Application of Wavefront Sensorless Methods to Shack-Hartmann Patterns for Wavefront Reconstruction

A Two-step Method for Single-Frame Phase Retrieval

Liam Cools

TUDelft Delft University of Technology

Application of Wavefront Sensorless Methods to Shack-Hartmann Patterns for Wavefront Reconstruction

A Two-step Method for Single-Frame Phase Retrieval

MASTER OF SCIENCE THESIS

Liam Cools

November 25, 2023

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
DELFT CENTER FOR SYSTEMS AND CONTROL

The undersigned hereby certify that they have read and recommend to the Faculty of Mechanical, Maritime and Materials Engineering (3mE) for acceptance a thesis entitled

APPLICATION OF WAVEFRONT SENSORLESS METHODS TO SHACK-HARTMANN PATTERNS FOR WAVEFRONT RECONSTRUCTION by

LIAM COOLS

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE SYSTEMS & CONTROL

Dated: <u>November 25, 2023</u>

Supervisor(s):

—————————————————
dr. O.A. Soloviev

—————————————————
prof.dr.ir. M.H.G. Verhaegen

Reader(s):

—————————————————
dr. O.A. Soloviev

—————————————————
prof.dr.ir. M.H.G. Verhaegen

—————————————————
dr. C.S. Smith

Abstract

Counter-acting image quality degradation caused by phase aberrations through physical correction requires the phase field to be known. As imaging hardware captures real-valued intensity, defined as the wave amplitude squared, obtaining this lost phase information is known as the phase retrieval problem and is a non-linear and non-convex optimisation problem.

Literature treats this problem from two points of view: reconstruction using indirect phase sampling, and reconstruction using Fourier amplitude sampling. The former employs wavefront sensors such as the Shack-Hartmann sensor, which encodes phase gradient information in the form of the displacement of imaged spots. These methods, while fast, discard information such as interference and are limited to low-order reconstruction. The latter, also called wavefront sensorless phase retrieval, uses the full point-spread function (PSF) to obtain high-accuracy reconstruction at the cost of computational speed, number of intensity images required, or limited aberration magnitude. These two points of view have remained largely separated, but can be made compatible through a modelling technique called Shack-Hartmann diversity. This thesis explores the merging of wavefront sensorless methods with Shack-Hartmann intensity patterns to leverage more information from a single captured image frame.

Firstly, a low-order modal reconstruction technique is presented applied to phase gradient fields obtained from Fourier demodulation of a Shack-Hartmann intensity pattern. A method of minimising the amount of redundant data-points used for reconstruction is illustrated through the removal of Fourier-interpolated data to speed up performance.

Secondly, a novel extension of the Fourier demodulation technique to hexagonal Shack-Hartmann arrays is presented, allowing phase gradient extraction and modal reconstruction of hexagonal array intensity images using Fourier demodulation.

Thirdly, Shack-Hartmann diversity is used to extend intensity-based modal phase retrieval using Taylor approximation of the intensity function to Shack-Hartmann intensity patterns. This bridges the gap between wavefront sensorless methods and Shack-Hartmann intensity patterns.

Lastly, a novel hybrid method is presented for high-accuracy phase reconstruction based on applying the above intensity-based method to a single Shack-Hartmann intensity pattern with low-order pre-conditioning obtained from Fourier demodulation. The method is demonstrated to obtain highly accurate reconstruction on low-order aberrations, and better reconstruction accuracy on small-magnitude high-order aberrations with dominating large-magnitude low-order terms than traditional methods. Potential use cases are discussed, such as open-loop turbulence reconstruction for use in turbulence modelling.

Table of Contents

List of Figures

Acknowledgements

Before diving into the wonderous world of optical phase retrieval, there are several people without whom this project would not have been possible that I would like to explicitly thank.

Firstly, I would like to thank my daily supervisor dr. Oleg Soloviev, whose advice, feedback, and support has been invaluable during my thesis project. The weekly seminar meetings not only gave inspiration to try new methods and solve faced problems, but it also acted as a motivator to make progress each week. I was given a lot of freedom to explore what to research, and how to solve the research questions that arose for which I am incredibly grateful. I would also like to thank dr. Dimitri Kromm for the fruitful discussions during the seminars, friendly chats, advice, and insight on the practical challenges of phase retrieval.

Furthermore, I would like to thank my main supervisor prof.dr.ir. Michel Verhaegen for the feedback, advice, and ideas provided at every milestone throughout the project. This advice not only helped shape the project into what it has become, but also helped me improve in how I could best show it to others.

I would also like to thank my friends Knut, Ali, Sachin, and Arvis. Thank you for standing by me, supporting me, advising me, tolerating my ramblings. Thank you for the work we tackled together, for the challenges successfully faced both in the Bachelors and in the Masters, and for teaching me so many different things.

Lastly, I would like to thank my ever-supportive family: my mother Ilona, father Paul, and brother Jonas. They supported me all throughout my journey, all throughout my life. Thank you to my parents for caring for me, for supporting me, for teaching me the value of hard work. Thank you to my brother for teaching me tenacity, to stand up again after falling, and to finish what you have started.

Delft, University of Technology
November 25, 2023

Liam Cools

Chapter 1

Introduction

Astronomy, microscopy, laser-based telecommunication, earth-observation, ophthalmology, crystallography, bio-medicine, and countless others. While these fields appear to have very little in common with one another, they all rely on the capturing and analysis of images, and have a vested interest in improving the quality of these images[2].

When light is modelled by a wave, degradation in the quality of such captured images can be explained by local delays introduced to the wave, variations in the wave phase field. These delays, called aberrations, influence the propagation of the light and thereby the wave field at the time of capturing its intensity. Misalignment, the shape of reflective surfaces, lenses (in the eye or in the optical path), even differences in density of the medium through which the light travels are all potential sources of delay. This last source is often called turbulence, the major source of inaccuracy in astronomy already identified by Newton as being the main limiting factor to achievable resolution with telescopes[3]. Should these delays be known, correction of the light could be performed using adaptive optics through phase conjugation, the local introduction of delays that cancel out the original aberration[2, 4], before it is ever captured. Correction prior to capture avoids the detrimental effects of pixel discretisation to numerical image improvement algorithms, and as such is the best path to improving image quality.

However, as image intensity is wholly defined by wave amplitude at the detector plane, all phase information is lost at moment of capture. Retrieving this lost phase information from intensity is called the phase retrieval problem, it is non-linear and non-convex, has no analytical solution, and has a long and challenging history with new discoveries being made all the time[5–7].

Broadly, the field of phase retrieval is divided into two categories. The first are algorithms using additional optical components to extract phase information from the intensity image. These methods are called wavefront sensor phase retrieval methods, with a popular sensor being the Shack-Hartmann wavefront sensor[8] which will also be used in this thesis, and a popular algorithm being centroiding[9]. These algorithms are often based on simple concepts such as the displacement of imaged spots, and therefore discard a lot of information encoded in the sensor point-spread function (PSF) in favour of obtaining a fast low-accuracy approximation of the aberration. The second category contains the wavefront sensorless methods utilising only the camera module image, a single-lens ideal system which can be modelled using a single Fourier transform. Due to the lack of readily encoded phase information, these methods use the full PSF and as such can obtain high-accuracy phase reconstruction.

However, this comes at the cost of convergence guarantees[10], regions of applicability[11], or curse of dimensionality[12].

This thesis proposes the merging of these two categories using a Shack-Hartmann sensor modelling technique called Shack-Hartmann diversity[13]. Using this diversity, the Shack-Hartmann intensity pattern can be seen as the image captured by an augmented single-lens ideal system. This allows the use of wavefront sensorless methods on Shack-Hartmann intensity patterns.

The resulting novel method is a hybrid method, using wavefont sensor phase retrieval to obtain an initial phase estimate as a pre-conditioner to a wavefront sensorless method. However, additional novelties are such as the application of Fourier demodulation to hexagonal Shack-Hartmann intensity patterns, and the use of modal reconstruction on multi-directional shrunken derivative fields will also be presented.

The structure of this thesis is as follows. Theoretical background information required for this thesis will be provided in Chapter 2, with Section 2-1 discussing optical propagation methods as well as introducing the phase retrieval problem, Section 2-2 discussing the modal representation of phase fields, and Section 2-3 introducing the design and reconstruction principles behind the Shack-Hartmann wavefront sensor.

Subsequently, a brief overview of the state-of-the-art in phase retrieval will be provided in Chapter 3. This chapter will discusses both methods for phase retrieval with wavefront sensors, as well as phase retrieval for single-lens ideal systems (with)out phase diversities in Section 3-1 and Section 3-2 respectively. The research question will also be introduced in Section 3-3.

The contents of Chapter 4 will detail the classical model of a Shack-Hartmann sensor in Section 4-1, followed by an alternative phase diversity model in Section 4-2 known as the Shack-Hartmann diversity model.

Afterwards, Chapter 5 will focus on the explanation and derivation of a low-order modal phase retrieval method based on the fitting of subsampled derivative fields obtained from Fourier demodulation. The concept of Fourier demodulation will be explained, its numerical implementation discussed, and the reduction of redundant data through trimming of the spectral domain will be treated. Additionally, a novel extension of Fourier demodulation to hexagonal Shack-Hartmann intensity patterns will be discussed.

The contents of Chapter 6 will detail the concept of Taylor intensity approximation for high-accuracy and high-order modal phase retrieval for small-magnitude phases. This will also be extended to Shack-Hartmann patterns through the use of Shack-Hartmann diversity, forming a novel application of Taylor intensity approximation.

Lastly, a novel hybrid method combining the low-order retrieval of Chapter 5 with the high-order Taylor Shack-Hartmann intensity approximation method using pre-conditioning will be discussed in Chapter 7. Potential use-cases will be provided in Section 7-2 and results discussed in Section 7-3.

An overall conclusion and outlook will be provided in Chapter 8.

Chapter 2

Theoretical Fundamentals

The aim of this chapter is to provide the reader with the fundamental basics required to understand the field of optical phase retrieval in the context of this thesis. Firstly, optical propagation methods and the phase retrieval problem will be introduced in Section 2-1. Secondly, Section 2-2 will discuss the representation of wavefront phase using basis functions. Lastly, Section 2-3 will detail the base operating principle of the Shack-Hartmann wavefront sensor and related wavefront reconstruction methodology.

2-1   Optical Propagation and the Phase Retrieval Problem

When considering the propagation of light through an optical system, two interpretations are traditionally considered: modelling light as a ray whose direction is altered by optical components, and modelling light as a wave function.

The former, known as geometrical optics, is suited for use-cases such as the creation of lenses, or the general analysis of where images are expected to be positioned after propagation. However, it fails to capture real-world observed phenomena such as light interference patterns, diffraction, or light coherence. The latter is called wave optics or Fourier optics [14] and provides a framework to capture these phenomena mathematically.

The purpose of this section is to provide a fundamental basis in Fourier optics propagation equations as used in phase retrieval, for more in-depth information the reader is referred to relevant literature[14].

While light is a three-dimensional wavefield, an immediate simplification will be made: the wave field will be considered at two-dimensional slices along the assumed axis of propagation. The wavefront is expressed at each slice along this propagation axis using Equation 2-1 where $\mathcal{A}(x, y)$ contains the wave amplitude at each point and $\phi(x, y)$ is the phase field. Explicit time dependence, whose presence would be expected due to the oscillatory nature of wave functions, has also been removed. The rationale behind this removal is the capture rate of optical detectors. Visible light, made up of wavelengths in the 600nm range, oscillates at frequencies around $\frac{c}{\lambda} \approx \frac{3 \cdot 10^8}{6 \cdot 10^2 10^{-9}} \propto 10^{14}$ times per second where c is

the speed of light. As no detector is capable of capturing this, multiple oscillations are averaged and time dependence is ignored[15].

$$W(x,y) = \mathcal{A}(x,y)e^{i\phi(x,y)} \tag{2-1}$$

Wavefront propagation is performed using propagation equations, making use of the Fourier domain to simplify calculations[14]. The most general propagation method is the Angular Spectrum method, based on Rayleigh-Sommerfeld diffraction. Its working principle is the the decomposition of the wavefront into differently oriented plane waves, followed by propagation and re-combination at the requested distance z. The propagation equation for this method is given inEquation 2-2[14]. It is the most computationally expensive as well as the most analytically complex propagation method, and as such many simplifications can be applied to it.

$$\begin{cases} W(x,y,z) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{W(x,y,0)\right\}H(f_x,f_y,z)\right\} \\ H(f_x,f_y,z) = e^{i2\pi z\sqrt{\frac{1}{\lambda^2}-f_x^2-f_y^2}} \end{cases} \tag{2-2}$$

The Fresnel propagation method is the first of such simplifications[14]. By making the approximation $\frac{1}{\lambda}\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2} \approx \frac{1}{\lambda}\left(1-\frac{(\lambda f_x)^2}{2}-\frac{(\lambda f_y)^2}{2}\right)$, Equation 2-2 is simplified to Equation 2-3. This simplification makes use of the paraxial approximation, meaning light travelling nearly parallel to the optical propagation axis is assumed. As light emitted from a point-source does so in a spherical manner, this means the point-source is either considered to be further away to obtain a small curvature, or the wave is analysed close to the propagation axis. As can be seen from the equation, this method is computationally slightly cheaper and analytically simpler to understand, but it too requires a double Fourier transformation. Further simplification is desirable. For this purpose, the Fresnel diffraction integral is given in Equation 2-4 where (x_o, y_o) are the coordinates in the object plane, and (x_i, y_i) are the coordinates in the image plane.

$$\begin{cases} W(x,y,z) = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{W(x,y,0)\right\}H(f_x,f_y,z)\right\} \\ H(f_x,f_y,z) = e^{ikz}e^{-i\pi\lambda z(f_x^2+f_y^2)} \end{cases} \tag{2-3}$$

$$W(x_i,y_i,z) = \frac{e^{ikz}}{i\lambda z}\int W(x_o,y_o,0)e^{i\frac{k}{2z}\left[(x_i-x_o)^2+(y_i-y_o)^2\right]}dx_o dy_o \tag{2-4}$$

If the propagation distance z is assumed to significantly exceed the wavelength such that $z >> \lambda$, the approximation $e^{i\frac{k}{2z}(x_o^2+y_o^2)} \approx 1$ can be made. This simplifies Equation 2-4 to Equation 2-5. This simplified propagation equation is known as Fraunhofer propagation, and is usually employed when considering far-field propagation[14]. As Equation 2-5 shows, propagation is performed using a single Fourier transform and a multiplication with a pre-factor. The field obtained through Fraunhofer propagation is also closely related to an ideal single-lens system imaging from the front to the back focal plane[14]. If pre-factors are ignored, a common practice, this results in a perfect Fourier transform relation for both far-field diffraction and single-lens system imaging. This method is computationally very cheap due to efficient Fourier transform implementations such as the fast Fourier transform (FFT) algorithm, and is analytically highly interesting.

$$\begin{cases} W(x_i, y_i, z) = \frac{e^{ikz}}{i\lambda z} e^{-i\frac{k}{2z}(x_i^2 + y_i^2)} \underbrace{\int W(x_o, y_o, 0) e^{-i\frac{2\pi}{\lambda z}(x_i x_o + y_i y_o)} dx_o dy_o}_{\mathcal{F}\{W(x,y,0)\}} \\ x_i = \lambda z f_x \\ y_i = \lambda z f_y \end{cases} \tag{2-5}$$

The direct Fourier transform relation obtained for propagation through an ideal single-lens system is also used to obtain the wavefront at the detector plane of a camera module, assumed to consist of a detector chip placed in the focal plane of a single-lens system. These detector chips measure light intensity, i.e. the number of photons hitting individual pixels over a given period of time, modelled mathematically as $I(x, y) = W^*(x, y)W(x, y) = |\mathcal{A}(x, y)|^2$ using the wave description given in Equation 2-1. Due to the single-lens model of a camera, this means the incoming wavefront intensity is captured by a camera module with focal length f as $I(x, y, z = f)\left(\mathcal{F}\{W(x, y, 0)\}\right)^* \mathcal{F}\{W(x, y, 0)\}$.

Unfortunately, this real-valued intensity does not explicitly capture the phase field $\phi(x, y)$, which is present in the complex wavefront. This is problematic as degradation of the captured image quality can mostly be attributed to phase delays incurred along the optical path, also called aberrations. These aberrations are included in the phase field $\phi(x, y)$, where a perfect plane wave has $\phi(x, y) = 0$ and generates the highest quality image. If the aberrated phase $\phi(x, y)$ is known, the principle of optical conjugation can be used to correct the aberration. This is the base principle on which the field of adaptive optics is founded[2]. Through a corrective element, for example a deformable mirror, the phase $\phi_{corrector}(x, y) = -\phi(x, y)$ can be introduced to the optical path of the wavefront which cancels out the original aberration.

The retrieval of this lost phase information from an intensity image is a non-convex and highly non-linear problem called the phase-retrieval (PR) problem. The optimisation problem given in Equation 2-6 describes PR mathematically, where the general optimisation parameter x then contains the complex wavefront from which the phase can be calculated. Due to the non-linear and non-convex nature of this problem, it is still a highly active area of research[2, 5–7].

$$\begin{cases} |\mathcal{F}\{x\}| = A \\ |x| = a \\ x \in \mathbb{C}^N \end{cases} \tag{2-6}$$

Solving the PR problem directly from intensity images registered with single-lens ideal systems is called wavefront sensorless phase retrieval. These methods are often iterative, require multiple images of the wavefront with introduced phase diversities, or attempt to reduce or eliminate non-linearity or non-convexity at the cost of aberration magnitude and problem dimensionality[5]. A brief overview of these methods in literature will be given in Section 3-2.

Phase information can be captured in the intensity image if alternative optical setups are used. An example would be the use of holography, which registers a complete phase profile in the form of interference patterns but requires a specialised setup[5]. Instead, so called wavefront sensors are employed which capture phase information indirectly and can be incorporated more easily within an optical setup. A well-known wavefront sensor in the field of adaptive optics is the Shack-Hartmann sensor[8], designed to capture average phase gradient information in the form of spot displacements. More information on the operating principle of the Shack-Hartmann sensor will be provided in Section 2-3, and phase retrieval based on Shack-Hartmann intensity patterns will be discussed in Section 3-1.

2-2   Modal Basis Decomposition

The representation of a wavefront phase for use in phase retrieval methods can be done zonally or modally. As the name suggests, zonal representation defines the wavefront over a set of zones as being either piece-wise constant or piece-wise linear. The size of such a zone can range from the full aperture to a single pixel, where the latter is the most common in numerical zonal phase retrieval algorithms such as alternating projections[10]. The higher the zone count or pixel grid becomes, the more data is required to represent the wavefront phase making zonal representation the least efficient in terms of storage if retrieved phases are to be saved. Each zone is treated independently meaning a smooth phase is not guaranteed, but conditions on continuity can be applied between neighbouring zones if desired.

Modal representation, on the other hand, expresses the phase field globally as a weighted sum of basis functions as shown in Equation 2-7 where K refers to the number of basis functions used in the reconstruction. An advantage of modal representation is that, given a smooth set of basis functions, the resulting modal wavefront will also be smooth. As common naturally ocurring aberrations such as defocus, astigmatism and tip-tilt are smooth functions, modal representation is especially suited. Additionally, saving the phase for later use only requires saving a set of coefficients which is much more efficient in terms of consumed memory.

$$\phi(x,y) = \sum_{k=1}^{K} \alpha_k f_k(x,y) \tag{2-7}$$

A highly popular basis for phase representation and the basis used in this thesis is the Zernike basis which is complete as well as orthonormal on the unit circle and as such used for circular apertures[16]. In polar coordinates, each basis function is defined by Equation 2-8 where n is the radial order and m the azimuthal order, $n, m \geq 0$, and $m \leq n$. The basis functions are usually referred to by their (n, m) denomination, but other orderings are also used in practice such as Noll's ordering using a single index associated to each $(n, \pm m)$ pair[17].

$$\begin{cases} Z_n^{\pm m}(\rho, \theta) & = R_n^m \begin{cases} cos(m\theta), & \pm m > 0 \\ sin(m\theta), & \pm m < 0 \\ 1, & m = 0 \end{cases} \\ R_n^m(\rho) & = \sum_{s=0}^{\frac{n-m}{2}} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!} \rho^{n-2s} \end{cases} \tag{2-8}$$

The Zernike basis has a number of desirable properties. Firstly, the complexity of the basis functions increases with their order. As the Zernike functions are real-valued, complexity in this context refers to the amount of variation or frequency of the function over the studied domain and as such allows the modelling of more complex aberrations with higher orders. Secondly, as mentioned above, the basis is orthonormal on the unit circle and complete which means any aberration on a circular aperture can theoretically be expressed using a weighted sum of Zernike polynomials. It is worth noting, however, that orthogonality is a mathematical concept applicable in continuous space and on the full infinite basis, i.e. with $K = \infty$ in Equation 2-7.

If a truncated ($K < \infty$) basis is used, which is unavoidable when implemented numerically, orthogonality is not guaranteed to hold[18]. In fact, it is known not to hold for the Zernike basis. However, aberrations can often be assumed to be below a particular order of complexity meaning truncation of

the Zernike basis beyond this order is expected to minimise the effects of loss of orthogonality. An additional problem presents itself in the form of gradient orthogonality. Methods such as centroiding[4] or the method which will be discussed in Section 5-4, attempt to find a set of coefficients $\vec{\alpha}$ such that $\nabla\phi = \sum_{k=1}^{K} \alpha_k \nabla f_k$. However, if the gradient fields are not orthogonal to one another, this does not necessarily mean $\phi = \sum_{k=1}^{K} \alpha_k f_k$. A good example of this is the Zernike basis discussed above, whose gradient fields are not orthogonal.

There are bases in literature whose gradient fields are orthogonal, or are designed to be used for fitting gradients prior to translation to a basis such as Zernike. An example of this would be the vector polynomial basis derived by Zhao and Burge[19, 20]. This basis, however also loses orthogonality when truncated.

If orthogonality under truncation and orthogonality in gradient is desired, an additional criterion must be met which was derived by Soloviev[18]. According to this criterion, given in Equation 2-9, each basis function f_n in the basis must be a solution to the 2D Helmholtz equation in the aperture with homogeneous boundary conditions. For circular apertures, an example of such a basis would be the set of circular Bessel functions[18].

$$\Delta f_n + k_n^2 f_n = 0 \qquad , k_n \in \mathbb{R} \tag{2-9}$$

Many other bases for use in phase retrieval exist in literature such as the complex-valued Extended Nijboer-Zernike basis[21], the set of Gaussian radial basis functions, B-splines et cetera which are implemented for specific use-cases. The method developed in this thesis, however, uses the classic Zernike basis due to its popularity in optics.

Computation of the Zernike basis can be performed using Equation 2-8 if desired. However, Cartesian defintions of the polynomials exist and recurrence relations between the polynomials as well as their gradients have been exploited to speed up calculations significantly. The work in this thesis uses such an algorithm, developed originally by Andersen and translated to Python for use in this thesis[22].

2-3 Shack-Hartmann Sensor Principles

The aim of this section is to briefly introduce the Shack-Hartmann wavefront sensor and how it is traditionally used in phase retrieval. The sensor was designed in its current form in the year 1971, when Shack proposed to build an optical component comprised of an array of lenslets, called a micro-lens array (MLA), and a detector chip placed in the focal plane[8]. This design was a variation of the Hartmann screen, an opaque surface with pinholes spaced at regular intervals, proposed by Hartmann in 1904[23].

Each lenslet in the MLA images the incoming wavefront as a spot on the detector plane, but with a shift which was observed to be proportional to the average tip/tilt over each individual sub-aperture. As the average tip/tilt equals the average phase gradient over the sub-aperture, the registered intensity indirectly contains phase gradient information in the form of these spot shifts. An illustration of this principle is provided in Figure 2-1.

Calculating these spot displacements is done through a procedure called centroiding, resulting in two displacements $s_{x,i}$ and $s_{y,i}$ per sub-aperture. As can be seen in Figure 2-1 the average tip/tilt would then be $\frac{-s_{x,i}}{f}$ and $\frac{-s_{y,i}}{f}$ per sub-aperture. It is worth noting that the displacements are assumed to be in physical dimensions, meaning the physical pixel size must be known. The most common algorithm
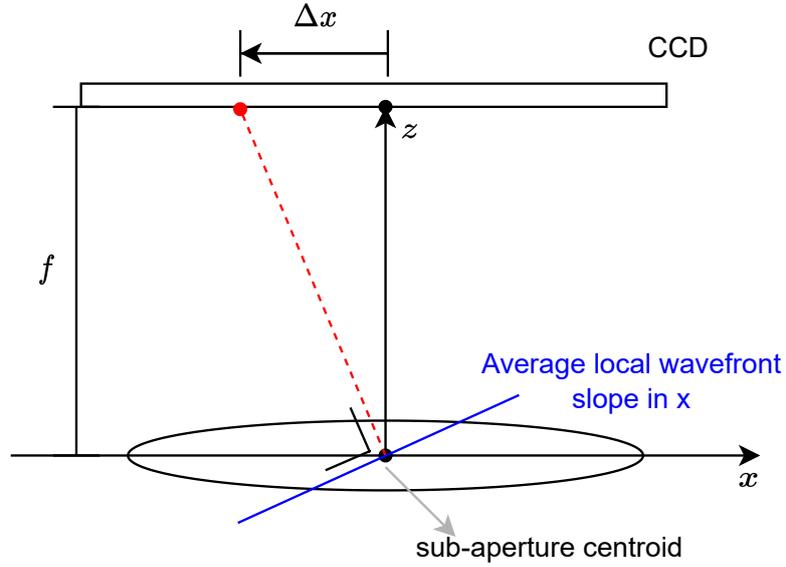
Figure 2-1: Operating principles of a Shack-Hartmann wavefront sensor. The observed shifting of the spot imaged by a lenslet is shown to be proportional to the wavefront gradient over the sub-aperture

for centroiding is based on the calculation of the first image moment per sub-aperture, but alternative methods will be briefly discussed later in Section 3-1.

If the wavefront phase is reconstructed zonally, meaning pixel-wise or region-wise, the average tip/tilts in each sub-aperture can simply be stitched together to form a piece-wise linear approximation of the wavefront[4]. This reconstruction technique often lacks a smooth phase, which may not be desirable due to the smoothness of common aberrations such as defocus, tip/tilt and astigmatism.

More commonly, however, the phase is reconstructed modally, which results in a smooth phase if smooth basis functions are used. By calculating the x and y gradients of the known basis functions in each sub-aperture centre, a gradient matrix Z can be obtained. Stacking the calculated displacements s_x and s_y into a vector $\vec{s}$ and multiplying by $\frac{1}{f}$ then provides a vector of phase gradient estimates. As this estimated gradient should correspond to a weighted combination of the basis gradients, a problem of the form given in Equation 2-10 is obtained, where the Zernike basis is used as an illustration. The best fit solution for this problem can be calculated with the least-squares solution $\vec{\alpha} = \frac{1}{f}(Z^T Z)^{-1} Z^T \vec{s}$.

$$
\begin{bmatrix}
\frac{\partial Z_1^{-1}}{\partial x}(x_0, y_0) & \dots & \frac{\partial Z_n^m}{\partial x}(x_0, y_0) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial x}(x_N, y_N) & \dots & \frac{\partial Z_n^m}{\partial x}(x_N, y_N) \\
\frac{\partial Z_1^{-1}}{\partial y}(x_0, y_0) & \dots & \frac{\partial Z_n^m}{\partial y}(x_0, y_0) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial y}(x_N, y_N) & \dots & \frac{\partial Z_n^m}{\partial y}(x_N, y_N)
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\
\vdots \\
\alpha_{m,n}
\end{bmatrix}
= \frac{1}{f}
\begin{bmatrix}
s_{x,1} \\
\vdots \\
s_{x,N_{MLA}} \\
s_{y,1} \\
\vdots \\
s_{y,N_{MLA}}
\end{bmatrix}
\tag{2-10}
$$

As centroiding provides two displacement values per sub-aperture, the maximum number of modes that can be retrieved using this least-squares solution is equal to $2N_{MLA}$ where N_{MLA} is equal to the

number of spots present in the intensity image. As large lenslet arrays are expensive to manufacture, this number is traditionally rather limited and as such centroiding-based reconstruction results in low-order approximations of the wavefront phase.

From the discussion above it becomes clear that the classic use of the Shack-Hartmann sensor disregards a lot of information encoded in the form of interference and spot shape, i.e. the point-spread function (PSF) nature of a Shack-Hartmann intensity pattern is ignored. As will become clear from the discussions in Section 6-2 and Chapter 7, however, exploiting this information results in much higher accuracy in wavefront phase reconstruction.

Chapter 3

Current State-of-the-art in Phase Retrieval

This chapter will discuss current state-of-the-art phase retrieval methods with and without the use of Shack-Hartmann wavefront sensors. Firstly, phase retrieval using wavefront sensors will be discussed in Section 3-1. Secondly, phase retrieval for single-lens ideal systems (with)out phase diversities will be discussed in Section 3-2. Lastly, the thesis research question will be formulated in Section 3-3.

3-1   Wavefront Reconstruction through Phase Sampling

As mentioned earlier in Section 2-1, the phase retrieval problem is both non-convex and non-linear as intensity patterns captured by traditional image registration only consider wavefront amplitude information at the detector plane, losing any explicit information of the phase. However, by altering the optical setup, direct or indirect registration of phase information is possible which substantially simplifies the phase retrieval problem.

One possible alteration is the use of holography, which captures the phase directly in the form of interference patterns between a reference beam and the aberrated beam[5, 24]. This, however, is a highly specific use-case and requires significant alterations to an existing optical setup. Instead, indirect phase registration can be performed using so called wavefront sensors. One common wavefront sensor is the Shack-Harmann sensor[8], consisting of a micro-lens array (MLA) where each lenslet images the wavefront as a spot on the focal plane with a spatial displacement proportional to the average phase gradient over the lenslet sub-aperture. As such, an approximation of the wavefront can be obtained which, depending on the wavefront aberration being low-order and thus changing minimally over a sub-aperture, can be used to correct aberration arbitrarily well. As the accuracy of the average gradient information is tied directly to the accuracy of the calculated displacements, literature focuses on this aspect.

The calculation of displacements is traditionally done via the procedure of centroiding, i.e. finding the spot centroid in each sub-aperture will yield the displacement with respect to the sub-aperture centre

where an unaberrated wavefront spot centroid would be located. If the intensity in a sub-aperture is denoted $I_{sub}(x, y)$, its corresponding centroid can then be obtained via calculation of the first image moment as given in Equation 3-1[4, 9]. After performing this for every sub-aperture, the procedure in Section 2-3 can be used to reconstruct the wavefront.

$$\begin{cases} x_{c,sub} & = \dfrac{\int\int I_{sub}(x,y)x\,dxdy}{\int\int I_{sub}(x,y)dxdy} \\ y_{c,sub} & = \dfrac{\int\int I_{sub}(x,y)y\,dxdy}{\int\int I_{sub}(x,y)dxdy} \end{cases} \tag{3-1}$$

As the main source of error in these centroid calculations come from the effects of noise and spatial discretisation to pixels, methods discussed in literature aim to improve upon the classic first-moment calculation. An example of this would be the work of Nightingale and Gordeyev, in which both pre-processing methods and alternative centroid calculation methods are discussed[9]. In this paper, three pre-processing methods were discussed for the removal of noise: intensity thresholding, windowing around the brightest pixel, and gamma correction. An illustration of their effects is shown in Figure 3-1. Additionally, the paper discussed three alternative centroid calculation methods. The first was first-moment calculation of the sub-aperture intensity convolved with a 2D Gaussian function. The second, fitting of a 2D Gaussian to the sub-aperture profile. The third being an iterative weighted first-moment calculation with a Gaussian weighting function. More details can be found in the original paper, but simulations showed the best results for iterative weighted first-moment calculation and the regular first-moment centroiding when paired with fourth-order gamma correction.

Besides low-order limitations of centroiding-based reconstruction, the aberrations that can be reconstructed are also limited in magnitude. The culprit for this limitation is the independence assumption between sub-apertures, i.e. each sub-aperture is assumed to contain only its own spot. The reason for this is the non-determinism that would arise if multiple spots share a single sub-aperture: it would be uncertain which spot corresponds to which sub-aperture and as such displacement calculation is non-deterministic. Methods have been proposed in literature for associating identified spots with their sub-aperture through, a process called spot indexing. One method is the iterative Zernike reconstruction of the phase based on a set of known and associated spots, followed by virtually projecting the centroids to their estimated positions according and associating unmatched spots within a given radius of the projected centroids to their corresponding sub-apertures[1]. A diagram of this method is provided in Figure 3-2. A second method is based on the observation that the rectangular or hexagonal spot lattice structure is locally preserved under aberrations, meaning neighbouring matched spots can be used to associate unmatched spots with the correct sub-aperture, or even approximate the location of missing spots[25].
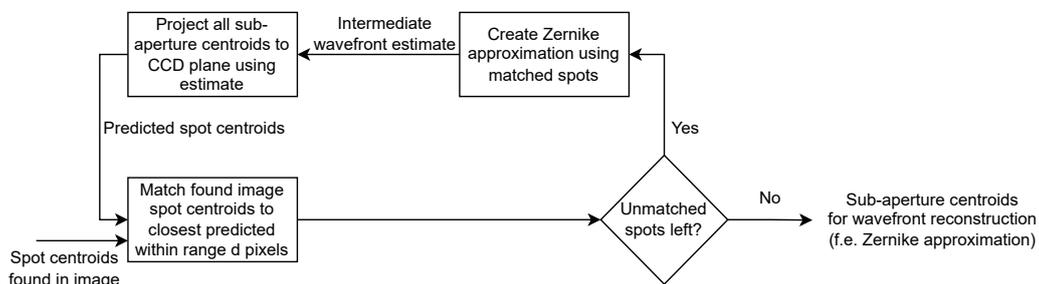
Figure 3-2: Diagram the of iterative Zernike reconstruction spot indexing[1]

Spot

Spot with thresholding

Spot with windowing

(a) Sub-aperture with a noisy spot

(b) Application of intensity thresholding (25%)

(c) Application of windowing around brightest pixel

Spot with gamma correction

(d) Application of gamma correction ($\gamma = 4$)
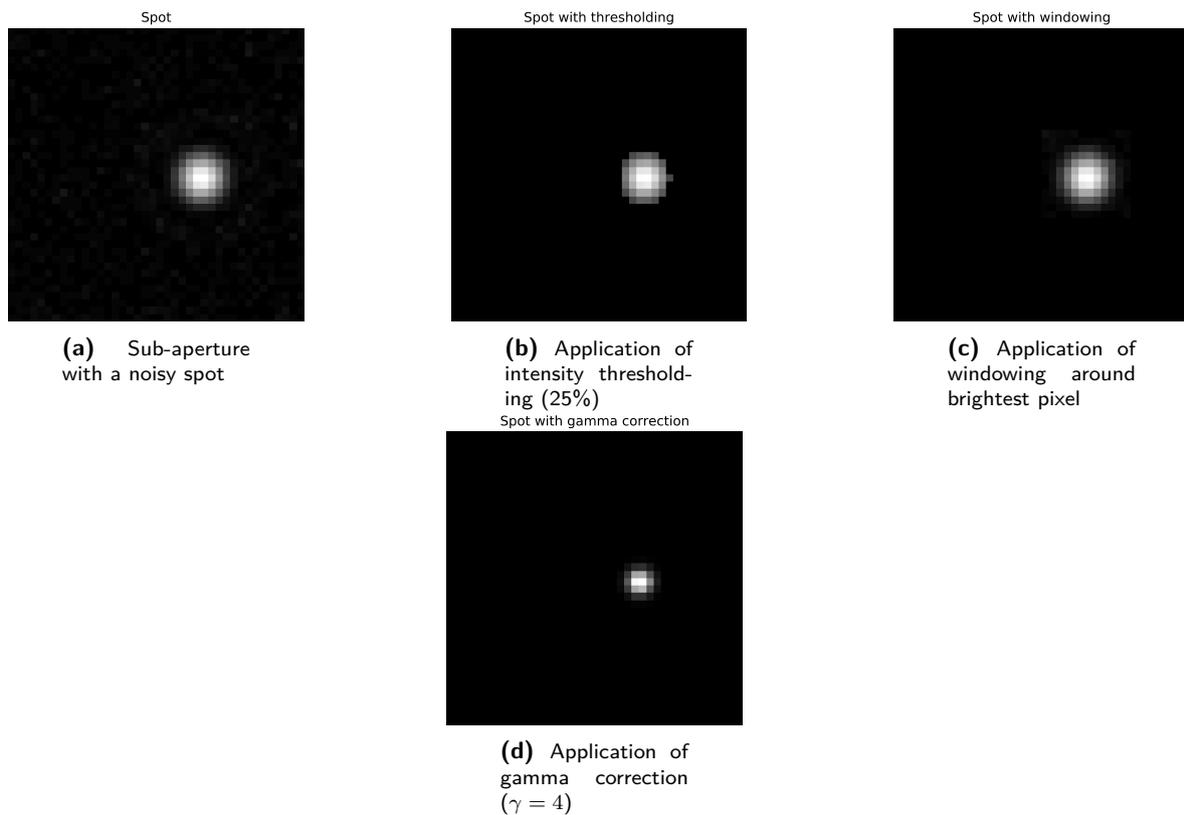
Figure 3-1: Illustration of pre-processing techniques on a noisy spot to limit influence on centroid calculation

It is worth noting that centroiding-based methods have also been used in literature to obtain curvature estimates. An example of this would be the work of Zou et al. in which three Shack-Hartmann sensors with two being offset in x and y respectively were used to obtain a curvature estimate through finite differences[26]. Additionally, centroiding-based reconstruction has been generalised to extended scenes by calculation of the displacement using optical flow[27, 28].

A powerful alternative to centroiding treated in literature for phase gradient extraction from Shack-Hartmann patterns is the method of Fourier demodulation[29–33]. This method, based in the theoretical model of a Shack-Hartmann sensor as a grating interferometer derived by Primot, treats the spot displacements as a modulation of the bi-periodic grid structure caused by the MLA geometry and will be discussed in more detail in Chapter 5. It is worth noting that centroid calculation is in fact related to Fourier demodulation, meaning the two methods are not dissimilar in terms of information extracted[29, 34]. However, the Fourier demodulation method supports higher-magnitude aberrations by design, meaning spots moving outside of their sub-apertures is no longer a limiting factor and making it more versatile. Additionally, non-orthogonal gradients can be extracted, providing information redundancy[29].

As all above methods only extract average phase gradient information through the displacement of spots, a significant amount of data in the form of inter-spot interference or spot shape in the point-spread function (PSF) is discarded. As such, the achievable reconstruction accuracy is limited, more specifically it is limited to $2N_{MLA}$ modes where N_{MLA} refers to the number of spots within the

aperture whose displacements have been calculated[4]. Any aberrations with phases whose change over a sub-aperture cannot be accurately approximated by a tip/tilt will also be aliased to lower-order modes. Finally, spot displacement techniques are also blind to aberrations with zero average phase over a sub-aperture such as perfect sinusoidal aberrations with a period equal to a multiple of the lenslet pitch. A higher reconstruction accuracy could be obtained by using an MLA with a higher lenslet count, but these are expensive to manufacture and will fundamentally limit the aberration magnitude due to spots having less room for movement.

Though lacking in reconstruction accuracy, computational speed for these methods is high if a feasible number of lenslets is used and expensive spot indexing algorithms can be largely avoided. Additionally, an approximation of the wavefront is readily obtained from a single intensity image. As will be shown in Chapter 7, this makes for an excellent pre-conditioning method for high-accuracy phase retrieval through PSF fitting.

3-2   Wavefront Reconstruction through Fourier Amplitude Sampling

Fourier amplitude sampling refers to the use of the full discretised intensity pattern, or PSF, generated by the aberration. As these methods are applied to wavefront sensorless systems, i.e. single-lens ideal systems, the generated intensity pattern is therefore the square of the Fourier amplitude.

Wavefront reconstruction methods for wavefront sensorless systems aim to solve the phase retrieval problem discussed in Section 2-1 either directly in its original form, or following some relaxation or re-casting of the problem. As such a large number of differing approaches have been worked out in literature.

The first of the discussed approaches is the method of alternating projections, aiming to solve the original problem without simplification or alteration. In this iterative method, two complex fields S and T are solved for knowing the mapping $\mathcal{F}\{S\} = T$ and having absolute amplitude information on both fields[5–7]. As such it is a projection between the sub-spaces on which S and T are defined by substituting the amplitude information at every step until the mapping is satisfied. The most popular implementation of alternating projections is the Gerchberg-Saxton algorithm[10], whose base principle of operation is shown in Figure 3-3. As the absolute amplitude of a complex field is not a one-to-one mapping even if the Fourier transform $\mathcal{F}$ is, convergence to the correct solution is not guaranteed unless additional intensity images using phase diversities are provided as has been done for the phase diversity phase retrieval algorithm[35]. This latter algorithm can be cast to the Gerchberg-Saxton framework through the use of three-dimensional Fourier transformation. Alternatives of the phase diversity phase retrieval algorithm have been proposed to be applied to Shack-Hartmann intensity patterns in the past, employing the full Shack-Hartmann transmission function model of Section 4-1[36]. However, more recently the Shack-Hartmann diversity model presented in Section 4-2 has also been proposed as a candidate for use in the original phase diversity phase retrieval algorithm[13]. The main issue with alternating projections algorithms is the lack of convergence guarantees or the need for multiple observations with diversities, as well as the fact that it is fundamentally a zonal reconstruction algorithm.
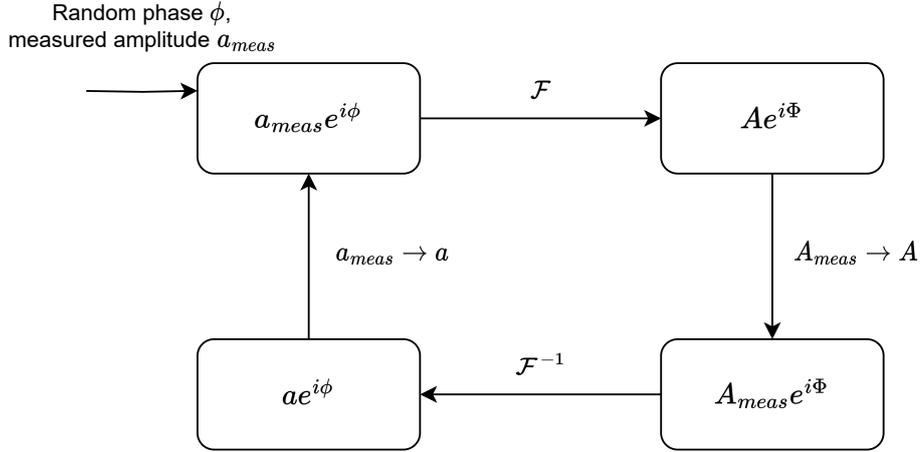
Figure 3-3: Diagram showing the operating principle of the Gerchberg-Saxton phase retrieval algorithm

A second method is known as dimensionality lifting, recasting the phase retrieval problem to a higher dimension to obtain convexity in the form of a semidefinite programming problem at the cost of squared computational complexity and storage requirements[5, 12, 37]. Implementations of this method express the intensity measurements in an intensity image as $\vec{y} = |\langle \vec{a}, \vec{x} \rangle|^2$[5]. This notation allows the notation given in Equation 3-2 to be used where multiple images $k = 1, \dots, M$ are assumed, which is linear in a higher dimension. Obtaining the wavefront would then amount to solving the semidefinite programming optimisation problem in Equation 3-3. Alternatives to rank minimisation with other heuristics have also been proposed in literature[37].

$$\begin{cases} \vec{y_k} = \vec{x}^* \vec{a_k} \vec{a_k}^* \vec{x} = \vec{x}^* A_k \vec{x} = Tr(A_k X) & k = 1, \dots, M \\ X \succeq 0 \\ rank(X) = 1 \end{cases} \qquad (3\text{-}2)$$

$$\begin{aligned} & \min_X \left(\text{rank(X)} \right) \approx \min_X Tr(X) \\ s.t. \quad & \vec{y_k} = Tr(A_k X) \qquad k = 1, \dots, M \\ & X \succeq 0 \end{aligned} \qquad (3\text{-}3)$$

As this method scales quadratically in computational complexity with increasing image size, practical implementation is challenging. However, alternative semidefinite problem definitions have been proposed whose problem dimension does not scale quadratically with image size through the use of convex relaxations[38]. Overall, however, this method remains computationally demanding and challenging to implement.

The last of the discussed methods is approximation of the intensity function. In intensity approximation, the non-linear and non-convex intensity function is locally approximated using Taylor approximation, usually around zero aberration as closed-loop conditions or small aberrations are assumed[11, 39]. One example of approximation are the methods by Silva et al., de Visser and Verhaegen, and Viegers et al., which is applied to a (Shack-)Hartmann intensity pattern and uses first-order Taylor approximation with a B-spline basis[39–41]. While these methods retrieve high-order aberrations, an independence of sub-apertures is assumed making its applicability to aberrations with a larger magnitude

limited even if Taylor approximation around a previous wavefront estimate is used. Similarly, the method by Marinica et al. also employs Taylor approximation, but applies it to the usual wavefront sensorless optical setup, showing reconstruction for both a first-order and second-order Taylor approximation with an arbitrary modal basis[11] and use of phase diversities to circumvent ambiguity in the solution. This approximation method is highly promising as it is numerically fast to solve, requiring a single least-squares solution for the first-order approximation, and is built to support phase diversities as well as initial estimates. As such, it will be further discussed in Chapter 6.

It is worth noting that alternative methods also exist based on physical reconstruction of the wavefront in actuator space. These methods do not retrieve the phase numerically from intensity patterns, but instead aim to directly correct for aberrations using image metrics. Two examples are correction based on minimisation of the second image moment, which is related to the average square phase gradient over the aperture[21, 42], and correction based on minimisation of the spot radius[43]. As these methods require physical iteration with multiple images and a phase corrector such as a deformable mirror, they do not qualify as numerical phase retrieval algorithms, nor as candidates for single-frame phase retrieval unless aberration dynamics are assumed, and are not studied further.

3-3   Research Question

As can be deduced from Section 3-1 and Section 3-2, reconstruction methods employing Shack-Hartmann wavefront sensors and wavefront sensorless reconstruction methods are complementary to one another.

While methods such as centroiding or Fourier demodulation are limited in reconstruction accuracy and as such are meant for low-order aberrations, they can provide a phase estimate from a singular image frame. In addition, they discard a large portion of information in the form of PSF characteristics such as spot shape or interference.

Conversely, wavefront sensorless methods can retrieve high-order aberrations as all available information in the form of interference is used. However, these methods suffer from ambiguity of the solution unless multiple images with phase diversities are used and/or are limited to small deviations around an approximation point.

The question then arises whether wavefront sensorless methods can be applied to Shack-Hartmann intensity patterns to reconstruct high-order aberrations while also including any phase information encoded within these patterns extracted by phase sampling methods like centroiding or Fourier demodulation. Additionally, could this be achieved using only a single image frame as would be the case for phase sampling methods. Using the hybrid Shack-Hartmann sensor model discussed in Chapter 4 consisting of a single-lens and phase diversity, known as the Shack-Hartmann diversity, this becomes a possibility which is explored in this thesis. Formally, the research question can therefore be phrased as follows:

How can we obtain higher-accuracy reconstructions of high-order aberrations from a single frame by considering the full Shack-Hartmann pattern as a point-spread function through Shack-Hartmann diversity?

Chapter 4

Modelling of a Shack-Hartmann Wavefront Sensor

This chapter will discuss two modelling techniques for a Shack-Hartmann wavefront sensor. First, the classic model viewing the Shack-Hartmann sensor as an array of lenses will be discussed in Section 4-1. Subsequently, an alternative view of the Shack-Hartmann sensor as a single-lens ideal system with a phase diversity will be discussed in Section 4-2 and a brief conclusion will be provided in Section 4-3.

4-1   Optical Transmission Function of a Micro-lens array

The influence of an optical component on the light wave passing through it is captured in the optical transfer function (OTF) of the optical component. The wavefield after passing is then given by multiplication of the incoming wave with the OTF[15].

For an ideal rectangular lens with focal length f and size $d \times d$ this transmission function is given by Equation 4-1[36]. Insertion of this formula in the Rayleigh-Sommerfeld propagation integral ultimately results in the emergence of a direct Fourier transform relation to obtain the wavefront in the focal plane of a single-lens ideal system[15].

$$t_0(x,y) = \Pi\left(\frac{x}{d}, \frac{y}{d}\right) e^{-i\frac{k}{2f}\left(x^2+y^2\right)} \tag{4-1}$$

As a micro-lens array (MLA) is an array of micro-lenses, Equation 4-1 must be cloned to all centroid positions. For a rectangular MLA, this means they must be placed at a distance p from one another in both x and y direction with p being the lenslet pitch. Additionally, the lenslet sub-aperture will have size $p \times p$. This results in Equation 4-2[36] where $\text{comb}_{p,p}$ is a 2D comb function consisting of Dirac delta functions spaced at intervals of p.

$$t(x,y) = t_0(x,y) * \text{comb}_{p,p}(x,y) = \Pi\left(\frac{x}{p}, \frac{y}{p}\right) e^{-i\frac{k}{2f}\left(x^2+y^2\right)} * \text{comb}_{p,p}(x,y) \tag{4-2}$$

While the above equation only shows the transmission function of a rectangular array, a hexagonal array is straightforwardly obtained as a hexagonal lens cloned on a hexagonal grid, which can be decomposed as a sum of two comb functions.

Applying Equation 4-2 to the incoming wavefront does not yet result in the wavefront at the detector plane of a Shack-Hartmann sensor. To obtain the wavefront at the detector, explicit propagation must be performed over a distance f with, for example, the angular spectrum propagation method[36].

Unlike for a single-lens system, propagation to the focal plane can no longer be achieved with a singular Fourier transform. This is undesirable for two reasons. Firstly, it increases the computational effort for simulation of Shack-Hartmann patterns. Secondly, as will be discussed in Chapter 6, most wavefront sensor-less methods solving the phase retrieval problem introduced in Section 2-1 rely on direct Fourier relations between the incoming wavefront and the captured wavefront and as such are not directly compatible with Shack-Hartmann sensors.

Clever modelling of an MLA, however, can be used to circumvent this apparent incompatibility. This will be further discussed in Section 4-2.

4-2   Shack-Hartmann Wavefront Sensor as a Phase Diversity

This section will discuss an alternative method of modelling a Shack-Hartmann sensor, namely as a phase diversity. This method, developed in parallel by Soloviev et al. and Hénault and Pannetier, proposes to split the Shack-Hartmann sensor micro-lens array into a single-lens ideal system and a piece-wise linear phase diversity dubbed the Shack-Hartmann diversity[13, 44]. As it is an extension of a single-lens system, this interpretation allows the use of direct Fourier transformation of the original wavefront with the additional phase term to obtain the wavefront at the Shack-Hartmann sensor detector plane, a numerically fast method of performing forward simulation of Shack-Hartmann patterns. Additionally, it allows the application of wavefront sensorless phase retrieval methods to Shack-Hartmann sensor intensity patterns. An example of such an application given by Soloviev et al. is the use of Shack-Hartmann diversity as a diversity in phase diversity phase retrieval, an alternating projections method[13, 35]. However, as will be shown in Section 6-2 and Chapter 7, it can be used to apply intensity approximation phase retrieval methods to Shack-Hartmann patterns as well.

As both Soloviev et al. and Hénault and Pannetier developed a Shack-Hartmann diversity separately, both will be briefly discussed. It is worth noting that the diversity obtained by Soloviev et al. will be used in this thesis, the rationale for which will become clear later in this section.

Observing the transmission functions of a lens and a micro-lens array, shown before in Equation 4-1 and Equation 4-2 respectively, the transmission function phase fields are given by Equation 4-3 and Equation 4-4. In these equations the full pupil and sub-aperture pupil functions have been replaced by domains S and S_i respectively, where S_i and centroid $\vec{x}_{c,i}$ refer to the sub-aperture and centroid position of the i'th lenslet. Additionally, inner-product notation is used to simplify notation.

$$\phi_L(x,y) = -\frac{k}{2f}\left(x^2 + y^2\right) = -\frac{\pi}{\lambda f}\left\langle\vec{x},\vec{x}\right\rangle, \vec{x} \in S \tag{4-3}$$

$$\phi_{MLA,i}(x,y) = -\frac{\pi}{\lambda f}\left\langle\vec{x}-\vec{x}_{c,i}, \vec{x}-\vec{x}_{c,i}\right\rangle, \vec{x} \in S_i \in S \tag{4-4}$$

The difference between these phase fields can then be used to obtain the piece-wise linear Shack-Hartmann diversity in Equation 4-5 whose formulation depends on the MLA focal length f, the used

wavelength λ, and the MLA geometry. As shown in Equation 4-6, adding this diversity term to the phase of the incoming wavefront $W(x, y)$ prior to ideal lens propagation through Fourier transformation will result in the wavefront at the detector plane.

$$\begin{cases} \phi_L(x,y) & = -\frac{\pi}{\lambda f} \langle \vec{x}, \vec{x} \rangle, \vec{x} \in S \\ \phi_{MLA,i}(x,y) & = -\frac{\pi}{\lambda f} \langle \vec{x} - \vec{x}_{c,i}, \vec{x} - \vec{x}_{c,i} \rangle, \vec{x} \in S_i \in S \\ \phi_{SH}(x,y) & = \phi_{MLA}(x,y) - \phi_L(x,y) \end{cases}$$

$$\Longleftrightarrow \phi_{SH} = \frac{\pi}{\lambda f} \langle \vec{x}_{c,i}, 2\vec{x} - \vec{x}_{c,i} \rangle, \qquad \vec{x} \in S_i \tag{4-5}$$

$$W_{detector}(x,y) = \mathcal{F}\left\{ W(x,y)e^{i\phi_{SH}(x,y)} \right\} \tag{4-6}$$

It is worth noting that the notation of Equation 4-5 applies for both rectangular as well as hexagonal Shack-Hartmann sensors micro-lens arrays.

This Shack-Hartmann diversity concept was implemented in a Python framework for this thesis for both rectangular and hexagonal arrays and used for forward Shack-Hartmann sensor simulation as well as phase retrieval. The framework was designed for use as an optical component within the HCIpy[45] Python optical propagation framework and as such is straightforward to re-use for later projects. An illustration of the obtained diversities for both geometries is given in Figure 4-1.

(a) Wrapped Shack-Hartmann diversity for a rectangular micro-lens array in radians

(b) Wrapped Shack-Hartmann diversity for a hexagonal micro-lens array in radians
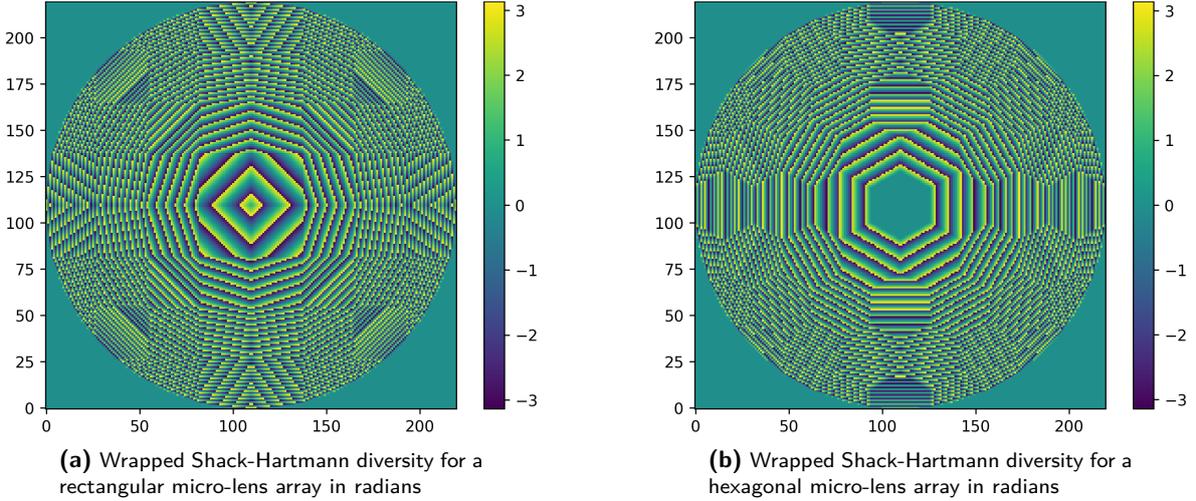
Figure 4-1: Illustratory wrapped Shack-Hartmann diversities for rectangular and hexagonal micro-lens arrays

To illustrate forward simulation Figure 4-2 shows the obtained intensity patterns for a rectangular and hexagonal Shack-Hartmann sensor under zero-aberration conditions.

4-3 Conclusion

The main purpose of this chapter was to introduce an alternative method of modelling a Shack-Hartmann sensor as a single-lens ideal system with an added piece-wise linear phase diversity.

Shack-Hartmann intensity pattern with zero aberration

Shack-Hartmann intensity pattern with zero aberration

(a) Obtained intensity pattern for rectangular Shack-Hartmann configuration

(b) Obtained intensity pattern for hexagonal Shack-Hartmann configuration
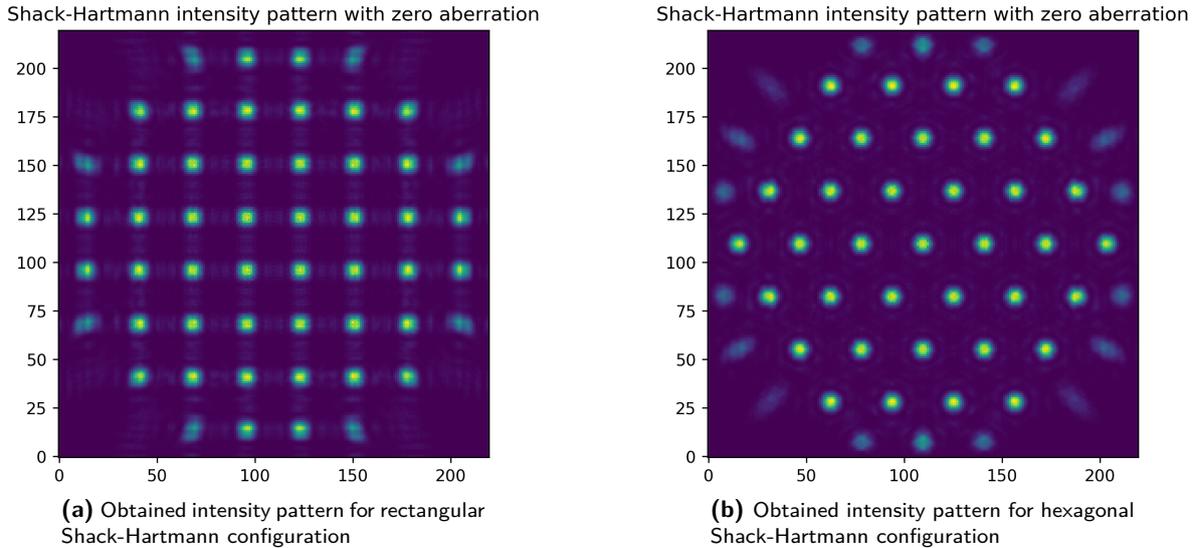
Figure 4-2: Intensity pattern obtained for zero-aberration on Shack-Hartmann sensors with focal length 10mm, a circular pupil, and aperture diameter 1.2mm

First, the classic model of a Shack-Hartmann sensor as a singular optical component with a micro-lens array was discussed in Section 4-1. While this model is the best mathematical expression of the real physical system, the description in Equation 4-2 does not allow simplification of the optical propagation equations. As such, explicit propagation is required using, for example, the angular spectrum method which incurs a higher computational cost instead of the computationally cheaper direct Fourier relation obtained with single-lens system as is the case for a wavefront sensorless system. More importantly, however, the lack of a direct Fourier relation makes wavefront sensorless phase retrieval methods incompatible with Shack-Hartmann intensity patterns in this model.

The theory in Section 4-2 described an alternative view of a Shack-Hartmann sensor, namely as a single-lens ideal system with an additional piece-wise linear phase diversity dubbed the Shack-Hartmann diversity. The model is applicable to both rectangular and hexagonal Shack-Hartmann sensor MLA geometries and as such is versatile. As a single-lens ideal system does employ a direct Fourier transform relation for propagation through the system, it has two interesting consequences. Firstly, the use of direct Fourier transformation results in a computationally cheap method to perform forward simulation wavefronts through Shack-Hartmann sensors. Secondly, it bridges the gap between wavefront sensorless phase retrieval methods and Shack-Hartmann intensity patterns. This latter consequence will be illustrated in Section 6-2 where the application of Taylor approximation to the intensity function is shown to be compatible with Shack-Hartmann patterns, and Chapter 7 where a method for high-accuracy phase retrieval from a single Shack-Hartmann image frame is shown.

Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

This chapter will discuss phase gradient extraction using an augmented Fourier demodulation method and subsequent modal wavefront reconstruction. Firstly, a brief overview of the shearing interferometer model of a Shack-Hartmann wavefront sensor will be provided in Section 5-1. This is followed by the theoretical explanation, details of numerical implementation, and augmentation of Fourier demodulation to rectangular spot grids in Section 5-2. Section 5-3 will then discuss the extension of this methodology to hexagonal spot grids. Furthermore, the modal phase reconstruction methodology and basis considerations will be discussed in Section 5-4. Lastly, reconstruction results and a brief conclusion will be provided in Section 5-5.

5-1 Shack-Hartmann Wavefront Sensor as a Shearing Interferometer

The principle behind the Fourier demodulation technique relies on a specific model of the Shack-Hartmann derived by Primot, that of the Shack-Hartmann sensor as a shearing interferometer. This model allows the displacements of the spots generated by each lenslet caused by the average phase gradient over each sub-aperture to be interpreted as a modulation of the underlying lattice structure of the spots in a regular rectangular Shack-Hartmann configuration [34]. It is worth noting that the procedures from paper of Primot will be used throughout this section, but the original work should be consulted for a more in-depth analysis[34].

22

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors

The derivation of this model starts from the transmission function, or grating equation, of a micro-lens array (MLA) in Equation 5-1 where $f_{\mu L}$ and $p_{\mu L}$ are the MLA focal length and lenslet pitch respectively. The rect function $\Pi\left(\frac{x}{p_{\mu L}}, \frac{y}{p_{\mu L}}\right)$ defines the rectangular sub-aperture and has the value 1 inside a $p_{\mu L} \times p_{\mu L}$ window around the origin and 0 outside of it. The $*$ operator denotes convolution, and the $\mathrm{comb}_{p_{\mu L}, p_{\mu L}}$ function is a comb function consisting of a 2D grid of Dirac delta functions spaced at a distance $p_{\mu L}$ apart from one another in both x and y. As such, due to the sifting property of Dirac delta functions this convolution results in an ideal lens transmission function with a $p_{\mu L} \times p_{\mu L}$ sub-aperture being cloned at $p_{\mu L}$ intervals, keeping true to its definition as a micro-lens array.

$$G(x,y) = \left[e^{i\pi \frac{x^2+y^2}{\lambda f_{\mu L}}} \Pi\left(\frac{x}{p_{\mu L}}, \frac{y}{p_{\mu L}}\right)\right] * \mathrm{comb}_{p_{\mu L}, p_{\mu L}}(x,y) \tag{5-1}$$

An assumption is made that the MLA is of infinite extension in both x and y direction, making the grating equation bi-periodic. As such, the grating equation can be represented by an infinite 2D Fourier series as illustrated in Equation 5-2 where the coefficients $c_{n,m}$ need to be found.

$$G(x,y) = \frac{1}{p_{\mu L}^2} \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} c_{n,m} e^{\frac{i2\pi}{p_{\mu L}}(nx+my)} \tag{5-2}$$

These coefficients $c_{n,m}$ are given by Equation 5-3 where $\Psi_{\mu L}$ is defined by Equation 5-4.

$$
\begin{aligned}
c_{n,m} &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{i\pi \frac{x^2+y^2}{\lambda f_{\mu L}}} \Pi\left(\frac{x}{p_{\mu L}}, \frac{y}{p_{\mu L}}\right) e^{\frac{-i2\pi}{p_{\mu L}}(nx+my)} dx dy \\
&= \mathcal{F}\left\{e^{i\pi \frac{x^2+y^2}{\lambda f_{\mu L}}} \Pi\left(\frac{x}{p_{\mu L}}, \frac{y}{p_{\mu L}}\right)\right\} \left(\frac{n}{p_{\mu L}}, \frac{m}{p_{\mu L}}\right) \\
&= \Psi_{\mu L}\left(\frac{n}{p}, \frac{m}{p}\right)
\end{aligned}
\tag{5-3}
$$

$$\Psi_{\mu L} = \left(\mathrm{sinc}(pu)\,\mathrm{sinc}(pv)\right) * e^{i\pi \lambda f_{\mu L}(u^2+v^2)} \tag{5-4}$$

The shearing interferometer model owes its name to the behaviour captured in these equations, as Primot observed that upon illumination with a monochromatic near-plane wave this results in copies of the wavefront being sent along different wave-vectors $k_{n,m}$ given by Equation 5-5 after leaving the MLA and interfering at the Shack-Hartmann detector plane. An observation is also made by Primot that Equation 5-4 can also be seen as the focal plane amplitude of the point $(n\lambda, m\lambda)$ for a F/1 square lens with aperture size $p_{\mu L}$ imaging a point source at $f_{\mu L}$. This observation will be used later to simplify the expression in Equation 5-3.

$$
\begin{cases}
k_{n,m} &= \frac{2\pi}{\lambda} \begin{bmatrix} n\alpha \\ m\alpha \\ \sqrt{1-(n^2+m^2)\alpha^2} \end{bmatrix} \\
\alpha &= \frac{\lambda}{p}
\end{cases}
\tag{5-5}
$$

Assuming a sensor with a large number of sub-apertures, a simplified wavefront amplitude at the detector plane is given by Equation 5-6 where H is given by Equation 5-7 and $H_{NG} \geq H$ is dependent on H with more details available in the original paper[34].

$$A_{detector}(x,y) = \sum_{n=-H_{NG}}^{H_{NG}} \sum_{m=-H_{NG}}^{H_{NG}} c_{n,m} e^{\frac{-i\pi}{2H}(n^2+m^2)} e^{i\phi(x+\frac{np_{\mu L}}{2H},y+\frac{mp_{\mu L}}{2H})} e^{\frac{i2\pi}{p_{\mu L}}(nx+my)} \tag{5-6}$$

$$H = \left\lfloor \frac{p_{\mu L}^2}{2\lambda f_{\mu L}} \right\rfloor \tag{5-7}$$

Depending on the relative rates of variation of the convolved functions in Equation 5-4, the paper by Primot derives the generated intensity image in one of two domains: independence domain in which lenslets are assumed independent of one another, and cross-talk domain.

In the independence domain a small F number system is assumed and the earlier mentioned analogy between Equation 5-4 and an F/1 lens system can be used to simplify Equation 5-3 through geometrical optics. In this simplification, $c_{n,m} = e^{i\pi\frac{\lambda f_{\mu L}}{p_{\mu L}^2}}$ for all $-H < n, m < H$ and 0 otherwise, simplifying the intensity equation obtained from the amplitude in Equation 5-6 to Equation 5-9. In this last simplification, an additional assumption is made that the wavefront phase varies little at the scale of a sub-aperture, meaning the approximation in Equation 5-8 holds.

$$\phi(x+\frac{np_{\mu L}}{2H}, y+\frac{mp_{\mu L}}{2H}) = \phi(x,y) + \frac{np}{2H}\frac{\partial\phi}{\partial x}(x,y) + \frac{mp}{2H}\frac{\partial\phi}{\partial y}(x,y) \tag{5-8}$$

$$\begin{cases} I_G(x,y) & = \sum_{n,m,n',m'=-H}^{H} e^{i\left(\phi(x+\frac{np_{\mu L}}{2H},y+\frac{mp_{\mu L}}{2H})-\phi(x+\frac{n'p_{\mu L}}{2H},y+\frac{m'p_{\mu L}}{2H})\right)} e^{\frac{i2\pi}{p_{\mu L}}((n-n')x+(m-m')y)} \\ & = \sum_{k=-2H}^{2H}\sum_{l=-2H}^{2H} Harm_{k,l}(x,y) e^{i\frac{2\pi}{p_{\mu L}}(kx+ly)} \qquad , k=n-n', l=m-m' \\ Harm_{k,l}(x,y) & = (2H+1-|k|)(2H+1-|l|)e^{i\frac{p_{\mu L}}{2H}\left(k\frac{\partial\phi}{\partial x}(x,y)+l\frac{\partial\phi}{\partial y}(x,y)\right)} \end{cases} \tag{5-9}$$

The intensity pattern in Equation 5-9 is built up of a sum of harmonics filtered each depending on the phase gradient along a particular direction. Upon close inspection, however, taking the Fourier transform of the intensity pattern will separate the spectra of the harmonics at intervals of $\frac{1}{p_{\mu L}}$ apart in the frequency domain, allowing straightforward extraction if the assumption is made that the spectra have minimal overlap. This is the definition of the intensity function upon which gradient extraction through Fourier demodulation in Section 5-2 will be based, though Primot warns of the limited region of validity for which this intensity function truly resembles the actual one. However, as will become apparent later this does not pose a significant issue during simulation.

If a high F number system is assumed instead, the geometrical simplification of $c_{n,m}$ is no longer possible and, after a number of intermediate steps and assumptions which will not be listed here as well as the assumed small phase variation over a sub-aperture, a much more complex definition of the intensity function $I_{NG}(x,y)$ is obtained. This more complex definition is provided in Equation 5-10. Extraction of gradients is no longer straightforward as harmonics contain a weighted sum of gradient components. However, the observation is made by Primot that the weighting is small. Therefore, it is assumed that systems in the region of applicability of Equation 5-10 may be processed using gradient extraction techniques which will be discussed for Equation 5-9.

$$
\begin{cases}
I_{NG}(x,y) & = \sum_{k=-H_{NG}}^{H_{NG}} \sum_{l=-H_{NG}}^{H_{NG}} Harm_{k,l}^{NG}(x,y) e^{i\frac{2\pi}{p_{\mu L}}(kx+ly)} \\
Harm_{k,l}^{NG}(x,y) & = \Gamma_{\mu L}\left(\frac{k}{p_{\mu L}}, \frac{l}{p_{\mu L}}\right) e^{i\frac{p_{\mu L}}{2H}\left(k\frac{\partial\phi}{\partial x}(x,y)+l\frac{\partial\phi}{\partial y}(x,y)\right)} \\
\Gamma_{\mu L}(u,v) & = \text{autocorrelation}\left(\psi_{\mu L}(u,v) e^{\frac{-i\pi}{2H}(u^2+v^2)}\right)
\end{cases}
\tag{5-10}
$$

One final important observation is made by Primot. As the Shack-Hartmann intensity pattern is undersampled, aliasing occurs between the highest and lowest harmonics in the intensity spectrum. However, the harmonics rapidly, making the observable effect of the aliasing minimal.

With the theoretical foundation provided in this section, the extraction of phase gradient information from harmonics can be discussed as will be done in Section 5-2 for rectangular arrays. While the work of Primot only considered rectangular Shack-Hartmann arrays[34], Section 5-3 will illustrate that a similar procedure can be applied to hexagonal arrays as well.

5-2 Phase Gradient Extraction on Rectangular Spot Grids using Fourier Demodulation on a Trimmed Spectral Domain

For the extraction of gradients, the work of Rouzé et al. can be used[29]. This work is based largely on the theoretical model described above in Section 5-1, but presents the theory in a much more condensed manner.

A change of parameters and notation is performed in this section to stay close with the notation in the work of Rouzé et al.. The parameter H defined in Equation 5-7 will be included in an alternative coefficient K given by Equation 5-11. Furthermore, $Harm_{k,l}$ will be shortened to the briefer notation $H_{k,l}$ with an identical meaning.

$$
K = 2H + 1 = 2\left\lfloor \frac{p_{\mu L}^2}{2\lambda f_{\mu L}} \right\rfloor + 1
\tag{5-11}
$$

As previously defined in Equation 5-9, the (k, l) harmonic is related to the gradient along the (kx, ly) direction by Equation 5-12 where the notation of Rouzé et al. is used. The parameter ∂p in this equation is given by Equation 5-13.

$$
H_{k,l}(x,y) = (K-\mid k \mid)(K-\mid l \mid)e^{-i\lambda\sqrt{k^2+l^2}\partial p \frac{\partial\phi}{\partial x_{k,l}}}
\tag{5-12}
$$

$$
\partial p = \frac{f_{\mu L}}{p_{\mu L}}
\tag{5-13}
$$

As mentioned previously in Section 5-1, the harmonic spectra are spaced at intervals of $\frac{1}{p_{\mu L}}$ apart from one another in the spectrum of the Shack-Hartmann intensity pattern. This means that obtaining the harmonic is achieved by inverse Fourier transformation of the extracted harmonic spectrum, obtained by defining a $\frac{1}{p_{\mu L}} \times \frac{1}{p_{\mu L}}$ window around the harmonic spectrum and moving it to the centre of the frequency space. Mathematically, this can be written as given in Equation 5-14.

$$H_{k,l}(x,y) = \mathcal{F}^{-1}\left\{\text{shift}\left(\text{rect}\left(\frac{f_x - k/p_{\mu L}}{1/p_{\mu L}}, \frac{f_y - l/p_{\mu L}}{1/p_{\mu L}}\right)\mathcal{F}\{I_{SH}(x,y)\}\right)\right\} \qquad (5\text{-}14)$$

Rearrangement of Equation 5-12 leads to the formula used for extraction of the phase gradient along the $x_{k,l}$ direction in Equation 5-15. If the gradient is assumed normalised and small, the angle calculation can be substituted by taking the imaginary component of the harmonic by virtue of the small angle approximation, an approach used in other works employing Fourier demodulation [32]. However, this simplification will not be employed in the work of this thesis.

$$
\begin{aligned}
\frac{\partial \phi}{\partial x_{k,l}} &= \frac{1}{-i\lambda\sqrt{k^2 + l^2}\partial p}\log\left(\frac{H_{k,l}(x,y)}{(K - |k|)(K - |l|)}\right) \\
&= -\frac{1}{\lambda\sqrt{k^2 + l^2}\partial p}\text{angle}\left(\frac{H_{k,l}(x,y)}{(K - |k|)(K - |l|)}\right) \\
&= -\frac{1}{\lambda}\text{angle}\left(\left(\frac{H_{k,l}(x,y)}{(K - |k|)(K - |l|)}\right)^{\frac{1}{\sqrt{k^2 + l^2}\partial p}}\right)
\end{aligned}
\qquad (5\text{-}15)
$$

The methodology by Rouzé et al. presented above was derived for continuous Fourier transforms, and as such does not fully consider the specifics and challenges that come with implementing this for discrete Fourier transforms as one would use in practice on computing hardware. Therefore, the continuation of this subsection will discuss the numerical implementation of the above gradient extraction method as was performed for the thesis project.

The first observation that must be made is the difference in interpretation between a discrete and continuous Fourier transform. While continuous Fourier transformation assumes the period of repetition to be infinite, discrete Fourier transformation instead assumes the function under consideration to repeat with a period equal to the studied domain, i.e. if the function to be Fourier transformed consists of 20×20 data-points or pixels, the assumed period of repetition is 20 pixels and a spectrum is provided with 20×20 discrete frequency bins. An illustration is provided in Figure 5-1 for clarity.

This means that patterns with asymmetry at the edges give rise to high-frequency components in the resulting frequency spectrum to account for the abrupt change when the function is tiled. To remedy this, functions are either apodized near the edges prior to discrete Fourier transformation, thereby destroying data, or, more commonly, zero-padded. The amount of zero-padding used is not fixed, for example [33] suggests zero-padding such that the padded domain is twice that of the original function. The work considered in this MSc thesis will instead assume padding to three times the original dimensions, resulting in a padded intensity image I_{padded} of size $3N \times 3N$ obtained from Equation 5-16. This choice of padding is mostly for ease of implementation as it employs a Kronecker product and can easily be reversed. Overall, a padding factor c_{pad} will be defined such that the padded size $N_{zp} = c_{pad}N$.

$$I_{padded} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \otimes I \qquad (5\text{-}16)$$

26

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors
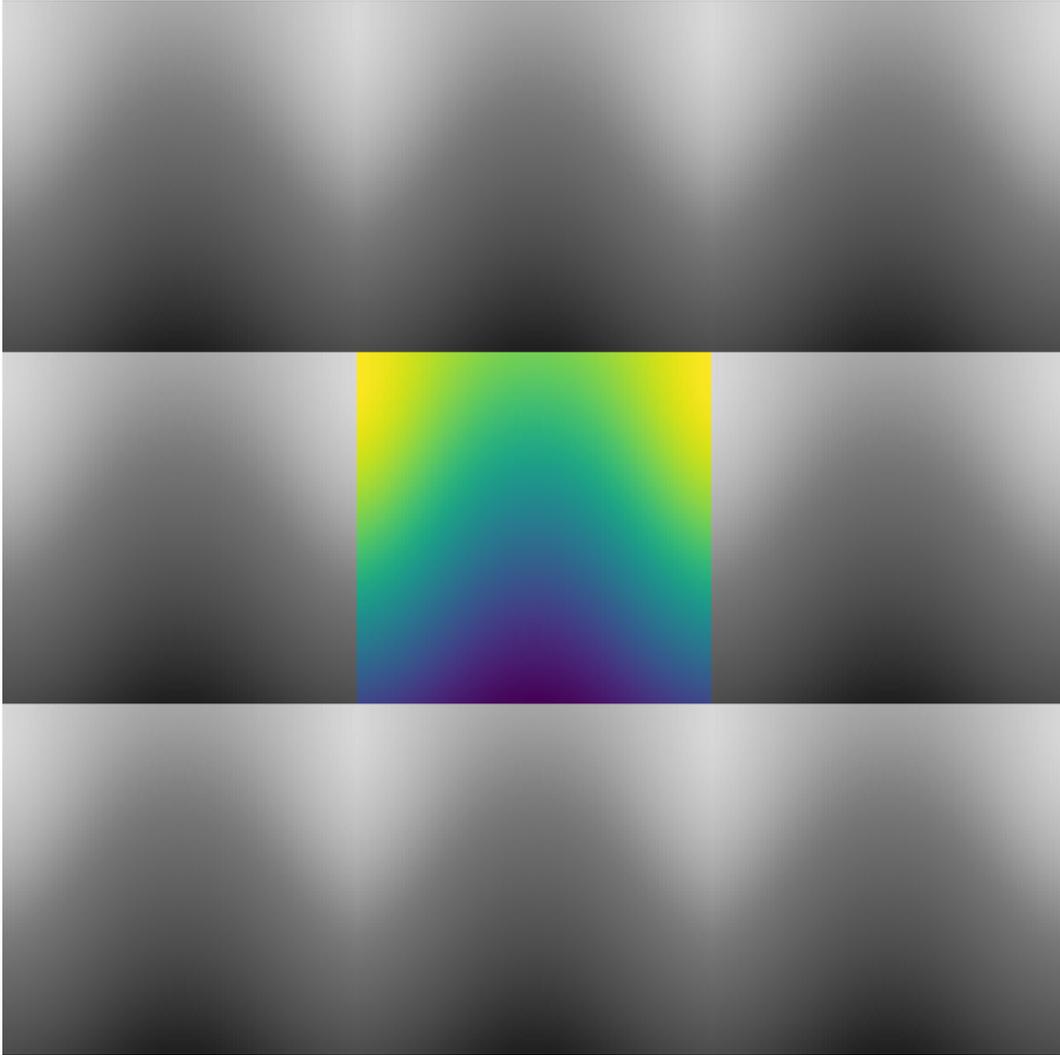
Figure 5-1: Illustration of the perceived pattern of repetition by discrete Fourier transformation

After the padded discrete Fourier spectrum DFT $\{I_{padded}\}$ is obtained, extraction of the relevant harmonic must be considered. The continuous-time equations discussed above refer to an extraction window of size $\frac{1}{p_{\mu L}} \times \frac{1}{p_{\mu L}}$ in 2D frequency space centred at the (k,l)'th multiple of pitch frequency $\frac{1}{p_{\mu L}}$, which must be converted to pixels to adhere to the discrete frequencies used in the discrete 2D Fourier transform. For a discrete Fourier spectrum of $N_{zp} \times N_{zp}$ pixels, where $N_{zp} = c_{pad}N = 3N$ is the zero-padded size, the discrete approximation of $\frac{1}{p_{\mu L}}$ in pixels $w_{f,pix}$ can be obtained using Equation 5-17 in which d is the physical pixel size in meters. The extraction window for the Fourier transform of the (k,l)'th harmonic can then be centred at an offset of $(kw_{f,pix}, lw_{f,pix})$ pixels from the DC component of the full $N_{zp} \times N_{zp}$ padded intensity image frequency spectrum. It is worth noting that, for rectangular arrays, $w_{f,pix}$ will be the same as the number of lenslets in each row of the Shack-Hartmann array multiplied by the padding factor, meaning an 8×8 array will have $w_{f,pix} = c_{pad}8 = 24$ for a padding factor $c_{pad} = 3$.

$$w_{f,pix} = \left\lfloor \frac{\frac{1}{p_{\mu L}}}{\frac{1}{N_{zp}d}} \right\rfloor = \left\lfloor \frac{N_{zp}d}{p_{\mu L}} \right\rfloor \tag{5-17}$$

The size of the window to be extracted would then also be $w_{f,pix} \times w_{f,pix}$ pixels. However, there is an important implementation detail that must be taken into account: potential energy leaks in the extracted spectrum depending on the even- or oddness of $w_{f,pix}$.

The centre pixel of the extraction window represents the DC component of the extracted harmonic's spectrum, meaning the remaining $w_{f,pix} - 1$ pixels in each direction correspond to the symmetric non-DC frequency bins. To illustrate the energy leak, one can consider the zero-aberration case. If the harmonic relation of Equation 5-12 is assumed to hold, then a plane wavefront should result in zero phase gradient and as such a fully real harmonic, requiring a fully symmetric spectrum and as such each non-zero negative frequency bin must have a corresponding and equal non-zero positive frequency bin. For an odd $w_{f,pix}$ this holds as one pixel corresponds to the DC component and the remaining number of pixels is even in each direction, corresponding to the symmetric positive and negative frequency bins. For an even $w_{f,pix}$, however, this is not necessarily true as one frequency bin in each direction will not have a symmetric counter-part. Per the numerical algorithm used for the discrete Fourier transform, namely the fast Fourier transform (FFT) algorithm, these will be the negative frequencies. If these uncompensated frequency bins are non-zero, the spectrum is no longer symmetric and the resulting harmonic will not be real, violating the relation in Equation 5-12. This energy leak influences non-zero aberrations as well, though these will by definition not have real harmonics when their gradients are non-zero and instead will simply be corrupted by the energy leak.

This can be mitigated in one of two ways. One possibility is to set the values of the uncompensated frequency bins to zero, i.e. the first row and column of the extracted discrete 2D spectrum. Another option, and the option chosen for this MSc thesis, is to increase the extraction window size by one if $w_{f,pix}$ is even, always resulting in an odd-sized extraction window. The rationale behind this approach is the preservation of data as the uncompensated frequency bins do still contain harmonic information and thereby gradient information. As mentioned in Section 5-1 the spectra of the harmonics do suffer from aliasing, meaning the frequency bin will not be perfectly compensated if an extra row and column is taken. However, the corruption caused by this overlap from aliasing is assumed to be small enough not to have a significantly more impactful effect than it has on the rest of the extracted spectrum.

The use of discrete spectra rather than continuous spectra brings with it one more issue. As mentioned before the harmonics are spaced at $\frac{1}{p_{\mu L}}$ from one another in the intensity image frequency spectrum, which does not necessarily correspond to the centre of a discrete Fourier frequency bin. In fact, Equation 5-17 shows that $w_{f,pix}$ is rounded down to the centre of the nearest frequency bin to $\frac{1}{p_{\mu L}}$ unless $\frac{N_{zp}d}{p_{\mu L}}$ is already an integer. A discrepancy in this assumed centre will lead to the emergence of a non-zero gradient even when no aberration is present and an ideal system with no static aberration is assumed, an erroneous virtual tip/tilt not present in the actual aberration. Literature deals with this in different ways, for example the work of Ribak et al. limits the application of the method to Shack-Hartmann setups for which the integer assumption holds thereby circumventing the problem, meanwhile Rouzé et al. chooses not to mention this problem explicitly [29, 33]. The method that is used in the work presented in this MSc thesis is to instead include a reference pattern, i.e. the Shack-Hartmann pattern for zero outside aberration, undergoing the same harmonic extraction procedure as the aberrated pattern. The added benefit of employing a reference pattern is the immediate elimination of static aberrations within the system, i.e. those caused by mirror misalignment, mirror sag, and lens defoci, in addition to the elimination of the virtual tip/tilt caused by discretisation. In practice this means the phase seen by the discrete Fourier demodulation method is $\phi_{obs} = \phi + \phi_{static} + \phi_{virtual} =$

28

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

$\phi + \phi_{ref}$ where ϕ_{ref} would be captured by the reference pattern. To avoid wrapping issues with the angle() operator, the static and virtual aberrations should be eliminated before the angle is calculated, thereby leading to Equation 5-18.

$$
\begin{aligned}
H_{obs_{k,l}}(x_i, y_i) &= (K- \mid k \mid)(K- \mid l \mid)e^{-i\sqrt{k^2+l^2}\partial p\lambda \frac{\partial\phi_{obs}}{\partial x_{k,l}}} \\
&= (K- \mid k \mid)(K- \mid l \mid)e^{-i\sqrt{k^2+l^2}\partial p\lambda\left(\frac{\partial\phi}{\partial x_{k,l}} + \frac{\partial\phi_{static}}{\partial x_{k,l}} + \frac{\partial\phi_{virtual}}{\partial x_{k,l}}\right)} \\
&= H_{k,l}(x_i, y_i)H_{ref_{k,l}}(x_i, y_i) \\
\Longleftrightarrow \quad H_{k,l}(x_i, y_i) &= \frac{H_{obs_{k,l}}(x_i, y_i)}{H_{ref_{k,l}}(x_i, y_i)} \\
\Longleftrightarrow \quad \frac{\partial\phi}{\partial x_{k,l}}(x_i, y_i) &= -\frac{1}{\lambda}\text{angle}\left(\left(\frac{H_{obs_{k,l}}(x_i, y_i)}{H_{ref_{k,l}}(x_i, y_i)}\right)^{\frac{1}{\sqrt{k^2+l^2}\partial p}}\right)
\end{aligned}
\tag{5-18}
$$

One last point of discussion remains, namely the zero-padding of the discrete extracted harmonic spectrum prior to inverse Fourier transformation. In the continuous-time relation shown in Equation 5-14 it is shown that the extracted spectrum of the harmonic is to be shifted to the centre of the frequency spectrum and inverse Fourier transformed. In discrete Fourier transformation this would mean placing the extracted discrete spectrum in the centre of an $N_{zp} \times N_{zp}$ empty spectrum. Doing this and then employing Equation 5-18 results in an $N_{zp} \times N_{zp}$ padded gradient field, for which the padding can be easily reversed to obtain the $N \times N$ gradient field which can be multiplied with the aperture to remove artefacts at the edges. An example of gradients extracted from the $(1,0)$ and $(1,1)$ harmonics is given in Figure 5-2 for a spherical aberration whose gradient field is known to be a coma oriented along the gradient direction.

(a) Extracted gradient along the y direction

(b) Extracted gradient along the $x + y$ direction
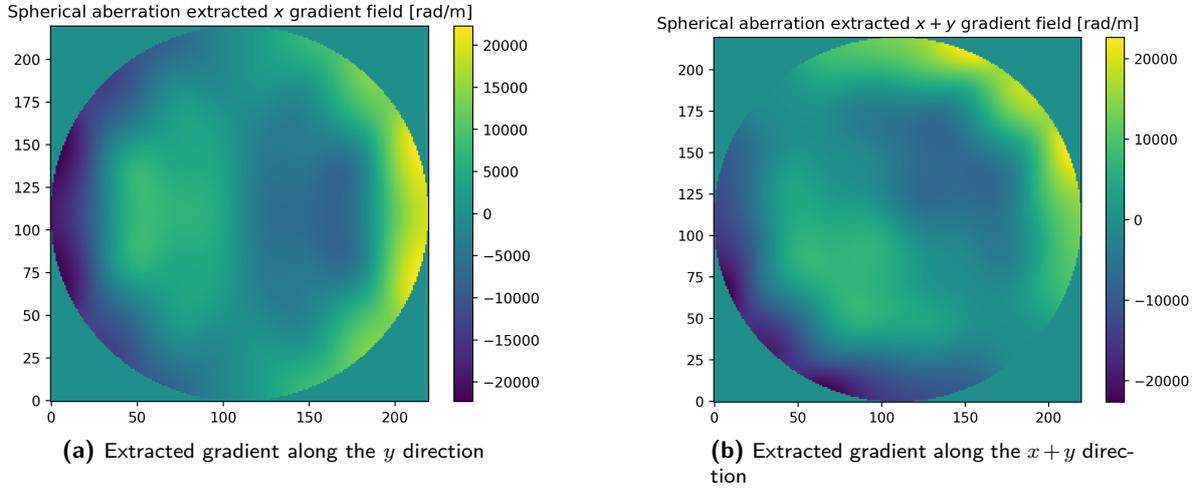
Figure 5-2: Example of extracted gradients using Fourier demodulation of a spherical aberration, analytically known to be comas

It is important to realise, however, that zero-padding the spectrum does not create additional data, but merely expands the reconstructed function in x and y direction through interpolation using sinc

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors

29

functions. Reconstruction methods performing zonal reconstruction such as Fourier or numerical integration require a gradient field of equal size to the original phase to be reconstructed, necessitating the padding procedure described above. However, this is not necessarily true for other reconstruction techniques and a reduced padding may be beneficial for these. For example in the modal reconstruction technique discussed in Section 5-4, the computational cost reduces drastically if a lower padding is used. A similar observation of lack of additional data is used by Glazer et al. to speed up adaptive optical control by building a controller calibration matrix directly from the extracted harmonic, but this work does not perform explicit reconstruction of the phase [32].

The padded harmonic spectrum is defined to be the original extracted harmonic spectrum zero-padded to a size $N_{zph} \times N_{zph}$ with $N_{zph} \geq w_{f,pix} + w_{f,pix} \mod 2$ where the modulo results from the spectrum energy leak mitigation strategy discussed above.

The value of N_{zph} is subject to a few conditions:

1. $N_{zph} \geq w_{f,pix} + w_{f,pix} \mod 2$

2. $N_{zph} \geq N_{zph,lower}$

3. $N_{zph} \mod c_{pad} = 0$

4. $N_{zph} \leq N_{zp} = c_{pad}N$

5. $N_{zp} \mod N_{zph} = 0$

Conditions 1 and 2 set a lower bound for N_{zph} where the first follows from the definition of N_{zph} and the second allows for an artificial lower bound if desired. Condition 3 ensures that the padding procedure used to mitigate the emergence of high-frequency components in the original intensity image spectrum due to discrete Fourier transformation can be reversed for the reduced gradient field dimension. Condition 4 sets an upper bound equal to the original problem dimension. Condition 5, however, requires further explanation which is provided below.

The spatial discretisation caused by the physical imaging hardware results in an evenly spaced spatial grid $\mathcal{G}_1$ defined by the set of coordinates $S \times S$ with $S = \{-\frac{D_{pupil}}{2} + nd, \forall n = 0, \dots, N\}$ and $d = \frac{D_{pupil}}{N}$ being the physical pixel size. One would prefer the reconstructed phase, the used basis functions, and the extracted gradients to be defined on the same grid $\mathcal{G}_1$ or a sub-region thereof. This not only for the sake of consistency in terms of applied theory, but also numerical implementation as other grids can result in re-computation of the aforementioned functions.

If the full non-zero frequency spectrum of a sampled function is assumed to be captured within a window, then the zero-padding or removing of padding on this window prior to inverse discrete Fourier transformation would be equivalent to expanding or shrinking the continuous function in the spatial domain respectively before sampling it again. This will be referred to as trimming the spectral domain. The amount of shrinking or expanding of the function is proportional to $\frac{N_{zp}}{N_{zph}}$ where N_{zp} will be related to the sampling grid, so if this ratio is an integer, the sampled version of the shrunk or expanded function consists of a sub-sampling of the original sample function with no interpolation required. As such, choosing $N_{zp} \mod N_{zph} = 0$ means a sub-set of values on the original sampling grid $\mathcal{G}_1$ will be used, keeping the used grid consistent. It is worth noting that in numerical implementation special care must be taken that the resulting function is scaled correctly according to the new spectrum size.

An example of the behaviour explained above is given in Figure 5-3 where a sinusoid in Figure 5-3a, whose frequency spectrum is known to be containable within a window, is subjected to the Fourier

spectral trimming procedure above and the resulting reconstructed function is compared to a sub-sampled version of the original sinusoid. As can be seen, the results are identical.

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

31

(a) Densely sampled sinusoidal function

(b) Spectrum of the sinusoidal function in (a)

(c) An extracted version of the spectrum in (b) with an integer ratio of the size

(d) Inverse Fourier transform of the extracted spectrum in (c) showing a reconstructed sinusoid clearly containing sub-sampled values of (a)

(e) A sub-sampled version of (a) with sub-sampling step corresponding to the size ratio of (b) and (c)

Figure 5-3: Illustration of the Fourier spectrum trimming procedure applied to a sinusoidal function

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors

32

Removing the unnecessary padding reduces computational cost for both gradient calculation, as well as phase reconstruction and was found to have only a small effect on accuracy. As an illustration, Figure 5-4 shows the root-mean-square error (RMSE) of a low-order aberration reconstructed using all valid N_{zph} values between 8 and 220 for a 220×220 pixel detector and an 8×8 rectangular Shack-Hartmann sensor. The reconstruction procedure used is discussed in Section 5-4.
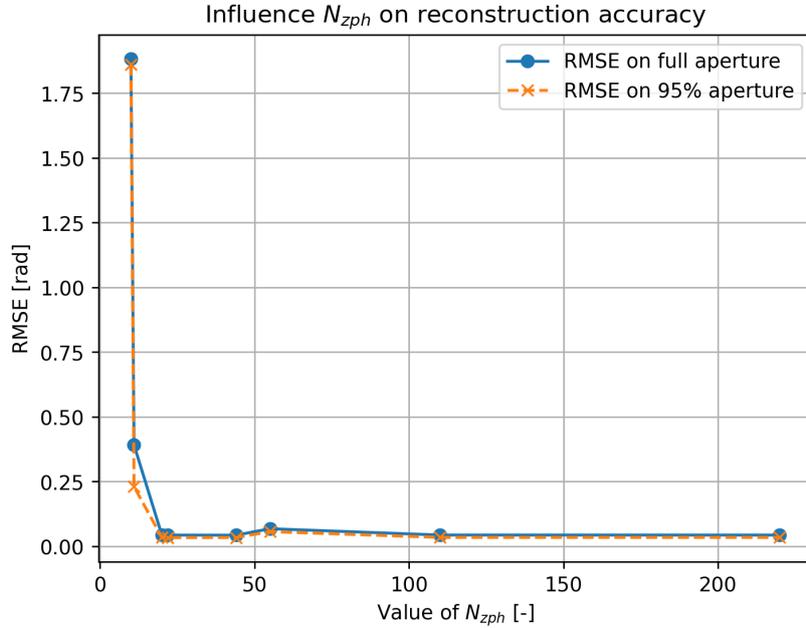
Figure 5-4: Effect of N_{zph} on reconstruction accuracy. Note: the higher N_{zph} is, the more computationally expensive gradient extraction and modal phase retrieval become

5-3 Phase Gradient Extraction on Hexagonal Spot Grids using Fourier Demodulation on a Trimmed Spectral Domain

The theory discussed in Section 5-1 and Section 5-2 considers only a rectangular MLA[29, 34], but hexagonal arrays are interesting for phase retrieval due to their higher spot density. A question that can be posed, therefore, is whether a similar procedure can be defined for hexagonal arrays. This question was found to be unanswered as no prior literature was found that attempted to perform Fourier demodulation, based on the shearing interferometer model of Primot, on hexagonal arrays.

The basic properties of the Shack-Hartmann pattern leading to the derivation of the shearing interferometer model also apply to hexagonal arrays. In fact, when observing the grating equation of a hexagonal array given in Equation 5-19 the similarity to Equation 5-1 is difficult to ignore. As can be observed from this equation, a hexagonal grid too is a bi-periodic lattice that can be expressed as a simple combination of 2D comb functions, be it scaled and shifted comb functions. Similarly to how a Fourier transform of a rectangular lattice with pitch p is again a rectangular lattice with pitch $\frac{1}{p}$, the Fourier transform of a hexagonal lattice becomes a hexagonal lattice with inverse spacing. To illustrate this, the spectrum of a hexagonal Shack-Hartmann intensity pattern is provided in Figure 5-6.
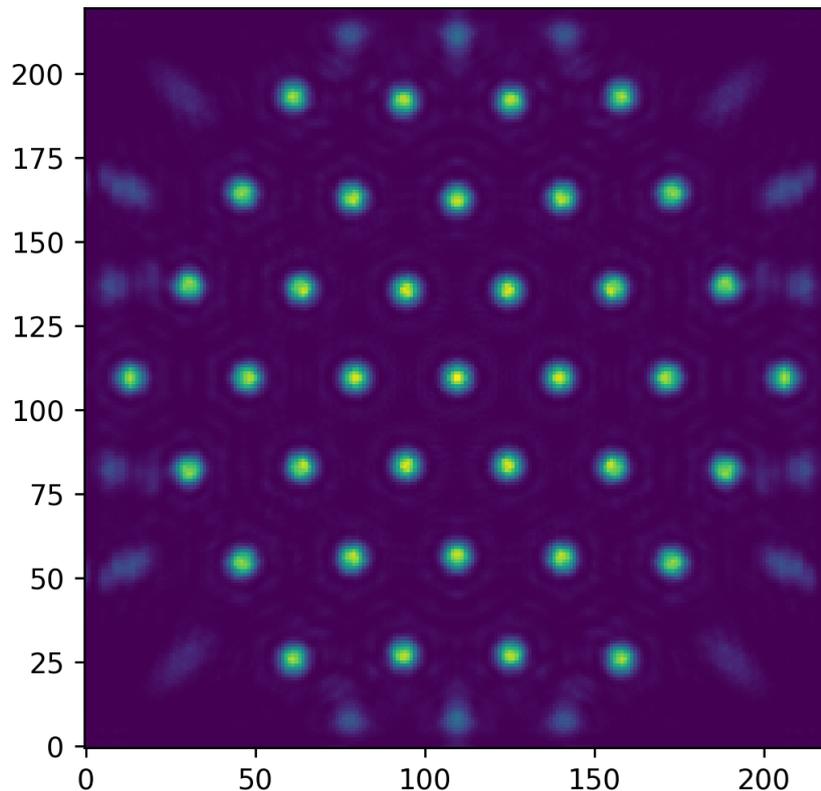
Figure 5-5: Hexagonal Shack-Hartmann pattern obtained from a spherical aberration

A Shack-Hartmann intensity pattern obtained for a hexagonal array is provided in Figure 5-5 for a spherical aberration.

$$
G(x,y) = \left[e^{i\pi \frac{x^2+y^2}{\lambda f_\mu L}} \operatorname{Hex}_{D_{h,o}}(x,y) \right] * \left(\operatorname{comb}_{D_{h,i},\frac{3}{2}D_{h,o}}(x,y) + \operatorname{comb}_{D_{h,i},\frac{3}{2}D_{h,o}}\left(x - \frac{D_{h,o}}{2}, y - \frac{3D_{h,o}}{4} \right) \right)
$$

$$(5\text{-}19)$$

Looking at Figure 5-6, what could be interpreted as harmonic spectra are present at each hexagonal lattice point. To prove that these are indeed harmonics, a proof by induction can be performed. For this purpose, a Shack-Hartmann pattern generated by a spherical aberration will be considered whose spectrum will be a variation of the one given in Figure 5-6.

If the Fourier demodulation method is extendable to hexagonal arrays, the phase relation in Equation 5-12 must also be applicable.

However, as this equation was derived for a rectangular grid, some alterations are required before it can be used. Firstly, a new set of indices is proposed: (k, l, m) where k and l are along the horizontal and vertical axis respectively, and m is used to traverse diagonally between the two shifted rectangular lattices. This means the spectrum of harmonic $H_{k,l,m}$ is located k and l spots horizontally and vertically along the non-shifted rectangular lattice, and m spots shifted diagonally along the lines connecting the two rectangular lattices.

34

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors
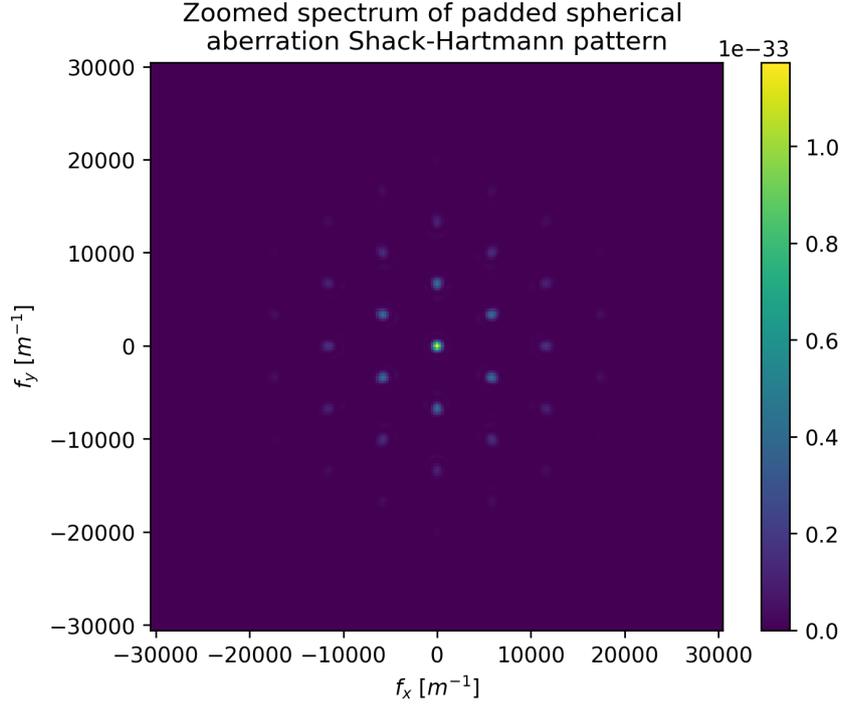
Figure 5-6: Zoomed spectrum of a hexagonal Shack-Hartmann intensity pattern showing a hexagonal lattice structure with inverse spacing

In a hexagon, the relation between the inner and outer diameter is given by $D_{h,i} = \frac{\sqrt{3}}{2} D_{h,0}$. For a hexagonal array consisting of hexagons with their corners pointed upwards, the horizontal spacing between hexagons on the same row is $D_{h,i}$, the vertical spacing between hexagons on the same rectangular lattice is given by $\frac{3}{2} D_{h,o}$. Upon Fourier transformation, it was found that the resulting hexagonal lattice has a horizontal spacing of $\frac{2}{D_{h,i}}$ and a vertical spacing of $\frac{2}{\frac{3}{2} D_{h,o}}$. It is worth noting that flipping the hexagonal grid on its side would result in swapping of these horizontal and vertical spacings everywhere in the following equations.

Using this knowledge and the proposed indices, and expressing all terms with respect to a common length $D_{h,o}$, the centre of the (k, l, m) harmonic spectrum can be found to be located at frequency $\left(\frac{4}{\sqrt{3}} \left(k + \frac{1}{2} m \right), \frac{4}{3} \left(l + \frac{1}{2} m \right) \right) \frac{1}{D_{h,o}} = (k_{hex}, l_{hex}) \frac{1}{D_{h,o}}$. As this coordinate system has its unit length expressed as frequency step $\frac{1}{D_{h,o}}$, it can be made analogous to the $\frac{1}{p_{\mu L}}$ frequency step used in Section 5-2. Recalling the definition of ∂p in Equation 5-13, being $\frac{f_{\mu L}}{p_{\mu L}}$, this can also be replaced by $\frac{f_{\mu L}}{D_{h,o}}$.

Performing these substitutions turns the relation between the harmonic and the phase gradient into the one provided in Equation 5-20. Or, for discrete Fourier transform implementation, Equation 5-21 to account for virtual tip/tilt caused by discretisation.

$$
\begin{aligned}
\frac{\partial \phi}{\partial x_{k_{hex},l_{hex}}} &= -\frac{1}{\lambda\sqrt{k_{hex}^2+l_{hex}^2}\,\frac{f_{\mu L}}{D_{outer}}}\mathrm{angle}\left(\frac{H_{k,l,m}(x,y)}{(K-\mid k\mid)(K-\mid l\mid)}\right)\\
&= -\frac{1}{\lambda}\mathrm{angle}\left(\left(\frac{H_{k,l,m}(x,y)}{(K-\mid k\mid)(K-\mid l\mid)}\right)^{\frac{1}{\sqrt{k_{hex}^2+l_{hex}^2}\,\frac{f_{\mu L}}{D_{outer}}}}\right)
\end{aligned}
\tag{5-20}
$$

$$
\begin{aligned}
H_{obs_{k,l,m}}(x_i,y_i) &= (K-\mid k\mid)(K-\mid l\mid)e^{-i\sqrt{k_{hex}^2+l_{hex}^2}\frac{f_{\mu L}}{D_{outer}}\lambda\frac{\partial \phi_{obs}}{\partial x_{k_{hex},l_{hex}}}}\\
&= (K-\mid k\mid)(K-\mid l\mid)e^{-i\sqrt{k_{hex}^2+l_{hex}^2}\frac{f_{\mu L}}{D_{outer}}\lambda\left(\frac{\partial \phi}{\partial x_{k_{hex},l_{hex}}}+\frac{\partial \phi_{static}}{\partial x_{k_{hex},l_{hex}}}+\frac{\partial \phi_{virtual}}{\partial x_{k_{hex},l_{hex}}}\right)}\\
&= H_{k,l,m}(x_i,y_i)H_{ref_{k,l,m}}(x_i,y_i)\\
\Longleftrightarrow H_{k,l,m}(x_i,y_i) &= \frac{H_{obs_{k,l,m}}(x_i,y_i)}{H_{ref_{k,l,m}}(x_i,y_i)}\\
\Longleftrightarrow \frac{\partial \phi}{\partial x_{k_{hex},l_{hex}}}(x_i,y_i) &= -\frac{1}{\lambda}\mathrm{angle}\left(\left(\frac{H_{obs_{k,l,m}}(x_i,y_i)}{H_{ref_{k,l,m}}(x_i,y_i)}\right)^{\frac{1}{\sqrt{k_{hex}^2+l_{hex}^2}\frac{f_{\mu L}}{D_{outer}}}}\right)
\end{aligned}
\tag{5-21}
$$

The extraction of the gradient itself is also slight different from the one provided in Section 5-2. As the harmonics are no longer spaced apart in a rectangular grid, a rectangular extraction window would not be desirable. The main reason for this is the potential overlap between harmonic spectra. As such, the optimal extraction window shape in a hexagonal grid is, of course, a hexagon. Due to the inverse spacing in frequency domain, the hexagonal mask used is always the flipped sub-aperture shape, meaning a sub-aperture with pointy top has an extraction mask with a flat top.

It is important to note that the same extraction procedure is used as discussed in Section 5-2 before a hexagonal mask is applied. This ensures energy leaks are avoided. Similarly, the same observations about zero-padding harmonic spectra to $N_{zph} \times N_{zph}$ also apply to hexagonal harmonics.

An example of an extracted harmonic spectrum is provided in Figure 5-7a where the $(0,0,1)$ harmonic was extracted. Applying Equation 5-21 results in Figure 5-7b. For comparison, Figure 5-7c shows the analytical gradient corresponding to the $(0,0,1)$ harmonic. As can be seen, these functions are highly similar and as such Fourier demodulation for hexagonal arrays has hereby been proven by induction.

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal
Shack-Hartmann Wavefront Sensors

36

(a) Extracted and padded $(0, 0, 1)$ harmonic with $N_{zph} = 20$

(b) Extracted gradient from $(0, 0, 1)$ harmonic of spherical aberration resulting in coma shape

(c) Analytical gradient of a spherical aberration oriented along the $(0, 0, 1)$ harmonic direction

Figure 5-7: Extracted hexagonal harmonic with corresponding gradient and analytical definition. The similarity serves as a proof by induction. Simulation performed on a hexagonal Shack-Hartmann sensor with 3 lenslet rings, $D_{pupil} = 1.2$mm and focal length 10mm

One important observation must be made about the retrieved gradient fields: as a hexagonal mask is used on a small pixel grid, it is likely that some relevant high frequency components will be accidentally eliminated and the gradient will be smoothed out. As such, the gradients obtained using hexagonal Fourier demodulation may cause slight inaccuracies during phase retrieval.

5-4 Modal Phase Retrieval from Phase Gradient Fields

This section will treat modal phase retrieval using phase gradient fields. Firstly, the reconstruction procedure will be discussed in Section 5-4-1, followed by a brief discussion on reconstruction bases in Section 5-4-2.

5-4-1 Reconstruction Procedure

The procedures discussed in Section 5-2 and Section 5-3 result in approximative phase gradient fields along both orthogonal and non-orthogonal directions. The approximative nature follows from the base principle upon which the Fourier demodulation is built: modulation of the lenslet centroid lattice[29] by the gradient field. This principle ties the Fourier demodulation method to centroiding and as such, the extracted gradient can only be assumed to be close to the real gradient for aberrations whose change over a lenslet sub-aperture is small[29, 34]. This would result in the average gradient over the sub-aperture, which is interpreted as a tip/tilt causing the spot shifts, being close to the actual gradient. Any aberrations of higher order will have their gradient field aliased to lower orders upon gradient extraction and therefore only approximate the real one. Therefore, reconstruction using gradients provided by the Fourier demodulation method does not reconstruct more modes than classic centroiding, it can simply deal with higher magnitude aberrations.

Assuming a set of (non-)orthogonal phase gradient fields is given, the procedure for reconstructing the phase itself can be performed in one of two ways: zonally or modally.

In zonal reconstruction, the wavefront is reconstructed region- or pixel-wise. An example of the former would be stitching piece-wise linear phases together from average tip/tilt over sub-apertures obtained from centroiding. An example of the latter is point-wise integration of the phase gradient fields as this considers individual gradient values per pixel. If the full gradient field is present, one possible integration method would be Fourier integration. Fourier integration makes use of the Fourier derivative property, shown for an arbitrary derivative in Equation 5-22 where $\vec{v}$ is a unit vector, which can be written as a projection problem for multiple gradient fields as illustrated in Equation 5-23 where γ is a regularisation term to deal with zero-division at the DC component. Alternatively, $\gamma = 0$ can be used and this DC component can be set to zero manually. This method is used in various pieces of literature dealing with integration of phase gradients in literature such as the work of Carmon and Ribak and Velghe et al.[30, 46].

$$\mathcal{F}\left\{ \frac{\partial f}{\partial \left(\vec{v} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \right)} \right\} = i \left(\vec{v} \cdot \begin{bmatrix} \omega \\ \xi \end{bmatrix} \right) \mathcal{F}\{f\} \tag{5-22}$$

$$\mathcal{F}\{\phi(x,y)\} = \frac{\sum_{n=1}^{N} \left[\left(i \left(\vec{v_n} \cdot \begin{bmatrix} \omega \\ \xi \end{bmatrix} \right) \right)^* \mathcal{F}\left\{ \frac{\partial \phi(x,y)}{\partial \vec{v} \cdot \begin{bmatrix} x \\ y \end{bmatrix}} \right\} \right]}{\sum_{n=1}^{N} \left| i \left(\vec{v_n} \cdot \begin{bmatrix} \omega \\ \xi \end{bmatrix} \right) \right|^2 + \gamma}, \qquad \gamma \geq 0 \tag{5-23}$$

Inverse Fourier transformation would then yield the phase $\phi(x,y)$. Fourier integration has a problem, however, in that it is easily corrupted by gradient field inconsistencies. In addition, application to non-rectangular apertures requires the definition of boundary conditions[33], adding complexity. Numerical integration, for example through finite difference approximation as has been used in literature by Legarda-Sáenz et al. who used a quadratic optimisation problem to obtain a robust numerical integration algorithm[47], suffers from similar gradient inconsistency and boundary condition problems.

Zonally reconstructed phases also lack guaranteed continuity and/or smoothness, while common aberrations such as tip/tilt or defocus do possess these properties.

It is for these reasons that zonal reconstruction was cast side, and modal reconstruction was studied instead, an approach rarely treated for Fourier demodulation. As mentioned before in Section 2-2, modal phase representation expresses the phase as a sum of weighted basis functions $f_k(x,y)$ as shown in Equation 5-24. While this basis usually contains infinite functions in theory, numerical limitations require a truncation of the basis to K terms, a problem which was previously discussed in Section 2-2.

$$\phi(x,y) = \sum_{k=0}^{K} \alpha_k f_k(x,y) \tag{5-24}$$

If the basis functions used are smooth and continuous, the represented phase will be as well, which is a desirable quality for representing common aberrations. In addition, the phase can be represented by a single set of coefficients, which will be much smaller in memory consumption than a numerical pixel-wise representation of the phase obtained in zonal reconstruction.

Reconstruction using centroid displacement and the Zernike basis was discussed before in Section 3-1. This came down to solving Equation 5-25 for the coefficient vector $\vec{\alpha}$ using a total of N_{MLA} lenslet

38

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

centroids within the aperture where s_{x,c_i} and s_{y,c_i} are the displacements of the centroids from their reference positions in x and y direction respectively. As each centroid has 2 displacements, this results in a total of $2N_{MLA}$ modes that can theoretically be retrieved before the least-squares problem becomes under-determined.

$$
\begin{bmatrix}
\frac{\partial Z_1^{-1}}{\partial x}(x_{c_0}, y_{c_0}) & \cdots & \frac{\partial Z_n^m}{\partial x}(x_{c_0}, y_{c_0}) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial x}(x_{c_{N_{MLA}}}, y_{c_{N_{MLA}}}) & \cdots & \frac{\partial Z_n^m}{\partial x}(x_{c_N}, y_{c_N}) \\
\frac{\partial Z_1^{-1}}{\partial y}(x_{c_0}, y_{c_0}) & \cdots & \frac{\partial Z_n^m}{\partial y}(x_{c_0}, y_{c_0}) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial y}(x_{c_{N_{MLA}}}, y_{c_{N_{MLA}}}) & \cdots & \frac{\partial Z_n^m}{\partial y}(x_{c_{N_{MLA}}}, y_{c_{N_{MLA}}})
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\ \vdots \\ \alpha_{m,n}
\end{bmatrix}
= \frac{-1}{f}
\begin{bmatrix}
s_{x,c_0} \\ \vdots \\ s_{x,c_{N_{MLA}}} \\ s_{y,c_0} \\ \vdots \\ s_{y,c_{N_{MLA}}}
\end{bmatrix}
\tag{5-25}
$$

However, the Fourier demodulation method discussed in this chapter provides gradient fields, not centroid displacements. Additionally, it can provide non-orthogonal gradients. Re-writing Equation 5-25 to take into account all N_{pix} pixels within the aperture, as well as all n_{deriv} provided gradient fields instead of centroid displacements, results in Equation 5-26. It is worth noting that this only requires the x and y gradient fields of the Zernike functions to be analytically computed before-hand, as all other gradients can be expressed as a linear combination of these computed fields. Solving for the coefficient vector $\vec{\alpha}$ would again yield the modal coefficients of the phase.

$$
\begin{bmatrix}
\frac{\partial Z_1^{-1}}{\partial \vec{v}_0 \cdot \vec{x}}(x_0, y_0) & \cdots & \frac{\partial Z_n^m}{\partial \vec{v}_0 \cdot \vec{x}}(x_0, y_0) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_0, y_0) & \cdots & \frac{\partial Z_n^m}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_0, y_0) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial \vec{v}_0 \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}}) & \cdots & \frac{\partial Z_n^m}{\partial \vec{v}_0 \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}}) \\
\vdots & & \vdots \\
\frac{\partial Z_1^{-1}}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}}) & \cdots & \frac{\partial Z_n^m}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}})
\end{bmatrix}
\begin{bmatrix}
\alpha_1 \\ \vdots \\ \alpha_{m,n}
\end{bmatrix}
=
\begin{bmatrix}
\frac{\partial \phi}{\partial \vec{v}_0 \cdot \vec{x}}(x_0, y_0) \\
\vdots \\
\frac{\partial \phi}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_0, y_0) \\
\vdots \\
\frac{\partial \phi}{\partial \vec{v}_0 \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}}) \\
\vdots \\
\frac{\partial \phi}{\partial \vec{v}_{n_{deriv}} \cdot \vec{x}}(x_{N_{pix}}, y_{N_{pix}})
\end{bmatrix}
\tag{5-26}
$$

$$\iff A\vec{\alpha} = \vec{b}$$

$$\iff (A^T A)\vec{\alpha} = A^T \vec{b}$$

An important observation must be made about this equation, however. As the amount of data is much larger than with centroiding one could expect the number of retrievable modes to be much larger as well: for n_{deriv} derivatives and N_{pix} gradient values within the aperture one would expect $n_{deriv}N_{pix}$ modes to be retrievable.

This, however, is incorrect. One must remember that the gradient fields can be expressed as a linear combination of the x and y gradient fields, meaning any provided gradient fields other than the x and y gradient does not provide additional mode insights, just more observations of the same gradients. This would already decrease the number of retrievable modes to a maximum of $2 \cdot N_{pix}$. Additionally, the theory on which Fourier demodulation is based in modulation of the spot grid, which contains only N_{MLA} spots. As such, the number retrievable modes is actually $2N_{MLA}$, similar to centroiding. In practice, the value of N_{pix} will allow a slightly larger number of modes to retrieved if chosen such that the system is over-determined even with two gradients.

5-4-2 Reconstruction Bases

After discussing the reconstruction procedure in Section 5-4-1, a brief note will be given on reconstruction bases. As mentioned in Section 5-4-1 the used basis for reconstruction is the Zernike basis. While this choice is fixed in this thesis, alternatives will be briefly discussed.

The Zernike basis, while common in literature for centroiding-based reconstruction[4, 15], is flawed. It is orthonormal on the unit circle and therefore well-suited for the representation of wavefronts, but its gradients are not orthogonal. This non-orthogonality leads to mode cross-talk when the gradients are fitted to an extracted gradient field or centroid displacements. Mathematically, this means that coefficients $\vec{\alpha}$ making up the actual wavefront $\phi = \sum_{k=1}^{K} \alpha_k f_k$ will not correspond to the retrieved coefficients $\hat{\vec{\alpha}}$ obtained from fitting the gradient to the basis gradients $\nabla\phi = \sum_{k=1}^{K} \alpha_k \nabla f_k$ while for a basis with orthogonality in the basis and its gradient this would be the case. This effect is usually especially noticeable around the edges of the reconstructed phase.

There is an additional source of error: truncation. As discussed earlier in Section 2-2, the Zernike basis is only orthogonal when the full basis is considered, upon truncation to a finite number of terms this orthogonality is no longer guaranteed.

In an effort to mitigate the lack of gradient orthogonality, the vector polynomial basis by Zhao and Burge[19, 20] was implemented in code. This basis was derived specifically for orthogalisation of the Zernike gradients. The polynomials in this set are defined by Equation 5-27 and are dependent on the Zernike gradients, which themselves are related to the actual Zernike polynomials. It is worth noting that the j indices in this equation follow a different ordering from Noll's ordering[17], meaning careful book-keeping is required.

$$
\begin{cases}
\vec{S}_j & = \frac{1}{\sqrt{2n(n+1)}} \nabla Z_{ANSI,j} \qquad ,n=m \\
\vec{S}_j & = \frac{1}{\sqrt{4n(n+1)}} \left(\nabla Z_{ANSI,j} - \sqrt{\frac{n+1}{n-1}} \nabla Z_{ANSI,j(n'=n-2,m'=m)} \right) \qquad ,n \neq m \\
j - j' \text{ even when } m \neq 0
\end{cases}
$$
(5-27)

Upon fitting the vector polynomial to the extracted gradient fields, resulting in a set of coefficients $\vec{\gamma}$, the corresponding Zernike coefficients $\vec{\alpha}$ of the phase could be calculated using Equation 5-28. A problem becomes immediately apparent in this equation: a future-relation in the coefficients. In fact, this basis also loses full orthogonality upon truncation and as such the cross-talk issue persists. An example of an S polynomial is given in Figure 5-8a.

$$
\begin{cases}
\alpha_j = \frac{\gamma_{j(n,m)}}{\sqrt{2n(n+1)}} \qquad ,n=m \\
\alpha_j = \frac{\gamma_{j(n,m)}}{\sqrt{4n(n+1)}} - \frac{\gamma_{j'(n+2,m)}}{\sqrt{4(n+1)(n+2)}}
\end{cases}
$$
(5-28)

An additional source of error could be the fact that the gradient fields obtained from Fourier demodulation consist of average gradient values extracted from the spot displacement caused by the average tip/tilt over each sub-aperture. To obtain a closer resemblance between the basis gradients and these averaged gradients, the analytical gradients defined on a $N \times N$ pixel grid were averaged down to a $N_{zph} \times N_{zph}$ grid. An example of a coma which has undergone this averaging procedure is given in Figure 5-8b. The effect of this operation on mode cross-talk was negligible. Similarly, a re-sizing of the gradients via the spectrum trimming method discussed in Section 5-2 was implemented, an example of which can be seen in Figure 5-8c. This alteration, however, had unintended side-effects such as the re-sized piston function in Figure 5-8e and as such had a detrimental impact on mode cross-talk.

Ultimately, it was decided to simply use the sub-sampled Zernike gradient basis as shown in Figure 5-8d. Part of the rationale for this decision was the fact that the low-order method discussed in this chapter is meant as a pre-conditioner for the hybrid method which will be presented in Chapter 7. As this hybrid method performs a secondary phase-fitting using intensity approximation, it uses the Zernike functions themselves instead of their non-orthogonal gradients which is expected to partially alleviate cross-talk.

In future implementations of the algorithm, the use of a basis with orthogonal gradient fields as well as conservation of orthogonality under truncation would be suggested. One such basis would be the set of circular Bessel functions, discussed previously in Section 2-2[18].

(a) Example of a vector polynomial basis function for gradient fitting

(b) Coma function averaged from $N = 220$ to $N_{zph} = 20$

(c) Resized coma function obtained using trimming of the image spectrum

(d) Sub-sampled coma function

(e) Resized piston function obtained using trimming of the image spectrum
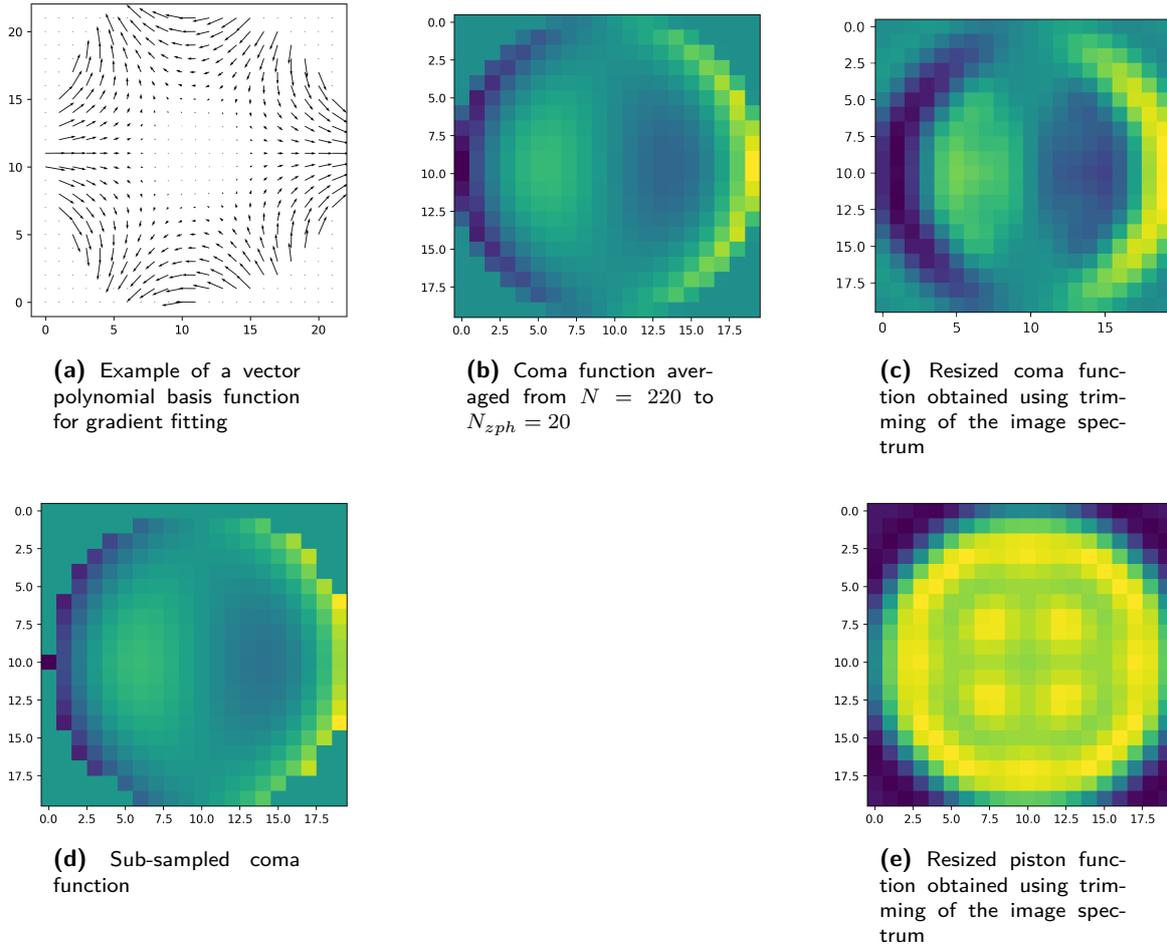
Figure 5-8: Attempted Zernike basis alterations for gradient fitting

5-5   Reconstruction Results and Conclusion

The purpose of this chapter was to present a modal gradient-based phase retrieval method with an augmented practical implementation of Fourier demodulation for approximative gradient field extraction.

After a brief presentation of the shearing interferometer model in Section 5-1, the classic rectangular Fourier demodulation method was explained in Section 5-2, followed by implementation specifics for use with discrete Fourier transformation, and lastly an augmentation to obtain shrunken derivative fields for modal reconstruction. It was found that shrinking the retrieved derivative has little influence on reconstruction accuracy as demonstrated in Figure 5-4 while helping speed up reconstruction.

Subsequently, a novel application of Fourier demodulation to hexagonal Shack-Hartmann wavefront sensors was presented in Section 5-3 while also implementing the same shrunken derivative field augmentation.

Modal phase reconstruction was discussed in Section 5-4-1 where first the disadvantages of using zonal reconstruction techniques such as Fourier integration were discussed, followed by the discussion of modal reconstruction using gradient fields. In addition, it was found that the number of retrievable modes is similar between Fourier demodulation and centroiding. Lastly, reconstruction bases were treated in Section 5-4-2 where it was found that subsampling of the regular Zernike gradient basis provided sufficiently accurate results, while alternative implemented bases either had negligible effect or suffered from truncation issues. A suggestion was made to attempt the implementation of circular Bessel functions as a basis for future work.

With the methodology treated in this chapter, the retrieval of low-order phase from both rectangular and hexagonal arrays can be performed. While Fourier demodulation is a previously existing method, its application to hexagonal arrays is novel and the use of modal reconstruction is uncommon for gradient field methods. As such, a comparison to literature will be performed by using a centroiding algorithm with modal reconstruction. The centroiding methodology used is classic first-moment centroid calculation per sub-aperture with gamma correction applied for noise reduction which was found to have highly accurate displacement calculations in literature[9].

All methodology discussed in this chapter have been implemented into a custom Python simulation and phase reconstruction framework, optical propagation is performed using the HCIpy library[45].

All simulations will be performed for a wavelength of 632.8nm, a MLA focal length of 10mm, a circular pupil with diameter 1.2mm, and a 220×220 detector pixel grid. Rectangular Shack-Hartmann patterns will be obtained for an 8×8 MLA with a circular pupil, while hexagonal patterns will be obtained for an MLA with three lenslet rings. Performance will be measured using RMSE with respect to the ground truth over the full aperture as well as over only 95% of the aperture to reduce the effect of mode cross-talk error on the fringes.

To obtain a fair comparison and limit the influence of mode cross-talk due to basis truncation, the reconstruction will be performed using a small number of modes, namely 40, for all methods. It is worth noting, however, that the gradient field method is capable of reaching higher accuracy when more modes are allowed while centroiding suffers significantly from mode cross-talk. All reconstructions will be performed for a spherical aberration of quarter-wavelength magnitude. The magnitude was chosen to be small in order to ensure spot containment within each sub-aperture as additional spot indexing methods would introduce additional uncertainty.

The second harmonics $(2, 0)$ and $(0, 2)$ and the equivalent hexagonal harmonics will be used for Fourier demodulation based reconstruction due to their proven higher excitability in literature[29].

Firstly, the Shack-Hartmann patterns together with identified spot and reference centroid locations are given in Figure 5-9.

(a) Obtained intensity pattern for rectangular Shack-Hartmann configuration with obtained centroid locations indicated

(b) Obtained intensity pattern for hexagonal Shack-Hartmann configuration with obtained centroid locations indicated
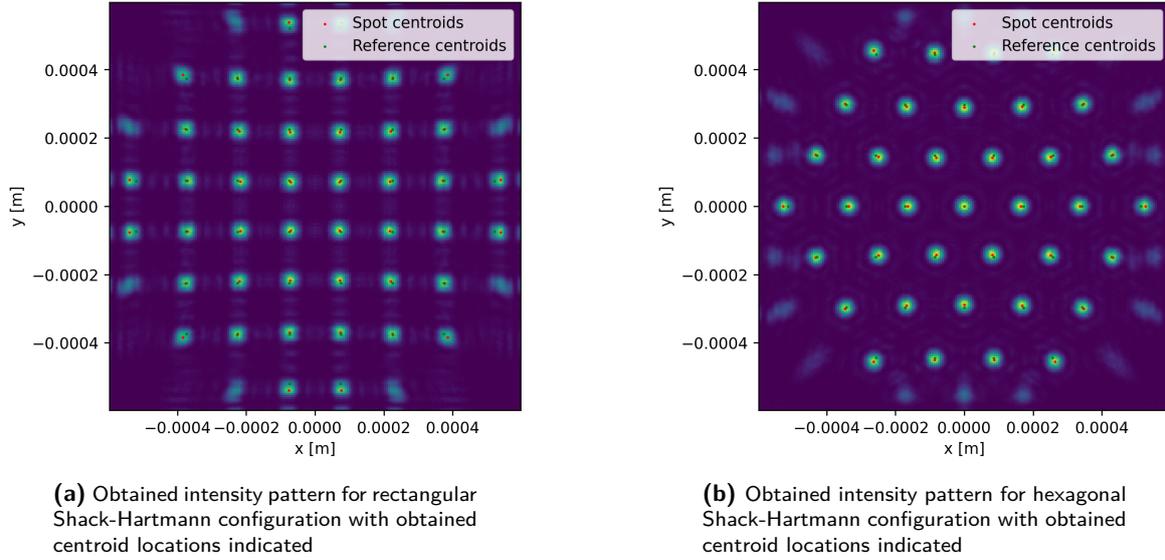
Figure 5-9: Intensity pattern obtained for spherical aberration of quarter wavelength magnitude on Shack-Hartmann sensors with focal length 10mm, a circular pupil, and aperture diameter 1.2mm

Reconstruction results for the rectangular pattern are provided in Figure 5-10. As can be seen, the Fourier demodulation gradient field modal reconstruction method is capable of retrieving the phase with a lower error than centroiding. Due to this, the former will be used as a benchmark for later chapters. The main source of error in the centroiding algorithm appears to be caused by mode cross-talk due to the limited number of data-points available, this was also noticed during simulation when the number of retrieved modes was increased. The Fourier demodulation-based method, however, fared much better with increasing mode count due to the slight data-point redundancy caused by the padding to $N_{zph} \times N_{zph}$ values.

Reconstruction results for the hexagonal pattern are provided in Figure 5-11. As can be seen, centroiding outperforms the Fourier demodulation gradient field modal reconstruction. The cause is likely to be the elimination of high-frequency components of the gradient field as previously discussed in Section 5-3 on one hand, and the clearer spot centroids for hexagonal spot patterns on the other. As will become apparent in Chapter 7, hexagonal Fourier demodulation still forms a promising method for pre-conditioning. In addition, Fourier demodulation is not bound by sub-aperture independence, giving it a wider range of applicability than centroiding.

In terms of speed, centroiding is difficult to beat. While no computational complexity analysis was performed, reconstruction times were recorded. For the simulation parameters above and values $N_{zph} = 20$ and $c_{pad} = 3$, the average combined double gradient extraction and modal reconstruction with 110 coefficients took 0.6s on an Intel Core I7-7700HQ CPU at 2.81 GHz (110 coefficients were chosen for consistency with later chapters). Conversely, centroiding took only between 0.15-0.2s for 110 coefficients. This, however, is unsurprising for multiple reasons: centroiding does not require multiple Fourier transform operations nor a lot of array manipulation, in addition the code is at this time unoptimised. However, it is unlikely to exceed computational efficiency of centroiding unless a very large number of sub-apertures is used.

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

43

LO only, RMSE95 0.0828 rad, RMSE full 0.1097 rad
Reconstructed Zernike from 40 coefficients [rad]

Centroiding RMSE95 0.1019 rad, full RMSE 0.111 rad
Reconstructed Zernike from 40 coefficients [rad]

(a) Obtained modal reconstruction using Fourier demodulation gradients, 40 modes

(b) Obtained modal reconstruction using centroiding, 40 modes

Figure 5-10: Reconstruction results for a rectangular Shack-Hartmann sensor using 40 modes

LO only, RMSE95 0.2492 rad, RMSE full 0.2778 rad
Reconstructed Zernike from 40 coefficients [rad]

Centroiding RMSE95 0.1057 rad, full RMSE 0.1261 rad
Reconstructed Zernike from 40 coefficients [rad]

(a) Obtained modal reconstruction using Fourier demodulation gradients, 40 modes

(b) Obtained modal reconstruction using centroiding, 40 modes
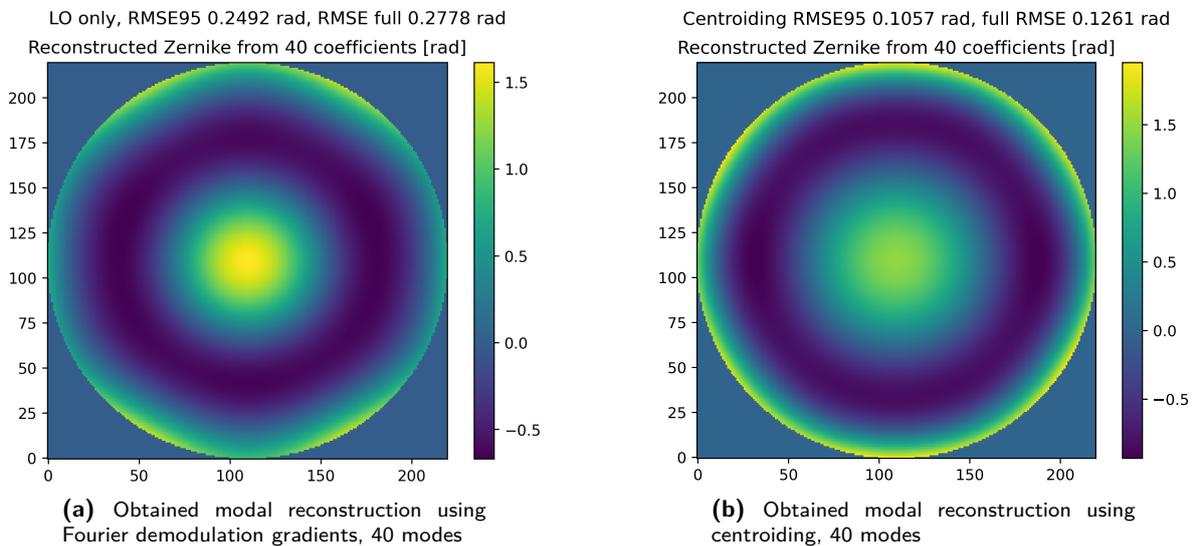
Figure 5-11: Reconstruction results for a hexagonal Shack-Hartmann sensor using 40 modes

44

5. Modal Phase Retrieval on Augmented Fourier Demodulation Gradient Fields for Rectangular and Hexagonal Shack-Hartmann Wavefront Sensors

Modal Phase Retrieval on Full Point-Spread Functions using Taylor Approximated Intensity Function Fitting and Inclusion of Shack-Hartmann Diversity

This chapter will discuss the use of Taylor approximation of the intensity function for modal wavefront reconstruction. Firstly, the base principles and work-out of Taylor intensity approximation will be provided in Section 6-1, followed by the inclusion of diversity and the application to Shack-Hartmann patterns in Section 6-2. The retrieval of modal coefficients will then be discussed in Section 6-3, and a brief conclusion will be provided in Section 6-4.

6-1 Intensity Function Taylor Approximation

The intensity image generated at the detector plane of a single-lens system by a complex wavefront $W = \mathcal{A}(x,y)e^{i\phi(x,y)}$ is given by Equation 6-1 where the direct Fourier relation is a direct consequence of the ideal single-lens assumption [15].

$$ h(x,y) = \mid \mathcal{F}\left\{W\right\} \mid^2 = \mathcal{F}\left\{W\right\} \mathcal{F}\left\{W\right\}^* \tag{6-1} $$

Retrieval of the phase $\phi(x,y)$ from intensity measurements $h(x,y)$ has been previously stated to be a highly non-linear, non-convex and ill-posed problem, with methods in Section 3-2 solving this iteratively with multiple phase diversities, or using a casting to a much higher dimensional problem to obtain convexity.

Conversely, the work of Marinica et al. proposes to solve this problem using first or second order Taylor approximation [11]. If the phase is assumed to be modally represented as a set of weighted basis functions with weights vector $\vec{\alpha}$ then Equation 6-1 can be approximated by Equation 6-2. If a small aberration is assumed, the linearisation point $\vec{\alpha}_L$ can be set to $\vec{\alpha}_L = 0$.

$$h(x,y) \approx h(x,y) \mid_{\vec{\alpha}=\vec{\alpha}_L} + \left(\frac{\partial}{\partial \vec{\alpha}} h(x,y) \bigg|_{\vec{\alpha}=\vec{\alpha}_L} \right)^T \vec{\alpha} + \mathcal{O}(\vec{\alpha}^2) \tag{6-2}$$

This requires the linearisation point intensity $h(x,y) \mid_{\vec{\alpha}=\vec{\alpha}_L}$ and the Jacobian $\frac{\partial}{\partial \vec{\alpha}} h(x,y) \big|_{\vec{\alpha}=\vec{\alpha}_L}$ to be worked out. For complex intensity function definitions, automatic differentiation could be used as discussed in the work of Jurling and Fienup, Wong et al.[48, 49]. While this approach was considered and investigated during the development of the code written for this thesis, compatibility issues and the relative simplicity of the intensity function lead to a preference for analytical derivation.

The modally represented wavefront will be written as Equation 6-3, where the wave amplitude of the incoming wave $\mathcal{A}(x,y)$ is assumed to be constant, and the wavefront is considered at the time of passing the pupil with pupil function $P(x,y)$. In addition, a truncated Zernike basis with $K-1$ terms and a circular pupil function are assumed, but other bases and pupil functions can be substituted if desired without affecting further theory.

$$W(x,y) = P(x,y) e^{i \sum_{k=1}^{K} \alpha_k Z_k(x,y)} \tag{6-3}$$

The notation of Equation 6-3 can then be used to re-write Equation 6-1 to Equation 6-4, where the coordinates x and y will henceforth be dropped for conciseness.

$$h = \left| \mathcal{F} \left\{ P e^{i \sum_{k=1}^{K} \alpha_k Z_k} \right\} \right|^2 = \mathcal{F} \left\{ P e^{i \sum_{k=1}^{K} \alpha_k Z_k} \right\} \left(\mathcal{F} \left\{ P e^{i \sum_{k=1}^{K} \alpha_k Z_k} \right\} \right)^* \tag{6-4}$$

This notation allows the linearisation point intensity to be directly calculated using Equation 6-5.

$$h(x,y)(\vec{\alpha} = \vec{\alpha}_L) = \left| \mathcal{F} \left\{ P e^{i \sum_{k=1}^{K} \alpha_{L,k} Z_k} \right\} \right|^2 \overset{\vec{\alpha}_L=0}{=} |\mathcal{F}\{P\}|^2 \tag{6-5}$$

As Equation 6-4 must be derived with respect to the coefficients $\vec{\alpha}$ instead of spatial coordinates x and y, the Fourier operators can be treated as parentheses. This heavily simplifies the apparent complexity of the intensity function for differentiation.

For the Jacobian, $\frac{\partial}{\partial \vec{\alpha}} h(x,y) \big|_{\vec{\alpha}=\vec{\alpha}_L}$ can be calculated for each α coefficient separately. Employing the standard differentiation rule $\frac{\partial}{\partial \alpha_j} e^{i \sum_{k=1}^{K} \alpha_k Z_k} = i Z_j e^{i \sum_{k=1}^{K} \alpha_k Z_k}$, this results in Equation 6-6. As can be seen, the linearisation point $\vec{\alpha}_L$ is set to zero, this convention will be used for consistency with both the original paper by Marinica et al., and an alternative notation for re-linearisation which will be discussed later in Section 6-2 as well as Chapter 7. It is worth noting that the use of the $\mathfrak{Re}$ operator not only mathematically simplifies the equation, it also noticeably affects both computational speed as well as digital storage requirements as complex arrays require approximately double the memory capacity of regular floating point arrays.

$$
\begin{aligned}
\frac{\partial h}{\partial \alpha_j}(\vec{\alpha} = \vec{\alpha}_L) =& \mathcal{F}\left\{iPZ_j e^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\left(\mathcal{F}\left\{Pe^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\right)^* \\
&+ \mathcal{F}\left\{Pe^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\left(\mathcal{F}\left\{iPZ_j e^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\right)^* \\
=& 2\Re\left\{\left(\mathcal{F}\left\{Pe^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\right)^* \mathcal{F}\left\{iPZ_j e^{i\sum_{k=1}^{K}\alpha_{L,k}Z_k}\right\}\right\} \\
\overset{\vec{\alpha}_L=0}{=}& 2\Re\left\{\left(\mathcal{F}\{P\}\right)^* \mathcal{F}\{iPZ_j\}\right\}
\end{aligned}
$$

(6-6)

Upon closer inspection of Equation 6-5 and Equation 6-6 and Equation B-2, similarities can be seen which can be used to reduce re-computations and thereby speed up the generation of the Jacobian. These are shown in Equation 6-7.

$$
\begin{cases}
h(x,y)(\vec{\alpha} = 0) &= \left|\underbrace{\mathcal{F}\{P\}}_{a}\right|^2 \\[2em]
\frac{\partial h}{\partial \alpha_j}(\vec{\alpha} = 0) &= 2\Re\left\{\underbrace{(\mathcal{F}\{P\})^*}_{a^*}\mathcal{F}\left\{\underbrace{iPZ_j}_{b}\right\}\right\}
\end{cases}
$$

(6-7)

The solving of Equation 6-2 for coefficient vector $\vec{\alpha}$ will be discussed in Section 6-3 for the first-order approximation. Readers interested in the solution of the second-order approximation can find more information in Chapter B. However, an important observation must already be made: ambiguity in the solution. While the Fourier transform is a one-to-one mapping by virtue of the Fourier duality principle and as such is unique for each function, the power spectrum obtained from this Fourier transform is not a one-to-one mapping. As the power spectrum is the same as the intensity image, multiple aberrations can generate the same intensity image. For example, aberrations with symmetric phase fields $\phi_{symm}(x,y)$ will generate an identical intensity image to aberrations with phase fields $-\phi_{symm}(x,y)$. For this reason methods developed in literature, including the work of Marinica et al. used to construct the equations above, use phase diversities to differentiate between these aberrations, an approach discussed further in Section 6-2.

It is also important to note that the use of Taylor approximation as presented in this section fundamentally limits the accuracy of the method to aberrations in close proximity to the linearisation point. As no prior information about the wavefront is assumed, this limits the method to aberrations close to zero and as such closed-loop conditions. However, if more prior information can be obtained about the incoming wavefront, the linearisation point can be set accordingly and broaden the range of retrievable aberrations. An example discussed in the work of [11] is the use of previous phase estimates as linearisation points. This, however, simply moves the problem back in time and still requires closed-loop conditions. Instead, an open-loop single-frame retrieval method will be discussed in Chapter 7 based on Taylor intensity approximation with phase diversity discussed in Section 6-2, which allows open-loop retrieval under common operational conditions.

6-2   Intensity Function Taylor Approximation with (Shack-Hartmann) Diversity

As discussed at the end of Section 6-1 the power spectrum obtained from Fourier transformation, which is equivalent to the intensity image generated by a single-lens ideal system, is not a one-to-one mapping. In practice this means that multiple functions, or aberrations in the context of optics, can generate the same intensity image. A given example was the set of symmetric aberrations with symmetric phase fields $\phi_{symm}(x, y)$ generating the same intensity image as those with phase fields $-\phi_{symm}(x, y)$. Many common aberrations suffer from this ambiguity, for example spherical, astigmatism and defocus aberrations all generate identical spectra to their negatives, which negatively impacts any wavefront correction. An example of this ambiguity is shown in Figure 6-1.

(a) Phase field of a defocus aberration

(b) Phase field of a negative defocus aberration

(c) Intensity pattern of defocus aberration

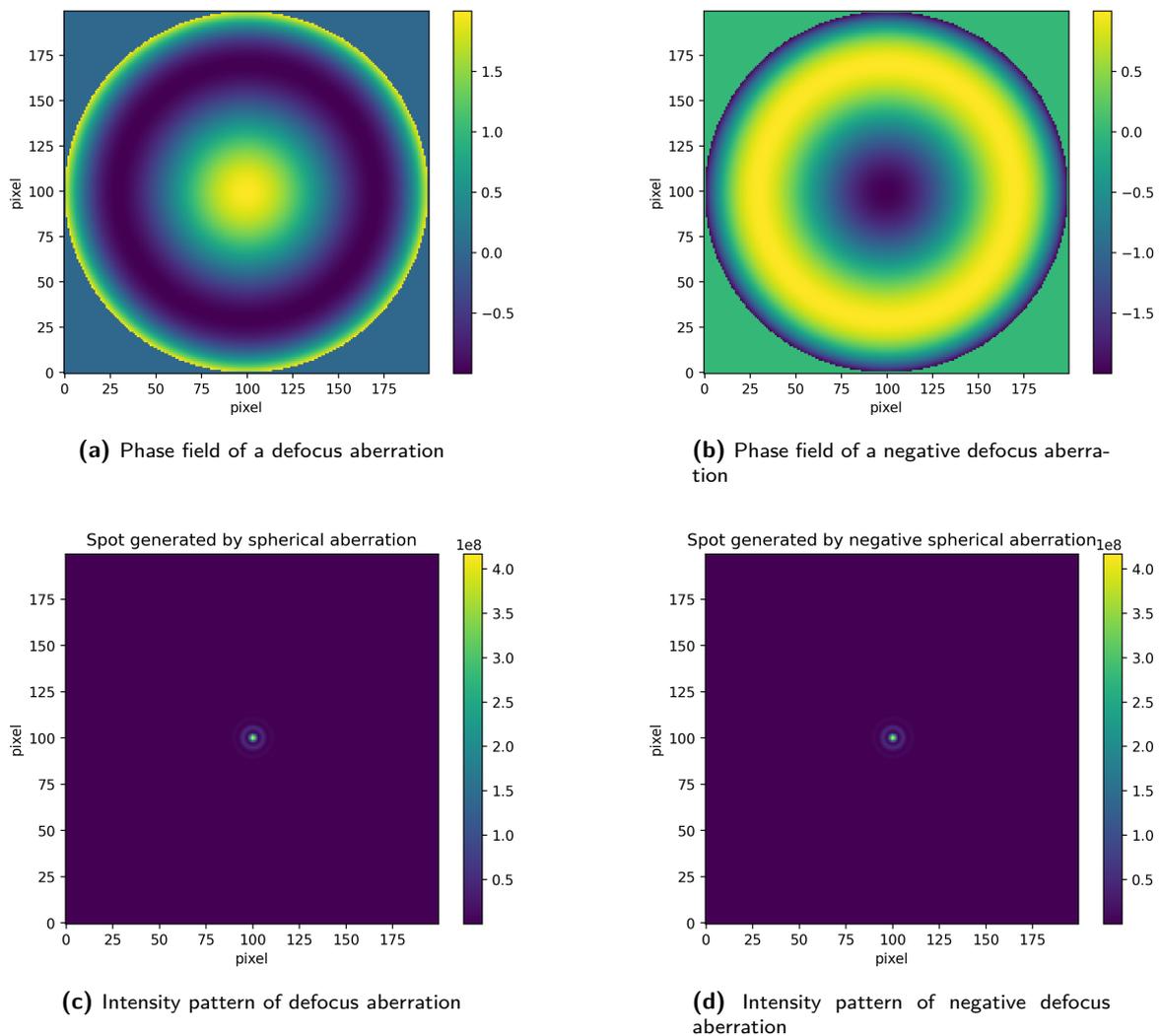**(d)** Intensity pattern of negative defocus aberration

Figure 6-1: Illustration of phase ambiguity in intensity pattern for a defocus aberration and its negative

This ambiguity can be remedied through the use of a phase diversity, i.e. an additional phase delay $\phi_d(x, y)$ which is introduced to the incoming wavefront prior to capture of the intensity pattern by the camera module. As phase delays are additive, the incoming wavefront becomes of the form given in Equation 6-8 where the amplitude is assumed constant and the wavefront at the time of passing the diversity element and the pupil with pupil function $P(x, y)$ is considered.

$$W_{diversity}(x, y) = P(x, y)e^{i\phi_{new}(x,y)} = P(x, y)e^{i(\phi(x,y)+\phi_d(x,y))} \tag{6-8}$$

As can be seen from this equation, the negative of the original phase field $-\phi(x, y)$ will not lead to $-\phi_{new}(x, y)$ as $-\phi(x, y) + \phi_d(x, y) \neq -\phi_{new}(x, y) = -\phi(x, y) - \phi_d(x, y)$ for any non-zero $\phi_d(x, y)$. As such, the intensity patterns captured will not be the same and the ambiguity is avoided. This method is also employed in the work of Marinica et al. where multiple diversities are used to generate multiple intensity images for phase retrieval[11]. The diversities $\phi_{d,i}(x, y)$ can be represented by modal coefficients $\vec{\beta}_i$. In this way, each diversity can be interpreted as changing the linearisation point in Equation 6-2 and further equations of Section 6-1 to $\vec{\alpha}_L = \vec{\beta}_i$. This is the approach followed by Marinica et al..

However, this creates confusion for diversities expressed in different modal bases or arbitrary diversities as these would first need to be cast to the modal basis used for coefficients $\vec{\alpha}$ in the equations of Section 6-1. Therefore, a different notation is proposed instead. As the diversity phase $\phi_d(x, y)$ is independent of the incoming wavefront and as such the phase modal coefficients $\vec{\alpha}$, a slight abuse of wavefront notation can be used to express the diversity wavefront as Equation 6-9 and the linearisation point $\vec{\alpha}_L$ can be kept at zero.

$$W_{diversity}(x, y) = P(x, y)e^{i(\phi(x,y)+\phi_d(x,y))} = P(x, y)e^{i\phi_d(x,y)}e^{i\phi(x,y)} = P_d(x, y)e^{i\phi(x,y)} \tag{6-9}$$

If the phase is again assumed to be modally represented by a Zernike basis as in Equation 6-3 and $P_d(x, y)$ is substituted for $P(x, y)$, the calculus of Section 6-1 can be repeated which results in expressions for the linearisation point intensity, Jacobian and Hessian given in Equation 6-10 where the coordinates x and y have been dropped for conciseness and the similaries have again been highlighted. As can be seen, the only difference between Equation 6-10 and Equation 6-7 is the substitution of P with $P_d = Pe^{i\phi_d}$. This makes introducing arbitrary diversities straightforward and also allows the inclusion of arbitrary phase pre-conditioning as will be discussed in Chapter 7.

$$\begin{cases} h(x,y)(\vec{\alpha}=0) & = \left| \underbrace{\mathcal{F}\{P_d\}}_{a} \right|^2 \\ \frac{\partial h}{\partial \alpha_j}(\vec{\alpha}=0) & = 2\Re\mathfrak{e}\left\{ \underbrace{(\mathcal{F}\{P_d\})^*}_{a^*} \mathcal{F}\left\{ \underbrace{iP_d Z_j}_{b} \right\} \right\} \end{cases} \tag{6-10}$$

Commonly used diversities for this phase diversity Taylor approximation method are defocus aberrations by introducing lenses with known characteristics into the optical path[11], though arbitrary diversities are supported.

This support for arbitrary diversities opens up a highly interesting possibility when recalling the Shack-Hartmann diversity model of Chapter 4. As discussed in Chapter 4, the Shack-Hartmann sensor can be modelled as an ideal single-lens system with a piece-wise linear phase diversity $\phi_{SH}(x, y)$ known as

the Shack-Hartmann diversity. As such, this interpretation allows the direct application of the Taylor intensity approximation method with phase diversity discussed in this section to Shack-Hartmann intensity patterns by setting $P_d(x,y) = P(x,y)e^{i\phi_{SH}(x,y)}$.

The use of Shack-Hartmann diversity as a phase diversity in this method is interesting for a number of reasons. Firstly, Shack-Hartmann sensor patterns contain information redundancy in the form of filtered copies of the original wavefrontpoint-spread function (PSF) generated by each micro-lens. These copies are spread out spatially, reducing the effect of localised distortions in the optical path or electrical noise of the camera on the retrieved modal coefficients. Secondly, the spreading of information in practice raises the signal-to-noise ratio of the captured image. Lastly, a change in wavefront not only has a clear influence on the interference pattern between Shack-Hartmann spots and the spots themselves, but also a spatial influence as the location of the generated spots will move noticeably with different aberrations.

Perhaps more importantly, however, is the fact that low-order phase retrieval methods like those discussed in Chapter 5 can be used to obtain an initial estimate of the phase ϕ from the Shack-Hartmann pattern which can be used to pre-condition the Taylor intensity approximation method. This can be performed on a single image frame and does not require previous estimates, opening the path to open-loop retrieval. This two-step method will be discussed in Chapter 7 and forms the main contribution of this thesis to the field of phase retrieval.

The solution of the first-order Taylor intensity approximation optimisation problem for coefficients $\vec{\alpha}$ is discussed in Section 6-3. For more information on second-order Taylor approximation, the reader is referred to Chapter B.

6-3   Solving Intensity Approximation for Modal Coefficients

As shown in Section 6-1 and Section 6-2, the non-linear and non-convex intensity function generated by a (modally represented) wavefront can be approximated through the use of Taylor approximation in each point of the generated intensity pattern. This section will demonstrate how this approximation can be solved for the modal coefficients $\vec{\alpha}$ if a captured intensity pattern is given. Only the solution to first-order Taylor intensity approximation will be discussed in this section, information on second-order approximation can be found in Chapter B.

Assuming an intensity image with a total of N^2 pixels, the captured image can be related to the first-order Taylor approximation through Equation 6-11 where η_i is a noise term in the i'th pixel. As this equation is linear in coefficients $\vec{\alpha}$, the minimum variance unbiased estimator is given by the least-squares solution[50].

$$I(x_i, j_i) = h(x_i, y_j) + \eta_{i,j} \approx h(x_i, j_i, \vec{\alpha} = 0) + \left(\frac{\partial h}{\partial \vec{\alpha}}\right)^T (x_i, j_i, \vec{\alpha} = 0)\vec{\alpha} + \eta_i \qquad i = 1, \dots N^2 \quad (6\text{-}11)$$

To simplify notation, the measurement vector $\vec{y}$ can be defined as the vectorised representation of the intensity image I as given in Equation 6-12. As can be seen in this equation, the vector is normalised with respect to the total measured image intensity, resulting in vector $\bar{y}$. This normalisation is done to match the theoretical intensity levels and the numerically stored intensity which is usually expressed in eight-bit integer values in the 0-255 range.

$$
\begin{cases}
\vec{y} & = \mathrm{vec}\,(I) = \begin{bmatrix} I(x_1, y_1) \\ \vdots \\ I(x_N, y_N) \end{bmatrix} \\
\bar{\bar{y}} & = \frac{1}{\sum I_{i,j}} \vec{y}
\end{cases}
\tag{6-12}
$$

Similarly, the predicted zero-aberration intensity in each pixel can be stacked into vector $\vec{b}_{fo}$ and normalised with respect to the predicted total intensity to $\bar{\vec{b}}_{fo}$. The assumption is made that aberrations only re-distribute intensity and as such do not change the total intensity, making the normalisation of measurement and prediction compatible. Like the measurement vector and predicted zero-aberration intensity, the pixel-wise Jacobians can also be stacked and normalised as has been done in Equation 6-14.

$$
\begin{cases}
\vec{b}_{fo} & = \begin{bmatrix} h(x_1, y_1, \vec{\alpha} = 0) \\ \vdots \\ h(x_N, y_N, \vec{\alpha} = 0) \end{bmatrix} \\
\bar{\vec{b}}_{fo} & = \frac{1}{\sum b_{fo,i}} \vec{b}_{fo}
\end{cases}
\tag{6-13}
$$

$$
\begin{cases}
A_{fo} & = \begin{bmatrix} \left(\frac{\partial h}{\partial \vec{\alpha}}(x_1, y_1, \vec{\alpha} = 0)\right)^T \\ \vdots \\ \left(\frac{\partial h}{\partial \vec{\alpha}}(x_N, y_N, \vec{\alpha} = 0)\right)^T \end{bmatrix} = \begin{bmatrix} \frac{\partial h}{\partial \alpha_1}(x_1, y_1, \vec{\alpha} = 0) & \cdots & \frac{\partial h}{\partial \alpha_N}(x_1, y_1, \vec{\alpha} = 0) \\ \vdots & \ddots & \vdots \\ \frac{\partial h}{\partial \alpha_1}(x_N, y_N, \vec{\alpha} = 0) & \cdots & \frac{\partial h}{\partial \alpha_N}(x_N, y_N, \vec{\alpha} = 0) \end{bmatrix} \\
\bar{A}_{fo} & = \frac{1}{\sum b_{fo,i}} A_{fo}
\end{cases}
\tag{6-14}
$$

With the normalised vectors $\bar{\bar{y}}$ and $\bar{\vec{b}}_{fo}$ and the normalised stacked Jacobian matrix $\bar{A}_{fo}$ defined, the best estimate for the modal coefficients $\hat{\vec{\alpha}}$ is given by the normal equation in Equation 6-15. As the full intensity image is used in this method, the number of retrievable modes is much higher than low-order methods such as centroiding or phase extraction through Fourier demodulation of Shack-Hartmann patterns. While technically the number of modes could be as high as N^2 before the system of equations becomes under-determined, this also includes pixels outside of the aperture or pixels with little information and a lower number of modes should be used. The exact amount is not discussed in literature[11], but was set to around 400-1000 in this thesis for $N = 220$.

$$
\hat{\vec{\alpha}} = (\bar{A}_{fo}^T \bar{A}_{fo})^{-1} \bar{A}_{fo}^T \left(\bar{\bar{y}} - \bar{\vec{b}}_{fo} \right)
\tag{6-15}
$$

If multiple measurements would be taken, the vectors and matrices can simply be stacked vertically before applying the normal equation. It is worth noting that this solution method applies to intensity patterns with and without diversities, with the former suffering from possible ambiguity.

6-4   Conclusion

This chapter discussed the method of modal phase reconstruction through local first-order Taylor approximation of the intensity function. After providing the general framework for Taylor approximation with and without a diversity for resolving phase ambiguity caused by the non-uniqueness of Fourier power spectra, the introduction of Shack-Hartmann diversity was discussed. By introducing Shack-Hartmann diversity, the PSF is spread out spatially over the detector, reducing local effects of noise. In addition, information redundancy is obtained as the wavefront is copied and filtered to each Shack-Hartmann centroid location.

In its current form, the aberration magnitude is still limited to small aberrations unless iterative re-linearisation is performed[11], however this is an long and computationally expensive procedure to perform if the aberration is large in magnitude. Instead, this form serves a stepping-stone for the hybrid method presented in Chapter 7.

While reconstruction accuracy could already be discussed and compared to Chapter 5 in this section, a larger comparison between all three methods will instead be performed in Section 7-3.

Novel Hybrid Modal Phase Retrieval Method using Low-Order Pre-Conditioning of Taylor Approximated Intensity Function Fitting on Rectangular and Hexagonal Shack-Hartmann Intensity Patterns

This chapter will discuss a novel hybrid phase retrieval method combining gradient-based low-order phase approximation from Shack-Hartmann patterns with Taylor intensity approximation for a high-accuracy reconstruction on a single Shack-Hartmann frame. Firstly, the method will be discussed in Section 7-1, followed by potential use-cases in Section 7-2. Phase retrieval results will then be provided in Section 7-3 and compared to the individual performance of the low-order and intensity-based methods.

7-1   Introducing Pre-conditioning to Taylor Intensity Approximation

The potential use of pre-conditioning for Taylor intensity approximation as derived in Chapter 6 was a notion already hinted at in the original paper of Marinica et al. on which the method in this thesis is partly based[11]. In this paper, the pre-conditioning would be obtained from previous estimates of the wavefront thereby requiring a re-computation of the intensity function Jacobian (and Hessian if applicable) at every iteration step. However, Taylor approximation is only valid near the

approximation point, and the lack of assumed prior information requires the first iteration to have a zero aberration approximation point. As such, convergence to aberrations with high magnitude components would require a large number of steps, incur a large computational cost for retrieving a single phase, and could get even get stuck in local minima.

The aim of this chapter, on the other hand, is to show that merging the low-order reconstruction methodology of Chapter 5 with the Taylor intensity approximation using Shack-Hartmann diversity of Chapter 6 results in a novel hybrid method capable of reconstructing aberrations comprised of high-magnitude low-order components and low-magnitude high-order components with high-accuracy using only two consecutive reconstruction steps on a single Shack-Hartmann intensity pattern.

As mentioned above, the hybrid method is a combination of the low-order reconstruction method discussed in Chapter 5 and the Taylor intensity approximation methodology of Chapter 6. The way in which these are combined is discussed below.

In the interest of reducing computational complexity, a first-order Taylor approximation of the intensity function is used given in Equation 7-1.

$$h(x,y) \approx h(x,y)\mid_{\vec{\alpha}=0} + \left(\left. \frac{\partial}{\partial \vec{\alpha}} h(x,y) \right|_{\vec{\alpha}=0} \right)^T \vec{\alpha} + \mathcal{O}(\vec{\alpha}^2) \tag{7-1}$$

The assumption will now be made that the phase seen by the detector, ϕ_{obs}, can be written as $\phi_{obs} = \phi + \phi_{SH} = \phi_{taylor} + \phi_{LO} + \phi_{SH}$ where ϕ_{SH} is the Shack-Hartmann diversity of Section 4-2 and ϕ_{LO} is a low-order estimate of the phase. If ϕ_{LO} and ϕ_{SH} are known and ϕ_{taylor} is assumed to be modally represented, this means the wavefront can be modelled using Equation 7-2. This results in the intensity function being defined by Equation 7-3 where the term $P_{nc}(x,y)$ encompasses all information which is independent of the unknown modal coefficients. The phase estimate ϕ_{LO} can theoretically come from any low-order wavefront sensor phase gradient sampling method, but the work in this thesis will assume the modal reconstruction technique based on gradient fields from augmented Fourier demodulation is used discussed in Chapter 5.

$$W_{ts}(x,y) = \underbrace{P(x,y)e^{i\phi_{SH}(x,y)}e^{i\phi_{LO}(x,y)}}_{P_{nc}(x,y)} e^{i \sum_{k=1}^{K} \alpha_k Z_k(x,y)} = P_{nc}(x,y)e^{i \sum_{k=1}^{K} \alpha_k Z_k(x,y)} \tag{7-2}$$

$$\begin{aligned} h(x,y) &= \mathcal{F}\{W_{ts}\}\left(\mathcal{F}\{W_{ts}\}\right)^* \\ &= \mathcal{F}\left\{ P_{nc}(x,y)e^{i \sum_{k=1}^{K} \alpha_k Z_k(x,y)} \right\} \left(\mathcal{F}\left\{ P_{nc}(x,y)e^{i \sum_{k=1}^{K} \alpha_k Z_k(x,y)} \right\} \right)^* \end{aligned} \tag{7-3}$$

Similarly to the inclusion of phase diversity to Taylor approximation in Section 6-2, the Jacobian and approximation point intensity for Taylor approximation with pre-conditioning and phase diversity can be obtained by straightforward substitution of $P_{nc}(x,y)$ in place of $P(x,y)$ in the procedure of Section 6-1. This results in Equation 7-4.

$$\begin{cases} h(x,y)(\vec{\alpha}=0) &= \left| \underbrace{\mathcal{F}\{P_{nc}\}}_{a} \right|^2 \\ \frac{\partial h}{\partial \alpha_j}(\vec{\alpha}=0) &= 2\mathfrak{Re}\left\{ \underbrace{(\mathcal{F}\{P_{nc}\})^*}_{a^*} \mathcal{F}\left\{ \underbrace{iP_{nc}Z_j}_{b} \right\} \right\} \end{cases} \tag{7-4}$$

Following an identical procedure for solving the resulting least-squares problem as previously discussed in Section 6-3 with the definitions in Equation 7-4 results in an estimate of the modal coefficients $\hat{\vec{\alpha}}_{taylor}$. It is important to note, however, that these coefficients capture the phase deviation from the linearisation point. As such, the true full phase coefficients of the two-step method, denoted $\hat{\vec{\alpha}}_{ts}$, is given by Equation 7-5. In this equation, the low-order estimate of the phase ϕ_{LO} is assumed to be modally decomposed in the same basis used for the Taylor approximation in order for summation of coefficients to be valid. In the context of this thesis these would both be expressed in the Zernike basis, but this is not necessarily required and instead a mixed-basis expression of the wavefront phase or a zonal representation of the final retrieved phase could be used.

$$
\begin{cases}
\hat{\vec{\alpha}}_{ts} = \begin{bmatrix} \hat{\vec{\alpha}}_{LO} \\ \vec{0} \end{bmatrix} + \hat{\vec{\alpha}}_{taylor} & , \dim(\hat{\vec{\alpha}}_{LO}) < \dim(\hat{\vec{\alpha}}_{taylor}) \\
\hat{\vec{\alpha}}_{ts} = \hat{\vec{\alpha}}_{LO} + \hat{\vec{\alpha}}_{taylor} & , \dim(\hat{\vec{\alpha}}_{LO}) = \dim(\hat{\vec{\alpha}}_{taylor})
\end{cases}
\tag{7-5}
$$

An additional remark can be made: while the method described above is a two-step method, it was found during testing that an initial refinement of the low-order estimate could yield slightly better results at the cost of a single extra reconstruction step. What this means in practice is that two Taylor approximation steps are done, the first of which estimates the same number of coefficients as the low-order method, while the second estimates a larger number of coefficients with the updated pre-conditioning phase. The observed accuracy gain from this additional pre-refinement is small, in the order of 10^{-2} rad, and as such may not be worth the extra computational cost. However, for illustration purposes this extra refinement will be the method used in Section 7-3.

It is important to note that, while powerful, this hybrid method is blind to specific aberrations. As the low-order estimate is constructed using the average phase gradient over each sub-aperture, aberrations with zero average gradient such as sinusoids with periods equal to multiples of the lenslet pitch or waffle modes are invisible to average gradient-based modal reconstruction while they are visible to intensity methods. As such, if these components have a large magnitude the hybrid method will not perform well.

7-2   Proposed Use-cases

The hybrid method discussed in Section 7-1 is primarily designed for aberrations outside of the magnitude range of Taylor intensity approximation, and out of the modal of classic Shack-Hartmann wavefront sensor reconstruction methods, i.e. high order aberrations with dominating low-order components. This aberration description is readily applicable to real-world aberrations, namely in the form of turbulence[21].

As real-time reconstruction for use in turbulence correction is a challenging endeavour, many methods employ the use of turbulence modelling to reduce the computational cost and speed up the controller[51]. A common model for capturing the behaviour of turbulence is a vector auto-regressive model as given in Equation 7-6, where the capture of a single turbulence layer is done with an order $r = 1$ model. These models are based on the Taylor frozen-flow assumption stating that turbulence shape varies slow over the captured field of view and as such can be assumed to have a constant profile, but shifts spatially due to wind[51]. The system matrix A and movement noise co-variance matrix Q are defined by Equation 7-7 and Equation 7-8 respectively where $C_{\phi,1}$ is given by Equation 7-9.

$$\vec{\phi}(k+1) = \sum_{i=0}^{r-1} A_i \vec{\phi}(k-i) + \vec{w}(k), \qquad \vec{w} \sim \mathcal{N}(0, Q) \tag{7-6}$$

$$A = C_{\phi,1} C_\phi^{-1} \tag{7-7}$$

$$Q = C_\phi - A C_\phi A^T \tag{7-8}$$

$$C_{\phi,1} = E\left[\vec{\phi}(k)\vec{\phi}^T(k-1)\right] \tag{7-9}$$

As can be seen from these equations, there is a dependence on the turbulence wavefront covariance matrix C_ϕ. This matrix is often obtained using using data-driven system identification, for which a naive equation is provided in Equation 7-10[21]. This data-driven approach requires accurate phase retrieval on open-loop phase data, for which the hybrid method discussed in this chapter is uniquely suited.

$$
\begin{aligned}
C_\phi &= \frac{1}{K} \sum_{i=1}^{K} \vec{\phi}(i)\vec{\phi}^T(i) \\
C_{\phi,1} &= \frac{1}{K-1} \sum_{i=1}^{K} \vec{\phi}(i+1)\vec{\phi}^T(i)
\end{aligned}
\tag{7-10}
$$

However, turbulence is not the only use-case of this hybrid method. Any system suffering from low-order dominated aberrations with small high-order components can benefit from the application of this method.

7-3   Reconstruction Results and Comparison

This chapter introduced a novel hybrid phase retrieval method based on low-order pre-conditioning of Taylor intensity approximation applied to a Shack-Hartmann intensity pattern.

The base idea behind the approach was circumventing the small aberration limitation of the basic Taylor intensity approximation method as well as the steep computational cost if that method were to be applied iteratively. By applying the Taylor approximation method to Shack-Hartmann patterns, both ambiguity of the retrieved phase is avoided, and access to phase gradient information is provided. As such, the small aberration limitation can be circumvented by using the phase gradient information to construct a low-order estimate to re-linearise the Taylor approximation around and reconstruct aberrations in the vicinity of this estimate.

Because of this, the method is meant to be used with aberrations whose high-order components are small while low-order components can be large in magnitude. However, its general applicability will be studied.

For the simulations in this section, both a rectangular and hexagonal Shack-Hartmann wavefront sensor were simulated. Each has a circular aperture, a pupil diameter of 1.2mm, a focal length of

10mm, a detector pixel grid of 220×220 pixels and assumes incoming light with wavelength 632.8nm.
The rectangular array consists of 8×8 lenslets and the hexagonal of a central lenslet and 3 rings.
The padding factor $c_{pad} = 3$ and $N_{zph} = 20$. These configurations were chosen as the generated
spot patterns remained relatively free of artefacts. The low-order algorithm will be configured to fit
110 coefficients using the multi-gradient modal reconstruction algorithm discussed in Chapter 5, the
intensity approximation method 400 coefficients, and an additional refinement step will be performed
on the low-order estimate prior to estimation of the 400 coefficients. The low-order method will make
use of the $(0, 2)$ and $(2, 0)$ harmonics and its hexagonal equivalent. This is done as these are said to
be highly excitable according to Rouzé et al.[29].

The simulation is implemented in the Python programming language as a custom simulation frame-
work, but it makes use of the HCIpy optical library for wavefront propagation[45].

Four different aberrations will be studied:

1. A low-order and relatively low-magnitude aberration consisting of Zernike orders $(2, 3, 14)$ with
   magnitudes $(0.1, 0.15, 0.2)\lambda$

2. A high-order but low-amplitude aberration consisting of an order 320 Zernike term with magni-
   tude 0.1λ

3. A high-order and high-amplitude aberration consisting of an order 320 Zernike term with mag-
   nitude 0.6λ

4. An aberration with dominating low-order components (orders $(2, 11)$, magnitudes $(0.35, 0.4)\lambda$)
   and low-amplitude high-order components (orders $(260, 320)$, magnitudes $(0.1, 0.15)\lambda$)

Each of these are meant to represent a possible environment of operation where the largest impact
of the hybrid method is expected to occur for the first and last environment. The third type is not
expected to be common as high-order aberrations are usually found in sources such as turbulence,
which is known to be decomposed more like the fourth aberration[21].

As can be seen in Figure 7-1 and Figure 7-2, the hybrid method outperforms both the low-order
approximation and the Taylor approximation method without pre-conditioning with a root-mean-
square error (RMSE) which is an order of magnitude smaller. The reason behind this result can
be explained as follows. While the aberration is not as large as some other aberrations given, it is
near the edge of where Taylor approximation around zero aberration is still valid. Therefore, while
the Taylor approximation gives an estimate close to reality, linearisation error will cause inaccuracy.
The low-order estimate, on the other hand, is interesting because the aberration is supposed to be
fully within the reconstructable regime for the low-order method. However, this is where the effects
of cross-talk surface. As mentioned before, the used reconstruction basis is the Zernike basis, which
while orthonormal in its basis functions (ignoring the effects of truncation for a moment), is far
from orthogonal in its gradients. As such, mode cross-talk is unavoidable when this basis used for
reconstruction using gradient fields. Because mode cross-talk surfaces mostly around the fringes, its
effects were deemed inconsequential in Chapter 5. The main reason for this decision was the fact that
the method was meant to be used in pre-conditioning of the Taylor intensity approximation method,
which does in fact use the orthogonal basis functions themselves instead of their gradients and could be
expected to undo some of the damage done by cross-talk. This behaviour is exactly what is observed
in Figure 7-1d and Figure 7-2d where the low-order estimate now places the Taylor approximation
region of validity right around the already accurate phase estimate.

The reconstruction results for the second aberration can be seen in Figure 7-3 and Figure 7-4. In
these figures, the high-order aberration is clearly outside the modal range of the Shack-Hartmann

wavefront sensor and is aliased to lower modes for the low-order retrieval algorithm. While the RMSE is low, visually the phase can be considered to be largely inaccurate, though it may still be suited for control purposes. As the aberration is small, it is entirely within the range of applicability of the Taylor intensity approximation without pre-conditioning. As such, the difference between the two-step hybrid method and the intensity-only method is negligible and the low-order pre-conditioning can even be detrimental to the accuracy. This behaviour was expected.

The third aberration is predicted to have the worst results. These can be observed in Figure 7-5 and Figure 7-6. As the aberration is high-order, it is aliased to low-order modes in the low-order reconstruction. In addition, the high amplitude means it is outside the region of validity of the intensity-only method meaning this too has low reconstruction accuracy. The hybrid method is able to obtain a better estimate in the inner 95% of the aperture but otherwise also has a large RMSE error of between 0.64 and 0.72 radians which can be considered bad with the hexagonal results faring even worse.

The last aberration is the closest to the operational environments of interest such as turbulence. As expected, the low-order retrieval algorithm is capable of capturing the overall behaviour of the aberration as it is low-order dominated. However, the fine detail is lost. As the low-order terms are such high amplitude, the aberration falls outside of the range of validity of the intensity-only method and as such the reconstruction is not faithful to the ground truth with RMSE values over 1.2 radians. But the hybrid method is capable of restoring the high-order details lost during low-order approximation quite accurately. Most of the inaccuracy is located at the fringes as becomes clear from the RMSE values, but this may be a consequence of the RMSE calculations being performed on the wrapped output as the aberration looks visually nearly identical to the ground truth and the RMSE value in the inner 95% of the aperture is 1-2 order of magnitude smaller than the two individual reconstruction methods for both the rectangular and hexagonal arrays.

As has evident from this section, the hybrid method is able to outperform both or perform equally well as one of its constituent low-order and intensity-only reconstruction algorithms in all but the most unlikely operational environment: high-order high-amplitude aberrations. Its main downside is computational time. On an Intel Core I7-7700HQ CPU at 2.81 GHz, low-order reconstruction took on average 0.6s, intensity-only 1.65s, and the hybrid method including refinement of the pre-conditioning phase on average 2.6s. While the code is largely unoptimised for speed and faster reconstruction can definitely be achieved, the general ratio should remain the same making the hybrid method, understandably, the slowest method.

In addition, these results show the effectiveness of the novel hexagonal Fourier demodulation method when combined with Taylor approximation, undoing the inaccuracy caused by gradient smoothing discussed in Chapter 5.

While further simulation studies are recommended and alternative hybrid implementations can be considered using Shack-Hartmann diversity, the hybrid method developed in this chapter has been shown to perform admirably.

(a) Wrapped phase profile of a Low-order phase aberration to be re-constructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step

Figure 7-1: Phase retrieval performed on a low-order phase aberration consisting of Zernike orders $(2, 3, 14)$ with magnitudes $(0.1, 0.15, 0.2)\lambda$ captured by an 8×8 rectangular Shack-Hartmann wavefront sensor

(a) Wrapped phase profile of a Low-order phase aberration to be re-constructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step
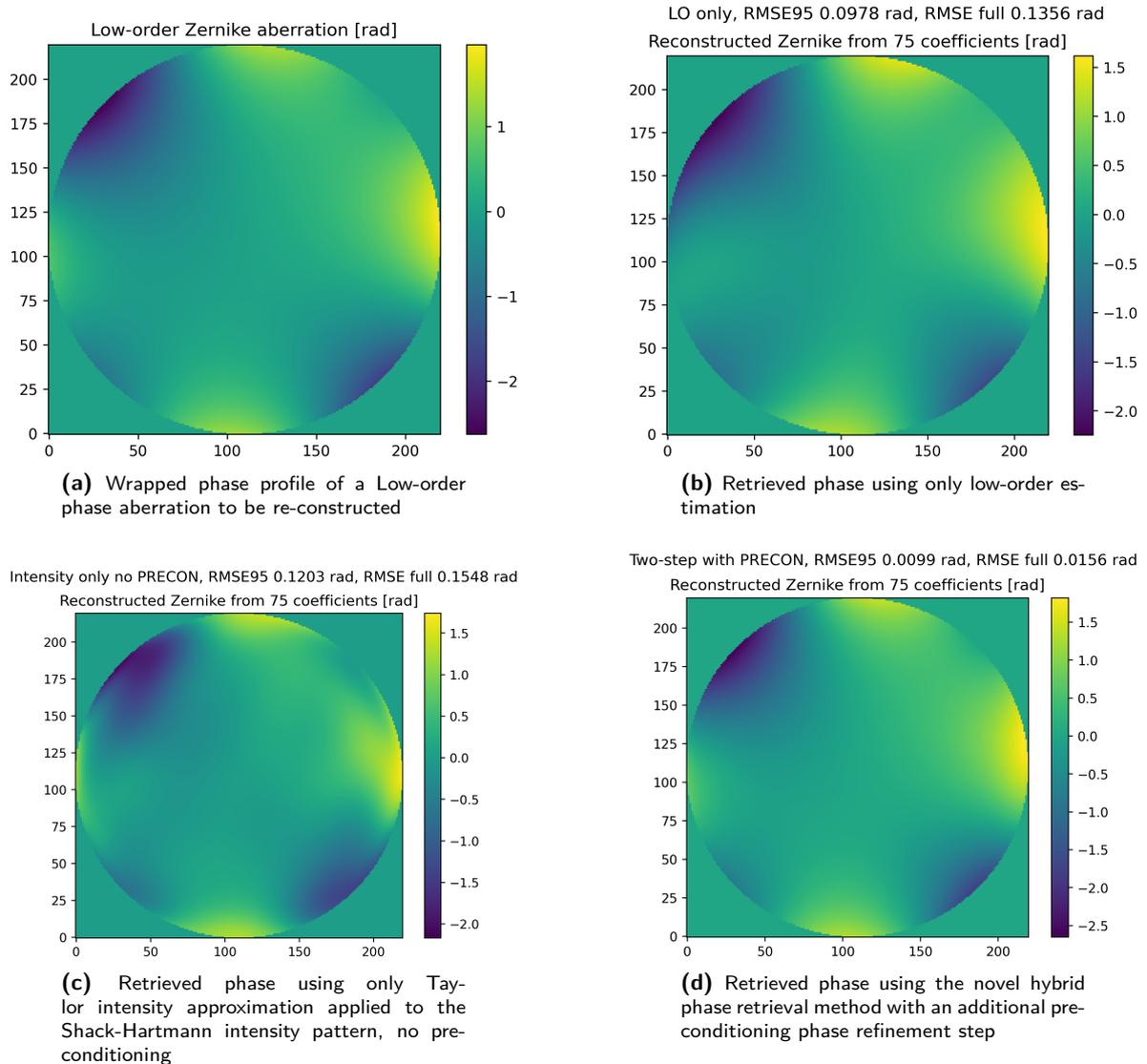
Figure 7-2: Phase retrieval performed on a low-order phase aberration consisting of Zernike orders $(2, 3, 14)$ with magnitudes $(0.1, 0.15, 0.2)\lambda$ captured by a hexagonal Shack-Hartmann wavefront sensor with 3 lenslet rings

Zernike order 320 with small 0.1λ magnitude [rad]

LO only, RMSE95 0.0889 rad, RMSE full 0.0978 rad
Reconstructed Zernike from 110 coefficients [rad]

(a) Wrapped phase profile of a High-order low-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

Intensity only no PRECON, RMSE95 0.0257 rad, RMSE full 0.0525 rad
Reconstructed Zernike from 400 coefficients [rad]

Two-step with PRECON, RMSE95 0.0256 rad, RMSE full 0.0524 rad
Reconstructed Zernike from 400 coefficients [rad]

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step

Figure 7-3: Phase retrieval performed on a high-order (Zernike order 320) low-amplitude (0.1λ) phase aberration captured by an 8×8 rectangular Shack-Hartmann wavefront sensor

62

7. Novel Hybrid Modal Phase Retrieval Method using Low-Order Pre-Conditioning of Taylor Approximated Intensity Function Fitting on Rectangular and Hexagonal Shack-Hartmann Intensity Patterns

(a) Wrapped phase profile of a High-order low-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step

Figure 7-4: Phase retrieval performed on a high-order (Zernike order 320) low-amplitude (0.6λ) phase aberration captured by a hexagonal Shack-Hartmann wavefront sensor with 3 lenslet rings

(a) Wrapped phase profile of a High-order high-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no preconditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional preconditioning phase refinement step

Figure 7-5: Phase retrieval performed on a high-order (Zernike order 320) high-amplitude (0.6λ) phase aberration captured by an 8×8 rectangular Shack-Hartmann wavefront sensor

(a) Wrapped phase profile of a High-order high-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step
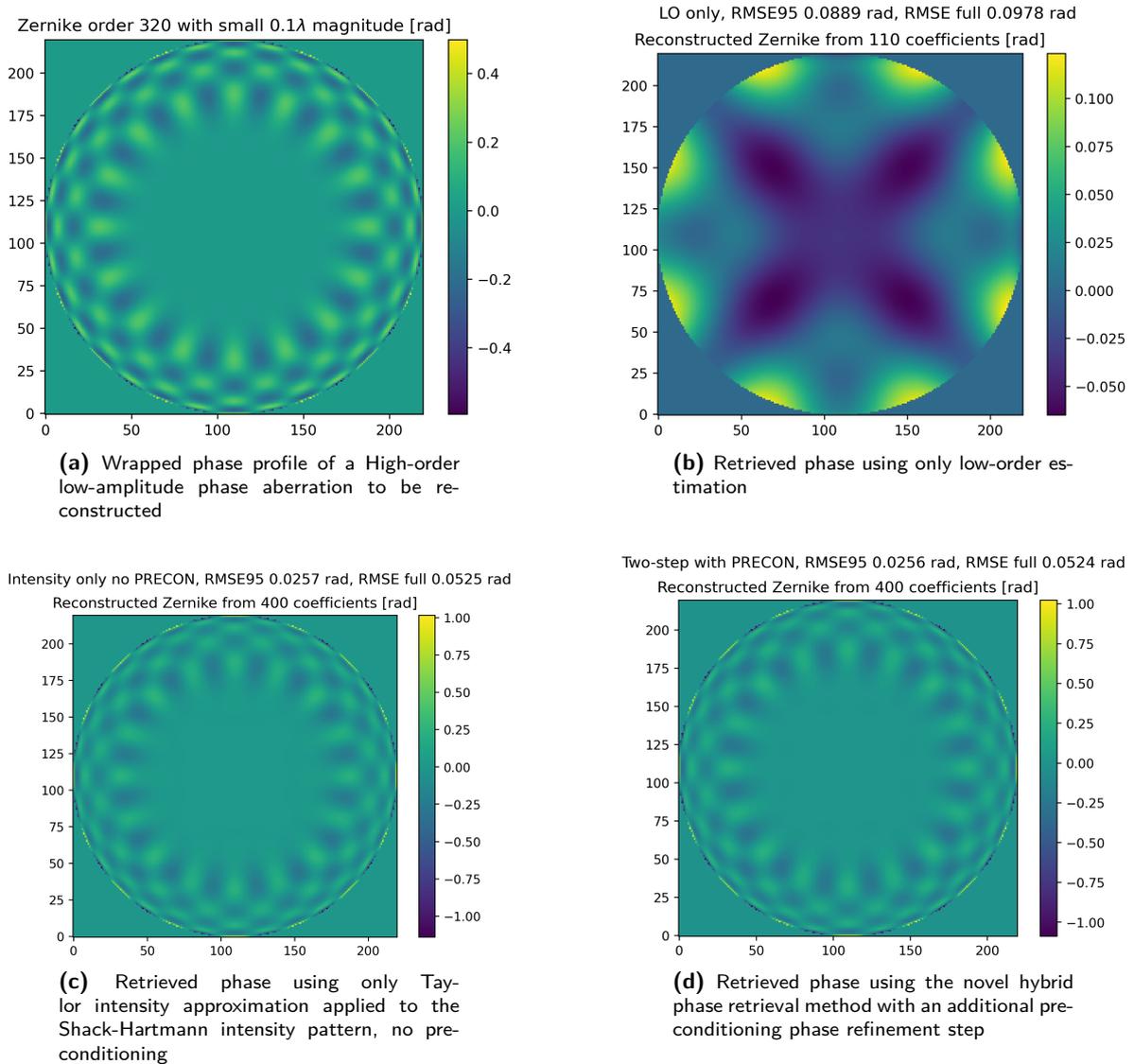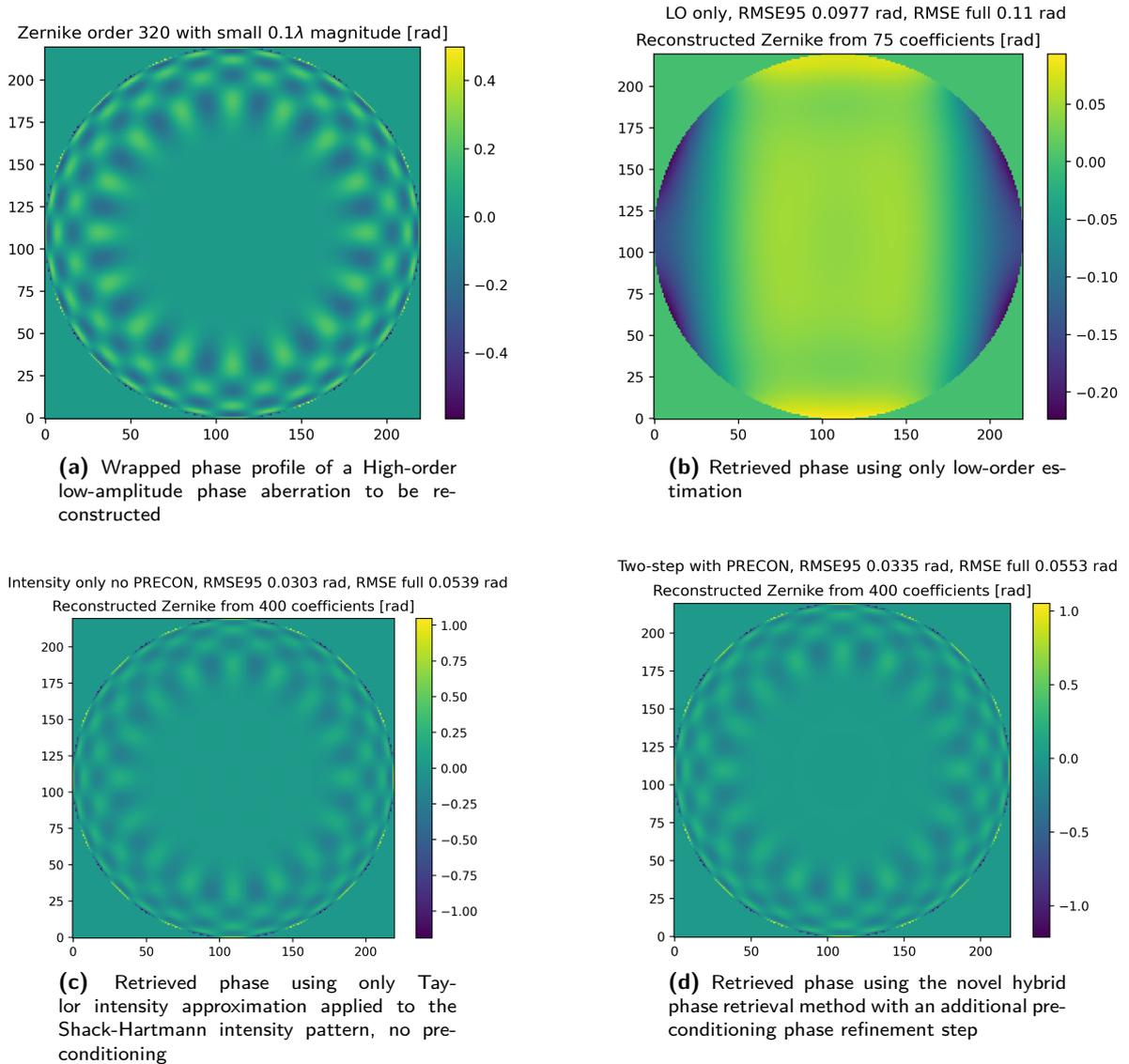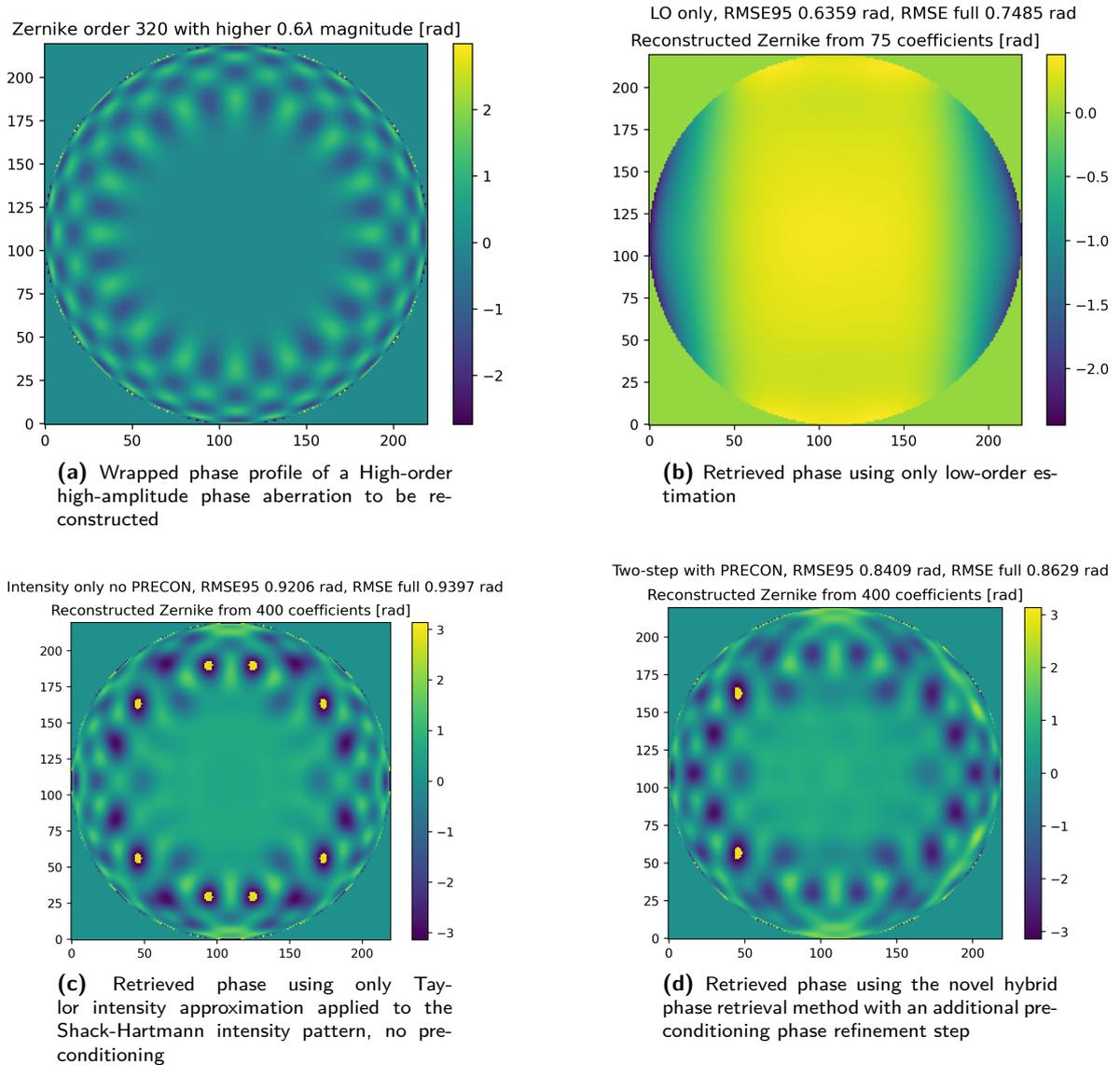
Figure 7-6: Phase retrieval performed on a high-order (Zernike order 320) high-amplitude (0.6λ) phase aberration captured by a hexagonal Shack-Hartmann wavefront sensor with 3 lenslet rings
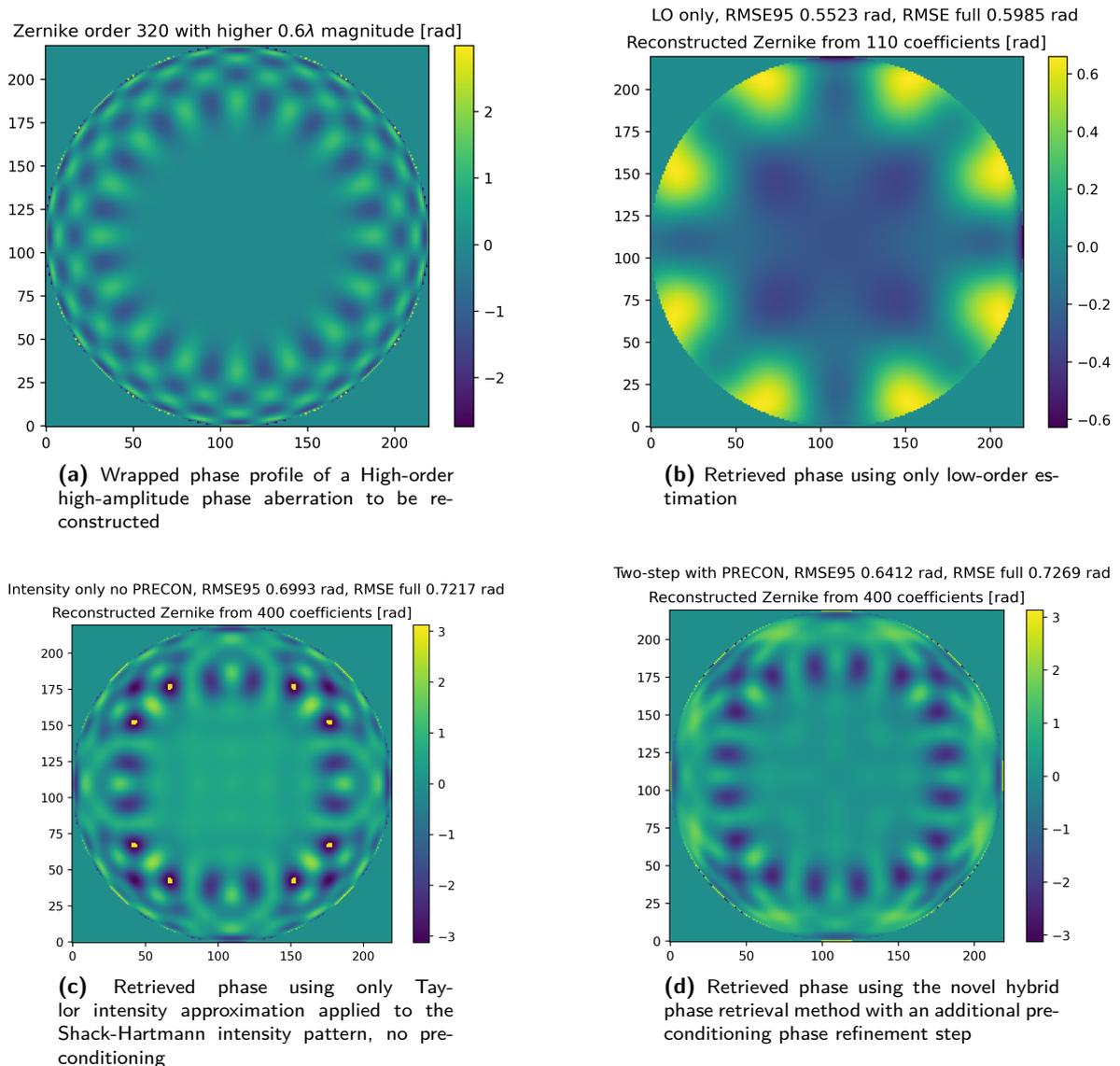
(a) Wrapped phase profile of a High-order high-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step

Figure 7-7: Phase retrieval performed on a phase aberration consisting of dominating low-order components (orders $(2, 11)$, magnitudes $(0.35, 0.4)\lambda$) and low-amplitude high-order components (orders $(260, 320)$, magnitudes $(0.1, 0.15)\lambda$) captured by an 8×8 rectangular Shack-Hartmann wavefront sensor

(a) Wrapped phase profile of a High-order high-amplitude phase aberration to be reconstructed

(b) Retrieved phase using only low-order estimation

(c) Retrieved phase using only Taylor intensity approximation applied to the Shack-Hartmann intensity pattern, no pre-conditioning

(d) Retrieved phase using the novel hybrid phase retrieval method with an additional pre-conditioning phase refinement step
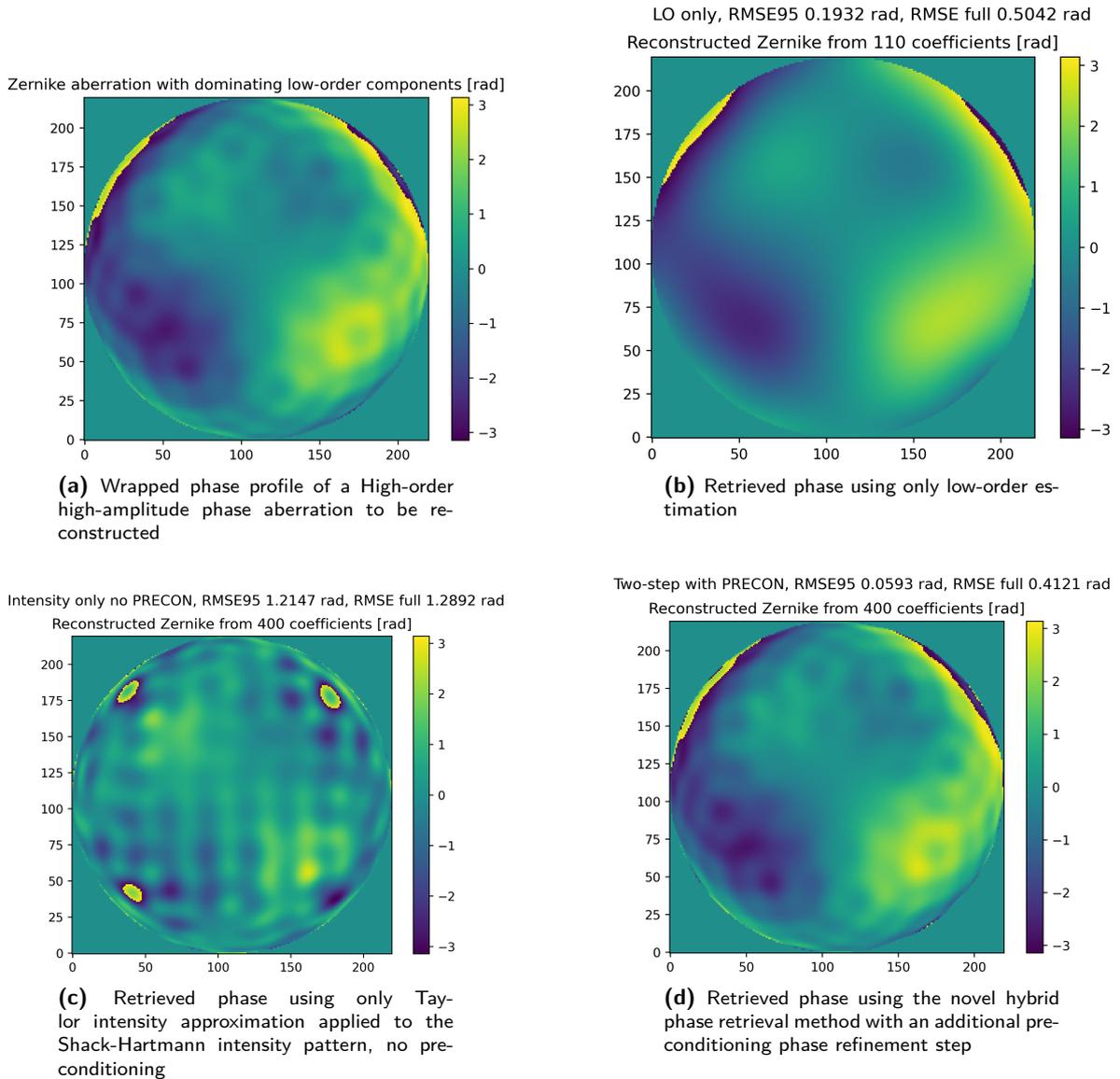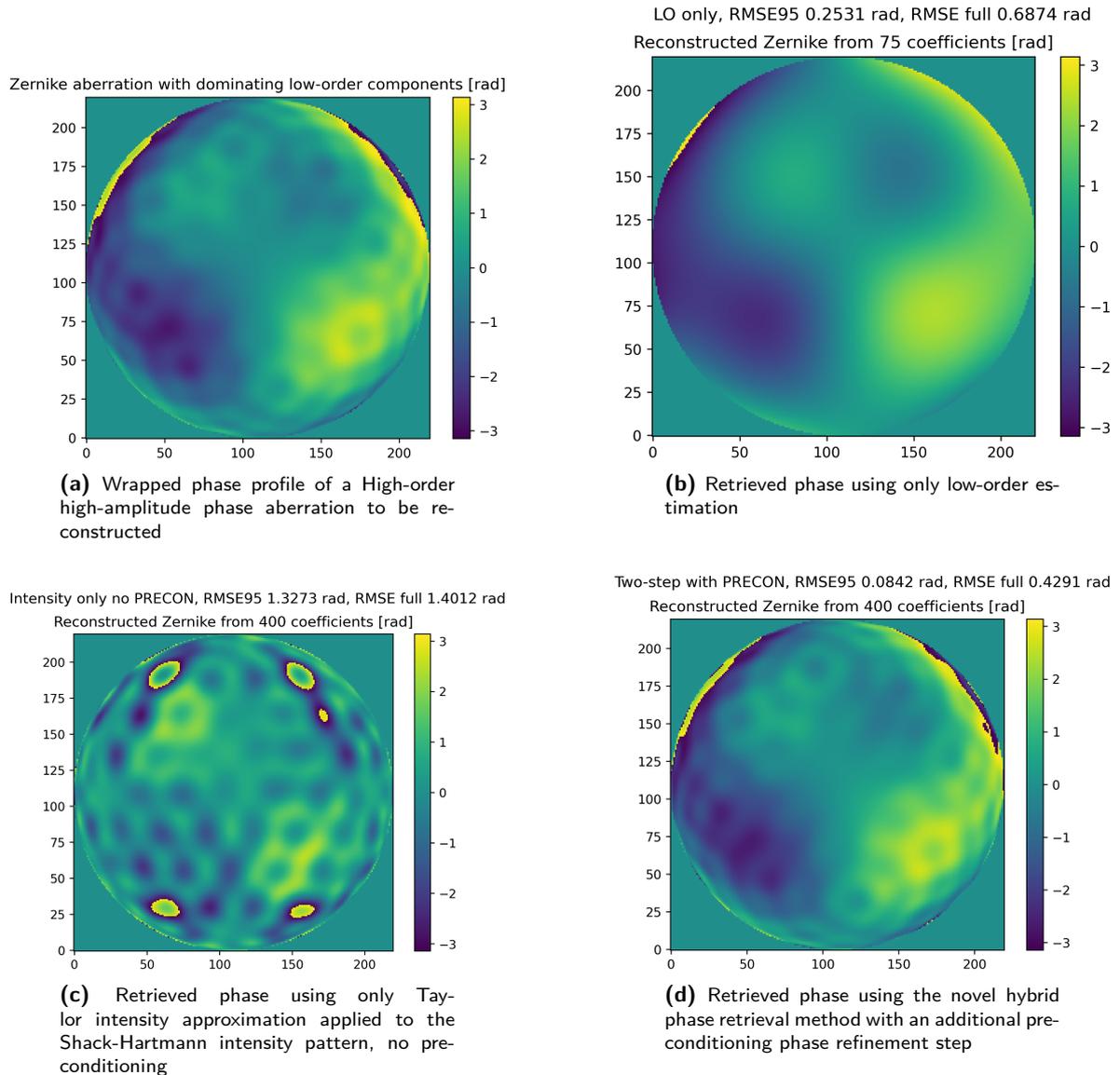
Figure 7-8: Phase retrieval performed on a phase aberration consisting of dominating low-order components (orders $(2, 11)$, magnitudes $(0.35, 0.4)\lambda$) and low-amplitude high-order components (orders $(260, 320)$, magnitudes $(0.1, 0.15)\lambda$) captured by a hexagonal Shack-Hartmann wavefront sensor with 3 lenslet rings

Chapter 8

Conclusion and Outlook

This chapter will present the general conclusions drawn from the thesis in Section 8-1. Additionally, avenues of exploration for future research will be provided in Section 8-2.

8-1   Conclusion

The work in this thesis was focused on the derivation of high-accuracy phase retrieval algorithms applied to Shack-Hartmann intensity patterns. To achieve this, novel additions to low-order phase retrieval methods were made, the application of wavefront sensorless phase retrieval methods to Shack-Hartmann intensity patterns was achieved for intensity approximation methods, and a novel hybrid method was presented for high-accuracy phase retrieval using a single open-loop Shack-Hartmann intensity frame.

For low-order phase retrieval, the Fourier demodulation gradient extraction method was extended to intensity patterns generated using hexagonal Shack-Hartmann wavefront sensors. In addition, modal reconstruction was achieved based on multi-directional derivative fields obtained from trimmed spectral harmonics. This reconstruction algorithm was shown to outperform centroiding accuracy in the case of rectangular arrays, or perform similarly to centroiding in the case of hexagonal arrays. However, a larger aberration magnitude is supported than centroiding.

For modal phase retrieval of small-magnitude aberrations of low- and high-order, first-order Taylor intensity approximation was adapted to be applicable to Shack-Hartmann intensity patterns through the use of Shack-Hartmann diversity. Through this diversity, unambiguous phase reconstruction was achieved on single Shack-Hartmann image frames.

Lastly, the Taylor approximation method was combined with low-order phase retrieval to obtain a novel hybrid method capable of retrieving aberrations consisting of dominating low-order components and smaller high-order components with high accuracy from a single Shack-Hartmann intensity frame. This method extracts not only phase gradient information encoded in Shack-Hartmann patterns, but also the interference information embedded in the point-spread function (PSF). A suggested use-case was the identification of turbulence models in open-loop conditions, high order aberrations dominated

by low-order components. Simulation experiments showed the hybrid method capable of outperforming both of its constituent parts. The main down-side of this hybrid method is the higher computational time as its execution time is tied to the sum of its two constituent methods.

All the above mentioned methods were implemented in a custom Python simulation and phase retrieval framework, making use of the HCIpy library for optical propagation[45].

8-2 Outlook

For future research, the prime candidate for improvement is the methodology presented in Chapter 5. Firstly, the hexagonal Fourier demodulation suffers from cancelled high-frequency components in the extracted harmonics. Because of this, some gradient information is smoothed out, and lost. Identifying ways to avoid this destruction of data this would increase the accuracy of the gradients that can be obtained using this method. Secondly, the trimming of the spectrum brings with it a risk of under-sampling when wrapping of the phase gradient occurs in-between sub-sampling points, after which phase unwrapping algorithms will alias the resulting gradient. As multiple harmonics can be extracted, each containing information about the way the gradient evolves, a multi-grid method could potentially be developed to avoid this issue. Thirdly, the use of reconstruction bases with orthogonal gradients and conservation of orthogonality under truncation would benefit the accuracy of the reconstructed phase. As this low-order method was meant to pre-condition an intensity-based method which does not suffer from non-orthogonality of the gradient, the need for such a basis was less prevalent. However, it could improve overall reconstruction accuracy should it be implemented.

For the intensity approximation method, the implementation of second-order Taylor approximation is an interesting avenue for exploration. Its implementation is briefly discussed in Chapter B.

Lastly, the hybrid method could be refined by performing multiple iterations after the initial reconstruction. While this would no longer result in a two-step algorithm, a trade-off should be made with the potential accuracy improvement.

An overall improvement would be the optimisation of the code base for better phase retrieval performance.

Appendix A

Shack-Hartmann Diversity by Hénault

As mentioned before in Section 4-2, Hénault and Pannetier derived a similar concept for Shack-Hartmann diversity which will be called Hénault diversity in this appendix[44]. While the original work only considered rectangular micro-lens arrays, the approach can be generalised to hexagonal arrays as was done in Equation A-1. To differentiate this Hénault diversity with the Shack-Hartmann diversity used in the continuation of this thesis and to keep with the notation of the original paper, $\Gamma(x, y)$ will be used to denote the phase field.

$$
\begin{aligned}
\Gamma(x, y) &= \frac{2\pi}{\lambda f} \underbrace{\sum_{m=-M}^{M} \sum_{n=-M}^{M} \Pi_{p,p}(x - mp, y - np)}_{\vec{x} \in S_i} \left(\underbrace{mp}_{x_{c,i}}(x - mp) + \underbrace{np}_{y_{c,i}}(y - np) \right) \\
\iff \Gamma_i(x, y) &= \frac{2\pi}{\lambda f} \left(x_{c,i}(x - x_{c,i}) + y_{c,i}(y - y_{c,i}) \right), \qquad \vec{x} \in S_i \\
&= \frac{\pi}{\lambda f} \langle \vec{x}_{c,i}, 2\vec{x} - 2\vec{x}_{c,i} \rangle \\
&= \phi_{SH} - \langle \vec{x}_{c,i}, \vec{x}_{c,i} \rangle
\end{aligned}
\tag{A-1}
$$

As can be observed in Equation A-1, the two diversities differ by a constant piston phase per sub-aperture depending on the centroid position. Due to this difference, adding the Hénault diversity to the ideal lens transmission function phase of Equation 4-3 does not result in the micro-lens array (MLA) transmission function phase given in Equation 4-4 and as such gives a physically incorrect point-spread function. Due to the constant phase delay per sub-aperture, spot positions will not be largely affected and as such the Hénault diversity could be used in centroiding-based phase retrieval methods where interference patterns are unused. However, any method employing interference information in the captured intensity pattern, i.e. the actual point-spread function (PSF) as a whole, will not give correct results when the Hénault diversity is used. As such it cannot be used with any intensity-based phase retrieval method discussed in Chapter 6 and as a result also not the hybrid method developed in this thesis presented in Chapter 7.

Instead, the earlier given Shack-Hartmann diversity presented in Equation 4-5 derived by Soloviev et al. is used for both forward simulation of Shack-Hartmann patterns as well as phase retrieval[13].

Appendix B

Second-order Taylor Intensity Approximation and Solution

The intensity approximation discussed in Chapter 6 only treated first-order Taylor approximation. However, the work of Marinica et al. discusses the use of second-order approximation as well[11]. Continuing on the contents of Section 6-1, the approximation can also be written as Equation B-1.

$$
h(x,y) \approx h(x,y)\left.\right|_{\vec{\alpha}=\vec{\alpha}_L} + \left(\left.\frac{\partial}{\partial\vec{\alpha}}h(x,y)\right|_{\vec{\alpha}=\vec{\alpha}_L}\right)^T \vec{\alpha} + \vec{\alpha}^T \left(\left.\frac{\partial^2}{\partial\vec{\alpha}\partial\vec{\alpha}^T}h(x,y)\right|_{\vec{\alpha}=\vec{\alpha}_L}\right)\vec{\alpha} + \mathcal{O}(\vec{\alpha}^3) \quad \text{(B-1)}
$$

As can be observed, Equation 6-2 also shows a second order derivative, a Hessian matrix, with respect to α coefficients. Each element in the Hessian is readily obtained by deriving Equation 6-6 with respect to coefficient α_k where α_j may equal α_k. As the Hessian is symmetric, this operation only has to be performed for half the coefficient pairs, i.e. elements $(i,j)\ \forall i = 1,\ldots,K$ and $\forall j = 1,\ldots i$. Using standard differentiation rules and setting $\vec{\alpha}_L = 0$, the elements of the Hessian can be obtained from Equation B-2.

$$
\begin{aligned}
\frac{\partial^2 h}{\partial\alpha_j\partial\alpha_k}(\vec{\alpha} = \vec{\alpha}_L = 0) &= \mathcal{F}\left\{iiPZ_jZ_k\right\}\left(\mathcal{F}\left\{P\right\}\right)^* + \mathcal{F}\left\{iPZ_j\right\}\left(\mathcal{F}\left\{iPZ_k\right\}\right) \\
&\quad + \mathcal{F}\left\{iPZ_k\right\}\left(\mathcal{F}\left\{iPZ_j\right\}\right)^* + \mathcal{F}\left\{P\right\}\left(\mathcal{F}\left\{iiPZ_jZ_k\right\}\right)^* \\
&= 2\mathfrak{Re}\left\{\left(\mathcal{F}\left\{P\right\}\right)^*\mathcal{F}\left\{iiPZ_jZ_k\right\}\right\} + 2\mathfrak{Re}\left\{\left(\mathcal{F}\left\{iPZ_j\right\}\right)^*\mathcal{F}\left\{iPZ_k\right\}\right\} \\
&= 2\mathfrak{Re}\left\{\left(\mathcal{F}\left\{P\right\}\right)^*\mathcal{F}\left\{iiPZ_jZ_k\right\} + \left(\mathcal{F}\left\{iPZ_j\right\}\right)^*\mathcal{F}\left\{iPZ_k\right\}\right\}
\end{aligned} \quad \text{(B-2)}
$$

Combining this with the Jacobian of Equation 6-6 and the linearisation point intensity in Equation 6-5 gives Equation B-3 where again similarities are indicated which can be used to reduce the number of computations. As with Section 6-2 and Section 7-1, P can simply substituted with P_d or P_{nc} without further alterations.

$$
\begin{cases}
h(x,y)(\vec{\alpha}=0) & = \left| \underbrace{\mathcal{F}\{P\}}_{a} \right|^2 \\[2ex]
\dfrac{\partial h}{\partial \alpha_j}(\vec{\alpha}=0) & = 2\Re\mathfrak{e}\left\{ \underbrace{(\mathcal{F}\{P\})^*}_{a^*} \mathcal{F}\left\{ \underbrace{iPZ_j}_{b} \right\} \right\} \\[2ex]
\dfrac{\partial^2 h}{\partial \alpha_j \partial \alpha_k}(\vec{\alpha}=0) & = 2\Re\mathfrak{e}\left\{ \underbrace{(\mathcal{F}\{P\})^*}_{a*} \mathcal{F}\left\{ \underbrace{iiPZ_jZ_k}_{ibZ_k} \right\} + \underbrace{(\mathcal{F}\{iPZ_j\})^*}_{b} \mathcal{F}\{iPZ_k\} \right\}
\end{cases}
\tag{B-3}
$$

This problem can then be re-written as presented in Equation B-4 which can be used to cast the problem into a semidefinite programming problem as given in Equation B-5. It is worth noting that, while it is not shown in these equations, normalisation similar to that performed in Section 6-3 must be applied.

$$
\begin{aligned}
h(x,y) = \begin{bmatrix} 1 & \vec{\alpha}^T \end{bmatrix} &\underbrace{\begin{bmatrix} h(x,y,\vec{\alpha}=0) & \frac{1}{2}\left(\frac{\partial h}{\partial \vec{\alpha}}(x,y,\vec{\alpha}=0) \right)^T \\ \frac{1}{2}\frac{\partial h}{\partial \vec{\alpha}}(x,y,\vec{\alpha}=0) & \frac{\partial^2 h}{\partial \vec{\alpha} \partial \vec{\alpha}^T}(x,y,\vec{\alpha}=0) \end{bmatrix}}_{A_H(x,y)} \begin{bmatrix} 1 \\ \vec{\alpha} \end{bmatrix} \\[3ex]
&= Tr\left(A_H(x,y) \underbrace{\begin{bmatrix} 1 \\ \vec{\alpha} \end{bmatrix} \begin{bmatrix} 1 & \vec{\alpha}^T \end{bmatrix}}_{X} \right)
\end{aligned}
\tag{B-4}
$$

$$
\begin{aligned}
&\min_X \left(Tr(X) \right) \\
s.t. \quad &y_i = Tr(A_{H,i}X) \\
&X_{1,1} = 1 \\
&X \succeq 0
\end{aligned}
\tag{B-5}
$$

Due to the lifting of dimensionality, this optimisation problem is computationally very demanding. In fact, using the full image $N \times N$ image will result in $N^2(K^2 + K + 1)$ elements that need to be stored in computer memory where K is the number of coefficients to retrieve. This quickly takes up more than the entire RAM capacity of conventional computer hardware. For this reason, Marinica et al. instead suggests to only apply the second-order approximation to a subset of the brightest pixels as these are more likely to contain usable data.

While the second-order constraint generation as well as the optimisation problem were implemented in Python code for this thesis project, it was ultimately abandoned due to convergence issues with the used numerical solver as well as concerns about the computational and memory cost versus the gained accuracy. However, future research is recommended.

Bibliography

[1] C. Leroux and C. Dainty, "A simple and robust method to extend the dynamic range of an aberrometer," *Optics Express*, vol. 17, no. 21, p. 19055, 10 2009.

[2] R. K. Tyson and B. W. Frazier, *Principles of Adaptive Optics*. Boca Raton: CRC Press, 1 2022.

[3] I. Newton, *Opticks: Or, a treatise of the reflections, refractions, inflections, and colours of light*, 1979.

[4] M. Verhaegen, *Lecture notes on Control for High Resolution Imaging (SC42030)*, 4 2017.

[5] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase Retrieval with Application to Optical Imaging: A contemporary overview," *IEEE Signal Processing Magazine*, vol. 32, no. 3, pp. 87–109, 5 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7078985

[6] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Applied Optics*, vol. 21, no. 15, p. 2758, 8 1982. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ao-21-15-2758

[7] D. Luke, "Phase Retrieval, What's New?" *SIAM SIAG/OPT*, vol. 25, p. 1, 11 2017.

[8] R. Shack and B. Platt, "Production and use of a lenticular Hartmann screen," *J. Opt. Soc. Am.*, vol. 61, pp. 656–660, 4 1971.

[9] A. M. Nightingale and S. Gordeyev, "Shack-Hartmann wavefront sensor image analysis: a comparison of centroiding methods and image-processing techniques," *Optical Engineering*, vol. 52, no. 7, p. 071413, 3 2013.

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

[11] R. Marinica, C. S. Smith, and M. Verhaegen, "State feedback control with quadratic output for wavefront correction in adaptive optics," in *52nd IEEE Conference on Decision and Control*. IEEE, 12 2013, pp. 3475–3480. [Online]. Available: http://ieeexplore.ieee.org/document/6760416/

[12] E. J. Candes, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and Stable Signal Recovery from Magnitude Measurements via Convex Programming," 9 2011. [Online]. Available: http://arxiv.org/abs/1109.4499

[13] O. Soloviev, H. T. Nguyen, V. Bezzubik, N. Belashenkov, G. Vdovin, and M. Verhaegen, "Shack-Hartmann sensor as an imaging system with a phase diversity," *Journal of the Optical Society of America A*.

[14] J. W. Goodman, *Introduction to Fourier Optics*, ser. McGraw-Hill physical and quantum electronics series. W. H. Freeman, 2005. [Online]. Available: https://books.google.be/books?id=ow5xs_Rtt9AC

[15] J. Kalkman and K. S. Grußmayer, *Lecture notes on Advanced Optical Imaging (AP3122)*, 2 2022.

[16] M. Born, E. Wolf, A. B. Bhatia, P. C. Clemmow, D. Gabor, A. R. Stokes, A. M. Taylor, P. A. Wayman, and W. L. Wilcock, *Principles of Optics*. Cambridge University Press, 10 1999.

[17] R. J. Noll, "Zernike polynomials and atmospheric turbulence*," *Journal of the Optical Society of America*, vol. 66, no. 3, p. 207, 3 1976. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josa-66-3-207

[18] O. Soloviev, "Alias-free basis for modal sensorless adaptive optics using the second moment of intensity," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 20, no. 03, 5 2022. [Online]. Available: https://www.worldscientific.com/doi/10.1142/S0219691320400081

[19] C. Zhao and J. H. Burge, "Orthonormal vector polynomials in a unit circle, Part I: basis set derived from gradients of Zernike polynomials," *Optics Express*, vol. 15, no. 26, p. 18014, 2007. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?uri=oe-15-26-18014

[20] ——, "Orthonormal vector polynomials in a unit circle, Part II : completing the basis set," *Optics Express*, vol. 16, no. 9, p. 6586, 4 2008.

[21] P. J. Piscaer, "High-resolution focal-plane wavefront sensing for time-varying aberrations," Ph.D. dissertation, 2021.

[22] T. B. Andersen, "Efficient and robust recurrence relations for the Zernike circle polynomials and their derivatives in Cartesian coordinates," *Optics Express*, vol. 26, no. 15, p. 18878, 7 2018.

[23] J. Hartmann, "Objektivuntersuchungen," *Zeitschrift für Instrumentenkunde*, vol. 24, 1904.

[24] D. Gabor, "A New Microscopic Principle," *Nature*, vol. 161, pp. 777–778, 1948.

[25] D. G. Smith and J. E. Greivenkamp, "Generalized method for sorting Shack-Hartmann spot patterns using local similarity," *Applied Optics*, vol. 47, no. 25, p. 4548, 9 2008. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ao-47-25-4548

[26] W. Zou, K. P. Thompson, and J. P. Rolland, "Differential Shack-Hartmann curvature sensor: local principal curvature measurements," *Journal of the Optical Society of America A*, vol. 25, no. 9, p. 2331, 9 2008.

[27] M. Rais, J.-M. Morel, C. Thiebaut, J.-M. Delvit, and G. Facciolo, "Improving wavefront sensing with a Shack–Hartmann device," *Applied Optics*, vol. 55, no. 28, p. 7836, 10 2016. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ao-55-28-7836

[28] M. Rais, J.-M. Morel, and G. Facciolo, "Iterative Gradient-Based Shift Estimation: To Multiscale or Not to Multiscale?" in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, A. Pardo and J. Kittler, Eds. Cham: Springer International Publishing, 2015, pp. 416–423.

[29] B. Rouzé, G. Giakoumakis, A. Stolidi, J. Jaeck, C. Bellanger, and J. Primot, "Extracting more than two orthogonal derivatives from a Shack-Hartmann wavefront sensor," *Optics Express*, vol. 29, no. 4, p. 5193, 2 2021. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=oe-29-4-5193

[30] Y. Carmon and E. N. Ribak, "Phase retrieval by demodulation of a Hartmann–Shack sensor," *Optics Communications*, vol. 215, no. 4-6, pp. 285–288, 1 2003. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S003040180202254X

[31] A. Talmi and E. N. Ribak, "Direct demodulation of Hartmann–Shack patterns," *Journal of the Optical Society of America A*, vol. 21, no. 4, p. 632, 4 2004. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josaa-21-4-632

[32] O. Glazer, E. N. Ribak, and L. Mirkin, "Adaptive optics implementation with a Fourier reconstructor," *Applied Optics*, vol. 46, no. 4, p. 574, 2 2007. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ao-46-4-574

[33] E. N. Ribak, Y. Carmon, A. Talmi, O. Glazer, O. Srour, and N. Zon, "Full wave front reconstruction in the Fourier domain," 6 2006, p. 627254. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.684930

[34] J. Primot, "Theoretical description of Shack–Hartmann wave-front sensor," *Optics Communications*, vol. 222, no. 1-6, pp. 81–92, 7 2003.

[35] R. A. Gonsalves, "Phase Retrieval And Diversity In Adaptive Optics," *Optical Engineering*, vol. 21, no. 5, 10 1982.

[36] J. Li, Y. Gong, H. Chen, and X. Hu, "Wave-front reconstruction with Hartmann–Shack sensor using a phase-retrieval method," *Optics Communications*, vol. 336, pp. 127–133, 2 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0030401814009249

[37] E. J. Candes, Y. Eldar, T. Strohmer, and V. Voroninski, "Phase Retrieval via Matrix Completion," 2011.

[38] R. Doelman, N. H. Thao, and M. Verhaegen, "Solving large-scale general phase retrieval problems via a sequence of convex relaxations," *Journal of the Optical Society of America A*, vol. 35, no. 8, p. 1410, 8 2018. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=josaa-35-8-1410

[39] J. Silva, E. Brunner, A. Polo, C. de Visser, and M. Verhaegen, "Wavefront reconstruction using intensity measurements for real-time adaptive optics," in *2014 European Control Conference (ECC)*. IEEE, 6 2014, pp. 2412–2417. [Online]. Available: http://ieeexplore.ieee.org/document/6862405/

[40] C. C. de Visser and M. Verhaegen, "Wavefront reconstruction in adaptive optics systems using nonlinear multivariate splines," *Journal of the Optical Society of America A*, vol. 30, no. 1, p. 82, 1 2013.

[41] M. Viegers, E. Brunner, O. Soloviev, C. C. de Visser, and M. Verhaegen, "Nonlinear spline wavefront reconstruction through moment-based Shack-Hartmann sensor measurements," *Optics Express*, vol. 25, no. 10, p. 11514, 5 2017.

[42] H. Yang, O. Soloviev, and M. Verhaegen, "Model-based wavefront sensorless adaptive optics system for large aberrations and extended objects," *Optics Express*, vol. 23, no. 19, p. 24587, 9 2015. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=oe-23-19-24587

[43] Z. Zhang, Y. Luo, H. Yang, H. Su, and J. Liu, "CoolMomentum-SPGD Algorithm for Wavefront Sensor-Less Adaptive Optics Systems," *Photonics*, vol. 10, no. 2, p. 102, 1 2023.

[44] F. B. Hénault and C. Pannetier, "Hartmann vs reverse Hartmann test: a Fourier optics point of view," in *Applied Optical Metrology III*, E. Novak and J. D. Trolinger, Eds. SPIE, 9 2019, p. 11. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11102/2527895/Hartmann-vs-reverse-Hartmann-test--a-Fourier-optics-point/10.1117/12.2527895.full

[45] E. H. Por, S. Y. Haffert, V. M. Radhakrishnan, D. S. Doelman, M. Van Kooten, and S. P. Bos, "High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator," in *Adaptive Optics Systems VI*, ser. Proc. SPIE, vol. 10703, 2018. [Online]. Available: https://doi.org/10.1117/12.2314407

[46] S. Velghe, J. Primot, N. Guérineau, M. Cohen, and B. Wattellier, "Wave-front reconstruction from multidirectional phase derivatives generated by multilateral shearing interferometers," *Optics Letters*, vol. 30, no. 3, p. 245, 2 2005. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ol-30-3-245

[47] R. Legarda-Sáenz, M. Rivera, R. Rodríguez-Vera, and G. Trujillo-Schiaffino, "Robust wave-front estimation from multiple directional derivatives," *Optics Letters*, vol. 25, no. 15, p. 1089, 8 2000. [Online]. Available: https://opg.optica.org/abstract.cfm?URI=ol-25-15-1089

[48] A. S. Jurling and J. R. Fienup, "Applications of algorithmic differentiation to phase retrieval algorithms," *Journal of the Optical Society of America A*, vol. 31, no. 7, p. 1348, 7 2014.

[49] A. Wong, B. Pope, L. Desdoigts, P. Tuthill, B. Norris, and C. Betters, "Phase retrieval and design with automatic differentiation: tutorial," *Journal of the Optical Society of America B*, vol. 38, no. 9, p. 2465, 9 2021.

[50] M. Verhaegen and V. Verdult, *Filtering and system identification.* Cambridge, England: Cambridge University Press, 7 2012.

[51] P. Piscaer, O. Soloviev, and M. Verhaegen, "Predictive wavefront sensorless adaptive optics for time-varying aberrations," *Journal of the Optical Society of America A*, vol. 36, no. 11, p. 1810, 10 2019.

Glossary

List of Acronyms

PSF	point-spread function
PR	phase-retrieval
FFT	fast Fourier transform
MLA	micro-lens array
OTF	optical transfer function
RMSE	root-mean-square error

List of Symbols

λ	Wavelength
$\mathcal{A}$	Wave amplitude function
ω	Frequency coordinate in the x direction
ϕ	Phase
$\vec{\alpha}$	Modal coefficient vector
ξ	Frequency coordinate in the y direction
c_{pad}	Padding factor
d	Physical pixel size
$D_{h,i}$	Hexagon inscribed circle diameter
$D_{h,o}$	Hexagon circumscribed circle diameter
D_{pupil}	Pupil diameter
f	Focal length
$f_{\mu L}$	Lenslet focal length
h	Intensity function
$H_{k,l,m}$	(k,l,m)'th harmonic in the spectrum of a hexagonal Shack-Hartmann intensity pattern
$H_{k,l}$	(k,l)'th harmonic in the spectrum of a rectangular Shack-Hartmann intensity pattern
I	Intensity image
N	Pixel count in a row for an unpadded image
N_{MLA}	Number of micro-lens array centroids within the aperture
N_{pix}	Number of gradient field values within the aperture
N_{zph}	Pixel count in a row for a padded harmonic in Fourier demodulation

N_{zp}	Pixel count in a row for a padded image
P	Pupil function
$p, p_{\mu L}$	Lenslet pitch
P_d	Pupil function including a phase diversity term
P_{nc}	Pupil function including a phase diversity and pre-conditioning phase
W	Wavefront
$w_{f,pix}$	Lenslet pitch frequency in pixels
x	Spatial coordinate in the horizontal direction
y	Spatial coordinate in the vertical direction
Z_j, Z_m^n	Zernike basis function of Noll's order j or radial-azimuthal order (n, m)