Engineering and Applied Science Theses & Dissertations

McKelvey School of Engineering

Spring 5-15-2016

Double Alternating Minimization (DAM) for Phase Retrieval in the Presence of Poisson Noise and Pixelation

Weimin Zhou
Washington University in St Louis

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Part of the Signal Processing Commons

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Science

Department of Electrical and Systems Engineering

Thesis Examination Committee:

Joseph A. O'Sullivan

Matthew D. Lew

Mark A. Anastasio

Double Alternating Minimization (DAM) for Phase Retrieval

in the Presence of Poisson Noise and Pixelation

by

Weimin Zhou

A thesis presented to the School of Engineering

of Washington University in St. Louis in partial fulfillment of the

requirements for the degree of

Master of Science

May 2016

Saint Louis, Missouri

Contents

List of Figures

List of Abbreviations

DAM :.. Double Alternating Minimization

DFT :..Discrete Fourier Transform

DTFT :.. Discrete-Time Fourier Transform

GFP :...Green Fluorescent Protein

G-S :...Gerchberg-Saxton algorithm

LS :...Least Square

MLE :.. Maximum Likelihood Estimation

NA :..Numerical Aperture

NNI/NI :..Nearest Neighbor Interpolation

PSF :... Point Spread Function

SMLM :...Single Molecule Localization Microscopy

SNR :..Signal to Noise Ratio

Acknowledgments

I would first like to thank my research advisor, who is also my academic advisor, Professor Joseph A. O' Sullivan. He spent many of his time to meet up with me and guided me in a right direction of my thesis works. Besides, he also gave me great advising for my future academic career.

I would also like to thank Professor Matthew D. Lew. He also spent lots of time for helping me to improve my thesis. Without his kindness help, the progress of my thesis would not be smooth.

<div align="right">Weimin Zhou</div>

Washington University in St. Louis
May 2016

ABSTRACT

Double Alternating Minimization (DAM) for Phase Retrieval

in the Presence of Poisson Noise and Pixelation

by

Weimin Zhou

Master of Science in Electrical Engineering

Washington University in St. Louis, 2016

Research Advisors: Dr. Joseph A. O'Sullivan

Dr. Matthew D. Lew

Optical detectors, such as photodiodes and CMOS cameras, can only read intensity information, and thus phase information of wavefronts is lost. Phase retrieval algorithms are used to estimate the lost phase and reconstruct an accurate effective pupil function, where the squared modulus of its Fourier transform is detected by a camera. However, current algorithms such as the Gerchberg-Saxton algorithm and Fienup-style algorithm do not consider the detector sampling rate and shot noise introduced by photon detection. If the sampling rate is low, we must interpolate the detected image in order to accurately reconstruct its pupil function. Here, we develop an appropriate estimation method for interpolating the detected image by using penalized I-divergence and then use the interpolated image for phase retrieval. In our simulation, after 300 iterations of our DAM algorithm, the phase-retrieved pupil function has a root-mean-squared error of about 43±3% less than Fienup-style algorithm with nearest neighbor interpolation when one hundred million photons are collected.

Chapter 1

Introduction

The following sections will introduce the motivation of this research and give some related background.

1.1   Motivation

The resolution of light microscopy is naturally limited by the diffraction barrier (around half of the wavelength), which was thought to be unbreakable for a long time. However, due to the key innovation of switchable fluorophores, the Single Molecule Localization Microscopy (SMLM) was born [1-4] and allows us to observe nanoscale details of biological structures and activities, without damaging living cells [5-8]. The other important aspect of SMLM is the powerful detection and estimation algorithm, which helps us to localize single molecules accurately and precisely [6]. In order to find the location of single molecules, there are two main methods used: Least Squares fitting and Maximum Likelihood Estimation (MLE). Both of these two methods need to use the predicted model of measured data, which is the defocused Point Spread Function (PSF) of an optical system. The Least Square algorithm is to find the parameters (in this problem it is the localizations of light emitters) by minimizing the squared error between measurements and the model. Thus, the model mismatch will generate inaccurate results, especially when we use weighted least squares, in which the mismatch of PSF tails has significant impact on the results [6]. The MLE algorithm estimates parameters by maximizing the likelihood of data given an accurate model. Besides, there are also other algorithms that need to use an accurate PSF model, such as DAOSTROM [9], which is to estimate high density excited molecules' positions in order to decrease the data acquisition time. Thus, in order to have a better estimation of locations of single molecules, the precise defocused PSFs have to be explored.

Ideally, we can calculate the PSF analytically from the pupil function in an optical system (since the PSF is the squared magnitude of a scaled Fourier transform of pupil function). However, the theoretically calculation does not account for aberrations in the optic system, so the PSF model is inaccurate for later use in localization algorithms [10]. Thus, in order to obtain more precise image models, we usually use calibration images that are obtained from the measurements of a single fluorophore on different pre-known axial positions. A popular way to reconstruct the experimental PSFs in continuous axial positions is to calculate the pupil function based on multiple defocused planes by using Fienup-style phase retrieval [11-13]. However, this method does not consider the Poisson noise of the measurements and the pixelation effects. Besides, this popular method is also lacking mathematical analysis. In this thesis, I invented a new method for solving phase retrieval problem in SMLM with specifically handling the Poisson noise and pixelation. The direct use of the Fienup-style algorithm [11-13] is just a special case in our generalized algorithm.

1.2   Background

The most important contribution of SMLM is allowing people to observe nanoscale structures in living cells and bacteria by breaking the diffraction limit of light microscopy. As an interdisciplinary technology, SMLM combines optics, chemistry, biology and information processing to help scientists pinpoint the positions of single molecules. The following are some important knowledge related to SMLM.

1.2.1   Abbe's diffraction limit and the principle of SMLM

Ever since the invention of microscope in 1590 by Hans and Zacharias Janssen [29], it has played an important role in the study of biology by allowing the visualization of tiny structures of cells and bacteria. For example, bacteria were first being observed by Anton Van Leeuwenhoek by using a microscope with one lens in 1675 [29]. In 1911, Oskar Heimstadt created the first fluorescence microscope to image auto-fluorescence within bacteria [30]. The invention of fluorescence microscope improved the contrast of biological images (backgrounds signals are eliminated by a dichroic mirror).

However, all optical imaging modalities are limited by the diffraction barrier---Abbe's diffraction limit. In 1873, Ernst Abbe first stated the relationship between microscopy resolution and wavelength [14]. Due to the limited size of a collection aperture, it is impossible to capture all of the light that is emitted from a point source. By Fourier optics, we know that any light distribution can be decomposed into a summation of an infinite number of plane waves, and each plane wave carries the information about Fourier transform of the image on object plane [27]. Thus, the effect of the limited size of aperture is equivalent to multiplying by a pupil or aperture function, which could be treated as some kind of window function or filter. This process implies that the impulse response of optical systems is related to the Fourier transform of the pupil function, which is an infinitely spread function on the image plane. For example, if the pupil is a circular function with diameter D, then the impulse response of the focused optical system with focus f and light wave length λ is an Airy disk with radius $\rho_A = 1.22\lambda \frac{f}{D} \approx \frac{1.22\lambda}{2NA} = \frac{0.61\lambda}{NA}$, where NA is the numerical aperture. This equation defines diffraction limit that defines the resolution of optical systems. The diffraction limit has always been thought to be impossible to break. Fortunately, Single Molecule Localization Microscopy has since been developed. In 1997, Professor W. E. Moerner observed the blinking behavior of a mutant of Green Fluorescent Protein (GFP). Excited by 488 nm light, GFP emitted fluorescence for some cycles, and after blinking for some time, GFP transitioned to a dark state [15]. Fluorescence was reactivated by illuminating the sample with 405 nm light [15]. This property of fluorescent proteins provides a new way to overcome the diffraction limit. After labelling biological samples with fluorescent proteins, we can activate a sparse set of proteins at any point in time, and since the activated molecules are well separated, the locations of each single molecule can be estimated accurately and precisely far beyond the diffraction limit. Then we repeat this process many times to obtain a series of molecule locations. Finally, all of these locations can be combined together to form a super-resolved reconstruction of the structure of interest.

1.2.2   Estimation methods for localizing single molecules.

The estimation process for locating positions of single molecules is comparing the received data with the image model. The decision rules for solving this estimation problem could be either least squares (LS) criterion or maximum-likelihood criterion (ML). LS criterion needs no information about the noise in the received signals; it aims to minimize the difference between the vectors of measurements and models in parameter space. In addition, the least squares are often weighted by dividing the expected variance of data [6]. This strategy utilizes the idea that the error should have more weight if the variance of the received signal is small. This assumption is reasonable since the Poisson noise model indicates that small signals have small variance. Thus, inaccurate modeling of PSF tails will introduce problems since the weight is large, affecting the estimated location. The other method, ML estimation, is widely used because it can theoretically approach the Cramer-Rao lower bound (CRLB) [7]. The principle of ML is to maximize the likelihood of the measured data, given the parameters of the imaging model. Thus, mismatch of the image formation model will lead to inaccurate estimations.

1.3   Our contributions

One of the most popular methods for calculating experimental PSFs is the Fienup-style phase retrieval algorithm. To my best knowledge, the first use of this Fienup-style method was in [12] and it is the modified version of Gerchberg-Saxton algorithm [12]. However, the rigorous mathematic validation of monotonicity for this method used with multiple defocused Fourier modulus constraints is still lacking. Our first contribution is that we discussed a rigorous mathematic framework to prove the feasibility of the Fienup-style method. Second, due to the size of detectors cannot be ignored, the measurements are the results of pixelation. Thus, if the detector size is large, the measurement vector will have lower dimension than the reconstruction vector, and this digitization and sampling process will introduce aliasing of the spatial domain function (and the measurements are in frequency domain), as discussed in Chapter 2. Under this scenario, the Fienup-style phase retrieval algorithm cannot be directly used since Fourier constraint information is incomplete. Thus, before the use of the Fienup-style phase retrieval method, we need first to interpolate the Fourier modulus data that is missed

during the measurement process. In order to overcome this problem, we use the minimum I-divergence estimation by taking the Poisson noise of received photons and pixelation of detectors into account, and then we obtain more precise model of experimental PSF.

Chapter 2

Alternating Projections for Phase retrieval problem

This chapter discusses two main methods for phase retrieval: Gerchberg-Saxton algorithm and the Fienup-style algorithm.

2.1   Phase Retrieval Problem

In Fourier optics, according to Fresnel approximation [27] discussed in Section 2.1.2, light waves transmitted through a single lens will generate the scaled Fourier transform of original input signals. Ideally, if we can measure the Fourier transform from output, we can restore input perfectly. However, the measurement process is not perfect. Due to the fact that optical detector can only measure the intensity of light, which is the squared magnitude of the defocused Fourier transform, the phase information is lost.

2.1.1   Importance of phase information

Phase determines the fundamental shape of the image since the phase carries information about the displacement of each sinusoid signals, which are the base functions for constructing image. The magnitudes of the Fourier transform, which construct the Fourier spectrum, are responsible for the intensity distribution in the image. The effects of Fourier phase and Fourier spectrum are shown in Figure 2.1. By comparing Figure 2.1 (b) and (c), which are reconstructed by only Fourier modulus and Fourier phase respectively, we can observe that the shape information of the image is restored by using phase of Fourier transform. Then, I introduce a new image which is the face of a mandrill. I calculated the Discrete Fourier Transform (DFT) of both Lena and Mandrill images and then switched

their Fourier spectrum and phase angle with each other. Figure 2.1 (e) shows the reconstructed image from the combination of Lena's Fourier spectrum and Mandrill's phase, and it looks like just a Mandrill's face with some artifacts. Figure 2.1 (f) shows the reconstruction from combination of Mandrill's magnitude and Lena's phase, and it looks like Lena image with some artifacts. Thus, it is reasonable to say that the phase information of images has an important influence on image reconstruction.
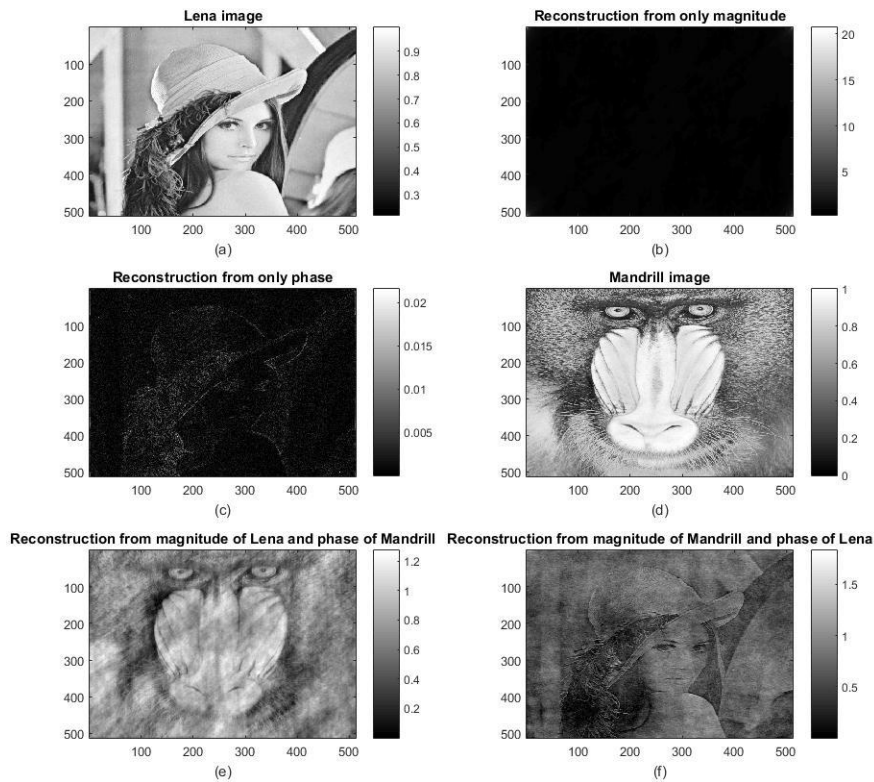
Figure 2.1 Image reconstruction from different parts and combinations of Fourier transforms. (a) Original Lena image; (b) Reconstruct image by only using Fourier Spectrum; (c) Reconstruct image by using phase and set the Spectrum equal to one; (d) Original Mandrill image; (e) Reconstructed image by combining the Fourier spectrum of Lena and phase from Mandrill; (f) Reconstructed image by combining the Fourier spectrum of Mandrill and phase from Lena.

2.1.2 Phase Retrieval problem in SMLM

In this sub-section, I will introduce some basic knowledge of Fourier optics for deriving the Point Spread Function of SMLM. Then I will discuss Gerchberg-Saxton algorithm and the Fienup-style algorithm for solving phase retrieval problem.

2.1.2.1 Fourier Optics

In all parts of this paper, with all mention of lights and optics, we are only concerned with the monochromatic light, which is represented by the complex amplitude:

$$U(r) = a(r)\exp[\theta(r)], \tag{2.1}$$

the corresponding complex wave function is:

$$u(r,t) = U(r)\exp(-j2\pi vt). \tag{2.2}$$

In this equation, r represents position in space. The magnitude of complex amplitude is $|U(r)| = a(r)$, the phase is its argument $\arg\{U(r)\} = \theta(r)$. The optical intensity is $I(r) = |U(r)|^2$.

Suppose a plane wave $U(x, y, z) = A\exp[j(kz)]$ travels through a plane with transmittance $f(x, y)$, where k is a wavenumber $k = \frac{2\pi}{\lambda}$ and λ is the wavelength. Taking the Fourier transform of $f(x, y)$, we can get

$$F(u, v) = \iint_{x,y} f(x, y)e^{-j2\pi(xu+yv)}dxdy,$$

$$f(x, y) = \iint_{u,v} F(u, v)e^{j2\pi(xu+yv)}dudv. \tag{2.3}$$

The wave complex amplitude immediately after the transmittance plane is:

$$U(x, y, 0) = A\iint_{u,v} F(u, v)e^{j2\pi(xu+yv)}dudv. \tag{2.4}$$

Where A is just the amplitude. Then, the transmitted wave will be

$$U(x, y, z) = A \iint_{u,v} F(u, v) e^{j2\pi(xu+yv)} e^{jk_z z} dudv, \qquad (2.5)$$

where $k_z = 2\pi\sqrt{\lambda^{-2} - u^2 - v^2}$. Thus, the transmittance separates the incident plane wave into infinite many plane waves with angles for each frequency u, v to be:

$$\varphi_x = \sin^{-1}\lambda u \approx \lambda u,$$

$$\varphi_y = \sin^{-1}\lambda v \approx \lambda v. \qquad (2.6)$$

2.1.2.2 Fourier transform by using single lens

Suppose there is a single lens system shown in Figure 2.2, where f is the focal length and d is the distance from input plane to the lens; $f(x, y)$ is the input and $g(x, y)$ is the output.

Lens

d

f

z

f(x, y)

g(x, y)
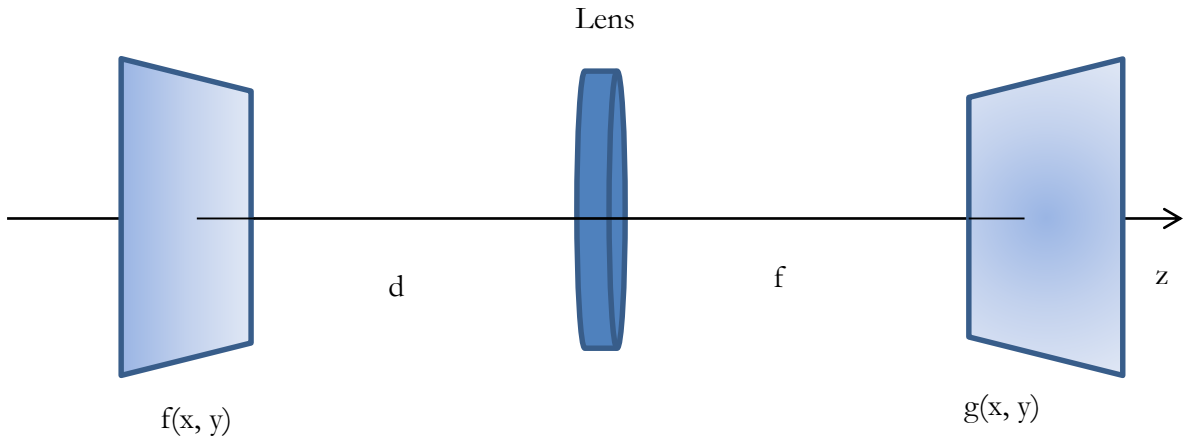
Figure 2.2 Single lens microscopy model. $f(x, y)$ is the transmittance on input plane and $g(x, y)$ is the wave complex amplitude on output plane.

According to the discussion in Section 2.1.2.1, the input wave is decomposed into different plane waves with different angles. Thus we can approximate the process of transmitting waves through a single lens as the superposition of separated processes for each plane wave. The transmittance of lens

is $t(x, y) \approx h_t \exp\left[-jk\frac{x^2+y^2}{2f}\right]$, where $h_t = e^{-jnkd_0}$, d_0 is the thickness of the lens, and n is the refractive index.

Lemma 2.1:

The plane waves transmitted through the lens will be transferred into a parabolic wave converging to a point on back focal plane. The proof is shown below.

Proof:

Assume the lens is located at the position $z = -f$. Then the incident plane wave is:

$$U(x, y, -f) = A\exp(-jkf). \tag{2.7}$$

The output of the lens will be:

$$U_{lens}(x, y, -f) = Ah_t\exp(-jkf)exp\left[-jk\frac{x^2+y^2}{2f}\right]. \tag{2.8}$$

Since the Fresnel approximation of the diverging spherical wave is:

$$U(x, y, z) \approx \frac{A}{z}\exp(jkz)exp\left[jk\frac{x^2+y^2}{2z}\right], \tag{2.9}$$

the waves outputted from the lens will converge to a point at the focal plane. In this case, the input plane wave is parallel with the optic axis, so the convergent point on focal plane is $(0,0)$. Let us now assume that the lens is placed at position $z = 0$, and then the result of the convergence shows that the superposition of all decomposed plane waves will converge to the central point on the focal plane $z = f$. If we rotate the plane wave from the input to the incident angle $(\varphi_{x0} \approx \lambda u_0, \varphi_{y0} \approx \lambda v_0)$, then we can treat this problem as a modulating transmittance of lens:

$$t_m(\mathrm{x}, \mathrm{y}) = At(x, y) \times \exp(j2\pi(u_0 x + v_0 y)$$

$$= A \iint\limits_{u,v} F(u, v)e^{j2\pi(x(u+u_0)+y(v+v_0))}dudv \tag{2.10}$$

From Equation (2.10), we can observe that each plane wave components of transmittance have also been rotated by angle $(\varphi_{x0}, \varphi_{y0})$. Thus, the superimposition of these rotated components will also

rotate at the same angle; under the paraxial approximation, the convergence point is relocated at $(\varphi_{x0}f \approx \lambda f u_0, \varphi_{y0}f \approx \lambda f v_0)$. The output will be [27]:

$$g(x,y) \propto F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \tag{2.11}$$

The following expression then gives the precise expression for the output of a single lens system [27]:

$$g(x,y) = h_l \exp\left[-j\pi \frac{(x^2+y^2)(d-f)}{\lambda f^2}\right] F\left(\frac{x}{\lambda f}, \frac{y}{\lambda f}\right). \tag{2.12}$$

Where $h_l = -\frac{j}{\lambda f}\exp(jk(d+f))$.

To prove this, we need to first introduce a lemma.

Lemma 2.2:

$$\int_{-\infty}^{+\infty} f(x,y)\exp\left[j\pi \frac{(x'-x)^2 + (y'-y)^2}{\lambda d}\right]dxdy$$

$$= \exp(j\pi \frac{x'^2 + y'^2}{\lambda d})\mathcal{F}\left[f(x,y)\exp\left(j\pi \frac{x^2+y^2}{\lambda d}\right)\right]\Big|_{u=\frac{x'}{\lambda d}, v=\frac{y'}{\lambda d}} \tag{2.13}$$

This shows that the convolution of $f(x,y)$ and $\exp\left(j\pi \frac{x^2+y^2}{\lambda d}\right)$ is in the form of the Fourier transform of the product of them.

Proof:

$$\mathcal{F}\left[f(x,y)\exp\left(j\pi \frac{x^2+y^2}{\lambda d}\right)\right]\Big|_{u=\frac{x'}{\lambda d}, v=\frac{y'}{\lambda d}}$$

$$= \iint f(x,y)\exp\left(j\pi \frac{x^2+y^2}{\lambda d}\right)\exp\left(-j\pi \frac{(2xx'+2yy')}{\lambda d}\right)dxdy$$

$$= \exp(-j\pi \frac{x'^2+y'^2}{\lambda d})\iint f(x,y)\exp\left(j\pi \frac{(x-x')^2 + (y'-y)^2}{\lambda d}\right)dxdy \tag{2.14}$$

Multiply it by $\exp\left(j\pi \frac{x'^2+y'^2}{\lambda d}\right)$ will complete this proof:

11

$$\exp\left(j\pi \frac{x'^2 + y'^2}{\lambda d}\right) \mathcal{F}\left[f(x,y)\exp\left(j\pi \frac{x^2 + y^2}{\lambda d}\right)\right]\Big|_{u=\frac{x'}{\lambda d}, v=\frac{y'}{\lambda d}}$$

$$= \iint f(x,y)\exp\left(j\pi \frac{(x-x')^2 + (y'-y)^2}{\lambda d}\right)dxdy \qquad (2.15)$$

Now, we can move forward to the proof of Fourier transform property of a single lens.

Proof of the Fourier transform by a single lens:

Suppose the input of the system is $f(x)$. The waves firstly transmit distance d in free space to arrive at the lens. According to Fresnel Approximation, the impulse response of free space is [27]:

$$h(x,y) \approx h_0 \exp\left(jk\frac{x^2 + y^2}{2d}\right),$$

$$(2.16)$$

$$h_0 = \left(-\frac{j}{\lambda d}\right)\exp(jkd).$$

Thus, the complex amplitude of the wave immediately before the lens is:

$$g_{lens-in}(x',y') = \int\int h_0 f(x,y)\exp[jk\frac{(x'-x)^2 + (y'-y)^2}{2d}]dxdy$$

$$= h_0 f(x,y) * \exp\left(j\pi \frac{x'^2 + y'^2}{\lambda d}\right). \qquad (2.17)$$

After waves transmitted through the lens, the complex amplitude will multiply by the transmittance of the lens, so the wave immediately after the lens is:

$$g_{lens-out}(x',y') = h_t \exp\left[-j\pi\left(\frac{x'^2 + y'^2}{\lambda f}\right)\right] h_0 f(x,y) * \exp\left(j\pi \frac{x'^2 + y'^2}{\lambda d}\right). \qquad (2.18)$$

Then, propagate $g_{lens-out}(x')$ in free space of distance f, we will get the output of the system:

$$g_{out}(x'',y'') = h_1 g_{lens-out}(x'',y'') * \exp\left[jk\frac{x''^2 + y''^2}{2f}\right]. \qquad (2.19)$$

From Lemma 2.2:

$$g_{out}(x'', y'')$$

$$= h_1 \exp\left[jk \frac{x''^2 + y''^2}{2f}\right]$$

$$\times \int g_{lens-out}(x', y') \exp\left[j\pi \frac{x'^2 + y'^2}{\lambda f}\right] \exp\left[-j2\pi \frac{x'x'' + y'y''}{\lambda f}\right] dx'dy'$$

$$\approx h_1 \exp\left[jk \frac{x''^2 + y''^2}{2f}\right] \mathcal{F}[g_{lens-in}(x', y')]\Big|_{u=\frac{x''}{\lambda f}, v=\frac{y''}{\lambda f}}. \tag{2.20}$$

Where $h_1 = \left(-\frac{j}{\lambda f}\right) \exp(jkf)$. The approximation holds when we suppose the thickness of lens is sufficiently small, $h_t \to 1$.

Then we calculate the Fourier transform of $g_{lens-in}(x', y')$:

$$\mathcal{F}[g_{lens-in}(x', y')] = h_0 \mathcal{F}[f(x', y')] \mathcal{F}\left[\left(\exp\left[j\pi \frac{x'^2 + y'^2}{\lambda d}\right]\right)\right]. \tag{2.21}$$

The function $\exp[j\pi \frac{x'^2}{\lambda d}]$ is a scaled Chirp function, so the Fourier transform of it is [27]:

$$\mathcal{F}\left[\left(\exp\left[j\pi \frac{x'^2}{\lambda d}\right]\right)\right] = e^{\frac{j\pi}{4}} \sqrt{\lambda d} \exp(-j\pi\lambda du^2). \tag{2.22}$$

Thus,

$$\mathcal{F}\left[\left(\exp\left[j\pi \frac{x'^2 + y'^2}{\lambda d}\right]\right)\right] = j\lambda d \exp(-j\pi\lambda d(u^2 + v^2)). \tag{2.23}$$

$$\mathcal{F}[g_{lens-in}(x', y')] = h_0 \mathcal{F}[f(x', y')] j\lambda d \exp(-j\pi\lambda d(u^2 + v^2)) \tag{2.24}$$

Multiply it in Equation (2.19):

$$g_{out}(x'') =$$

13

$$h_1 \exp\left[jk\frac{x''^2 + y''^2}{2f}\right] j\lambda d\left(-\frac{j}{\lambda d}\right) \exp\left(j\frac{2\pi d}{\lambda}\right) \exp(-j\pi\lambda d(u^2$$

$$+ v^2))F\left(\frac{x''}{\lambda f}, \frac{y''}{\lambda f}\right)$$

$$= h_1 \exp\left(j\frac{2\pi d}{\lambda}\right) \exp\left[j\pi\frac{x''^2}{\lambda f} - j\pi\frac{dx''^2}{\lambda f^2}\right] \exp\left[j\pi\frac{y''^2}{\lambda f} - j\pi\frac{dy''^2}{\lambda f^2}\right] F\left(\frac{x''}{\lambda f}, \frac{y''}{\lambda f}\right)$$

$$= \left(-\frac{j}{\lambda f}\right) \exp[jk(f+d)] \exp\left[-j\pi\frac{(d-f)(x''^2 + y''^2)}{\lambda f^2}\right] F\left(\frac{x''}{\lambda f}, \frac{y''}{\lambda f}\right).$$

$$= h_l \exp\left[-j\pi\frac{(d-f)(x''^2 + (y''^2)}{\lambda f^2}\right] F\left(\frac{x''}{\lambda f}, \frac{y''}{\lambda f}\right). \tag{2.25}$$

Then Equation (2.12) has been proved.

2.1.2.3 Imaging model for phase retrieval

In SMLM, in order the variance of estimated 3D positions of single emitters to approach the Cremer-Rao lower bound [11], we usually add an engineered pupil mask on the pupil plane, which is the image plane of the first single lens system in SMLM. There are some popular pupil masks that can be used to generate such PSFs such as astigmatism PSF and double-helix PSF [16] [17]. Also, the method for optimizing pupil masks over Zernike polynomial parameters [18] for achieving the Cremer-Rao lower bound and the corresponding resultant Cat Mask PSF can be found in Supplemental material of [11]. Ideally, we can know the designed pupil mask exactly, and thus we can theoretically calculate the defocused PSFs for use in the estimation process of locating single molecules. However, due to the imperfection of optical systems, and there may exist some rays that do not satisfy paraxial approximation, the wavefronts aberrations will be inevitably introduced. The image model for estimating the locations of emitters is thus not optimal if we do not account for the wavefronts aberration. Therefore, the more realistic image models need to be discovered. We can obtain the accurate optical system model by knowing the overall effect of the phase term, which is the

14

multiplication of the aberrated phase distortion and the pupil mask. In this paper, we used effective pupil function to represent the combination of an engineered pupil mask and optical aberrations. In order to recover the effective pupil function, we need to know the phase information of the wavefronts. However, the data measured by detectors lose the phase information. Therefore, the phase retrieval algorithm is needed for solving this problem.

In fluorescent microscopy, 4-f system is used, which is constructed by cascading two 2-f systems. The principle of 4-f system is that we can reconstruct the original image by taking the inverse Fourier transform (by using second 2-f system) of its Fourier transform (by using first 2-f system). If we move the fluorophore a small distance along the optical axis, the system becomes a cascade of a single lens system and a 2-f system, and then we can observe the defocused response on the image plane. From Equation (2.25), we can derive the expression of defocused PSFs. Suppose the fluorophore is placed at distance z_k from the focus position, and then the output of the wave will be [27]:

$$g_k(x',y') = h_{l2}\mathcal{F}\left\{h_{l1}\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]p(x'',y'')\right\}\Big|_{u=\frac{x'}{\lambda f},v=\frac{y'}{\lambda f}}$$

$$h_{l1} = \left(-\frac{j}{\lambda f}\right)\exp[jk(2f+z_k)]$$

$$h_{l2} = \left(-\frac{j}{\lambda f}\right)\exp[jk(2f)]$$

(2.26)

Thus, we can simplify it as:

$$g_k(x',y') \propto \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]p(x'',y'')\right\}$$

(2.27)

The measurement at the detector plane is then given as:

$$I_k(x',y') \propto \left|\mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]p(x'',y'')\right\}\right|^2,$$

(2.28)

where $k = 1,2,\ldots,K$ are ordered numbers for different axial positions of the point source. The series of these images $\{I_k(x',y'): k = 1,2,\ldots K\}$ is called calibration image for the use of the phase retrieval

15

algorithm. The problem then becomes reconstructing the effective pupil function by using the known calibration images.

2.1.2.4 Phase retrieval problem and conventional algorithms

The problem of reconstructing effective pupil function comes from the fact that optical detectors can only measure the intensity of light, so that the phase information is lost during the measurement process. If we can know the exact phase information of the measurements, the Fourier transform of effective pupil function can be identified, and then the effective pupil function can be reconstructed by using inverse Fourier transform. However, since the phase information is lost in detection process, we cannot restore the effective pupil function directly.

A single measured image from detector defines infinitely many pupil functions since we can assign arbitrary phase for each pixel to generate different Fourier transforms. All possible pupil functions, defined by this single measurement, are in a subset of the pupil function space, which is defined by possible functions. By introducing other measurements from propagations at different defocus positions, different subsets of constraints are defined. The goal is then to retrieve the pupil function that can comply with all measurement constraints. Thus, the intersection of all subsets defined by each constraint is the solution.

Before we continue to discuss the phase retrieval problem, let us discuss the relationship between the projection operator and the least square problem.

In least square problems, suppose the measurement is $Y = (y_1, y_2, y_3, \dots, y_n)^T$, we want to find the parameter $\theta = \{\theta_1, \theta_2, \theta_3 \dots, \theta_m\}$ to minimize the squared error between Y and $X = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_m x_m$. Let $T = \{x_1, x_2, \dots, x_m\}$ be a set of independent n-dimensional vectors. This problem can be treated as finding the element in the subspace $V = \text{span}(T)$ that has the minimum distance to the measurement vector Y. Since the projection of a vector onto a space is defined as the minimum distance between vector and space, the least square problem is equivalent to find the projection of the measurement onto the model space.

Let us go back to our pupil retrieval problem and introduce the famous Gerchberg-Saxton algorithm [19] [20]. In Gerchberg-Saxton algorithm, two constraints are introduced. One is the function

constraint at spatial domain; the other is the Fourier constraint defined by the modulus of the Fourier transform at the frequency domain. Suppose that the constraint of the function domain defines a set of functions that the modulus must equal to $|f(x,y)|$, the Fourier constraint defines another set of functions that the modulus of Fourier transform have to be $|F(u,v)|$. Then there are four main steps in each iteration to solve this phase retrieval problem [20]:

(1) The estimated function at m-th iteration of function domain is inputted:

$$f^{(m)}(x,y) = |f(x,y)|e^{j\theta^{(m)}(x,y)} \tag{2.29}$$

Fourier transform it we get:

$$\mathcal{F}\{f^{(m)}(x,y)\} = G^{(m)}(u,v) = |G^{(m)}(u,v)|e^{j\varphi^{(m)}(u,v)} \tag{2.30}$$

(2) Force the magnitude of $\mathcal{F}\{f^{(m)}(x,y)\}$ to be the Fourier constraint $|F(u,v)|$ and then get the revised Fourier transform:

$$\hat{G}^{(m)} = |F(u,v)|e^{j\varphi^{(m)}(u,v)} \tag{2.31}$$

(3) Inverse Fourier transform of $\hat{G}^{(m)}$ and get the $\hat{f}^{(m)}$:

$$\hat{f}^{(m)}(x,y) = \mathcal{F}^{-1}\{\hat{G}^{(m)}\} = |\hat{f}^{(m)}(x,y)|e^{j\theta^{(m+1)}(x,y)} \tag{2.32}$$

(4) Force the function domain constraint to $\hat{f}^{(m)}$ and get the updated $f^{(m+1)}$:

$$f^{(m+1)}(x,y) = |f(x,y)|e^{j\theta^{(m+1)}(x,y)} \tag{2.33}$$

In the above algorithm, Step (2) and Step (4) are actually projection operators that minimize the squared error of $G^{(m)}(u,v)$ and $\hat{f}^{(m)}(x,y)$ with each corresponding constraint set respectively. The reason is followed by the following lemma.

Lemma 2.3

$$\min_{\varphi}|ae^{j\varphi} - be^{j\theta}|^2 = (a-b)^2 \tag{2.34}$$

Where a and b are real positive numbers, φ and θ are their corresponding phase angle. The optimal φ^* to let the equation holds is $\varphi^* = \theta$.

Proof:

Suppose the difference between φ and θ is δ, $0 \leq \delta \leq \pi$. According to Law of Cosines:

$$\left| ae^{j\varphi} - be^{j\theta} \right|^2 = a^2 + b^2 - 2ab \cos \delta \tag{2.35}$$

Since $\cos \delta \leq 1$, $a^2 + b^2 - 2ab \cos \delta \geq a^2 + b^2 - 2ab$. Thus, $\left| ae^{j\varphi} - be^{j\theta} \right|^2 \geq (a-b)^2$. The equation holds only when $\cos \delta = 1$, i.e. $\varphi^* = \theta$.

Thus, the Gerchberg-Saxton algorithm is an Alternating Projection algorithm. The Alternating Projection is widely used for solving problems of convex sets, and the proof for its convergence was shown in [22]. However, the Fourier constraints in phase retrieval problem are obviously not convex sets. The simple example for proving the non-convex property of Fourier constraints is shown below. Suppose the Fourier constraint is $|F(u,v)|$, assign an arbitrary phase angle $\varphi(u,v)$ we can get one element $G(u,v) = |F(u,v)|e^{j\varphi(u,v)}$ in the defined constraint set. We set another phase angle $\emptyset(u,v) = \varphi(u,v) + \pi$. Then we get another element $G'(u,v) = -|F(u,v)|e^{j\varphi(u,v)}$ in that Fourier constraint set. It is obviously that the sum of $G(u,v)$ and $G'(u,v)$ is just 0 and it is not in this Fourier constraint set. Therefore, the Fourier constraint sets are non-convex, and the convergence proof for convex Alternating Projection in [22] cannot be used in Gerchberg-Saxton algorithm. However, the Gerchberg-Saxton algorithm is also an error reduction algorithm, so we can prove the monotonicity by proving the property of the decreasing of the squared error. The proof of monotonicity of Gerchberg-Saxton algorithm is shown in Appendix A and it can also be found in [20].

For our phase retrieval problem discussed in this paper, there are multiple defocused Fourier constraints and no function constraint. Taking two Fourier constraints as an example, we can treat one as the function constraint, which defines a set of functions that satisfy the defocused Fourier modulus constraint. The Gerchberg-Saxton algorithm for our problem then can be extended to be [10]:

(1) In m-th iteration the input is $f^{(m)}(x,y)$ of which the Fourier transform satisfies corresponding modulus constraint.

(2) Multiply the first propagator and then take the Fourier transform of this defocused function.

(3) Replace the modulus of output from (2) by the corresponding constraint values.

(4) Inverse Fourier transform of (3) and divided it by corresponding propagator phase term.

(5) Multiply (4) by the second defocus propagator term and take Fourier transform.

(6) Replace the modulus of (5) by the corresponding constraint values, take Inverse Fourier transform and divided by second propagator to get the updated function $f^{(m+1)}(x,y)$.

We will observe that the performance of Gerchberg-Saxton algorithm is good for noise-free data. However, if the noise is in the presence, the Gerchberg-Saxton algorithm cannot converge to the desired solution. The simulation results will be given in Section 2.2.

From the basic Alternating Projection between constraints, there is an extension called Averaged Projection. This modified algorithm is just taking projections of the function into each constraint set and then average these projections to get the updated function. The principle of this averaged projection method can be understood as reformulating the minimization problem of finding a vector consistent with all constraints to the problem that find the multiple same vectors that consistent with corresponding constraints.

Let us define N constraint sets $\{C_1, \dots, C_N\}$, and matrix $M = (x_1, \dots, x_N)$, where each x_i is a column vector in the corresponding constraint set C_i: $x_i \in C_i$. The solution we are finding is $x_1 = x_2 = \cdots = x_N$. It is equivalent to find the intersection of sets:

$$C = \{x_1, \dots, x_N : x_1 \in C_1, \dots, x_N \in C_N\} \text{ and } E = \{x_1, \dots, x_N : x_1 = x_2 = \cdots = x_N\}.$$

Then, by using projections between C and E, we have:

$$\left\{ x_1^{(m+1)}, \dots x_N^{(m+1)} \right\} = P_E \left\{ P_C \left\{ x_1^{(m)}, \dots x_N^{(m)} \right\} \right\}$$

19

$$= \left\{ \frac{1}{N} \left(P_{C_1}\{x_1\} + \cdots + P_{C_N}\{x_N\} \right), \ldots, \frac{1}{N} \left(P_{C_1}\{x_1\} + \cdots + P_{C_N}\{x_N\} \right) \right\} \qquad (2.36)$$

For phase retrieval in SMLM, It can be shown from the Figure 2.4 that the Gerchberg-Saxton algorithm will have a faster convergence rate than the Fienup-style algorithm when no noise introduced in data. This result is consistent with the result of comparison between Projection onto Convex Set (POCS) and Parallel Projection Method (PPM) shown in [23]. The application of Gerchberg-Saxton algorithm for reconstructing experimental PSFs in SMLM can be found in [10]. However, in [10] they did not just take only one calibration image at each axial position, but instead they taken several calibration images at each position and averaged them to reduce the noise before input them into Gerchberg-Saxton algorithm. If the noise is presented in data, the Fienup-style algorithm performs better than Gerchberg-Saxton algorithm. From Figure 2.5, we can observe that the log of squared error resulted from Gerchberg-Saxton algorithm will decrease in the first several iterations and then tends to oscillate around a constant. This shows that Gerchberg-Saxton algorithm is severely impacted by noise. The intuitive reason is that since each of the measurements suffered from noise at each iteration, and the next projection is depended on the previous projection, so if, at each stage, there are some deviations from the noise, the final output may have large deviation and the squared error cannot be further reduced. In contrast, in Figure 2.5, we can see that the result of Fienup-style algorithm can always reduce the log of squared error even there is noise introduced, so the preferred phase retrieval method for reconstructing pupil mask in SMLM is Fienup-style averaged projection. If the Gerchberg-Saxton algorithm is used, the preprocessing of the measurements, such as taking multiple images at each defocused position and averaging them, is often used to reduce the noise effect before input the data into the algorithm [10]. Although there are many papers that used the averaged projection method, they did not give the relationship between this method and least square problem. Then, the proof of this robust performance of averaged projection will be given below by proving that the averaged projection is exactly the same as the least square problem. Before we prove this, we will firstly introduce a lemma, which was previously used in [24].

Lemma 2.4

Suppose a and b are real positive numbers, then

$$|a - b|^2 = \left|a - |b e^{j\theta}|\right|^2 = \min_{\varphi \in (-\pi, \pi]} \left|a e^{j\varphi} - b e^{j\theta}\right|^2 \tag{2.37}$$

$$\arg \min_{\varphi \in (-\pi, \pi]} \left|a e^{j\varphi} - b e^{j\theta}\right| = \theta \tag{2.38}$$

This can be proved directly by extending Lemma 2.3.

Then, we will prove the equivalence between averaged projection method and least square problem.

Proof:

First of all, let us give the least square problem:

$$\min_{f(x',y')} \left\{ \sum_k \sum_{x',y'} |\,|F_k(x',y')| - |\mathcal{F}_k[f](x',y')|\,|^2 \right\} \tag{2.39}$$

Where $F_k(x', y')$ is k-th Fourier constraint, operator $\mathcal{F}_k[f]$ is the k-th defocused Fourier transform: $\mathcal{F}\left\{\exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f(x'', y'')\right\}$, where $f(x'', y'')$ is the pupil mask. Direct solving this equation is very hard because the absolute operator is used for Fourier transform. In order to simplify if, we can rewriting the above least square problem by using Lemma 2.4:

$$\min_{f(x',y')} \left\{ \sum_k \sum_{x',y'} |\,|F_k(x',y')| - |\mathcal{F}_k[f](x',y')|\,|^2 \right\}$$

$$= \min_{f(x'',y'')} \min_{\varphi_k(x',y')} \sum_k \sum_{x',y'} \left| |F_k(x',y')| e^{j\varphi_k(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f(x'', y'')\right\} \right|^2 \tag{2.40}$$

Thus, the least square problem has been lifted to alternating minimization problem. Suppose we know $f(x'', y'')$, then the problem is to find $\varphi_k(x', y')$. According to Lemma 2.4, this should be the phase

angle of $\mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]f(x'',y'')\right\}$. Then by the updated $\varphi_k(x',y')$, we can find the updated $f(x'',y'')$ by solving the normal least square problem:

$$f^* = \underset{f(x'',y'')}{\operatorname{argmin}} \sum_k \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]f(x'',y'')\right\}\right|^2$$

$$(2.41)$$

Writing it as the matrix form and approximate the Fourier transform as the Discrete Fourier Transform, then we have:

$$\sum_k \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]f(x'',y'')\right\}\right|^2$$

$$= \left\| \begin{bmatrix} F_1 e^{j\varphi_1} \\ F_2 e^{j\varphi_2} \\ \dots \\ F_K e^{j\varphi_K} \end{bmatrix} - \begin{bmatrix} HD_1 \\ HD_2 \\ \dots \\ HD_K \end{bmatrix} f \right\|^2.$$

$$(2.42)$$

Where $F_1, F_2, \dots, F_K$ are column vectors reshaped from $F_k(x',y')$ column by column; φ_k and f are also vectored by the same way. $H \in C^{N\times N}$ is the forward matrix that calculate the 2D Fourier transform of f, so it has the property:

$$H^+H = NI. \qquad (2.43)$$

Where H^+ is the conjugate transpose of H, I is the identity matrix and N is the total number of pixels in the detector plane. $D_k \in C^{N\times N}$ is the diagonal matrix that account for the defocused propagation phase,

$$D_k = \operatorname{diag}\left(\exp\left[-j\pi\frac{z_k(x''^2+y''^2)}{\lambda f^2}\right]\right) \qquad (2.44)$$

The solution of the above least square problem can be easily solved by either using the Orthogonally Principle or Error Minimization via Gradients. The solution will be:

$$f^* = (\mathcal{H}^+\mathcal{H})^{-1}\mathcal{H}^+\mathbb{F}, \tag{2.45}$$

where $\mathcal{H} = \begin{bmatrix} HD_1 \\ HD_2 \\ \cdots \\ HD_K \end{bmatrix}$, $\mathbb{F} = \begin{bmatrix} F_1 e^{j\varphi_1} \\ F_2 e^{j\varphi_2} \\ \cdots \\ F_K e^{j\varphi_k} \end{bmatrix}$.

And then,

$$\mathcal{H}^+\mathcal{H} = [(HD_1)^+, (HD_2)^+, \dots, (HD_K)^+]\begin{bmatrix} HD_1 \\ HD_2 \\ \cdots \\ HD_K \end{bmatrix}$$

$$= D_1{}^+H^+HD_1 + D_2{}^+H^+HD_2 + \cdots + D_K{}^+H^+HD_K$$

$$= KNI. \tag{2.46}$$

The above result is also from the fact that $D_i^+ D_i = I$, this is because D_i is a diagonal matrix with all elements have unit modulus. Thus, $(\mathcal{H}^+\mathcal{H})^{-1} = \frac{1}{KN}I$.

Then we have,

$$\mathcal{H}^+\mathbb{F} = [(HD_1)^+, (HD_2)^+, \dots, (HD_K)^+]\begin{bmatrix} F_1 e^{j\varphi_1} \\ F_2 e^{j\varphi_2} \\ \cdots \\ F_K e^{j\varphi_k} \end{bmatrix}$$

$$= D_1{}^+H^+F_1 e^{j\varphi_1} + D_2{}^+H^+F_2 e^{j\varphi_2} + \cdots + D_K{}^+H^+F_K e^{j\varphi_k}. \tag{2.47}$$

23

Let $\frac{1}{N}D_i{}^+H^+F_ie^{j\varphi_i} = \hat{f}_i$. This is the inverse Fourier transform of $F_ie^{j\varphi_i}$ and divided by the

corresponding defocused phase term: $\exp\left[-j\pi\frac{z_i(x''^2+y''^2)}{\lambda f^2}\right]$.

Thus, we get the solution for this least square problem:

$$f^* = (\mathcal{H}^+\mathcal{H})^{-1}\mathcal{H}^+\mathbb{F} = \frac{1}{KN}I\left(N\hat{f}_1 + N\hat{f}_2 + \cdots + N\hat{f}_K\right)$$

$$(2.48)$$

$$= \frac{1}{K}\left(\hat{f}_1 + \hat{f}_2 + \cdots + \hat{f}_K\right).$$

This is the average of estimated pupil functions from the projection at each defocused Fourier constraint.

From the above proof, we know that the Fienup-style averaged projection in SMLM for phase retrieval is equivalent to the specific least square problem solving by alternating minimization. Then we will prove the monotonicity decreasing of this alternating minimization.

Proof of the monotonicity of averaged projection:

Suppose at m-th iteration we know $f^{(m)}(x'', y'')$ and

$$\varphi_k{}^{(m)}(x', y') = \arg\left(\mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right]f^{(m)}(x'', y'')\right\}\right).$$

$$(2.49)$$

By solving least square problem showing above, we have the solution $f^{(m+1)}(x'', y'')$ that minimizes:

$$\sum_k\sum_{x',y'}\left||F_k(x', y')|e^{j\varphi_k{}^{(m)}(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right]f(x'', y'')\right\}\right|^2.$$

$$(2.50)$$

So we have:

$$\sum_k\sum_{x',y'}\left||F_k(x', y')|e^{j\varphi_k{}^{(m)}(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right]f^{(m+1)}(x'', y'')\right\}\right|^2$$

$$\leq \sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k{}^{(m)}(x',y')} - \mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m)}(x'',y'') \right\} \right|^2$$

(2.51)

Then we update φ_k :

$$\varphi_k{}^{(m+1)} = \arg\left(\mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m+1)}(x'',y'') \right\} \right).$$

(2.52)

From Lemma 2.4:

$$\sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k{}^{(m+1)}(x',y')} - \mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m+1)}(x'',y'') \right\} \right|^2$$

$$\leq \sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k{}^{(m)}(x',y')} - \mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m+1)}(x'',y'') \right\} \right|^2$$

(2.53)

So we will have:

$$\sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k{}^{(m+1)}(x',y')} - \mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m+1)}(x'',y'') \right\} \right|^2$$

$$\leq \sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k{}^{(m)}(x',y')} - \mathcal{F}\left\{ \exp\left[-j\pi \frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m+1)}(x'',y'') \right\} \right|^2$$

$$\leq \sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k^{(m)}(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m)}(x'',y'')\right\} \right|^2$$

Because

$$\sum_{k} \sum_{x',y'} \left| |F_k(x',y')|e^{j\varphi_k^{(m)}(x',y')} - \mathcal{F}\left\{\exp\left[-j\pi\frac{z_k(x''^2 + y''^2)}{\lambda f^2}\right] f^{(m)}(x'',y'')\right\} \right|^2$$

$$= \sum_{k} \sum_{x',y'} \left| |F_k(x',y')| - \left|\mathcal{F}_k[f^{(m)}](x',y')\right| \right|^2 \tag{2.54}$$

We will have:

$$\sum_{k} \sum_{x',y'} \left| |F_k(x',y')| - \left|\mathcal{F}_k[f^{(m+1)}](x',y')\right| \right|^2$$

$$\leq \sum_{k} \sum_{x',y'} \left| |F_k(x',y')| - \left|\mathcal{F}_k[f^{(m)}](x',y')\right| \right|^2 \tag{2.55}$$

So we can guarantee that the squared error is monotonically decreased by using averaged projection method. The simulated results are consistent with this statement shown in next section.

2.2 Simulation results for Gerchberg-Saxton algorithm and Fienup-style algorithm

In this section, the data simulation process is the same as the process in Section 3.3. To compare these two methods, I first simulate the noise-free data from the generated pupil function used in Section 3.3.1, and then we can see from Figure 2.4 that when no noise is in presence, the convergence rate of Gerchberg-Saxton algorithm is faster than Fienup-Style algorithm.
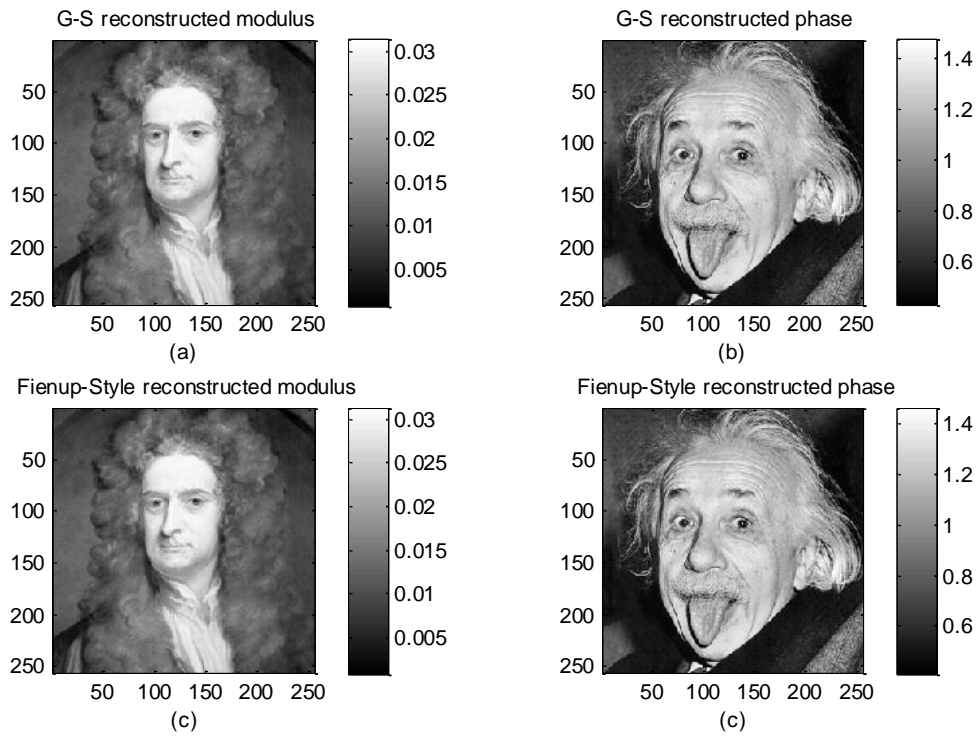
Figure 2.3 G-S algorithm and Fienup-style algorithm with no noise. Top line shows the G-S algorithm reconstructed modulus and phase with no noise presented; bottom line shows the Fienup-Style reconstructed modulus and phase under no noise present. Both of these methods can solve the phase retrieval problem.
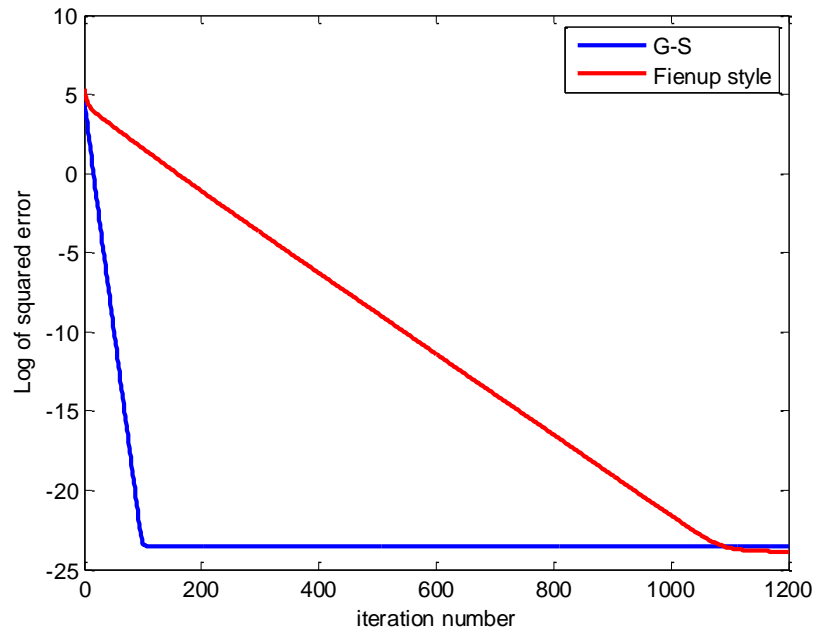
Figure 2.4 Log of squared error between measurements and reconstructed defocused Fourier modulus: G-S algorithm and Fienup-style algorithm. G-S algorithm converges faster than Finup-style algorithm

However, if we add Poisson noise to the measurements, we can see that the performance of Fienup-style algorithm is better than G-S algorithm. The following figure shows the results when one million photons are collected.

G-S reconstructed modulus

G-S reconstructed phase

(a)

(b)

Fienup-Style reconstructed modulus

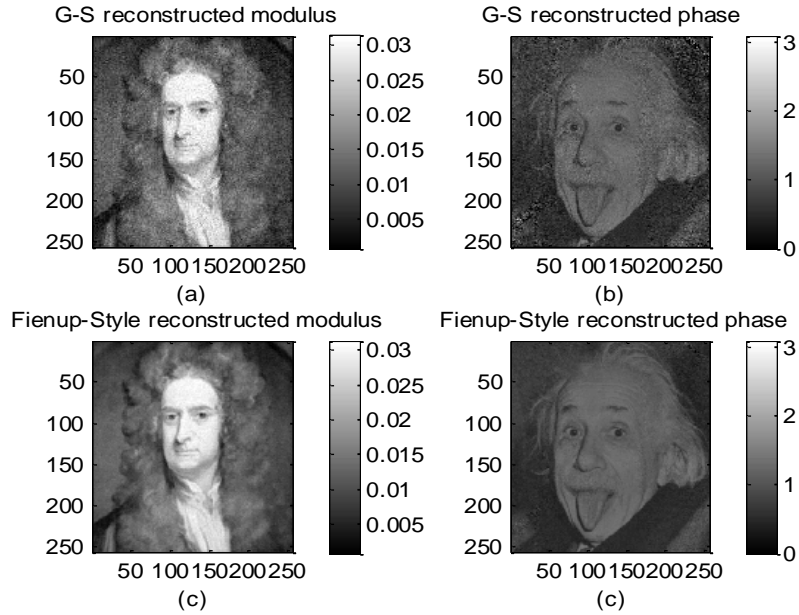Fienup-Style reconstructed phase

(c)

(c)

Figure 2.5 G-S algorithm and Fienup style algorithm with Poisson noise. Top line shows the G-S algorithm reconstructed modulus and phase when Poisson noise is presented; bottom line shows the Fienup-Style reconstructed modulus and phase when Poisson noise is presented
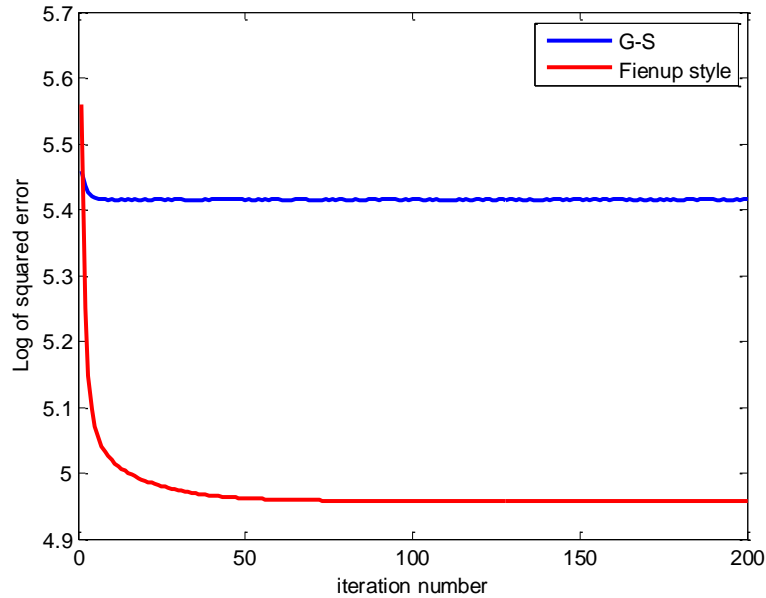

Figure 2.6 Log of squared error for G-S algorithm and Fienup algorithm with Poisson noise. The Fienup-style algorithm have lower squared error than G-S algorithm.

29

In Figure 2.6, we observed that the G-S algorithm cannot continue to decrease the squared error after some iterations. In comparison, Fienup-style algorithm can guarantee the squared error monotonically decreasing for more iterations. This result is consistent with our analysis.

Chapter 3

Phase Retrieval by Double Alternating Minimization

In Chapter 2 we discussed the phase retrieval problem in SMLM and the popular methods for solving it. However, no method discusses the sampling issues in this problem. As we know, cameras sample the distribution of intensity of light, which is squared modulus of defocused Fourier transform of the effective pupil function. As the sampling theorem indicates [28], we need to sample at a high enough rate to avoid aliasing. Thus, if the sampling aperture of detector is too large to sample at high enough rates, then it is impossible to reconstruct pupil mask at function domain. This is because the inverse Fourier transform of loosely sampled Fourier transform is aliased at function domain. In order to solve this problem, we need to interpolate or estimate the missed Fourier modulus. However, if we just consider about defocused Fourier modulus, it is obvious that there are infinite many solutions for maximum likelihood estimation. This is because that the ideal detector (sample aperture equals to sample space and each photon detected by pixel does not affect surrounding pixels) measures integral of all light arrived at it and each detector is independent with others. If we only consider the Poisson noise, the maximum likelihood can be achieved just when the summation of parameters equals to the measurement from the corresponding detector. For example, suppose we have M detectors, we want to estimate $N = 4M$ number of parameters, so at each detector, we have four unknown parameters and just one equation. The problem is underdetermined and thus will have infinite many solutions. However, there may not exist such pupil function that generate those estimations of defocused Fourier modulus. Then, we estimate the defocused Fourier modulus from measurements by adding a penalty that penalize the solutions from which we cannot find an appropriate pupil function.

In this chapter, we will discuss the pixelation of perfect detectors and sampling issues. Then we will build our model of algorithms for solving this phase retrieval problem.

31

3.1   Sampling and Pixelation

Due to the fact that simulations are worked on by computer, the discrete signal processing needs to be addressed. In this section, I will discuss the sampling issue on spatial domain for approximating our continuous pupil function, and then discuss the sampling issue on Fourier domain in order to address the problem of pixelation.

3.1.1   Sampling on spatial domain

Due to the fact that pupil function is continuous and the aperture size on corresponding spatial plane is finite, the Fourier spectrum of it is infinitely spread. Then any sampling process will introduce aliasing effect in frequency domain. Besides, the camera size is also finite, so there always information about high frequency components lost. However, we can assume that the pupil function is sampled at high enough rate such that the aliasing effect can be ignored, and also assume that the camera size is large enough to record all data about the Fourier spectrum. Suppose now we want to reconstruct a pupil function sampled in separations Δx and Δy at each direction. We also suppose that the sampling rate is high and the Discrete-Time Fourier Transform (DTFT) of sampled pupil function can approximate to the Fourier transform of the continuous function. Then the range of frequency component in the base period of Fourier transform in frequency domain is $\left[\frac{-1}{2\Delta x}, \frac{1}{2\Delta x}\right] \times \left[\frac{-1}{2\Delta y}, \frac{1}{2\Delta y}\right]$. Assume the camera size is greater or equal to $\frac{1}{\Delta x} \times \frac{1}{\Delta y}$ and the sampling period of detectors is small enough , then all information about the sampled pupil function are known and thus the pupil function can be reconstructed by using phase retrieval algorithms.

3.1.2   Sampling on frequency domain and pixelation

Under the assumptions from Section 3.1.1, we can reconstruct the sampled pupil function. However, we still need to sample on the frequency domain by the detectors. The sampling process in the frequency domain will result in a summation of periodic displacements on spatial domain, this

property is just similar as sampling process on spatial domain. Therefore, in order to avoid aliasing effect on spatial domain, we still need to sample at a high enough rate in the frequency domain. Suppose the aperture size is $\frac{A}{2} \times \frac{A}{2}$ (the shape of aperture is circle and we will use here as rectangular for simulation purpose and we can also pad zeros around the aperture to construct a rectangular image), if there is no aliasing by the sampling process on frequency domain, then the sampling period should be $\frac{1}{A}$ and $\frac{1}{A}$ at each direction. Assume the sampled pupil function has size $M \times N$ pixels. Then $\Delta x = \frac{A}{M}, \Delta y = \frac{A}{N}$. Thus, the maximum sampling period on frequency domain is $\frac{1}{M\Delta x}$ and $\frac{1}{N\Delta y}$ at each direction. The number of the samples on frequency domain at each direction is $\frac{\frac{1}{\Delta x}}{\frac{1}{M\Delta x}} = M$ and $\frac{\frac{1}{\Delta y}}{\frac{1}{N\Delta y}} = N$.

The above sampling process can guarantee that the reconstructed pupil function will not be aliased. From the above sampling processes on both spatial domain and frequency domain, we can observe the approximation of output on frequency domain is Discrete Fourier Transform (DFT). Thus, we use DFT in our simulations.

Suppose we use CMOS camera with perfect detectors, then the measurement read from each detector is the summation of photons distributed on this detector. Thus, our sampling process will have pixelation effect. If the detector size is small (sample period is small), then the sampling rate is high enough and the pixelation effect can be ignored. However, some times the detector size may be too large to sample at a high rate on frequency domain for reconstructing large apertures. Thus, the pixelation needs to be taken into account.

3.2   Estimation process for phase retrieval

In order to solve the low sampling rate problem discussed in section 3.1.2, I come up with the method to estimate the missed values of defocused Fourier modulus by minimizing the penalized I-divergence, which is equivalent to maximizing penalized likelihood.

3.2.1 The optimization problem

Let $d = \{d_1, d_2, \ldots, d_K\}$ be the measurement, where d_i is a measured image when the point source is placed at i-th position and suppose d_i is at the coordinate $\{(u, v): u = m\Delta u, v = n\Delta v; m = 1, \ldots, M; n = 1, \ldots, N\}$, where Δu and Δv are height and width of rectangular detectors. We want to reconstruct a sampled pupil mask that has $M' \times N'$ samples, where $M' > M, N' > N$. From the discussion in section 3.1.2, we know that the number of points in Fourier domain needs to be at least $M' \times N'$. Let $\mu = \{\mu_1, \mu_2, \ldots, \mu_K\}$ be our estimation parameter, μ_i is estimation of parameter when the point source placed at i-th defocused plane and μ_i is at the coordinate $\{(u', v'): u' = m\Delta u', v = n\Delta v'; m = 1, \ldots, M'; n = 1, \ldots, N'\}$. Therefore, each detector contains $\frac{M'}{M} \times \frac{N'}{N}$ parameters. For the pixelation of perfect detector, the measurement of each detector is the summation of squares of all these $\frac{M'}{M} \times \frac{N'}{N}$ parameters. Then we define the response function of detectors:

$$h(u, v | u', v') = \begin{cases} 1, & \text{if } (u', v') \text{ is in the region defined by detectors at } (u, v) \\ 0, & (u', v') \text{ is not in the region defined by detectors at } (u, v) \end{cases}$$

(3.1)

Thus, we have the expected value of measurement given parameter μ:

$$g_i(u, v) = \sum_{u', v'} h(u, v | u', v') \mu_i^2 (u', v')$$

(3.2)

Since the measurement is suffered from shot noise when we use optical detectors, the data received is Poisson distributed: $d_k(x, y) \sim \text{Poisson}(g_k(x, y))$. The likelihood for each detector is:

$$l[d_i(u, v) | g_i(u, v)] = e^{-g_i(u,v)} \frac{g_i(u, v)^{d_i(u,v)}}{d_i(u, v)!}$$

(3.3)

Since all detectors are independent with each other and the measurements from different defocused positions are also independent, then the likelihood of all data is [21]:

$$l_d(g) = \prod_i \prod_{u,v} e^{-g_i(u,v)} \frac{g_i(u,v)^{d_i(u,v)}}{d_i(u,v)!} \tag{3.4}$$

The log-likelihood is:

$$L_d(g) = \sum_i \sum_{u,v} d_k(u,v)\log[g_i(u,v)] - g_i(u,v) - \log d_i(u,v)! \tag{3.5}$$

Then we want to find $\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$, which is estimated defocused Fourier modulus, to maximize $L_d(g)$ and also to guarantee that there exists such pupil function that can generate the estimated μ.

If $\mu = \{\mu_1, \mu_2, \dots, \mu_K\}$ can be generated from some pupil function $f^*(x,y)$, then the following expression must hold:

$$\sum_i \sum_{u',v'} \|\mu_i(u',v') - |\mathcal{F}_k[f^*](u',v')|\|^2 = 0 \tag{3.6}$$

Thus, $\min_f \|\mu_i(u',v') - |\mathcal{F}_k[f](u',v')|\|^2 = 0$ is the constraint that must hold for solving the maximum likelihood problem.

We could also change the maximum likelihood to minimum I-divergence problem under the Poisson noise condition [25]. The I-divergence is the generalized relative entropy for any two non-negative distributions and the expression is given by [25]:

$$I(d||g) = \sum_i \sum_{u,v} d_i(u,v)\log\frac{d_i(u,v)}{g_i(u,v)} - d_i(u,v) + g_i(u,v) \tag{3.7}$$

Since $d_i(u,v)$ does not depend on our estimation parameter, thus:

$$\max_\mu \sum_i \sum_{u,v} d_i(u,v)\log[g_i(u,v)] - g_i(u,v) - \log d_i(u,v)!$$

$$= \min_\mu \sum_i \sum_{u,v} d_i(u,v) \log\left[\frac{d_i(u,v)}{g_i(u,v)}\right] - d_i(u,v) + g_i(u,v) \tag{3.8}$$

35

Where

$$g_i(u, v) = \sum_{u', v'} h(u, v | u', v') \mu_i{}^2(u', v') \tag{3.9}$$

Then we can write our problem as the following:

$$\min_{\mu} \sum_i \sum_{u,v} [d_i(u, v) \ln \frac{d_i(u, v)}{\sum_{u', v'} h(u, v | u', v') \mu_i{}^2(u', v')} - d_i(u, v)$$

$$+ \sum_{u', v'} h(u, v | u', v') \mu_i{}^2(u', v')]$$

Such that:

$$\min_f \sum_i \sum_{u', v'} \left| \left| \mu_i(u', v') \right| - |\mathcal{F}_i[f](u', v')| \right|^2 = 0. \tag{3.10}$$

This is a constrained optimization problem with one equality constraint and the feasible region is given by $\left\{ \mu : \min_f \sum_i \sum_{u', v'} \left| \left| \mu_i(u', v') \right| - |\mathcal{F}_i[f](u', v')| \right|^2 = 0 \right\}$. We see that this problem is very difficult to solve if we formulate the Lagrangian and try to find the point that force the gradient to be zero. Thus, we use the penalty method to solve it and find the optimal point iteratively by using alternating minimizations.

3.2.2 DAM for Solving Phase Retrieval Problem in SMLM

By using penalty method to approximate the constrained optimization problem, we formulate our objective function followed by the penalty term as:

$$J(\mu) = \left\{ \sum_i \sum_{u,v} d_i(u,v) \ln \frac{d_i(u,v)}{\sum_{u',v'} |\mu_i(u',u')|^2 h(u,v|u',v')} - d_i(u,v) \right.$$

$$\left. + \sum_{u',v'} \mu_i(u',v')^2 h(u,v|u',v') \right\} +$$

$$\beta \left\{ \min_f \sum_i \sum_{x',y'} |\, \mu_i(u',v') - |\mathcal{F}_i[f](u',v')|\,|^2 \right\} \tag{3.11}$$

Where β is the regularization parameter and $\sum_i \sum_{x',y'} |\,|\mu_i(u',v')| - |\mathcal{F}_i[f](u',v')|\,|^2$ is the penalty term. Thus, we expand our optimization problem constrained in feasible region $\left\{ \mu: \min_f \sum_i \sum_{u',v'} \|\mu_i(u',v') - |\mathcal{F}_i[f](u',v')|\|^2 = 0 \right\}$ to the optimization problem without such constraint by adding penalty to the objective function. Thus, our problem becomes just:

$$\min_\mu J(\mu) \tag{3.12}$$

It is also impossible to solve our reformulated optimization problem directly. Thus, in order to solve it, we need to use alternating minimization algorithm.

First, we can further rewrite our problem:

$$\min_\mu J(\mu)$$

$$= \min_\mu \sum_i \sum_{u,v} [d_i(u,v) \ln \frac{d_i(u,v)}{\sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')} - d_i(u,v)$$

$$+ \sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')]$$

$$+ \min_f \sum_i \sum_{u',v'} |\, \mu_i(u',v') - |\mathcal{F}_i[f](u',v')|\,|^2$$

$$= \min_{\mu} \min_{f} \sum_{i} \sum_{u,v} [d_i(u,v)\ln\frac{d_i(u,v)}{\sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')} - d_i(u,v)$$

$$+ \sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')]$$

$$+ \beta \sum_{i} \sum_{u',v'} |\mu_i(u',v') - |\mathcal{F}_i[f](u',v')||^2 \tag{3.13}$$

Thus, this problem is naturally lifted to higher dimensions optimization and the alternating minimization should be applied.

Define our new objective function with respect to alternating minimization:

$$Q(\mu,f) = \sum_{i} \sum_{u,v} [d_i(u,v)\ln\frac{d_i(u,v)}{\sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')} - d_i(u,v)$$

$$+ \sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')]$$

$$+ \beta \sum_{i} \sum_{u',v'} |\mu_i(u',v') - |\mathcal{F}_i[f](u',v')||^2 \tag{3.14}$$

The idea of alternating minimization is to minimize $Q(\mu,f)$ with respect to one variable at a time by fixing the other variable, and then assume that the updated variable is fixed to minimize $Q(\mu,f)$ with respect to another variable. Updating both two variables iteratively decrease the value of $Q(\mu,f)$.

