ESTIMATING THE IMPULSE RESPONSE OF LINEAR, SHIFT-INVARIANT, IMAGE DEGRADING SYSTEMS

by

Anthony Edmund Filip

B.S., Wayne State University
(1967)

M.S., Wayne State University
(1968)

SUBMITTED IN PARTIAL FULFILLMENT OF

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

October, 1972 (i. e. Feb. 1973)

Signature of Author _____
  Department of Electrical Engineering, October 16, 1972

Certified by _____
                                          Thesis Supervisor

Accepted by _____
  Chairman, Departmental Committee on Graduate Students

ESTIMATING THE IMPULSE RESPONSE OF LINEAR,
SHIFT-INVARIANT, IMAGE DEGRADING SYSTEMS

by

Anthony Edmund Filip

ABSTRACT

This study presents a novel method for estimating the
impulse response of a linear, shift-invariant, image de-
grading system. It is unusual because the estimate is
obtained directly from the received image, using only
minimal information about the original, unblurred image.
The technique is based upon the theory of homomorphic
systems, whereby the problem is mapped into a space where
the desired signal component is corrupted by additive
noise. The estimation procedure can be divided into two
parts: estimating the log-magnitude function and the phase
function of the degrading system. While the estimation of
the log-magnitude function is fairly straightforward,
estimation of the phase angle is made difficult because of
the multiple-valued nature of the phase function.

To overcome this problem, a phase smoothing alorithm
was used to make the phase angle continuous. However, the
algorithm requires a large number of phase samples, a
number that is difficult to obtain for two-dimensional
signals. Also, the smoothing algorithm is likely to make
errors whenever the magnitude approaches zero, and these
errors are propagated undiminished by the smoothing
algorithm.

In order to test the accuracy of the estimation
procedure, blurred images were simulated on a computer
using a number of different degrading functions. The
homomorphic estimation procedure was then used to estimate
the degrading function. The accuracy of the estimate was

determined in two ways. A quantitative error measure was obtained by computing the normalized root-mean-squared error, while a subjective judgement on the value of the estimate was made by using it to create an inverse filter to deblur the received image. The effects of additive noise in the received image, coarse quantization in the received image, and windowing were also investigated to determine to what extent they affected the resulting estimate. The overall result was that good quality estimates were obtained for the log-magnitude function, but the estimate of the phase function was rather poor.

THESIS SUPERVISOR: T. S. Huang
TITLE: Associate Professor of Electrical Engineering

to Bonnie

ACKNOWLEDGEMENT

The following people have all contributed to a greater or lesser degree toward the successful completion of this study. First and foremost, I would like to thank my thesis supervisor, Professor T. S. Huang, as well as the other members of my thesis committee: Professor A. V. Oppenheim, Professor W. F. Schreiber, Dr. O. J. Tretiak, and Professor I. T. Young. All of these men contributed materially to this study by offering innumerable comments, suggestions, and ideas.

Also to be commended are the students and staff of the Cognitive Information Processing Group of the Research Laboratory of Electronics. Don Hartman, Jack Bowie, and Eric Jensen were particularly helpful during the early days as I was becoming familiar with the PDP-9 computer system. The staff technicians, normally the unsung heroes of many research programs, also deserve credit for keeping the computer down time to a minimum.

Finally, I would like to thank my beautiful wife, Bonnie, whose unwavering support and encouragement helped me through these long years of study.

TABLE OF CONTENTS

6

7

1. INTRODUCTION

One problem that often arises in the field of image enhancement is that of restoring an image after it has been degraded by a linear, shift-invariant (LSI) system. Such degradations may occur through several mechanisms, such as diffraction limited imaging, defocused lenses, relative motion between the object and the recording device, and long duration exposures through turbulent media. Because of the LSI character of these degrading mechanisms, they are "easily" analyzed using linear system theory. Due to the ease of analysis and the wide applicability of the LSI model, many papers have appeared in recent years offering a number of unique and interesting solutions to the problem of image restoration, assuming that the degrading mechanism is LSI. A good review of these techniques is contained in references (4,16,17,45).

One common tie among the various restoration procedures is the fact that they all depend upon knowing the degrading system impulse response or its Fourier transform. The only exception being Slepian's work (43), where he treats the impulse response as a stochastic process and formulates his solution in terms of the second order moments of the Fourier transform of the impulse response of the degrading system.

While it is possible in many cases to calculate analytically the impulse response of a degrading system if the degrading mechanism is known exactly, one is often confronted with the case where the exact degrading mechanism is unknown. For example, camera motion aboard a satellite may be due to vibration of undetermined origin or perhaps random meteor strikes. In such cases, it is impossible to calculate the impulse response of the degrading system. Thus, one is faced with the problem of finding or estimating the impulse response before any actual image restoration can begin.

This paper presents a technique that allows one to estimate the degrading system impulse response using only the received image and the second order moments of the original image process. The estimating procedure uses a technique taken from Oppenheim's theory of homomorphic systems (29). It basically involves using a nonlinear operator in the frequency domain to transform the estimation problem to a space where the signal and noise components are combined additively. The practical utility of the homomorphic estimation scheme is tested by first convolving an image with a known impulse response and then trying to estimate the impulse response from the blurred image. The quality of the estimate is judged in two ways. First, the normalized, root-mean-squared error is calcu-

lated as a quantitative measure of the error in the estimate. Then the estimated degrading function is used to restore the degraded image and the resulting restoration judged qualitatively.

This study is organized in a logical sequence, beginning with a review, in Chapters 2 and 3, of the necessary background material. Chapter 2 discusses basic image enhancement techniques for LSI degrading systems, while Chapter 3 presents some typical degrading functions that arise from both diffraction limited systems and motion-degraded images. In contrast with these cases, where the degrading function can be computed analytically, are those situations in which one attempts to derive the degrading function from observations on the received image. Chapter 4 presents one such technique based upon the known (a priori) presence of certain key objects in the original image such as lines, edges, or pinpoints of light.

Finally, the last three chapters present a method for estimating the degrading system impulse using homomorphic systems. Chapter 5 discusses the theoretical basis for the method as well as the practical considerations necessary for its implementation. Chapter 6 is purely exemplary in nature, presenting several examples in the use of homomorphic estimation. The last chapter (7) presents three possible alternatives to estimating the phase angle function

11

of the degrading system, which turns out to be the most
difficult part of the homomorphic estimation procedure.

2. IMAGE ENHANCEMENT

The goal of any image enhancement technique is to improve the subjective quality of a degraded image. Image quality may suffer due to a number of factors, such as poor contrast, lack of sharpness, geometric distortion, lack of detail in the shadow and highlight areas, additive noise components, etc. However, in this paper I shall concentrate on degradations which can be modelled as a linear, shift-invariant filter. This type of degradation is usually low-pass in nature and, hence, manifests itself as a fuzziness or lack of sharpness in the image.

Under this model, the received image, $r(n,m)$, is obtained by convolving the original image, $s(n,m)$ with the impulse response, or point spread function (PSF), of the degrading system, i.e.

$$r(n,m) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} s(k,l)\, h(n-k,m-l)$$

$$= s(n,m) * h(n,m) \tag{2.1}$$

Now, in order to recover s from r, given h, one must, in general, solve an infinite number of linear, simultaneous equations for an infinite number of unknowns, $s(n,m)$, clearly an impossible task. If one requires that both s and h be nonzero over only a finite portion of the n,m

13

plane, then the problem reduces to one of solving N equations in N unknowns. However, even for pictures of moderate size, say 64 x 64, the number of equations to be solved is 4096, a formidable, if not impossible, task on today's computers. Another time-domain recovery technique was developed by Silverman at Brown University in his doctoral thesis (42). This technique develops the answer recursively, and furthermore allows either s or h (but not both) to be of infinite extent.

Let us now approach the problem in the frequency domain. By taking the Fourier transform of both sides of Eqn. (2.1), one obtains the familiar result

$$R(u,v) = S(u,v) \cdot H(u,v) \tag{2.2}$$

where capital letters denote the digital Fourier transform of the signal having the corresponding small letter and is defined as

$$H(u,v) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h(n,m)e^{-j(un+vm)} \tag{2.3}$$

Thus, in order to recover s(n,m), or equivalently, S(u,v), one transforms both r and h and computes the quotient of the two transforms,

$$S(u,v) = \frac{R(u,v)}{H(u,v)} \tag{2.4}$$

14

This technique, which I call simple inverse filtering, is
indeterminate at those locations where H(u,v) = 0. The
problem becomes even more acute when one includes the ever-
present additive noise component in the model for the de-
grading system as follows:

$$r(n,m) = s(n,m) * h(n,m) + n(n,m) \qquad (2.5)$$

Taking Fourier transforms, one obtains

$$R(u,v) = S(u,v) H(u,v) + N(u,v) \qquad (2.6)$$

and performing simple inverse filtering yields

$$\widetilde{S}(u,v) = \frac{R(u,v)}{H(u,v)} = S(u,v) + \frac{N(u,v)}{H(u,v)} \qquad (2.7)$$

Now, the estimate $\widetilde{S}(u,v)$ suffers in two respects. First,
assuming the H(u,v) is basically a low-pass filter, and
that N(u,v) has a fairly flat spectral distribution, it is
clear that Eqn. (2.7) tends to emphasize high frequency
noise. Furthermore, the restoration scheme breaks down at
those points where H(u,v) = 0. Although simple inverse
filtering has serious drawbacks, some useful restorations
have been achieved for certain special cases, such as
Harris (14).

In an effort to overcome the problems inherent in
simple inverse filtering, researchers have turned to ran-

15

dom process theory. Using this approach, both the images and the noise are treated as random processes whose first and second order moments are known. The goal is to design a filter that will optimize the estimate, s(n,m), according to some error criterion. To be most effective, this error criterion should take into account the properties of the human visual system. Unfortunately, the current state of the art in image processing has not yet reached the point of defining a quantitive measure that accurately gauges subjective image quality for a large class of images. Indeed, because of the many factors that affect picture quality (39), it seems doubtful if a widely applicable, yet mathematically tractable measure will ever be found (9,13, 36,37,40). Thus, it has been conventional to use a minimum, mean squared error (MMSE) criteria, not because it is a good measure of subjective quality, but because its simplicity allows a solution to the filtering problem to be computed.

Using the MMSE criteria, and assuming that the noise is uncorrelated with the original image, Helstrom (15) has shown that the Fourier transform of the optimum inverse filter is

$$G(u,v) = \frac{H^*(u,v)}{\left|H(u,v)\right|^2 + \Phi_{nn}(u,v)/\Phi_{ss}(u,v)} \qquad (2.8)$$

where $\Phi_{ss}(u,v)$ and $\Phi_{nn}(u,v)$ are the power spectral densities of the signal and noise respectively. Examination of Eqn. (2.8) reveals that the optimum inverse filter is free of the two drawbacks of the simple inverse filter, i.e. high frequency noise is attenuated and zeroes in $H(u,v)$ do not result in $G(u,v)$ blowing up. Another interesting point is that for the noise-free case ($\Phi_{nn}=0$), $G(u,v)$ reduces to the simple inverse filter.

The optimum inverse solution was extended by Slepian (43) to the case where h is considered to be a sample function of a random process. His solution is

$$G(u,v) = \frac{E\left\{H^*(u,v)\right\}}{E\left\{|H|^2\right\} + \Phi_{nn}/\Phi_{ss}} \tag{2.9}$$

where $E\{\}$ denotes expectation over the ensemble of sample functions.

Finally, several researchers have attempted to estimate $S(u,v)$ outside the passband of $H(u,v)$ by extrapolating the frequency response with prolate spheroidal wave functions (3,10,17,35).

Thus, with the exception of Slepian's work, it should be noted that regardless of the level of sophistication of the restoration procedure, either the impulse response of the degrading system, or its Fourier transform, must be known. A block diagram for the general degrading and

17

enhancing system is shown in Fig. 2.1.

Figure 2.1

18
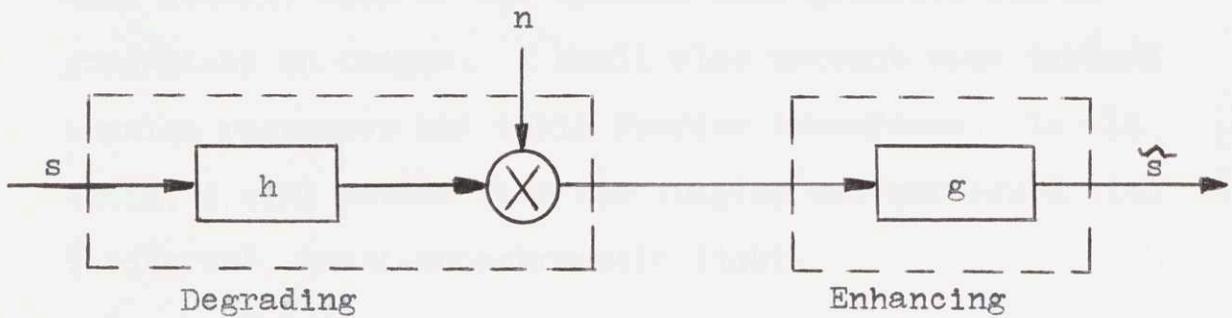
3. LSI DEGRADING SYSTEMS

In this chapter I would like to discuss, in somewhat more detail, some of the systems that generate LSI degradations in images. I shall also present some typical impulse responses and their Fourier transforms. In all cases, I will assume that the imaging was performed with incoherent, quasi-monochromatic light.

3.1 Diffraction Limited Systems

Let us begin by examining degradations in diffraction limited optical imaging systems. This type of degradation is fundamental, being present even in the most carefully designed systems. Its effect, however, usually goes unnoticed because it is buried beneath the more visible effect of lnes aberration or geometric distortion. Assuming that the lenses are perfect, one finds that the image degradation can be modelled as a result of diffraction effects caused by the finite size of either the entrance or exit pupils of the imaging system. The Fourier transform of the degrading impulse response function, usually called the optical transfer function (OTF), can be shown (11) to equal

$$H(u,v) = \frac{\iint\limits_{-\infty}^{\infty} P(\alpha - \frac{\lambda d_1}{2}u, \beta - \frac{\lambda d_1}{2}v)P^*(\alpha + \frac{\lambda d_1}{2}u, \beta + \frac{\lambda d_1}{2}v)\,d\alpha\,d\beta}{\iint\limits_{-\infty}^{\infty} \left|P(\alpha,\beta)\right|^2 \,d\alpha\,d\beta}$$

(3.1)

where λ is the wavelength of light, d_i is the distance between the exit pupil and the image, and $P(x,y)$ is the exit pupil function. The integral in the denominator of Eqn. (3.1) is a normalization constant conventionally employed to remove effects of background illumination.

For aberration-free systems, $P(x,y)$ is either 0 or 1 depending on whether the point (x,y) is outside or inside the pupil. Under this condition, the denominator in Eqn. (3.1) is merely the total area of the exit pupil. There is also a convenient geometrical interpretation for the numerator. Take the pupil function and shift it first to the co-ordinates $(\frac{\lambda d_i}{2} u, \frac{\lambda d_i}{2} v)$ and then to the co-ordinates $(\frac{-\lambda d_i}{2} u, \frac{-\lambda d_i}{2} v)$. The integral of the product of the shifted pupil functions is then equal to the area of overlap between the two shifted pupil functions. So Eqn. (3.1) can be restated in purely geometrical terms as

$$H(u,v) = \frac{\text{Area of Overlap}}{\text{Total Area}} \tag{3.2}$$

If this formula is applied to a circular aperture of radius R, one finds that $H(u,v)$ is also circularly symmetric, and equal to

$$H(r) = \begin{cases} \dfrac{2}{\pi}\left[\cos^{-1}\!\left(\dfrac{r}{r_o}\right) - \left(\dfrac{r}{r_o}\right)\sqrt{1 - \left(\dfrac{r}{r_o}\right)^2}\,\right] & r \leq r_o \\[3mm] 0 & r \geq r_o \end{cases}$$

$$(3.3)$$

where $r_o = \dfrac{2R}{\lambda d_i}$ and r is a radial frequency variable equal

to $r = \sqrt{u^2 + v^2}$. A cross section of this transfer function

is shown in Fig. 3.1 and is seen to be of a low-pass type

with cutoff frequency r_o.
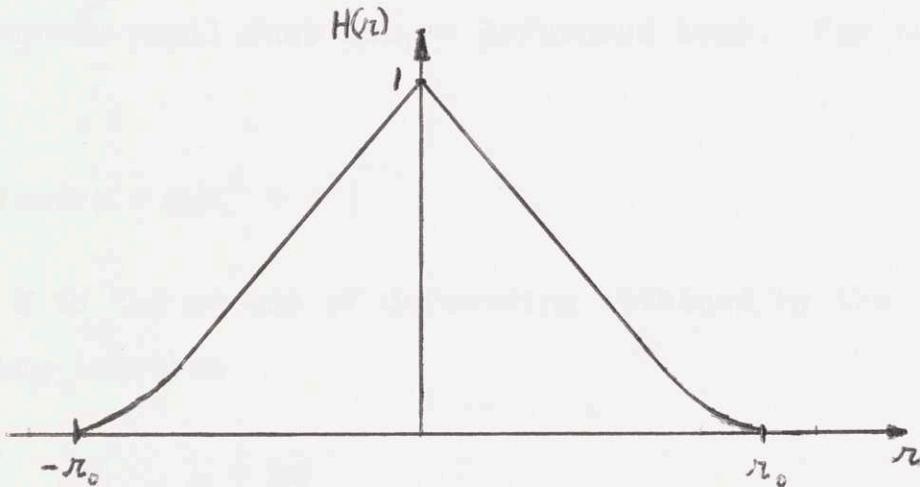
Figure 3.1

The impulse response of this OTF is also circularly

symmetric and equals

$$h(\rho) = \frac{1}{\pi} \left[\frac{J_1(r_0\pi\rho)}{\rho} \right]^2 \tag{3.4}$$

where ρ is a circular spatial variable, $\rho^2 = x^2 + y^2$.

A more interesting class of transfer functions is obtained if we generalize the form of $P(x,y)$ to include a complex exponential factor, i.e.

$$P(x,y) = A(x,y) \exp(jkW(x,y)). \tag{3.5}$$

In most cases, A is assumed to be unity inside the pupil and zero outside the pupil. Consider, as an example of the generalized pupil function, a defocused lens. For this case,

$$W(x,y) = \epsilon(x^2 + y^2) \tag{3.6}$$

where ϵ is the amount of defocusing obtained by the focusing codition

$$\frac{1}{d_o} + \frac{1}{d_i} - \frac{1}{f} = 2\epsilon \tag{3.7}$$

To find $H(u,v)$, the integral in Eqn. (3.1) must be evaluated. For a circular aperture of radius R, the resulting OTF is again circularly symmetric and equals

22

$$H(r) = \begin{cases} \dfrac{4}{\pi} \displaystyle\int_{\beta}^{1} \sqrt{1 - u^2} \; \cos \gamma \beta (u + \beta) \, du & \beta \leq 1 \\[4mm] 0 & \beta \geq 1 \end{cases}$$

$$(3.8)$$

where

$$\beta = \frac{r}{r_o}, \qquad r_o = \frac{2R}{\lambda d_i}$$

$$(3.9)$$

and

$$\gamma = \frac{2\pi \epsilon D^2}{\lambda}, \qquad D = 2R$$

This rather cumbersome expression can be expressed in terms of the incomplete Poisson integral, $P_o(x,y)$, which has been tabulated by Steel (47). With

$$P_o(x,y) = \frac{4}{\pi} \int_{y}^{1} \sqrt{1 - t^2} \; e^{jxt} \, dt,$$

$H(r)$ becomes

$$H(r) = \begin{cases} Re \left[e^{j\gamma \beta^2} \; P_o(\gamma \beta, \beta) \right] & \beta < 1 \\[4mm] 0 & \beta \geq 1 \end{cases} \qquad (3.10)$$

Note that Eqn. (3.8) reduces to Eqn. (3.3) when the system is perfectly focused; i.e., $\epsilon = 0$. When plotting Eqn. (3.8), the usual procedure is to draw a family of curves with parameter $2\pi \epsilon D^2$, expressed in fractions of a wavelength. These

23

curves are sketched in Fig. 3.2. Thus, although a defocused system has the same absolute cutoff frequency, $r = r_o$, the response falls off much more quickly than in the focused case, and, in addition, may contain areas of contrast reversal, i.e. $H(r) < 0$.

$H(r)$

$2\pi \epsilon D^2 = 0$

$.1\lambda$

1.0λ

r_o

r
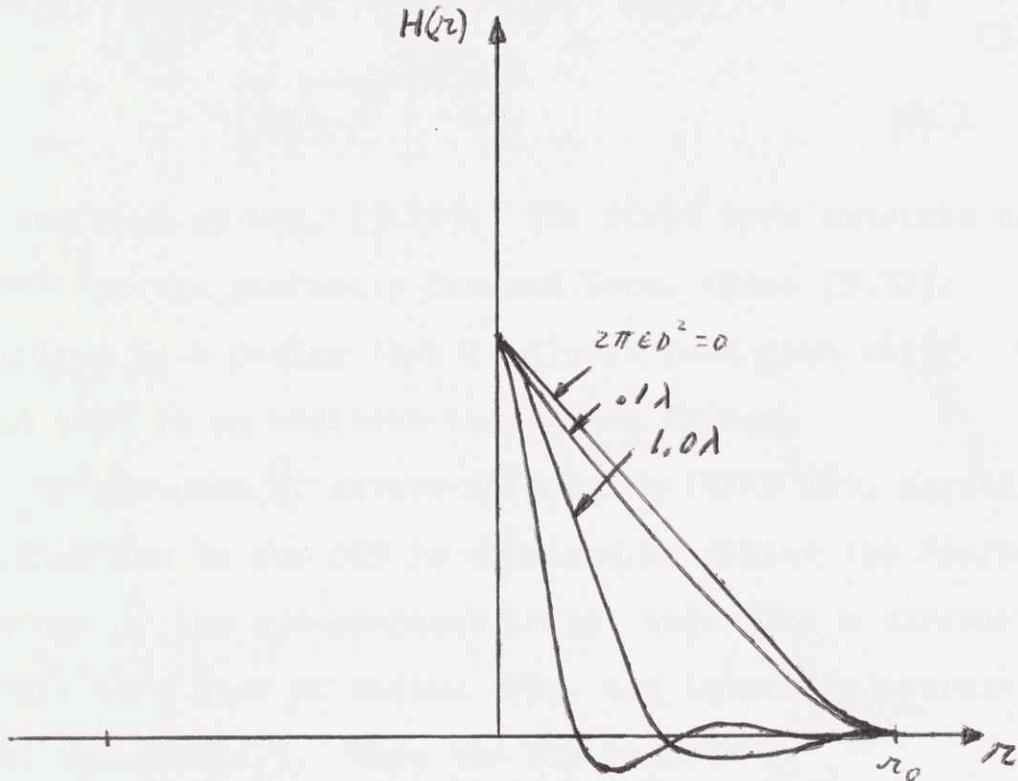
Figure 3.2

For the limiting cases of minor and severe defocusing, simplified expressions for the OTF can be found. For minor defocusing, i.e. $2\pi \epsilon D^2 << \lambda$, the cosine factor in Eqn. (3.8) can be approximated by the first two terms of its Taylor series expansion. Thus,

24

$$H(r) \simeq \frac{4}{\pi} \int_\beta^1 \sqrt{1-u^2}\left[1 - \frac{[\gamma\beta(u+\beta)]^2}{2}\right] du \qquad \beta < 1$$

$$(3.11)$$

After a rather lengthy integration, this yields

$$H(r) = \frac{2}{\pi}\left[\cos^{-1}\beta - \beta\sqrt{1-\beta^2}\right]\left[1 - \frac{\gamma^2\beta^2}{8}(4\beta^2+1)\right]$$

$$(3.12)$$

$$+ \left[\beta\sqrt{1-\beta^2}\right]^3 \frac{5\gamma^2}{6\pi} \qquad \beta < 1$$

Note the form of Eqn. (3.12). The first term consists of the OTF for the perfectly focused lens, (Eqn. (3.3)), multiplied by a factor that is always less than unity. The second term is an additive correction factor.

For the case of severe defocusing ($2\pi\epsilon D^2 > \lambda$), a good approximation to the OTF is obtained by taking the Fourier transform of the out-of-focus image, which for a circular aperture is a disk of radius $2\epsilon R d_i$ and intensity proportional to $1/(2\epsilon R d_i)$. Then the OTF becomes

$$H(r) = \frac{J_1(4\pi\epsilon R d_i r)}{r}$$

$$(3.13)$$

Finally, Steel (48) has presented an empirically derived approximation which gives results accurate to within 2% for all values of defocusing. The equation is

$$H(r) \simeq \frac{2}{\pi}\left[\cos^{-1}\beta - \beta\sqrt{1-\beta^2}\right]\frac{2J_1(\xi)}{\xi} \qquad \beta < 1 \qquad (3.14)$$

where

$$\mathcal{E} = \gamma \beta (1 - \beta) \qquad (3.15)$$

3.2 Motion-Degraded Images

As a second area, let us consider motion-degraded images. Assume that the recording device (e.g. camera) moves perpendicularly to the line joining it to the object while the image is being recorded. One convenient way of describing the motion is to decompose it into its x and y components, each of which is expressed in terms of the time parameter, t.

$$\overline{R}(t) = x(t)\overline{i} + y(t)\overline{j} \qquad (3.16)$$

If s denotes distance along the path R, the velocity along that path is

$$\frac{ds}{dt} = \left|\frac{dR}{dt}\right| = (x'^2(t) + y'^2(t))^{\frac{1}{2}} \qquad (3.17)$$

Now, for a point object, the film exposure is, by definition, the impulse response of the motion degraded image. The exposure, in turn, is proportional to the incremental time spent at the location (x,y). Thus,

$$E(x,y) = I_0 \Delta t(x,y) \qquad (3.18)$$

where E is the exposure as a function of position, and I_0 is a proportionality constant related to the intensity of the light impinging on the film. Now,

26

$$\Delta t \approx \frac{\Delta s}{ds/dt} \tag{3.19}$$

$$E(x,y) = \frac{K}{\left[x'^2(t) + y'^2(t)\right]^{\frac{1}{2}}} = h(x,y) \tag{3.20}$$

where K is a constant equal to $I_o \Delta s$.

This is the general equation for the impulse response in terms of the equation of relative motion between the object and recording device. Some examples will be presented below to illustrate the type of impulse response that can arise due to image motion.

1) Linear Motion in One Dimension

$$\text{Let } x(t) = \begin{cases} v_o t & -T/2 \leq t \leq T/2 \\ 0 & \text{Elsewhere} \end{cases} \tag{3.21}$$

and let $y(t)$ be constant. From Eqn. (3.8), we get

$$h(x) = \frac{K}{v_o} \qquad -(T/2)v_o \leq x \leq (T/2)v_o \tag{3.22}$$

which is shown in Fig. 3.3. The Fourier transform of $h(x)$ is

$$H(u,v) = K \frac{\sin \pi T v_o u}{\pi v_o u} \tag{3.23}$$
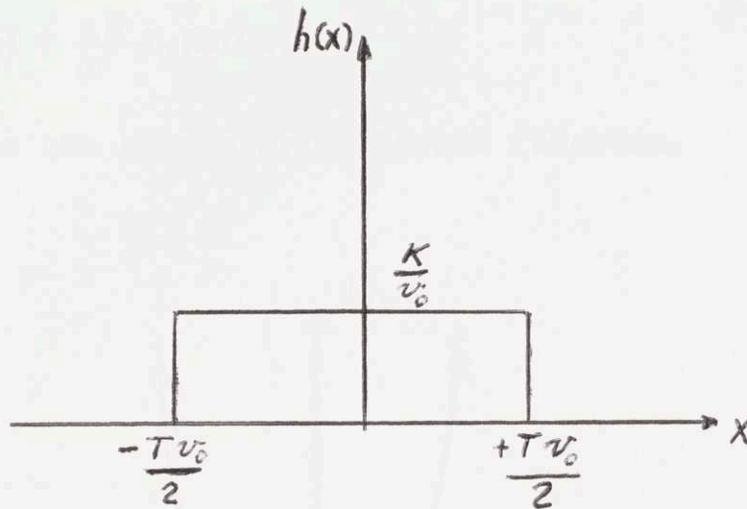
Figure 3.3

2) Harmonic Motion in One Dimension

$$x(t) = A \sin \omega t \qquad (3.24)$$

Plugging the expressions for x(t) and y(t) (y(t) is constant) into Eqn. (3.20), one obtains

$$h(x) = \frac{K}{\omega \sqrt{A^2 - x^2}} \qquad -A \leq x \leq A \qquad (3.25)$$

as shown in Fig. 3.4. Strictly speaking, this result assumes that an integral number of cycles occurred during the exposure time, T_e. However, if $\frac{2\pi}{\omega} \ll T_e$, then Eqn. (3.25) is a good approximation, even if $n\frac{2\pi}{\omega} \neq T_e$. The Fourier transform of h(x) is

28

$$H(u,v) = \frac{K\pi}{\omega} J_O(2\pi uA) \tag{3.26}$$

where J_O is the zeroth order Bessel function.
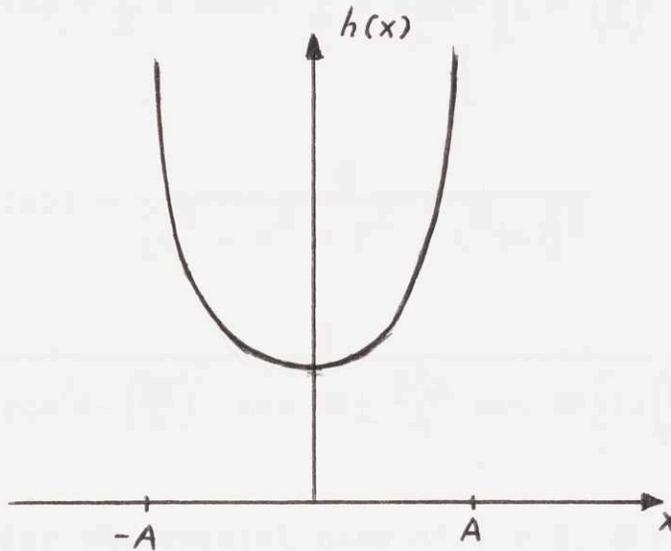
Figure 3.4

3) Harmonic Motion in Two Dimensions

Assume the equations of relative motion are:

$$x(t) = A \sin\omega t, \qquad y(t) = B \sin(\omega t + \theta) \tag{3.27}$$

Thus, the camera traces out an elliptical path during the exposure period. Plugging into Eqn. (3.20), one obtains

$$h(x,y) = \frac{K}{(A^2 - x^2 + B^2 - y^2)^{\frac{1}{2}}} \tag{3.28}$$

29

Now, y is not free to take on all values in the domain of h because it bears a fixed relationship to the independent variable x through the parameter t. That is,

$$y = g(x) = \frac{B}{A} x \cos\theta \pm B \sin\theta \left[1 - \left(\frac{x}{A}\right)^2 \right]^{\frac{1}{2}} \tag{3.29}$$

Thus,

$$h(x, g(x)) = \frac{K}{\left[A^2 - x^2 + B^2 - g^2(x) \right]^{\frac{1}{2}}} \tag{3.30}$$

$$= \frac{K}{\left[A^2 - x^2 + B^2 \cos^2\theta - \left(\frac{Bx}{A}\right)^2 \cos 2\theta \pm \frac{B^2 x}{A} \sin 2\theta \left[1 - \left(\frac{x}{A}\right)^2 \right]^{\frac{1}{2}} \right]^{\frac{1}{2}}}$$

Note that for the special case of $A = B$, $\theta = \frac{\pi}{2}$

$$h(x, g(x)) = h(x, \pm (A^2 - x^2)^{\frac{1}{2}}) = \frac{K}{\omega A} \tag{3.31}$$

i.e. $h(x, g(x))$ is an impulse ring of radius A and magnitude K/A. The Fourier transform for this special case is

$$H(r) = \frac{2\pi K}{\omega} J_0(2\pi A r) \tag{3.32}$$

There is another approach to the characterization of motion degraded imaging systems. This method yields an expression for the OTF directly in terms of the equations of motion. We begin by expressing $h(x,y)$ as an integral:

30

$$h(\overline{x}, \overline{y}) = \int_{0}^{T} u_O \; (\overline{x} - x(t), \; \overline{y} - y(t)) \; dt \qquad (3.33)$$

where u_O is a two-dimensional impulse located at $(x(t),$ $y(t))$. The OTF for $h(x,y)$ is obtained by simply utilizing the sifting property of impulses. Thus,

$$H(u,v) = \int_{0}^{T} e^{-j2\pi(ux(t)+vy(t))} dt \qquad (3.34)$$

Sometimes this approach is easier than using Eqn. (3.20) and then computing the OTF. A more thorough discussion of motion-degraded images can be found in the references (28, 41,44).

From the preceding examples, it is apparent that the analytic calculation of the impulse response, or its Fourier transform, rapidly becomes extremely difficult. However, numerical methods can and have been applied with some success to this problem. In particular, the use of the Fast Fourier Transform (FFT) algorithm and high speed convolution has been of tremendous value in the computation of OTF's for diffraction limited systems (12,23).

4. DIRECT ESTIMATION OF THE IMPULSE RESPONSE

In the previous chapters, several examples were pre-
sented showing how the impulse response of an image de-
grading system, often called the point spread function
(PSF), can be determined either analytically or numeri-
cally, provided one has an accurate, detailed model for
the degrading mechanism. The procedures were based upon
having maximum a priori information about the degrading
system, but nothing was stated about the nature of the
original picture. In this chapter, I would like to review
three methods which allow the PSF to be estimated directly
from the received image. These methods require very little
a priori information about the impulse response but do make
certain assumptions about the nature of the original image.

The simplest case is one in which the original image
consists of one or more widely spaced pinpoints of light,
such as astronomical photographs. If these pinpoints of
light can be modelled as impulses, then the image on the
photographic plate is, by definition, the impulse response
of the image degrading system.

The only restriction on the impulse response is that
its effective width be narrow enough so that there is no
overlap between the images of different points. There are
two practical problems with this procedure, however. The

first is that the received image usually has an additive
noise component which creates large errors in the measure-
ment of the impulse response from a single pinpoint. This
problem can be alleviated by averaging the measured impulse
response over many pinpoints.

The second problem is, of course, the rather limited
number of scenes that are known to contain widely separated
impulse-like objects. However, there are two other shapes
that are useful in estimating PSF's from the received image.
These are straight lines and edges (Fig. 4.1a and 4.1b).
In addition, these shapes are frequently found in images,
particularly those of man-made objects.

Clearly, the straight line, or impulse sheet, is
simply the directional derivative of the edge function.
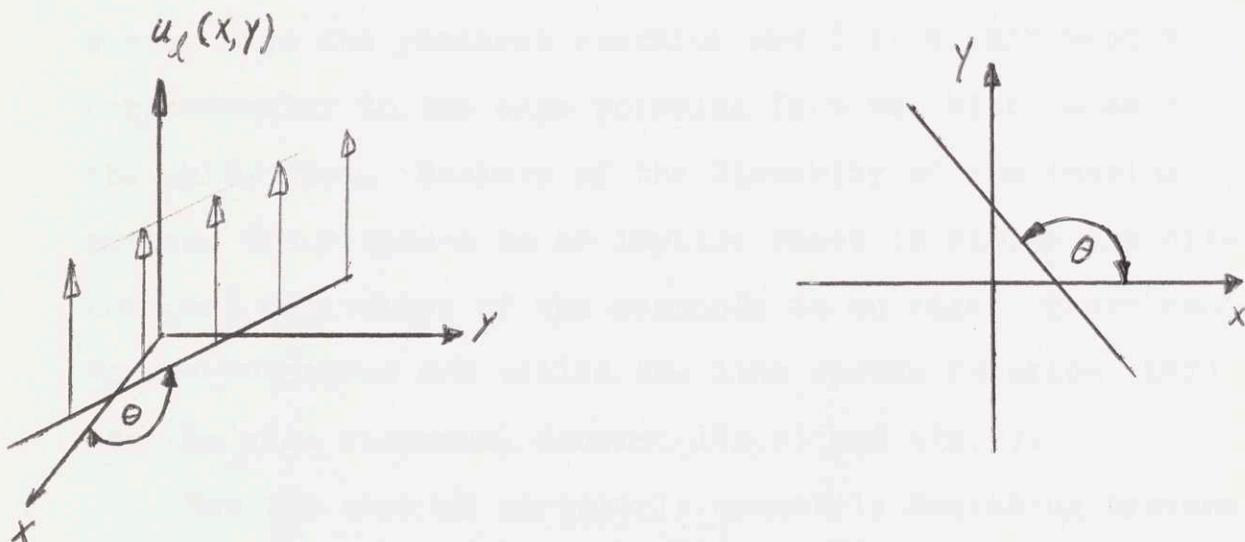
Figure 4.1a
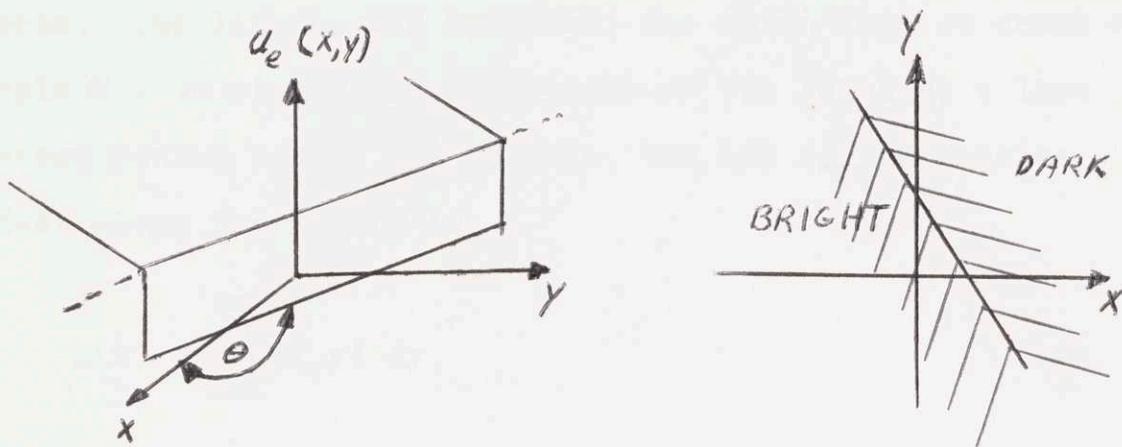
Figure 4.1b

Thus, if $u_e(x,y)$ denotes the edge function and $u_1(x,y)$
denotes the impulse sheet, then

$$u_1(x,y) = \overline{\nabla} u_e(x,y) \cdot \overline{p} \qquad (4.1)$$

where $\overline{\nabla}$ is the gradient operator and $\overline{p}$ is a unit vector
perpendicular to the edge pointing from the black area to
the white area. Because of the linearity of the imaging
system, the response to an impulse sheet is simply the dir-
ectional derivative of the response to an edge. These re-
sponse functions are called the line spread function (LSF)
and the edge response, denoted $l(x,y)$ and $e(x,y)$.

For the case of circularly symmetric degrading systems,
it is possible to calculate the PSF from the LSF and vice

34

versa. The line spread function, due to an impulse sheet at
angle θ , is simply the projection of the PSF onto a line
perpendicular to θ. For example, the LSF for an impulse
sheet along the y axis is

$$l(x) = \int_{-\infty}^{\infty} h(x,y) \, dy \qquad (4.2)$$

which is the projection of h along the x axis. For systems
having circular symmetry, Eqn. (4.2) becomes

$$l(x) = \int_{-\infty}^{\infty} h(\sqrt{x^2 + y^2}) dy \qquad (4.3)$$

Eqn. (4.3) can be inverted to give h(r) in terms of l(x)
(2)

$$h(r) = \frac{-1}{\pi r^2} \int_{r}^{\infty} \frac{x \, l(x) + x^2 \, l'(x)}{\sqrt{x^2 - r^2}} \, dx \qquad (4.4)$$

Thus, for those systems having circular symmetry, the
procedure for estimating the PSF is the following. First,
measure either the LSF or the edge response directly from
the received image. Then use Eqns. (4.1 - 4.4) to find
the PSF.

In those cases where the image degrading system is not
circularly symmetric, we require that the original image
contains lines and/or edges at many different orientations.

35

For example, if in a certain region, it is assumed that the original signal was an impulse sheet with slope a,

$$u_1(x,y) = u_0(y-ax) \tag{4.5}$$

then in that region,

$$r(x,y) = u_1(x,y) * h(x,y) \tag{4.6}$$

or, equivalently,

$$R(u,v) = U_1(u,v) \cdot H(u,v) \tag{4.7}$$

Now

$$U_1(u,v) = \iint\limits_{-\infty}^{\infty} u_1(x,y)e^{-j2\pi(ux+vy)}\,dxdy \tag{4.8}$$

$$= u_0(u + av)$$

This result is depicted in Fig. 4.2.

Figure 4.2

36

So,

$$R(u,v) = u_o(u + av) \cdot H(u,v)$$
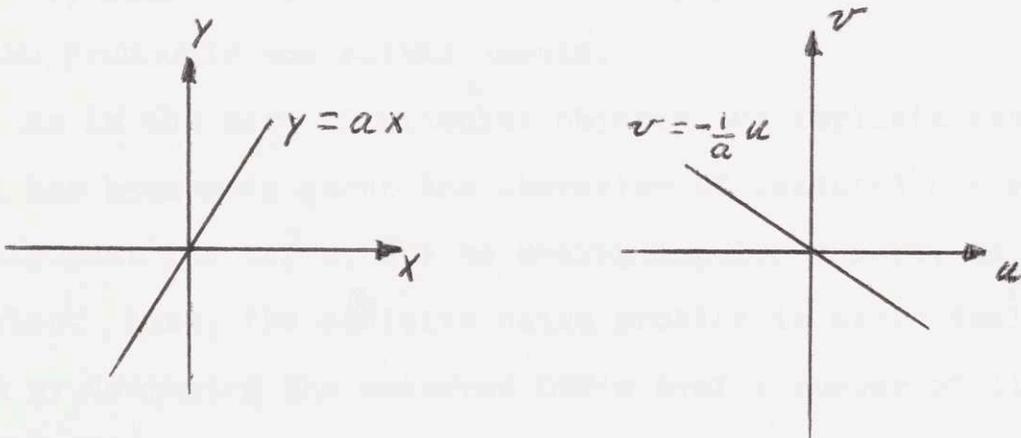
$$= u_o(u + av) \cdot H(u, -u/a) \tag{4.9}$$

That is, the Fourier transform of the response to an impulse sheet with slope a is the Fourier transform of the impulse response sampled along the line, u + av = 0.

Thus, if the picture contains impulse sheets at many orientations, one can very simply obtain samples of the Fourier transform of the unknown impulse response along radial lines. Then, in order to obtain an estimate for the impulse response, the space between the radial samples must be filled in using some interpolation scheme to yield $H(u,v)$ and finally this must be inverse Fourier transformed (51). The interpolation procedure is a non-trivial task. However, Mersereau (27) has shown that for a certain limited class of functions, it is theoretically possible to obtain $H(u,v)$ from only one radial sample.

As in the case of pinpoint objects, an implicit assumption has been made about the character of measured LSF's, namely that the LSF's, due to nearby impulse sheets, do not overlap. Also, the additive noise problem is again dealt with by averaging the measured LSF's over a number of lines as in (5).

However, as the image starts to contain more detail, the direct estimation method breaks down because of the inability to find isolated points, lines, or edges. For highly detailed images, another method must be found.

5. ESTIMATING THE IMPULSE RESPONSE USING HOMOMORPHIC SYSTEMS

5.1 Introduction

In the last chapter it was shown how the impulse response can be measured directly from the received image. The procedure depended upon the original scene containing certain key objects that were reasonable approximations to either an impulse, an impulse sheet, or an ideal edge. There are, however, many images that don't contain any of these objects, or the key objects may be so close together that the PSF's or LSF's overlap and, hence, cannot be accurately measured. For these highly detailed images, a novel approach to estimating the PSF is presented, based upon homomorphic systems. This new method was first used by Stockham in an attempt to restore old Caruso recordings (49). It compliments the direct measurement technique in that its utility tends to increase as the objects contain more random structure. Indeed, this method models the original image as a two-dimensional random field. Before discussing the homomorphic filtering approach, I shall review two results which also model the image as a random field to illustrate what homomorphic filtering offers over the more conventional approach.

The first is the well known result in random process theory relating the power spectral density (PSD) of the output function to the PSD of the input function,

$$\Phi_{rr}(u,v) = \Phi_{ss}(u,v)\left|H(u,v)\right|^2 \tag{5.1}$$

Solving for $|H|$, one obtains

$$\left|H(u,v)\right| = \left\{\Phi_{rr}(u,v)/\Phi_{ss}(u,v)\right\}^{\frac{1}{2}} \tag{5.2}$$

Thus, assuming one knows the PSD of the original image process, and one can estimate the PSD of the received image, the magnitude of $H(u,v)$ can be found directly. Notice, however, that no information about the phase angle of $H(u,v)$ is obtained through Eqn. (5.2).

A second approach is to use maximum liklihood estimation to obtain a point-by-point estimate of $H(u,v)$. The problem is the following: given N observations of the form

$$R_r^k(u,v) = H_r S_r^k - H_i S_i^k$$
$$\qquad\qquad\qquad\qquad k=1,\dots N \tag{5.3}$$
$$R_i^k(u,v) = H_r S_i^k + H_i S_r^k$$

estimate H_r and H_i, assuming the joint statistics of S_r and S_i are known. The subscripts "r" and "i" indicate the real and imaginary parts of a function, and the superscript "k" denotes the kth sample. I shall further assume that S_i and S_r are independent, zero mean, Gaussian random variables, with equal variance. These assumptions are justified on the basis that they are in agreement with experimentally derived statistics. For example, using the new crowd scene (CIPG # 11) as our standard image (Fig. 5.1),

histograms of S_r and S_i and also $\log|S|$ and $\arctan\left(\dfrac{S_i}{S_r}\right)$

were computed. These histograms are shown in Figure 5.2(a)
for S_i (top) and S_r (bottom) and in Figure 5.2(b) for $\log|S|$
(bottom) and $\arctan\left(\dfrac{S_i}{S_r}\right)$ (top). The Gaussian character of

S_r and S_i is apparent. Also, the uniform distribution of
the phase angle indicates that S_r and S_i are, indeed, un-
correlated and, thus, independent. Any correlation between
S_r and S_i would manifest itself as one or more peaks in the
phase angle histogram.

Returning to the maximum liklihood estimation problem,
a few manipulations yield the result that

$$|H(u,v)|^2 = \frac{1}{N\sigma^2} \sum_{k=1}^{N} |R^k(u,v)|^2 \qquad (5.4)$$

Note that once again no information has been obtained
about the phase angle of H. It is this deficiency that I
shall attempt to overcome by means of homomorphic filtering.

One might ask at this point if it is really necessary
to know the phase angle of H(u,v) in order to obtain a
reasonably good restoration. If it turns out that phase
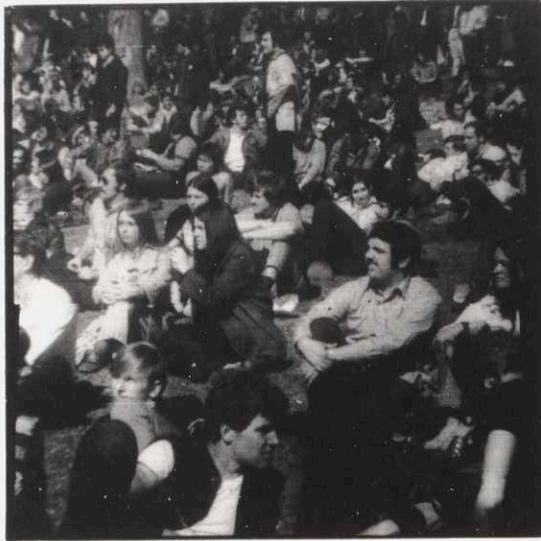errors play only a minor role in determining subjective
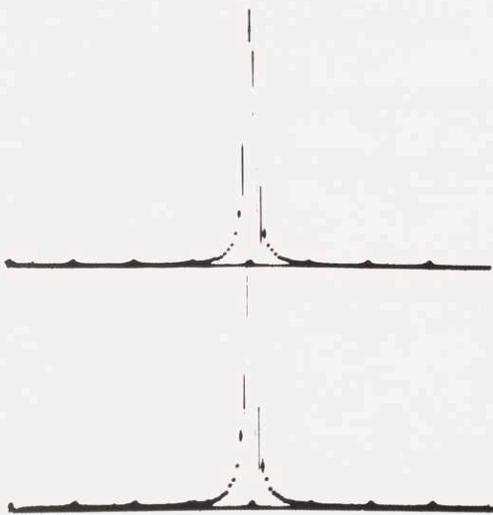
41

Figure 5.1
New Crowd Scene (CIPG #11)

Fig. 5.2(a), Histograms of
S_r (bottom) and S_1 (top)
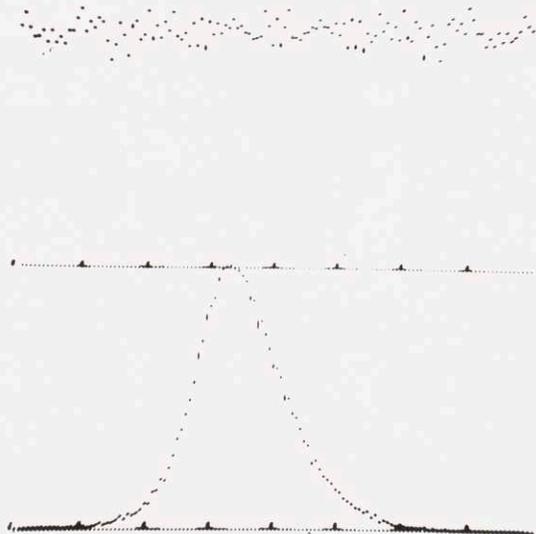
Fig. 5.2(b), Histograms of
$\log |S|$ (bottom) and $PV\angle S$ (top)

picture quality, then one can use either Eqn. (5.2) or Eqn. (5.4) to construct a zero-phase inverse filter, and the homomorphic filtering technique described below need never be applied. Obviously any uncorrected phase errors will result in a less-than-perfect restoration, but it is not clear to what extent phase errors influence picture quality. Anderson's study (1) of the effects of additive noise in the frequency domain indicates that, for small noise levels, phase errors had a less deleterious effect on image quality than an equal amount of noise added to the magnitude function. However, the Fourier transform of a short-duration impulse response, as was assumed in this study, is quite smooth. Hence, the error introduced into the magnitude and phase functions is better described as being systemmatic rather than random in nature.

To learn more about the effects of systemmatic phase errors, a two-dimensional, all-pass filter was created. The Z-transform of a bilinear type all-pass filter is

$$H(z_1, z_2) = \frac{-az_1z_2 + cz_1 + bz_2 + 1}{z_1z_2 + bz_1 + cz_2 - a} \qquad (5.5)$$

where

$$\left| H(e^{ju}, e^{jv}) \right| = 1 \qquad \forall \quad u, v$$

Thus, the effect of the all-pass system is to distort only the phase angle of the signal; the magnitude function is

43

unchanged. The following constraints on the coefficients are necessary to insure stability of the filter when recursing in the +n, +m direction.

$$\left|\frac{a}{b}\right|<1, \qquad \left|\frac{b-a}{c+1}\right|<1, \qquad \left|\frac{b+a}{c-1}\right|<1 \qquad (5.6)$$

Figrue 5.3(a) shows the frequency response of a typical all-pass filter with a = .625, b = .75, c = -.75. The top half of the figure shows the log-magnitude function. The black area is less than .4dB down from the white area, indicating that the magnitude has a maximum variation of about 5% over the entire u-v plane. The bottom half of Figure 5.3(a) illustrates the phase angle with -π shown as black and +π as white. This filter was then implemented recursively and used to distort the phase angle of the image shown in Figure 5.3(b), with the result shown in Figure 5.3(c). The most obvious result is a reduction of contrast accompanied by a gradient-like, edge enhancement effect. Clearly, phase distortion has significantly lowered the subjective picture quality.

As a more extreme example of systemmatic phase error, the image of Figure 5.4(a) was distorted in two, admittedly special ways. The image in Figure 5.4(b) was obtained by setting the magnitude function equal to a constant, while leaving the phase undisturbed. The effect was to high pass the picture, accentuating the edges. But despite the dis-
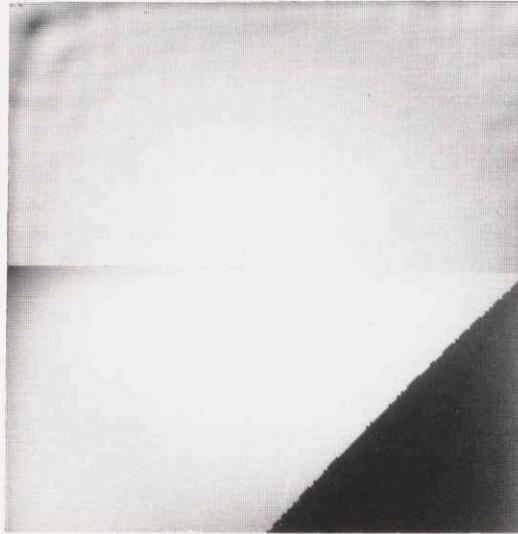
44

Figure 5.3(a)
All Pass Frequency Response
(a=.625, b=.75, c=-.75)
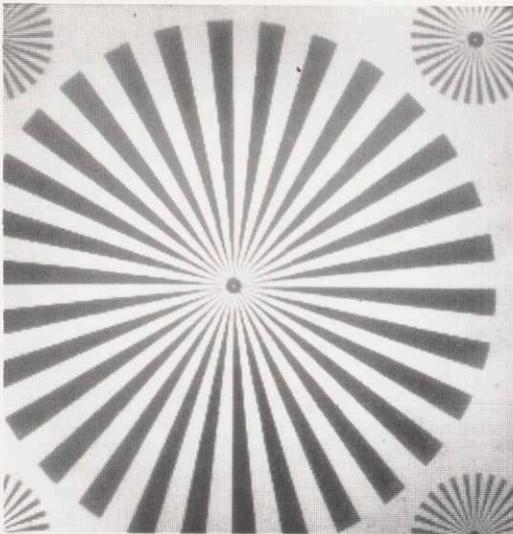.4 dB variation in Log-mag. (top)

Figure 5.3(b)
Input to All-Pass System
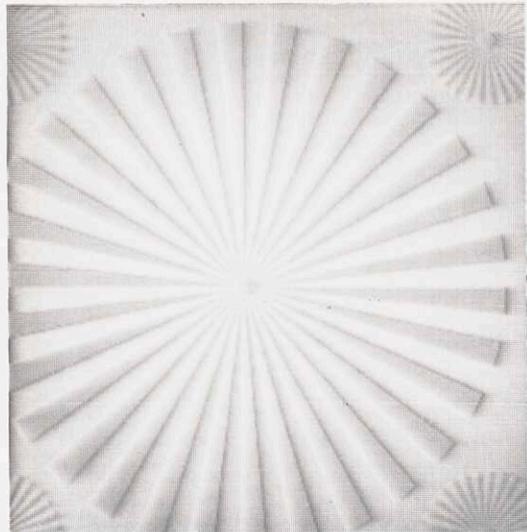
Figure 5.3(c)
Output of All-Pass System
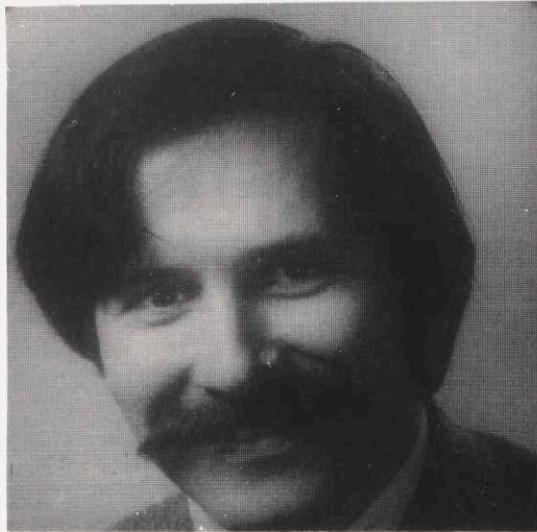
Figure 5.4(a)
Oleh (CIPG #2)
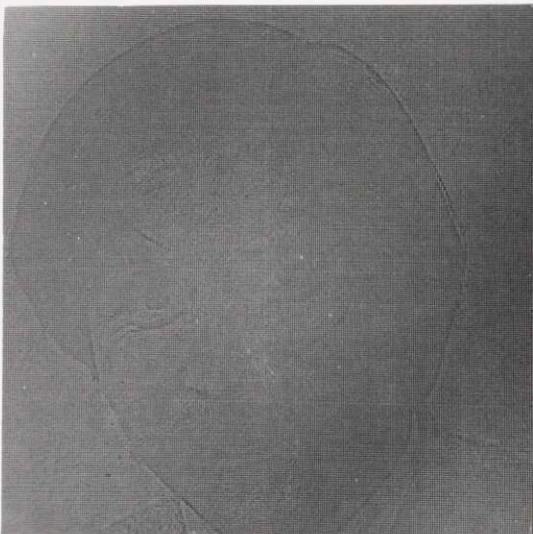
Fig. 5.4(b), Magnitude =
constant, Phase unchanged
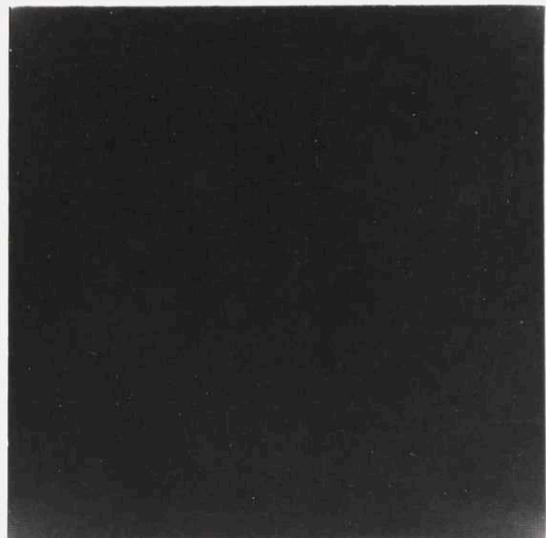
Fig. 5.4(c), Magnitude
unchanged, Phase=0

tortion, the essential character of the image is unchanged; i.e., it is still clearly a face. However, if one now sets the phase angle to a constant (0) and leaves the magnitude unchanged, the effect is dramatically different, as shown in Figure 5.4(c). Now the resulting image is totally unrelated to the original and appears only as four lightened areas on an otherwise featureless background.

What should be apparent from these examples is that the phase function plays an important role in determining image quality and cannot be neglected. As will be shown in the next section, homomorphic filtering provides a means for estimating the phase angle of the degrading filter, and this in turn can be used to construct a restoration filter which corrects for both magnitude and phase distortion.

5.2 Homomorphic Filtering

The usual assumption made in estimation theory is that the desired signal is corrupted by an additive noise component; e.g., $r(t) = s(t) + n(t)$. However there are situations in which the desired signal is corrupted by multiplicative noise or, as in the case of image degrading systems, convolutional noise. For the purposes of this study, the desired signal component is $h(x,y)$ while $s(x,y)$ is considered to be the corrupting convolutional noise whose statistics are known. Since recovery of the original signal, $s(x,y)$, is the ultimate goal, and since the received signal is

47

symmetric with respect to the convolution operator, it would appear desirable to estimate $s(x,y)$ directly. However, as will be shown later, the algorithm used to estimate $h(x,y)$ requires an input data structure quite different from that needed to estimate $s(x,y)$. But more on that later.

The basic problem of estimating a signal corrupted by noise is the same as that found in the classical estimation problem; the only difference is the manner in which the signals are combined. For the classical case, the combination is additive, while for image degrading systems, the method of combination is convolution. Given the basic similarity between the two problems, it would seem desirable to convert the convolutional noise to additive noise so that one could make use of the extensive techniques available for estimating a signal in an additive noise environment. Homomorphic filtering (HF) is the technique used to achieve this conversion.

Before proceeding, I shall adopt the convention that all signals be denoted as discrete functions, in keeping with the fact that all of the processing was done on a digital computer. Also, the estimate of a signal will be denoted by a squiggle over the letter. Thus, $h(x,y)$ becomes $h(n,m)$ and its estimate is denoted by $\tilde{h}(n,m)$, etc.

Homomorphic filtering, or generalized linear filtering, was first studied by Oppenheim (29). Extensions of the

theory and some practical applications were later carried out by Schafer (38), and Oppenheim, et al. (30). What follows is intended only as a brief introduction to the theory of homomorphic filtering since this study uses HF as a tool rather than an end in itself. The interested reader is referred to the references for additional material.

A homomorphic system is literally a linear mapping between vector spaces. Figure 5.5 is a block diagram of the overall generalized linear filter. The boxes labelled T and T^{-1} are homomorphic systems which serve to map from the convolution space to the additive space and inversely. The subsystem L is a standard linear filter which performs the actual estimation in the additive space. T has the property that

$$T\left[s(n,m) * h(n,m)\right] = T\left[s(n,m)\right] + T\left[h(n,m)\right]$$
$$= \hat{S} + \hat{H} \tag{5.7}$$

while T and T^{-1} satisfy the relation

$$T^{-1}\left[T(f)\right] = T\left[T^{-1}(f)\right] \tag{5.8}$$

The nature of the mapping to and from the additive space suggests that T and T^{-1} can each be broken into two separate subsystems shown in Figure 5.6, where FT and IFT denote the two-dimensional digital Fourier transform and

49
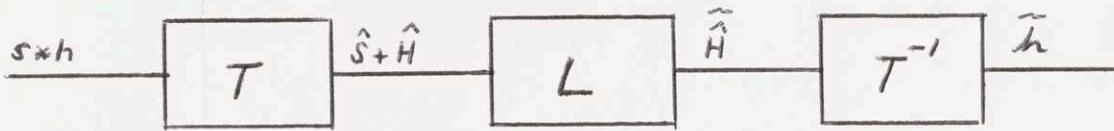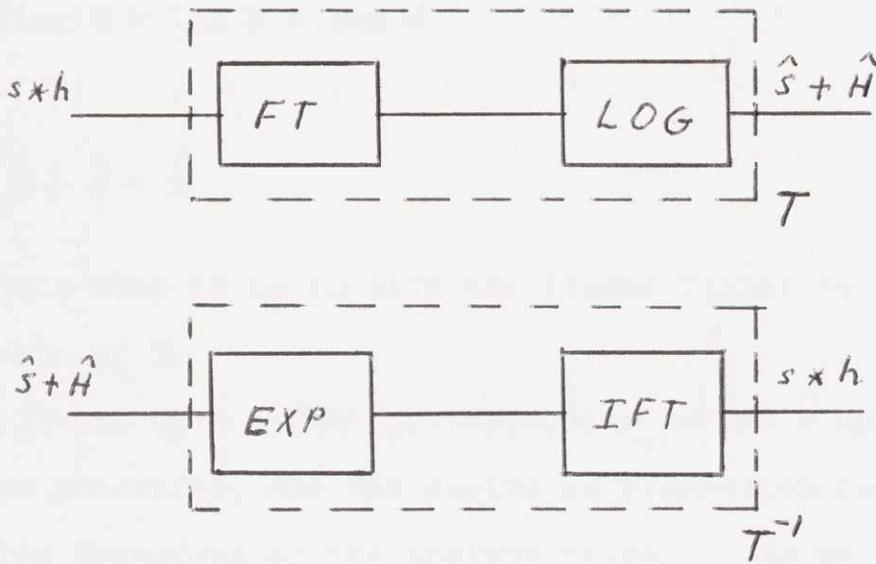
Figure 5.5

Figure 5.6

its inverse:

$$H(u,v) = \sum_{n} \sum_{m} h(n,m)\, e^{-j(un+vm)} \tag{5.9}$$

$$h(n,m) = \frac{1}{(2\pi)^2} \iint_{-\pi}^{\pi} H(u,v)\, e^{j(un+vm)} du\, dv$$

Thus, in going to the additive space, first one com-

putes the Fourier transform of r,

$$r(n,m) = s(n,m) * h(n,m) \tag{5.10}$$

to yield

$$R(u,v) = S(u,v) \cdot H(u,v) \tag{5.11}$$

and then takes the complex logarithm of R,

$$\log R = \log S + \log H \tag{5.12}$$

or

$$\hat{R} = \hat{S} + \hat{H}$$

The next step is to specify the linear filter to extract an estimate of H.

If one uses a MMSE criteria, and models $\hat{S}$ and $\hat{H}$ as random processes, one can derive an expression for the Fourier transform of the optimum filter in terms of the power spectral density functions of $\hat{S}$ and $\hat{H}$. However, these functions are usually unknown and are quite difficult to compute, so a second, simpler procedure was used.

This procedure requires no a priori information about h(n,m), other than assuming that its effective width is relatively narrow as in the direct estimation procedure of Chapter 4. Then one merely computes $\hat{R}$ for a number of different samples and averages the result on a point-by-point basis over the u-v plane. That is,

$$\frac{1}{N} \sum_{i=1}^{N} \hat{R}_i = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{S}_i + \hat{\tilde{H}} \right) = \hat{\tilde{H}}_B(u,v)$$

$$\text{(5.13)}$$

$$= B(u,v) + \hat{\tilde{H}}(u,v)$$

where $B(u,v)$ corresponds to the bias in the estimate and equals the average value of $\hat{S}$ at all points in the u-v plane. Assuming that the mean value of $\hat{S}(u,v)$ is known, the quantity $m_{\hat{S}}(u,v)$ can be subtracted from $\hat{\tilde{H}}_B$ to yield an unbiased, consistent estimate of $\hat{H}$, i.e.

$$E \left\{ \hat{\tilde{H}}_B - m_{\hat{S}} \right\} = \hat{\tilde{H}}$$

$$\text{(5.14)}$$

$$\text{var} \left\{ \hat{\tilde{H}}_B - m_{\hat{S}} \right\} = \frac{1}{N} \text{var} \left(\hat{S} \right)$$

Finally, in order to obtain $h(n,m)$, one must first exponentiate and then inverse Fourier transform the quantity $(\hat{\tilde{H}}_B - m_S)$ by passing it through T^{-1}.

5.3 Subdividing the Image

The first problem in implementing the averaging procedure is where to obtain the N sample functions. There would be no problem if the $\hat{S}_i(u,v)$ corresponded to N different images that had been degraded by the same system. Usually, however, there is only one image available. The technique used is to sub-divide the received image into N subsections. The manner in which the image is divided

depends upon the nature of the degrading system.   In the
most general case, the image is divided into N nonoverlap-
ping squares. So

$$r(n,m) = s(n,m) * h(n,m)$$

$$= \left[\sum_{i=1}^{N} s_i(n,m) \right] * h(n,m)$$

(5.15)

where the s_i are nonoverlapping regions of the original
image, s. Now, if the effective width of h is much less
than the size of a subsection, Eqn. (5.15) can be approxima-
ted by

$$r(n,m) \approx \sum_{i=1}^{N} \left(s_i(n,m) \ * h(n,m) \right)$$

(5.16)

or

$$r_i(n,m) = s_i(n,m) * h(n,m)$$

(5.17)

The only error in the approximation occurs near the boundary
of each subsection where the interaction between subsections
in the convolution Eqn. (5.15) has been neglected. Since
one of the later steps in the procedure involves multiplying
each subsection by a Hanning window, this tends to further
suppress the boundary errors. The process of subdividing
an image also involves a performance trade-off. On the one
hand, one would like the subsection size to be large to re-
duce the relative error at the boundary, or admit degrading

53

functions with larger effective widths. But using large subsections means that N, the total number of sample functions in the averaging procedure, will be smaller, and, hence, from Eqn. (5.14), the variance in the estimate will be higher. The only alternative is to allow the subsecti to overlap to some extent. This introduces another problem in that the subsections will be more highly correlated, thus requiring that N be made even larger before a good average is obtained.

If one is willing to assume that $h(n,m)$ is separable, i.e.

$$h(n,m) = h_1(n) \cdot h_2(m) \tag{5.18}$$

then the problem becomes considerably easier. The reason is that the problem can now be considered as two, one-dimensional problems. Consider what needs to be done to estimate $h_1(n)$. Following Eqn. (5.13) we have

$$H_1(u) = \frac{1}{N} \sum_{i=1}^{N} R_i(u,v) - m_g(u,v) \tag{5.19}$$

The advantage is that the $r_i(n,m)$ could be taken to be one line of the received image, in which case

$$r_i(n,m) = r(n,i) \tag{5.20}$$

Or the $r_i(n,m)$ could be taken to be nonoverlapping sections

54

of lines, etc. For although the same conditions on boundary error, variance of the estimate, and correlation among the subsamples apply also to the separable case, there are so many more ways in which to subdivide the image that these constraints are easily satisfied. Similarly, in estimating $h_2(m)$, one might choose to subdivide r into columns, etc.

It should be obvious at this point why this averaging procedure is not generally applicable to the direct estimation of the original image. Use of this procedure would require that the same image be degraded by a large number of uncorrelated degrading functions. One situation, however, that does meet this criteria is short time exposures of the same object taken through turbulent media and, indeed, some work has been done in this area (22,25,26).

Having discussed the method by which the N sample functions are obtained for the averaging, I now return to Eqn. (5.12) to discuss in detail the averaging mechanism and related issues. Separating Eqn. (5.12) into its real and imaginary parts, one obtains

$$\log |R_i| = \log |S_i| + \log |H|$$

$$(5.21)$$

$$\text{ang}(R_i) = \text{ang}(S_i) + \text{ang}(H)$$

Clearly, the problem can be viewed as two separate problems, one of estimating the magnitude of H ($|H|$) and the other of estimating the phase angle of H ($\angle H$). As it turns

55

out, the problems and procedures associated with the two quantities, $|H|$ and $\angle H$, differ to such an extent that a separate analysis of each will be presented.

5.4 Estimating $|H|$

Considering only the real part of Eqn. (5.21), and averaging over N samples, one obtains

$$\frac{1}{N} \sum \log |R_i| = \frac{1}{N} \sum \log |S_i| + \log |H|$$

$$\tag{5.22}$$

$$= B(u,v) + \log |H|$$

where $B(u,v)$ represents the bias in estimating the magnitude. The bias, however, is directly related to the power spectral density of the original image process. In fact, one can show (19) that under appropriate conditions,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \log |S_i| = \tfrac{1}{2} \log \Phi_{ss}(u,v) \tag{5.23}$$

where the limit is in the mean-square sense. Thus, the bias term can be removed by subtracting $\tfrac{1}{2} \log \Phi_{ss}(u,v)$, a quantity that is assumed to be known a priori. For the special case of white noise processes of spectral height N_o, the bias term need not be removed, since the constant, $\tfrac{1}{2} \log N_o$, merely appears as a multiplicative constant in the final result; i.e.

$$\tilde{h}(n,m) = (\tfrac{1}{2} \log N_o) \, h(n,m) \qquad\qquad (5.24)$$

Unfortunately, most pictures are distinctly non-white.
They are most often of a low pass nature. For this report,
I modelled $\bar{\Phi}_{ss}$ to be of the form

$$\Phi_{ss}(u,v) = K \exp(-a|u| - b|v|) \qquad\qquad (5.25)$$

in which case the bias term becomes

$$B(u,v) = \tfrac{1}{2}(\log K - a|u| - b|v|) \qquad\qquad (5.26)$$

The parameters a and b were varied until the best result
was obtained.

Two practical problems arise in estimating the magni-
tude. First, if any of the $S_i(u,v)$ are zero at a point
where $H(u,v)$ is non-zero, the value of $H(u,v)$ at that
point can never be found. And the same effect occurs if the
$|R|$ drops below the dynamic range of the computer due to the
finite register length. For the PDP-9 computer, assuming
15 noise-free bits, the dynamic range is about 90 dB. Thus,
at any point where $|R_i|$ is more than 90 dB down from its
maximum value, it is impossible to estimate $H(u,v)$. How-
ever, these limitations did not prevent good estimates from
being obtained. Indeed, in some cases, using N=4 gave sur-
prisingly good results. The details are discussed in
Chapter 6.

5.5 Estimating the Phase Angle of H

Returning now to the imaginary part of Eqn. (5.21), one is faced with a tougher, though more subtle problem. The problem arises due to the fact that the phase angle of a signal is multivalued. Furthermore, the arctangent function used to calculate the phase angle has a range of only $\pm \pi$, i.e., only the principle value (PV) of the phase angle is available. And the unfortunate fact is that

$$PV(\angle R_1) = PV(\angle S_1 + \angle H)$$
$$\neq PV(\angle S_1) + PV(\angle H)$$

(5.27)

That is, the $PV(\angle R_1)$ is not generally equal to the sum of the PV of the angles of S_1 and H. In other words, the problem is no longer additive. For example, let $\angle S_1 = \pi/4$, $\angle H = 3\pi/4$. Then $\angle S_1 + \angle H = \angle R = 2\pi$, but $PV(\angle R) = 0$.

Another way to view the problem is to realize that the sum $(\angle S_1 + \angle H)$ has been passed through the picewise linear, non-invertible operator shown in Figure 5.7.

The problem can be postponed for a time by noting that

$$PV(A + B) = PV(PV(A) + PV(B))$$

(5.28)

and, thus

$$PV \sum_{i=1}^{N} PV(\angle R_1) = PV \sum_{i=1}^{N} PV(\angle S_1 + \angle H)$$

$$= PV \sum_{i=1}^{N} (\angle S_1 + \angle H) \qquad (5.29)$$
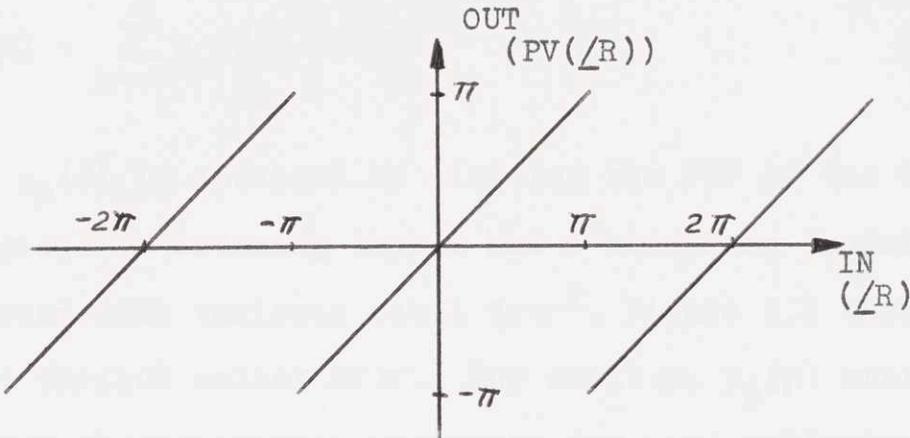
Figure 5.7

That is, when summing over N samples, the PV of the sum of $\angle R_1$ can be obtained by taking the PV of the sum of $PV(\angle R_1)$. However, when attempting to divide by N, one is again faced with

$$\frac{1}{N} PV \left[\sum_{i=1}^{N} (\angle S_1 + \angle H) \right] \neq PV \left[\frac{1}{N} \sum_{i=1}^{N} \angle S_1 + \angle H \right] \qquad (5.30)$$

i.e., the PV operator and the divide-by-N operator do not commute.

Another way in which the PV problem manifests itself is through the statistics of PV$\angle R$. Generally, when trying to estimate a constant (in this case $\angle H$) buried in additive noise ($\angle S_1$), best results are obtained if the noise has a "peaky" or leptokurtic probability density function (PDF). But consider the effect of the PV operator on the PDF of $PV \angle R \equiv \phi$. With θ denoting the value of the phase angle at

59

the input to the PV operator, we have

$$p_\phi(\phi) = \sum_{k=-\infty}^{\infty} p_\theta(\phi + 2\pi k) \qquad (5.31)$$

That is, $p_\phi(\phi)$ is obtained by aliasing the PDF of the true phase angle, θ. Assuming that θ has a zero-mean Gaussian distribution with variance equal to σ^2, Figure 5.8 shows $p_\phi(\phi)$ for various values of σ. For small σ, $p_\phi(\phi)$ exhibits the "peaky" characteristic necessary for good estimates. However, as σ approaches π and higher values, $p_\phi(\phi)$ becomes essentially flat, and is thus of no value in the estimation procedure. For example, consider what happens when a constant, a, is added to ϕ. For $\sigma \geq \pi$, there is essentially no change in the PDF of ϕ, and, hence, a cannot be accurately estimated. On the other hand, if $\sigma < \pi$, then the peak in the PDF of ϕ will shift to a point equal to PV(a). Unfortunately, experimental evidence such as the histograms of Figure 5.2 indicate that ϕ is essentially uniformly distributed, and thus, one cannot estimate $\angle H$ directly from PV$\angle R$.

The remainder of this chapter deals with a technique, called phase unwrapping, that handles the problems associated with the PV operator. Because the one-dimensional problem is conceptually more straightforward, I shall begin with it.
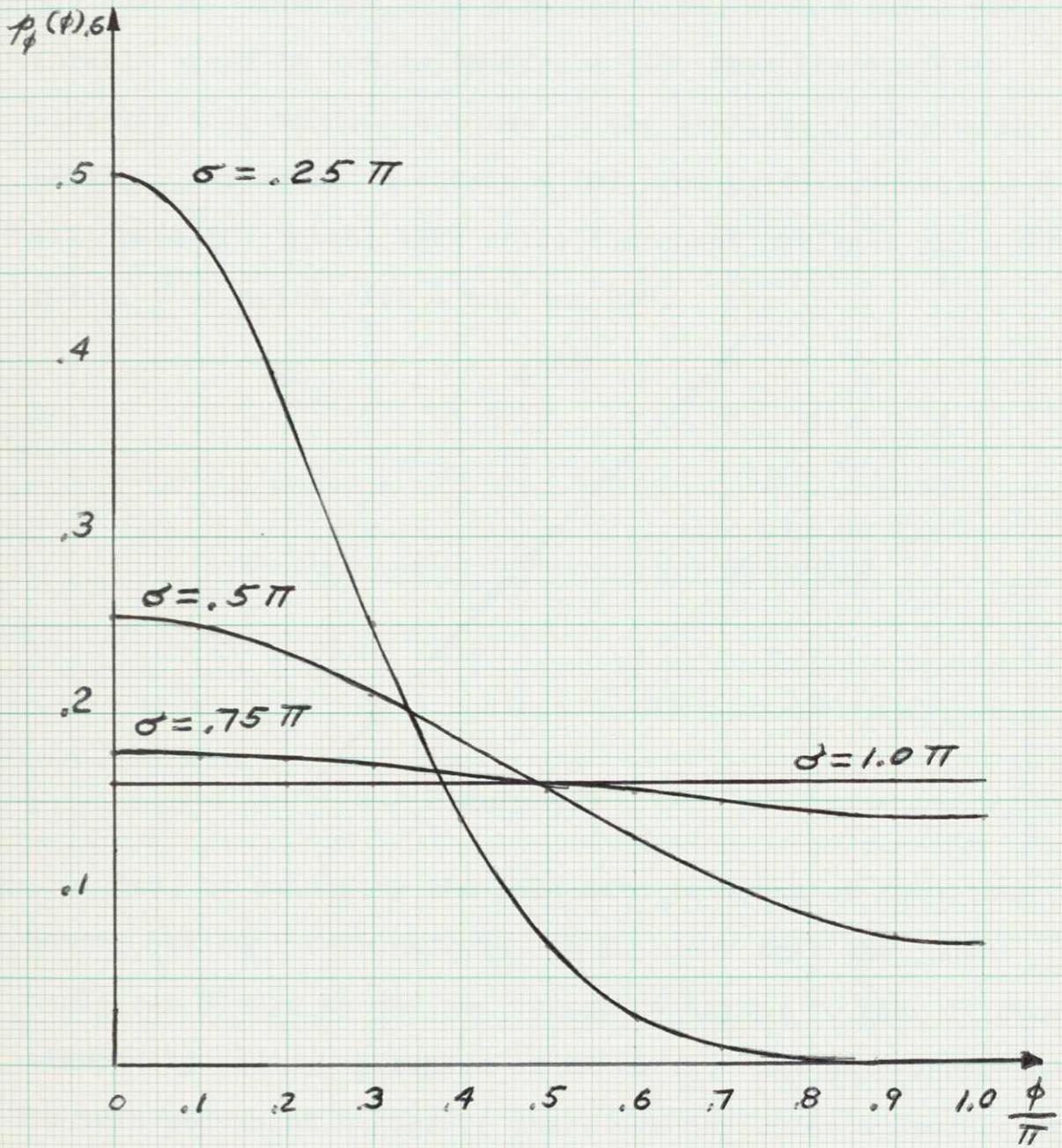
The basic problem with the PV operator is that the

FIGURE 5.8

measurable quantity, $PV\angle R_1$, is not equal to the sum of its components. If the additive property can be restored to the phase angle, then the problem is solved. Thus, if the non-invertible nature of the PV operator as shown in Figure 5.7 can be converted to a straight line, with a slope preferably equal to one as in Figure 5.9, then the additive property will hold. In terms of the statistics discussed earlier, one seeks to obtain $p_\Theta(\theta)$ from $p_\phi(\phi)$ so that one will have to contend with noise component having a peaky PDF, rather than noise with a uniform PDF.

Figure 5.9

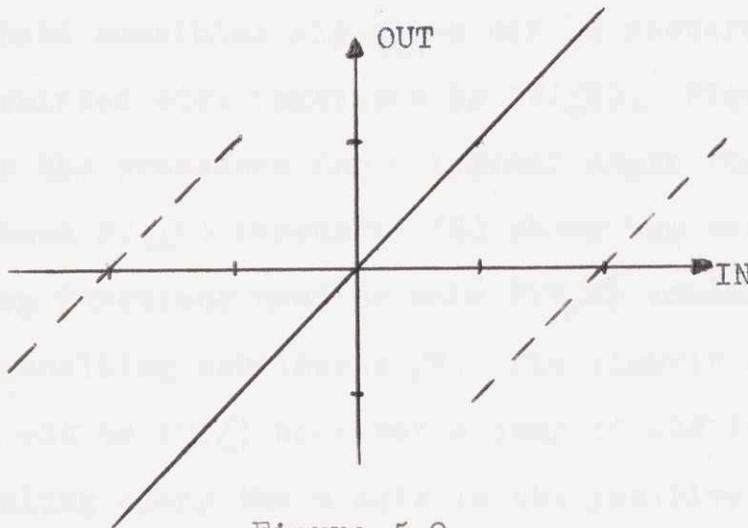Now, by examining the pole-zero patterns of functions having rational Z-transforms, it is clear that the phase angle can always be defined in such a way as to make it continuous by adding multiples of 2π to it. The only exception to this statement occurs if there are any poles or zeroes of odd multiplicity on the unit circle, in which case the phase angle would have jumps of $\pm\pi$ in it. If the function r(n)

62

is assumed not to have any odd order poles or zeroes on the unit circle, then the $\angle R$ must be continuous for all frequencies. This fact provides the condition necessary to find $\angle R$ from $PV(\angle R)$. All one must do is add the appropriate multiple of 2π to $PV(\angle R)$ at each frequency such that the result is a continuous function of frequency. This is exactly how Figure 5.9 was obtained. Actually, even odd order poles or zeroes on the unit circle are allowed if one only seeks to correct discontinuities with magnitude greater than π.

The phase smoothing algorithm can be restated in terms of adding shifted step functions to $PV(\angle R)$. Figure 5.10 illustrates the procedure for a typical angle function. Part (a) shows $PV(\angle R)$ versus u; (b) shows the set of shifted step functions used to make $PV(\angle R)$ continuous; (c) shows the resulting continuous $\angle R$. The algorithm is to add a constant $+2\pi$ to $PV(\angle)$ whenever a jump of -2π is detected while traveling along the u axis in the positive direction. If a jump of $+2\pi$ is found, then a constant -2π is added to $PV(\angle)$.

While the phase unwrapping procedure is conceptually straightforward, a serious practical problem arises due to the fact that, in practice, only samples of the $PV(\angle)$ are available, which, in turn, means that it would be quite possible to miss one of the $\pm$ 2π jumps in the $PV(\angle)$, which

63
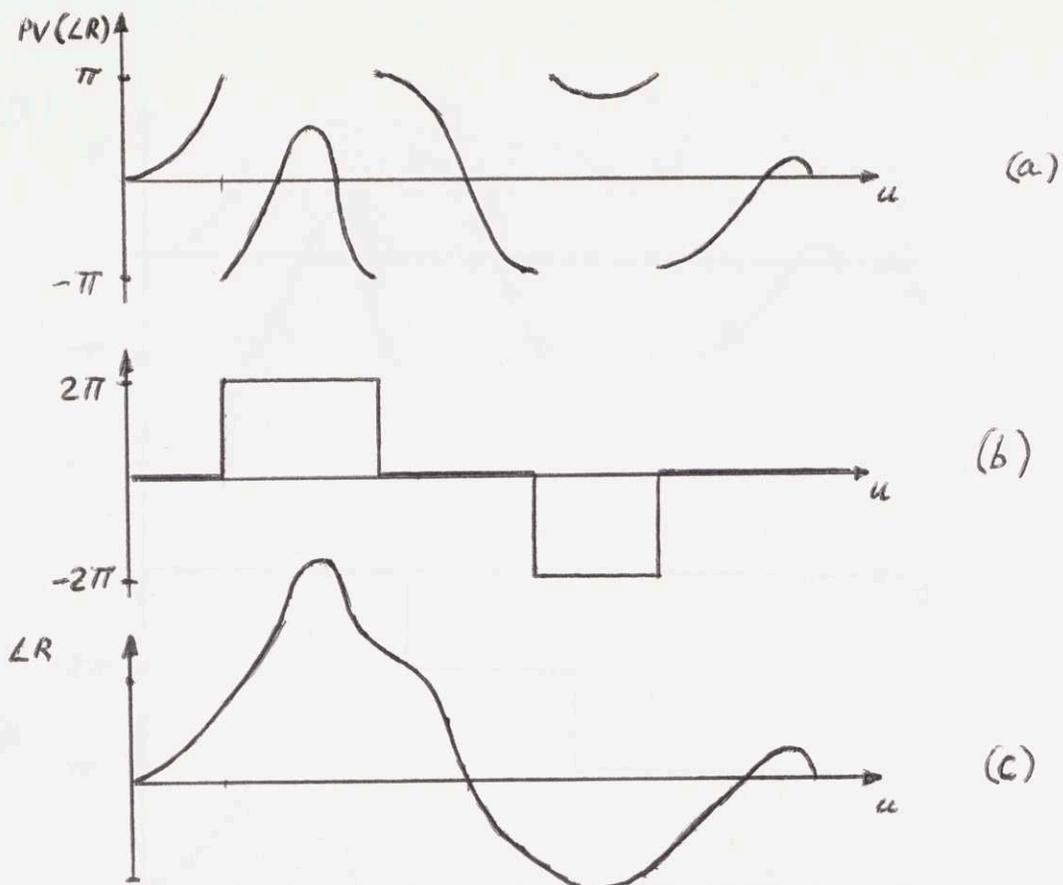
Figure 5.10

would lead to a completely erroneous result. For example,
consider Figure 5.11, which is a sampled reproduction of
Figure 5.10.

Notice that the algorithm failed to identify the first
and forth jumps in the PV($\angle$) with the result that the
"smoothed" phase angle is completely in error. Indeed, this
is an inherent problem of phase unwrapping, namely, that an
error made at any point along the u axis, whether by failing
to identify a jump (miss) or by claiming a jump when, in
fact, none was present (false alarm), tends to propagate,
undiminished, along the u axis. Thus, it is of utmost

PV(∠R)
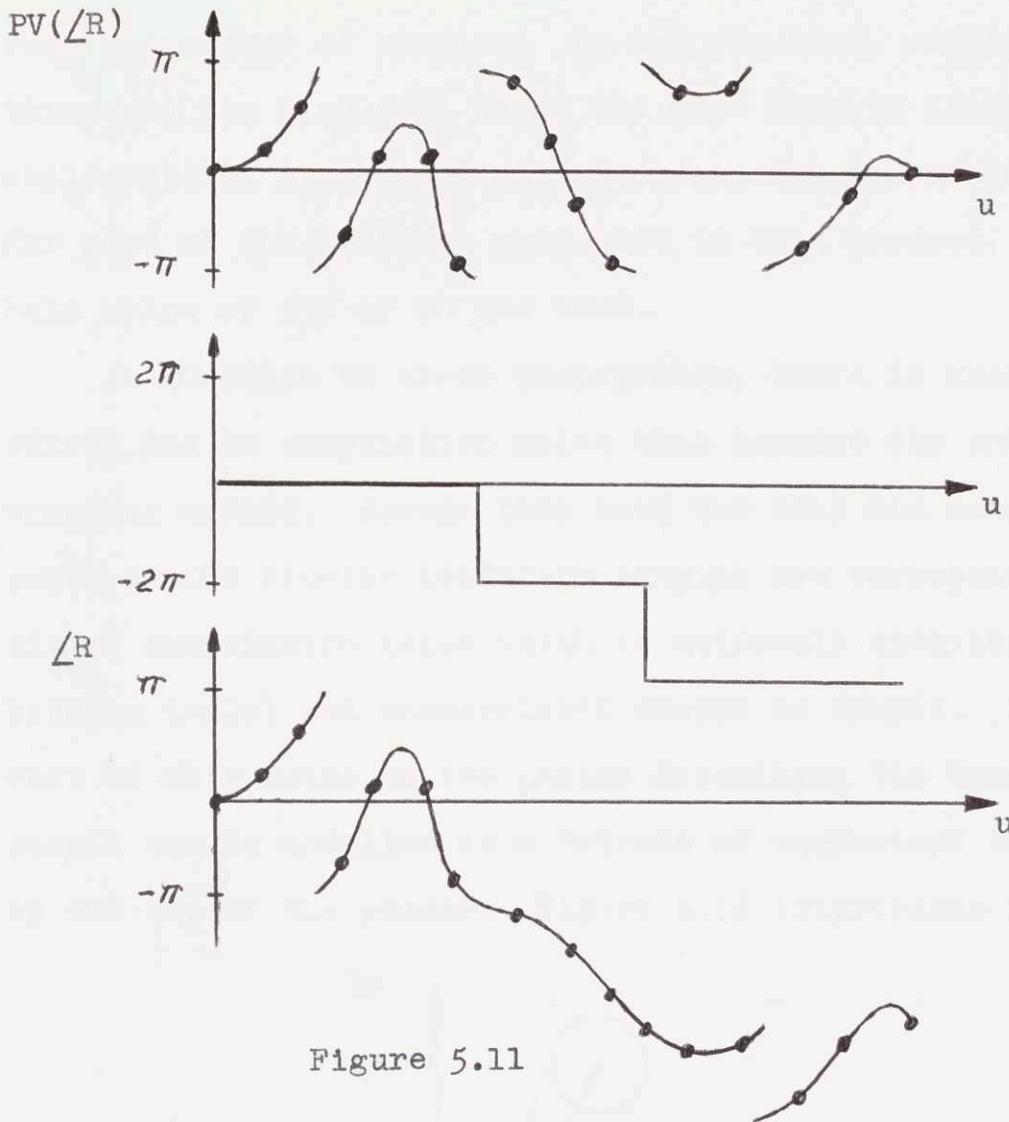
Figure 5.11

importance to correctly identify the jumps. One means by
which the error rate can be reduced is to increase the
number of samples. Examples run using picture data indicate
that 2048 samples along one line in the transform plane seem
to give accurate results. Another helpful technique is to
identify as ± 2π jumps, those places where adjacent phase
samples differ by more than a fixed fraction of 2π. This
involves a performance trade-off and also influences the

required number of samples. On the one hand, setting the threshold too high will cause the miss rate to increase, while setting it too low will increase the false alarm rate. For most of the examples presented in this project, a threshold value of 53% of 2π was used.

In addition to error propagation, there is another effect due to computation noise that hampers the phase unwrapping method. Assume that both the real and imaginary parts of the Fourier transform samples are corrupted by additive computation noise which is uniformly distributed between (-a,a) and uncorrelated sample to sample. The effect of this noise on the phasor describing the transform sample can be modelled as a "circle of confusion" centered on the tip of the phasor. Figure 5.12 illustrates several

Figure 5.12

66
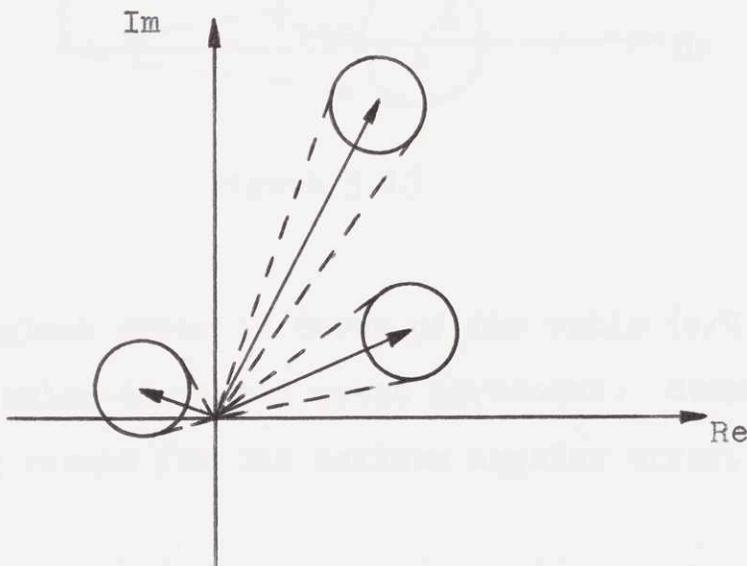
frequency samples and their confusion circles. Note that
the radius of the confusion circles is assumed to be the
same for all samples. Clearly, as the magnitude of the
phasor decreases, the range over which possible phase angle
errors can occur increases, as indicated by the dotted lines
emanating from the origin and tangent to the confusion
circles. To demonstrate this effect analytically, consider
Figure 5.13 where a phasor of length R has been arbitrarily
placed along the Re axis. An expression is derived for the
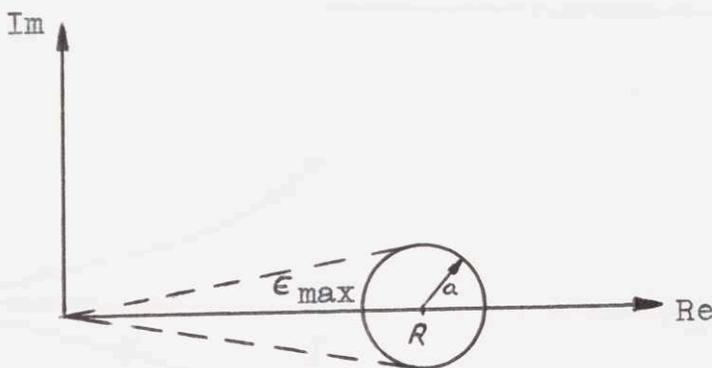
Figure 5.13

maximum angluar error in terms of the ratio (a/R), which
acts as a noise-to-signal ratio parameter. Simple geometry
yields the result for the maximum angular error, ϵ_{max},
which is

$$\epsilon_{max} = \begin{cases} \pm \tan^{-1} \dfrac{(a/R)}{\sqrt{1-(a/R)^2}} & |a/R| < 1 \\[4mm] \pm \pi & |a/R| \geq 1 \end{cases} \qquad (5.32)$$

A plot of ϵ_{max} vs a/R is shown in Figure 5.14. Notice that the solution contains a discontinuity at the point a = R. Once the (a/R) ratio reaches 1, the maximum angular error remains at $\pm \pi$. This result implies that for transform
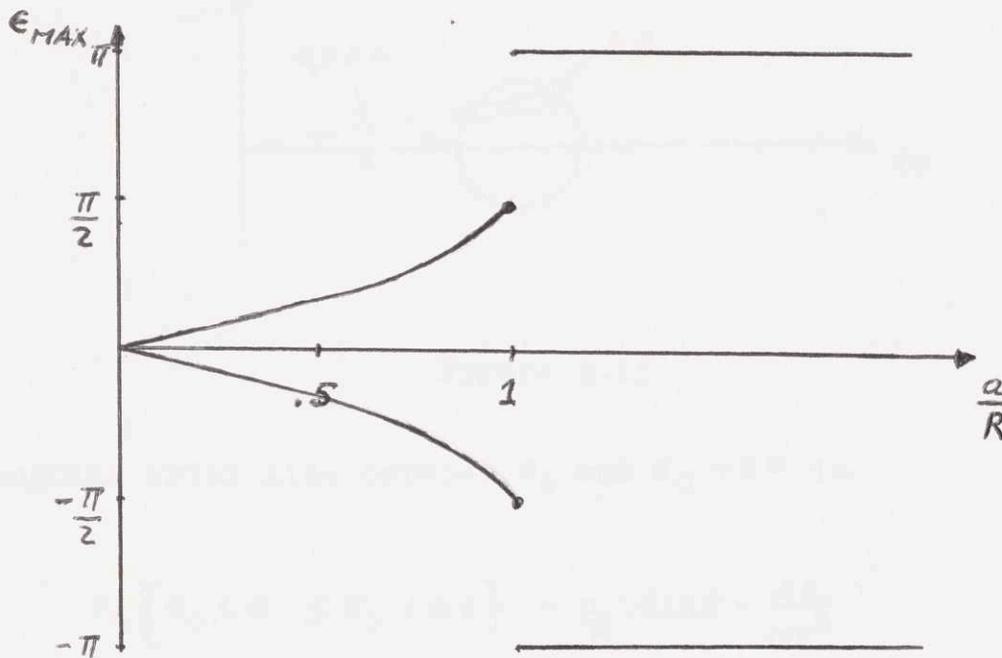
Figure 5.14

samples having small magnitude, the angular error may be quite large and, furthermore, due to the phase unwrapping, the error may propagate through succeeding phase samples.

This result can be somewhat misleading in that it doesn't provide any insight into the distribution of angular error and, indeed, in this light the problem doesn't appear quite as severe. With reference to Figure 5.15, the question is what is the distribution of angular error between the limits of ϵ_{max}. Assuming that each point in the confusion circle is equally likely, then the probability that the

Figure 5.15

angular error lies between ϵ_0 and $\epsilon_0 + \Delta\epsilon$ is

$$P_r \left\{ \epsilon_0 < \epsilon \leq \epsilon_0 + \Delta\epsilon \right\} = p_\epsilon (\epsilon) \Delta\epsilon = \frac{\Delta A}{\pi r^2} \tag{5.33}$$

After a few manipulations, one obtains

$$p_\epsilon (\epsilon) = \frac{2}{\pi} \frac{1}{1+m^2} \frac{R}{a} \sqrt{1+m^2 - \left(\frac{R}{a} m\right)^2} \tag{5.34}$$

where $\quad \tan \epsilon = m \quad$ and $\quad \epsilon_{max} = \tan^{-1} \dfrac{a/R}{\sqrt{1-(a/R)^2}}$

Eqn. (5.34) is valid only for $\left|\dfrac{a}{R}\right| < 1$. For the case of $\left|\dfrac{a}{R}\right| > 1$, $p_\epsilon(\epsilon)$ becomes
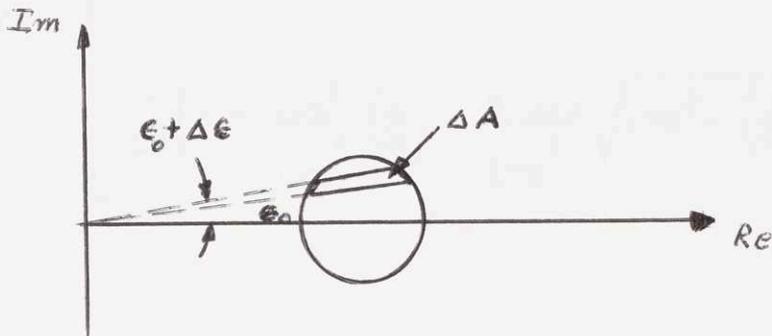
$$
p_\epsilon(\epsilon) =
\begin{cases}
\dfrac{1}{2\pi}\left[1 + \dfrac{1-m^2}{1+m^2}\left(\dfrac{R}{a}\right)^2 + \dfrac{2\,R/a}{1+m^2}\sqrt{1+a^2-\left(\dfrac{R}{a}m\right)^2}\,\right] & |\epsilon| < \dfrac{\pi}{2} \\[4ex]
\dfrac{1}{2\pi}\left[1 + \dfrac{1-m^2}{1+m^2}\left(\dfrac{R}{a}\right)^2 - \dfrac{2\,R/a}{1+m^2}\sqrt{1+a^2-\left(\dfrac{R}{a}m\right)^2}\,\right] & \dfrac{\pi}{2} < |\epsilon| < \pi
\end{cases}
$$

(5.35)

Using the fact that $\tan \epsilon = m$, Eqns. (5.34) and (5.35) can be rewritten more compactly, directly in terms of ϵ. Eqn. (5.34) becomes

$$
p_\epsilon(\epsilon) = \frac{2}{\pi}\,\frac{R}{a}\,\cos^2\epsilon\,\sqrt{\sec^2\epsilon - (R/a)^2\tan^2\epsilon} \qquad \left|\frac{a}{R}\right| < 1
$$

(5.36)

where, again, $\epsilon_{max} = \tan^{-1} \dfrac{a/R}{\sqrt{1-(a/R)^2}}$.

Similarly, for $\left|\dfrac{a}{R}\right| > 1$, Eqn. (5.35) becomes

70

$$p(\ ) = \begin{cases} \dfrac{1}{2\pi}\left[1+\cos2\ \left(\dfrac{R}{a}\right)^2 + 2R/a\ \cos^2\epsilon\sqrt{\sec^2\epsilon-(R/a\ \tan\epsilon)^2}\ \right] \\ \hspace{8cm} |\epsilon| < \dfrac{\pi}{2} \\[1em] \dfrac{1}{2\pi}\left[1+\cos2\epsilon\left(\dfrac{R}{a}\right)^2 - 2R/a\ \cos^2\epsilon\sqrt{\sec^2\epsilon-(R/a\ \tan\epsilon)^2}\ \right] \\ \hspace{8cm} \dfrac{\pi}{2} \le |\epsilon| \le \pi \end{cases}$$

$$(5.37)$$

In Figure 5.16, a family of curves is plotted for several values of the parameter (R/a). In general, the curves are highly peaked about $\epsilon = 0$ for (R/a) large, and gradually become broader as the value of (R/a) is decreased. It is clear that for all values of $(R/a) \ne 0$, the maximum angular error is incurred relatively infrequently. However, significant values of angular error are likely to occur whenever $(R/a) \le 1$. Thus, when performing phase unwrapping, special precautions should be taken for those frequency samples having a small magnitude (on the order of the computation noise).

Now that a means is available to calculate the unwrapped phase angle, its PDF will be investigated to determine whether or not it conforms to the Gaussian hypothesis used to generate the curves in Figure 5.8. To this end, 1024 unwrapped phase angle samples were selected at random from a total of over 65,000 samples. The phase unwrapping was based
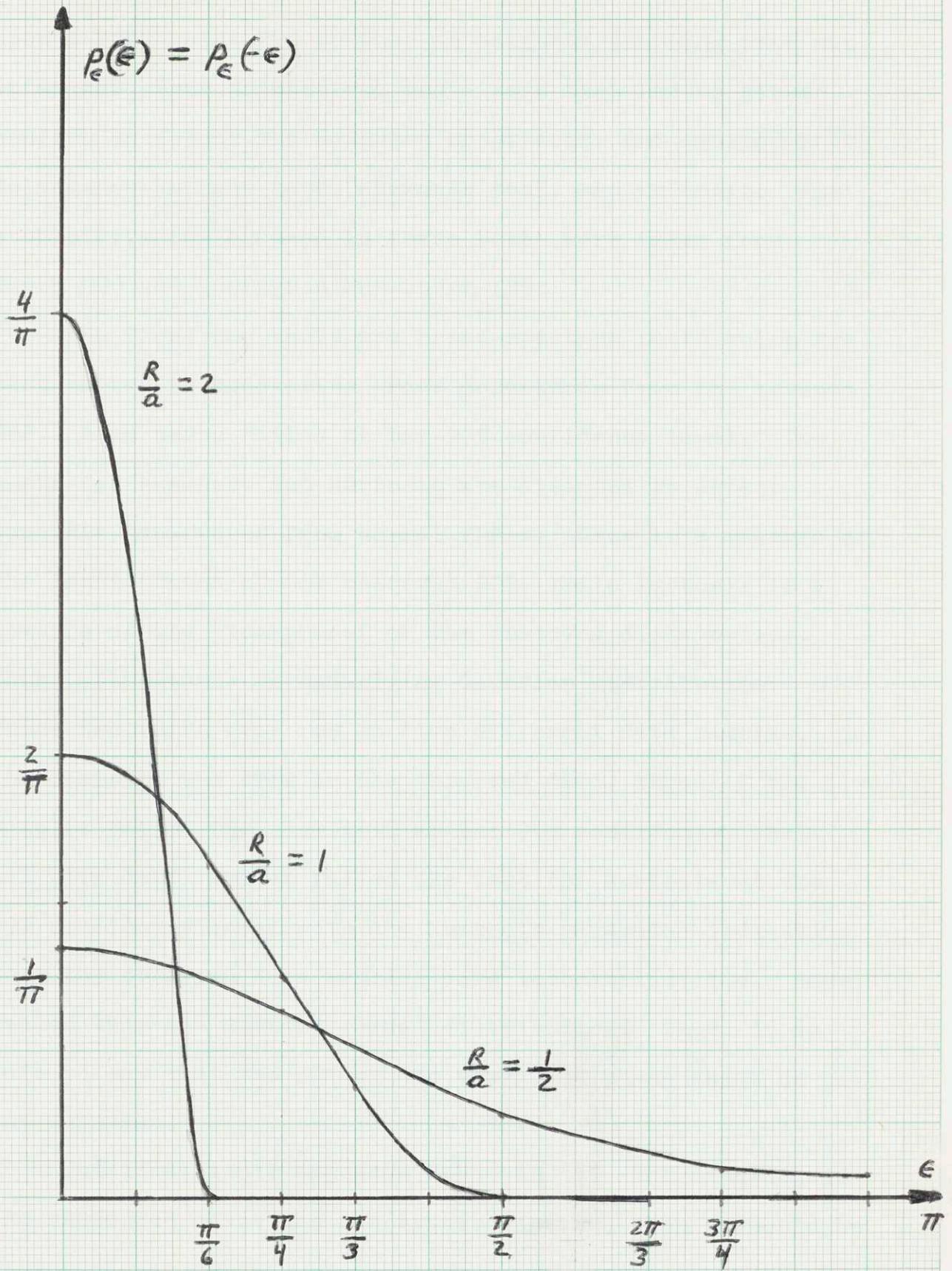
Figure 5.16

72

upon a 2048 point FFT so that an adequate sampling rate was attained. Four statistical tests were then run to either confirm or reject the Gaussian hypothesis; chi-squared test, Kolmogorov-Smirnov test, skewness test, and kurtosis. The probability tables used in these tests were taken from Knuth (21) and Remington and Shork (34).

The first two tests are broad-spectrum tests which are best suited to uncovering gross departures from the assumed PDF. Both tests failed at the 99% confidence level. The tests for skewness and kurtosis are disigned to detect specific departures from normal PDF's. Skew is a measure of the lopsidedness of the observed PDF, while kurtosis is a measure of the "peakyness" of the observed distribution. Both tests assume an underlying Gaussian distribution. These two tests were run on five different data samples. On the basis of the skew in the observed data, three of the five samples failed the skew test at the 99% confidence level. All five data samples failed the test for kurtosis at the 99% confidence level with the kurtosis parameter indicating that the observed data samples were leptokurtic; i.e., more peaked than the assumed normal distribution. Thus, it is safe to say that the PDF of the unwrapped phase angle is not Gaussian.

But if it isn't Gaussian, what is it? For example, if the unwrapped phase was Cauchy distributed, then it would be

pointless to try and estimate the phase since, in general, none of the moments of a Cauchy distribution are defined. Hence, the variance of the phase angle estimate would be infinite. However, the fact that all five of the data samples were leptokurtic raises doubt about the Cauchy assumption since the Cauchy density function is much broader than the Gaussian and would thus tend to be platykurtic. However, to get a more quantitative measure of whether or not the Cauchy assumption was justified, the nonparametric Kolmogorov-Smirnov test was performed. The Cauchy hypothesis failed the test for all five samples at the 99% level.

Having rejected both the Gaussian and the Cauchy hypothesis, how does this affect the family of curves shown in Figure 5.8 which were based upon a Guassian distribution? The answer is, not in any significant way. The reason is that although all of the samples had leptokurtic distributions, the average of the sample standard deviations was approximately equal to 2π. This value of the standard deviation is large enough to make $p_\phi(\phi)$ essentially uniform even though $p_\theta(\theta)$, based upon the sample histograms, is somewhat more peaky than a normal distribution.

Finally, before proceeding to the two-dimensional case, some examples of phase unwrapping in one-dimension will be presented. Figure 5.17(a) shows a plot of typical log-magnitude (top), PV($\angle R$) (center), and smoothed phase angle (bottom) functions. These functions were computed at 2048

74

samples, only half of which are shown due to symmetry. The
origin of the graphs is at the left of the base line. The
most striking feature of PV($\angle$R) and $\angle$R is the strong linear
component in the phase functions. Indeed, both functions
appear to consist chiefly of straight lines.

