PHASE RETRIEVAL WITH APPLICATION

TO INTENSITY CORRELATION INTERFEROMETERS

A Thesis

by

RUSSELL EDWARD TRAHAN III

Submitted to the Office of Graduate Studies of
Texas A&M University
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee:	David Hyland
Committee Members:	Suman Chakravorty
	Alexey Belyanin
Head of Department:	Rodney Bowersox

December 2012

Major Subject: Aerospace Engineering

ABSTRACT

As astronomers and astrophysicists seek to view ever-increasingly distant celestial objects, the desired angular resolution of telescopes is constantly being increased. Classical optics, however, has shown a proportional relationship between the size of an optical telescope and the possible angular resolution. Experience has also shown that prohibitive cost accompanies large optical systems. With these limitations on classical optical systems and with the drastic increase in computational power over the past decade, intensity correlation interferometry (ICI) has seen renewed interest since the 1950's and 60's when it was initially conceived by Hanbury Brown and Twiss. Intensity correlation interferometry has the advantage of less stringent equipment precision and less equipment cost when compared to most other forms of interferometry. ICI is thus attractive as a solution to the desire for high angular resolution imaging especially in space based imaging systems.

Optical interferometry works by gathering information about the Fourier transform of the geometry of an optical source. An ICI system, however, can only detect the magnitude of the Fourier components. The phase of the Fourier components must be recovered through some computational means and typically some *a priori* knowledge of the optical source.

This thesis gives the physics and mathematical basis of the intensity correlation interferometer. Since the ICI system cannot detect the phase of an optical source's Fourier transform, some known methods for recovering the phase information are

discussed. The primary method of interest here is the error-reduction algorithm by Gerchberg-Saxton which was adapted by Fienup to phase retrieval. This algorithm works by using known qualities of the image as constraints; however, sometimes it can be difficult to know what these constraints are supposed to be. A method of adaptively discovering these constraints is presented, and its performance is evaluated in the presence of noise. Additionally, an algorithm is presented to adapt to the presence of noise in the Fourier modulus data. Finally, the effects of the initial condition of the error-reduction algorithm are shown and a method of mitigating its effect by averaging several independent solutions together is shown.

DEDICATION

I would like to dedicate this thesis and my masters work to my grandfathers, Russell Trahan Sr. (1919-2012) and Bernard Callegan (1924-2012), who have passed during the time I've been working on my master's degree. I credit them with the ingenuity, curiosity, and motivation that have made my education successful. Without their loving influence I would not have had the opportunity to be where I am today.

ACKNOWLEDGEMENTS

I would like to thank Dr. David Hyland for his support since my senior year of undergraduate studies and through my graduate studies. His support, guidance, and encouragement are much appreciated. I also thank Dr. Suman Chakravorty and Dr. Alexey Belyanin for their time participating in my master's committee. I would also like to thank the gentlemen at the Air Force Research Laboratory who I have had the opportunity to work with. I thank the AFRL- Consortium for Autonomous Space Systems for funding my past two years of graduate study. I would also like to sincerely thank Dr. Hyland's students, namely Greg Kelderman, Daniel Fitch, and Micaela Landivar, for our unending discussions that have been priceless in our studies and research.

NOMENCLATURE

$\langle ... \rangle$	Ensemble average operator
$\langle ... \rangle_T$	Time average operator
D	Aperture diameter
E	Electromagnetic field
γ	Mutual coherence magnitude
I	Electromagnetic field intensity
ICI	Intensity correlation interferometry
J	Mutual coherence
λ	Light wavelength
MTF	Modulation transfer function
O	Observation plane
OTF, $\hat{O}(...)$	Optical transfer function
ω	Angular frequency of light
φ	Phase of a complex value
RMS	Root Mean Squared
S	Far field source object's plane
SNR	Signal-to-noise ratio
θ	Angular view coordinate
UV	Fourier domain coordinates

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

1. INTRODUCTION

An interferometer is an apparatus that measures the coherence between two points in a wave field. The knowledge of the wave field's coherence at the two spatial locations can yield knowledge about the source of the field. When the physics of wave propagation are known and the interference pattern of a wave field is known, the geometry of the wave field's source can be determined with great accuracy.

Interferometry has evolved over the past three centuries from the classical double-slit experiment to optical interferometers such the Michelson interferometer. The object of study here is the intensity correlation interferometer (ICI) as developed by Hanbury Brown and Twiss in the 1950s [1, 2, 3]. The ICI has several advantages over traditional optical interferometers; however, it does not capture the phase of the complex valued interference pattern. Rather, it only captures the magnitude. This leads to a need to estimate or retrieve the phase information—the so called 'phase retrieval' problem. The phase retrieval problem has been studied by physicists and mathematicians for various applications in astronomy, crystallography, and electron microscopy to name a few. Here the phase retrieval problem is applied to forming an image of astronomical bodies.

In section 2 the physics behind wave propagation and interference is presented. These fundamentals are used to derive the theory behind a general interferometer. The relationship between the geometry of a light source and its Fourier transform are discussed in section 2.3. The reason interferometry is attractive over traditional optics is

discussed in section 2.4. Section 2.4 uses the fundamentals explained previously and applies them to the intensity correlation interferometer. Some conventional methods for phase retrieval and several fields of study where phase retrieval is used are discussed in Section 3. Section 4 presents the collapsing foreground algorithm, the relaxed constraint algorithm, and the average image algorithm for phase retrieval. These algorithms build on the methods presented in section 3. Section 5 shows the performance of these three algorithms and attempts to qualify them as viable methods for phase retrieval. Some methods for programing these algorithms to decrease their computation time are also briefly discussed. Finally, Section 6 concludes by summarizing the phase retrieval problem and the algorithms presented in this thesis.

2. INTERFEROMETRY

Interferometry owes its existence to the wave properties of light. The wave theory of light was developed through the observation of fringe patterns seen when interfering light from two sources. Some of the first observations and theory were developed by Hooke, Huygens, and Frensel [4]. In 1665 Hooke observed wave interference phenomena in soap bubbles and oil on water; however, he was not able to definitively explain his observations in his book Micrographia [5]. Hooke first observed 'Newton's rings' which are simple interference patterns visible when a convex lens is placed next to a flat surface; however, Newton's explanations followed the flawed corpuscular theory (particle theory) of light [6]. In 1678, Huygens proposed the wave theory of light. His theory includes the statement that every point illuminated by light becomes a point source. Fresnel's work built on Huygens's work and made some big contributions to the understanding of light by using the wave model to explain the laws of reflection, refraction, and double refraction. Their combined work, later called the Huygens-Fresnel principle, rigorously explained the wave nature of light. It was Thomas Young who experimentally captured the wave nature of light in his famous double-slit experiment which he presented to the Royal Society of London in 1801 [7]. He observed that the interference pattern of light from a source corresponds to the geometry of the source. In the 1930s it was found that taking the inverse Fourier transform of the interference pattern can recover the geometry of the light source according to the van Cittert-Zernike theorem thus forming the basis of modern interferometry [8, 9].

2.1. Double-Slit Experiment

Thomas Young developed an experiment to demonstrate that light behaves indisputably as a wave and not solely as a particle as was the belief in his era. His experiment consisted of a light source, two walls with small opening for the light to pass through, and a solid plane to observe the light after passing through the slits [7]. This comprised the first and simplest interferometer.

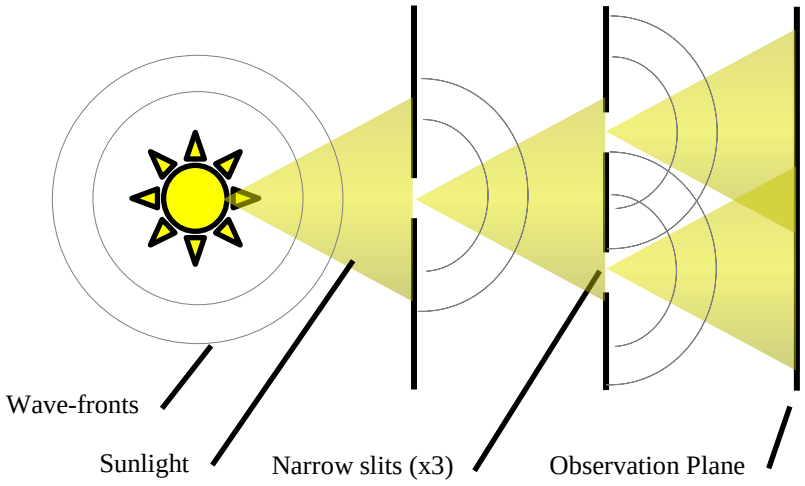


Fig. 1. Schematic of Thomas Young's double-slit experiment.

The purpose of the first single slit is to arrange the waves of light from the sun into wave-fronts, i.e. the light appears to come from a single point source located at the first slit as explained by Huygens's theory of light [10]. These wave-fronts then hit the second wall with the double slits. The waves then emanate from the two slits and hit the observation plane. Again, the second slits behave as point sources of light. Looking at the geometry of this experiment in Fig. 1, no light would be visible in the center of the

observation plane if light behaves according to Newton's particle theory of light. The particle theory states that the light would travel in straight lines; however, Young saw an interference pattern across the observation plane. The details of the interference pattern will be explained further in section 2.2.

A simplistic explanation of the double-slit experiment from the particle view can be developed with a statistical representation of photons [6]. A single photon going through the first slit has equal probability of going through either slit in the second wall. Max Born's 1954 Nobel Prize work described the state of the interferometer at a point in time as the probability of finding a photon at a given position. The photon must then not be considered as a particle with some location, but rather as a particle with probability of being somewhere in space. With this concept in mind, the photon can be thought of as having equal probability of being in two places at the same time. The probabilistic field that describes the position of the photon allows the photon to essentially interfere with itself by going through both slits. As subsequent photons go through the interferometer, an interference pattern emerges even though only one photon goes through the slits at a time. A caveat in proving this concept experimentally lies in trying to track a single photon through the interferometer. The probability field concept breaks down, because only a single realization of the probability field is detected. This breakdown is described by the Heisenberg-uncertainty principle; by knowing that the photon exists, i.e. detecting its momentum, its location is unknown. For this reason, it is more convenient to consider the wave concept of light and not think in terms of photons.

Young's experiment, observations, and explanations have had a profound effect on how light is understood. There is much more to say about Young's experiment, but here it suffices to explain his experimental setup and to state his conclusion that light indeed behaves as a wave.

2.2. Wave Interference & Propagation

The simplest case of wave interference to consider is that of two monochromatic waves from point sources as is the case in Thomas Young's double-slit experiment. Additionally, the waves are polarized in the same plane reducing the problem to two dimensions. The electromagnetic field waveform at a given point for the waves is described by

$$\begin{aligned} E_1 &= e_1 \exp\left(-i\phi_1\right) \\ E_2 &= e_2 \exp\left(-i\phi_2\right) \end{aligned} \tag{2.1}$$

where e_i and φ_i are the amplitudes and phases respectively [11]. The total intensity at this point of interest is

$$\begin{aligned} I &= \left|E_1 + E_2\right|^2 \\ &= \left|E_1\right|^2 + \left|E_2\right|^2 + E_1 E_2^* + E_1^* E_2 \\ &= I_1 + I_2 + 2\left(I_1 I_2\right)^{1/2} \cos\left(\phi_1 - \phi_2\right). \end{aligned} \tag{2.2}$$

In equation (2.2) the I_1 and I_2 terms denote the intensity of each wave considered independently at the point of interest [9, 12]. The intensity I is a function of the phase difference between the two waves, $\varphi_1 - \varphi_2$, which is a function of the distance from the source and time [13]. In practicality, light is never truly monochromatic nor emitted from a point source, so the visibility of the interference is defined as [10]

$$\nu = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}. \tag{2.3}$$

The maximum and minimum phase differences give the maximum and minimum total intensities respectively according to

$$I_{max} = I_1 + I_2 + 2\left(I_1 I_2\right)^{1/2} \tag{2.4}$$

and

$$I_{min} = I_1 + I_2 - 2\left(I_1 I_2\right)^{1/2}. \tag{2.5}$$

These two values, the maximum and minimum total intensity, refer to constructive and destructive interference, respectively as graphically shown in Fig. 2. The quality of interference in this example is unity. This means the constructive interference (a) results in double the amplitude of the contributing waves and the destructive interference (b) results in zero intensity. These wave interference concepts can be expanded to include multiple sources and even continuous (non-point) sources. For illustration, a field can be plotted showing the waveforms.

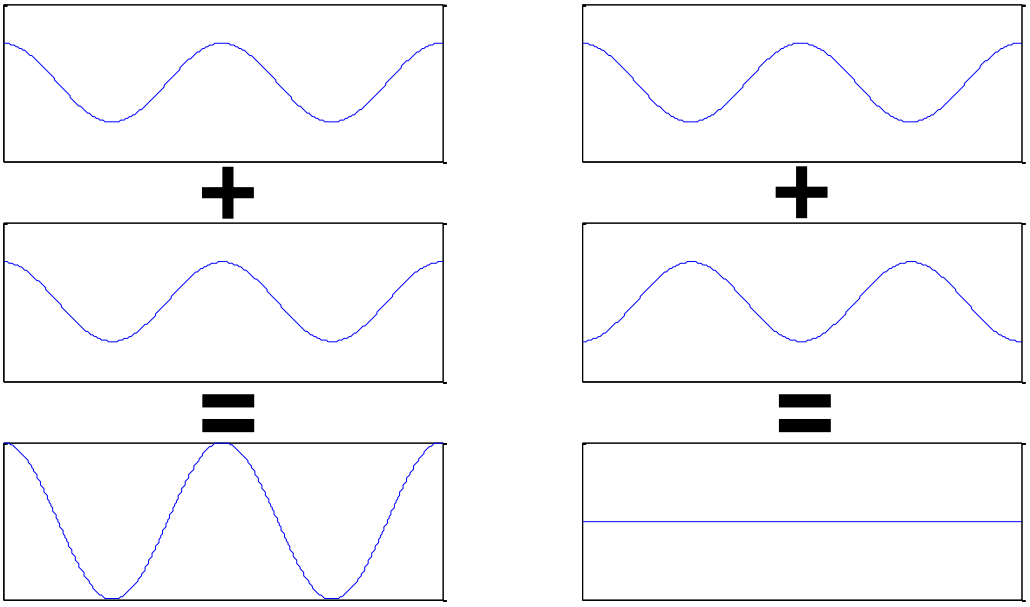


Fig. 2. Examples of (a) constructive interference and (b) destructive Interference.

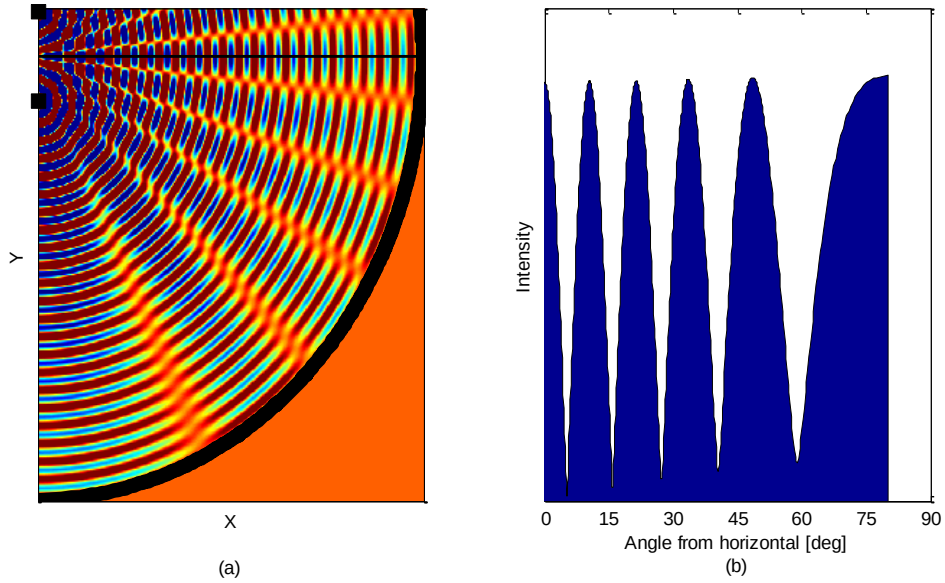


Fig. 3. (a) Wave propagation from two point sources (top left). (b) Interference pattern at the circular boundary in (a). The wave field is cropped at the top due to symmetry.

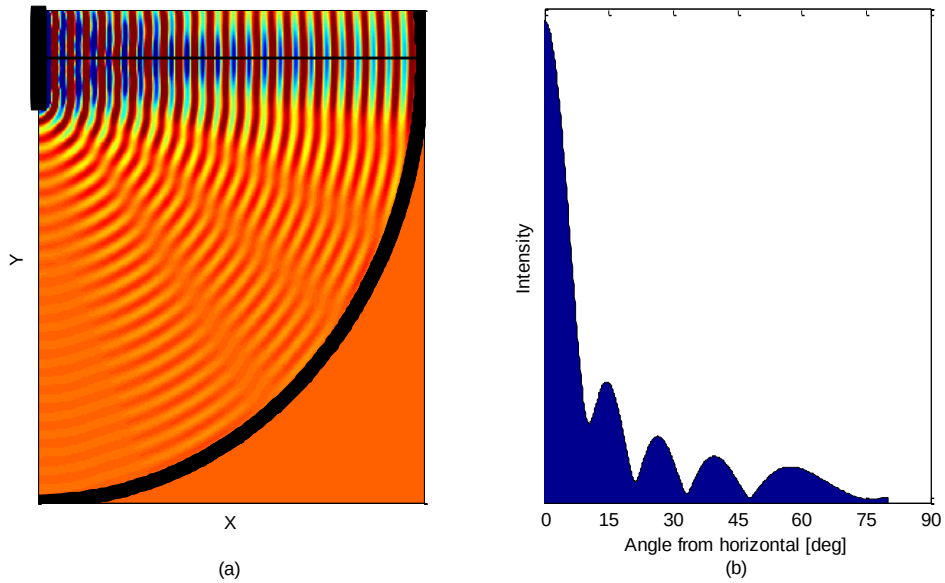


Fig. 4. (a) Wave propagation from a continuous source (top left). (b) Interference pattern at the circular boundary in (a).

Fig. 3a shows the wave field surrounding two point sources in the upper-left corner as is the case in Thomas Young's double-slit experiment. Fig. 3b shows the time average of the interference intensity at the circular boundary in Fig. 3a as a function of the angular position relative to horizontal. Over the averaging time, some regions see no intensity of light, i.e. complete destructive interference. Other regions see large amplitudes from the constructive interference between the waves. Fig. 4a shows the wave field surrounding a continuous source in the upper-left corner. The continuous source is simply several discrete point sources adjacent to each other. The interference pattern is similar to that of the two point sources but has a decaying component.

If the radius of the viewing circle is sufficiently large and only the region of the viewing circle near the horizontal line in the center of the light sources is considered, the interference pattern assumes the shape of the Fourier transform of the source according to the Huygens-Fresnel Principle [13].

2.2.1. 1-D Fourier transforms

The light source in Fig. 4a is a rectangle function over the y-coordinate of the plot. Explicitly the wave-source amplitude is

$$a(y) = rect\left(\frac{y}{L}\right) = \begin{cases} 1, & 0 < |y| < L/2 \\ 0, & L/2 < |y| < \infty \end{cases} \tag{2.6}$$

where L is the length of the wave source [14]. The Fourier transform of the rectangle function [13] is

$$\begin{aligned}\int_{-\infty}^{\infty} a(y)\, e^{-i2\pi fyL}\, dy &= L\int_{-\infty}^{\infty} \mathrm{rect}\left(\frac{y}{L}\right) e^{-i2\pi fy'L}\, dy' \\ &= L\int_{-1/2}^{1/2} e^{-i2\pi fy'L}\, dy' \\ &= L\,\frac{\sin(\pi fL)}{\pi fL} \\ &= L\,\mathrm{sinc}(fL)\end{aligned} \tag{2.7}$$

which is proportional to the sinc function[1]. The form of these functions is shown graphically in Fig. 5.

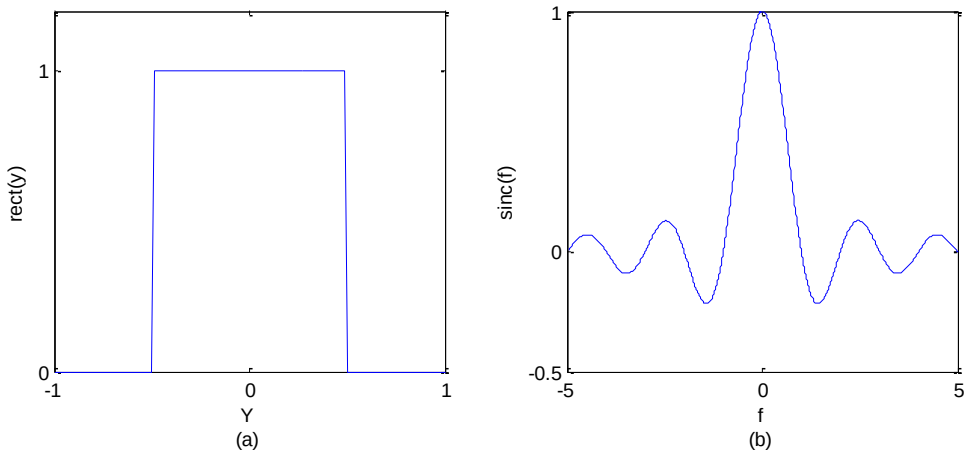


Fig. 5. (a) The rectangle function. (b) The *sinc* function.

The magnitude of the sinc function represents the interference pattern of a slit when viewed from far away from the source [13]. Compare the interference pattern in Fig. 4 with Fig. 5. This concept can be generalized for any source even in two dimensions.

[1] Here the sinc function is defined as $\mathrm{sinc}(x) = \sin(\pi x)/\pi x$; however, some texts define it as $\mathrm{sinc}(x) = \sin(x)/x$.

2.2.2. 2-D Fourier transforms

In the one dimensional case, the source could be represented in a functional form, and the interference pattern was represented by the Fourier transform. A two-dimensional source can be transformed to a frequency domain in the same way [12]. A simple case to illustrate a 2d Fourier transform is a simple rectangular source

$$f(x,y) = \begin{cases} 1, & -a/2 < x < -a/2 \;\&\; -b/2 < y < b/2 \\ 0, & \text{otherwise.} \end{cases} \tag{2.8}$$

In general the two-dimensional Fourier transform has the form [15]

$$F(u,v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\, e^{-i2\pi(ux+vy)} dx dy \tag{2.9}$$

and the inverse

$$f(x,y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u,v)\, e^{i2\pi(ux+vy)} du dv. \tag{2.10}$$

This standard form of the two-dimensional Fourier transform denotes the frequency domain as the UV plane and the Euclidian domain as the XY plane. Evaluating the case proposed in equation (2.8) gives the Fourier transform

$$\begin{aligned} F(u,v) &= -\frac{1}{\pi^2 uv}\left(e^{i\pi(ua+vb)} + e^{-i\pi(ua+vb)} - e^{-i\pi(ua-vb)} - e^{i2\pi(ua-vb)}\right) \\ &= -\frac{1}{\pi^2 uv}\left(2\cos(\pi(ua+vb)) - 2\cos(\pi(ua-vb))\right) \\ &= \frac{\sin(\pi ua)}{\pi u}\frac{\sin(\pi vb)}{\pi v} \\ &= ab\,\mathrm{sinc}(\pi ua)\,\mathrm{sinc}(\pi vb). \end{aligned} \tag{2.11}$$

The Fourier transform of the rectangle is a convolution of sinc functions in the U and V directions with the periodicity scaled by the size of the rectangle.

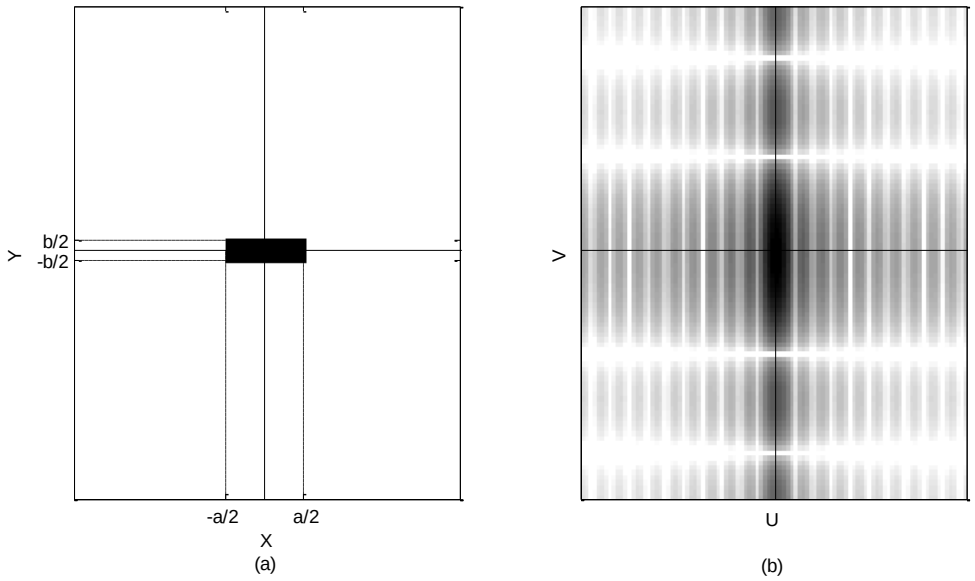


Fig. 6. (a) Rectangular source. (b) Fourier transform magnitude of a rectangular source.

Fig. 6a shows the rectangle image from equation (2.8) with $a = 20$ and $b = 5$. Fig. 6b shows the magnitude of the Fourier transform which includes the two sinc waveforms in equation (2.11). Note how the length scales in the source's X and Y directions are inverted in the Fourier transform's U and V directions respectively. Essentially, the small vertical dimension of the rectangle leads to a long wavelength sinc function and vice-versa for the horizontal rectangle dimension.

2.3. Imaging and the Fourier Domain

It has been shown graphically in section 2.2 that the Fourier transform of a light source is similar to the interference pattern of the light waves emitted by the source. Here the relationship between a source-object's Fourier transform and its interference pattern is derived. Just as the double-slit experiment revealed an interference pattern from two point sources on an observation plane, if optical apertures are placed on an

observation plane they can detect the interference pattern formed by the various point sources of light on a distant source. The interference pattern can then be inverse Fourier transformed to reveal the original source's geometry. This derivation is based on the van Cittert-Zernike theorem [9, 16].

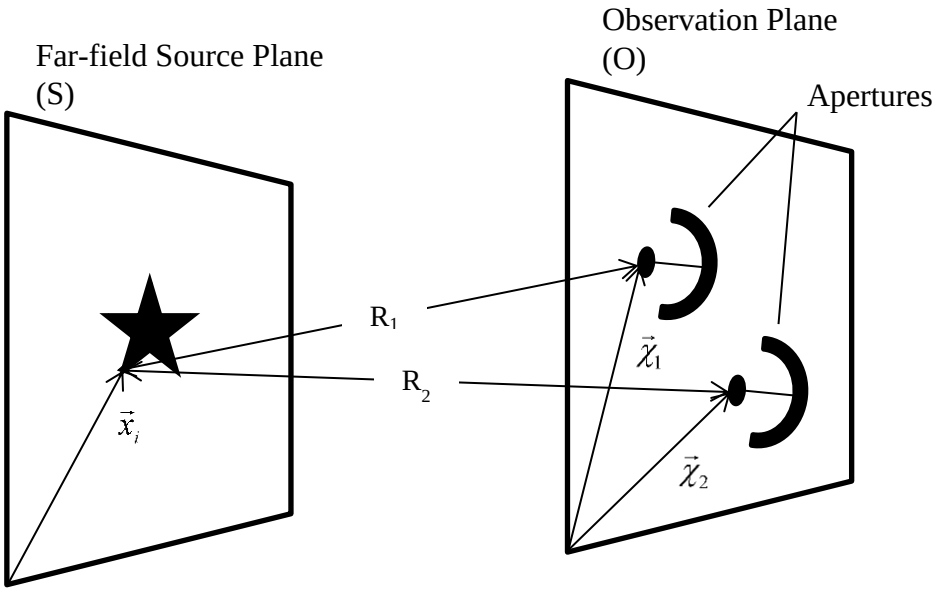


Fig. 7. Two optical apertures in an observation plane and the light source in the far-field plane. Both planes are parallel.

Shown in Fig. 7, there are light sources in the far-field plane located at $\vec{x}_i$ and two apertures located in the observation plane at positions $\vec{\chi}_1$ and $\vec{\chi}_2$. The electric field at $\vec{\chi}_1$ due to the source at $\vec{x}_i$ is

$$E_1\left(\vec{x}_i, t\right) = A\left(\vec{x}_i, t - \frac{R_1}{c}\right)\frac{1}{R_i}\exp\left(-\frac{i2\pi}{\lambda}\left(t - \frac{R_1}{c}\right)\right) \tag{2.12}$$

where A is the strength of the field, c is the speed of light, R_1 is the distance from the source to aperture 1, and λ is the average wavelength of light considered[2]. An analogous equation exists for aperture 2. The time average of the product of the field from the same source at the two apertures is

$$\left\langle E_1\left(\vec{x}_1,t\right)E_2^*\left(\vec{x}_1,t\right)\right\rangle$$
$$=\left\langle A\left(\vec{x}_1,t-\frac{R_1}{c}\right)A\left(\vec{x}_1,t-\frac{R_2}{c}\right)\right\rangle\frac{\exp\left(-\frac{i2\pi}{\lambda}\left(t-\frac{R_1}{c}\right)\right)}{R_1}\frac{\exp\left(\frac{i2\pi}{\lambda}\left(t-\frac{R_2}{c}\right)\right)}{R_2}. \quad (2.13)$$

Time can be shifted and terms rearranged such that

$$\left\langle E_1\left(\vec{x}_1,t\right)E_2^*\left(\vec{x}_1,t\right)\right\rangle=\left\langle A\left(\vec{x}_1,t\right)A\left(\vec{x}_1,t-\frac{R_2-R_1}{c}\right)\right\rangle\frac{\exp\left(i2\pi\frac{R_1-R_2}{\lambda}\right)}{R_1R_2}. \quad (2.14)$$

If the distance the light travels in the measurement time interval is much larger than the difference in the aperture distances from the source, that is $(t_i-t_{i-1})c \gg \left|R_1-R_2\right|$, another simplification can be made to the amplitude terms such that

$$\left\langle E_1\left(\vec{x}_1,t\right)E_2^*\left(\vec{x}_1,t\right)\right\rangle=\left\langle A\left(\vec{x}_1,t\right)A\left(\vec{x}_1,t\right)\right\rangle\frac{\exp\left(i2\pi\frac{R_1-R_2}{\lambda}\right)}{R_1R_2}. \quad (2.15)$$

Recalling that the term $\left\langle A\left(\vec{x}_1,t\right)A\left(\vec{x}_1,t\right)\right\rangle$ is the average intensity of the field, the intensity term $I\left(\vec{x}_1\right)$ can be introduced giving

$$\left\langle E_1\left(\vec{x}_1,t\right)E_2^*\left(\vec{x}_1,t\right)\right\rangle=I\left(\vec{x}_1\right)\frac{\exp\left(i2\pi\frac{R_1-R_2}{\lambda}\right)}{R_1R_2}. \quad (2.16)$$

[2] The light entering the apertures is filtered to a narrow bandwidth. This is the quasi-monochromatic assumption.

Equation (2.16) is the coherence of the field from $\vec{x}_1$ at apertures 1 and 2. The coherence of the field from all sources across the far-field plane s is

$$J(\vec{\chi}_1, \vec{\chi}_2) = \iint_S I(\vec{x}) \frac{\exp\left(i2\pi \frac{R_1 - R_2}{\lambda}\right)}{R_1 R_2} d\vec{x}. \tag{2.17}$$

Next, define the components of $\vec{x}$ as x and y and the components of $\vec{\chi}_i$ as x_i and y_i. The Fourier wave-number plane (UV plane) is defined where each unit is the number of wavelengths between the apertures. Formally, the UV plane has components

$$\begin{aligned} u &= \frac{x_1 - x_2}{\lambda} \\ v &= \frac{y_1 - y_2}{\lambda}. \end{aligned} \tag{2.18}$$

When the observation plane is sufficiently far away from the source plane and higher order terms are neglected, it can be shown that

$$\frac{R_2 - R_1}{\lambda} = ux + vy \tag{2.19}$$

and the differential area $d\vec{x}$ in equation (2.17) can be approximated as

$$R^2 dx\, dy = d\vec{x} \tag{2.20}$$

giving the final simplification of the coherence as

$$J(u,v) = \iint_S I(x,y) \exp\left(-i2\pi (ux + vy)\right) dx\, dy. \tag{2.21}$$

The coherence is the Fourier transform of the source as a function of the baseline of the two apertures and their angular orientation relative to each other. The baseline is the

distance between the two apertures, i.e. $\left|\vec{\chi}_1 - \vec{\chi}_2\right|$. The coherence is typically plotted on a circular plot of the UV plane.

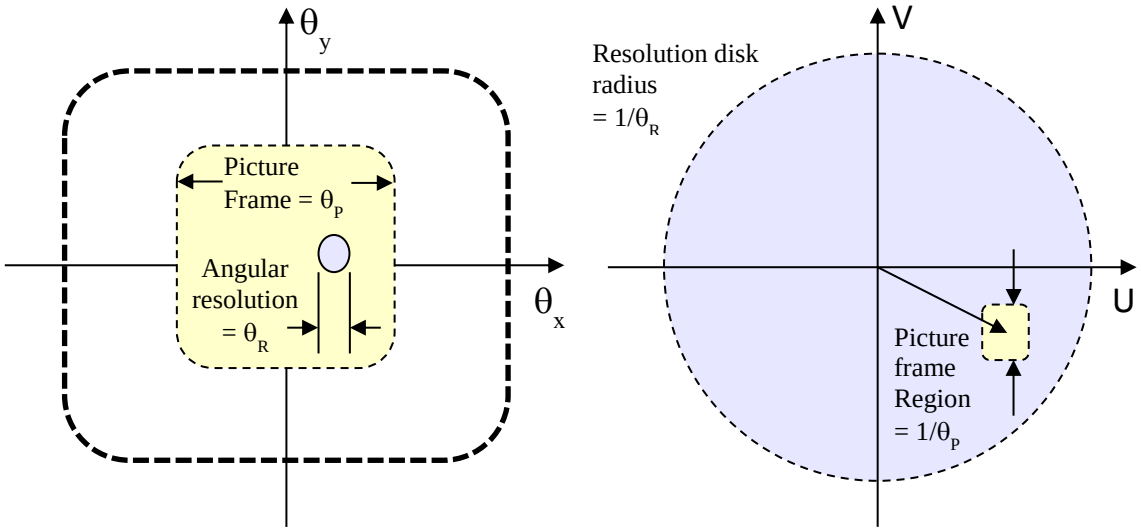


Fig. 8. Graphical representation of the θ angular spatial plane which contains the image and the Fourier UV wave number plane.

Fig. 8 shows the θ_{XY} plane which is the far-field XY plane from Fig. 7 represented in an angular view system. The angular resolution is defined as the minimum angular separation for which two objects can be resolved by an aperture as separate objects. The picture frame is the region in the image plane corresponding to the desired image, and it is again defined angularly as θ_P. If the two apertures are positioned at $\vec{\chi}_m$ and $\vec{\chi}_n$ in the observation plane, this represents the Fourier component in the neighborhood of

$$\vec{J} = \frac{\vec{\chi}_m - \vec{\chi}_n}{\lambda} \tag{2.22}$$

in the Fourier UV plane. Analogously, the baseline of the apertures is proportional to the radial distance of the Fourier component in the UV plane, and the angle between the apertures determines the angle from a reference in the UV plane. Because the picture frame is finite, this neighborhood in the UV plane has finite size. By this argument the UV plane can be discretized to contain 'coverage disks' that represent unique measurements of the Fourier component at various spatial positions of the apertures.

The apertures can be repositioned repeatedly and multiple sets of apertures can be used to collect the Fourier components across the UV plane assuming the source does not change during the entire imaging process. Reconstruction of the image does not require the entire UV plane to be covered; however, complete coverage gives a better quality image.

2.4. Intensity Correlation Interferometry

The intensity correlation interferometer is based on the Hanbury Brown and Twiss effect. In their 1957 paper [3], Hanbury Brown and Twiss discussed how the correlation magnitude of the wave field at two apertures is "the intensity fluctuation due to beats between waves of different frequency." They embraced the wave nature of light which was still somewhat misunderstood due to some still holding photons as classical particles. They discussed the photon 'bunching' effect of light which is based on photon arrival times not being independent of each other. They stated that their results could be justified with Bose-Einstein statistics of photons; however, they stated that the same results were more easily understood through the classical wave nature of light. They

stated that these fluctuations in the intensity of light from a coherent source observed at two spatial locations are correlated. The found that the cross-correlation of the intensity of the fluctuations recorded at these two locations is proportional to the square of the magnitude of the optical coherence at the two locations. A disadvantage of this method is that the narrow wavelength reduces the overall intensity of the light being measured. Having such a dim intensity of light to measure can introduce signal-to-noise ratio issues. More importantly, the proportionality constant multiplying the coherence magnitude is small, so long averaging times are needed to achieve a manageable the SNR for the coherence magnitude estimate. Additionally, the ICI system cannot provide the phase of the Fourier components. The phase must be recovered through computational algorithms which is the primary topic of this thesis.