Suppose at m-th iteration we have known values of $\mu^{(m)}$, then by minimizing the term: $\sum_i \sum_{u',v'} |\mu_i(u',v') - |\mathcal{F}_i[f](u',v')||^2$, we have $f: f^{(m)} = \underset{f}{\text{argmin}}\, Q(\mu^{(m)},f)$.

Then $J(\mu^{(m)}) = Q(\mu^{(m)},f^{(m)})$. Based on the updated $f^{(m)}$, we can find the value of μ: $\mu^{(m+1)} = \underset{\mu}{\text{argmin}}\, Q(\mu,f^{(m)})$, and we also have $Q(\mu^{(m+1)},f^{(m)}) \leq Q(\mu^{(m)},f^{(m)})$. By updating f again: $Q(\mu^{(m+1)},f^{(m+1)}) \leq Q(\mu^{(m+1)},f^{(m)})$. Thus we will have:

$$Q(\mu^{(m+1)},f^{(m+1)}) \leq Q(\mu^{(m+1)},f^{(m)}) \leq Q(\mu^{(m)},f^{(m)}) \tag{3.15}$$

Since $Q\left(\mu^{(m+1)}, f^{(m+1)}\right) = J\left(\mu^{(m+1)}\right)$, and $Q\left(\mu^{(m)}, f^{(m)}\right) = J\left(\mu^{(m)}\right)$, after each iteration, the objective function will be monotonically decreased.

The remaining problem now is to solve the minimization problems: $f^{(m)} = \underset{f}{\mathrm{argmin}}\, Q\left(\mu^{(m)}, f\right)$ and $\mu^{(m+1)} = \underset{\mu}{\mathrm{argmin}}\, Q\left(\mu, f^{(m)}\right)$. We will see both of these problems need to use iterative algorithm.

Thus, in each main iteration discussed above, we need two sub-iterations to solve these sub-problems respectively, and both of them use alternating minimization. Thus we call this method Double Alternating Minimizations (DAM).

We can also observe that during the process of minimizing $Q(\mu, f)$, the effective pupil function can also be reconstructed at the same time. Thus, DAM algorithm is also a phase retrieval and pupil reconstruction algorithm. In order to compare the results between DAM and other interpolation methods, we can just treat DAM as a special interpolation method. However, we need to keep in mind that DAM itself is a phase retrieval algorithm since it can reconstruct the effective pupil mask during the process of estimating defocused Fourier modulus.

3.2.2.1 First sub-iteration for updating Pupil function

At this sub-iteration, we suppose μ is known and then solve the problem:

$$\min_{f} \sum_{i} \sum_{u',v'} |\, \mu_i(u', v') - |\mathcal{F}_i[f](u', v')|\, |^2 \qquad (3.16)$$

From the discussion in Section 2.1.2.4, we know that this problem can be solved by the averaged projection algorithm. It is equivalent to the alternating minimization by introduce a new variable $\varphi_i(u', v')$, which is the phase angle of defocused Fourier transforms.

$$\underset{f}{\mathrm{Min}} \sum_{i} \sum_{u',v'} |\, \mu_i(u', v') - |\mathcal{F}_i[f](u', v')|\, |^2$$

$$= \min_{f} \min_{\varphi_i} \sum_{i} \sum_{u',v'} \left|\, \mu_i(u', v') e^{j\varphi_i(u',v')} - \mathcal{F}_i[f](u', v')\right|^2 \qquad (3.17)$$

The phase information is retrieved in this sub-iteration.

3.2.2.2 Second sub-iteration for updating Fourier Modulus

In second sub-iteration, we consider the pupil function is fixed to be $f^{(m)}$, then $|\mathcal{F}_i[f](u',v')|$ is fixed in objective function, we can simply denote it as $F_i(u',v')$. Then the problem is becoming minimize the following function respect to $\mu_i(u',v')$:

$$Q(\mu) = \sum_i \sum_{u,v} [d_i(u,v)\ln \frac{d_i(u,v)}{\sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')} - d_i(u,v)$$

$$+ \sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')]$$

$$+ \beta \sum_i \sum_{u',v'} |\mu_i(u',v') - F_i(u',v')|^2 \qquad (3.18)$$

If we take the derivative of $Q(\mu)$ directively, we will have coupled equations for all $\mu_i(u',v')$ since the summation operator $\sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')$ is in the $\ln$ function. In order to decouple the equations of $\mu_i(u',v')$, we need a special case of convex decomposition lemma to put the projection operator outside to $\ln$ function [25].

Lemma 3.1:

$$\ln\left(\sum_i q_i\right) = -\min_{p \in P} \sum_i p_i \ln \frac{p_i}{q_i} \qquad (3.19)$$

Where $P = \{p: p_i \geq 0, \ \sum_i p_i = 1\}$.

Proof:

Formulate the Lagrangian: $L(p, \lambda) = \sum_i p_i \ln \frac{p_i}{q_i} + \lambda(\sum_i p_i - 1)$ and solve the following equations:

$$\nabla_p L(p, \lambda) = 0,$$

$$\nabla_\lambda L(p, \lambda) = 0. \tag{3.20}$$

Thus we will have:

$$\ln\frac{p_i}{q_i} + 1 + \lambda = 0,$$

$$\left(\sum_i p_i - 1\right) = 0. \tag{3.21}$$

So, $p_i^* = q_i e^{-1-\lambda}$ and $\sum_i q_i e^{-1-\lambda} = 1$. Thus $e^{-1-\lambda} = \frac{1}{\sum_i q_i}$, $p_i^* = \frac{q_i}{\sum_i q_i}$. Substitute p_i^* into $\sum_i p_i \ln\frac{p_i}{q_i}$

we get:

$$\sum_i p_i^* \ln\frac{p_i^*}{q_i} = -\ln\left(\sum_i q_i\right). \tag{3.22}$$

The proof of this lemma is completed.

Then, by using Lemma 3.1, we can rewrite $Q(\mu)$ as:

$$Q(\mu) = \min_{p \in P} \sum_i \sum_{u,v} \sum_{u',v'} [d_i(u,v)p_i(u,v|u',v')\ln\frac{d_i(u,v)p_i(u,v|u',v')}{h(u,v|u',v')\mu_i^2(u',v')}$$

$$- d_i(u,v)$$

$$+ \sum_{u',v'} h(u,v|u',v')\mu_i^2(u',v')]$$

$$+ \beta(\sum_i \sum_{u',v'} |\mu_i(u',v') - F_i(u',v')|^2) \tag{3.23}$$

Thus, the minimization problem respect to μ becomes alternating minimization between μ and p.

Then we can build alternating minimization algorithm for this sub-iteration:

(1) Initialize $\mu^{(0)}$.

At m-th iteration:

(2) Calculate $p_i^*(u,v|u',v') = \frac{h(u,v|u',v')\mu_i^{(m)^2}(u',v')}{\sum_{u',v'} h(u,v|u',v')\mu_i^{(m)^2}(u',v')}$.

41

(3) Minimizing $Q(\mu, p^*)$ by taking derivative:

$$\frac{\partial Q(\mu, p^*)}{\partial \mu_i(u', v')} = \sum_{u,v} -2d_i(u, v)p_i^*(u, v|u', v')\frac{1}{\mu_i(u', v')}$$

$$+ 2h(u, v|u', v')\mu_i(u', v') + 2\beta\left(\mu_i(u', v') - F_i(u', v')\right)$$

$$= 0 \tag{3.24}$$

Then we will have the solution:

$$\mu_i^{(m+1)}(u', v')$$

$$= \frac{\beta F_i(u', v') + \sqrt{\left(\beta F_i(u', v')\right)^2 + 4\left[\sum_{u,v} h(u, v|u', v') + \beta\right]\left[\sum_{u,v} d_i(u, v)p_i^*(u, v|u', v')\right]}}{2\left[\sum_{u,v} h(u, v|u', v') + \beta\right]}$$

$$\tag{3.25}$$

(4) If not convergent, go back to (2).

Also, for the calculation efficiency, we do not calculate $p_i^*(u, v|u', v')$ explicitly, and the flow-cart for calculating the updated μ is shown in Figure 3.1, where $h_0(u', v') = \sum_{u,v} h(u, v|u', v')$.
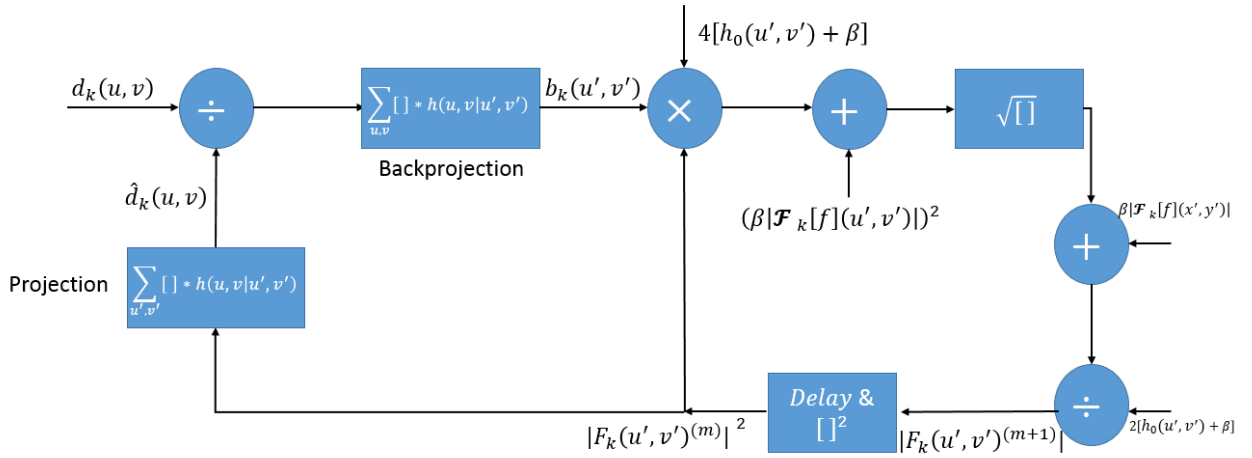
Figure 3.1 Flow-chart for sub-iteration 2.

3.3 Simulations and Results

In this section, I simulate the data by using the effective pupil mask generated by myself, and then I use our DAM algorithm to reconstruct it. By comparing the results from our Double Alternating Minimization algorithm to the currently used method, we see that our algorithm gives a much better result.

3.3.1 Data simulation

In the simulation, I supposed that the sampled pupil function we want to reconstruct is a complex function with size 256×256, which has magnitude to be an image of Newton and the phase angle to be an image of Einstein as shown in Figure 3.2. Then I scaled the magnitude of the generated pupil mask in order to control the number of collected photons. In the simulations, the collected total number of photons is set to be 10^8 and 10^5. Then I multiplied the generated effective pupil function with each defocused phase term, and make DFT of each defocused pupil function and pixelate it to generate the noise-free images on detector plane. The range of defocused position is from -25×10^{-4} to 25×10^{-4} with sampling distance 5×10^{-4}. The reason why I choose the order of 10^{-4} is because NA is approximated to be one, thus the term: $\frac{(x''^2 + y''^2)}{f^2}$ should close to 1. Besides, λ and Δz_k are few hundreds nanometers. Thus, $\frac{z_k(x''^2 + y''^2)}{\lambda f^2}$ should have order 0. However, in my simulation I just set all parameters to be one in order to be convenient to simulate. Since the simulated image size is 256×256, the order of $\max{(x''^2 + y''^2)}$ should be 4. Thus we need to set the order z_k approximately to -4. Finally, I added the Poisson noise to each of the detector pixels. The total detector number is 64×64.
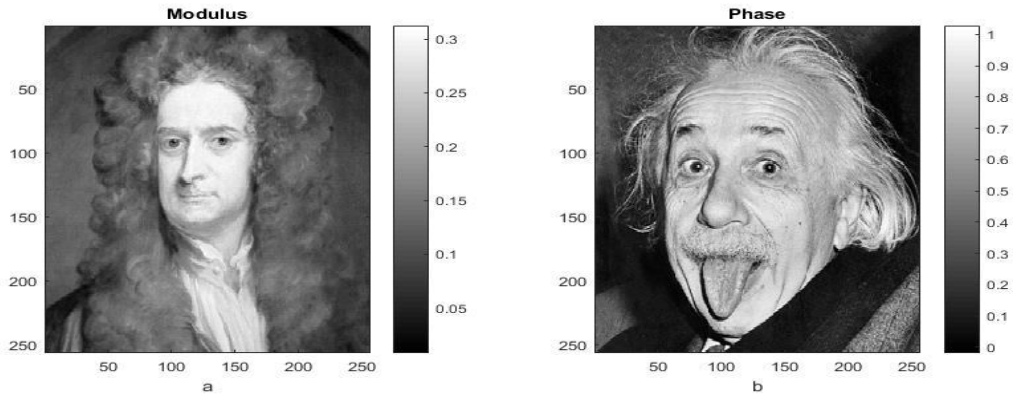
Figure 3.2 Ground-truth pupil to be reconstructed. The modulus is set to be Newton's image and the phase is set to be Einstein's image. The ranges of pixel values are shown in colorbars (phase is in units of radians).

3.3.2 Results and discussion

As we discussed in Chapter 2, we cannot reconstruct the image from the lower dimensional frequency domain. Thus, if we want to recover the image, we need to firstly interpolate the data in frequency domain. In this simulation, we want to reconstruct the image of size 256×256, but we only have 64×64 measurements. The first condition is that the total number of collected photons is 10^8. Then I reduced the photon number to be 10^5. The measured data and measured Fourier modulus at $z = -25 \times 10^{-4}$ from both collected photon numbers are shown in Figure 3.3.
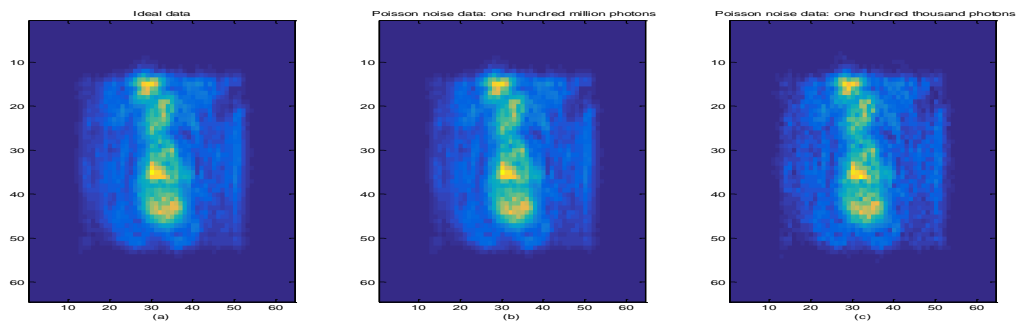
Figure 3.3 Measurements from different SNR. (a) Ideal data; (b) One hundred million photons collected, the maximum SNR of pixels is 80 dB; (c) One hundred thousand photons collected, the maximum SNR of pixels is 50 dB.

44

I ran 200 iterations of the Fienup-style algorithm after I interpolate the image by using nearest neighbor interpolation method. It will generate poor results as shown in Figure 3.4. Then, I applied our Double Alternating Minimization algorithm and ran 300 main iterations, in which each iteration has one sub-iteration 1 and one sub-iteration 2, to interpolate the image. And then I ran 200 iterations of Fienup-style algorithm for reconstructing pupil function. Our Double Alternating Minimization algorithm has much better results shown in Figure 3.5.
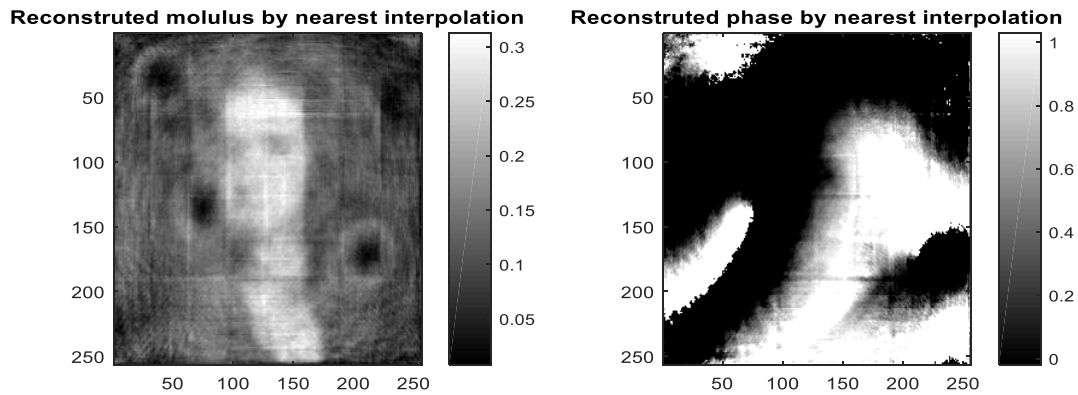
Figure 3.4 Reconstructed pupil function by using nearest neighbor interpolation. The modulus has severe artifacts. The phase cannot be reconstructed.
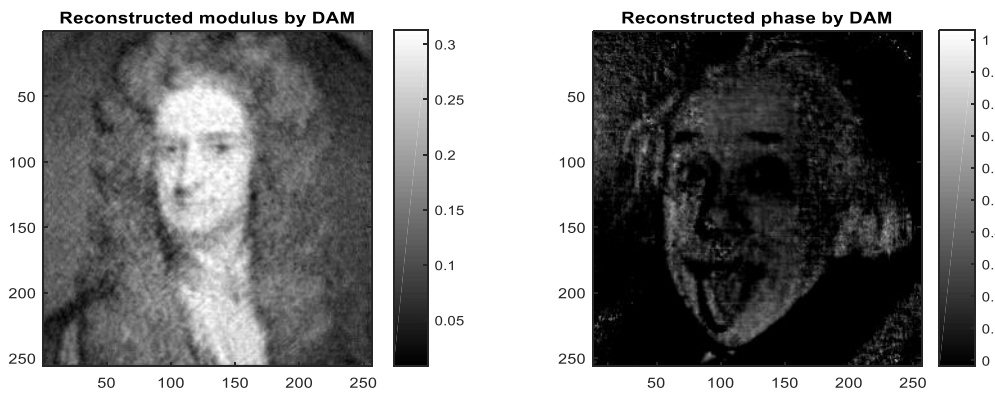
Figure 3.5 Reconstructed pupil function by using Double Alternating Minimizations. The modulus is clear, the shape of the phase image is reconstructed.

45

From Figure 3.5, we can directly observe that the results from DAM is much better than nearest neighbor interpolation. By using nearest neighbor interpolation, we observe that the reconstructed modulus has some black regions, which are corresponding to phase singularities, and the details are almost blurred; the reconstructed phase image shows almost no information about the original phase. The color bars in Figure 3.4 and Figure 3.5 are the same as the original image, this is because I changed the scale of the display of the color bar such that the values below the defined range are displaying dark and the values beyond the range are displaying white. The rescaled color bars images are also shown in Figure 3.6 and Figure 3.7, and the comparison between these two methods is still obvious.
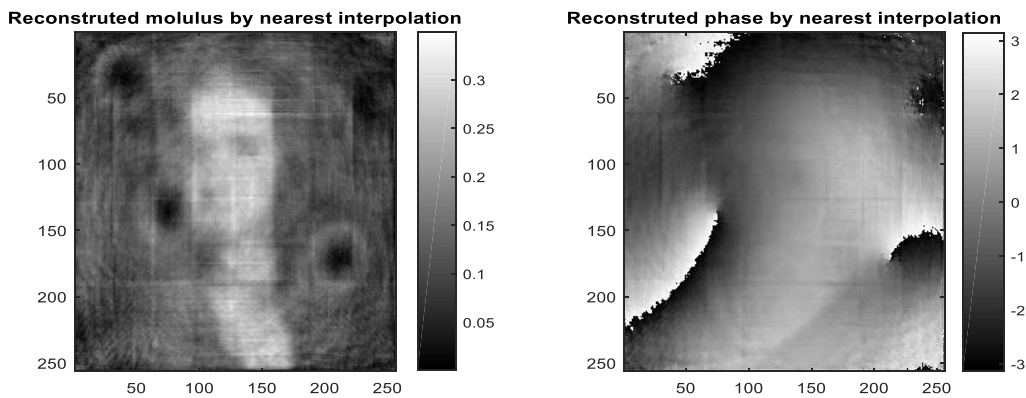
Figure 3.6 Reconstructed pupil function by using nearest interpolation: unchanged color bar. This is just using different colorbar with Figure 3.4.
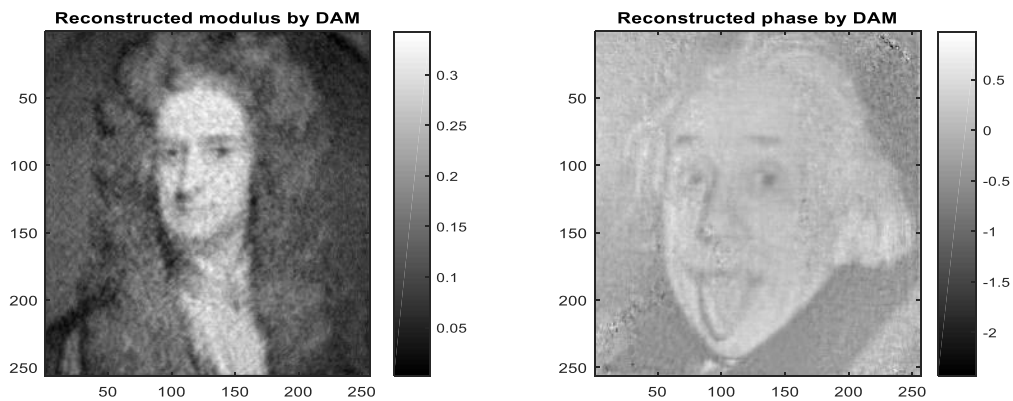
Figure 3.7 Reconstructed pupil function by using DAM: unchanged color bar. This is just using different colorbar with Figure 3.5

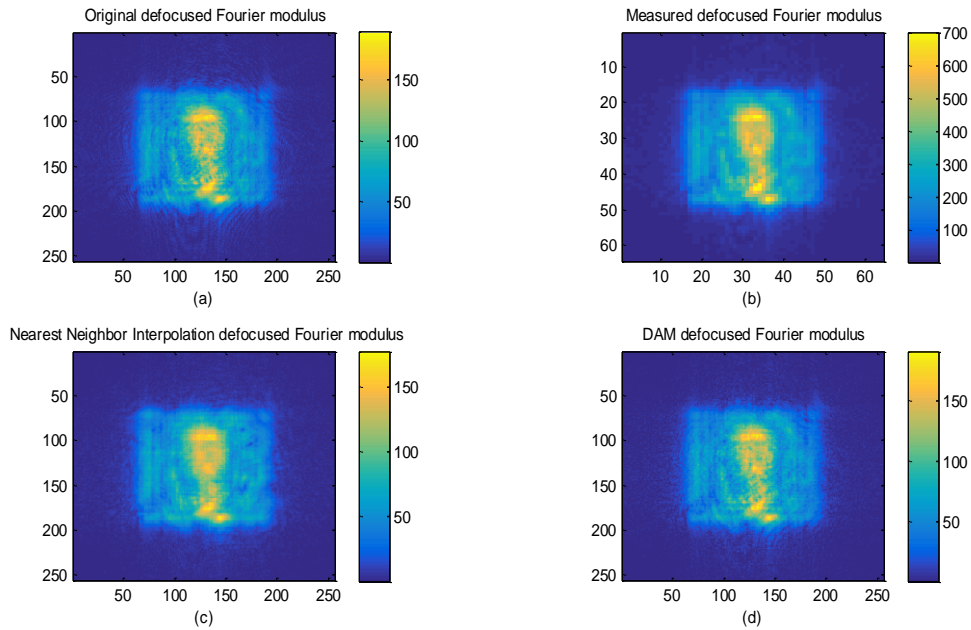We can also justify our method by comparing the defocused Fourier modulus, which is shown in Figure 3.8.

Figure 3.8 Reconstructed defocused Fourier modulus at one axial position. (a) Original defocused Fourier modulus at z=20e-4 position; (b) Measured defocused Fourier modulus at z=20e-4; (c) Reconstructed defocused Fourier modulus by using nearest neighbor interpolation before Fienup-style algorithm; (d) Reconstructed defocused Fourier modulus by using DAM before Fienup-style algorithm.

We can see from Figure 3.8 that the reconstructed defocused Fourier modulus, by using DAM, looks more similar to ground-truth than nearest neighbor interpolation reconstruction. Besides, the monotonicity of DAM estimation process is shown in Figure 3.9.

47
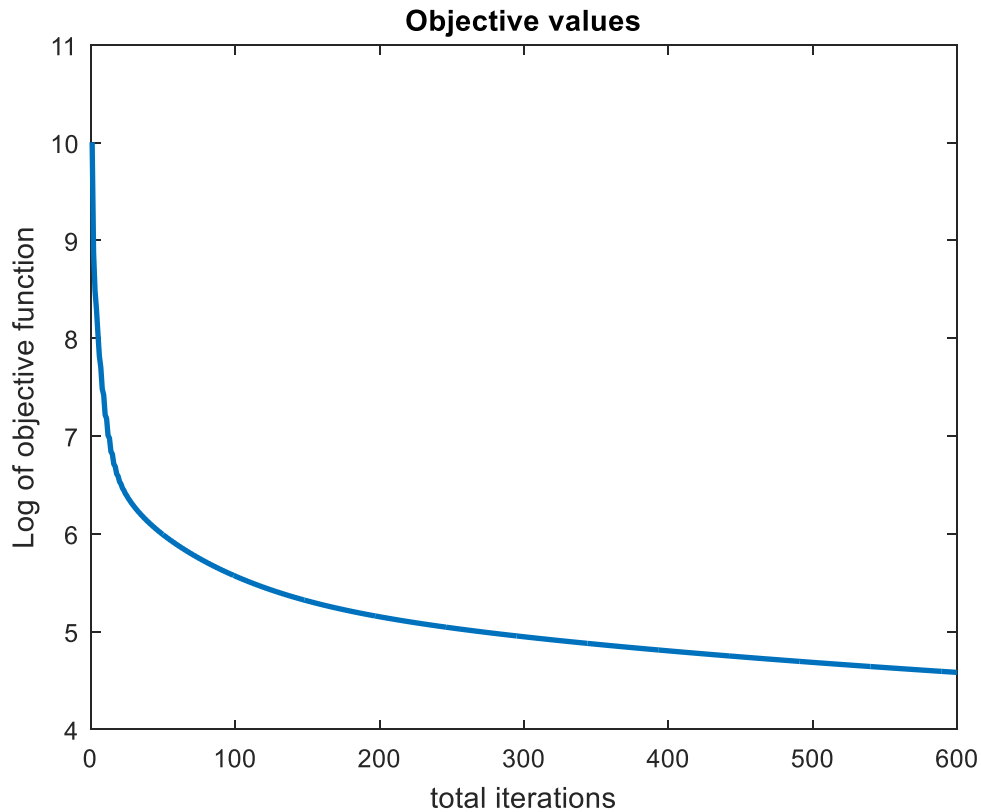
Figure 3.9 Objective values in DAM estimation process. The objective function is decreased after each sub-iteration.

In Figure 3.9, the number 600 is from 300 main iterations and each sub-iteration has been just run once in each main iteration. The reason why I choose this configuration will be discussed later in this section. It can be observed that the objective value is monotonically decreased after each sub-iteration, and thus it is monotonically decreasing after each main iteration.

Also, the quantitative squared error of both reconstruction algorithms is shown in Figure 3.10.
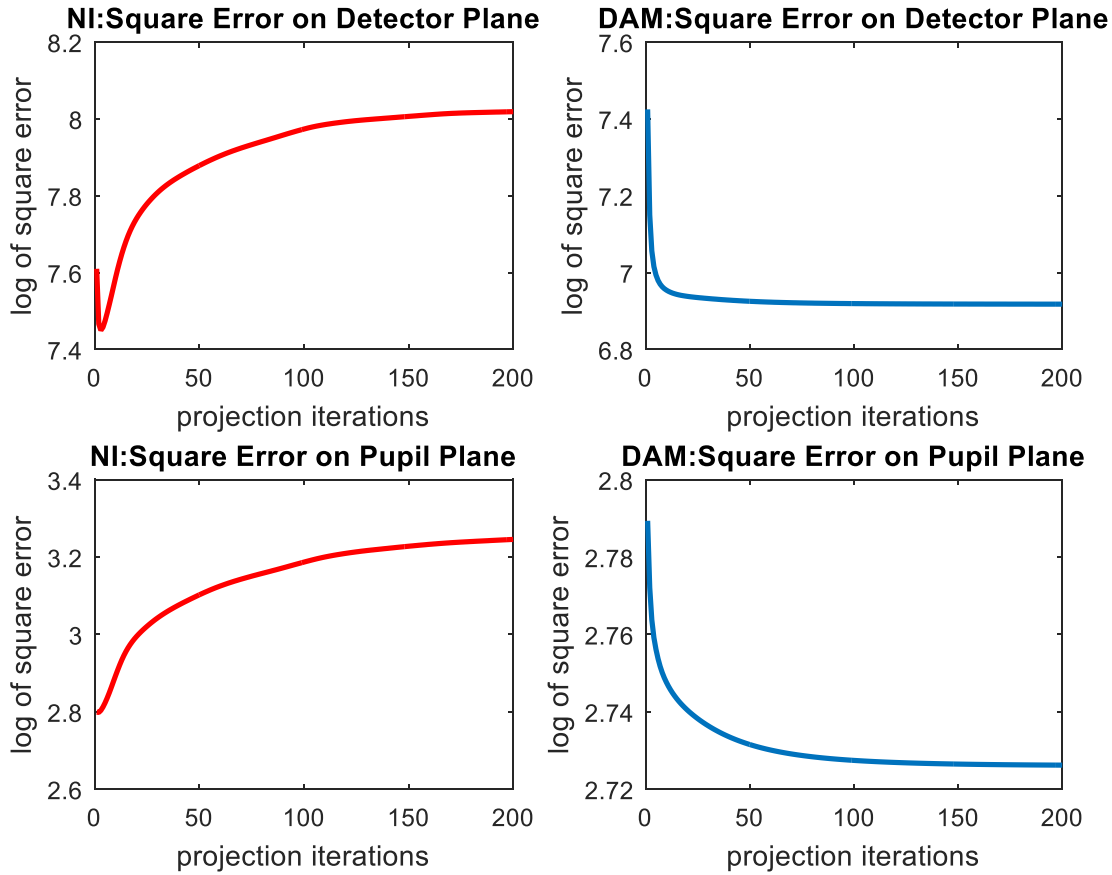
Figure 3.10 Log of squared error from Nearest Neighbor Interpolation and DAM on both pupil plane and detector plane. Top line shows the squared error on detector plane; bottom line shows the squared error on pupil plane.

From Figure 3.10, we can observe that the squared error between the defocused Fourier modulus and ideal Fourier modulus by using nearest neighbor interpolation will decrease in the first several iterations, and then it will even increase. The Fienup-style algorithm can just guarantee that the squared error between the estimation and input parameters is decreased. Thus, if the interpolated Fourier modulus has large deviations from the ground-truth Fourier modulus, the squared error between the estimation and the ground-truth Fourier modulus cannot be guaranteed to decrease. By using our DAM algorithm, a good estimate of the ground-truth Fourier modulus can be used as the input of

49

Fienup-style algorithm, so we have a much better result and we can see both of the squared error on detector plane and pupil plane are decreased in Fienup-style algorithm iterations.

In order to compare the squared error between nearest interpolation and DAM estimation, I plotted Figure 3.11:
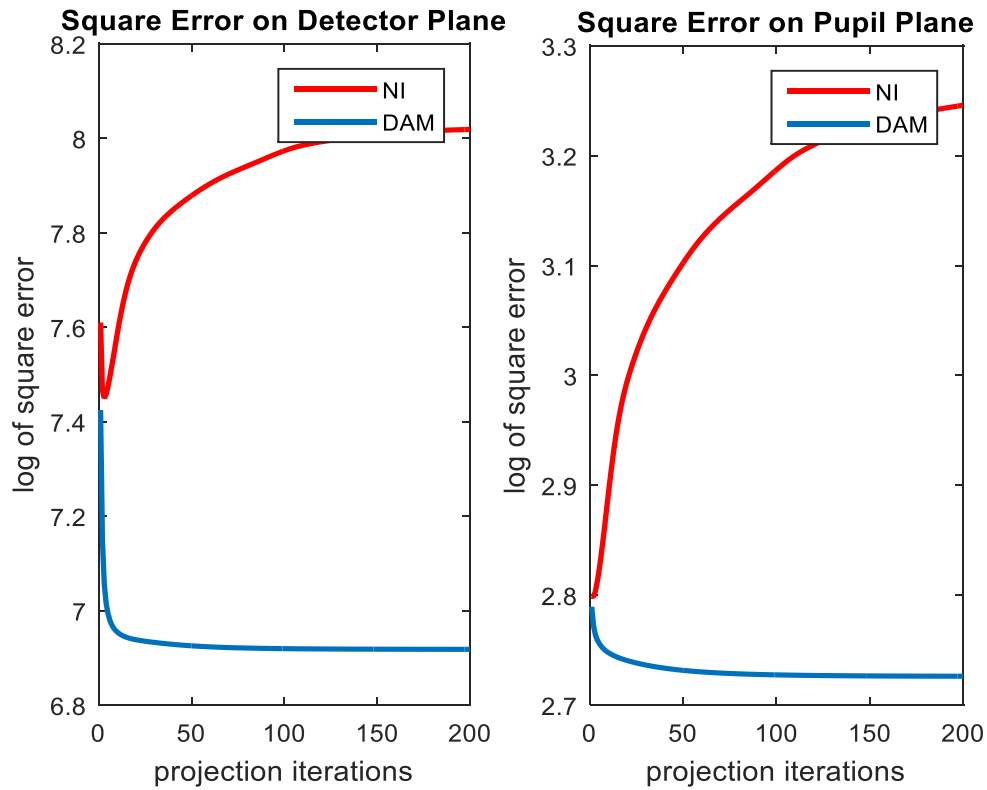

Figure 3.11 Log of squared error comparison between Nearest Interpolation and DAM

In Figure 3.11, we can observe that both squared errors on spatial domain and frequency domain have smaller values by using DAM than using nearest neighbor interpolation.

For a well-corrected optical imaging systems, we usually have phase ranged from $-\frac{1}{5}\pi$ to $\frac{1}{5}\pi$ [13]. Thus, after change the phase range of original pupil function to this range, we will have the reconstructed results showing in Figure 3.12.
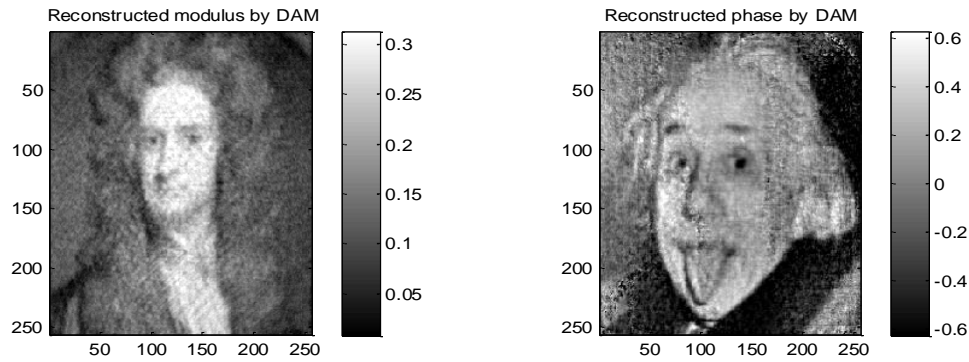
Figure 3.12 Reconstructed Pupil function by DAM: well-corrected. The phase is scaled to the range in well-corrected optical systems.

Also, the reconstruction by using nearest neighbor interpolation is shown in Figure 3.13.
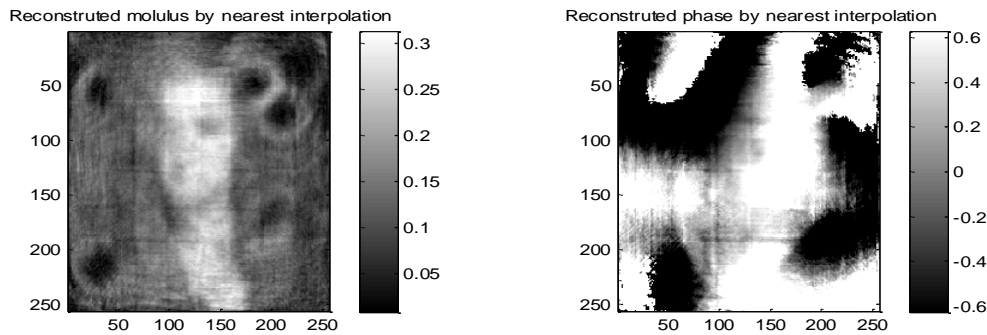

Figure 3.13 Reconstructed Pupil function by Nearest Neighbor Interpolation: well-corrected. The phase is scaled to the range in well-corrected optical systems.

From the above figures, we observe that DAM has a much better performance than the nearest neighbor interpolation when well-corrected optical systems are assumed to be used.

Suppose we have a circular aperture. To impose this function domain constraint, I just set the values of pupil function that are outside the aperture support domain to be zero. This is the process that project the estimated pupil function onto the function constraint set. Firstly, suppose we only know the aperture size, and the modulus and phase of pupil function are unknown. Then the simulated results are shown in Figure 3.14 and Figure 3.15.

51

Reconstruted molulus by nearest neighbor interpolation   Reconstructed phase by nearest neighbor interpolation
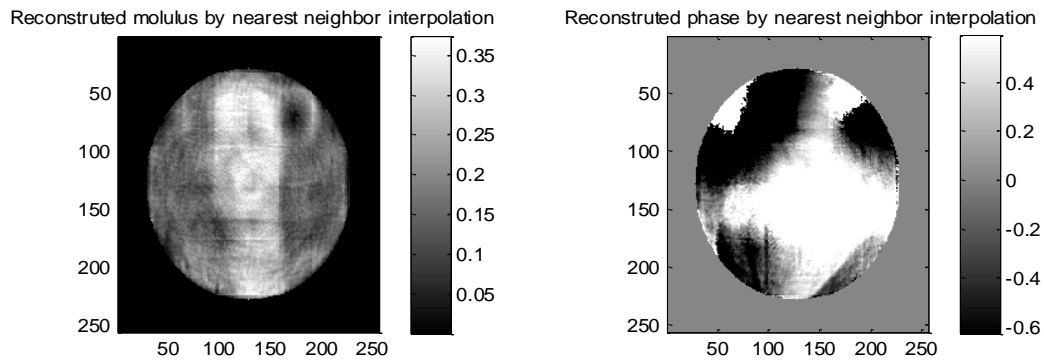
Figure 3.14 Reconstruction by nearest neighbor interpolation: circular aperture. The pupil function cannot be well reconstructed in the support domain.
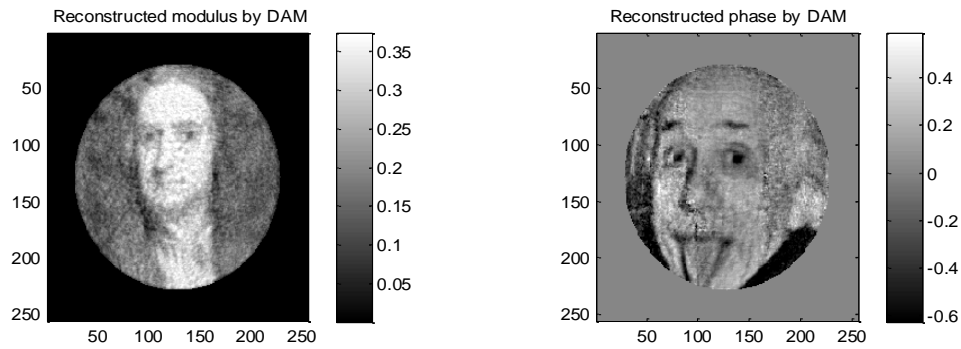

Reconstructed modulus by DAM   Reconstructed phase by DAM

Figure 3.15 Reconstruction by DAM: circular aperture. The reconstruction images are clear.

Then, I suppose that both aperture size and pupil modulus are known, the reconstructions are shown in Figure 3.16 and Figure 3.17.
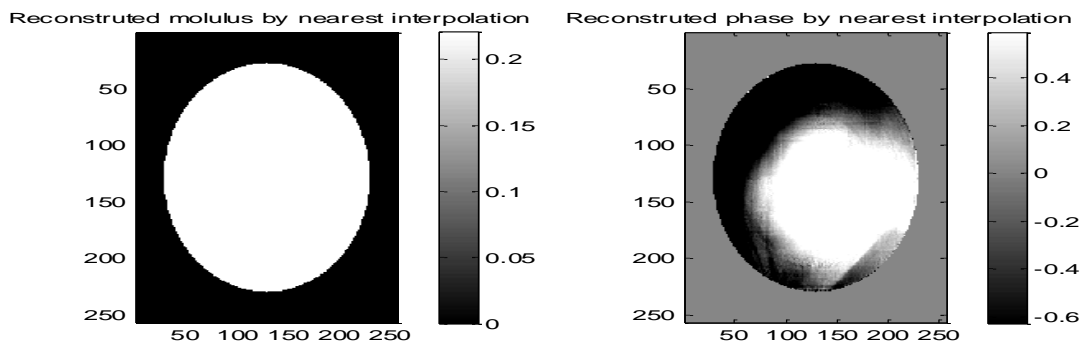

Reconstruted molulus by nearest interpolation   Reconstructed phase by nearest interpolation

Figure 3.16 Reconstruction by NNI: uniform known modulus. The phase cannot be reconstructed.
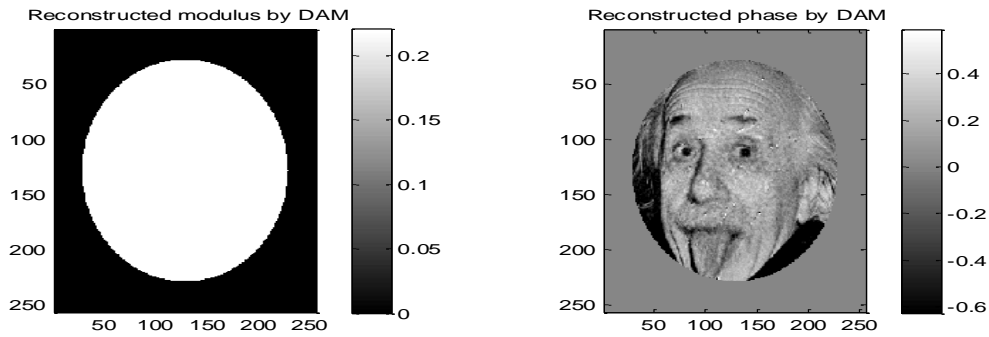
52

Figure 3.17 Reconstruction by DAM: uniform known modulus. The phase is reconstructed clearly.

The selected defocused Fourier modulus of known uniform modulus pupil function is shown in Figure 3.18 and the comparison on the frequency domain is also obvious.
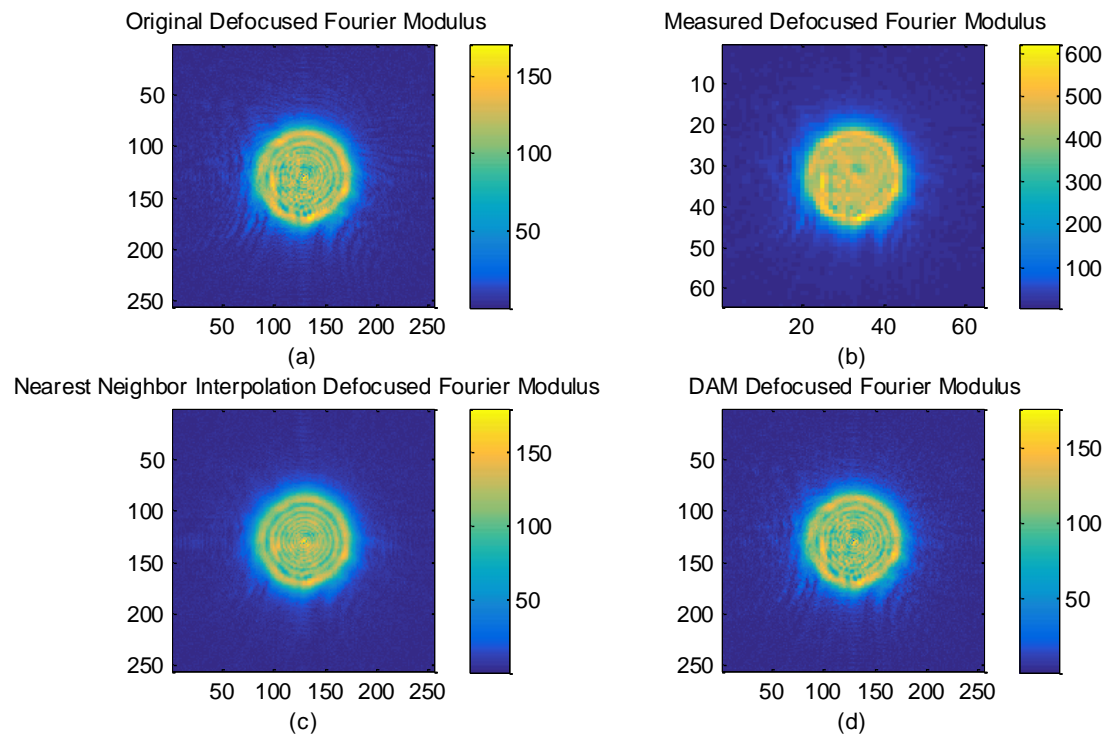

Figure 3.18 Reconstructed defocused Fourier modulus: uniform and circular aperture. (a) Original defocused Fourier modulus at z=20e-4 position; (b) Measured defocused Fourier modulus at z=20e-4; (c) Reconstructed defocused Fourier modulus by using nearest interpolation before Fienup style algorithm; (d) Reconstructed defocused Fourier modulus by using DAM before Fienup style algorithm.

In Figure 3.19, I also presented the results by using Fienup-style algorithm with the use of Linear Interpolation, Spline Interpolation and Bicubic Interpolation. We see that all these interpolation methods perform similar to the Nearest Neighbor Interpolation method, and they cannot generate good results.
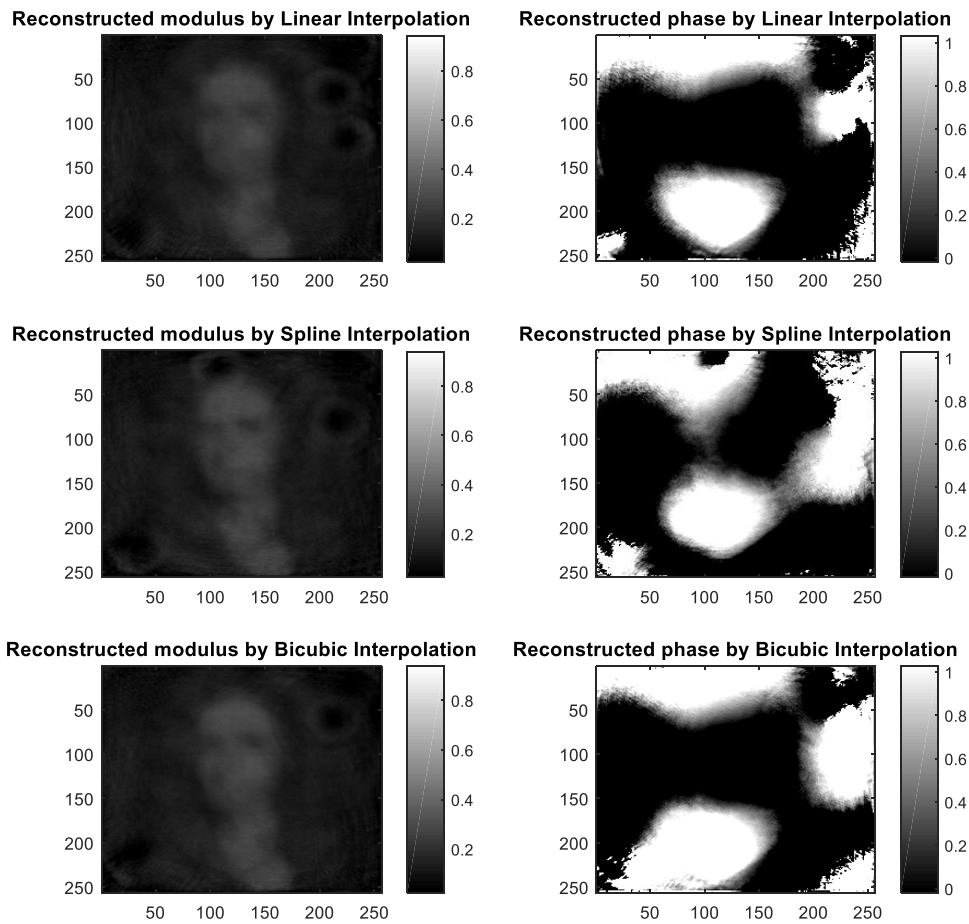

Figure 3.19 Reconstructions from other interpolation methods. Top line is the reconstructed pupil function by using Linear Interpolation; middle line is the reconstructed pupil function by using Spline Interpolation; bottom line is reconstructed pupil function by using Bicubic Interpolation.

Then, I decreased the total number of collected photons to be 10^5 and also run DAM 300 times, in which each running time also contains one sub-iteration 1 and one sub-iteration 2. The results are shown in Figure 3.20 and Figure 3.21. We can see the performance of DAM is still promising when we decreased SNR.
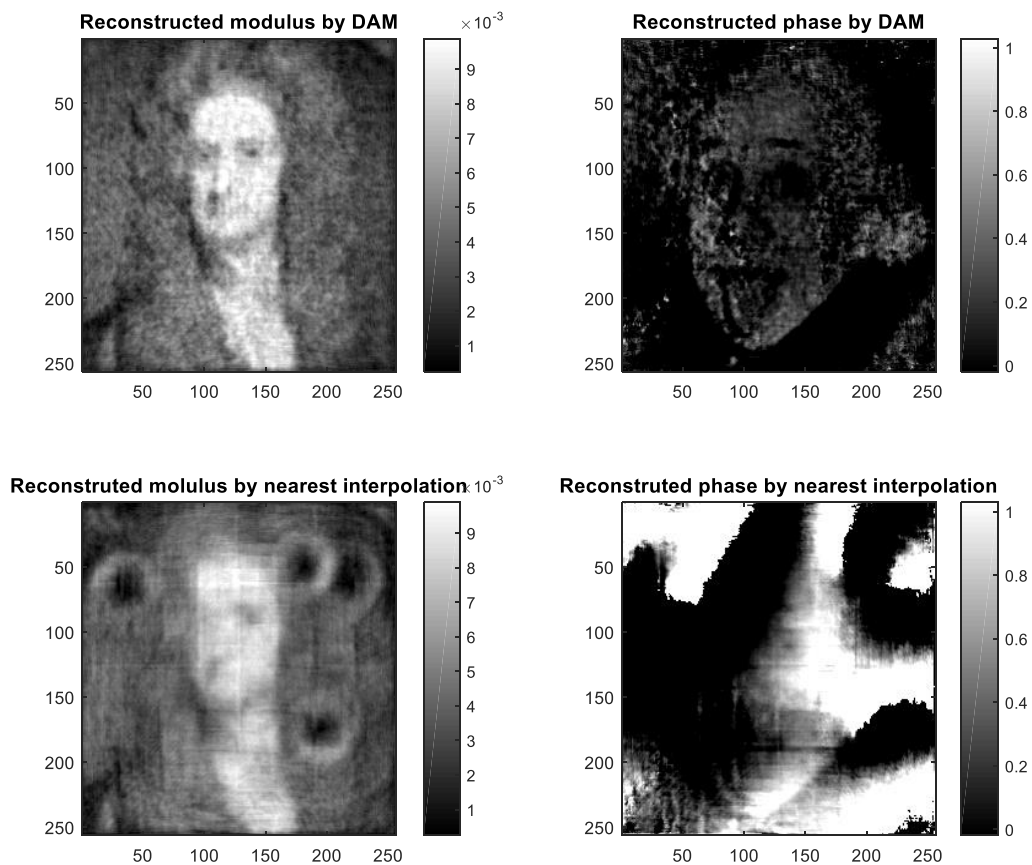

Figure 3.20 Reconstructed pupil function by DAM and nearest neighbor interpolation under lower SNR. The DAM still provides better results.
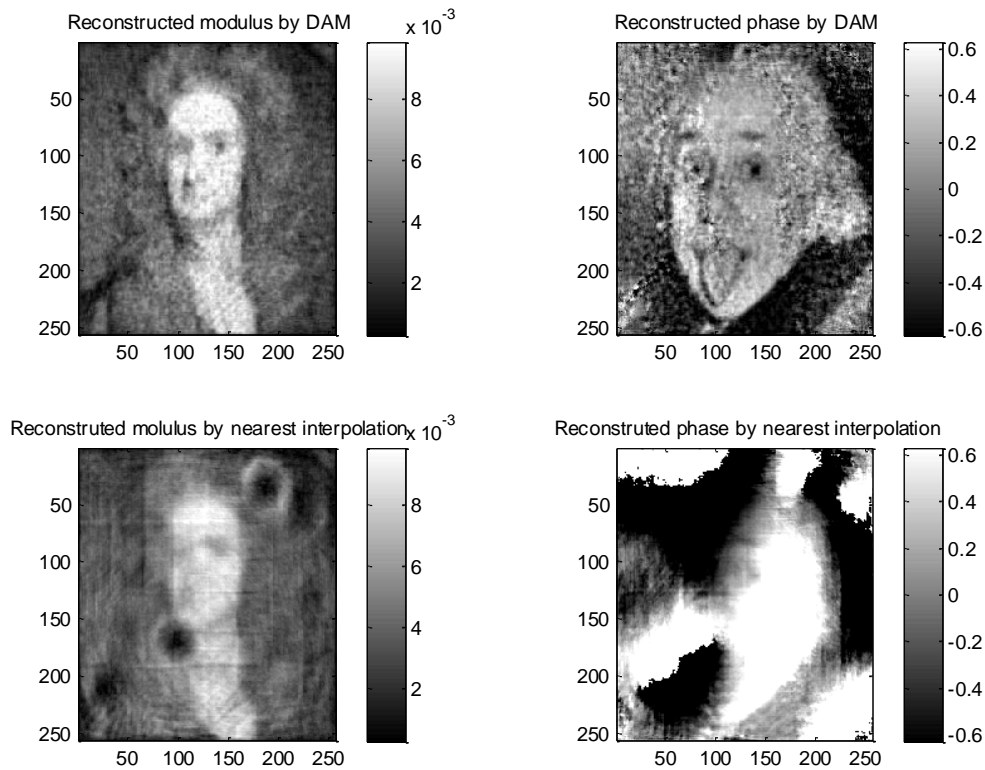
Figure 3.21 Reconstructed pupil function: well-corrected. DAM provides good results for both modulus and phase reconstruction.