Strong linear phase components are very common in phase
angle functions. In fact, one can show that the slope of
the linear phase component is equal to the number of zeroes
of the time function r(n) located outside of the unit circle
(38). From the standpoint of estimating the degrading
function, linear phase has no utility since it only contri-
butes a shift in the time domain estimate, which is of no
importance in image restoration. It does, however, pose a
practical problem in two ways. First, because the linear
phase component usually has such a steep slope, a high
sampling rate is necessary to accurately detect jumps.
Secondly, the information-bearing portion of the phase
function, the non-linear component, is all but buried be-
neath the large phase values attained by the linear component
which, in turn, tends to increase computation error due to
the limited dynamic range of the computer. For these reasons
it is desirable, although not necessary, to remove the linear
phase component. Clearly, this can be done by shifting the
time function r(n) an appropriate amount. The direction and
amount of shifting needed to eliminate the linear phase com-
ponent can be computed in the following manner. First,

75

smooth PV($\angle$R) as well as possible. Then compute $n_0 = \angle R(\pi)/\pi$. If n_0 is non-zero, shift the function r(n) by $-n_0$ units, and repeat from the first step. When n_0 equals zero, then the linear phase has been completely removed. Figure 5.17(b) shows the PV($\angle$R) (top) and $\angle$R (bottom) after having removed the linear phase components. The resulting phase structure is much more interesting in that it shows details that were completely obscured in part (a). Notice that smoothed phase function changes rapidly at those frequencies where the log-magnitude function becomes very small.

As an aside, the fact that the number of zeroes outside the unit circle is equal to $\angle R(\pi)/\pi$ can be used to determine the stability of discrete linear systems. The stability of discrete systems is usually determined by one of two methods. One may use the bilinear transformation to transform the z-plane to the s-plane where a Routh-Hurwitz test may be applied. However, this procedure is algebraically tedious, particularly for higher order systems. The other stability test requires the evaluation of Schur-Cohn determinants which may be computed directly from the coefficients of the system characteristic equation (20). For an n-th order system, the Schur-Cohn determinants consist of 2n determinants of order 1, 2, ... n-1, n; i.e., two first-order determinants, two second-order determinants, and so on up to two n-th order determinants. By noting the number of sign changes in this sequence of determinants,

76

the number of zeroes outside the unit circle can also be found.

Consider now the amount of computation required to perform this test. It can be shown (18) that the number of multiply and divide operations necessary to evaluate an n-th order determinant, by the Gauss-elimination technique, is equal to $\frac{n}{3}(n^2-1)$. Thus, to evaluate the sequence of 2n determinants requires

$$M = \frac{2}{3} \sum_{i=1}^{n} (i^3 - i^2)$$

$$= \frac{n}{6} (n+1)(n-1)(n+\frac{2}{3})$$

(5.38)

manipulations. By comparison, the number of operations necessary to determine both stability and the number of zeroes outside the unit circle, using phase unwrapping, is equal to $\frac{1}{2}(2048 \cdot 2 \cdot \log_2(2048) + 8 \cdot 2048) = 30720$, where 2048 is the number of phase samples available, 8 is the number of multiplications in the arctangent subroutine, and the factor of $\frac{1}{2}$ is included because of the odd-symmetry of the phase function. Comparing this last number with the expression from M in Eqn. (5.38), one finds that for $n \geq 21$, it is more efficient to use phase unwrapping to determine both stability and the number of zeroes outside the unit circle.

Of course, the unwrapped phase is still subject to possible errors at those frequencies where the magnitude is

77

small, which conceivably could lead to the wrong conclusion
being drawn about system stability. An interesting con-
jecture is whether or not the Schur-Cohn matrices become
ill-conditioned for such systems. If that is the case,
then this method might also yield erroneous results. How-
ever, such considerations lead too far afield from the
subject of this study.

Finally, Figure 5.17(c) shows an attempt at phase un-
wrapping using only 256 samples or one-eighth the number
used in parts (a) and (b). The linear phase component has
been removed. However, the resulting $\angle R$ looks quite unlike
that of part (b) because of errors made in the unwrapping
procedure due to an insufficient number of samples.

This last example should serve to indicate one of the
major obstacles to phase unwrapping in two-dimensions,
namely the inability to obtain an adequate number of samples.
The disk storage unit on the BIP, PDP-9 computer has a one
megaword storage capacity. As a practical matter, though,
the total number of phase samples is limited to about one
quarter million. Thus, one could store a phase array of
512 x 512, 256 x 1024, etc. However, from the experience
gained with one-dimensional phase unwrapping, it is apparent
that regardless of the array size, the sampling rate along
at least one side of the phase array will be inadequate. In
practice, what was one was to smooth the phase as well as

Figure 5.17(a)
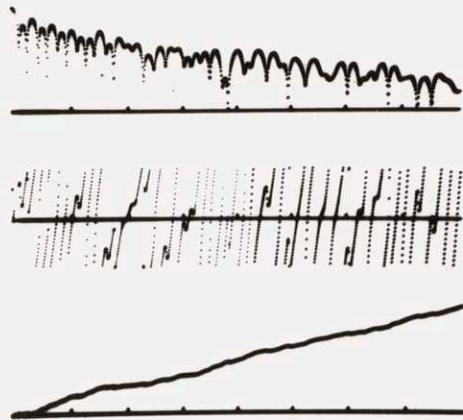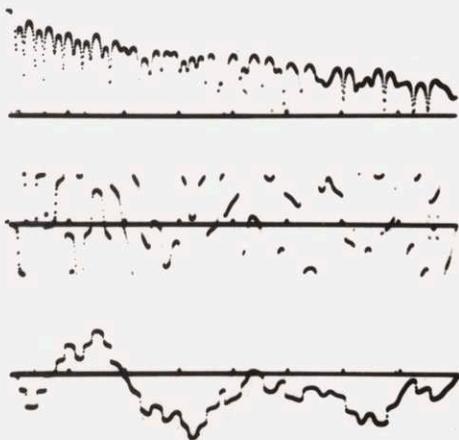Log-mag. (top), PV∠R (center), ∠R (bottom)
with linear phase, 2048 samples

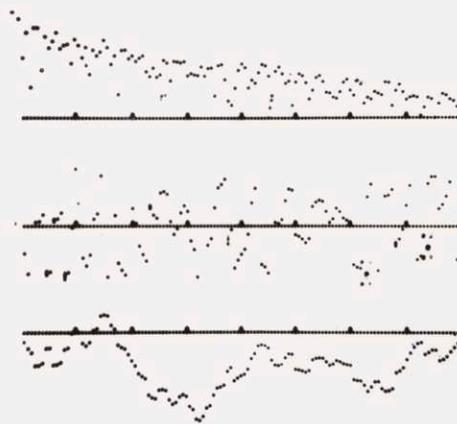Fig. 5.17(b), PV(∠R) (top),
∠R (bottom), with linear
phase removed, 2048 samples

Fig. 5.17(c), PV(∠R) (top),
∠R (bottom), with linear
phase removed, 256 samples

possible along one direction while accepting the smoothing errors that occurred along the other direction.

With the limited number of samples available, it is imperative to remove the linear phase component from the two-dimensional phase function. However, on this issue of linear phase, there is a significant departure from the one-dimensional theory in that the linear phase component cannot be completely removed merely by shifting r(n,m) by an appropriate amount. The reason is that each row and column of the phase array has a different linear component. To see this, expand the phase function $\Theta(u,v)$ in a two-dimensional power series, retaining only those terms which satisfy the odd-symmetry requirement on the phase angle of real functions. Thus,

$$
\begin{aligned}
\Theta(u,v) = \quad & u(a_{01} + a_{11}v^2 + a_{21}v^4 + \ldots) \\
+ \quad & u^3(a_{03} + a_{23}v^2 + \ldots) \\
& \quad \vdots \\
+ \quad & v(b_{01} + b_{11}u^2 + \ldots) \\
+ \quad & v^3(b_{03} + b_{23}u^2 + \ldots) \\
& \quad \vdots
\end{aligned}
$$

(5.39)

80

Note that the coefficient of u in the first term varies with v. Similarly, the coefficient of v in the third term varies with u. Thus, each row or column in the u-v plane has a different linear phase component. The only term that can be eliminated by shifting the picture is the ua_{01} and vb_{01} terms. These constitute a planar phase component passing through the origin. The amount of shift is determined in the following way.

First, calculate the projections of $r(n,m)$ along the n and m axes; i.e.

$$r_1(n) = \sum_m r(n,m)$$

$$r_2(m) = \sum_n r(n,m)$$

(5.40)

and compute the 2048 point transform of these functions, which yields

$$R_1(u) = R(u,0)$$

$$R_2(v) = R(0,v)$$

(5.41)

That is, the transforms of the projections correspond to the value of $R(u,v)$ along the u and v axes. R_1 and R_2 are now treated as separate one-dimensional functions and smoothed in accordance with the procedure described earlier. From the smoothed phase functions the amount of shift necessary to

81

eliminate the linear phase along the u and v axes can be
obtained. However, the linear phase components along other
rows and columns in the phase array will, in general, re-
main non-zero. Indeed, there is nothing to say that these
components have not become larger after shifting r(n,m).
Figure 5.18(a) shows a typical two-dimensional transform.
Again, the top half represents the log-magnitude while the
bottom half illustrates the phase function. The orienta-
tion is with the v axis running horizontally and the u axis
running vertically down the center with the origin in the
center of the top line of each function. Part (a) shows the
PV($\angle$R), while part (b) illustrates the smoothed phase func-
tion calculated without first removing the u-v axis linear
phase terms. The phase samples were computed on a 256 x
1024 array. Part (c) shows PV($\angle$R) obtained after first
shifting r(n,m) by the appropriate amount. Note that it
appears to be considerably smoother than the result obtained
without prior shifting. This is analogous to the one-
dimensional result shown in Figure 5.17(a) and (b).
Finally, part (d) illustrates the smoothed phase function
derived from part (c). The fact that it is totally differ-
ent from the result in part (b) indicates how many unwrapping
errors were made in the latter case.

An important point to remember about removing linear
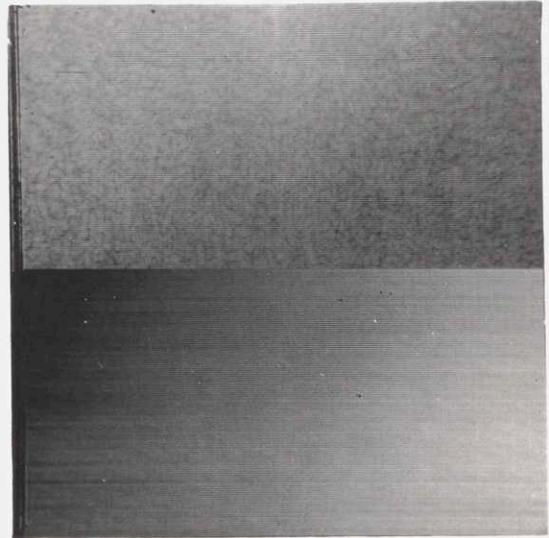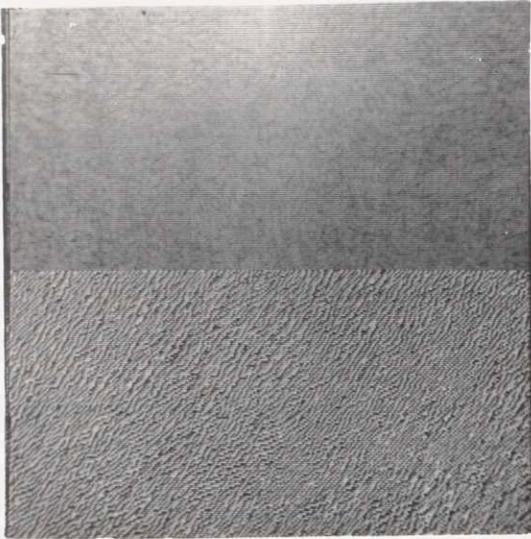phase components is that the proper shift distance is not

Fig. 5.18(a), Log-mag. (top),
PV($\angle$R) (bottom), 256x1024,
with linear phase

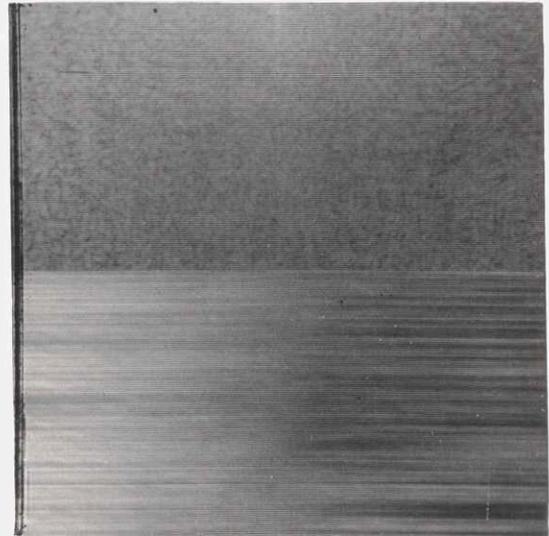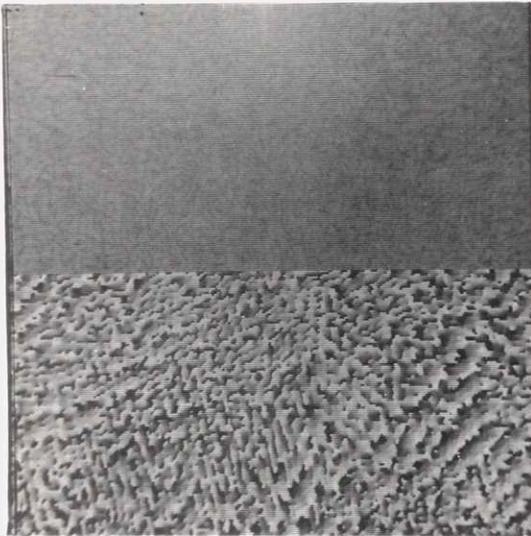Fig. 5.18(b), Log-mag (top),
$\angle$R (bottom), with linear phase

Fig. 5.18(c), Log-mag. (top),
PV($\angle$R) (bottom), 256x1024,
with linear phase removed
along u-v axes

Fig. 5.18(d), Log-mag. (top),
$\angle$R (bottom), with linear phase
removed along u-v axes

related to the center of gravity of the signal, but rather
to the number of zeroes located outside the unit circle.
For example, consider the exponential signal, $r(n) = a^n$,
where $a < 1$. This function has no zeroes outside the unit
circle and, indeed, a plot of the phase function shows that
$\angle R = 0$ when $u = \pi$, which conforms to 0 shift according to
the algorithm given above. However, $r(n)$ has a non-zero
center of gravity located at $a/(1-a)$.

In summary, homomorphic filtering allows one to esti-
mate $h(n,m)$ as though it were corrupted by additive noise.
The estimation procedure can be analyzed in two parts,
estimating the log-magnitude and the phase angle. The log-
magnitude function is a fairly well behaved function which,
given that the bias is known, yields a consistent estimate
of $\log|H(u,v)|$. The phase angle requires special treatment
to retain the desirable property of additivity. The tech-
nique is to make the phase angle continuous which, though
conceptually straightforward, is a procedure beset by
practical problems. The most obvious problem is the errors
generated by having an insufficient number of phase samples.
This is particularly true with two-dimensional signals where
the desired number of samples exceeds the bulk storage
capacity of the computer, although first shifting $r(n,m)$ by
an appropriate amount to remove the linear phase makes the
job considerably easier. Errors are most likely to be made

at those frequencies where the magnitude of R(u,v) becomes
so small that it is comparable to the computation noise.
Also, the phase unwrapping algorithm is rather intolerant
of errors since an error made at any point in frequency
space tends to propagate undiminished. The next chapter
illustrates the results of this procedure when applied to
actual image enhancement problems.

6. EXAMPLES OF HOMOMORPHIC ESTIMATION

In this chapter we shall use the homomorphic estimation procedure developed in Chapter 5 to first estimate the impulse response of the image degrading system, and then use this result to construct an inverse filter to obtain an improved image. In addition, some of the practical aspects of the problem will be discussed with examples presented to illustrate the material. This chapter is further intended to provide sufficient detail about the computational techniques to aid future researchers in the field.

Most of the examples in this chapter deal with images that have been filtered in one direction only. There are several reasons for this. First, the phase unwrapping technique is not hampered by lack of computer storage space. Having an adequate number of phase samples allows one to isolate and examine the effects of phase errors, error propagation, etc. Of course, computing time is also reduced. Secondly, using a one-dimensional blurring function allows one greater freedom in partitioning the received image and allows greater latitude in the chouce of N, the number of image subsections. Finally, one-dimensional signals can be displayed on standard x-y co-ordinates, whereas two-dimensional signals must either be plotted in a three-dimensional perspective format or have the dependent variable displayed as brightness variation on a two-dimensional CRT display.

Both of these methods have their drawbacks. The prespective
drawing becomes confusing when highly detailed signals must
be displayed. On the other hand, the variable brightness,
CRT display is deceptive because the information must
usually be greatly compressed to fit the dynamic range of
the display. This latter effect may be seen in Fig. 5.18(b)
and 5.18(c) where the phase angle appears to be almost linear.
In fact, there is substantial variation in the signal about
a linear mean value. This fact, however, is only apparent
in x-y plots of each line of the phase signal.

The first question to be considered in implementing the
homomorphic estimation procedure is exactly how does one
partition the received image. For one-dimensional proces-
sing, the 512 x 512 New Crowd Scene, shown in Fig. 5.1, was
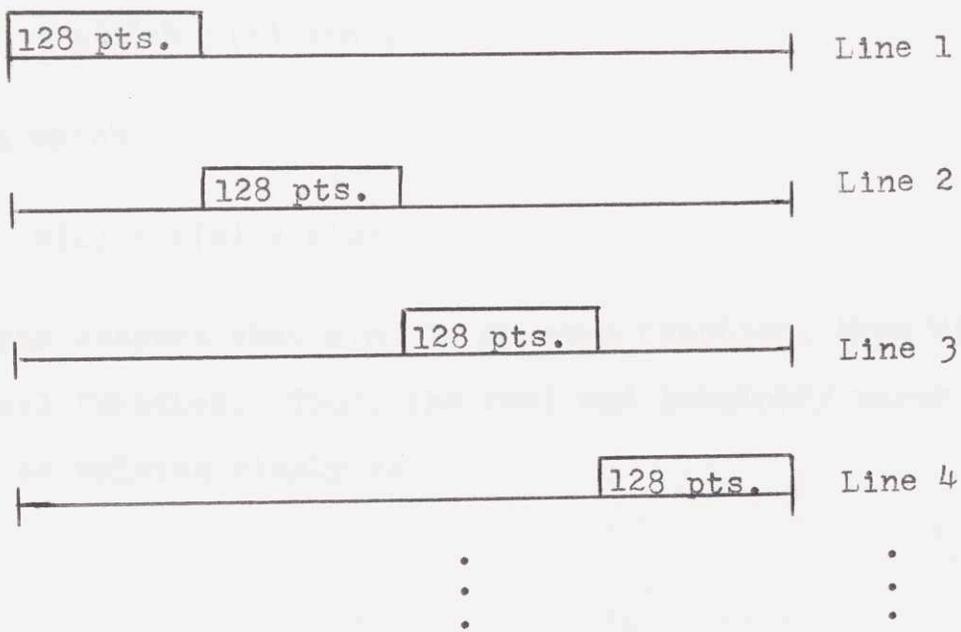partitioned as shown in Fig. 6.1.

Figure 6.1

87

This scheme was chosen to reduce the correlation between subsections. It also allows one to obtain up to 512 subsections from the blurred image.

The partitioning is equivalent to multiplying each line of the received data by a 128 point data window. Although the effects of data windows on the frequency response of a signal have been treated extensively, little is known about how data windows affect the magnitude and phase angle of a signal. As these two functions are basic to this study, I shall briefly review windowing theory with particular regard to its effect on magnitude and phase functions. The question is, given a signal $h(n)$ multiplied by a window function, $w(n)$, how are the magnitude and phase angle of the transform of the product related to the magnitude and phase angle of $h(n)$? Thus,

$$g(n) = h(n) \ w(n), \tag{6.1}$$

from which

$$G(u) = H(u) * W(u) \tag{6.2}$$

If one assumes that $w(n)$ is an even function, then $W(u)$ is a real function. Thus, the real and imaginary parts of $G(u)$ can be written simply as

88

$$G_r(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_r(\alpha) \, W(u-\alpha) \, d\alpha$$

(6.3)

$$G_i(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} H_i(\alpha) \, W(u-\alpha) \, d\alpha$$

Now,

$$|G|^2 = G_r^2 + G_i^2 = \frac{1}{4\pi^2}\left[\left(\int_{-\pi}^{\pi} H_r(\alpha) \, W(u-\alpha) \, d\alpha\right)^2\right.$$

(6.4)

$$\left.+ \left(\int_{-\pi}^{\pi} H_i(\alpha) \, W(u-\alpha) \, d\alpha\right)^2\right]$$

$$= \frac{1}{4\pi^2} \iint_{-\pi}^{\pi} \left[H_r(\alpha)H_r(\beta)+H_i(\alpha)H_i(\beta)\right] \, W(u-\alpha)W(u-\beta) \, d\alpha \, d\beta$$

and

$$\angle G = \tan^{-1}\left(\frac{\int_{-\pi}^{\pi} H_i(\alpha) \, W(u-\alpha) \, d\alpha}{\int_{-\pi}^{\pi} H_r(\alpha) \, W(u-\alpha) \, d\alpha}\right)$$

(6.5)

The important point to note about Eqns. (6.4) and (6.5) is that

$$|G|^2 \neq |H|^2 * |W|^2$$

(6.6)

$$\angle G \neq \angle H * \angle W$$

Thus the magnitude and phase angle of G cannot be obtained by convolving the magnitude and phase angle of H and W.

89

Equations (6.4) and (6.5) are not very useful in providing an intuitive idea of how a given window function will affect the magnitude and phase angle functions of the original data. About the only general statement that can be made is that these functions will in some sense be smoother.

Another point of view on the windowing issue is obtained by considering what effect windowing has on the zero locations of H(z). Multiplying h(n) by w(n) has the effect of modifying the coefficients in the polynomial representation of H(z). Altering the coefficients will, of course, change the locations of the zeroes of H(z). Now, since the zero locations are essentially what specify the magnitude and phase functions, it is clear that if one could predict how the zero locations shifted after windowing, one would be able to predict the new magnitude and phase functions.

Although a general solution to this problem has eluded mathematicians, Marden (24) has presented some interesting properties of this type of problem. In addition, this approach to the windowing problem breaks down in two dimensions because the concept of poles and zeroes does not extend to two-dimensional Z-transforms.

So, lacking any useful analytical approach, a set of experiments was run to observe the effects of various windows. The general window chosen was a variant of the Hanning window, with an added parameter.

$$w(n) = \begin{cases} 1 & 0 \le |n| \le n_0 \\ \frac{1}{2}\left[1 + \cos\dfrac{\pi(n-n_0)}{(N-n_0)}\right] & n_0 \le |n| \le N \\ 0 & N < |n| \end{cases} \qquad (6.7)$$

This window function is plotted in Fig. 6.2 for several values of n_0. Note that when $n_0 = N$ the $w(n)$ becomes a rectangular window of width $2N$ while for $n_0 = 0$, $w(n)$ becomes the standard Hanning window. Thus the primary purpose of including n_0 was to provide an adjustment for the roll-off rate at the edges of the data window.
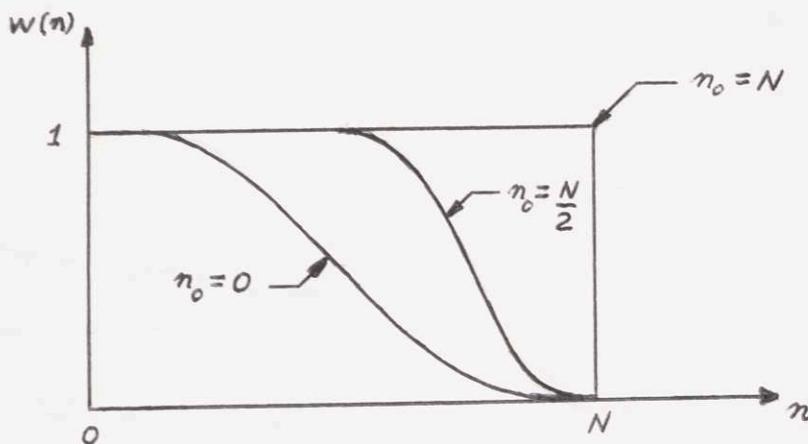

Figure 6.2

With N fixed at 64, $w(n)$ was applied to three data sections and n_0 varied to observe the effect on the magnitude and unwrapped phase functions. Figure 6.3 illustrates the log-magnitude, while Figure 6.4 shows the phase functions
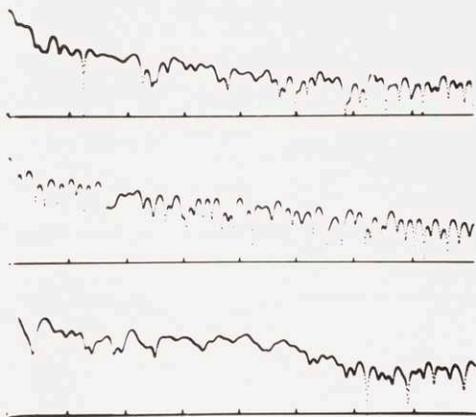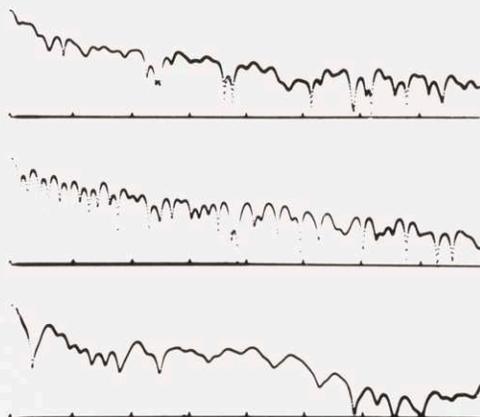
Figure 6.3(a), $n_0 = N$
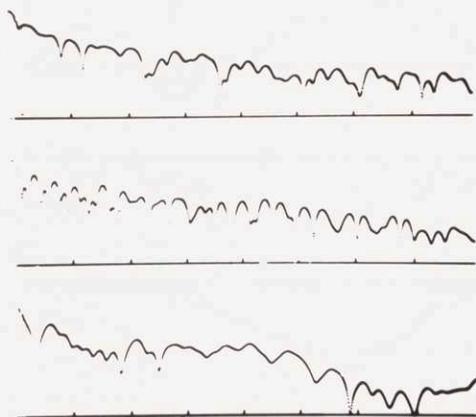

Figure 6.3(b), $n_0 = \frac{3}{4} N$

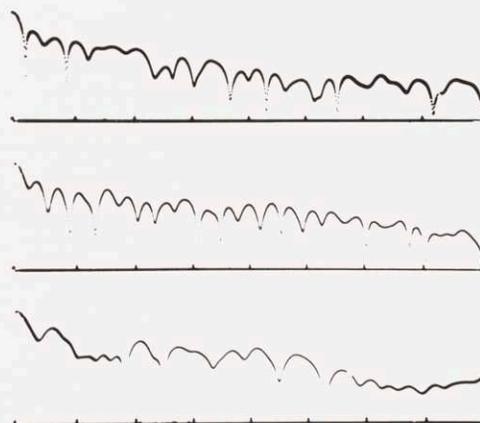Figure 6.3(c). $n_0 = \frac{1}{2} N$

Figure 6.3(d), $n_0 = 0$

Figure 6.3, Effects of Windowing on Log-magnitude Function
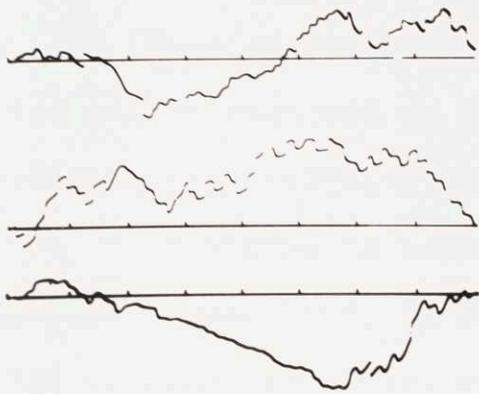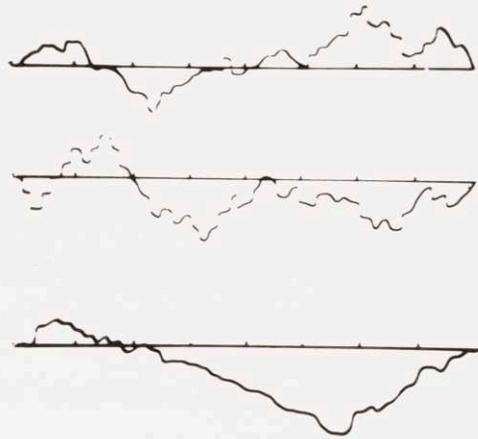
Figure 6.4(a), $n_0 = N$ Figure 6.4(b), $n_0 = \frac{3}{4} N$
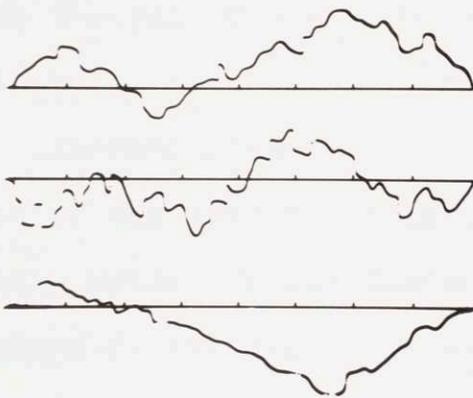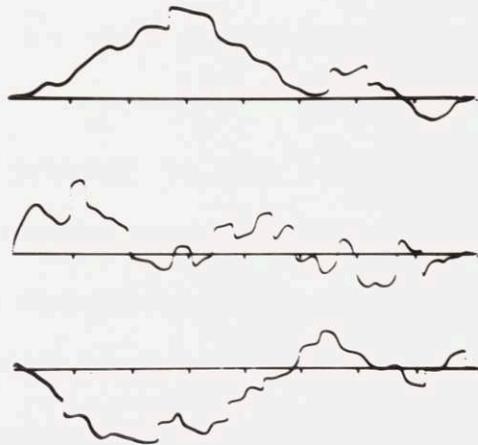
Figure 6.4(c), $n_0 = \frac{1}{2} N$ Figure 6.4(d), $n_0 = 0$

Figure 6.4
Effects of Windowing on Unwrapped Phase Function

for the three data sections. The value of n_0 was varied
from $n_0 = N$ in part (a) (rectangular window) to $n_0 = 0$ in
part (b) (Hanning window). Several interesting character-
istics can be observed in Fig. 6.3 and 6.4. As expected,
decreasing the effective window width results in a larger
degree of smoothing in both the log-magnitude and phase
functions. Note that, in addition to smoothing over zeroes
in the original function, at least one new zero is observ-
able in part (d), bottom line. This new zero near the unit
circle, in turn, causes a sharp jump in the otherwise smooth
phase function. For the remainder of this study n_0 was set
equal to $3N/4$. It was felt that any lower values might
distort or obscure important variations in the phase func-
tions while higher values of n_0 provided so little smoothing
that the job of phase unwrapping was made unnecessarily
difficult.