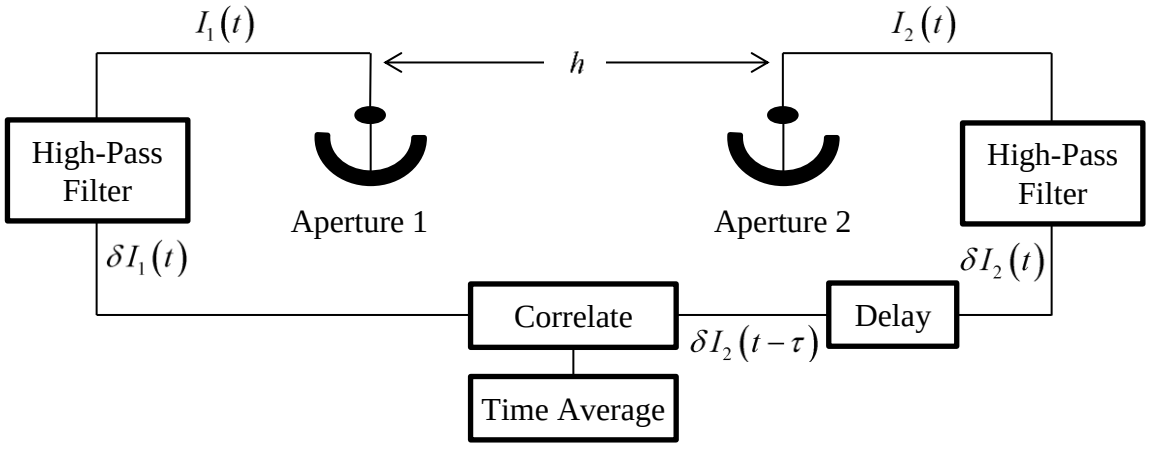


Fig. 9. Schematic of the Hanbury Brown and Twiss interferometer [2].

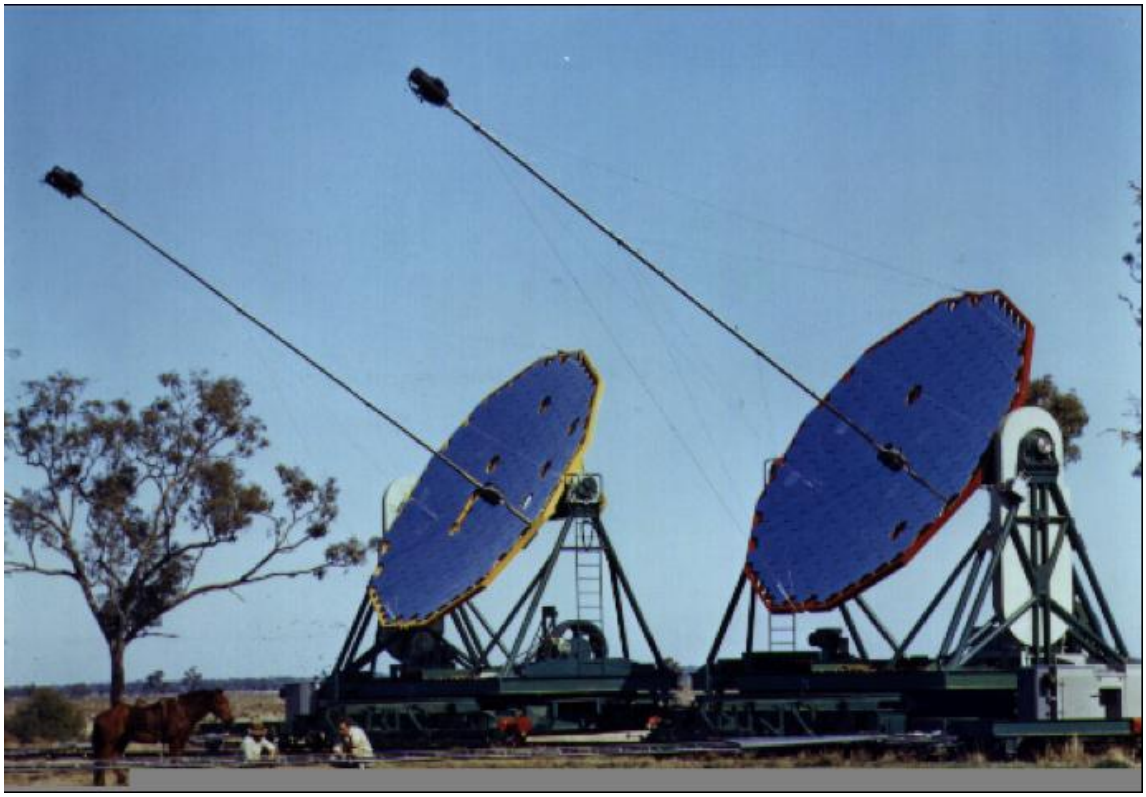


Fig. 10. The Hanbury Brown-Twiss interferometer at the Narrabri observatory. Shown are the two light collectors with photodiodes located on the center mast. The two apertures are on a circular track allowing freedom of baseline and orientation.

The interferometer designed by Hanbury Brown and Twiss is shown in Fig. 9 and Fig. 10. The basic components are two apertures with sensors that measure the total intensity of the light entering the aperture. The signal from these two photocells is high-pass filtered, delayed, correlated, and time averaged. The time averaged data is proportional to the square of the mutual coherence magnitude γ for the baseline h of the two apertures [17, 3] and is hence proportional to the square of the magnitude of the Fourier transform at the given baseline. The purpose of the high-pass filter is to remove the constant portion of the intensity measurement and only preserve the fluctuations in intensity. The normalized magnitude of the Fourier domain for the given baseline $\left|\gamma_{12}\right|^2$, the object of interest , is proportional to the SNR of the intensity,

$$\gamma_{12}^2 \propto \frac{\left\langle I_1 I_2 \right\rangle^2}{\left\langle \left(I_1 I_2 - \left\langle I_1 I_2 \right\rangle \right)^2 \right\rangle} \tag{2.23}$$

Because the signal-to-noise ratio can be quite low, the time average period in the mutual coherence determination is typically quite large to ensure convergence of the time averages. The long averaging period mandates a huge amount of data to be collected and processed. The ICI concept was not practical for many years due to these computational requirements; however, in recent years it has come of interest again because of increases in available computational power.

2.5. Applications to Astronomical Imaging

Astronomical imaging using interferometry usually takes two forms. Some applications, such as the work by Hanbury Brown and Twiss, pertain to precision measurement of the diameter of stars; however, they did not reproduce an image of the stars. They captured Fourier components until the large features of the star were evident, namely the overall circular shape of the star. Data from their measurements of Sirius are shown in Fig. 11. They captured the first oscillation of the jinc function resulting from the disk of the star. The rate of decay of the sinc function is proportional to the diameter of the star. Hanbury Brown and Twiss did not have the capability of measuring much more than the diameter of stars due to the nature of their equipment. Some more complex applications include measuring the separation between binary star systems. This application still doesn't require an entire image to be reproduced, but the Fourier components would contain many more features than a simple star disk would give.

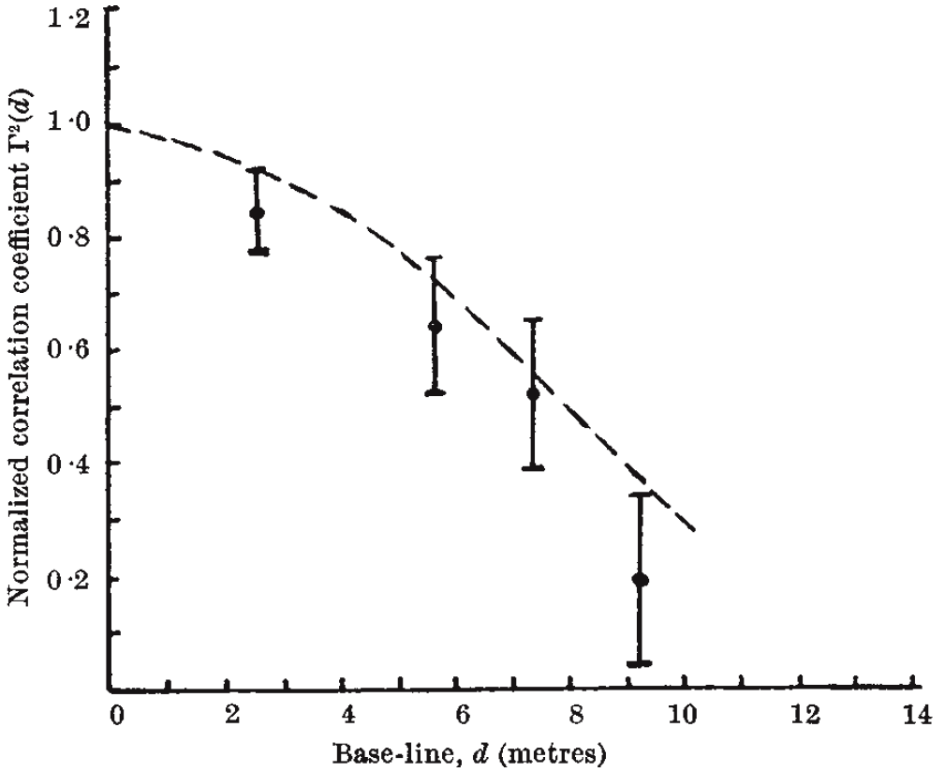


Fig. 11. Comparison between the values of the normalized correlation coefficient observed from Sirius and the theoretical values for a star of angular diameter 0.0063". The errors shown are the probable errors of the observations [2].

A more complex astronomical application of interferometry is starspot measurement and tracking. Starspots typically have a maximum diameter of 5% of the diameter of the star thus requiring resolution beyond any current optical system to resolve. Because of the complexity and interest in this application, a simplistic analysis will be given where the position and intensity of the starspots is measured. This example is adapted from the work contained in [18].

The scene of a star with radius r and N starspots can be given as

$$s(x,y) = \alpha_0 \operatorname{circ}\left(\frac{\sqrt{x^2+y^2}}{r}\right) + \sum_{k=1}^{N} \alpha_k \delta(x - x_k, y - y_k). \tag{2.24}$$

The first term is the disk of the star, the second term is a Dirac delta function approximation for each starspot, and the intensity of each feature is given by α_k. Note that the starspot α values will be negative. The delta function is appropriate to represent the starspots because the spots are assumed to be unresolvable by the apertures of the interferometer. The Fourier transform of the source is thus

$$S(u,v) = \alpha_0 r \frac{J_1\left(2\pi r\sqrt{u^2+v^2}\right)}{\sqrt{u^2+v^2}} + \sum_{k=1}^{N} \alpha_k \exp\left[-2\pi i(x_k u + y_k v)\right]. \tag{2.25}$$

The mutual coherence square magnitude is

$$
\begin{aligned}
M\left(u,v\right) = \alpha_0^2 r^2 \frac{J_1^2\left(2\pi r\sqrt{u^2+v^2}\right)}{u^2+v^2} \\
+2\alpha_0 r \frac{J_1\left(2\pi r\sqrt{u^2+v^2}\right)}{\sqrt{u^2+v^2}} \sum_{k=1}^{N} \alpha_k \cos\left(2\pi\left(x_k u + y_k v\right)\right) \\
+\sum_{k=1}^{N} \alpha_k^2 \\
+\sum_{k=1}^{N-1}\sum_{l=k+1}^{N} 2\alpha_k\alpha_l \cos\left(2\pi\left(\left(x_l - x_k\right)u + \left(y_l - y_k\right)v\right)\right).
\end{aligned}
\tag{2.26}
$$

The four terms in equation (2.26) have specific meanings. The first term is the interference due to the stellar disk. The second term is the interference between the stellar disk and the starspots. It contains the spot positions relative to the stellar disk in $x_k u + y_k v$. The tertiary term refers to the effect of the starspots on the scene, namely the dimming effect that each spot has on the overall scene. The fourth term refers to the interference between each of the spots and contains the spots positions relative to each other.

As an example, a case is considered with four starspots. The scene is normalized such that $\alpha_0 = r = 1$. This gives the star a total brightness of π. To represent a nominal starspot, the intensity is about 0.1% of the star brightness so $\alpha_k = 0.001\pi \approx 0.003$. The scene is displayed graphically but not to scale in Fig. 12.

Plotting the mutual coherence square magnitude and each of the four contributing terms as a function of the radial Fourier coordinate gives insight into the advantage of using intensity correlation interferometry for such an imaging task.

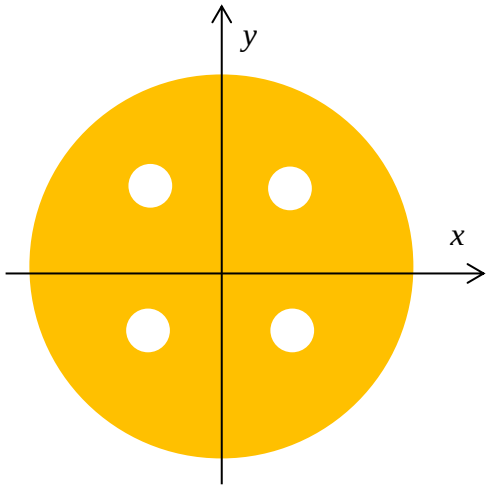


Fig. 12. Pictorial of the sample scene of a star with four starspots (not drawn to scale).

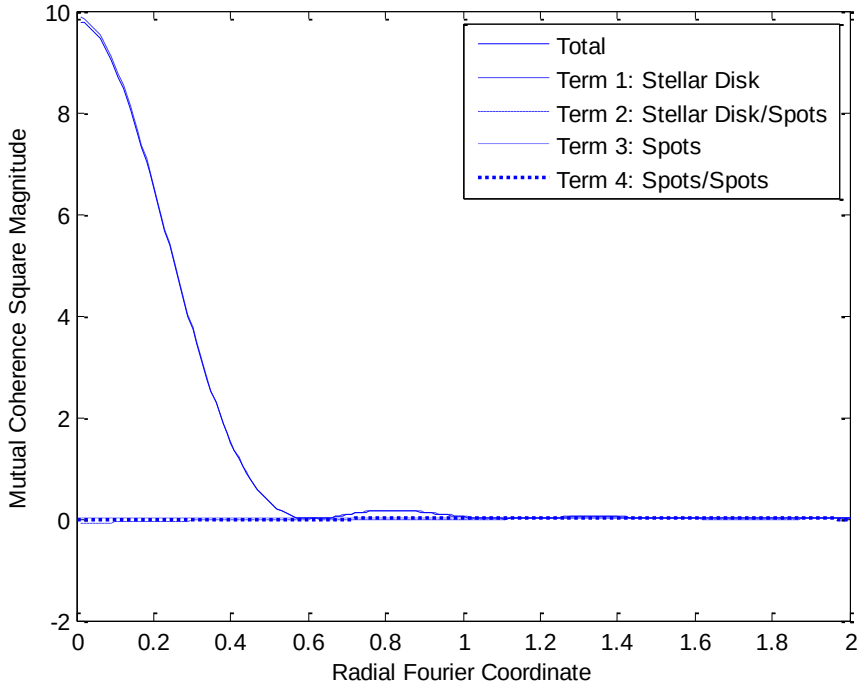


Fig. 13. Mutual coherence square magnitude and its contributing terms near the Fourier origin. Only the stellar disk term is significant.

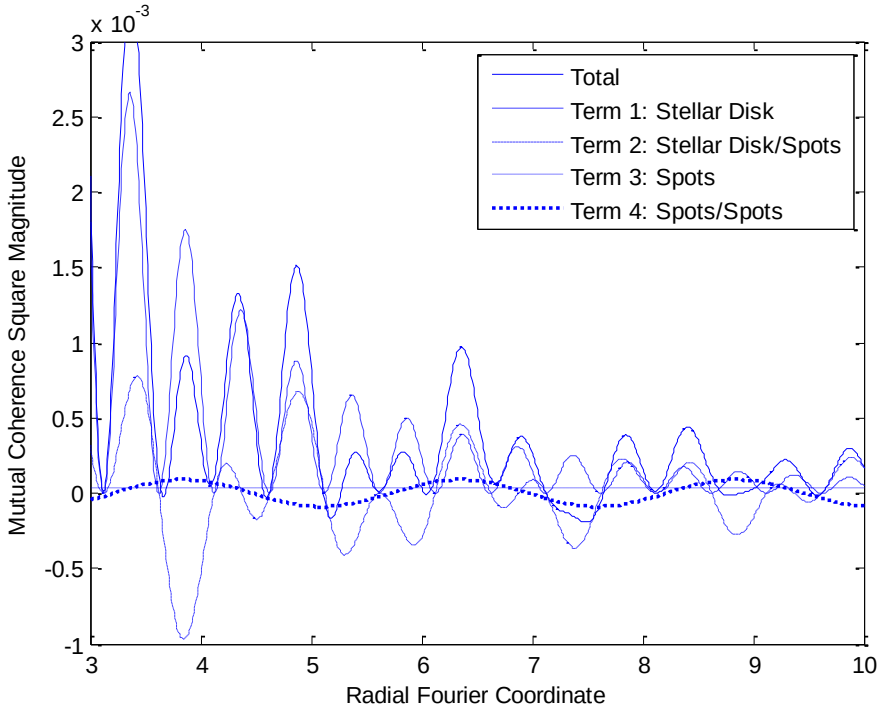


Fig. 14. Mutual coherence square magnitude and its contributing terms away from the Fourier origin.

Fig. 13 shows that for small radial Fourier coordinates the mutual coherence square magnitude is dominated by the stellar disk. The trend follows one very similar to the stellar diameter measurements done by Hanbury Brown and Twiss shown in Fig. 11. Fig. 14 shows that at larger radial Fourier coordinates the other terms begin to dominate. This implies that to gain knowledge of the position of the starspots and their intensities using an ICI for a star similar to our sun baselines must be approximately greater than $O\left(3r/\lambda\right) \approx O\left(1e5\right)$ meters, which would be impractical for a single aperture telescope. Additionally, measurements would need to be taken at intervals smaller than $O\left(0.25r/\lambda\right) \approx O\left(1e4\right)$ meters to capture the waveform. The number of measurements required to resolve information about the starspots would thus be very high. This example shows one of the primary complexities with high resolution imaging with intensity correlation interferometry; however, this insight highlights the potential of the intensity correlation interferometer when applied to astronomical imaging.

3. PHASE RETRIEVAL

Most interferometry systems correlate the electromagnetic field from two spatial locations to obtain the magnitude of the Fourier transform of the light source; however, some systems do not provide information about the phase of the complex Fourier transform values, such as the ICI. Phase retrieval is applicable to stellar interferometry as is discussed here. Additionally, two application of phase retrieval to small scale imaging are X-ray crystallography and high-resolution transmission electron microscopy. X-ray crystallography aims to map the structure of a crystal. The density of electrons is measured revealing the overall structure of the crystal and allowing inference about the molecular bonds present [19]. High-resolution transmission electron microscopy (HRTEM) is another method of observing a crystal structure at the atomic length scale. As a standard interferometer observes a light field, the HRTEM observes an electron field which is passed through a thin specimen. The atoms within the specimen essentially comprise the image of interest [20]. Phase retrieval is important in these fields because the entire UV plane is used to recreate an image. This is unlike most current astronomical interferometry applications which do not use the entire UV plane and do not aim to recreate an entire image.

There are several methods for recovering the phase information as outlined in [21] and [22]. The primary difficulty in recovering the phase that each method addresses is the large number of degrees of freedom. Every pixel in an image is a degree of

freedom, and for this reason the solutions are usually computationally intensive and rely on iterating to converge to a solution.

3.1. Exact Phase Solution

To formulate an exact solution to the phase retrieval problem, one can consider each square pixel in an $N \times N$ image to be a separate rectangle with some intensity value. This approach will lead to a convolution of the Fourier domain solution presented in section 2.2.2 for each square pixel. Convolution is possible due to the linearity of the Fourier transform, i.e. each feature in an image can be transformed separately and the results convoluted. Let the magnitude of each pixel across the (u,v) plane be $\left|A(u,v)\right|$. The convolution is thus [22]

$$\left|A(u,v)\right|\exp\left(i\varphi(u,v)\right) = \sum_{k=1}^{N}\sum_{m=1}^{N}\left|A(k,m)\right|\exp\left(i\varphi(k,m)\right)\operatorname{sinc}(u-k)\operatorname{sinc}(v-m). \tag{3.1}$$

Expressing the phase in trigonometric form and squaring both sides gives the non-linear system of equations

$$\left|A(u,v)\right|^2 = \left[\sum_{k=1}^{N}\sum_{m=1}^{N}\left|A(k,m)\right|\operatorname{sinc}(u-k)\operatorname{sinc}(v-m)\cos\left(\varphi(k,m)\right)\right]^2 + \left[\sum_{k=1}^{N}\sum_{m=1}^{N}\left|A(k,m)\right|\operatorname{sinc}(u-k)\operatorname{sinc}(v-m)\sin\left(\varphi(k,m)\right)\right]^2. \tag{3.2}$$

This system of equations has N^2 unknowns and N^2 solution sets. To constrain the system, any pixel can be constrained, i.e. $\varphi(1,1) = 0$, and a unique solution will exist. By constraining pixel $(1,1)$, equation (3.2) becomes

$$
\begin{aligned}
\left|A(u,v)\right|^2 = & \\
& \left[\sum_{k=1}^{N}{}^{*}\sum_{m=1}^{N}\left|A(k,m)\right|\operatorname{sinc}(u-k)\operatorname{sinc}(v-m)\cos(\varphi(k,m))\right. \\
& \left.+\left|A(1,1)\right|\ \operatorname{sinc}(u-1)\operatorname{sinc}(v-1)\right]^2 \\
& +\left[\sum_{k=1}^{N}{}^{*}\sum_{m=1}^{N}\left|A(k,m)\right|\operatorname{sinc}(u-k)\operatorname{sinc}(v-m)\sin(\varphi(k,m))\right]^2 .
\end{aligned}
\tag{3.3}
$$

The notation $\Sigma^{*}\Sigma$ means skip the summation when $k=1$ and $m=1$ because this is the constrained pixel. It is possible to solve this system, but the solution can be difficult to obtain due to many local minima and the large number of equations. It is thus more convenient to seek a more creative algorithm.

Before continuing, the non-uniqueness of the solution is an important caveat to discuss. There are three effects that the ambiguity in the solution has on the image. Two of these effects are considered to be trivial—translation and rotation. The pixels in the image can be translated vertically and horizontally similar to a mapping onto a torus. Pixels will jump from one side of the image to the opposing side if they move out of the image domain. The rotational ambiguity refers to a 180 degree rotation of the image. Both of these ambiguities can be fixed by simply translating or rotating the image. The third ambiguity, which is far less common, is attributed to the Fundamental Theorem of Algebra not existing in this 2d case. That is, the multivariable equation (3.3) can be factored in more than one way, thus non-unique solutions exist.

3.2. Error-Reduction Method

A relatively simple iterative method of phase retrieval is the Gerchberg-Saxton algorithm adapted to phase retrieval by Fienup which is commonly referred to as the error-reduction method [23, 24]. The algorithm uses constraints on the image domain and Fourier domain to iteratively handle the large number of degrees of freedom. The algorithm has four basic steps starting with an initial guess of the image: (1) the image is Fourier transformed to the Fourier domain; (2) constraints are imposed on the Fourier domain; (3) the result is inverse Fourier transformed to give an estimate of the image; (4) constraints are imposed on the image. The process is repeated until the true image is revealed. The process is shown as a block diagram in Fig. 15. This algorithm is favorable because each pixel is not solved via a system of equations. Instead, projections subject to constraints are made to the data until the convergence conditions are met.

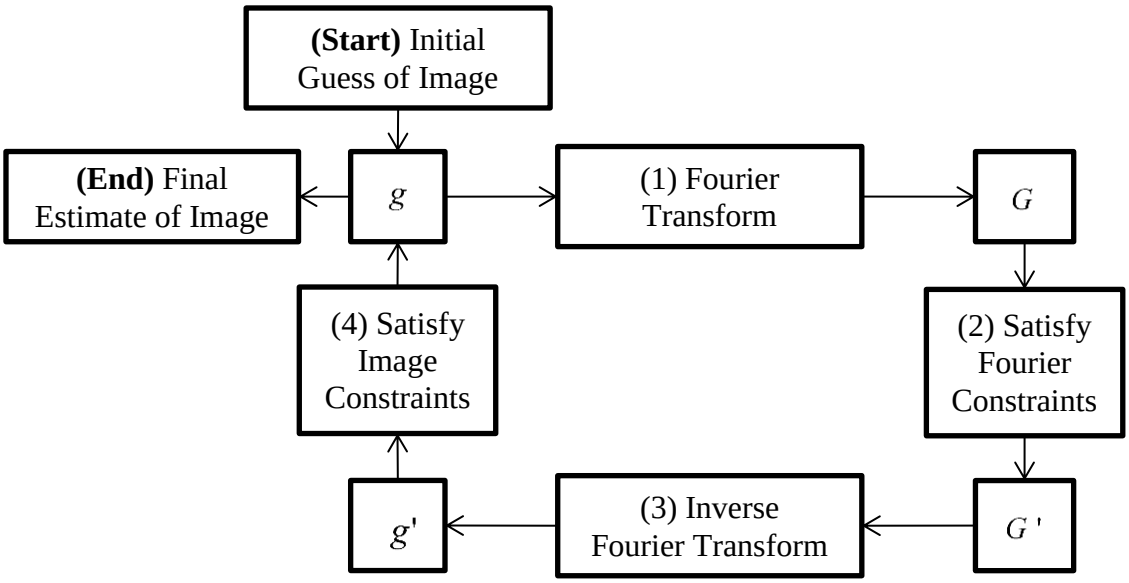


Fig. 15. Block diagram of the error-reduction approach [21]. The loop is continued until the constraint violations are sufficiently low, and the process terminates giving an estimate of the image.

The constraints on the Fourier transform are in the form of the known magnitude values from the interferometer measurements:

$$G'_k = \left|G_{measured}\right| \exp\left(i \arg\left(G_k\right)\right). \tag{3.4}$$

The constraints on the image are typically based on the following: the size of the object in the image is known approximately, the image has a blank background, and the pixel values must be real-valued. The background constraints are placed on the image at each iteration k such that

$$g_{k+1}(x,y) = \begin{cases} g'_k(x,y), & (x,y) \notin \gamma \\ 0, & \text{otherwise} \end{cases} \tag{3.5}$$

where γ is the region assumed to be the background of the image [23]. This is the form of the image constraint for the standard error-reduction method. The hybrid error-

reduction algorithm has a slightly different form for the constraints to expedite convergence. A gradient step is taken when imposing the constraints such that

$$g_{k+1}(x,y) = \begin{cases} g'_k(x,y), & (x,y) \notin \gamma \\ g_k(x,y) - \beta g'_k(x,y), & \text{otherwise} \end{cases} \tag{3.6}$$

The second line of equation (3.6) mimics the concept of negative feedback. This method tends to converge more quickly and tolerates noise better than the standard error-reduction algorithm.

Since this algorithm is iterative, it is convenient to quantify the convergence in terms of the mean-squared error. The error is defined by the sum of the squared constraint violations, i.e.

$$E^2_{0,k} = \sum_{\forall (x,y)} \left[g_{k+1}(x,y) - g_k'(x,y) \right]^2. \tag{3.7}$$

This error does not directly consider the difference between the true image's foreground and the estimated foreground. That is, the error does not actually quantify the appearance of the image as perceived by a person. Indirectly, errors in the foreground are considered in this definition due to the nature of this problem. The phase of a single pixel in the UV plane determines the phase of a sinusoid superimposed on the image at a given frequency. If there is an error in the phase of a UV pixel, the sinusoid related to that UV pixel will not be destructively interfered with by other UV pixel's sinusoids in the background. Simply stated, this definition of the error does capture errors in the foreground even though it only explicitly sums the pixels in the background of the image.

Fienup shows that the mean-squared error in the input-output algorithm (not necessarily the hybrid input-output algorithm) can only decrease which gives the algorithm its name, 'error-reduction' [23]. He defines the Fourier domain error,

$$\begin{aligned} E_{F,k}^2 &= N^{-2} \sum_{\forall (u,v)} \left| G_k(u,v) - G'_k(u,v) \right|^2 \\ &= \sum_{\forall (x,y)} \left| g_k(x,y) - g'_k(x,y) \right|^2, \end{aligned} \tag{3.8}$$

by using Parseval's theorem. Similarly, the image domain error in equation (3.7) can the transformed to give

$$\begin{aligned} E_{0,k}^2 &= \sum_{\forall (x,y)} \left[g_{k+1}(x,y) - g_k'(x,y) \right]^2 \\ &= N^{-2} \sum_{\forall (u,v)} \left[G_{k+1}(u,v) - G_k'(u,v) \right]^2. \end{aligned} \tag{3.9}$$

Because of the imposition of the image constraints in $E_{0,k}^2$ and the subsequent imposition of Fourier constraints in $E_{F,k+1}^2$, it holds that

$$\left| G_{k+1}(u,v) - G'_{k+1}(u,v) \right| \le \left| G_{k+1}(u,v) - G'_k(u,v) \right|. \tag{3.10}$$

which gives

$$E_{F,k+1}^2 \le E_{0,k}^2. \tag{3.11}$$

Expanding this expression to include an additional iteration gives

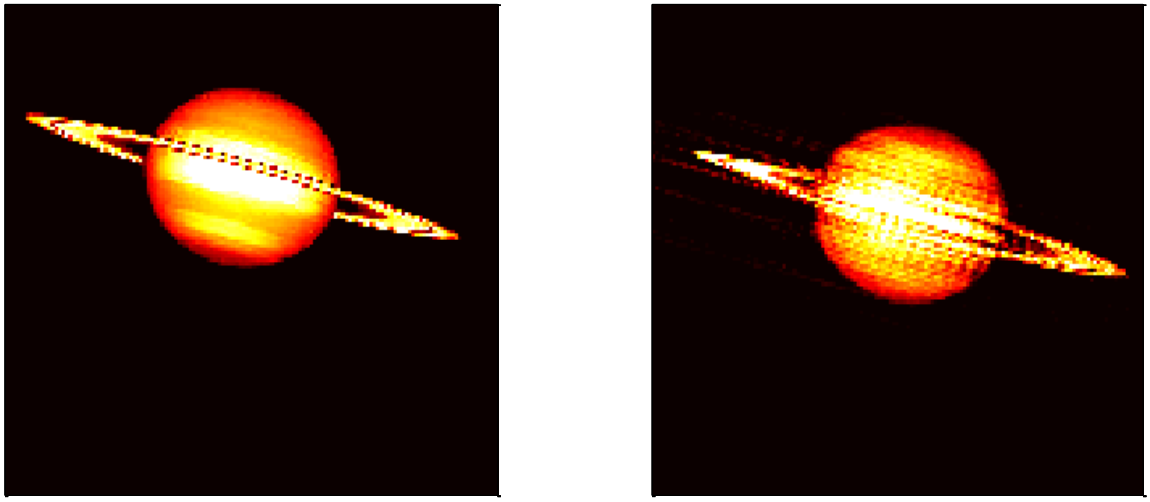
$$E_{0,k+1}^2 \le E_{F,k+1}^2 \le E_{0,k}^2 \le E_{F,k}^2 \tag{3.12}$$

which shows that both the image and Fourier domain error cannot increase at each iteration.

The discussions here-on do not use the standard error-reduction algorithm. Rather, the hybrid error-reduction algorithm is used. The standard algorithm has a proof that the error is monotonically decreasing; however, in practice the solution can assume local minima and take extremely long to converge. The hybrid error-reduction algorithm has no proof of convergence, but in practice shows better convergence behavior.

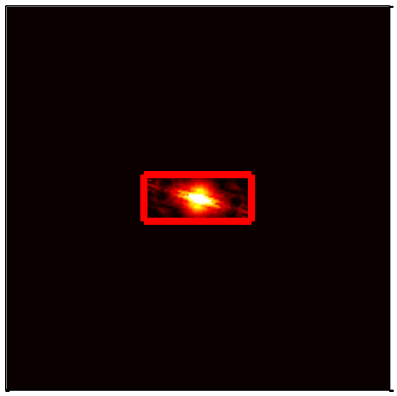
As an example of the hybrid error-reduction algorithm, the image of Saturn is considered. The true image and the estimated image from the hybrid error-reduction algorithm are shown in Fig. 16. Fig. 17 shows the results at various iterations. The red box indicates the boundary of the domain γ which corresponds the purely background region. Note that outside the boundary the background is nearly all black, but some artifacts[3] exist around the planet at some iterations. Furthermore, the planet is translated from its original position in the image frame. This is the result of the non-unique nature of the phase solution. The various phase solutions correspond to translation and rotation of the image. Fig. 18 shows the RMS error at each iteration. Notice that the error only decreases at each iteration.

[3] Artifact is a commonly used term in imaging and digital graphics that refers to any undesired abnormality. Its exact definition is quite vague.

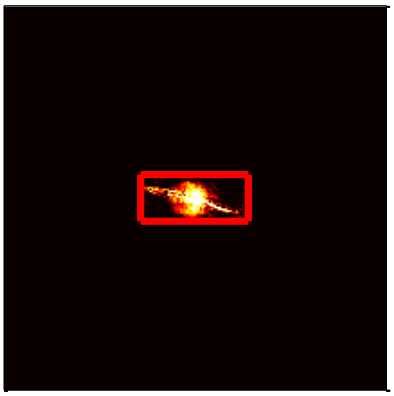


(a) True Image (b) Estimated Image

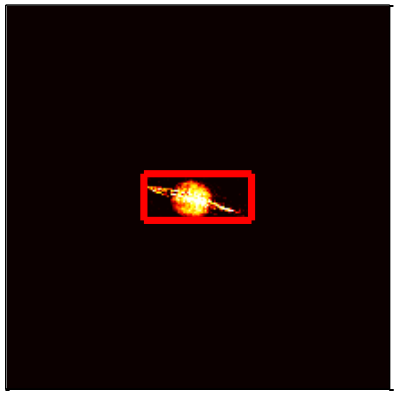
Fig. 16. (a) The test image of Saturn and (b) the results of the hybrid error-reduction algorithm.



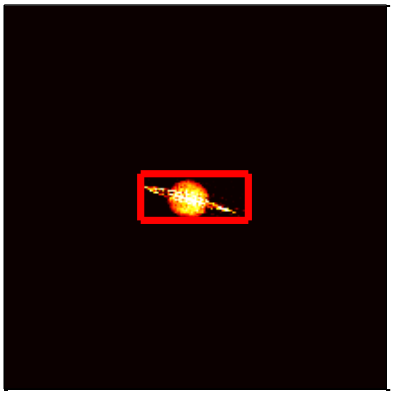
(a) Iteration 1



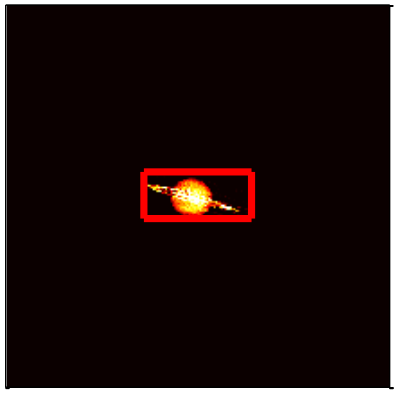
(b) Iteration 50



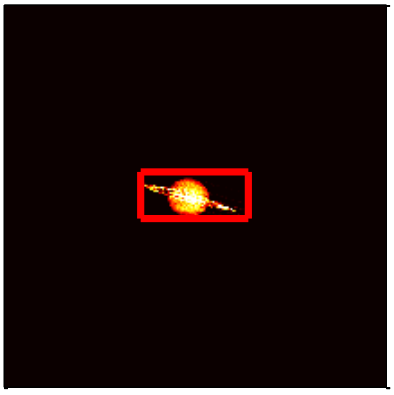
(c) Iteration 100



(d) Iteration 250



(e) Iteration 500



(f) Iteration 2000

Fig. 17. Reconstructed image of Saturn at various iteration steps using the error-reduction phase retrieval.

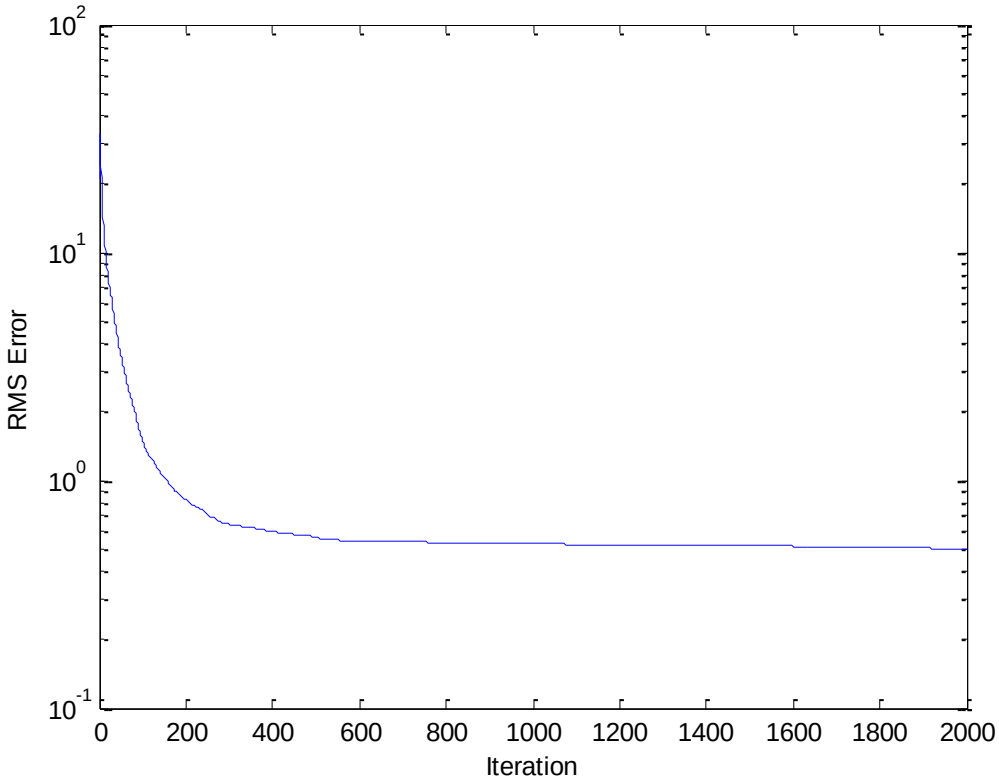


Fig. 18. RMS error at each iteration for the hybrid error-reduction algorithm used on the image of Saturn.

In general there is no unique solution for the phases; however, there are many solution sets which will provide a satisfactory image (discussed in section 3.1). The difficulty in using the error-reduction method is determining the exact constraint γ to impose on the image. This is the primary topic of discussion in section 4.1.

4. PHASE RETRIEVAL ALGORITHMS

In this section, algorithms for estimating the phase of a two-dimensional Fourier transformed image are discussed in detail. The astronomical imaging application is considered in which the image has a black background around the foreground object of interest.

The formal problem statement is the following. The magnitude of the 2d Fourier transform of an unknown image is known but noisy, and the phase is unknown. The background of the image is known to be black, but the exact size of the foreground object is unknown. The phase must be estimated and combined with the known magnitude. Then the result is to be inverse Fourier transformed to reveal an estimate of the unknown image.

The phase retrieval algorithm presented here is based on the hybrid error-reduction algorithm as presented by Fienup in [23]. As discussed in section 3.2, the algorithm has four steps which are repeated. Initially, a guess of the image is formed. Next, the image is Fourier transformed. Constraints are imposed on the Fourier transform. The result is inverse Fourier transformed, and finally constraints are imposed on the image. After many iterations, the number of corrections required to impose the constraints decreases, and the resulting image resembles the true image. The primary difficulty in this method is knowing what the constraints should be. The constraint on the Fourier modulus is the given ‘data.’ The obvious constraint on the image is the pixel values must be real and within a specified range, i.e. positive. The difficult constraint to

impose is the background region, because the size of the foreground is unknown. A secondary difficulty is forming the initial guess of the image. Both of these tasks are further complicated when noise is introduced into the Fourier modulus data.

4.1. The Collapsing Foreground Algorithm

The following are proposed as suitable solutions to determinging the background region in the phase retrieval problem without *a priori* knowledge of the size of the foreground. The foreground is initially assumed to be almost as big as the entire image. After the first few iterations and at each subsequent iteration, the pixels next to the background boundaries, called the ‘check region’, are inspected. If most of the pixels in the check region have a low value, the background boundary is moved inward slightly. If the check region contains too many pixels above a specified intensity, the background boundary will move outwards to prevent cutting off part of the foreground. This concept is shown in Fig. 19. The dashed line is the background boundary, and everything outside the boundary is constrained to be background. The light-gray region just inside the background boundary is the check region. As shown, the top edge of the background border is close to the image, so it will not move inward. The other three background borders are not close to the star so at each iteration they will move inward until their respective check regions touch the star. The moving background border gives the name of the algorithm: the collapsing foreground algorithm. This method essentially finds the foreground within the image. In practice, during the iterations the image is not a clear shape. It is rather many seemingly random patches of pixels as the image develops during the iterations.

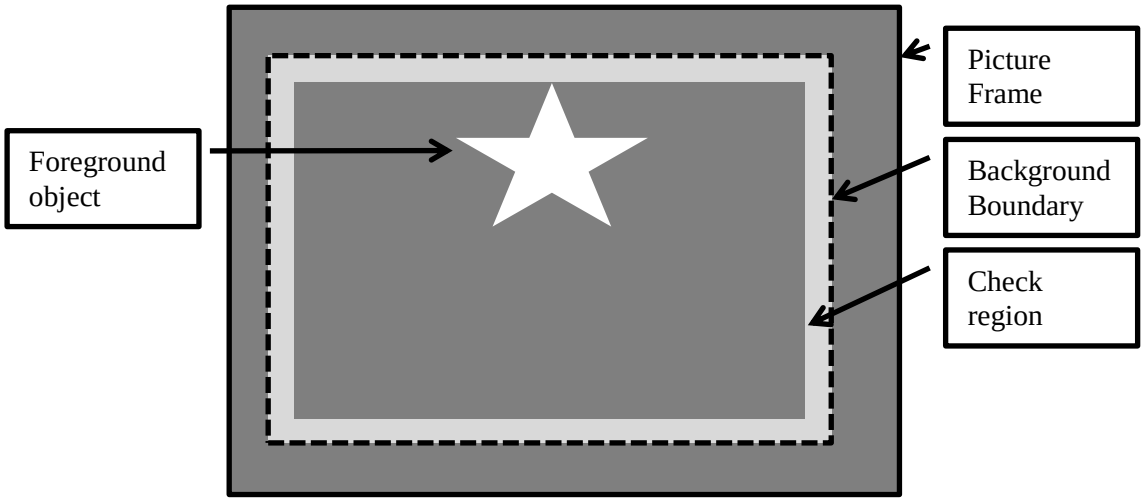


Fig. 19. Schematic of background boundary and check region.

It should be noted that the analysis of the RMS error given in section 3.2 does not entirely hold for this algorithm. The error is not always decreasing, because the argument that supports expression (3.10) does not hold if the background boundaries move inwards. Typically a sharp increase in the RMS error is seen when the background borders move inward; however, between movements the error is still monotonically decreasing.

Further detail in the form of a pseudo-code of this algorithm combined with the error-reduction algorithm is given in Algorithm 1 and Algorithm 2 in APPENDIX A.

4.2. Relaxed Constraints

When noise is introduced into the Fourier modulus data, the correct phase solution will include artifacts in the background of the image. The hybrid error-reduction allows for non-zero pixels in the background, effectively relaxing the background constraint. Because the true solution includes non-zero pixels in the background, forcing these pixels to be zero will drive the estimate away from the true phase solution. Likewise, the Fourier modulus data contains noise, so the true image cannot be obtained if it is rigidly constrained.

The Fourier modulus constraint can be relaxed by allowing it to drift over time. The constrained value, G', for the Fourier modulus can be defined according to

$$\left|G'_k(u,v)\right| = \varepsilon \left|G_k(u,v)\right| + (1-\varepsilon)\left|G_{measured}(u,v)\right| \tag{4.1}$$

Notice that the constrained Fourier modulus is comprised of the Fourier data before the constraints are imposed on the current iteration, G_k, and the initial modulus data, $G_{measured}$. Each of these components is important. The G_k term brings the image domain constraint's effect into the Fourier modulus; this is essentially where the noise cancelation comes from. The initial modulus term ensures stability and prevents the modulus value from running away. If the parameter ε is zero, the constraint is exactly in the error-reduction method's form.

4.3. Initial Condition and Image Averaging for Noisy Data

The initial condition has an effect on the final image. Fig. 20 shows how various initial conditions give different artifacts in the final images. It is thus proposed that for the initial guess the edges of the initial image are assumed to be background. This region

may be about ten percent of the total pixels in the image. The rest of the initial image contains random pixel values. Here the pixel values have a uniform probability distribution from zero to one.

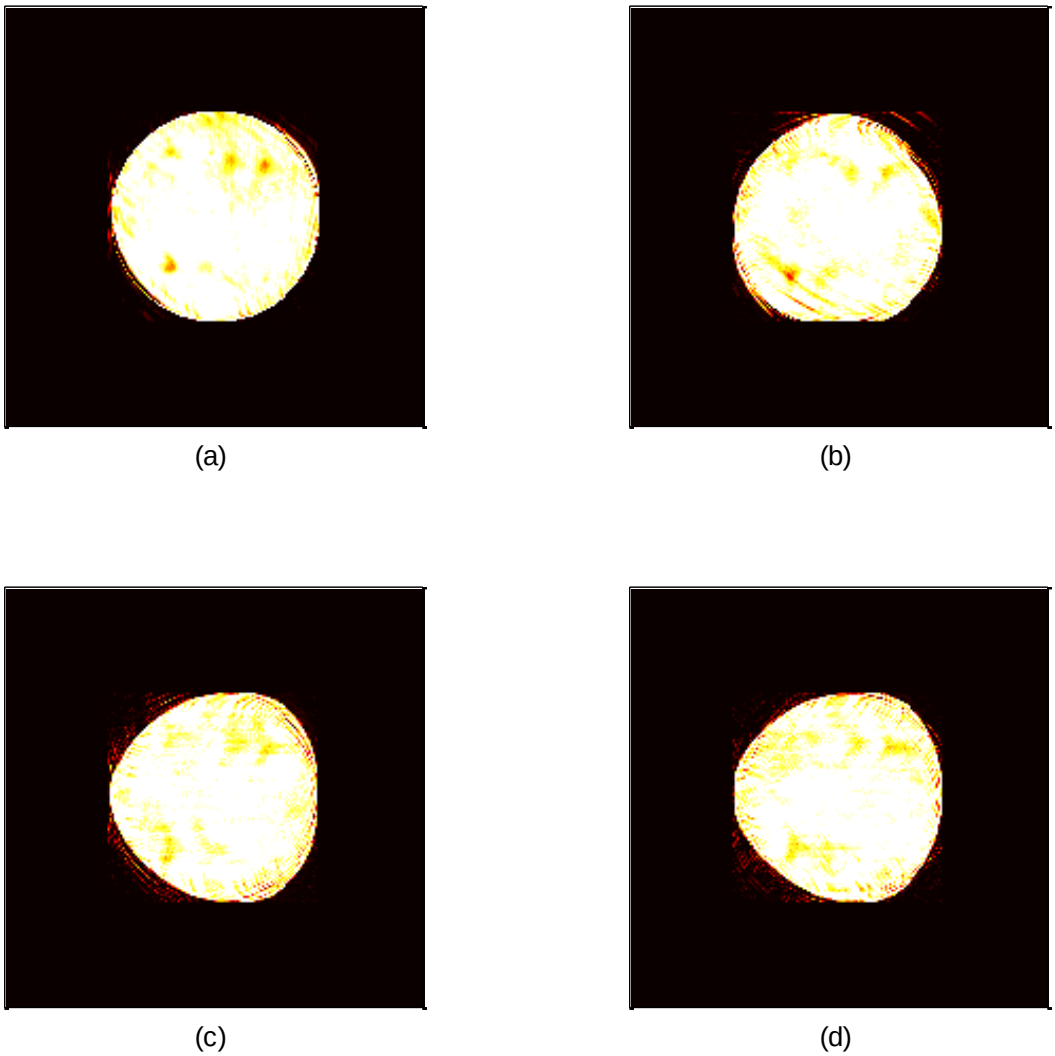


Fig. 20. Image estimates after 500 iterations (not fully converged) with four different initial conditions.

After using the random initial image in the error-reduction method, an estimate of the true image will be obtained. The image will contain many artifacts due to the noise in the Fourier modulus data and error caused by some of the initial images be poor choices. As shown in Fig. 20a, some initial conditions work well; however, the other three have much more distortion in the image and will require more iterations to correct the distortion. The artifacts can be nullified by generating several image estimates using different initial images. The resulting images can be averaged together to yield an image with less obvious artifacts.

To combine multiple image estimates, they must each be translated such that the foregrounds all line up. This can be done by finding the center of mass of the image matrix and circularly shifting the rows and columns of the matrix. The center of mass is found using

$$cm_x = \frac{\sum_{\forall x} x \left\lceil \sum_{\forall y} g(x,y) \right\rceil}{\sum_{\forall (x,y)} g(x,y)} \tag{4.2}$$

and

$$cm_y = \frac{\sum_{\forall y} y \left\lceil \sum_{\forall x} g(x,y) \right\rceil}{\sum_{\forall (x,y)} g(x,y)}. \tag{4.3}$$

The rows and columns are then shifted until the center of mass is in the middle of the matrix. The separate image estimates are then averaged.

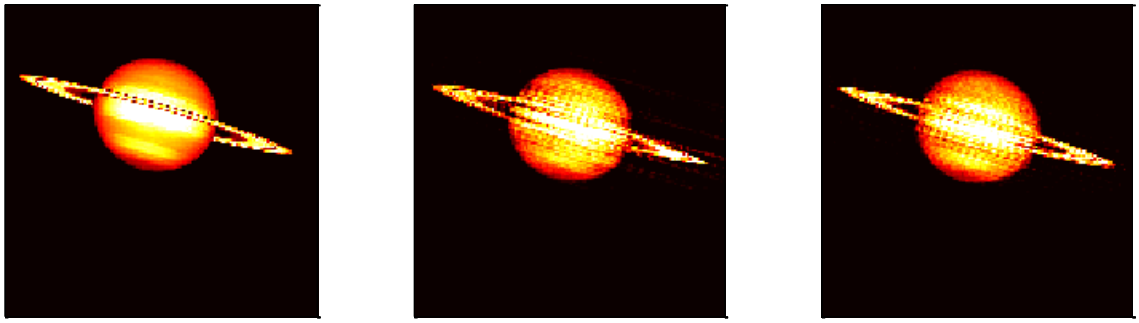
5. PERFORMANCE & RESULTS

The following sections show the performance of the collapsing foreground algorithm, the relaxed constraint algorithm, and the image averaging algorithm. The same images of Saturn and of a fictitious star with star-spots are considered. It should be noted that there exists an ambiguity in the phase solutions that results in a 180 degree rotation of the images in some results. The images in the following results are rotated to correct for this ambiguity when necessary.

5.1. Collapsing Foreground Algorithm with Noiseless Data

The collapsing foreground algorithm presented in section 4.1 proves to have performance comparable to the hybrid error-reduction method even considering the size of the foreground is unknown. The image of Saturn is used to compare the error-reduction method to the collapsing foreground algorithm. Fig. 21 is a comparison between the true image and the reproduced image with much of the background cropped. The close-up reveals some artifacts around the planet in the background. Most of the detail in the form of thin lines along the rings are preserved. The image estimate at various iteration steps is shown in Fig. 22. The foreground collapsed to the similar size as that used in the error-reduction method example. The RMS error shown in Fig. 23 reveals that the errors are comparable. Typically the collapsing foreground algorithm seems to show a quicker decrease in the RMS error. The collapsing foreground algorithm, however, has upward trends in the error value that are not present in the error-reduction method. These upward trends typically occur when the foreground is collapsed

because additional pixels are added to the background regions. The effect of the collapse on the error is understood by referring back to the definition of the RMS error given in equation (3.7).

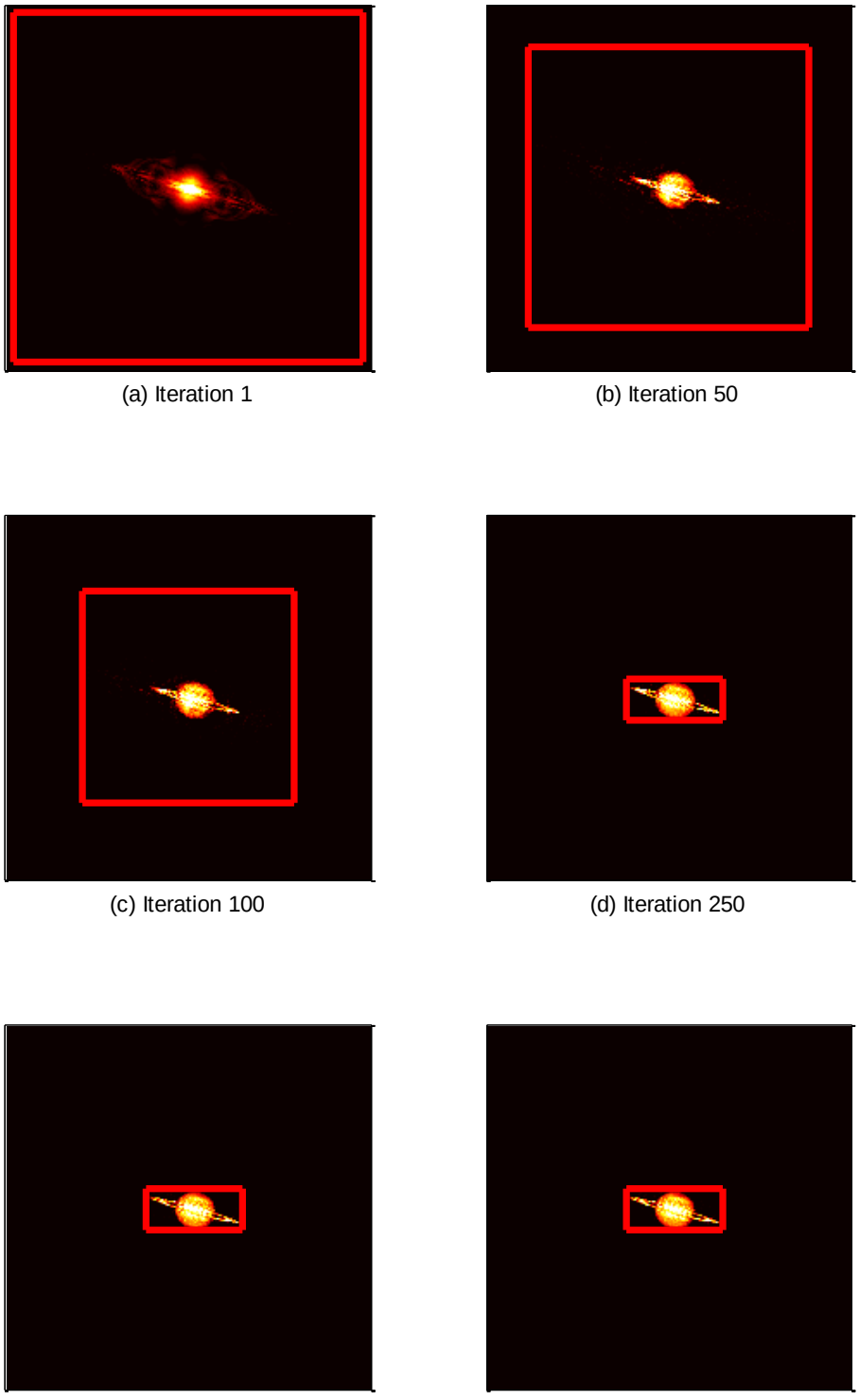


(a) True Image

(b) Estimated Image
Hybrid Error-Reduction

(c) Estimated Image
Collapsing Foreground

Fig. 21. Close-up comparison of true image of Saturn and the results of the collapsing foreground algorithm. Both the hybrid error-reduction and collapsing foreground algorithms were given the same initial guess and same number of iterations.



(a) Iteration 1

(b) Iteration 50

(c) Iteration 100

(d) Iteration 250

(e) Iteration 500

(f) Iteration 2000

Fig. 22. Reconstructed image of Saturn at various iteration steps using the collapsing foreground algorithm.

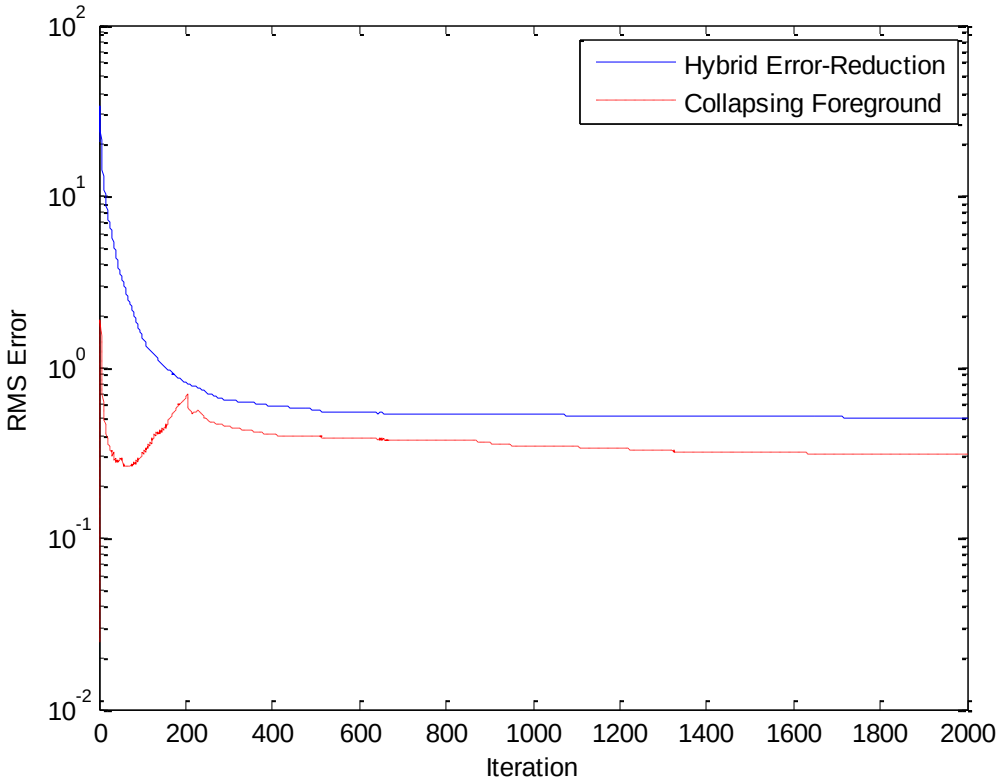


Fig. 23. Comparison between the RMS error for the collapsing foreground and error-reduction algorithms for the image of Saturn.