Let us go back to Figure 3.11, from the blue line we observed that the Fienup-style algorithm, after using DAM, decreased the objective function fast just in the first few iterations. Also, since sub-iteration 1 is just Fienup-style algorithm, and at the first few main iterations, the inputted Fourier modulus may not be close enough to the original one, so the squared error may be decreased only in the first few iterations and then it will increase. Thus we set sub-iteration 1 running once in each main iteration. Then, from the simulation results shown in Figure 3.22 and Figure 3.23, we can directly see the advantage of setting one sub-iteration 1 and one sub-iteration 2 in each main iteration.
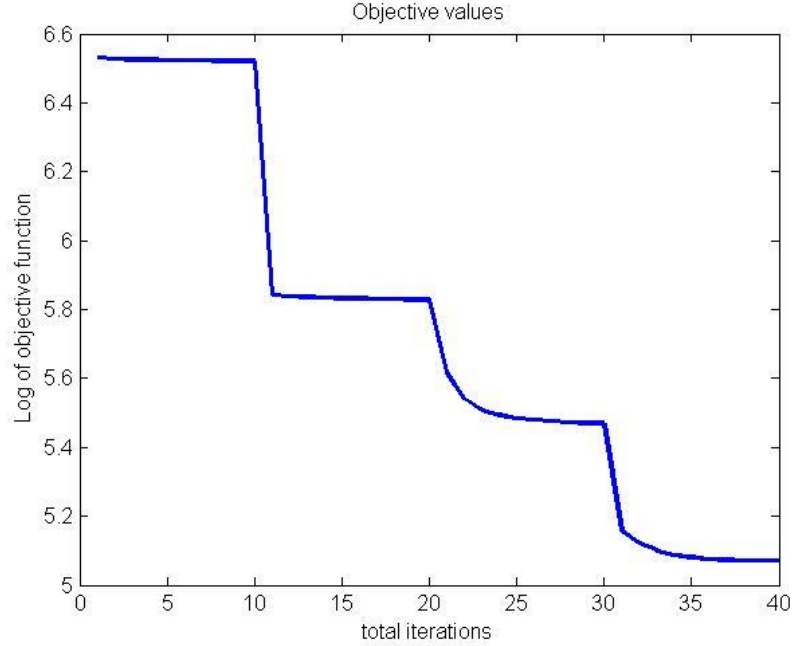
Figure 3.22 Two main iterations, each has 10 sub1 and 10 sub2. The low convergence rate shown in each sub-iteration.

I ran main iteration two times and each main iteration contains 10 sub-iteration 1 and 10 sub-iteration 2. In Figure 3.22, the convergence rate is slow in each sub-iterations and fast at the points where sub-iterations changed to another, such as at points 10, 20 and 30. By using my own computer with Intel(R) Core(TM) i7-4710MQ CPU @ 2.50 GHZ processor, the total running time for this configuration is 121.5611 seconds. Next, I measured the running time for one main iteration that has one sub-iteration 1 and one sub-iteration 2 to be around 7 seconds. Thus, in order to obtain almost the same running time of 121.5611 seconds, I ran the main iteration 17 times in this altered configuration, resulting in a running time of 116.4846 seconds. Then the log of objective function is shown in Figure 3.23 to illustrate the convergence performance.
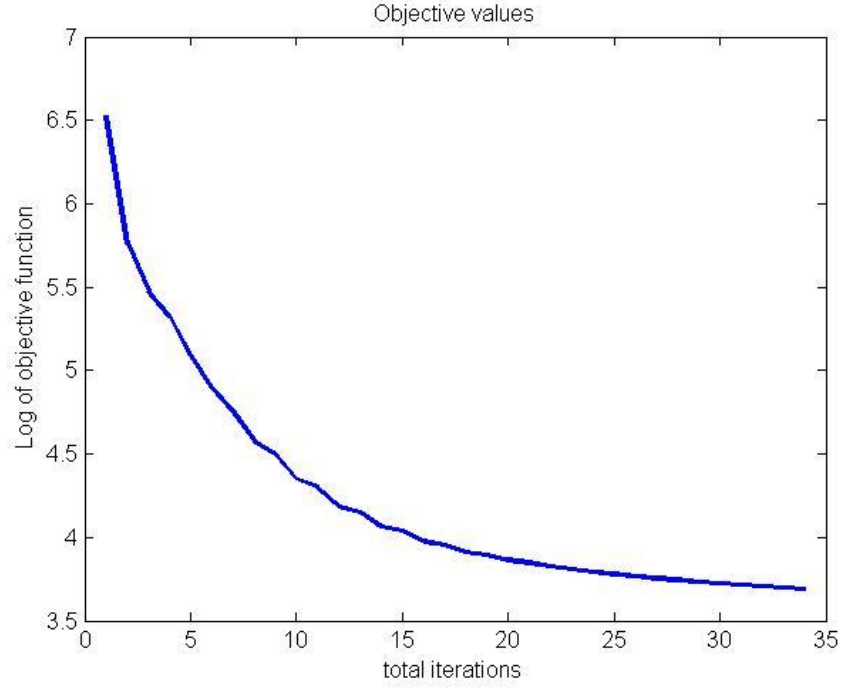
Figure 3.23 Main iterations with each has 1 sub-iteration 1 and 1 sub-iteration2. Reduce the number of sub-iteration to reduce the plateau.

From Figure 3.23, we see that the curve is smoother than the curve in Figure 3.22 and there is no plateau as shown in Figure 3.22. Thus, the convergence rate of the second configuration is faster than the first configuration. The reconstructed pupil functions from these two configurations are shown in Figure 3.24 and Figure 3.25.
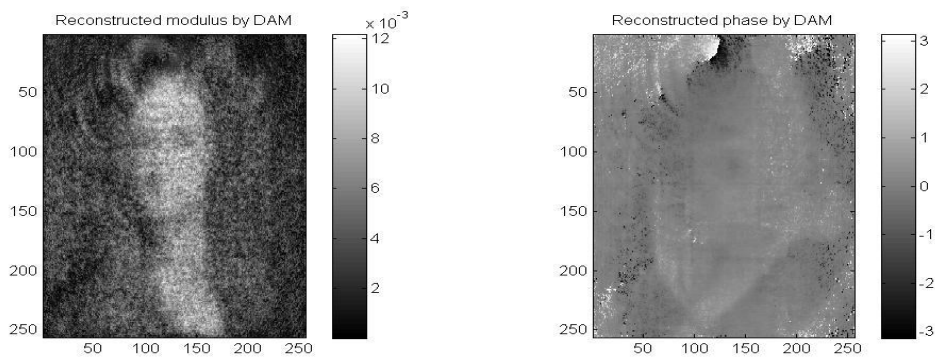

Figure 3.24 Reconstruction by 2 main iterations with 10 sub-iteration 1 and 10 sub-iteration 2.
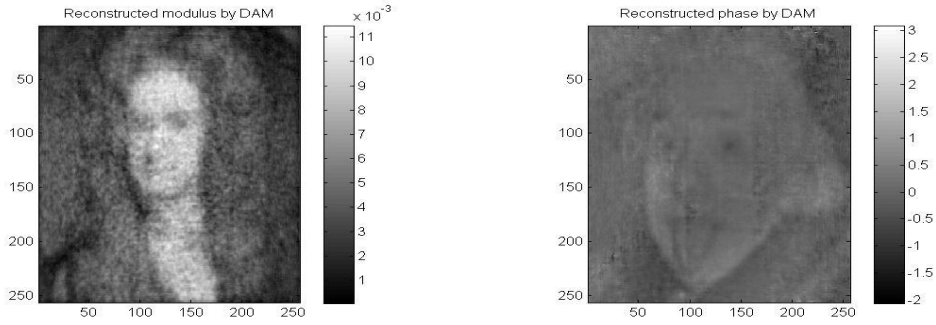
58

Figure 3.25 Reconstruction by 17 main iterations with 1 sub-iteration 1 and 1 sub-iteration 2

By comparing these two results, we observed that Figure 3.25 is more accurate than Figure 3.24. This result is consistent with the result that the final objective value of Figure 3.23 is less than Figure 3.22. The running time of both configurations is almost the same. Therefore, we prefer to choose the configuration in which each main iteration has only one sub-iteration 1 and one sub-iteration 2. However, the optimal choice of penalty parameter β is unclear, and we set it equal to one for all of the simulations. Although we did not discover the optimal choice of β, the weighting effect of it is obvious. The larger the β is chosen, the more rigorous constraints will be introduced. Thus, by the Penalty Convergence Theorem stated in [26], if the sequence of solutions of the penalized problem with different penalty parameters ordered from small to large will be convergent, then the convergent point is the solution of solving the constrained problem. Thus, if β is chosen sufficiently large, upon its convergence, the result will be the constrained problem solution. However, if we choose large β, the convergence rate becomes low since the update in sub-iteration 2 becomes mainly dependent on $F_i(u', v')$ in Equation (3.25). Thus, the balance between more accurate results by using a larger penalty parameter and the faster running time by choosing a smaller penalty parameter depends on the user's need.

Chapter 4

Conclusions and future works

Phase retrieval is a classical problem in microscopy. During the photon detection process, the phase information of wavefronts is lost and only the magnitude can be measured. Currently, two methods are widely used: the Fienup-style algorithm and the Gerchberg-Saxton algorithm. However, neither of these two methods consider the sampling rate at the detector plane. If the digitized pupil function has a higher dimension than the dimension of the detector arrays, the phase retrieval algorithm must be implemented after interpolating the measured images. In order to obtain the most probable pupil function, we need to find the most likely defocused Fourier modulus that consistent with the measured data. Here we developed a new algorithm that can estimate the defocused Fourier modulus by minimizing penalized I-divergence. By simulating the condition in which one hundred million photons are collected (ten trials), we show that the DAM-reconstructed effective pupil function has a root-mean-squared error of about 43±3% less than the effective pupil function reconstructed by Fienup-style algorithm with nearest neighbor interpolation.

In this thesis, I used simulated data. This new algorithm still needs to be tested with real data in order to verify its performance in the future. We can also use the reconstructed pupil function in the process of localizing single molecules, and then compare the variances of the estimated locations from using different reconstructed pupil functions in the future. The penalty parameter has been set to one for all of my simulations, so the performance of using different penalty parameters still need to be quantified. In addition, I simulated two phase ranges of pupil functions in this work, so the performance of this algorithm for other ranges of phase needs to be quantified in the future.

Given the promising reconstructed pupil functions estimated by our new algorithm in simulations, I believe that DAM has the ability to model imaging systems more accurately, and thus can further improve the image quality in microscopy.

Appendix A

Proof of Monotonicity of Gerchberg-Saxton algorithm

We consider the discrete case and the Fourier transform becomes Discrete Fourier Transform. The squared error in the object domain is:

$$E_o^{(m)} = \sum_{x,y} (\,|f(x,y)| - |\hat{f}^{(m)}(x,y)|\,)^2$$

$$= \sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m)}(x,y)|^2 \tag{A.1}$$

By Parseval's theorem, the squared error in the Fourier domain is:

$$E_F^{(m)} = N^{-2} \sum_{u,v} (|F(u,v)| - |G^{(m)}(u,v)|)^2$$

$$= N^{-2} \sum_{u,v} (\hat{G}^{(m)}(u,v) - G^{(m)}(u,v))^2$$

$$= \sum_{x,y} |f^{(m)}(x,y) - \hat{f}^{(m)}(x,y)|^2 \tag{A.2}$$

From the definition of forcing function constraint, it is finding the closest point in the constraint defined set to the $\hat{f}^{(m)}(x,y)$ at m-th iteration, we can have:

$$\sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^m(x,y)|^2 \leq \sum_{x,y} |f^{(m)}(x,y) - \hat{f}^{(m)}(x,y)|^2 \tag{A.3}$$

And from the definition of forcing Fourier constraint which is finding the closest point to $G^{(m+1)}(u,v)$ in the Fourier constraint set at (m+1)-th iteration, we have:

$$N^{-2} \sum_{u,v} (\hat{G}^{(m+1)}(u,v) - G^{(m+1)}(u,v))^2 \tag{A.4}$$

$$\leq N^{-2} \sum_{u,v} (\hat{G}^{(m)}(u,v) - G^{(m+1)}(u,v))^2$$

According to Parseval's theorem, we have:

$$\sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m+1)}(x,y)|^2 \leq \sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m)}(x,y)|^2 \tag{A.5}$$

Thus we have:

$$\sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m+1)}(x,y)|^2 \leq \sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m)}(x,y)|^2$$

$$\leq \sum_{x,y} |f^{(m)}(x,y) - \hat{f}^{(m)}(x,y)|^2 \tag{A.6}$$

So $\sum_{x,y} |f^{(m+1)}(x,y) - \hat{f}^{(m+1)}(x,y)|^2 \leq \sum_{x,y} |f^{(m)}(x,y) - \hat{f}^{(m)}(x,y)|^2$

By Parseval's theorem again, we can have:

$$N^{-2} \sum_{u,v} (\hat{G}^{(m+1)}(u,v) - G^{(m+1)}(u,v))^2 \leq N^{-2} \sum_{u,v} (\hat{G}^{(m)}(u,v) - G^{(m)}(u,v))^2 \tag{A.7}$$

Thus, $E_F^{(m+1)} \leq E_F^{(m)}$, the squared error will either decrease or stay the same after each iteration.

References

[1] Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* 313, 1642–5 2006.

[2] Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* 91, 4258–72 2006.

[3] Rust, M. M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* 3, 793–795 2006.

[4] Sharonov, A. & Hochstrasser, R. M. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18911–6 2006.

[5] M. A. Thompson, J. S. Biteen, S. J. Lord, N. R. Conley, and W. E. Moerner, "Molecules and Methods for Super-Resolution Imaging," *Methods in enzymology*, vol. 475, pp. 27–59, 2010.

[6] A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," *Nat Meth*, vol. 11, no. 3, pp. 267–279, Mar. 2014.

[7] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober, "Quantitative study of single molecule location estimation techniques," *Opt. Express*, vol. 17, no. 26, pp. 23352–23373, Dec. 2009.

[8] S. J. Sahl and W. Moerner, "Super-resolution fluorescence imaging with single molecules," *Current Opinion in Structural Biology*, vol. 23, no. 5, pp. 778–787, Oct. 2013.

[9] S. J. Holden, S. Uphoff, and A. N. Kapanidis, "DAOSTORM: an algorithm for high- density super-resolution microscopy," *Nat Meth*, vol. 8, no. 4, pp. 279–280, Apr. 2011.

[10] S. Quirin, S. R. P. Pavani, and R. Piestun, "Optimal 3D single-molecule localization for superresolution microscopy with aberrations and engineered point spread functions," *Proceedings of the National Academy of Sciences*, vol. 109, no. 3, pp. 675–679, Jan. 2012.

[11] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, "Optimal Point Spread Function Design for 3D Imaging," *Phys. Rev. Lett.*, vol. 113, no. 13, p. 133902, Sep. 2014.

[12] B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "Phase retrieval for high-numerical-aperture optical systems," *Opt. Lett.*, vol. 28, no. 10, pp. 801–803, May 2003.

[13] B. M. HANSER, M. G. L. GUSTAFSSON, D. A. AGARD, and J. W. SEDAT, "Phase-retrieved pupil functions in wide-field fluorescence microscopy," *Journal of Microscopy*, vol. 216, no. 1, pp. 32–48, 2004.

[14] E. Abbe, "Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung," *Archiv für mikroskopische Anatomie*, vol. 9, no. 1, pp. 413–418.

[15] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. E. Moerner, "On/off blinking and switching behaviour of single molecules of green fluorescent protein," *Nature*, vol. 388, no. 6640, pp. 355–358, Jul. 1997.

[16] M. Badieirostami, M. D. Lew, M. A. Thompson, and W. E. Moerner, "Three-dimensional localization precision of the double-helix point spread function versus astigmatism and biplane," *Applied Physics Letters*, vol. 97, no. 16, 2010.

[17] G. Grover, S. R. P. Pavani, and R. Piestun, "Performance limits on three-dimensional particle localization in photon-limited microscopy," *Opt. Lett.*, vol. 35, no. 19, pp. 3306–3308, Oct. 2010.

[18] V. N. Mahajan, "Zernike Circle Polynomials and Optical Aberrations of Systems with Circular Pupils," *Appl. Opt.*, vol. 33, no. 34, pp. 8121–8124, Dec. 1994.

[19] R. Gerchberg and O. Saxton, "A practical algorithm for the determination of the phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[20] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug. 1982.

[21] Richard G. Paxman, Timothy J. Schulz, and James R. Fienup, "Joint estimation of object and aberrations by using phase diversity," J. Opt. Soc. Am. A 9, 1072-1085 1992

[22] S. Boyd and J. Dattorro, "Alternating Projections." Autumn-2003.

[23] Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman, "On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints," *Computational Optimization and Applications*, vol. 51, no. 3, pp. 1065–1088, 2011.

[24] C. Preza and J. A. O'Sullivan, "Quantitative phase and amplitude imaging using Differential-Interference Contrast (DIC) microscopy," vol. 7246, pp. 724604–724604–11, 2009.

[25] J. A. O'Sullivan, D. L. Snyder and B. R. Whiting, "Alternating minimization algorithms for transmission tomography using energy detectors," Signals, Systems and Computers, 2002. Conference Record of the Thirty-Sixth Asilomar Conference on, Pacific Grove, CA, USA, pp. 144-147 vol.1. doi: 10.1109/ACSSC.2002.1197165, 2002

[26] Robert M. Freund , "Penalty and Barrier Methods for Constrained Optimization." February, 2004.

[27] Saleh, B. E. A., & Teich, M. C. Fundamentals of photonics. New York: Wiley, 1991.

[28] Rafael C. Gonzalez and Richard E. Woods. Digital Image Processing (3rd Edition). Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[29] "History of Microscopes", http://www.microscope.com/education-center/microscopes-101/history-of-microscopes/.

[30] Matthew D. Lew, "Lecture 19: Survey of Biological Imaging Techniques", ESE 582: Fundamentals and Applications of Modern Optical Imaging, Washington University in St. Louis, MO, USA, 2015.