Another practical point to consider is the implementa-
tion of the inverse filter. Due to the presence of compu-
tation noise, it was decided that Helstrom's optimum MMSE
inverse filter should be used rather than the direct in-
verse. This formulation is reproduced in Eqn. (6.8).

$$G(u) = \frac{H^*(u)}{|H(u)|^2 + N/S} \tag{6.8}$$

where N/S is the noise-to-signal ratio. For this study, the

94

N/S ratio was assumed to be of the form

$$\frac{N}{S}(u) \Big|_{u = \frac{2\pi n}{512}} = 2^K(1 - a^n) \qquad (6.9)$$

where K is an integer scaling constant, a is a positive
number less than one, and u is the frequency variable. N/S
is plotted in Fig. 6.5 for various values of a. This model
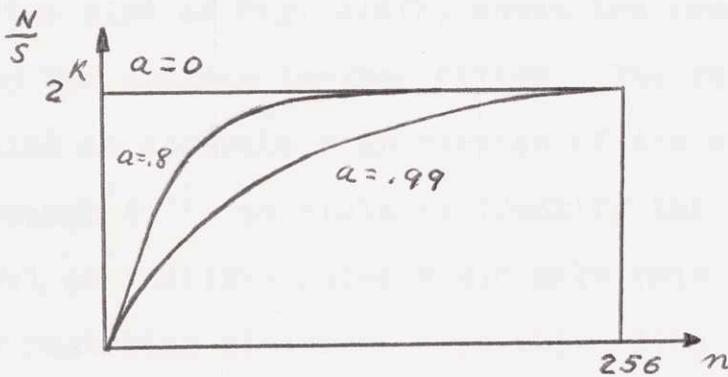
Figure 6.5

for the noise-to-signal ratio was chosen not only for its
flexibility but because it also reflects the fact that for
low frequencies the signal energy is much larger than the
noise energy while just the reverse is true at higher fre-
quencies.

To illustrate the utility of the optimum inverse
filter, some simple one-dimensional restoration experiments
were tried. Fig. 6.6(a) shows the original signal, a pulse,
and the result of convolving the pulse with a triangular im-
pulse response. This is a particularly difficult function
to invert since the inverse Fourier transform for the tri-

95

angle contains many singularities and is, thus, well
suited as an example of what happens to a restoration if
additive noise is ignored. The top line of Fig. 6.6(b)
shows the optimum inverse filter for a = 0, K = -34. The
only visible difference between this and the simple inverse
filter is at the points of singularity where, rather than
going to infinity, the optimum inverse is set to zero. The
center plot of Fig. 6.6(b) shows the result of deblurring
with the optimum inverse filter. The result can hardly be
called an accurate reproduction of the original pulse.
Although it is possible to identify the pulse, the high
level of additive noise would make this filter unsuitable
for restoring pictures. The third line of Fig. 6.6(b) shows
the result of setting all of the negative samples to zero.
This last step, passing the restored picture through a half-
wave rectifer, is often overlooked in image enhancement and
wastes a valuable bit of information that we have about the
original image: all of the picture sample values must be
positive. Although half-wave rectification didn't add much
to the quality of the restored pulse, it visibly improves the
quality of two-dimensional displays with variable brightness.
In this case, the half-wave rectification serves to reduce
the dynamic range of the signal to be displayed, thus making
the grey-scale compression less extreme.
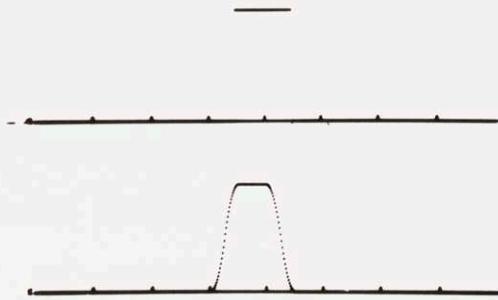
The next example, in Fig. 6.6(c), is with K = -16 and
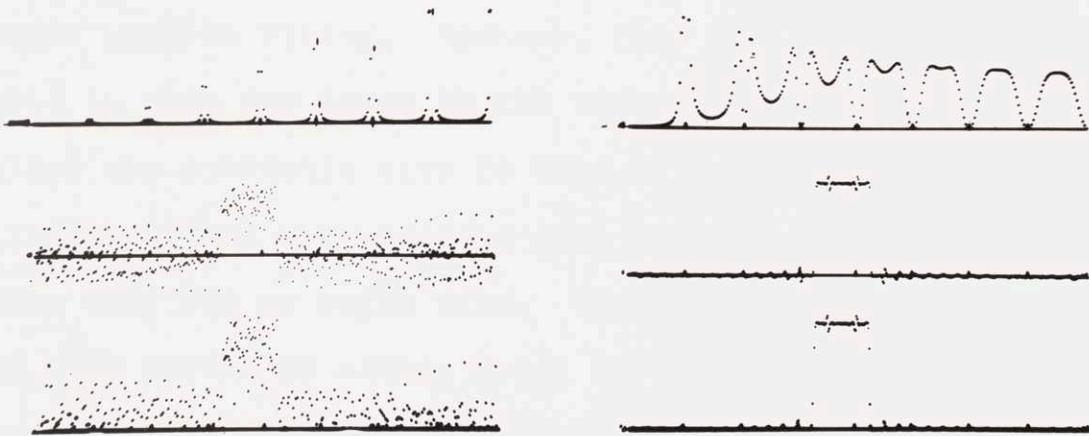
Figure 6.6(a), Original and Blurred Pulse

Fig. 6.6(b), K=-34, a=0   Fig. 6.6(c), K=-16, a=.996

Figure 6.6
Optimum Inverse Filtering

a = .996. Note how the frequency response of the optimum inverse filter is severly attenuated near the singular points. Also, for low frequencies, where N/S is still quite small, the response closely resembles that of the simple inverse filter in the previous example. Examination of the restored pulse shows that the additive noise has been drastically reduced. Of course, there is a penalty to be paid. For this particular example, the cost of noise reduction is a broadening of the transition bands of the pulse as well as some ringing near the edges of the pulse. Again, the third line shows the restored pulse after setting the negative sample points to zero.

The preceding examples clearly show the benefit to be derived from using the MMSE inverse filter rather than the simple inverse filter. However, they are somewhat unrealistic in that the input to the inverse filter, the blurred pulse, was available with 16 bits of precision. In a real-life situation, the received data is usually quantized using only six or eight bits. The coarse quantization has the same effect as adding noise to the received signal and, hence, can be expected to produce enhanced signals of lower quality than those obtained from a 16 bit received signal. Fig. 6.7(a), (b), and (c) illustrate this effect. Parts (a) and (b) are identical to Fig. 6.6(b) and (c) in all respects except that the received signal is quantized to only 8 bits.
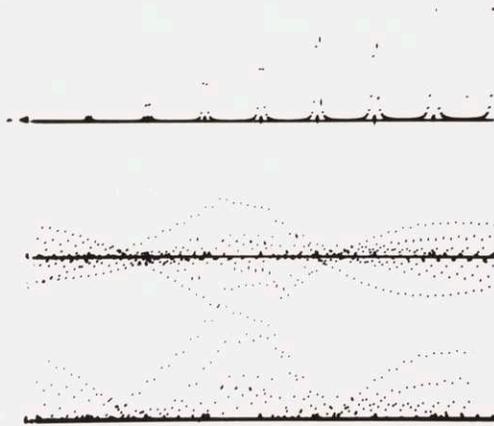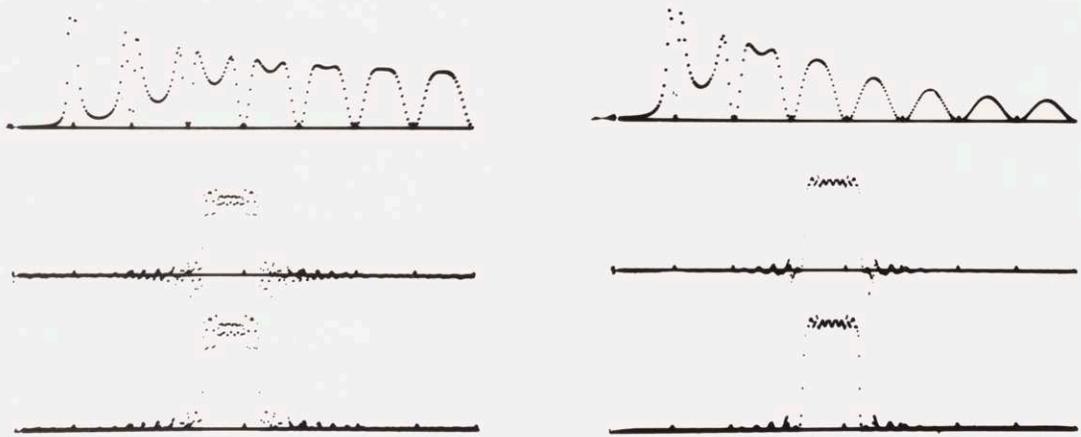
Figure 6.7(a), 8 bits, K=-34, a=0

Fig. 6.7(b), 8 bits
K=-16, a=.996

Fig. 6.7(c), 8 bits
K=-12, a=.996

Figure 6.7
Optimum Inverse Filtering of an 8-Bit Signal

The restored pulses are obviously of much poorer quality than the 16 bit restorations. Indeed, for part (a), it is difficult to even recognize the presence of a pulse, let alone ascertain the quality of the restoration. In Fig. 6.7(c), K is increased to -12, with a = .996, to reflect the increased N/S ratio. As might be expected, this results in the inverse filter having greater attenuation at all frequencies, particularly in the high end of the spectrum. The restored pulse obtained through this filter contains considerably less noise than that in part (b). The price for the reduced noise level is a somewhat greater broadening of the transition bands of the pulse. Thus, merely by increasing the N/S ratio in the optimum inverse filter, low noise restorations can be obtained from 8 bit signals. The performance penalty that must be paid is in the less-than-perfect restoration of the high frequency areas of the original signal.

Having considered and examined the effects of windowing, inverse filtering, and quantization, let us now return to the subject of this study: estimating the impulse response of the image degrading system. To this end, five examples are presented. The standard format of these examples is as follows: Part (a), the blurred image; Part (b), the restored image using known blurring function; Part (c), graphs showing actual and estimated impulse response functions and N, the

number of subsections used in the averaging; Part (d), the
restored image using the estimated impulse response func-
tion. Additional parts may be added to certain figures to
illustrate the effects of varying the N/S ratio or of
coarsely quantizing the received image, etc. These parts
will be explained as they arise. The reason for inverse
filtering the received picture with the known degrading
function (part (b)) was to observe the quality of restora-
tion that could be obtained under the best conditions. As
demonstrated when restoring the blurred pulse (Fig. 6.6),
certain blurring functions (in this case a triangular func-
tion) are rather difficult to invert, numerically, even when
they are known exactly.

Although the restored image (part (d)) is, of course,
the ultimate check on the overall quality of the estimation
procedure, the normalized, room-mean-squared (RMS) error
between the actual blurring function and the estimated im-
pulse response was also computed and is shown in part (c) of
the various figures. There are two reasons for doing this.
First of all, the RMS error is a single number which is
much more convenient for comparison purposes than the re-
stored picture, whose quality must be judged on a purely
subjective basis. Secondly, the RMS error is an indication
of the accuracy of the estimate that is totally independent
of the problems that arise during the actual inverse
filtering operation.

In order to simulate a degraded image, the New Crowd Scene was convolved with a series of one-dimensional impulse response functions. The convolution was actually carried out in the frequency domain by multiplying the Fourier transform of the original image with the transform of the impulse response and then inverse transforming. Due to the circular nature of the FFT algorithm used to calculate the Fourier transforms, vertical strips on the edges of the original image were set to zero so that the convolution would be linear rather than circular. See Gold and Rader (33) for a more detailed discussion of this technique.

Three blurring functions were used: Gaussian, triangle, and exponential. The Gaussian was chosen for its simplicity and also because many real-life degrading systems can be modelled as having a Gaussian impulse response. Another useful property is that it is a zero-phase function; i.e., the phase angle of a zero-mean Gaussian is everywhere zero. The triangle function was also chosen because its transform has zero phase everywhere, which greatly simplifies the estimation problem. However, it represents an added complexity in that the inverse filter constructed from the estimate will contain several almost-singular points and is, thus, a good test for the inverse filtering capabilities. Finally, the exponential signal was also chosen because of the properties of its Fourier transform. In particular, it has a

smoothly varying phase function which will test the ability
to estimate the phase angle of a degrading function.

The first example, shown in Fig. 6.8, uses a Gaussian
impulse response having a width of 32 points or a standard
deviation of 4 sample units. Thus, although the total
width of the Gaussian is equivalent to 1/16 of the picture
width, the effective width (defined as the number of samples
larger than 10% of the peak value) is much smaller, about
1/32 of the picture width. Thus, the error created by as-
suming independence of subsections in the convolution sum
will be quite small. Using an optimum inverse filter with
K = -20 and a = .996, a good quality reconstruction was
obtained and is shown in Fig. 6.8(b). Even though the ideal
inverse filter for the Gaussian is free of singularities,
the limited dynamic range of the computer registers (90 dB)
makes it necessary to use a non-zero value for the noise-to-
signal ratio to suppress noise effects at those frequencies
where the Gaussian function is more than 90 dB down from its
peak value. The result of the homomorphic estimation, over
256 subsections, is shown in Fig. 6.8(c), along with the
original Gaussian blurring function and the RMS error. The
only apparent errors in the estimate are the long tails and
the fact that the main lobe is somewhat wider than the ori-
ginal. As it turns out, this latter error has a signifi-
cant effect upon the restored image. Because the main lobe

Fig. 6.8(a), 16 pt. PSF.
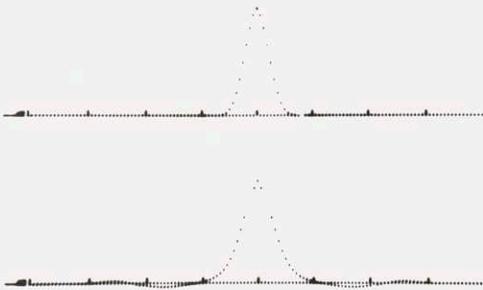(Gaussian)

Figure 6.8(b)
Ideal Restoration

Fig. 6.8(c), N=256, e=.261

Fig. 6.8(d), K=-16, a=.996
(Phase ≡ 0)
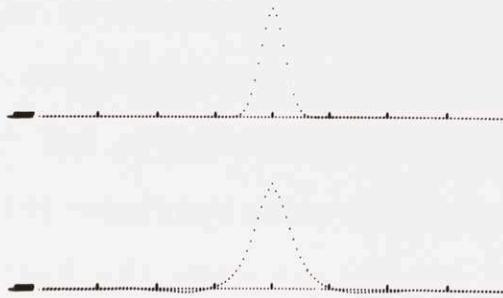
Fig. 6.8(e), N=64, e=.355
(Using phase estimate)

Fig. 6.8(f), K=-16, a=.996
(Using phase estimate)

is wider than the original, its frequency response will be
narrower than it should be. This, in turn, will result in
the inverse filter giving greater boost to the high frequen-
cies so that the subjective effect on the restoration will
be the same as if the original image was high-pass filtered.
Indeed, Fig. 6.8(d) confirms this prediction.

Fig. 6.8(c) and (d) made use of the fact that the de-
sired impulse response was known a priori to be a zero-
phase function. Fig. 6.8(e) and (f) were obtained without
the benefit of this assumption. That is, the phase angle
was estimated as accurately as possible. The resulting re-
storation shown in part (f) is only marginally poorer than
that of part (d), the most obvious flaw again being the
excess boost given to the high frequencies. Indeed, the
high frequencies appear to have been accentuated to a some-
what greater degree than in the assumed zero-phase case.

The next example, shown in Fig. 6.9, also uses a
Gaussian blurring function. The width, however, has been
doubled to 64 points, or a standard deviation of 8 samples.
This represents an effective width equal to approximately
1/16 of the total picture width and 1/4 of the selected sub-
section size. Notice how, for this case, even the ideal
restoration has significant additive noise. The reason, of
course, is that for wider Gaussian functions the frequency
response is narrower, which, in turn, means that an even

greater percentage of the spectrum will be more than 90 dB
down from the peak value. Of course, the noise could be
removed by boosting the N/S ratio in the inverse filter
formulation, which would then attenuate the high frequencies.
But, as shown in the one-dimensional examples (Fig. 6.6),
the cost of reducing the noise is a less-than-perfect rendi-
tion of the edges of the picture; i.e., some artificial
blurring would be introduced as an artifact of the inverse
filtering.

The only apparent defects in the estimated impulse re-
sponse (Fig. 6.9(c)) are the long tails on the Gaussian.
Unlike the first example, the effective width seems quite
close to that of the original. Thus, it is not surprising
that the restoration obtained from the estimated impulse
response (Fig. 6.9(d)) appears almost as good as the ideal
restoration. It is all the more surprising because the
effective width of the Gaussian is about 25% of the sub-
section size. Thus, either the Gaussian function is parti-
cularly well suited to homomorphic estimation, or the as-
sumption that the impulse response duration is much less than
the subsection size (Eqn. (5.16)) is overly restrictive.

As an example of the effects of quantization on two-
dimensional restorations, the estimated impulse response was
used to restore the blurred image after it had been quantized
using only 8 bits instead of the normal 16 bits. The result

107

Fig. 6.9(a), 32 pt. PSF.
(Gaussian)
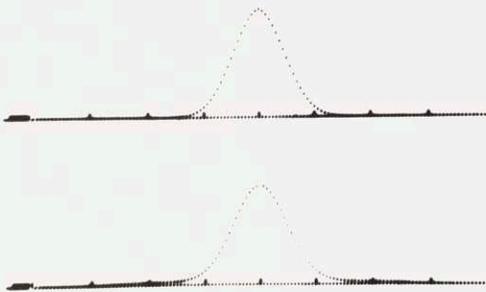

Figure 6.9(b)
Ideal Restoration

Fig. 6.9(c), N=256, e=.0712

Fig. 6.9(d), K=-18, a=.996
(Phase = 0)
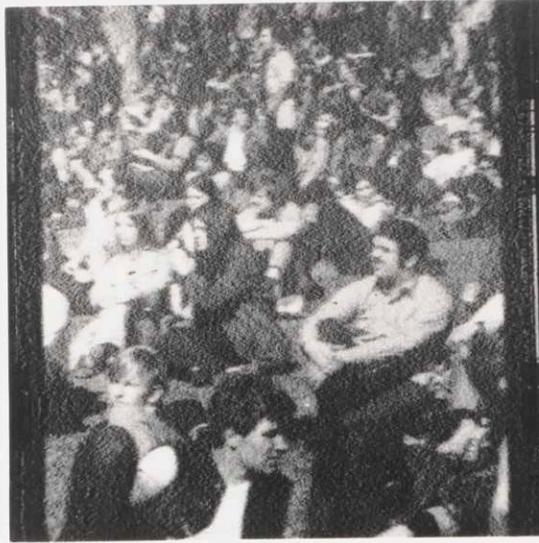
Fig. 6.9(e), K=-18, a=.996
(8 bits)

Fig. 6.9(f), K=-14, a=.996
(8 bits)

is shown in Fig. 6.9(e). As demonstrated in the earlier one-dimensional example (Fig. 6.7), the additive noise is much more pronounced in this case. Increasing K by 4 to take into account the increased noise levels yields the restoration of Fig. 6.9(f), which is of considerably higher quality than its predecessor.

The third example, Fig. 6.10, uses a 32 point triangle function to blur the original image. This function has an effective width of about 1/16 of the total picture width and 1/4 of the subsection size. The singularities in the inverse filter of the triangle function make it difficult to obtain even an ideal restoration, shown in Fog. 6.10(b). Not only is additive noise present in the restoration, but there is also an annoying periodic vertical stripe in the restored image. This is due to the fact that the periodic singularities in the inverse filter were set to zero rather than infinity.

The estimated impulse response, shown in Fig. 6.10(c), has a low RMS error of only .0657 even though the estimate has rather long tails not present in the actual blurring function. The effect of these tails on the restoration, however, is pronounced, with the restored image appearing as though it had been multiply exposed. In an effort to improve the restored image quality, the N/S ratio was increased by a factor of 16 and a changed to .75, with the result shown in Fig. 6.10(e). Although the echoes are still

Fig. 6.10(a), 32 pt. PSF.
(Triangle)

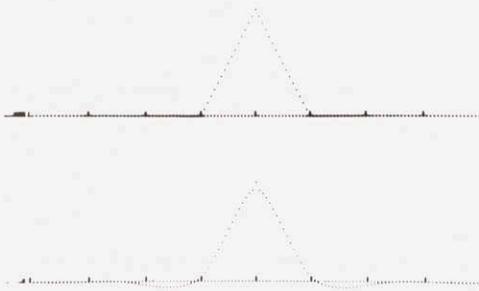Figure 6.10(b)
Ideal Restoration

Fig. 6.10(c), N=256, e=.0657

Fig. 6.10(d), K=-16, a=.996
(Phase = 0)

Fig. 6.10(e), K=-12, a=.75
(Phase = 0)
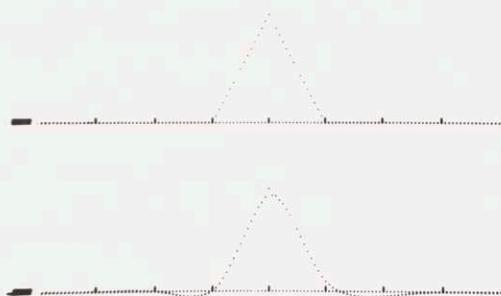

Fig. 6.10(f), N=64, e=.135
(Using phase estimate)

Fig. 6.10(g), K=-16, a=.996
(Using phase estimate)

visible, they have been reduced considerably. As usual, there is a performance trade-off in that the restored image now appears to be less sharp than that of Fig. 6.10(d).

The last two parts of this example illustrate the result obtained when the zero-phase nature of the triangle function is ignored and the phase function is estimated from the received image. Although the RMS error has more than doubled, the subjective quality of the restoration using the estimated phase appears to be about the same as that in Fig. 6.10(d).

It might be assumed from the quality of the restoration shown in the first and third examples that the job of estimating the phase angle was accomplished to a reasonable degree of accuracy. However, as evidenced by the next two examples, this was nothing more than a lucky fluke; lucky because both phase angles were actually zero. The next two examples show how much error occurs in the estimation of a non-zero phase function.

The fourth example uses a one-sided exponential signal to blur the original image. Thus

$$h(n) = b^n \qquad\qquad n \geq 0 \qquad\qquad\qquad (6.10)$$

With b = .75, the effective width of the impulse response was 8 samples or about 1/64 of the total picture width and 1/16 of the subsection size. The ideal restoration in Fig.

6.11(b) contains no visible defects. The estimated impulse response, on the other hand, has a considerable RMS error, which is obvious considering it is double sided while the original exponential is single sided. The fact that the estimate is almost symmetric about its center indicates that the imaginary part of its transform is almost zero, unlike that of the ideal exponential function. Using the estimated impulse response, the degraded image was restored as shown in Fig. 6.11(d). Fig. 6.11(e) shows the resulting restoration if the phase angle of the estimated impulse response is set to zero before inverse filtering. The primary effect of this total loss of phase information seems to be the loss of contrast of the restoration. Otherwise it appears about the same as that in Fog. 6.11(d). Finally, Fig. 6.11(f) plots the log-magnitude and phase function for both the ideal (top two) and estimated impulse responses (bottom two). As expected, the log-magnitude has been estimated quite accurately while the phase angle exhibits gross errors throughout the spectrum. Although the automatic scaling in the display programs make it appear as though the ideal and estimated phase functions have the same range, in fact, the ideal phase ranges over $[0, .27\pi]$ while the estimated phase is confined to $[-.04\pi, .02\pi]$.

The fifth and final example also uses an exponential signal, but b is increased to .906. This provides an effec-

Fig. 6.11(a), 8 pt. PSF.
(Exponential, b=.75)

Figure 6.11(b)
Ideal Restoration

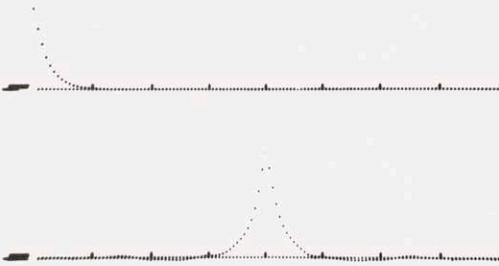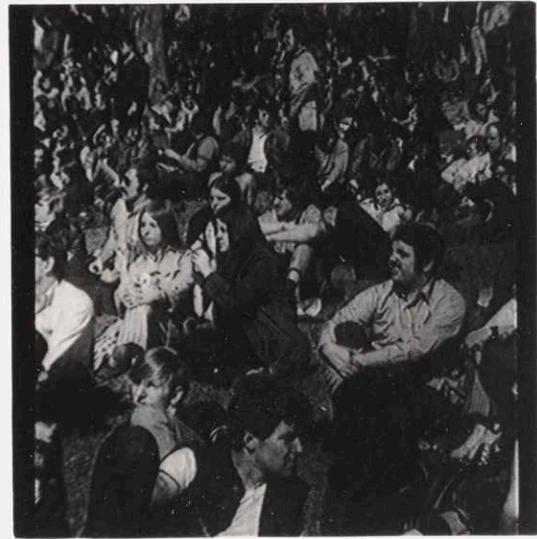Fig. 6.11(c), N=256, e=.852   Fig. 6.11(d), K=-10, a=.996

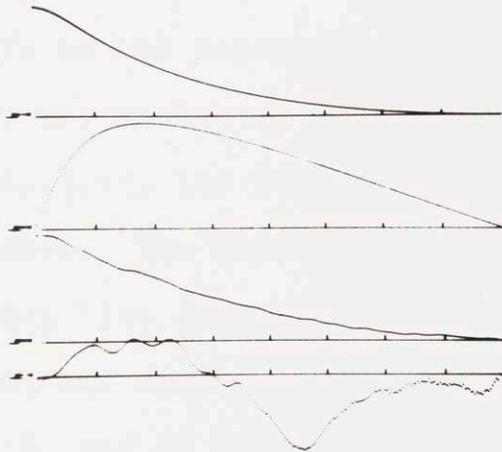Fig. 6.11(e), K=-10, a=.996
(Phase ≡ 0)

Fig. 6.11(f), Ideal (top) and
est'ed (bottom) log-mag and phase functions
Range of ideal phase - $(0, .27\pi)$
Range of est'ed phase - $(-.04\pi, .02\pi)$

tive width of 25 samples. Again, the estimated impulse response is very nearly an even function. A comparison of the log-magnitude and phase functions (Fig. 6.12(f)) again shows gross errors in the estimated phase function. The ideal phase angle ranges over $[0, .36\pi]$ while the estimated phase angle lies in the interval $[-.15\pi, 0]$. Again, a restoration was made with the phase angle set to zero (Fig. 6.12(e)). Unlike the previous example, the subjective quality of the zero-phase restoration is noticeably worse than the restoration obtained by using the (poorly) estimated phase angle. The last two parts of Fig. 6.12 show restorations obtained from an 8 bit blurred image. Part (g) uses the same N/S ratio as part (d) while in part (h), the N/S ratio was increased by a factor of 16.

One aspect of the estimation problem that hasn't been treated explicitly is how sensitive the estimated impulse response is to N, the number of subsections used in averaging. In Fig. 6.13(a), the 32 point Gaussian was used to test this dependence. The zero-phase assumption was used in all cases. The top line shows the original Gaussian while the second, third, and fourth lines show the estimates obtained using 4, 16, and 64 subsections respectively. Note that the quality of the estimate is quite good, even for N = 4. Fig. 6.13(b) shows a similar set of experiments based upon the exponential signal with a = .75. The display

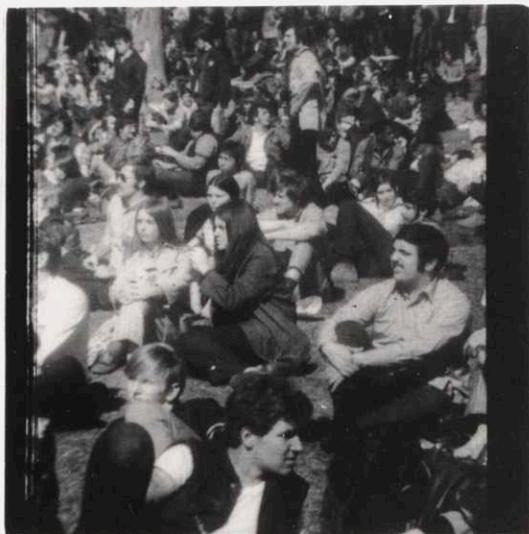Fig. 6.12(a), 25 pt. PSF.
(Exponential, b=.906)

Figure 6.12(b)
Ideal Restoration
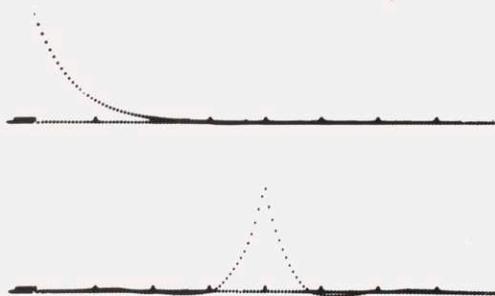

Fig. 6.12(c), N=64, e=.854

Fig. 6.12(d), K=-16, a=.996

Fig. 6.12(e), K=-16, a=.996
(Phase = 0)
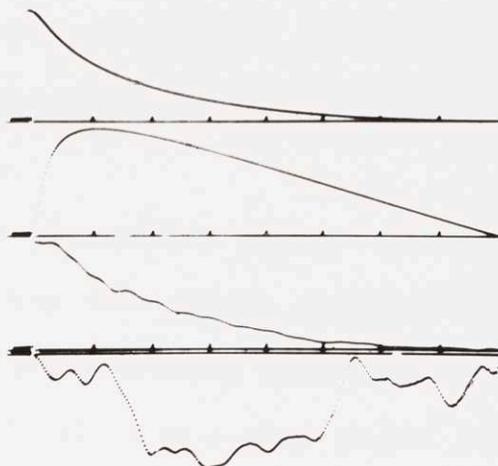
Fig. 6.12(f), Ideal (top)
and est'ed (bottom) log-mag
and phase functions
Ideal $\angle \in (0, .36\pi)$
Est'ed $\angle \in (-.15\pi, 0)$
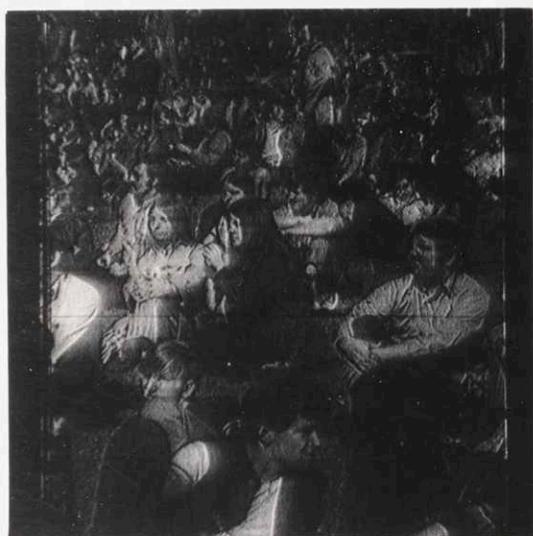
Fig. 6.12(g), K=-16, a=.996
(8 bits)
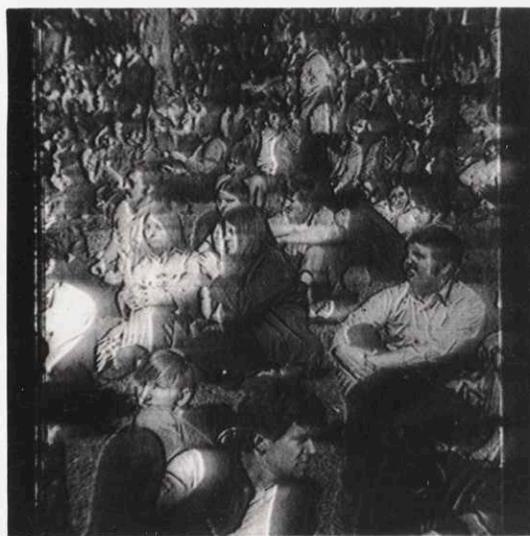
Fig. 6.12(h), K=-12, a=.996
(8 bits)

format is identical to part (a). Note, however, that none
of the examples, even N = 64, provide what might be called
an accurate estimate of the actual exponential function.
The reason, of course, is the errors incurred during the
estimation of the phase function.

In summary, the homomorphic estimation procedure appears
to be quite effective in estimating zero-phase impulse re-
sponse functions. However, it does a rather poor job when-
ever the unknown impulse response has a non-zero pahse func-
tion. Restated in terms of frequency domain functions,
homomorphic estimation does a good job of estimating log-
magnitude functions, but is unable to obtain a good phase
estimate. In some applications, this need not be a serious
handicap. For example, long exposure atmospheric degrada-
tion is often modelled as being Gaussian-like. Of course,
in those cases where the blurring function is assumed to be
Gaussian, it would probably be simpler to merely estimate
the variance of the curve rather than its value at every
point.

An interesting result of the two exponential examples
is that a restoration obtained by assuming zero-phase still
provides some improvement in image quality over the original
blurred image. Thus, homomorphic estimation does offer a
general means for obtaining some improvement in received
image quality, if only through its ability to accurately

Ideal

N=4, e=.372

N=16, e=.352

N=64, e=.358

Figure 6.13(a)

Ideal

N=4, e=1.74

N=16, e=.972

N=64, e=.807
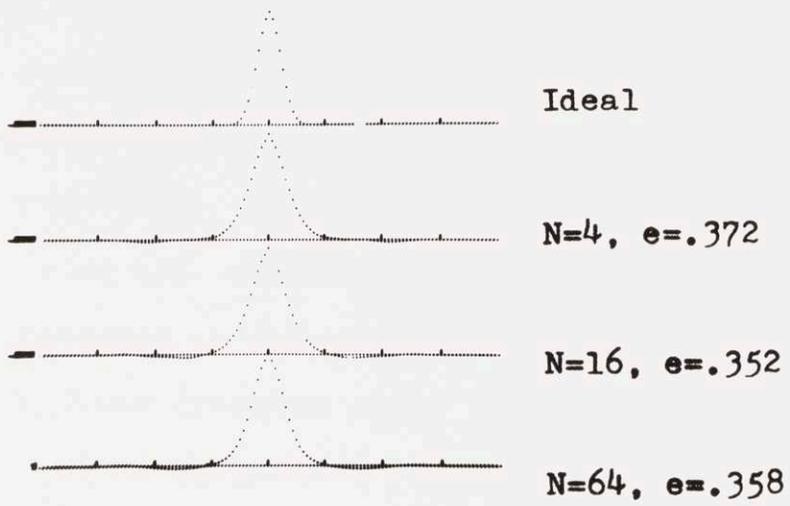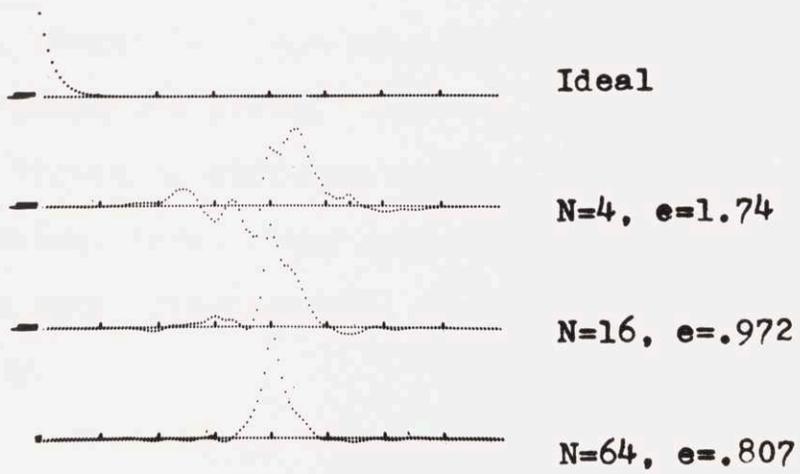
Figure 6.13(b)

estimate the log-magnitude function of the degrading impulse response.

The ability to accurately estimate the phase function is limited for two reasons: errors in computing $PV(\angle)$ when the magnitude is small and error propagation in the phase unwrapping algorithm. The first error primarily depends upon the size of the computer register and the type of arithmetic employed in the computation of the FFT. On the PDP-9, the FFT was computed using 18 bits and integer arithmetic. Other computers, for example, an IBM 360, use floating point arithmetic and 32 bit words in the FFT calculation. Since the FFT algorithm generates less noise when using floating point arithmetic (31), it is quite likely that this source of error can be effectively eliminated on a different computer. On the other hand, error propagation during phase unwrapping is not directly affected by changing computers. The only cure is to make no errors while unwrapping the phase.

In an effort to overcome these limitations on homomorphic estimation, three other approaches to estimating the phase angle were investigated. These methods are presented in Chapter 7.

7. ALTERNATE APPROACHES TO PHASE ESTIMATION

7.1 Introduction

As shown in Chapter 6, homomorphic estimation does a rather poor job of estimating the phase angle of the unknown blurring function. In this chapter, three alternate means of estimating the phase function are explored. The first of these is an offshoot of the homomorphic estimation procedure. The second limits itself to minimum phase functions, while the third explores a statistical approach to phase estimation. All three suffer from either theoretical or practical drawbacks and were, therefore, rejected in favor of the phase unwrapping technique. However, they are included here as an aid to future researchers.

7.2 Phase Derivatives

The method of phase derivatives makes use of the time (spatial) domain function corresponding to $\hat{R}(u,v) \equiv \log R(u,v)$. This function is commonly called the complex cepstrum of $r(n,m)$. The method is based upon the following result.

$$\hat{r}(n,m) \leftrightarrow \hat{R}(u,v) = \log R(u,v) \qquad (7.1)$$

$$n \cdot \hat{r}(n,m) \leftrightarrow j\frac{\partial}{\partial u}\hat{R}(u,v) = \frac{j\frac{\partial}{\partial u}R(u,v) \leftrightarrow nr(n,m)}{R(u,v) \leftrightarrow r(n,m)} \qquad (7.2)$$

where $\hat{r}(n,m)$ denotes the complex cepstrum of $r(n,m)$. The division of the Fourier transform, $j \frac{\partial}{\partial u} R(u,v)$ by $R(u,v)$ in Eqn. (7.2) represents the first practical problem with the method of phase derivatives. Previous experience in creating inverse filters should serve to indicate the type of problems that can arise when inverting Fourier transforms. Unfortunately for this case, there is nothing comparable to an optimum inverse filter that can be employed to attenuate the adverse effects of extraneous noise. However, if the computation could be performed using high precision, floating point arithmetic, rather than the 18 bit integer format of the PDP-9, then perhaps a useful result could be obtained. It should also be pointed out that the basic phase derivative result, Eqn. (7.2), can be reformulated in a complimentary manner as shown in Eqn. (7.3).

$$
\hat{mr}(n,m) \leftrightarrow j \frac{\partial}{\partial v} \hat{R}(u,v) = \frac{j \frac{\partial}{\partial v} R(u,v) \leftrightarrow mr(n,m)}{R(u,v) \leftrightarrow r(n,m)} \tag{7.3}
$$

Assuming for the moment that the division of transforms can be carried out satisfactorily, let us examine what opportunities this offers towards the problem of phase estimation. The most obvious benefit of this approach is the freedom from having to deal with the PV($\angle$). The phase derivatives are based upon the smoothed phase angle, not its principal value. Thus, since the averaging and derivative

124

operators commute, the derivative of the phase angle of
$H(u,v)$ can be estimated directly without having to perform
phase unwrapping. Thus

$$\overline{\frac{\partial}{\partial u} \angle H} = \frac{1}{N} \sum_i \frac{\partial}{\partial u} \angle R_i \tag{7.4}$$

Ultimately, of course, the $\angle H$ will have to be reconstructed
from $\frac{\partial}{\partial u} \angle H$. This creates another practical problem because
the accurate reconstruction of $\angle H$ from samples of $\frac{\partial}{\partial u} \angle H$
requires a large number of samples--perhaps as many as re-
quired in the phase unwrapping algorithm. This is parti-
cularly true in those areas where the phase function changes
rapidly. In addition, the integration of the phase deri-
vative also tends to propagate any errors that occur, just
as in phase unwrapping. The reason is that each point on
the integrated phase curve is a function of all those
points that preceded it. Thus, an error anywhere along a
line tends to propagate undiminished unless a compensating
error occurs.

The problems with low sampling rate and error propaga-
tion can be exchanged for other difficulties by working with
the complex cepstrum. Since the Fourier transform is a
linear operation, the averaging operation can just as well
be carried out in the time (spatial) domain as in the fre-
quency domain. Thus,

$$\widetilde{n\hat{h}}(n,m) = n \frac{1}{N} \sum_i \hat{r}_i(n,m) \qquad (7.5)$$

Now, $\widetilde{\hat{h}}(n,m)$ can be found from Eqn. (7.5) at all points
except along the n=0 axis. Of course, Eqn. (7.3) could be
used to find h(n,m) everywhere except along the m=0 line.
This result, combined with that derived from Eqn. (7.2),
could be used to find h(n,m) everywhere except at the origin.
However, this single point can be determined from the basic
definition of $\widetilde{\hat{h}}(n,m)$,

$$\widetilde{\hat{h}}(0,0) = \frac{1}{4\pi^2} \iint\limits_{-\pi}^{\pi} \log H(u,v) \, du \, dv \qquad (7.6)$$

Using the symmetry properties of the Fourier transform of
real functions, Eqn. (7.6) can be expressed in terms of the
$\log |H(u,v)|$ only, the phase angle having been eliminated by
virtue of its odd symmetry. Thus,

$$\widetilde{\hat{h}}(0,0) = \frac{1}{4\pi^2} \iint\limits_{-\pi}^{\pi} \log |H(u,v)| \, du \, dv \qquad (7.7)$$

Computationally, this procedure requires that for each sub-
section five Fourier transforms be computed plus two divi-
sion-of-transform operations, and finally, Eqn. (7.7) must
be evaluated. If one limits himself to the use of only Eqn.
(7.2), the computation time can be cut in half. The defini-

126

tion of $\widetilde{\widehat{h}}(n,m)$ can be used to find $h(0,m)$ as shown in Eqn. (7.8).

$$\widetilde{\widehat{h}}(0,m) = \frac{1}{4\pi^2} \iint\limits_{-\pi}^{\pi} \log H(u,v)\ e^{jvm}\ du\ dv \qquad (7.8)$$

which can be simplified to

$$\widetilde{\widehat{h}}(0,m) = \frac{1}{4\pi^2} \iint\limits_{-\pi}^{\pi} \log |H(u,v)| \cdot \cos vm - \angle H(u,v) \sin vm\ du\ dv$$
$$(7.9)$$

Now, Eqn. (7.9) requires the smoothed phase angle function which, in turn, means that the phase unwrapping algorithm must be used to find $\angle H$. Thus, the price for reducing the apparent computation requirements is the need to once again unwrap the phase angle, which defeats the primary purpose of introducing the phase derivative procedure.

There is yet an additional consideration when using the complex cepstrum for averaging. This has to do with the aliasing error in the complex cepstrum arising through the computation of the discrete Fourier transform. Rather than obtaining $n\widehat{r}_i(n,m)$, one actually obtains an aliased version: $(n+kM)\widehat{r}_i(n+kM,m+kM)$, where M is the size of the FFT, and k takes on all integer values. Thus, unless $\widehat{r}_i(n,m)$ falls off very quickly, particularly in the n direction, or M is made quite large, a substantial aliasing error will occur in $n\widehat{r}_i(n,m)$. This error, in turn, causes errors in the estimate of $\widehat{h}(n,m)$.

127

Thus, of the three possible uses of phase derivatives, it appears that estimating $\angle H$ from $\frac{\partial}{\partial u} \angle H$ is the easiest method. The "only" requirements are enough samples of the phase derivative and an accurate means of performing the division-of-transforms operation.

7.3 Minimum Phase Functions

A well known result of one-dimensional signal processing is that the phase angle of a signal can be uniquely recovered from the log-magnitude function provided the signal is a minimum phase (or maximum phase) signal. Since the homomorphic estimation procedure has a demonstrated capability for accurately estimating the log-magnitude, it appears that the minimum phase property could be used to find the unknown phase function, provided the unknown function is minimum (maximum) phase. That last qualification is one of the major limiting factors in using the mimimum phase property to determine the phase angle of blurring functions; i.e., real-life blurring functions are distinctly non-minimum phase.

There are three interchangeable definitions of one-dimensional, minimum phase (MP) functions. The basic definition is: A MP function is that function chosen from the set of all functions having the same autocorrelation function (i.e., the same magnitude of the Fourier transform) which has minimum phase for all frequencies. The second definition is given in terms of the pole-zero pattern of $f(n)$: $f(n)$ is MP

if all of its poles and zeroes lie inside the unit circle. This definition implies a necessary, though not sufficient, condition for $f(n)$ to be MP--$f(n)$ must be a causal function. The third definition is stated in terms of the complex cepstrum of $f(n)$: $f(n)$ is minimum phase if $\hat{f}(n)$ is causal; i.e., $\hat{f}(n)=0$, $n \leq 0$. This can be restated as: $f(n)$ is MP if $\hat{f}(n)$ exhibits even part sufficiency; i.e., $\hat{f}(n)$ can be determined from the even part of $\hat{f}(n)$.

Extending these definitions to two dimensions is not trivial. While the first definition can be carried over directly, it does not provide a practical means of testing whether a function is MP or generating such functions. The second definition breaks down because the concept of discrete poles and zeroes is no longer valid in two dimensions. The third definition extends directly, provided that "causal" is interpreted as strictly causal in both n and m, not just one or the other.

Now, assuming that the log-magnitude of a MP function is known, how does one compute the corresponding phase functions? In one dimension, three steps are required and are shown in Fig. 7.1, where

$$
w(n) = \begin{cases} 1 & n = 0 \\ 2 & n > 0 \\ 0 & n < 0 \end{cases} \tag{7.10}
$$

129
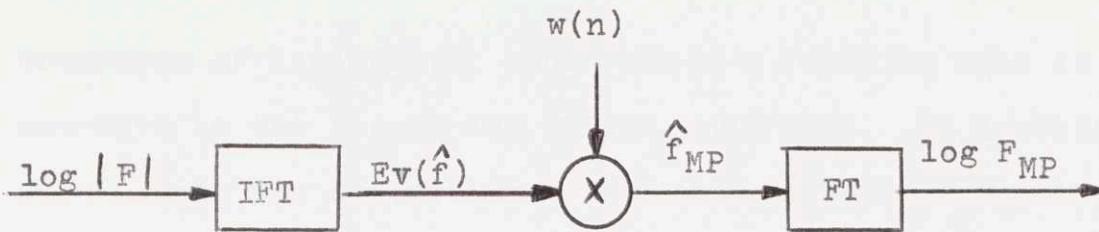
Figure 7.1

Thus, the even part of the complex cepstrum is obtained by inverse Fourier transforming the $\log|F|$. Then, using even-part sufficiency, $\hat{f}_{MP}(n)$ is computed from its even part. Finally, $\hat{f}(n)$ is Fourier transformed to yield $\log F_{MP}(u)$. Note that even if the original function that generated $\log|F|$ was not minimum phase, the output of Fig. 7.1 is a minimum phase function because of the utilization of the even-part sufficiency property.

The adaptation of this computational procedure to two-dimensional signals poses no difficulty if the complex cepstrum of the original signal is strictly causal. In that case, $w(n,m)$ is

$$w(n,m) = \begin{cases} 1 & n,m = 0 \\ 2 & n,m > 0 \\ 0 & \text{elsewhere} \end{cases} \tag{7.11}$$

However, it is quite conveivable that the inverse Fourier

130

transform of $\log|F(u,v)|$ could yield a function that is non-zero in the second and fourth quadrants. To provide for such cases, the definition of two-dimensional MP functions must be broadened to include those functions whose complex cepstrum is non-zero over no more than half of the n-m planes, including the first quadrant. But this definition introduces an ambiguity into the even part sufficiency property. Consider Fig. 7.2. Given

$$ \mathrm{Ev}\left[\, f(n,m)\right] = \tfrac{1}{2}\left[\, f(n,m) + f(-n,-m)\right] \,, \qquad (7.12) $$

one sees that $\mathrm{Ev}\left[\, f(n,m)\right]$ could have been generated by any one of the three functions shown in Fig. 7.2. The non-zero region of each function is shown as a shaded area.

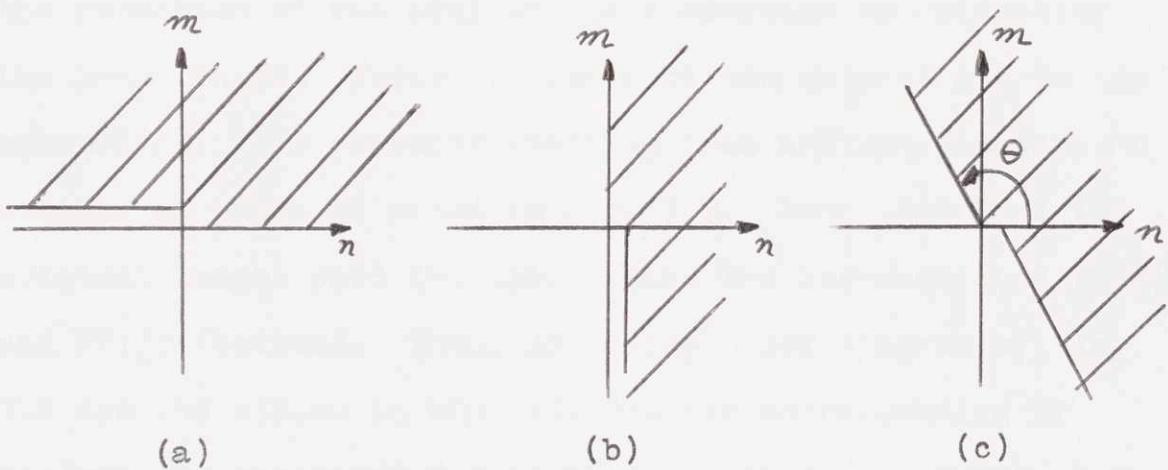(a)   (b)   (c)

Figure 7.2

The ambiguity is reflected in the three definitions for $w(n,m)$. For part (a),

$$w(n,m) = \begin{cases} 1 & n \geq 0, \quad m = 0 \\ 2 & m \geq 1 \\ 0 & \text{elsewhere} \end{cases} \qquad (7.13)$$

while for part (b),

$$w(n,m) = \begin{cases} 1 & m \geq 0, \quad n = 0 \\ 2 & n \geq 1 \\ 0 & \text{elsewhere} \end{cases} \qquad (7.14)$$

The expression for $w(n,m)$ corresponding to Fig. 7.2(c) is a rather complicated function of Θ.

This basic ambiguity, coupled with the fact that typical blurring functions are not minimum phase, led to the rejection of the minimum phase approach to estimating the phase angle. However, some work was carried out on the generation of MP pictures starting from ordinary images, an example of which is shown in Fig. 7.3. Part (a) shows the original image; part (b) illustrates the log-magnitude (top) and PV($\angle$) (bottom). Then, using the block diagram of Fig. 7.2 and the $w(n,m)$ in Eqn. (7.13), the corresponding MP picture was generated and is shown in part (c) of Fig. 7.3 along with its log-magnitude function (top) and phase function (bottom) in part (d). Reference (8) contains further details and more examples on the topic of generating MP pictures.
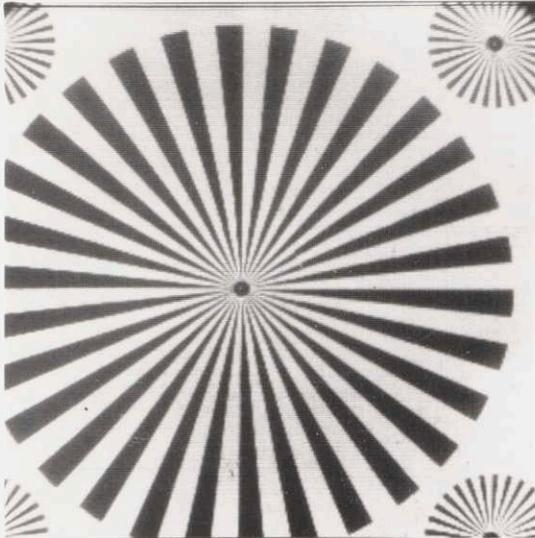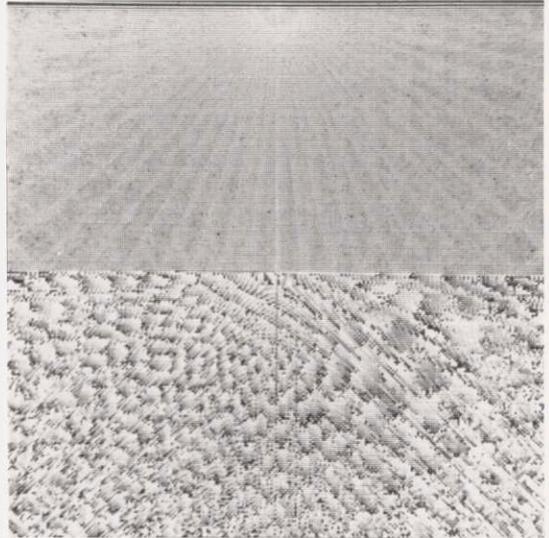
Figure 7.3(a)
(Original)
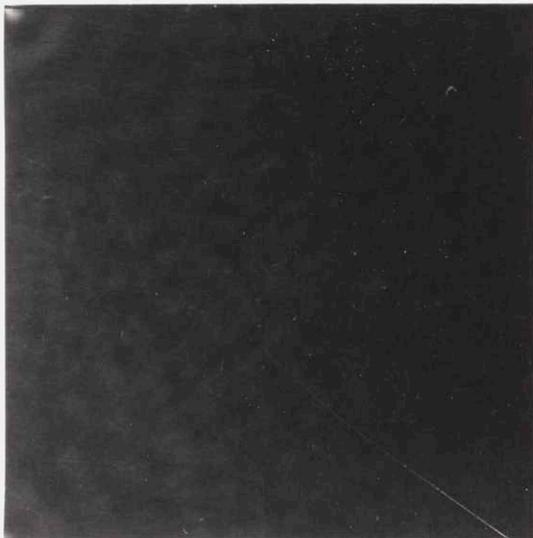

Fig. 7.3(b), Log-mag.
(top) & PV∠ (bottom)

Figure 7.3(c)
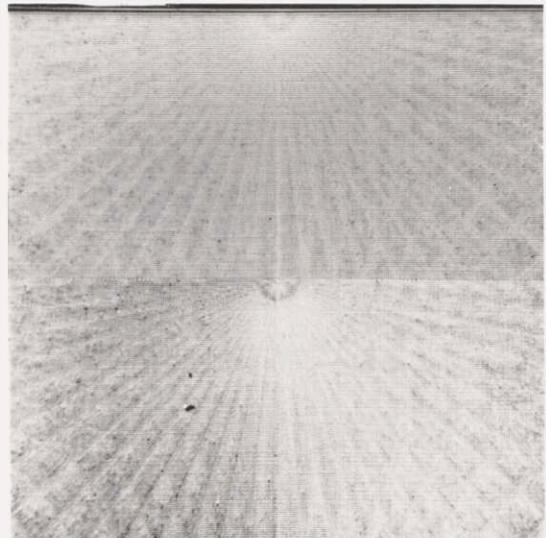(Minimum Phase Image)

Fig. 7.3(d), Log-mag.
(top) & phase angle for
minimum phase image

7.4 Filtering the Phase Angle

This approach to estimating the phase function was motivated by the observation that the autocorrelation of $\angle H$ had a much greater effective duration than the autocorrelation function of $\angle S$. It was hoped that a useful result could be obtained by treating the estimation problem as a problem in filtering two stochastic processes that had been passed through the non-linear, PV operator. Using $\phi_{\angle R}$ to denote the autocorrelation of the true sum of $\angle H$ and $\angle S$, i.e., the input to the PV operator, and $\phi_{PV\angle R}$ to denote the autocorrelation function of the output of the PV operator, one can show that under a Gaussian assumption

$$\phi_{PV\angle R}(u) = \sum_{k \text{ odd}} c_k \phi_{\angle R}^k(u) \qquad (7.15)$$

The constant c_k equals

$$c_k = \begin{cases} \frac{4}{k!}\left[\sum_{n=1}^{n} (-1)^n n^{k-1} \exp(-\tfrac{1}{2}\,\phi_{\angle R}(0)\cdot n^2)\right]^2 & k \text{ odd} \\ \\ 0 & k \text{ even} \end{cases} \qquad (7.16)$$

Eqns. (7.15) and (7.16) are derived in Appendix A.

Now, in order to explore how linear filtering might be employed to separate $\angle H$ from $\angle S$, let us examine the first few terms of Eqn. (7.15). In particular, we are interested in

134

the PSD functions associated with each term, $\phi^k_{\angle R}$. Using $\Phi_{\angle S}$ to denote the PSD for $\angle S$ and $\Phi_{\angle H}$ to denote the PSD function of $\angle H$, we have

$$\Phi_{PV\angle R} = \sum_{k \text{ odd}} c_k \; \Phi_{\angle R}^{\;k} * \Phi_{\angle R} \tag{7.17}$$

where the k over the convolution symbol indicates that $\Phi_{\angle R}$ is to be convolved with itself (k-1) times.

For example,

$$\Phi_{\angle R}^{\;1} * \Phi_{\angle R} = \Phi_{\angle R}$$

and $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ (7.18)

$$\Phi_{\angle R}^{\;3} * \Phi_{\angle R} = \Phi_{\angle R} * \Phi_{\angle R} * \Phi_{\angle R}$$

Using the fact that $\angle R = \angle S + \angle H$, and assuming independence between $\angle S$ and $\angle H$, $\Phi_{\angle R}$ can be written as

$$\Phi_{\angle R} = \Phi_{\angle S} + \Phi_{\angle H} \tag{7.19}$$

Also, the assumptions on the effective autocorrelation durations for $\angle S$ and $\angle H$ imply that the bandwidth of $\Phi_{\angle S}$ is much larger than the bandwidth of $\angle H$, as shown in Fig. 7.4. Thus, at least for the first term of Eqn. (7.17), simple low-pass filtering should serve to separate $\angle H$ from $\angle S$. However, the higher order terms in Eqn. (7.17) present a much more formidable task due to the interaction of signal components
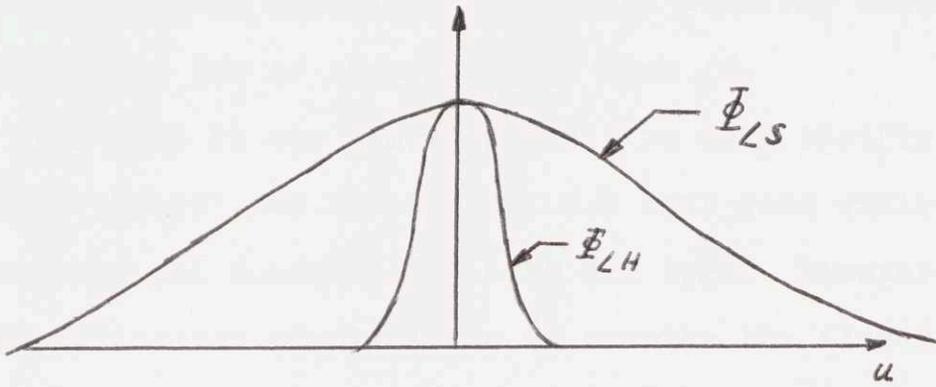
Figure 7.4

via the convolution operation. For example, Fig. 7.5 shows
the four distinct signal components contained in the second
term of Eqn. (7.17).
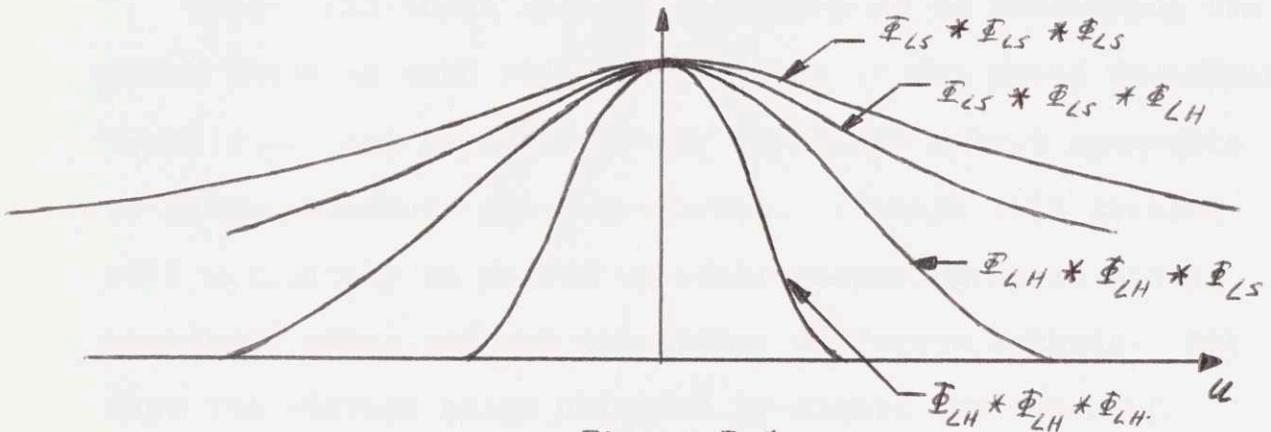
Figure 7.5

Notice that only one term is totally free of noise compon-
ents and that this term has three times the bandwidth of its
counterpart in Fig. 7.4. The situation becomes rapidly worse
for higher order terms with the bandwidth of the signal-only
component expanding to 5, 7, 9, ... times its original width.

Thus, it appears impossible to design a filter that can do a reasonable job of separating $\angle H$ from $\angle S$.

But what if one limits himself to only the first term of Eqn. (7.17)? We saw in Fig. 7.4 that good quality signal separation was possible for this one term. However, a moments reflection should serve to uncover the flaw in this reasoning. Approximating Eqn. (7.17) by only its first term is equivalent to approximating the PV operator by a straight line. This approximation completely ignores the fact that the PV operator does not preserve the desired property of additivity, which was the original motivation for developing the phase unwrapping procedure.

Thus, all three alternate approaches to estimating the phase function were rejected in favor of the phase unwrapping technique. But even the latter technique proved incapable of making accurate phase estimates. Perhaps this problem will ultimately be solved by other researchers, on other machines, using perhaps altogether different methods. But from the vantage point afforded by almost two years of research into the problems of phase estimation, I know that the ultimate solution to this problem requires not only a good theoretical understanding of the problems involved, but also a machine whose capabilities are up to this formidable computing task.

REFERENCES

1. G. Anderson, "Images and Fourier Transform", S.M. Thesis, MIT, 1967.

2. R. Barakat and A. Houston, "Line Spread Function and Cumulative Line Spread Function for Systems with Rotational Symmetry", J.O.S.A., vol. 54, June, 1964, p. 768.

3. C. W. Barnes, "Object Restoration in a Diffraction Limited Imaging System", J.O.S.A., vol. 56, May, 1966, pp. 575-578.

4. F. C. Billingsley, "Applications of Digital Image Processing", Applied Optics, vol. 9, February, 1970, p. 289.

5. E. S. Blackman, "Effects of Noise on the Determination of Photographic System Modulation Transfer Functions", Photographic Science and Engineering, vol. 12, September, 1968, p. 244.

6. W. B. Davenport and W. L. Root, Introduction to the Theory of Random Signals and Noise, Chapter 13, McGraw-Hill Co., New York, 1958.

7. R. Deutsch, Nonlinear Transformations of Random Processes, Prentice Hall Co., Englewood Cliffs, N.J., 1962.

8. A. E. Filip, "Generation of Minimum Phase Pictures", Q.P.R., No. 104, RLE, MIT, January, 1972, pp. 283-289.

9. B. R. Frieden, "Image Evaluation by Use of the Sampling Theorem", J.O.S.A., vol. 56, October, 1966, p. 1355.

10. B. R. Frieden, "Band Unlimited Reconstruction of Optical Objects and Spectra", J.O.S.A., vol. 57, August, 1967, p. 1013.

11. J. W. Goodman, Introduction to Fourier Optics, McGraw-Hill, New York, 1968.

12. J. H. Haines and J. P. King, "Approximate Computation of Physical MTF", S.P.I.E. Seminar on Modulation Transfer Functions, vol. 13, 1968, pp. 105-113.

13. J. L. Harris, "Resolving Power and Decision Theory", J.O.S.A., vol. 54, May, 1964, p. 606.

14. J. L. Harris, "Image Evaluation and Restoration", J.O.S.A., vol. 56, May, 1966, p. 569-574.

15. C. W. Helstrom, "Image Restoration by Method of Least Squares", J.O.S.A., vol. 57, March, 1967, p. 297.

16. T. S. Huang, "Image Enhancement: A Review", Opto-Electronics, vol. 1, January, 1969, p. 49.

17. T. S. Huang, W. F. Schreiber, and O. J. Tretiak, "Image Processing", Proc. IEEE, vol. 59, November, 1971, pp. 1586-1609.

18. E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Chapter 2, Wiley & Sons, New York, 1966.

19. G. M. Jenkins and D. G. Watts, Spectral Analysis and Its Applications, Chapter 6, Holden-Day Co., San Francisco, California, 1969.

20. E. I. Jury, "A Simplified Stability Criterion for Linear Discrete Systems", Proc. IRE, vol. 50, June, 1962, pp. 1493-1500.

21. D. E. Knuth, Seminumerical Algorithms, Chapter 3, Addison-Wesley, Reading, Mass., 1969.

22. A. Labeyrie, "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images", Astron. and Astrophys., vol. 6, 1970, pp. 85-87.

23. S. H. Lerman, "Application of the FFT to the Calculation of the OFT", S.P.I.E. Seminar on Modulation Transfer Functions, vol. 13, 1968, pp. 51-70.

24. M. Marden, The Geometry of the Zeros of a Polynomial in a Complex Variable, Math. Surveys, Vol. 3, Amer. Math. Soc., New York, 1949.

25. B. L. McGlamery, "Restoration of Turbulence Degraded Images", J.O.S.A., vol. 57, March, 1967, p. 293.

26. B. L. McGlamery, "Image Restoration Techniques Applied to Astronomical Photography", presented at the Princeton Seminar of Astronomical Use of TV-Type Image Sensors, 1970.

27. R. Mersereau, "Recent Advances in the Theory of Reconstructing Multidimensional Signals from Projections", Q.P.R., No. 105, RLE, MIT, April, 1972.

28. M. Nagel (ed.), Seminar of the Evaluation of Motion-Degraded Images, NASA/ERC, December, 1968, Cambridge, Mass.

29. A. V. Oppenheim, "Superposition in a Class of Nonlinear Systems", Sc.D. Thesis, MIT, 1964.

30. A. V. Oppenheim, R. W. Schaefer, and T. G. Stockham, "Nonlinear Filtering of Multiplied and Convolved Signals", Proc. IEEE, vol. 56, August, 1968, p. 1264.

31. A. V. Oppenheim and C. J. Weinstein, "Effects of Finite Register Length in Digital Filtering and the Fast Fourier Transform", Proc. IEEE, vol. 60, August, 1972, pp. 957-977.

32. A. Papoulis, "Approximations of Point Spreads for Deconvolution", J.O.S.A., vol. 62, January, 1972, pp. 77-80.

33. C. Rader and B. Gold, Digital Signal Processing, McGraw-Hill Co., New York, 1969.

34. R. D. Remington and M. A. Shork, Statistics with Applications to the Biological and Health Sciences, Prentice Hall, Englewood Cliffs, N.J., Chapters 8 and 9, 1970.

35. C. L. Rino, "Bandlimited Image Restoration by Linear Mean Square Estimation", J.O.S.A., vol. 59, May, 1969, p. 547.

36. P. G. Roetling, E. A. Trabka and R. E. Kinzly, "Theoretical Prediction of Image Quality", J.O.S.A., vol. 58, March, 1968, p. 342.

37. C. K. Rushforth and R. W. Harris, "Restoration, Resolution and Noise, J.O.S.A., vol. 58, April, 1968, p. 539.

38. R. W. Schaefer, "Echo Removal by Discrete Generalized Linear Filtering", Ph.D. Thesis, MIT, Cambridge, Mass., February, 1968.

39. W. F. Schreiber, "Picture Coding", Proc. IEEE, vol. 55, March, 1967.

40. F. Scott and R. E. Hufnagel, "Search for a Summary Measure of Image Quality", Annual Meeting of O.S.A., October, 1965.

41. R. Shack, "Influence of Image Motion and Shutter Operation on Photographic Transfer Function", J.O.S.A. vol. 54, March, 1964, p. 1171.

42. H. F. Silverman, "Identification of Linear Systems Using Fast Fourier Transform Techniques", Ph.D. Thesis, Brown Univ., 1971.

43. D. Slepian, "Linear Least Squares Filtering of Distorted Images", J.O.S.A., vol. 57, July, 1967, p. 918.

44. D. Slepian, "Restoration of Photographs Blurred by Image Motion, B.S.T.J., vol. 49, December, 1970, p. 2353.

45. M. M. Sondi, "Image Restoration: The Removal of Spatially Invariant Degradations", Proc. IEEE, vol. 60, July, 1972, pp. 842-853.

46. R. A. Sprague and B. J. Thompson, "Quantitative Visualization of Large Variation Phase Objects", App. Optics, vol. 11, July, 1972, pp. 1469-1477.

47. W. H. Steel and J. Y. Ward, "Incomplete Bessel and Sturve Functions", Proc. Camb. Phil. Soc., vol. 52, 1956, pp. 431-441.

48. W. H. Steel, "The Defocused Image of Sinusoidal Gratings", Optica Acta, vol. 3, June, 1956, pp. 65-73.

49. T. G. Stockham, 6.611 Lecture at MIT, April, 1970.

50. C. A. Taylor and H. Lipson, Optical Transforms - Their Preparation and Application to X-ray Diffraction Problems, Cornell Univ. Press, Ithaca, New York, 1965.

51. O. J. Tretiak, M. Eden, "Internal Structure from X-ray Images", Proc. of the 8th International Conference on Medical and Biological Engineering, 1969.

APPENDIX A

In this appendix, the derivation leading to Eqns.
(7.15) and (7.16) are presented. The theoretical background
for this material may be found in Davenport and Root (6)
or Deutsch (7). Given

$$y = PV(x) = g(x), \tag{A.1}$$

we seek to determine the autocorrelation of y in terms of
the autocorrelation of x. The function $g(x)$ is shown in
Fig. A.1. Thus,

$$\phi_y = \iint g(x_1)g(x_2)p(x_1,x_2) \, dx_1 \, dx_2 \tag{A.2}$$

where $p(x_1,x_2)$ denotes the joint probability density function
of x_1 and x_2, and

$$\begin{aligned} g(x_1) &= g\left[x(u_1)\right] \\ g(x_2) &= g\left[x(u_2)\right] \end{aligned} \tag{A.3}$$

Now, $g(x_1)$ and $g(x_2)$ can be expressed in terms of their
Fourier transforms as

$$y = \iint p(x_1,x_2) \left[\int G(f_1)e^{j2\pi f_1 x_1} \, df_1\right] \cdot$$
$$\left[\int G(f_2)e^{j2\pi f_2 x_2} \, df_2\right] dx_1 \, dx_2 \tag{A.4}$$

$$= \iint G(f_1)G(f_2) \left[\iint p(x_1,x_2)e^{j2\pi(f_1x_1+f_2x_2)}dx_1 \, dx_2 \right] \cdot$$

$$df_1 \, df_2 \qquad\qquad (A.5)$$

$$= \iint G(f_1)G(f_2) \, M_x(f_1,f_2) \, df_1 \, df_2 \qquad\qquad (A.6)$$

where $M_x(f_1,f_2)$ is the joint, moment-generating function for x_1 and x_2. Now, since $g(x)$ is a periodic function, its
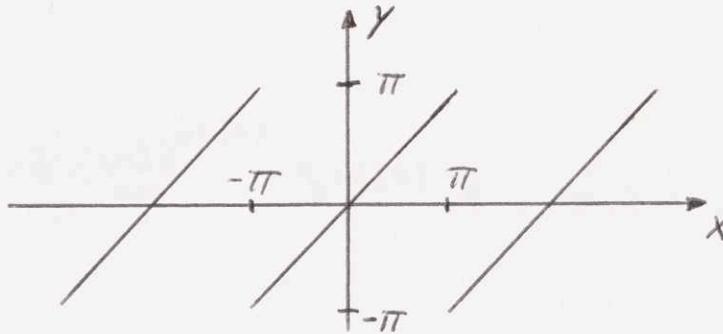

Figure A.1

transform, $G(f)$, will consist of only discrete frequencies. From Fig. A.1, one can easily show that

$$G(f) = \begin{cases} \displaystyle\sum_{n\neq o} \frac{(-1)^n}{jn} \, u_0(f-n) & n \neq 0 \\[4mm] 0 & n = 0 \end{cases} \qquad (A.7)$$

Furthermore, assume that x is the sum of two independent, zero-mean, Gaussian random processes. That is,

$$x = s + n, \qquad\qquad (A.8)$$

from which

$$\phi_x = \phi_s + \phi_n$$

The Gaussian property allows one to write M_x as

$$M_x(f_1, f_2) = \exp\left\{-\tfrac{1}{2}\left[\phi_x(0)f_1^2 + 2\phi_x(u)f_1f_2 + \phi_x(0)f_2^2\right]\right\} \quad (A.9)$$

Substituting these results into Eqn. (A.6), and using the sifting property of the impulse functions in Eqn. (A.7), one obtains

$$\phi_y(u) = \sum_{n}\sum_{m}\frac{(-1)^{n+m+1}}{nm} M_x(n,m) \quad (A.10)$$

$$\phi_y(u) = \sum_{\substack{n \\ \neq 0}}\sum_{\substack{m \\ \neq 0}}\frac{(-1)^{n+m+1}}{nm}\exp\left\{-\tfrac{1}{2}\left[\phi_x(0)(n^2 + m^2) + \right.\right. \quad (A.11)$$

$$\left.\left. 2\phi_x(u)nm\right]\right\}$$

Eqn. (A.11) offers a convenient check on the validity of the assumptions made thus far in the derivation. From other con- siderations one expects that y will be uniformly distributed between the values of $\pm\pi$. This information allows the variance to be computed as $\sigma_y^2 = \pi^2/3$ and compared against the value obtained from Eqn. (A.11). From Eqn. (A.11), we have

143

$$\sigma_y^2 = \phi_y(0) = \sum_{\substack{nm \\ \neq 0}} \frac{(-1)^{n+m+1}}{nm} \exp\left[-\tfrac{1}{2}\phi_x(0)(n+m)^2\right] \quad \text{(A.12)}$$

Retaining only those terms for which $n = -m$, $n = -m+1$, and $n = -m-1$, the double summation reduces to

$$\sigma_y^2 \approx \sum_{m \neq 0} \frac{1}{m^2} - e^{\frac{-\phi_x(0)}{2}} \left[1 + 4 \sum_{m=2}^{\infty} \frac{1}{m^2-1}\right] \quad \text{(A.13)}$$

Using the fact that

$$\sum_{m=1}^{\infty} \frac{1}{m^2} = \frac{\pi^2}{6} \qquad \text{and} \qquad \sum_{m=2}^{\infty} \frac{1}{m^2-1} = \frac{3}{4} \quad \text{(A.14)}$$

Eqn. (A.13) becomes

$$\sigma_y^2 \approx \frac{\pi^2}{3} - 4e^{\frac{-\phi_x(0)}{2}} \quad \text{(A.15)}$$

Assume $\phi_x(0) = (2\pi)^2$, in accordance with the sample standard derivation computed in Chapter V. In that case, σ_y^2 becomes

$$\sigma_y^2 \approx \frac{\pi^2}{3} - 4e^{-20} \quad \text{(A.16)}$$

which is amazingly close to the true value of $\pi^2/3$, considering all of the approximations which led to Eqn. (A.16).

The final step in the derivation is to expand the exponential in a Taylor series. This yields

$$\phi_y(u) = \sum_{k=0}^{\infty} \sum_{\substack{nm \\ \neq 0}} \frac{(-1)^{n+m+1}}{nm} e^{\frac{-\phi_x(0)(n^2+m^2)}{2}} \cdot$$

$$\frac{(-nm)^k}{k!} \phi_x^k(u) \qquad\qquad\qquad\text{(A.17)}$$

$$= \sum_{k=0}^{\infty} c_k \phi_x^k(u)$$

where

$$c_k = \frac{(-1)^{k-1}}{k!} \left[\sum_{n\neq 0} (-1)^n n^{k-1} \exp(-\tfrac{1}{2}\phi_x(0)n^2) \right]^2 \qquad\text{(A.18)}$$

Note that for k even, the sum goes to zero due to its odd symmetry, so that c_k can be further simplified to

$$c_k = \begin{cases} \dfrac{4}{k!} \left[\displaystyle\sum_{n\neq 0} (-1)^n n^{k-1} \exp(-\tfrac{1}{2}\phi_x(0)n^2) \right]^2 & k\ \text{odd} \\[20pt] 0 & k\ \text{even} \end{cases}$$

$$\text{(A.19)}$$

BIOGRAPHICAL NOTE

Anthony Edmund Filip received the B.S. and M.S. degree from the Electrical Engineering Department of Wayne State University in 1967 and 1968 respectively. While at WSU, he was twice honored as having the highest honor point average in the sophomore and senior classes. Upon coming to MIT in 1968, he became a member of the Research Laboratory of Electronics and conducted research in the fields of image processing, facsimile transmission, and digital holography, culminating in the Ph.D. degree.

In addition to several scholarships received during his undergraduate career, his graduate education was financed in part by an NDEA Fellowship (1967-68), an NSF Traineeship (1968-69), and a Sloan Fellowship (1972). Other funds were obtained through teaching assignments (1969-71) and working as a research assistant (1971-72).