Shown in Fig. 24 is an image of a star with star-spots used for additional evaluation of the collapsing foreground algorithm. Fig. 25 and Fig. 26 show the same algorithm as shown for the image of Saturn but for an image of a star with star-spots. The results are comparable to that of the Saturn image, and the star-spots are recognizable. The intensity and size, however, is slightly incorrect for some of the spots. Additional iterations can be performed to gain more detail of the star-spots.

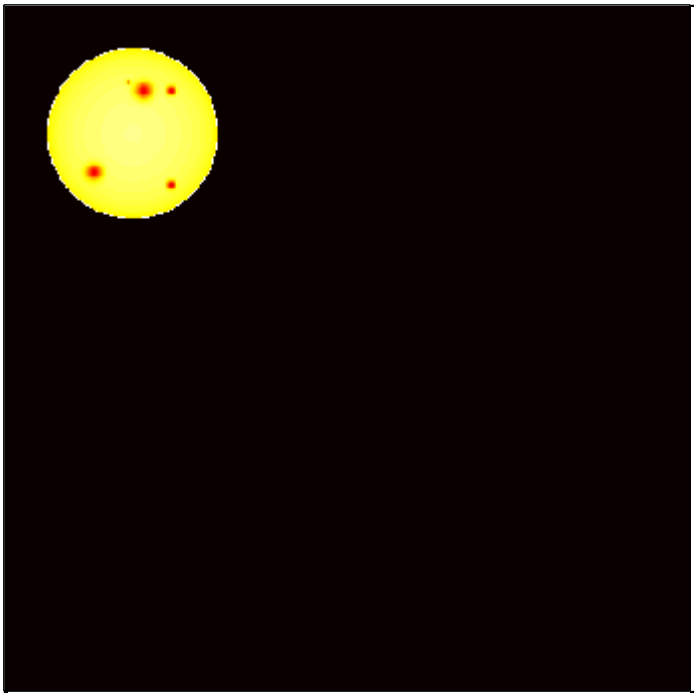
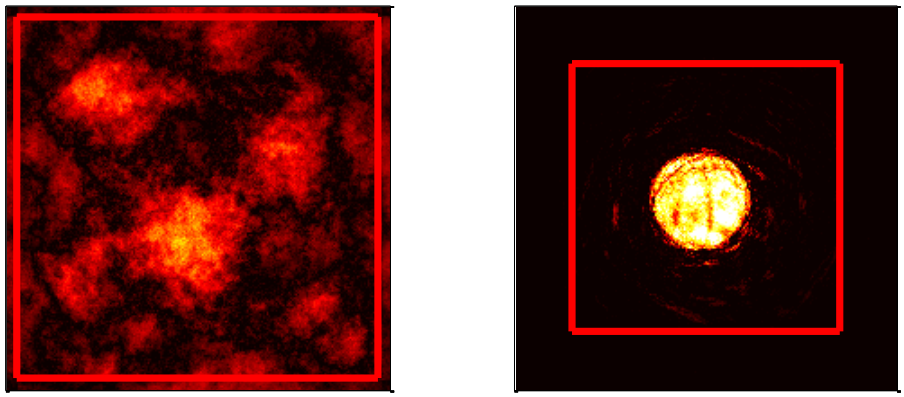


Fig. 24. Original image of a star with star-spots.



(a) Iteration 1

(b) Iteration 50

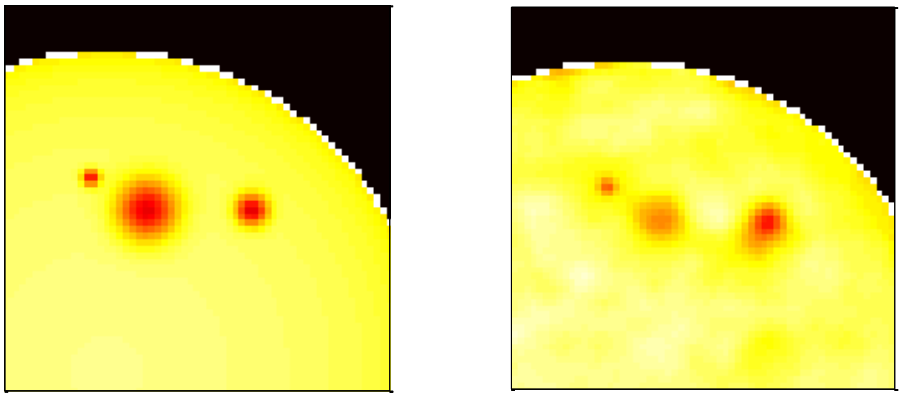
(c) Iteration 100

(d) Iteration 200

(e) Iteration 500

(f) Iteration 1000

Fig. 25. Reconstructed image of a star with star-spots at various iteration steps using the collapsing foreground algorithm.



(a) True Image (b) Estimated Image

Fig. 26. Close-up comparison of (a) the true image of a star with star-spots and (b) the results of the collapsing foreground algorithm.

5.2. Collapsing Foreground Algorithm with Noisy Data

To test the collapsing foreground algorithm with noisy data the image of Saturn is again considered. The modulus data has additive Gaussian noise such that the SNR is defined as

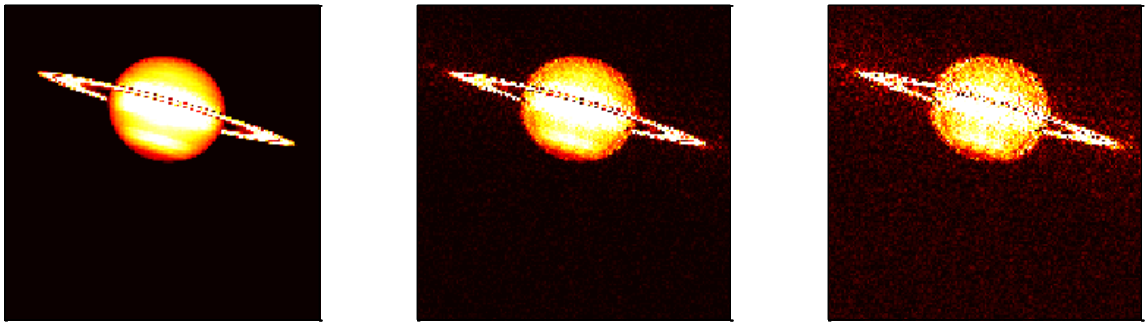
$$SNR = \frac{\mu(|G|)}{\sigma(N)} \tag{5.1}$$

which is the ratio of the mean of the Fourier modulus values $|G|$ to the standard deviation of the additive noise N. Fig. 27 shows (a) the original image of Saturn and (b and c) the image after the Fourier modulus has been corrupted with the additive Gaussian noise according to

$$|G_{Measured}(u,v)| = |G_{Measured}(-u,-v)| = |G_{True}(u,v)| + |N(o,\sigma)|. \tag{5.2}$$

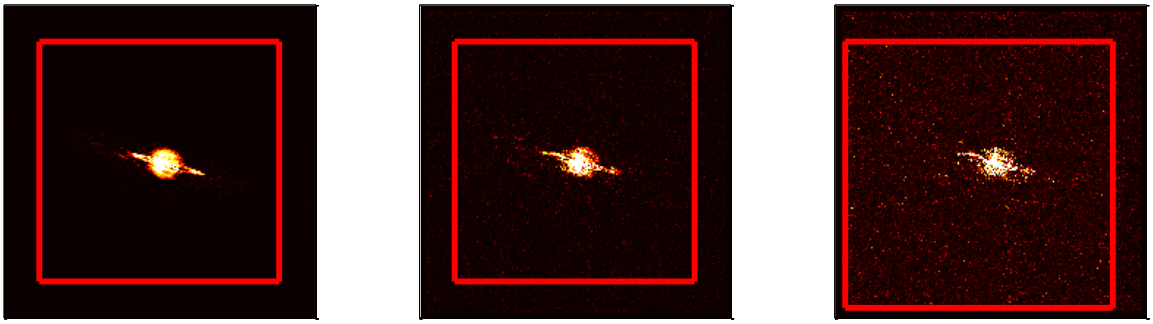
Fig. 28 shows the estimated image at several iteration steps. Note that there are some pixels with non-zero values in the background region at earlier iterations. This is because

the β value in equation (3.6) has a small value of 0.01. The small β helps the algorithm converge in the presence of noise by allowing pixel values in the background to be non-zero but non-increasing at each iteration. Fig. 29 shows the final image estimates after 1000 iterations. For these high levels of noise, the details of image are virtually lost; however, even at the high SNR of 0.42 the overall size of the planet is preserved. This result implies that the typical use of ICI, measuring sizes of celestial objects, is still practical. Fig. 30 shows the RMS error at each iteration for the three SNR levels. The upward trend in the noisy images is the result of the background boundaries moving inward. As might be expected, the SNR is inversely proportional to the final RMS error.

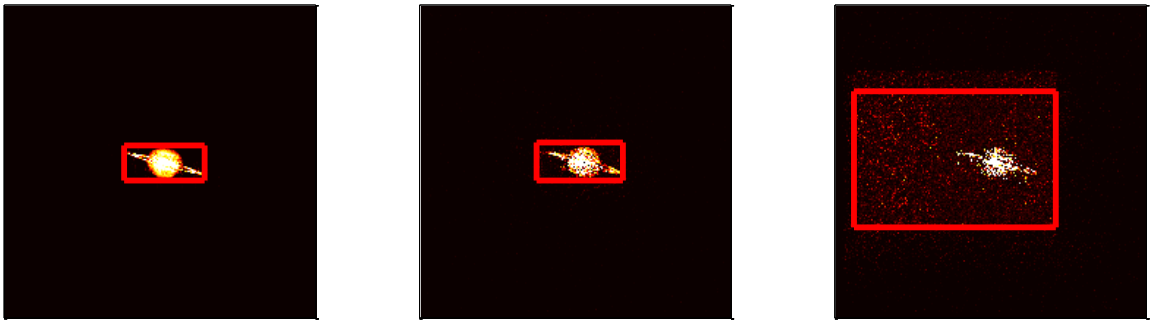


(a) SNR: Inf (b) SNR: 0.48 (c) SNR: 0.24

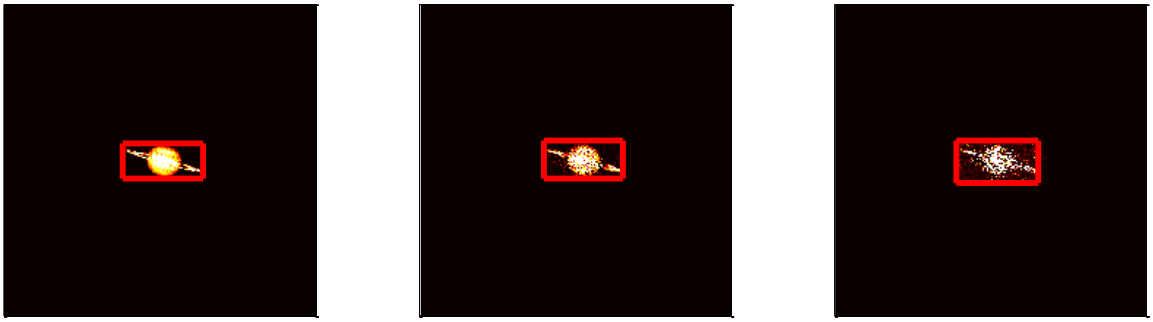
Fig. 27. Images corresponding to the noisy Fourier modulus and the true phase. Shown in (a) is the true image with no noise added. Depicted in (b) and (c) are the images corresponding to the corrupted Fourier modulus data and the true phase data. The SNR for (b) and (c) are 0.48 and 0.24 respectively. Some of the background is cropped.



(a) Iteration: 50 SNR: Inf (b) Iteration: 50 SNR: 0.48 (c) Iteration: 50 SNR: 0.24

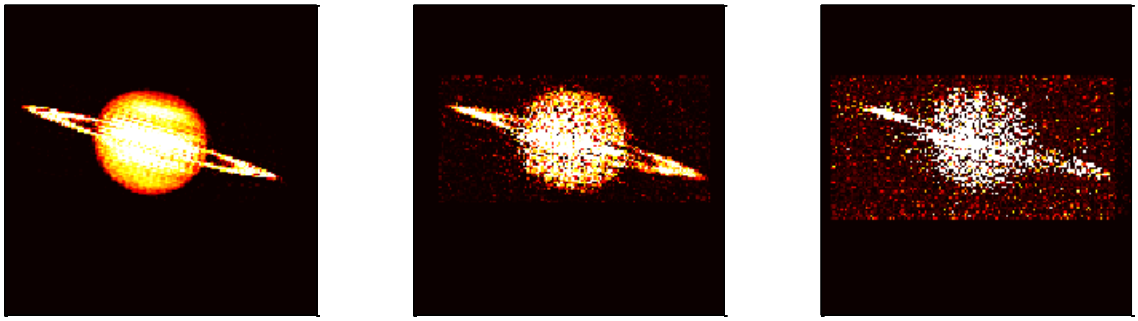


(d) Iteration: 250 SNR: Inf (e) Iteration: 250 SNR: 0.48 (f) Iteration: 250 SNR: 0.24



(g) Iteration: 1000 SNR: Inf (h) Iteration: 1000 SNR: 0.48 (i) Iteration: 1000 SNR: 0.24

Fig. 28. The image of Saturn reproduced from noisy modulus data at various iterations. (a,d, and g) have no additive noise. (b,e, and h) have an SNR of 0.48. (c,f, and i) have an SNR of 0.24.



(a) SNR: Inf (b) SNR: 0.48 (c) SNR: 0.24

Fig. 29. Final image estimates after 1000 iterations for three levels of additive Fourier modulus noise. Some of the background is cropped.

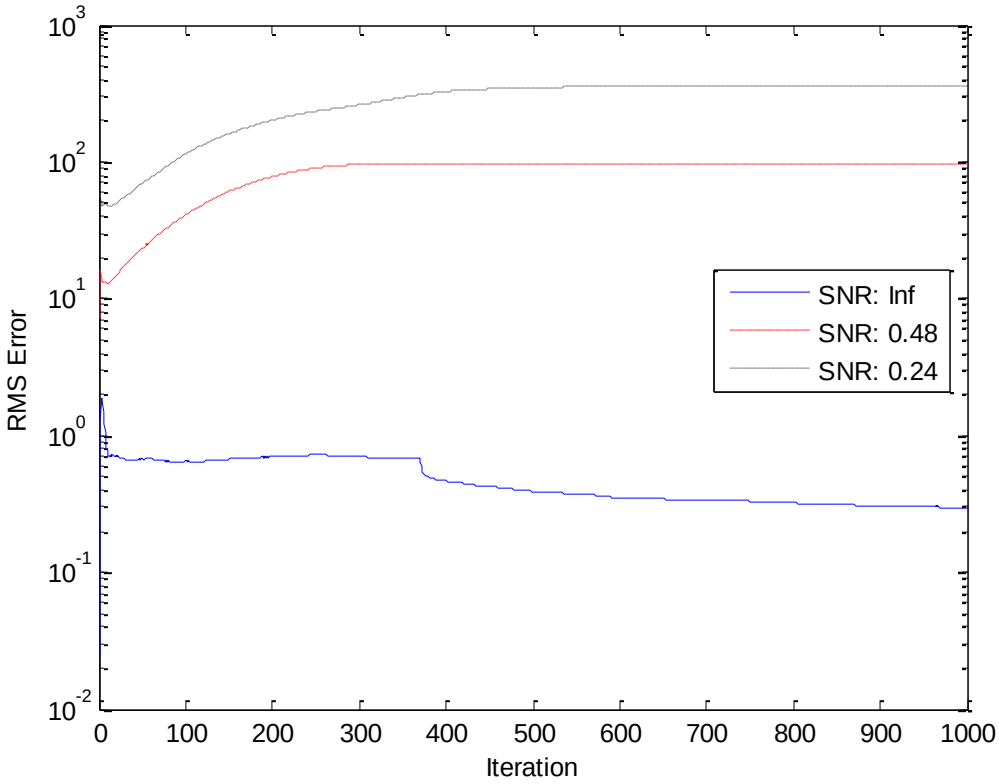
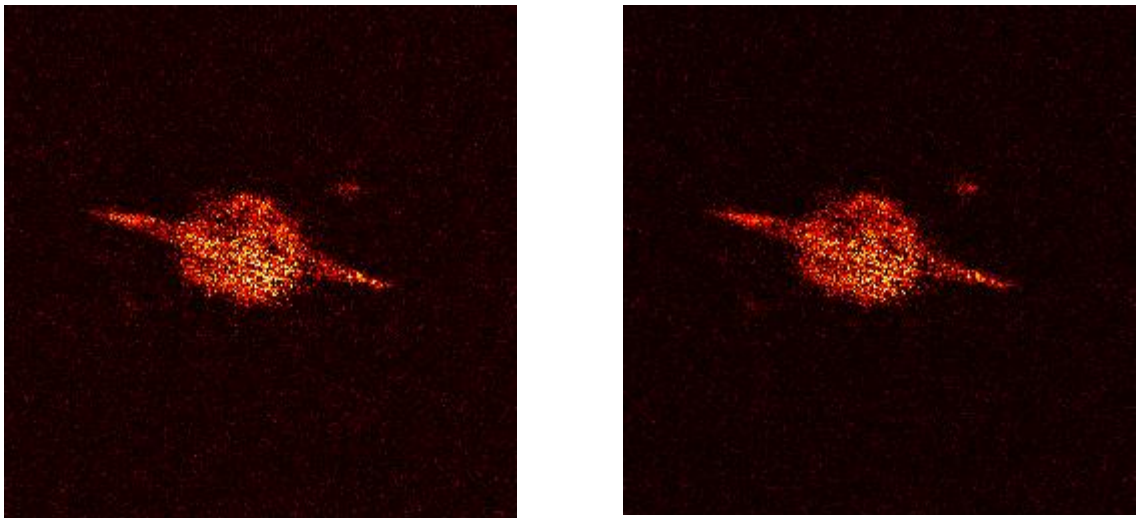


Fig. 30. RMS error at each iteration for the collapsing foreground algorithm used on noisy image of Saturn.

5.3. Relaxed Constraints

The relaxed constraint algorithm aims to remove some of the noise in the UV modulus constraint. Typically the SNR can be increased between 15% and 30%. Unfortunately, although the SNR is increased the image foreground typically doesn't change much through the relaxation process. The primary effect is a reduction in the background artifacts. The hybrid input-output allows pixels in the background to be non-zero which improves the foreground, but this can decrease the contrast between foreground and background. Fig. 31 shows a comparison of the background intensity before and after the constraint is relaxed. The effect is approximately a 50% reduction in the pixel values across the background with virtually no change to the foreground. The exact amount of error reduction is greatly dependent on the image, noise, relaxation parameter values, and the estimation algorithm. The amount of error reduction is also typically inversely-proportional to the parameter in the hybrid input-output algorithm.

The effect of relaxing the UV constraint on the error is shown in Fig. 32. The constraint was relaxed at iteration 1000. Notice a sharp drop in the background constraint error and the drop in the UV magnitude constraint error. After the constraint is relaxed and the UV error decreases, often an oscillation will be scene between the background and UV constraints. One error will increase slightly and the other will decrease repeatedly. Typically the range of oscillation is much less than the overall error reduction; however, at very high noise levels the oscillation can be substantial.



(a) Prior to constraint relaxation

(b) After constraint relaxation

Fig. 31. Comparison of the background pixel intensities before and after the UV constraint it relaxed.

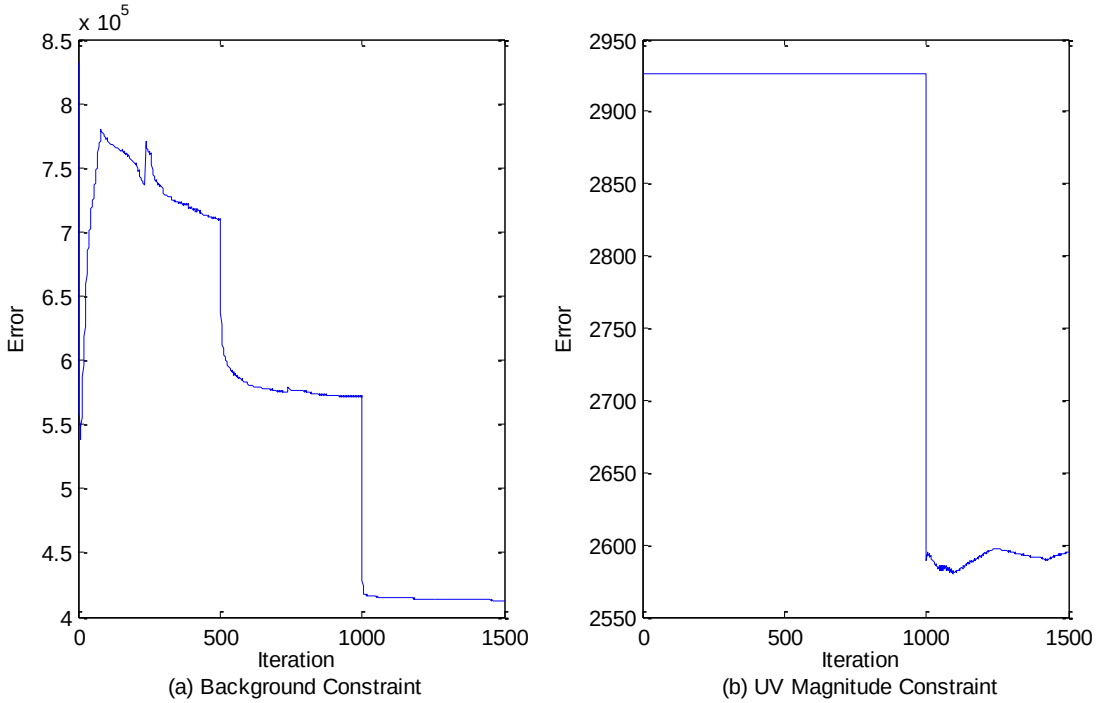


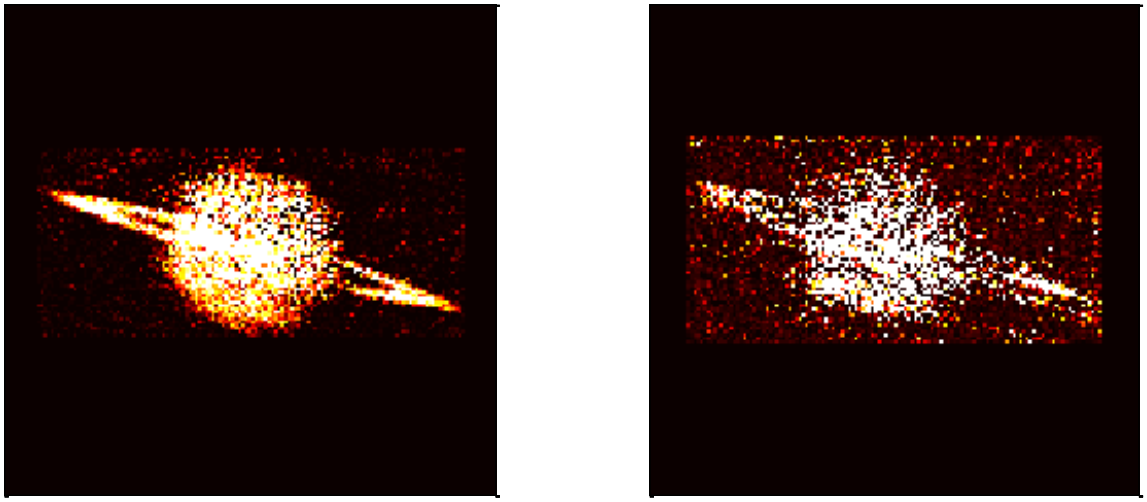
Fig. 32. The effect of relaxing the UV constraint on the UV magnitude error and the background constraint error. The UV magnitude constraint was relaxed at iteration 1000.

5.4. Image Averaging

The image averaging algorithm can be used to cancel out some artifacts that arise as the result of noise in the Fourier modulus data. To demonstrate this algorithm, estimates of the image of Saturn and the star with star-spots, like those shown in the previous section, are used. Twenty image estimates are formed using the collapsing foreground algorithm. Each is circularly shifted and then averaged together. The initial conditions for each estimate are independent, uniformly distributed random images.

For the image of Saturn, the Fourier modulus SNRs considered are the same as those used for the collapsing foreground algorithm examples: 0.48 and 0.24. Examples of the individual image estimates are shown in Fig. 33. The resulting averages of twenty of these estimates are shown in Fig. 34 and Fig. 35. The results show fewer artifacts in the disk of the planet. Also, the magnitude of the pixel values is closer to that of the true image; the colors are similar around the disk of the planet. These corrections to the image come at the expense of a blurring of the edges of the planet. In most cases this blurriness can be somewhat corrected through standard image sharpening techniques.

The image of the star with star-spots is more difficult to estimate because the small details, the spots, are of interest. As examples, the Fourier modulus data has an SNR of 1.5 and 0.75 in these two examples. The resulting images are shown in Fig. 36 and Fig. 37. The high SNR result shows that the outline of the star is very clear. The spots, however, are present but not very clear. Their positions are evident, but the size of the spots is difficult to measure. The low it shows less definition around the edge of the star and the spots are difficult to recognize.



(a) SNR: 0.48 (b) SNR: 0.24

Fig. 33. The first of the twenty image estimates used in the image averaging examples.

Fig. 34. The result of the image averaging algorithm using twenty image estimates with a Fourier modulus SNR of 0.48.

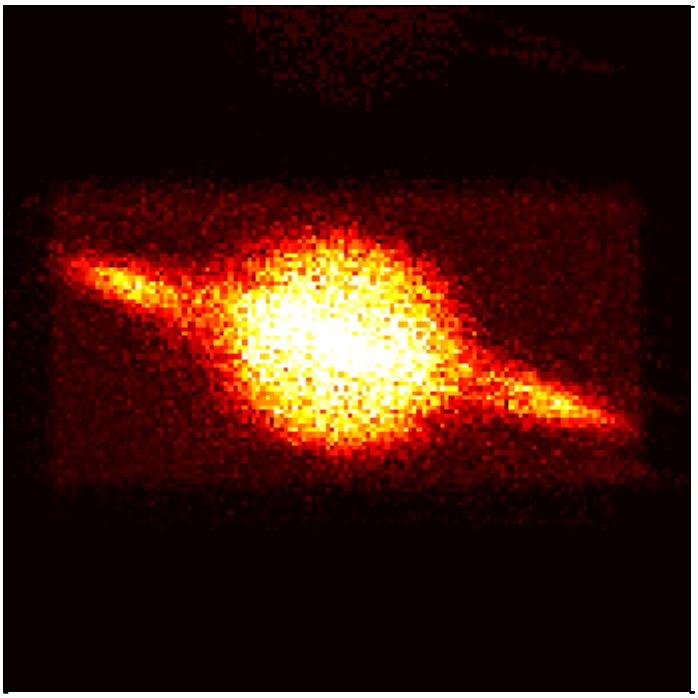


Fig. 35. The result of the image averaging algorithm using twenty image estimates with a Fourier modulus SNR of 0.24.

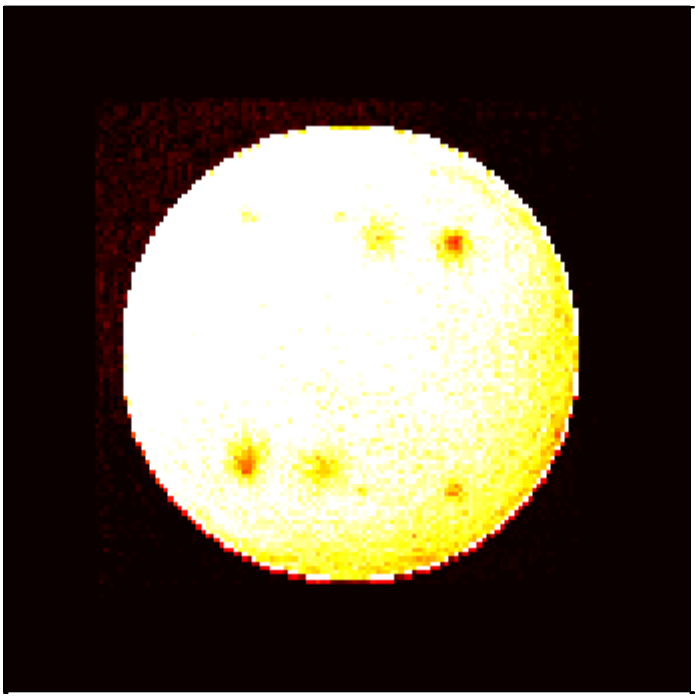


Fig. 36. The estimate of the image of a star using the image averaging algorithm using twenty image estimates with a Fourier modulus SNR of 1.5.

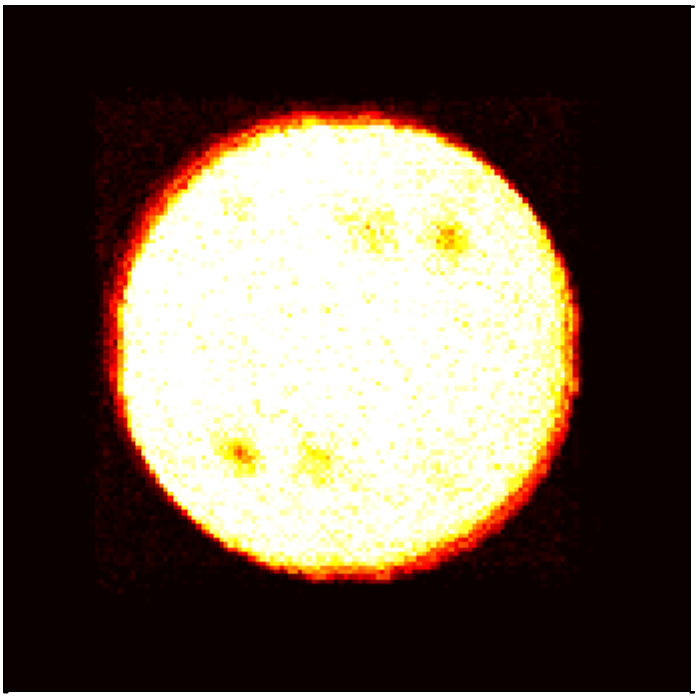


Fig. 37. The estimate of the image of a star using the image averaging algorithm using twenty image estimates with a Fourier modulus SNR of 0.75.

5.5. Performance Improvements

Most of the examples shown in this thesis were produced in a manner of minutes on a typical PC using Matlab. Utilizing a compiled language can increase the iteration speed of these algorithms. For the sake of comparison, these algorithms were implemented in a Windows forms application written in C#. This program includes a complete graphical user interface for managing the creating of the image estimates. The interface is shown in APPENDIX B. Through careful memory management and parallelization of the algorithms, multiple collapsing foreground trials can be run simultaneously an easily combined through the image averaging technique.

The crux of such a program's performance is the Fourier transform algorithm used. The best benchmarked FFT algorithm is the FFTW subroutine. It is a subroutine library written in C, which can be ported to C++ and C#, that out performs most other standard FFT methods in every case. It attains its efficiency by trying several different FFT methods ahead of time and checks which gives the fastest results for the problem being considered. The 'plan' is then utilized repeatedly through the iterations of the phase retrieval program.

To compare the performance of the C# program with that of the same algorithm in Matlab, the image of Saturn was run for 1000 iterations and is 256 by 256 pixels. The C# program performs approximately 120 iterations per second. In Matlab 45 iteration per second is typical. Additionally, when running multiple trials simultaneously, as is desired for using the image averaging algorithm, the C# program provides far superior

results. Although Matlab supports running several routines in parallel using the parallel computing toolbox, the iterations are by far slower.

While Matlab is the tool of choice for developing such algorithms, it is well known that it does not provide quick performance. These results give the notion that the practicality of these iterative algorithms can be greatly increased by paying further attention to the implementation of the algorithm.

6. CONCLUSIONS

This thesis proposed three adaptations of Fienup's phase retrieval algorithm that make phase retrieval of images feasible for Intensity Correlation Interferometry applications. The primary consideration was eliminating the need for knowledge of the size of the foreground of the image before attempting phase retrieval. It was shown that proper phase retrieval is possible with this information completely unknown by using the collapsing foreground algorithm. This algorithms performance was shown on ideal data and noisy data. The algorithm proved capable of providing a comprehendible image given Fourier modulus data contained additive, element-wise Gaussian noise with an SNR as low as 0.24. The addition of constraint relaxation also allowed for added compensation of noise. It should be noted, however, that the SNR metric is somewhat dependent on the content of the image.

Another consideration in this thesis was the issue of choosing an initial condition for the solver. It was shown that the initial condition affects the artifacts in the final image estimate. This led to the proposal of the image averaging algorithm where several estimates of the image are formed with various random initial conditions. The results were them averaged together to eliminate many of the artifacts unique to the initial condition for the trial. This algorithm also showed an improvement in eliminating some of the artifacts introduced by the Fourier modulus noise.

In the past the algorithms presented here would not have been feasible due to their computational expense; however, with advances in computational power and fast

Fourier transform algorithm efficiency the past few years, the computational cost in these algorithms is manageable. With additional attention paid to programming efficiency, parallelization, and compilation, the solution times can possibly be reduced to just several seconds.

Because intensity correlation interferometers suffer from very low SNR levels, an improvement to these algorithms may lie in the constraint relaxation algorithm. Since the noise in the Fourier modulus data will distort the image and cause artifacts in the background, both non-zero pixels in the background region and deviation from the given Fourier modulus data should be allowed. Relaxation of the constraints may help allow for higher tolerance of noise in the modulus data.

REFERENCES

[1] R. Hanbury Brown and R. Q. Twiss, "A New Type of Iterferometer for Use in Radio Astronomy," *Philosophical Magazine,* vol. 45, no. 366, 1954.

[2] R. Hanbury Brown and R. Q. Twiss, "A Test of a New Type of Stellar Interferometer on Sirius," *Nature,* vol. 178, no. 4541, pp. 1046-1048, 1956.

[3] R. Hanbury Brown and R. Q. Twiss, "Interferometry of the Intensity Fluctuations in Light," *Proceedings of the Royal Society of London,* vol. 242, no. 1230, pp. 300-324, 5 Nov 1957.

[4] A. Ghatak, Optics, 4th ed., West Patel Nagar: McGraw-Hill, 2009.

[5] R. Hooke, Micrographia: or, Some Physiological Descriptions of Minute Bodies Made by Magnifying Glasses, London: J. Martyn and J. Allestry, 1665.

[6] M. D. Fayer, Absolutely Small: How Quantum Theory Explains Our Everyday World, New York: American Management Association, 2010.

[7] T. Young, "On the Theory of Light and Colours," *Philosophical Transactions of the Royal Society of London,* vol. 92, pp. 12-48, 1802.

[8] M. Born and E. Wolf, Principles of Optics, 6th ed., Cambridge: Cambridge University Press, 1997.

[9] W. H. Steel, Interferometry, 2nd ed., New York, NY: Cambridge University Press, 1983.

[10] P. Hariharan, Basics of Interferometry, 2nd ed., San diego, Ca: Elsevier, 2007.

[11] E. Hecht, Optics, 4th ed., San Francisco, Ca: Peason Education, 2002.

[12] O. K. Ersoy, Diffraction, Fourier Optics and Imaging, New Jersey: Wiley, 2006.

[13] G. J. Gbur, Mathematical Methods for Optical Physics and Engineering, Cambridge, GBR: Cambridge University Press, 2010.

[14] R. Bracewell, Fourier Analysis and Imaging, New York: Klewer Academic/Plenum Publishers, 2003.

[15] E. H. Linfoot, Fourier Methods in Optical Image Evaluation, London: The Focal Press, 1964.

[16] P. H. van Cittert, "Die Wahrscheinliche Schwingung Verteilung in Einer von Einer Lichtquelle Direkt Oder Mittels Einer Linse Beleuchteten Ebene," *Physica,* vol. 1, p. 201, 1934.

[17] R. D. Guenther, Modern Optics, New York: Wiley, 1990.

[18] B. T. Young, *Phase Retrieval Using Estimation Methods for Intensity Correlation Imaging (Master's thesis)*, College Station, Tx: Texas A&M University, 2010.

[19] W. Clegg, Crystal Structure Determination, Cary, NC: Oxford University Press, 1998.

[20] H. H. Rose, "Optics of High-Performance Electron Microscopes," *Science and Technology of Advanced Materials,* vol. 9, no. 014107, pp. 1-30, 2008.

[21] J. R. Fienup, "Phase Retrieval Algorithms: A Comparison," *Applied Optics,* vol. 21, no. 15, pp. 2758-2769, 1 Aug 1982.

[22] H. A. Arsenault and K. Chalasinska-Macukow, "The Solution to the Phase Retrieval Problem Using the Sampling Theorem," *Optics Communications,* vol. 47, no. 6, pp. 380-386, Oct 1983.

[23] J. R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform," *Optics Letters,* vol. 3, no. 1, pp. 27-29, July 1978.

[24] R. W. Gerchbert and W. O. Saxton, "A Practical Algorithm for the Determination of the Phase from Image and Diffraction Plane Pictures," *Optik,* vol. 35, p. 237, 1972.

APPENDIX A

PHASE RETRIEVAL PSEUDO-CODES

The following are the pseudo-codes for the phase retrieval algorithms discussed in section 4. They contain some of the typical function names used in Matlab.

Algorithm 1. Standard Error-Reduction Algorithm with Collapsing Foreground

Input: The measured Fourier transform magnitudes stored in, *fourier_modulus*.
Output: The estimated image, *g_out*.

```
g_in = random matrix    //Create the initial guess of the image
while rms_error < some convergence criteria
    // Step 1 – Fourier Transform
    G = fft(g_in)

    // Step 2 – Impose Fourier Constraints
    G = fourier_modulus * angle(G)

    // Step 3 – Inverse Fourier Transform
    g_out = ifft(G)

    // Step 4 – Impose image constraints
    foreach (x,y)
        if g_out(x,y) > 1
            g_out(x,y) = 1
        if g_out(x,y) < 0
            g_out(x,y) = 0
        if (x,y) is in the background region
            g_out(x,y) = 0

    //Quantify error
    rms_error = sum( (g_out – g_in).^2 )

    //Move background boundaries inward
    if each g_out(x,y) in check region < threshold
        Move background boundary inward
return g_out
```

Algorithm 2. Hybrid Error-Reduction Algorithm with Collapsing Foreground

Input: The measured Fourier transform magnitudes stored in, *fourier_modulus*.
Output: The estimated image, *g_out*.

```
g_in = random matrix    //Create the initial guess of the image
while rms_error < some convergence criteria
    // Step 1 – Fourier Transform
    G = fft(g_in)

    // Step 2 – Impose Fourier Constraints
    G = fourier_modulus * angle(G)

    // Step 3 – Inverse Fourier Transform
    g_out = ifft(G)

    // Step 4 – Impose image constraints
    foreach (x,y)
        if g_out(x,y) > 1
            g_out(x,y) = 1
        if g_out(x,y) < 0
            g_out(x,y) = 0
        if (x,y) is in the background region
            g_out(x,y) = g_in(x,y) – beta * g_out(x,y)

    //Quantify error
    rms_error = sum( (g_out – g_in)^2 )

    //Move background boundaries inward
    if each g_out(x,y) in check region < threshold
        Move background boundary inward
return g_out
```

Algorithm 3. Image Averaging Algorithm

Input: Estimated images from the collapsing foreground image, *g[i]*.
Output: The estimated image, *g_out*.

```
foreach g[i]
    //Step 1 – Find the center of mass of the image matrix
    for x=1..Nx
        cmY += sum(g[i][x,:] .* [1,2,3,..,Ny]) / sum(sum(g[i]))
    for y=1..Ny
        cmX += sum(g[i][:,y] .* [1,2,3,..,Nx]) / sum(sum(g[i]))

    g[i] = circshift(g[i],-[cmX-Nx/2 cmY-Ny/2])

g_out = mean(g)

return g_out
```

APPENDIX B

THE PHASE RETRIEVER SOFTWARE

The following are screenshots from the software written to implement the algorithms introduced in this thesis. The program is titled ‘The Phase Retriever.’ It is written in C# to implement many tools and interfaces available through the windows .Net framework. This allowed for minimal time to be spent programming the interface and allowed more effort to be focused on the inner workings and optimization of the solver.

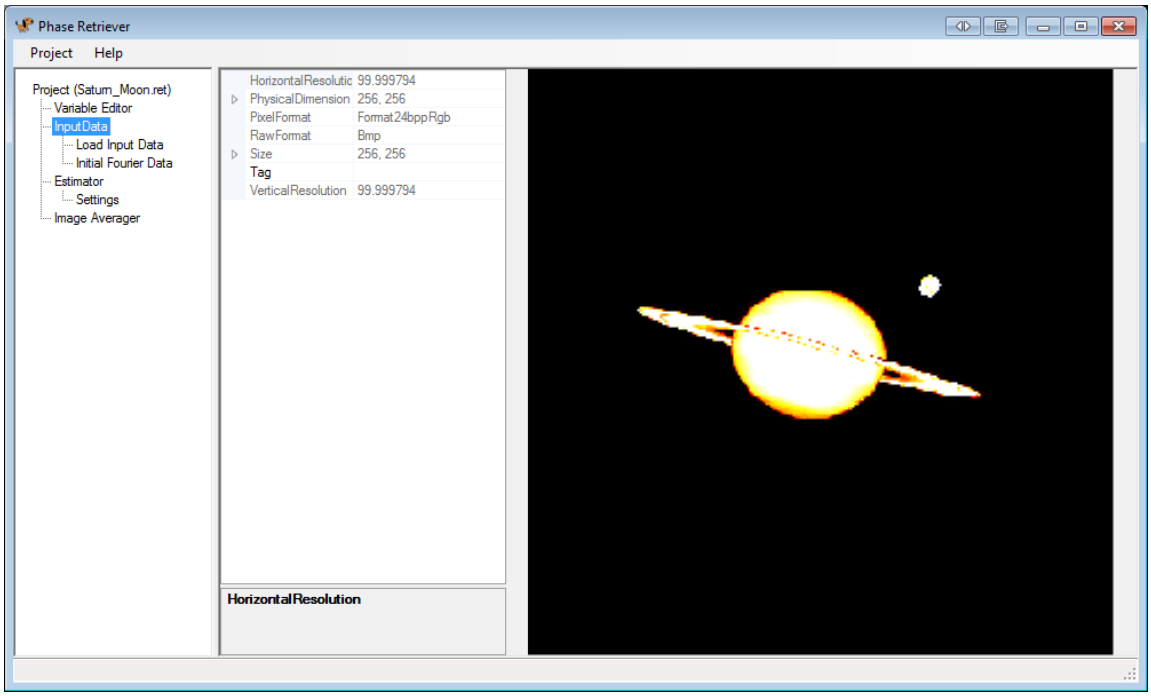


Fig. 38. The general Phase Retriever Interface. Shown are the input image and its properties.

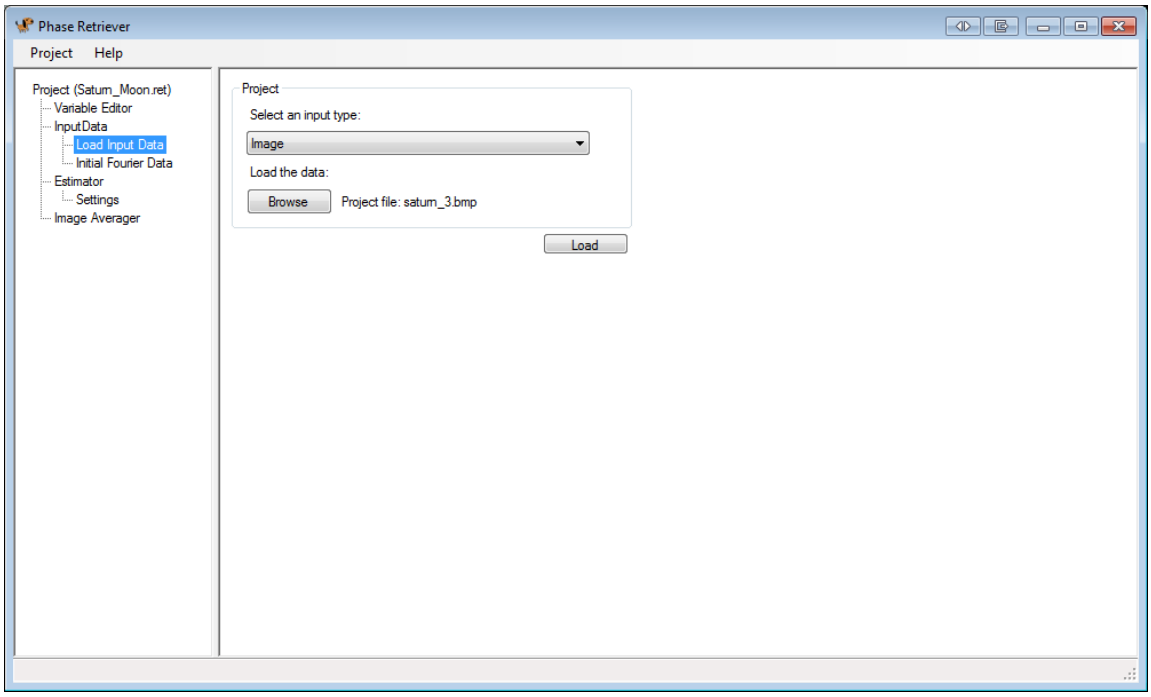


Fig. 39. Interface for loading images to create Fourier modulus data to import to the solver.

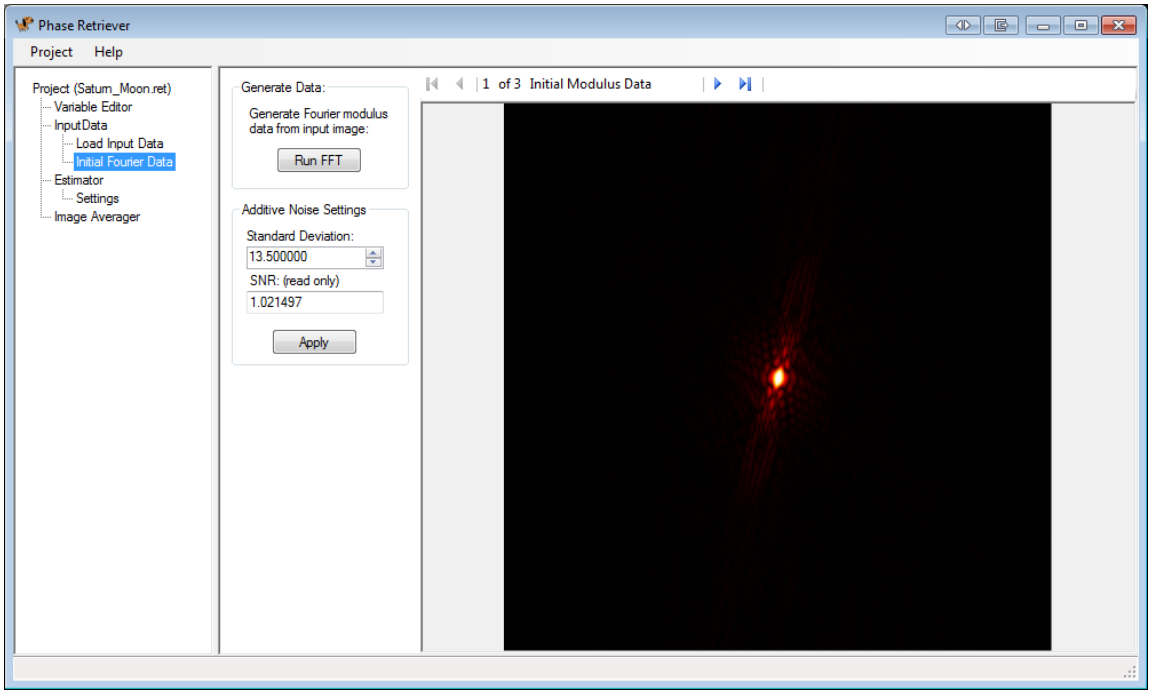


Fig. 40. Interface for running the FFT of the input image and adding noise to the modulus data.

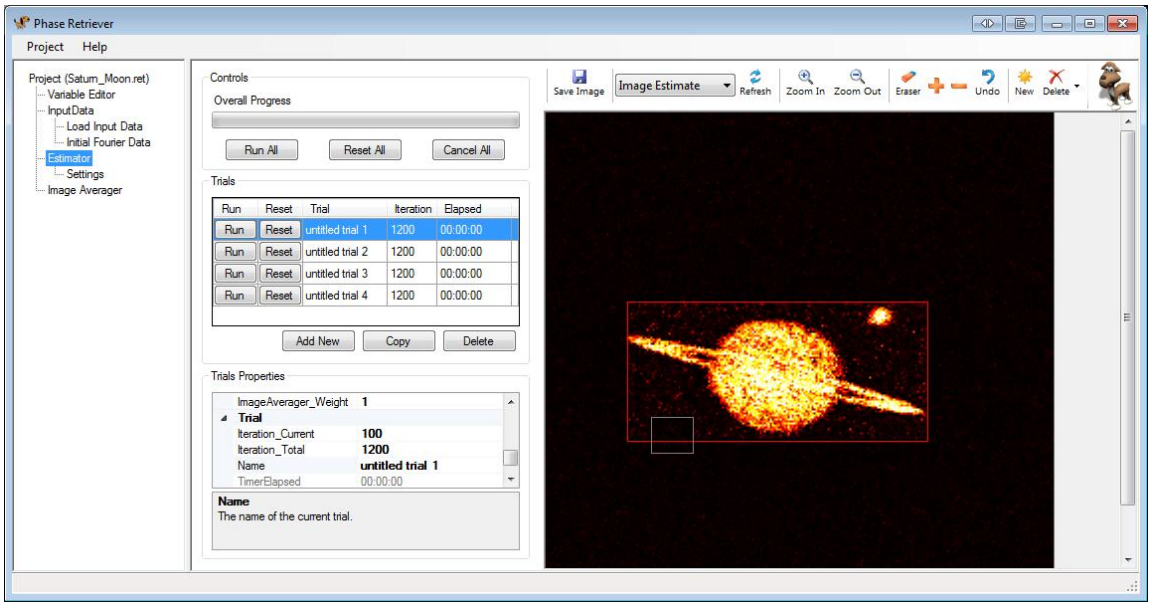


Fig. 41. Interface for running trials in the estimator. Includes tools for modifying the background region and editing the current iteration's image.

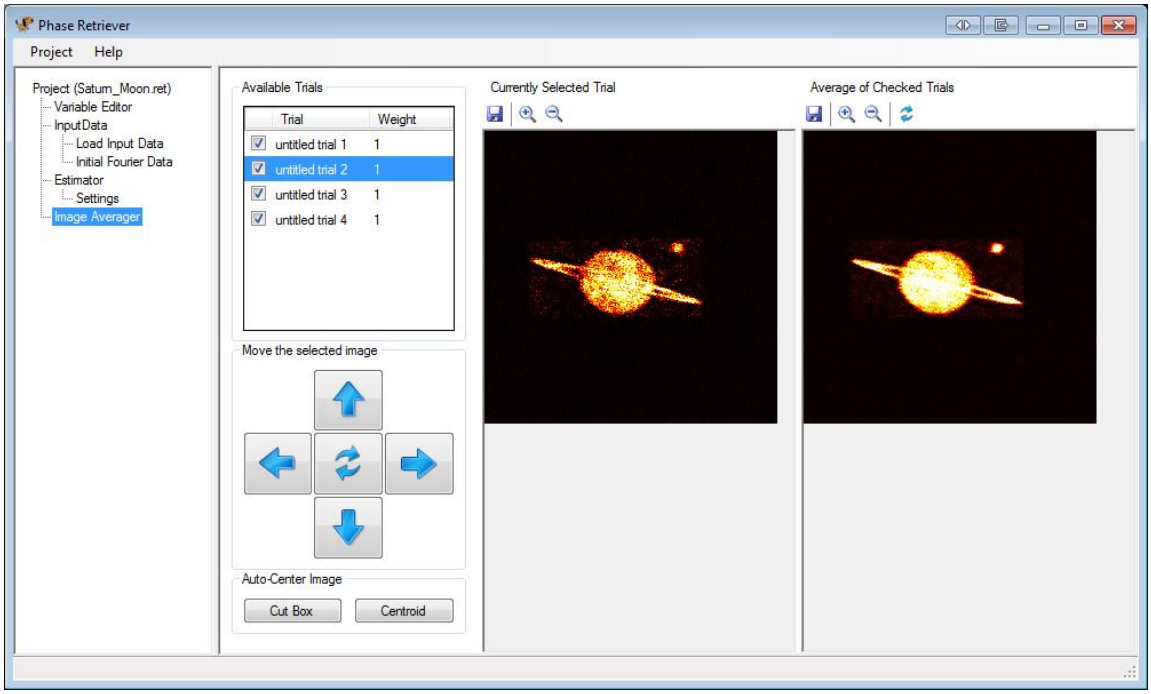


Fig. 42. Interface for performing image averaging. Buttons are included for translating and rotating the images from each trial.

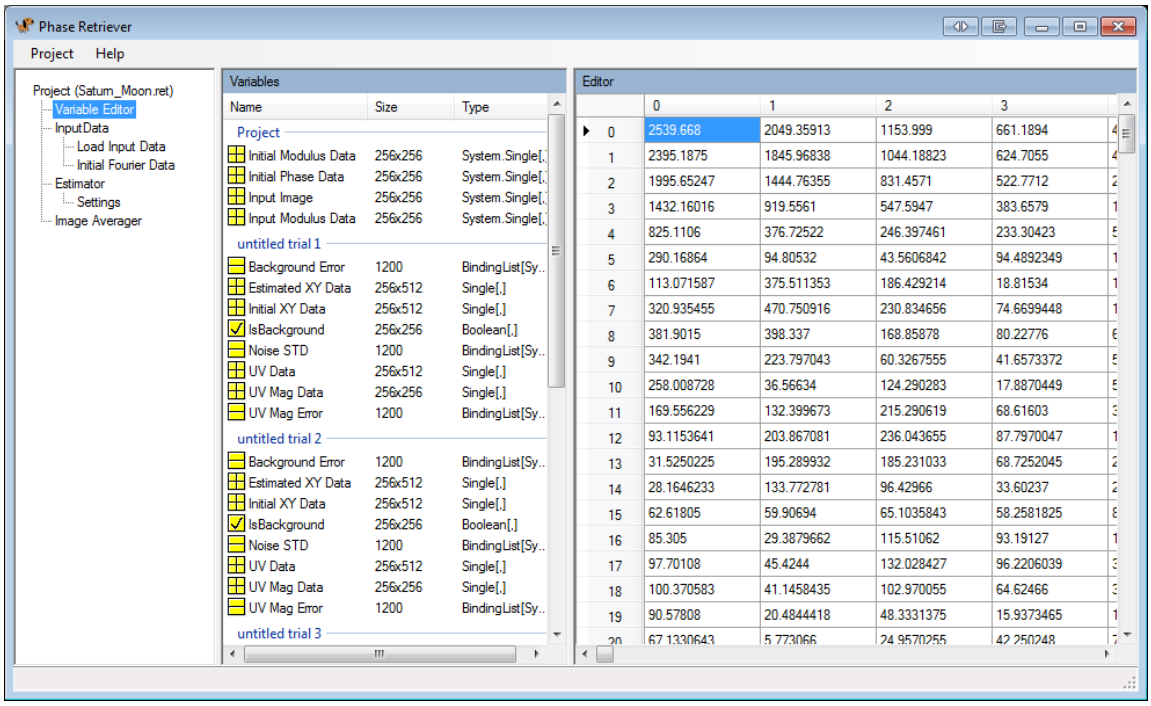


Fig. 43. Variable editor for viewing all working variables in each trial. This is useful for debugging and troubleshooting.