WAVEFRONT CONTROL TECHNIQUES FOR THE DIRECT IMAGING OF EXOPLANETS

by

Alexander T. Rodack

A Dissertation Submitted to the Faculty of the

JAMES C. WYANT COLLEGE OF OPTICAL SCIENCES

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2022

2

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by **Alexander T. Rodack**, titled ***Wavefront Control Techniques for the Direct Imaging of Exoplanets*** and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Jared R. Males
Dr. Jared R. Males — Date: 2022-03-14

Daewook Kim
Professor Daewook Kim — Date: 2022-3-14

OGuyon
Professor Olivier Guyon — Date: 2022-03-14

Richard A. Frazin
Dr. Richard A. Frazin — Date: 2022-03-14

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Jared R. Males
Dr. Jared R. Males — Date: 2022-04-29
Dissertation Committee Chair
Wyant College of Optical Sciences

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Alexander T. Rodack

ACKNOWLEDGEMENTS

To my advisers Dr. Jared Males and Dr. Richard Frazin, I extend a sincere and deep appreciation of your continued guidance, patience, support, encouragement, and understanding throughout the past several years. Without your joint pillars to stand upon, I would not only have been unable to complete this work, but would be less prepared for the future endeavors on which I am about to embark. Your advice has helped me to become a better scientist and engineer, and most importantly, a better person. For this I am deeply grateful. I wish you both continued success on your journeys through high contrast imaging and beyond.

I would also like to thank my first mentors in graduate school, Dr. Oliver Guyon and Dr. Johanan Codona. You both were there for my first steps in to high contrast imaging and adaptive optics, and nurtured a growing interest into a flourishing passion. It was on your advice and support that I applied in to the PhD program from the Master's program, so in a literal sense, as well as a figurative one, I would not be here today without you both. I will take this passion and curiosity with me going forward in to whatever challenges come my way, knowing that with the skills I have obtained I can accomplish anything I set my mind to.

Furthermore, I would like to thank my committee, Dr. Males, Dr. Frazin, Dr. Guyon, and Dr. Daewook Kim, as a whole. Your thoughtful comments and suggestions improved the story of my research in this dissertation, and provided much clarity for any future readers of this work. I appreciate the time and effort you each put in helping me to make this best possible reflection of the research I undertook.

To my original graduate school research colleagues, Dr. Kelsey Miller, Dr. Justin Knight, and Dr. Jennifer Lumbres, thank you for all the time and friendship throughout this long, 7 year process. The long discussions, whether about research or life in general, kept me sane, and served as motivation to continue on. The impact you have had on my life is profound, and I look forward to keeping in touch about the different paths we all are now walking. I would also like to thank all of my friends and classmates throughout the years of both undergraduate and graduate school that made surviving coursework possible. James Upton, Neil Momsen, Wente Yin, Jeff Wilhite, Lee Johnson, JJ Katz, Yufeng Yan, Jesse Jensen, Brian Hightower, James Gordon, and so many others; thank you for everything.

To the entire MagAO-X team and CAAO, thank you so much for all the support. You are all wonderful people that made this experience one of the best of my life. Specifically, I would like to name a few of you: Joseph Long, you are a wizard that made everything technological so much easier to handle, including on more than one occasion unbricking a laptop or computer of mine and saving me countless hours of work. Dr. Kyle van Gorkom, I enjoyed our discussions, and especially our journey in Canada at AO4ELT6. Alex Hedglen, I enjoyed not only the technical conversations, but also the ones we had diving in to music and Star Wars. Many if not all of your suggestions were in the playlist that was accompanying my writing of this dissertation. I am proud to have you take up the mantle as the most senior Alex of the team. To Kim Chapman, you did so much to make the administrative part of the journey so very simple and manageable for me. I always enjoyed our chats, even if they were few and far between sometimes. Without you and the rest of the staff in CAAO, I don't know where I would be.

I am also very grateful for the funding support I had throughout my graduate school experience. Without the support from the NSF and the Heising-Simons Foundation, as well as the scholarship I received from SPIE, my research and work would not have been possible.

Finally, I would like to thank my family. There are not words to truly express my gratitude for your love and support, but I will try. To my Aunt Cathy and Uncle Mark, your unwavering belief in me and my ability to reach this point was a cornerstone of my success. It feels great to add another PhD to the family in addition to yours. To my brother, Andy, thank you for always having my back, and for both your words of advice when I have needed them or the simple distraction of talking about sports or video games or life. Finally, to my parents, Jodi and Mike, thank you for setting the example of having curiosity, and always encouraging me to follow it, no matter where it might lead, even 7 years of graduate school getting a PhD. It has meant the world to me.

DEDICATION

To my parents, Jodi and Mike, and brother, Andy, without whose love and support I could never have made it through graduate school; and to my grandparents, Ruth and Wally, who always encouraged me to reach for the stars.

TABLE OF CONTENTS

TABLE OF CONTENTS – *Continued*

TABLE OF CONTENTS – *Continued*

LIST OF FIGURES

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF FIGURES – *Continued*

LIST OF TABLES

ABSTRACT

Over two decades ago, the first planet around a star other than the Sun was discovered. With each passing year, more and more such *exoplanets* are discovered as new technologies and methods of discovery are developed and enhanced. As these techniques continue to mature, humanity gets closer to finally being able to answer the question: are we alone in the universe? Improvements to Adaptive Optics (AO) have enabled ground-based observation to expand to including high-contrast imaging instruments called coronagraphs that are meant to make the direct imaging of exoplanet light possible. Direct imaging is a method of observation that gives astronomers the ability to determine if a planet exhibits signatures of life via spectroscopic analysis for biomarkers. This is a difficult task for three major reasons: the planet orbits very close to its host star if it is located in the so-called habitable zone, the planet light is up to 10^{-10} times fainter than the host star light, and static and quasi-static aberration being present during the observation degrades both coronagraph performance and post-processing technique efficacy.

In this dissertation, I explore two methods for estimating non-common path aberration (NCPA) in the science instrument of AO enabled, ground-based telescopes. The first is a method called the Differential Optical Transfer Function (dOTF), which is a simple, non-iterative, non-interferometric technique to estimate the complex amplitude field in the exit pupil of an optical system exploiting the properties of the functional derivative of the Optical Transfer Function. dOTF is demonstrated in both simulation and lab based experiments, showing several possible applications, including AO system self calibration, segment cophasing, and estimating systematic NCPA using an off-axis light source. The second method is known as Frazin's algorithm, which is a statistical regression framework that uses wavefront sensor (WFS) and science camera (SC) telemetry with advanced computational models of

optical systems to estimate any NCPA and any present exoplanet signals. I develop the history of the method starting from its inception in 2013 and its extension to potential real-time use in 2018, followed by the conception of an improved version that is fully realizable. Three separate estimators are presented within the framework, and then are demonstrated via comprehensive end-to-end simulation of an AO system running at 1kHz frame rate with a Lyot Coronagraph in the science arm. Finally, preliminary future extensions of the work done on Frazin's algorithm are presented to guide future steps to evolve the method to improve the current limits of ground-based direct imaging of exoplanets.

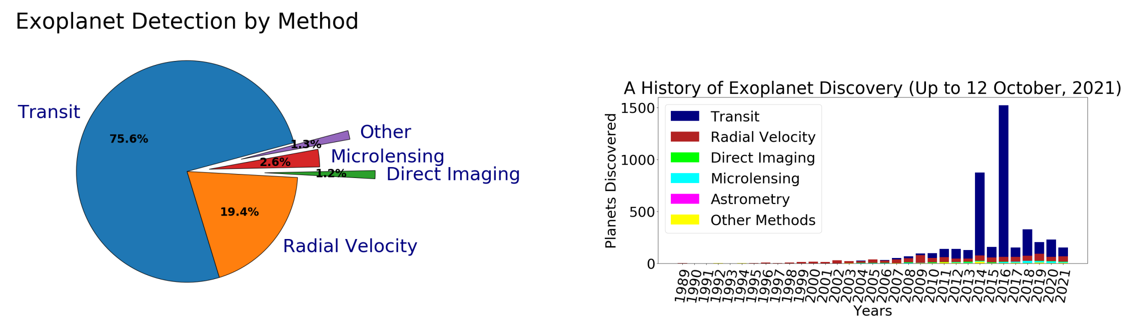
CHAPTER 1

Introduction: Exoplanets and high-contrast imaging

One of the prevailing challenges facing the fields of astronomy and applied optical engineering today is how to answer the question: is there other life in the universe? This question has provoked much thought from the philosophers and scientists alike all throughout history, and continued to inspire the human spirit of exploration, both physically and intellectually. In order to work towards the answer, many new technologies and algorithms have been developed and improved over the last three decades, including but not limited to: high precision wavefront measurement and control (Adaptive Optics (AO) and Coronagraphy) techniques, exquisite new detectors, and state of the art computational abilities that have opened the door for calculations once thought to be impossible to undertake. As the generation of ground based observatories known as *Extremely Large Telescopes (ELT)*, or the class of 25+ m diameter telescopes, arrives, alongside new space based observatories, our capability to resolve and characterize distant worlds is becoming a reality. With these instruments soon being available for use, the drive to push the limits of technology and algorithms to find the light from distant planets hiding underneath their host stars' bright shine has become a priority. This chapter will take a brief look into the history of exoplanet detection including various techniques that have been used. Further detail will then be provided for the direct imaging case, through the lens of high-contrast imaging as the example.

1.1 History of exoplanet discovery

The exploration and detection of *exoplanets* is a field that has rapidly developed over the last three decades. As of 12 October, 2021, 4528 such planets around stars other than the Sun (hence the name exoplanet meaning planet outside of our solar system) have been confirmed through various methods, some of which will be described with more detail later. Fig. 1.1 shows the percentage of confirmed exoplanets by method, as well as the year in which the the planets were confirmed. It is obvious from these plots that the *transit* method has been the most successful with respect to the total number of confirmed planets, followed by *radial velocity* (RV). The transit method has claimed the most confirmed exoplanets largely due to the Kepler and K2 missions, using the Kepler space telescope [4]. The Kepler space telescope has an observational field of view (FOV) of 115 square degrees, meaning it continuously can monitor the brightness of $150,000$ main-sequence stars. The number of exoplanets that Kepler has found (that have been further confirmed by other methods such as RV) looking at only a very small percentage of the estimated total number of stars in the universe tells us that planets are likely more common than was thought even as recently as the 1990s. As more planets have been found, the scientific community has pushed the limits of new and proven techniques to continue to refine the capabilities to search for Earth-like planets. Missions such as the Transiting Exoplanet Survey Satellite (TESS) [71] and the James Webb Space Telescope (JWST) [64] will continue the effort from space, while pioneering work is being done with the Spectro-Polarimetric High-Contrast Exoplanet Research instrument (SPHERE) [2], the Gemini Planet Imager (GPI) [45], the Subaru Coronagraphic Extreme Adaptive Optics (SCExAO) instrument [43], and the Magellan Extreme Adaptive Optics (MagAO-X) instrument [46] from the ground. All of these instruments are not only designed to help us discover more exoplanets, but also to learn more about their interesting characteristics. Understanding the mass, size, orbital radius, and even atmospheric composition of discovered exoplanets is vital to being able to characterize them, and also to make strides in answering one of humanity's most prevailing questions: are we alone in the universe?



(a) (b)

Figure 1.1: (a) The percentage of confirmed exoplanets discovered by different techniques. (b) The year of each confirmed exoplanet discovery and method by which it was detected. This data is made publicly available through the NASA Exoplanet Archive.

1.2 Observation of exoplanets

There are two main forms for observing exoplanets: direct and indirect detection. Indirect detection is observing phenomenon that occur to a host star for which a present exoplanet is a plausible/likely explanation, even though the exoplanet itself is not observed. Direct observation is, as is easy to surmise, the opposite, in which light from the exoplanet itself *is* measured. Each form of observation has several prevailing methods, of which a few are described below.

1.2.1 Indirect methods

The first means of indirect detection is the aforementioned transit method, which is the method used by the widely successful Kepler and K2 missions, among others such as TESS, and the citizen science project, PANOPTES [42]. In the transit method, the light photometry (or brightness) curve of a star is measured over a long period of time (from days to years). If the star being watched has an exoplanet in a face-on orbit with respect to the Earth (or more specifically the instrument observing the star's light), the planet will eventually cross between the star and the observer (a telescope on or near Earth). When this happens, the planet blocks some of the starlight, causing the light curve to drop corresponding to the reduction in brightness. When the planet then continues on its orbital path, and is no longer in front of the star, the light curve goes back up to its 'nominal' state. If this happens in a periodic, predictable way, it is considered a likely exoplanet in orbit around the star. The larger the planet is in diameter, the deeper the dip, and the longer the dip lasts, the further out the planet's orbit is from its star.

A second indirect method for discovering exoplanets is the Radial Velocity (RV) method. Planets have a gravitational affect on their host stars, which will cause them to wobble. As the planet orbits, if the star is pulled slightly away from Earth, the starlight will be redshifted, whereas if the star is pulled towards Earth, the starlight will be blueshifted. The more massive the planet, the larger the gravitational affect on the star, and thus the larger the shift in the frequency of the starlight. Precisely observing the spectrum of a star with an exoplanet along the line-of-sight over time with a spectograph will reveal a periodic variation in the wavelength of characteristic spectral lines corresponding to the Doppler shifts (increasing and decreasing). These measurements tell us the radial velocity of the star over that period of time, allowing us to plot them to see a characteristic sinusoidal shape that is indicative of a planet in orbit. Given the mass of the star being observed, and the period of this sinusoid of the star's radial velocity, the orbital radius of the planet can be determined, as well as the radial velocity of the planet. Thus we can also determine the minimum mass of the planet orbiting the star, as the inclination angle of the planet's orbit with respect to the perpendicular of the line-of-sight is often unknown, meaning the measured motion of the star may appear to be less than it actually is (as the measured velocity of the star is the true velocity scaled by the sine of the inclination angle). RV was the most successful detection method up until the rise of the transit method, and is also used as a means of providing confirmation of exoplanet candidates discovered by other methods because of its longer history of valid detections.

Although there are still other indirect methods, the last one that will be covered here, due to its interesting nature, is Gravitational Mircolensing. Light from a distant star is bent and focused by gravity as a planet passes between the star and Earth. Objects with mass warp space, causing light that travels through that region of space to appear to change direction. To an observer on Earth, this appears as a distant star that gets gradually brighter over time, and then fades away. If the star had a planet, a brief window may also be present where the light from the planet is focused by gravity and will appear during this process. Although this method is not as prolific as the other two brought up, due to the higher luck needed for an observer on Earth to be in-line with both the star bending light and the possible planet whose light is being bent behind it, Fig. 1.1(b) shows that such luck does happen a few times a year.

1.2.2 Direct imaging

While indirect methods are great for inferring or even confirming the presence of an exoplanet, and tell us some about its size, mass, and orbital period, they leave a lot of the more valuable information about an exoplanet off the table. In order to truly examine if a planet shows characteristic signatures of life, the light from the planet itself must be isolated and examined. Direct imaging techniques, as the name suggests, allows for just that, the observation of the planet light, which then allows for spectroscopic analysis. Such an analysis would look for the presence of *biomarkers* such as O_2, CH_4, and N_2O, alongside H_2O and CO_2 [40], that could inform us with great confidence of the presence of carbon-based, Earth-like life on the planet. As will be discussed, there are tough challenges to be overcome to allow for this type of observation, but the payoff is the best way forward for finding habitable worlds in the galaxy around us.

Coronagraphs

Direct imaging, which henceforth will be referred to as high-contrast imaging, generally employs the use of an optical system known as a *coronagraph.* A coronagraph is an optical system that is designed to suppress the on-axis light from a host star, but allow off-axis planet light to reach the detector largely “unharmed”. This is achieved by altering the diffraction pattern of the telescope either by manipulating

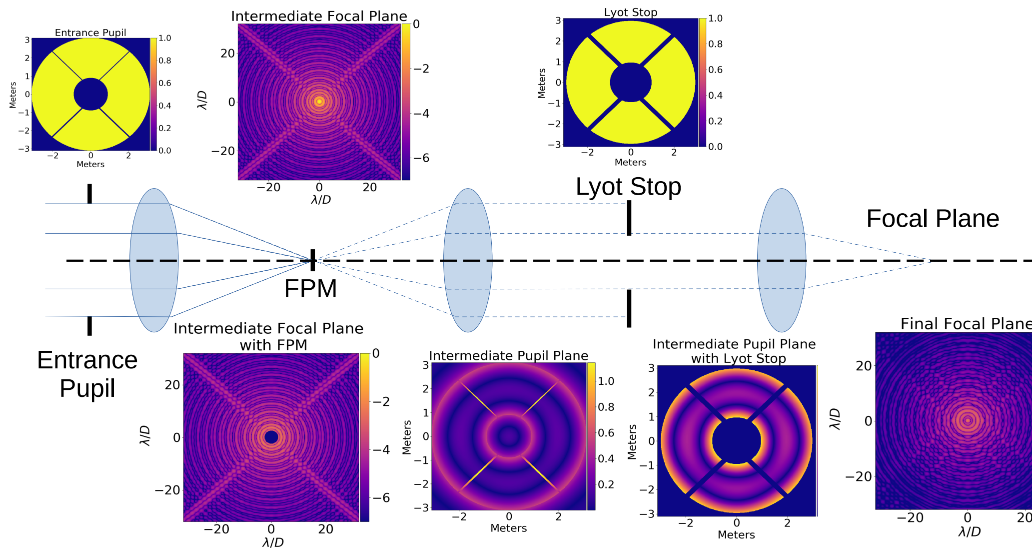


Figure 1.2: Cartoon of a Lyot Coronagraph with simulated images at various planes. Pictured are the entrance pupil shape for the Magellan Clay telescope, the intermediate focal plane with and without the FPM inserted, the Lyot Stop, the intermediate pupil plane with and without the Lyot Stop inserted, and the final focal plane.

the phase, the amplitude, or both as light travels through the system. The "canonical" coronagraph, called the Lyot coronagraph, which was developed in the 1930s by Bernard Lyot, consists of a small mask that is placed on-axis in an intermediate focal plane to block the core of the stellar PSF. This so-called *focal plane mask* (or FPM) has a maximum radius that is determined by not wanting to block all the off-axis light. One could imagine a coronagraph that blocks all the star light by simply placing a very large FPM in the intermediate focal plane. This however is very ineffective because the planet light throughput would then also be suppressed. The *inner working angle* (IWA) of a coronagraph is thus defined as the point for which the source throughput is half of the maximum throughput, and represents how close to the star (typically in angular units of λ/D, where λ is the observation wavelength and D is the diameter of the telescope being used) an exoplanet can be imaged. The example in Fig. 1.2 employs an FPM with a radius of $\approx 2\lambda/D$, corresponding to an IWA of approximately $2.5\lambda/D$. Propagating further in the Lyot coronagraph,

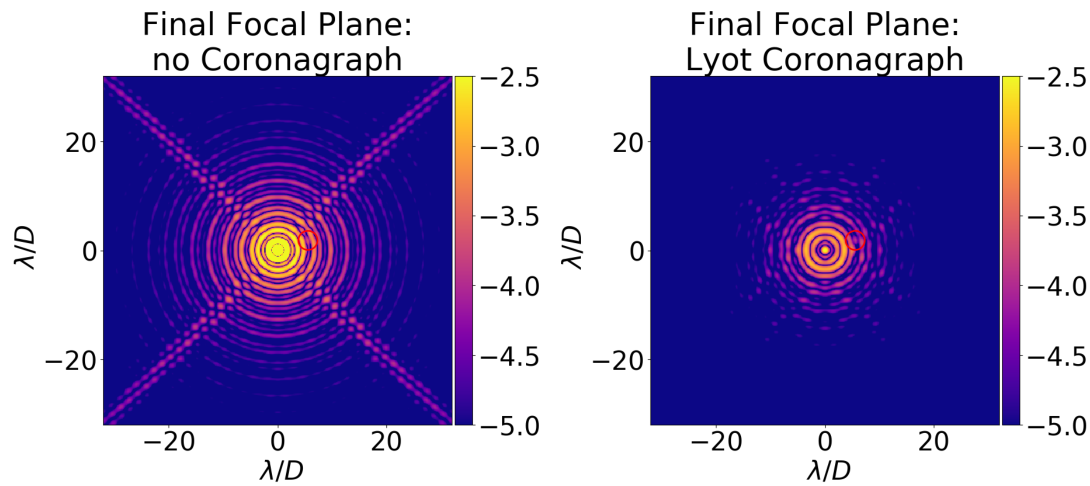


Figure 1.3: Focal plane images with and without a coronagraph. The color scale is chosen to maximize the scale of the coronagraphic image, meaning the coronagraph free image appears somewhat saturated. Both images have a 10^{-4} contrast planet injected at the center of the circled region. The exoplanet does not show well in the coronagraph free image, but does show in the coronagraphic image.

we arrive in an intermediate pupil plane following the FPM focal plane, in which another mask, or field stop, known as the *Lyot stop* (LS), is placed. The LS, as a field stop, serves the purpose of limiting the field of view of the coronagraph, thus setting the so-called *outer working angle* (OWA): the off-axis angle corresponding to where the exoplanet throughput is at least half of its maximum transmission. Together, then, the FPM and the LS work to block a majority of the stellar light from reaching the final, science focal plane, while defining the range of off-axis angles for which exoplanet light will reach the final focal plane with at least 50% transmission, making the exoplanet imageable. An example of the starlight suppression of a Lyot Coronagraph is shown in Fig. 1.3, while a simple ray diagram of it including an off-axis source can be seen in Fig. 1.4 .

The Lyot coronagraph is described here as its use will be employed in the simulations presented in Chapter 5. There are, however, many more types of coronagraphs that have been developed to improve on different parameters from the Lyot. These include Phase Induced Amplitude Apodization Complex Mask Coronagraphs (PI-

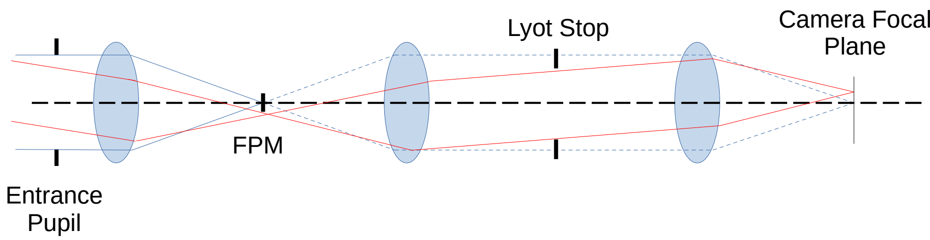


Figure 1.4: Schematic diagram of a Lyot coronagraph including an off-axis source. The inner and outer working angles are easier to understand by examining the size of the FPM and Lyot Stop in this figure.

AACMC) [32] and vector vortex coronagraphs [58]. In addition, other non-Lyot based coronagraphs, such as Apodizing Phase Plate (APP) coronagraphs [61; 65], have been developed. These types of coronagraphs improve on the starlight suppression and/or throughput of the planet light in different ways, but we will leave that to the reader to examine if they are so inclined.

1.3 Obstacles to direct imaging of an exoplanet

Among the many challenges of high-contrast imaging, three stand out as the largest obstacles to successfully observing exoplanet light directly. The first of these is the fact that the planet light is so much dimmer than the host starlight. If we define the metric *contrast* as the ratio of the maximum intensity of the light from the exoplanet divided by the maximum intensity of its host star, or $\frac{max(I_{planet})}{max(I_{star})}$, we can start to get a feeling for this problem. Although the required starlight suppression to achieve a contrast level to be able to observe a signal from the planet is heavily dependent on the host star type, as well as the observation wavelength, we can examine a couple of usual scenarios. A typical Earth-like planet around a Sun-like star, observing in visible wavelengths, will be approximately 1 to 10 billion times fainter than the star, or a contrast of 10^{-10}. If we swap to observing the thermal emission of the planet, the required contrast falls to roughly 10^{-6} [29]. This means that, for the purpose of discussion here, ignoring the capabilities of modern post-processing techniques that can improve upon the raw contrast achieved by the corongraph itself, the coronagraph has to suppress the starlight in a region of interest in the science focal plane by at least 6 orders of magnitude, up to 10, for the signal from the exoplanet to be observable. Although many coronagraph designs routinely can achieve contrast levels this deep in simulation and the lab in combination with several algorithms that will be discussed in Chapter 2, doing so on sky has proven to be quite difficult.

This also brings us to the second main obstacle that stands in the way of directly imaging light from an exoplanet: the angular separation of the planet from the star on sky. Stars are very far away from us here on Earth. The closest star to us, Alpha Centauri C (perhaps better known as Proxima Centauri), is 4.25 light years from Earth. Doing some simple trigonometry, we can get an idea of the angle separating an Earth-like exoplanet in an "Earth-like" orbit around it. This is to say, if we were to assume an exoplanet is in orbit around Proxima Cen such that it is in the so called *habitable zone*, or the range of semi-major orbital axes for which liquid water could exist on the surface of said planet (in fact, such a planet is confirmed to exist, named Proxima b [55]), we can compare the angular separation seen from Earth compared to the angular resolution of a typical modern telescope. Because Proxima Cen is an M-type red dwarf, its habitable zone is considerably closer to the star than the habitable zone of the Sun (a G-type main sequence star). This is because of the temperature and size of the stars are quite different. So while the habitable zone for the Sun is approximately 1 Astronomical Unit (AU, or the average distance between the Sun and the Earth of approximately 1.496×10^8 km), to consider an "Earth-like" orbit around Proxima Cen would be roughly 0.05 AU. This gives an angular separation for Proxima b, the exoplanet around Proxima Cen, as seen from Earth, of about 0.038 arcsecond. We can compare this value to the typical angular resolution of an $8m$ diameter telescope observing at $0.6\mu m$:

$$\frac{\lambda}{D} * 206265 = \frac{0.6 * 10^{-6}}{8} * 206265 \approx 0.0154 \text{ arcsecond.} \tag{1.1}$$

The consequence of this calculation reveals the difficulty of searching for exoplanets that are in their stars' habitable zone, as the separation on sky of such a planet in orbit around the nearest star to us is only separated from the star by about $2.5\times$ the angular resolution of the telescope. While this example may seem to provide some hope that our telescopes can resolve many habitable zone exoplanets, it does not, as this is examining the closest possible planet to us. Performing a simple calculation tells us that for a habitable zone that is roughly the same as that of Proxima Cen, the angular separation is only larger than the current state of the art angular resolution for stars less than 10.58 light years away. In fact, only 12 stars are within this distance from Earth, most of which are M-type stars (with the exceptions being Alpha Centauri AB, Sirius A, and Epsilon Eridani), so our assumption of the habitable zone radius likely is not a bad one for this comparison. Given this layout of our galaxy, current telescopes simply will struggle to directly image light from exoplanets that would be the most interesting to us in terms of habitability, even if the host starlight can be suppressed. For completeness, to show why the exoplanet community is so excited for the coming class of ELTs with much larger diameters, we can repeat this calculation for the Giant Magellan Telescope, with a maximum diameter of 25.448 m, and arrive at an angular resolution of 0.005 arcsecond, or approximately $7.6\times$ better than the current state of the art, and extends the corresponding distance from Earth to 33 light years.

And that brings us to the third obstacle, which will largely be the focus of the work in this dissertation, although the contrast and angular separation will still be discussed in Chapters 5 and 6 as a means to report on the results of Frazin's algorithm in useful observing scenarios. The above discussion assumes that the optical system being used around the coronagraph is *diffraction limited*, meaning the wavefront propagating through it is flat. But, as we all know, that is rarely the case, and aberration of various types creep in and degrade the performance of all optical systems. The most important factors limiting the optical systems for this work are atmospheric turbulence and non-common path aberration (NCPA), and will be discussed in Chapter 2.

CHAPTER 2

Technical background

To begin our discussion, we introduce two of the major contributing factors to the departure from diffraction limited imaging in ground–based astronomical observation: the atmosphere and non-common path aberration (NCPA). Although the work of this dissertation is primarily focused on the estimation of NCPA (in its static and quasi-static forms), understanding the concepts that arise when dealing with the effects of the atmosphere on light, how they are treated via Adaptive Optics (AO), and what the resulting optical system architecture and types of output telemetry are obtained, is paramount to the underlying mechanics of the method described in Chapter 4 and on. We start with a brief review of some important optics principles and definitions, and then move on to examining a simple AO system and how that can give rise to the possibility of uncompensated NCPA. Then, the problems NCPA cause for high-contrast imaging will be introduced, providing the motivation for the work that follows in this dissertation on methods for estimating the phase/Optical Path Length (OPL) underlying them so they can be eliminated. Finally, a brief introduction to some current techniques comparable to the work being presented here will be given to allow the reader to better understand the place of these methods.

2.1 A brief introduction to Fourier Optics: propagation of light

First, one must understand how light from a star (or its associated planet companion) is observed here on Earth. For illustrative purposes, we will return to the example of Proxima Centauri. Proxima Cen is the closest star to our own solar system, at a distance of approximately 4.25 light years (or to put that into a more easily graspable unit, $5,107,965,175$ round trips from Los Angeles to New York City by air). This means that light has to travel that distance (which as the unit of light year tells us, takes 4.25 years time) to arrive here on Earth to be observed by our telescopes. We will describe this light in terms of its electric field component, called ψ, which will represent the complex amplitude *field*.

If we take a moment and consider just how far away Proxima Cen (again the closest star system to Earth) is, a simple thought experiment can be conducted. From Earth's perspective, Proxima Cen will appear to be a point source, emitting spherical waves. As these spherical waves travel the great distance to reach Earth, they expand into ever increasingly larger radii. By the time the wave reaches the Earth, its radius of curvature is so large that where the wave intersects with the Earth can effectively be considered to now be planar (i.e. as a plane wave rather than a spherical wave). Ignoring the effects of the atmosphere for now (which will be treated in Chapter 2), we can use the entrance pupil of the telescope, typically the primary mirror, to impose a set of boundary conditions on the region of the wave we need to care about. Defining a simple pupil transmission function, $\mathcal{A}(\boldsymbol{r})$, where $\boldsymbol{r} = (x, y)$ is the vector representation of Cartesian coordinates, allows us to enforce its geometry on the field we wish to examine:

$$\psi(\boldsymbol{r}) = \mathcal{A}(\boldsymbol{r}) \exp(j\phi(\boldsymbol{r})) , \tag{2.1}$$

where j is $\sqrt{-1}$, and $\phi(\boldsymbol{r})$ is the phase of the field. Although we will typically only refer to the phase, ϕ, going forward, it is still important to understand what this means. The phase is represented as:

$$\phi(\boldsymbol{r}) = k \text{ OPL}(\boldsymbol{r}) = \frac{2\pi}{\lambda} \text{ OPL}(\boldsymbol{r}) , \tag{2.2}$$

where k is the wave number, λ is the wavelength, and $\text{OPL}(\boldsymbol{r})$ is a function describing the *Optical Path Length*. $\text{OPL}(\boldsymbol{r})$ represents the real world distance that light has to travel in the region represented by the boundary condition enforced by the pupil geometry being examined, with units of length, and can be the shape/surface of an optical element, or any spatial representation the path light takes. If we instead consider the *Optical Path Difference* (OPD), we can distinguish the path the light takes due to aberrations (errors with respect to the intended paths that would arise from the optical elements in the optical system). In this sense, we can propagate a field, $\psi(\boldsymbol{r}) = \mathcal{A}(\boldsymbol{r}) \exp(-j\frac{2\pi}{\lambda}\mathrm{OPD}(\boldsymbol{r}))$ by describing the aberration in the field through the OPD. Going forward, we define surfaces of constant ϕ as *wavefronts*, and this terminology will largely be used interchangeably with field. A perfectly flat wavefront corresponds to having $\mathrm{OPD}(\boldsymbol{r}) = 0$ for all $\boldsymbol{r}$.

We now consider the propagation of the wavefront incident in on the telescope:

$$u(x, y, z = 0) = u(\boldsymbol{r}) = \mathcal{A}(\boldsymbol{r}) \exp(j\phi(\boldsymbol{r}))\,. \tag{2.3}$$

Assuming an $\mathrm{OPD}(\boldsymbol{r}) = 0$, our field collapses to the much more simple to deal with pupil transmission function, $u(\boldsymbol{r}) = \mathcal{A}(\boldsymbol{r})$. For a complete derivation of scalar diffraction theory, I highly recommend the reader consult Chapters 3 and 4 in Goodman's *Introduction to Fourier Optics* [26]. To simplify the discussion here though, we will pick up the derivation at the Fraunhofer Approximation, equivalently known as far-field diffraction, given by the equation:

$$U(\xi, \eta \mid z) = \frac{\exp(jkz)\exp(j\frac{k}{2z}(\xi^2 + \eta^2))}{j\lambda z} \iint_{-\infty}^{\infty} u(x, y, 0) \exp[-j\frac{2\pi}{\lambda z}(\xi x + \eta y)]dxdy\,. \tag{2.4}$$

This integral is applicable under the rather stringent assumption that the propagation distance, z, is sufficiently long, following the relation:

$$z \gg \frac{k(x^2 + y^2)_{max}}{2}\,. \tag{2.5}$$

where $(x^2 + y^2)_{max}$ can be viewed, in the case of a circular pupil, as the square of the radius of the pupil. We can thus rework this relation as:

$$z \gg \frac{2\pi}{\lambda} \times \frac{r^2}{2} = \frac{\pi r^2}{\lambda}\,. \tag{2.6}$$

For a typical modern telescope (diameter of 8m), observing at visible wavelengths (for this purpose we will call out 0.6μm for the calculation), we see that $z \gg 167,550$ km, or to provide comparison to our above example, approximately 21.25 round trips from Los Angeles to New York City by air. And certainly, $5,107,965,175d \gg 21d$, where d is the distance from LA to NYC and back, is true, validating the use of the Fraunhofer Approximation for all star light.

Returning to Eq. (2.4), the trained reader will see that, ignoring the pre-multiplication factors, this integral is simply a *Fourier transform* of the input field, $u(\boldsymbol{r})$, evaluated at the spatial frequencies: $k_x = \frac{\xi}{\lambda z}$ and $k_y = \frac{\eta}{\lambda z}$. This allows us to rewrite Eq. (2.4) as:

$$U(\xi, \eta \mid z) = \frac{\exp(jkz)\exp(j\frac{k}{2z}(\xi^2 + \eta^2))}{j\lambda z} \mathscr{F}\left\{u(x, y)\right\}|_{k_x = \frac{\xi}{\lambda z}, k_y = \frac{\eta}{\lambda z}}\,, \tag{2.7}$$

where $\mathscr{F}$ is notation for the Fourier transform. This is an important equation for us going forward. It is important to also note that this equation is valid in another situation, which will be what helps to make it so useful to us. This situation is one in which a positive powered lens (or mirror) is properly placed between the planes being examined (the plane $u(x, y)$ is in, and the plane the resulting field will be observed). Examining the math (not shown), we can see that applying the Fresnel integral to propagate light focused by a lens a distance of the focal length, cancels out the spherical wave term, leaving us again at Eq. (2.4) [26]. The reason these two situations above are interesting to us in astronomical imaging is that we can, at least in a first-order modeling/simulation sense, ignore a lot of the details of a specific optical system, and instead treat the system as a series of conjugated intermediate pupil and focal planes. Following the above description of the propagation, the light can then be propagated between such planes with Fourier transforms, giving rise to the name *Fourier Optics*.

This leads to another important definition: the system *point spread function* (PSF). We can propagate the aberration-free field in the entrance pupil defined above, $\mathcal{A}(\boldsymbol{r})$, to its conjugated focal plane via Eq. (2.7). By doing this, we are assuming no further optical elements have to impart any further "disturbances" (planned or otherwise) to the field, so in essence the entrance pupil is equal to the exit pupil. However, at least at the time of writing, the phase of visible and near-IR wavelength light can not be measured; only intensity can be. The intensity of light, which *can* be measured (in this case assuming a noise-free detector) is the same mathematically as the PSF, and is given by:

$$i(\xi, \eta) = U(\xi, \eta) \times U^*(\xi, \eta) = \left|U(\xi, \eta)\right|^2 \equiv \mathrm{PSF}(\xi, \eta) = h(\xi, \eta)\,, \tag{2.8}$$

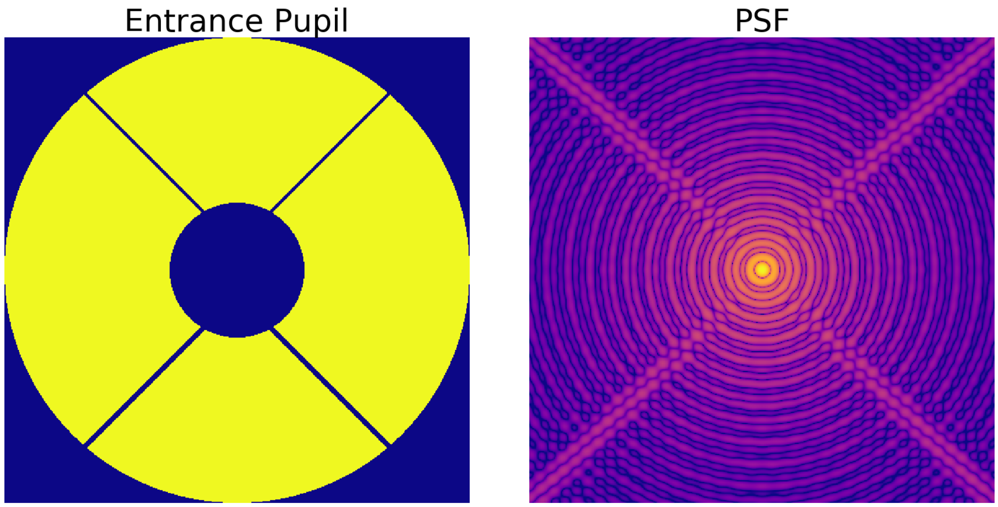


Figure 2.1: An example of conjugated pupil and focal planes for an aberration free optical field incident on the MagAO-X pupil geometry.

where * is the complex conjugate, and $U(\xi, \eta)$ is given in Eq. (2.7). Because this is formed from an aberration-free field, this is said to be *diffraction limited*. This means, in effect, the diffraction limited image of a star in the final focal plane is the PSF of the telescope used to focus the light on the detector. An example of the entrance pupil shape for the Magellan Extreme Adaptive Optics system (MagAO-X), and the associated PSF for an aberration-free field passing through it, is given in Fig. 2.1.

Looking at the PSF in the right frame of Fig. 2.1, several useful features can be pointed out. The *core* of the PSF is defined as the bright region at the center of PSF out to the first null, appearing as a disk. For a circular pupil, the first null occurs at an angular distance on the detector of 1.22 λ/D, given by the first zero of the jinc function defined as the Fourier transform pair of the circ function. This means the diameter of the core is double that amount. For modified circular pupil geometry, like for MagAO-X shown in Fig. 2.1, the central obscuration typical of the majority of telescopes, will reduce this diameter. The angular measurement λ/D will be used throughout this work as it is an easy way to relate sky angles to the physical attributes of the telescope / optical system and observing wavelength being used.

Although angular separations are usually given by measurements of arcseconds, being able to convert to units of λ/D allows the reader to quickly understand this angle in terms of the spatial resolution of the telescope (for example: resolving two point sources spatially is often defined by the Rayleigh criterion as the two sources being separated by 1.22 λ/D [35]). The conversion factor from units of radian to arcsecond is simple, via a multiplicative factor of $\approx 206265 \left[\frac{arcsecond}{radian}\right]$. This gives us λ/D in units of arcseconds, which can then be applied to compare to sky angles in measured data, as we saw in Sec. 1.3. Going out from the core, we see that the light is diffracted into many rings, and the *spiders* in the pupil, which are shadows of the struts used to hold the secondary mirror over the primary, cause diffraction spikes at associated angles. In Chapter 2, various sources of degradation to the PSF will be expanded upon that cause a departure from this diffraction limited case, but for now we will leave this discussion, and proceed further towards another useful quantity that arises from the linearity of the optical systems being used, the *Optical Transfer Function* (OTF).

The OTF first and foremost is allowed because the optical system is linear, shift-invariant (LSI) in intensity. This gives us the possibility to analyze the spatial frequency spectrum of the PSF, which can be quite useful. The unnormalized OTF is defined as:

$$\mathrm{OTF}(k_x, k_y) \equiv \mathscr{H}(k_x, k_y) = \mathscr{F}\left\{h(\xi, \eta)\right\} = \mathscr{F}\left\{|U(\xi, \eta)|^2\right\}. \tag{2.9}$$

The OTF is often considered as a normalized quantity, which is achieved by dividing by the maximum complex value of Eq. (2.9), and an example of it presented this way can be seen in Fig. 2.2. In the normalized state, several useful properties and interpretations of the OTF can be noted:

- $\mathscr{H}(0,0) = 1$
- $\mathscr{H}(k_x, k_y) \leq 1$ for all $(k_x, k_y) \neq 0$ due to the Cauchy-Schwarz inequality
- The OTF has Hermitian symmetry (as a consequence of being the Fourier transform of a strictly real quantity)

- The values of the OTF are the complex weights the optical system being analyzed applies to the spatial frequencies passing through it

By understanding the weighting the optical system applies to the spatial frequencies included in the wavefront that has been processed to the PSF, we can easily understand how the optical system works as a low pass filter. By examining Fig. 2.2, we see a very clear cutoff, where the optical system destroys any information outside a certain spatial frequency as the light goes from the entrance pupil to the conjugate focal plane. This is a key factor in many applications of optics, as these spatial frequencies are lost, and cannot (at least easily) be reconstructed. In the case of a typical circular aperture that is seen in astronomical applications, this cutoff frequency is given as:

$$\pm \frac{1}{\lambda f/\#}\,, \tag{2.10}$$

where $f/\#$ is the f-number (related to the focal length and diameter of the entrance pupil).

Furthermore, we can define the *Modulation Transfer Function* (MTF), as the modulus of the OTF. This is the quantity that Fig. 2.2 is actually showing, rather than the OTF itself, as for a diffraction limited case, the phase of the OTF is not that interesting. Chapter 3 will provide details on how to make use of the idea of the OTF to construct a simple wavefront sensor to measure both phase and amplitude errors in a static system. and provide a few experiments showing different use-cases. Now armed with the basics of light propagation, and some of the interesting results that fall out of linear systems theory that provide tools for analyzing light, we can move on to understanding the motivation for the work in this dissertation.

2.2 A brief introduction to Adaptive Optics

2.2.1 Atmospheric turbulence

As described in Chapter 1, we assume light from a distant star to be from a point source at infinity, meaning it arrives at the Earth's atmosphere as a plane wave.

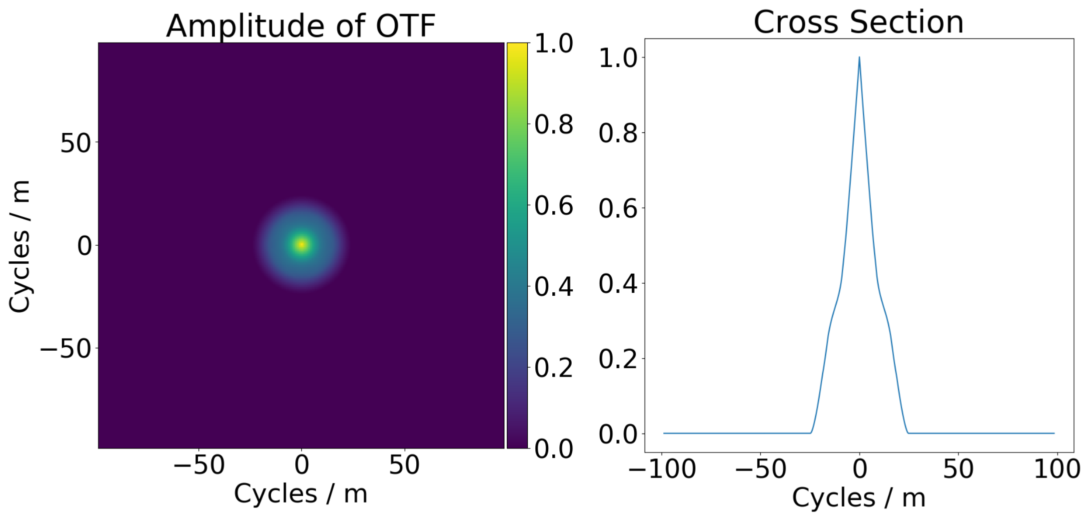


Figure 2.2: The corresponding Modulation Transfer Function ($|\mathrm{OTF}|$) for a diffraction limited MagAO-X optical system.

Now, the light must travel through the layers of air in the atmosphere to reach the primary mirror of the telescope observing light from the star. Conceptually, small temperature fluctuations in the air cause random variations in wind speeds and directions, leading to the formation of turbulence as eddies. Optically, the result of this randomly occurring displacement of air (via changing temperature and pressure) is that:

1. the refractive index varies randomly, leading to random refraction of the light as it travels; the reader can consider this as if the small / large scale eddies act as a long series of lenses the light must propagate through.

2. the wavelength dependence of the refractive index will cause differential refraction that is dependent on wavelength, that will disperse light as it propagates.

3. Molecules (air, dust, pollution, pollen, water vapor, etc.) can have size on the order of the wavelength of light, leading to scattering effects.

4. as the light propagates, all of these aberrations propagate as well, and can be coupled into amplitude effects called scintillation.

Though we do not wish to bog the reader down in all the theory of propagation through random media and turbulence, we can summarize the important terms and ideas that will influence the construction and understanding of the simulations that follow. There are many references that can be consulted for more information on these topics, namely Goodman (2015) and Hardy (1998), from which this section draws from.

The first terms to be familiar with are the outer and inner scales. The outer scale, referred to as L_0, is the size of the initial, largest eddy where energy is injected in the atmosphere. The inner scale, or l_0, is then the smallest scale where energy is dissipated into heat. The process of turbulence can then be described as the cascading of energy from the eddies at the shearing layers (of scale L_0) down to scales of l_0 as the energy is dissipated over some time scale. The power spectral density (PSD) that we will use to model this process is the von Kármán spectrum, that derives from additions to the seminal work on turbulence of Kolmogorov by Tatarski and von Kármán. We start by defining the outer scale as $L_0 = \frac{2\pi}{k_0}$, where k_0 is the critical wave number where the shape of the PSD is determined by the laws governing the breakup of large eddies into smaller ones. Typical values of L_0 range from 1 to 100m depending on the atmospheric condition. For $k > k_0$, or what is called the inertial subrange, Kolmogorov's work gives the PSD of the turbulence flow as:

$$\Phi_n(k) = 0.033\ C_n^2\ k^{-11/3}\,, \tag{2.11}$$

where C_n^2 is the *turbulence strength* or *structure constant* of the index of refraction fluctuations. Both Goodman (2015) and Hardy (1998) go in to great detail on the derivation of C_n^2 and the other structure functions that are useful for describing turbulence, but that is not of particular import to this discussion. Understanding that C_n^2 is related to the strength of the turbulence, or in other words, how much the turbulence will distort the plane wave as it propagates through the atmosphere, is all that is required. Now, the inner scale is included by defining k_m as the critical wave number for which $l_0 = \frac{2\pi}{k_m}$, and understand that for $k > k_m$, the turbulent eddies are small enough to dissipate energy as a result of viscous forces. Tatarski corrected the Kolmogorov PSD to better approximate what happens for $k > k_m$ by introducing a Gaussian tail:

$$\Phi_n(k) = 0.033\ C_n^2\ k^{-11/3}\ \exp\left(-\frac{k^2}{k_m^2}\right). \tag{2.12}$$

Finally, needing to account for the non-integrable pole at $k = 0$, as the finite amount of air in Earth's atmosphere enforces that $\Phi_n(k)$ cannot be infinitely large as $k \to 0$, von Kármán introduced the outer scale into the spectrum as:

$$\Phi_n(k) = 0.033\ C_n^2\ \left[k^2 + k_0^2\right]^{-11/6}\ \exp\left(-\frac{k^2}{k_m^2}\right). \tag{2.13}$$

With this PSD, we can fully define the spectrum of an atmospheric model in the region near $k_0 < k < k_m$ with just four parameters: the observing wavelength, the turbulence strength, the inner scale, and the outer scale.

The next important piece to understand is the *Fried parameter*, sometimes called the *seeing parameter*, r_0 [18]. r_0 derives from the effects of turbulence on optical phase, and is defined as:

$$r_0 = \left[0.423k^2(\sec(\zeta)) \int C_n^2(h)dh\right]^{-3/5}, \tag{2.14}$$

where h is the altitude path through the atmosphere and ζ is the zenith angle. This gives us a clue into C_n^2, as here the profile of C_n^2 is shown to have dependence on altitude, and can be integrated and scaled to arrive at a new parameter that will allow for an easier interpretation of what "turbulence strength" means. If we evaluate the mean square wavefront phase error over a circular telescope with a diameter of D at observation wavelength λ_0, we get:

$$\sigma_w^2 = 1.075\left(\frac{D}{r_0}\right)^{5/3}\ \mathrm{rad}^2. \tag{2.15}$$

This equation provides a physical interpretation of r_0, as the diameter of a telescope for which the atmospheric wavefront has an mean square error of approximately 1 radian. This also allows for some additional analysis, giving a rule of thumb that a telescope with a diameter D that is much smaller than r_0 will still appear to be close to diffraction–limited, but if D is much larger than r_0, it is said to be "seeing–limited", or dominated by the wavefront error induced by the atmosphere. A typical value for r_0 at an observing wavelength of 0.5μm is 20cm for a good site, and down to 5cm for a poor site. Considering modern telescopes are upwards of $6-8$m in diameter, with future telescopes pushing beyond 30m, Eq. (2.15), these typical r_0 values demonstrate why atmospheric turbulence is such a hindrance to ground–based astronomical observation, including high–contrast imaging. Evaluating Eq. (2.15) for $D = 6$m and $r_0 = 0.2$m, we see the resulting mean square phase error is around 300 rad^2, an enormous error. As r_0 depends on the integral of C_n^2, and is somewhat easier to interpret, it is often specified over C_n^2 when a summary of the turbulence strength is all that is necessary. It is also interesting to determine the peak wavefront excursion in terms of r_0. Removing any error caused by the angle of arrival (which accounts for tilt errors), for a 99.4% probability ($\pm 2.5\sigma$), the peak wavefront excursion is:

$$W_{pk} = \pm 0.149\lambda_0 \left(\frac{D}{r_0}\right)^{5/6} \text{ meters,} \tag{2.16}$$

where r_0 is specified at λ_0 [33]. For the example of a 6m diameter telescope observing in turbulence represented by an r_0 value of 0.2m, this corresponds to a peak excursion of $\pm 1.25\mu$m, or about 2.5 waves, for what is considered a good observing site.

Finally, we discuss the general way of modeling the von Kármán turbulence, that includes the assumption of "frozen" flow. In order to model an evolving atmospheric aberration, multiple layers are used, corresponding to different altitudes (and thus individual r_0 as well), which are realizations from Eq. (2.13) we call *phase screens*. