NL+NA
      I=IB-NC
      M=MB-NC
      LB=P-I
      NB=Q-M
      IF((LB.GE.NC).OR.(LB.LE.(-NC)))GO TO 100
      IF((NB.GE.NC).OR.(NB.LE.(-NC)))GO TO 100
      IF (LLFLAG) 30,90,40
30    AEXP=(-2.1*PI*((Z1+Z2)*(I*SINT/AX+M*SINP/AY)
     1+Z2*(LB*SINT/AX+NB*SINP/AY))
```

```
      GO TO 50
40    AEXP=(-2.)*PI*((Z1+Z2)*(LB*SINT/AX+NB*SINP/AY)
     1+Z2*(I*SINT/AX+M*SINP/AY))
50    ZZ=CMPLX(0.,AEXP)
      ZW=CEXP(ZZ)
      NG=ND*(LB+NC-1)+NB+NC
      IF(LLFLAG.GT.0)GO TO 150
      A(MM,NL)=A(NN,NR)=REAL(ZW*CMPLX(G(NG),G(NG+NA)))
      A(NN,NL)=AIMAG(ZW*CMPLX(G(NG),G(NG+NA)))
      A(MM,NR)=-A(NN,NL)
      GO TO 90
150   A(MM,NL)=A(NN,NR)=REAL(ZW*CMPLX(F(NG),F(NG+NA)))
      A(NN,NL)=AIMAG(ZW*CMPLX(F(NG),F(NG+NA)))
      A(MM,NR)=-A(NN,NL)
      GO TO 90
100   A(MM,NL)=A(NN,NR)=A(MM,NR)=A(NN,NL)=0.
90    CONTINUE
80    CONTINUE
70    CONTINUE
C
C     SOLVE FOR F USING LEAST SQUARES
      DO 110 I=1,IA
      B(I)=H(I)
110   CONTINUE
      TOL=.0001
      KBASIS=0
      IF(LLFLAG.GT.0)GO TO 160
      CALL LLSOF(A,IA,IA,NE,B,TOL,KBASIS,F,HW,IP,IER)
C  ROTATE F VECTOR SO IMAG(F(0,0))=0
      KA=KA+1
      IF(KA.NE.5) GO TO 140
      FZR=F(5)
      FZI=F(14)
      ZW=CMPLX(FZR,FZI)
      ZW=CONJG(ZW/CABS(ZW))
      PRINT 900,(F(I),I=1,NE)
      GO TO 170
160   CALL LLSOF(A,IA,IA,NE,B,TOL,KBASIS,G,HW,IP,IER)
      IF(KA.NE.5) GO TO 140
      KA=KA+1
      XMAG=0.
      DO 250 I=1,NE
      XMAG=XMAG+G(I)**2
250   CONTINUE
      DO 260 I=1,NE
      XNORM=G(I)/SQRT(XMAG)
      G(I)=XNORM
```

```
260   CONTINUE
      PRINT 902,(G(I),I=1,NE)
      KA=0
170   R=0.
      DO 120 I=1,IA
      R=R+B(I)**2
120   CONTINUE
      PRINT 130,R,TOL,KBASIS,LL
      IF(R.LT.1.0E-6)GO TO 5
140   CONTINUE
5     CONTINUE
130   FORMAT(3H R=,E10.4,5X,4HTOL=,E10.4,5X,7HKBASIS=,I3,
     15X,"LL=",I3)
900   FORMAT(1X,"REAL F=",9F7.4/1X,"IMAG F=",9F7.4)
901   FORMAT(1X,"X=",F5.3,2X,"REAL U",8F6.3)
902   FORMAT(1X,"REAL G=",9F7.4/1X,"IMAG G=",9F7.4)
903   FORMAT(10X,"IMAG U=",8F6.3)
904   FORMAT(10X,"REAL H=",9F6.3)
905   FORMAT(10X,"IMAG H=",9F6.3/)
906   FORMAT(1H1)
      STOP
      END
      SUBROUTINE DFT(M)
C  THIS SUBROUTINE USES THE 1-D FFT SUBROUTINE FFT2C FROM
C  THE IMSL LIBRARY TO PERFORM A 2-D FFT.
      COMPLEX AA(8,8),A(8)
      COMMON AA
      INTEGER IWK(4)
      N=2**M
C  LOAD ROWS OF AA INTO A AND FFT.  RETURN RESULTS TO AA.
      DO 30 I=1,N
      DO 40 J=1,N
      A(J)=AA(I,J)
40    CONTINUE
      CALL FFT2C(A,M,IWK)
      DO 50 J=1,N
      AA(I,J)=A(J)/N
50    CONTINUE
30    CONTINUE
C  LOAD COLUMNS OF AA INTO A AND FFT.  RETURN RESULTS TO AA.
      DO 60 I=1,N
      DO 70 J=1,N
      A(J)=AA(J,I)
70    CONTINUE
      CALL FFT2C(A,M,IWK)
      DO 80 J=1,N
      AA(J,I)=A(J)/N
80    CONTINUE
60    CONTINUE
      RETURN
      END
```

VITA

Clair S. Davis was born on 14 October 1948 in Preston, Idaho. He graduated from high school in Dayton, Idaho in 1966 and attended Utah State University where he completed the requirements for a Bachelor of Science degree in Electrical Engineering in December 1972 and for a Master of Electrical Engineering degree in June 1973. Both degrees were awarded in June 1973. He received a commission in the USAF through the ROTC program in December 1972, and came on active duty following his graduation. He was first assigned to a six month communications-electronics school at Keesler AFB, Mississippi, after which he served for 3½ years as a radar evaluation officer in the 4754th Radar Evaluation Squadron at Hill AFB, Utah. He entered the PhD program in the School of Engineering at the Air Force Institute of Technology in July 1977. He is married to the former Ada Cox of Clifton, Idaho, and they are the parents of five children.

Permanent address: Box 86
Clifton, Idaho
83228

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): AFIT/DS/EE/83-1	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Air Force Institute of Technology; 6b. OFFICE SYMBOL (If applicable): AFIT/EN	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code): Wright-Patterson AFB, Ohio 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: Rome Air Devel. Center; 8b. OFFICE SYMBOL (If applicable): RADC/OCSE	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code): Griffiss AFB, NY	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
62711E	C223	01	03

11. TITLE (Include Security Classification): DECONVOLUTION OF ABERRATIONS IN OPTICAL SYSTEMS (U)

12. PERSONAL AUTHOR(S): Davis, Clair S.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
PhD Dissertation	FROM 79/4/1 TO 83/12/9	83/12/9	150

16. SUPPLEMENTARY NOTATION

Approved for public release: IAW AFR 190-17
7 Feb 84
Lynn E. Wolaver
Dean for Research and Professional Development
Air Force Institute of Technology (AIC)
Wright-Patterson AFB OH 45433

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
20	6	11
20	6	4

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Deconvolution, Aberrations, HALO, Bilinear, Optical, Telescope, Adaptive

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Methods are developed for estimating aberrations on the elements of a two-element optical system based on knowledge of the input and output fields. The equation is written for the propagation of a scalar, quasi-monochromatic field through the system, and the output field is assumed measurable at a given plane. The resulting convolution integral contains the unknown aberration functions in the integrand. The integral equation is a Fredholm equation of the first kind, and the integrand is a nonlinear function of the aberrations. The integral is completed by the method of stationary phase. Methods for estimating the aberration functions are developed, and the effects of noisy measurements are considered. A computerized algorithm that will estimate the Fourier coefficients of the aberration functions is demonstrated for a simple deconvolution problem.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Capt Clair S. Davis	(513) 476-9022	AFWAL/AAD

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

END

FILMED

DTIC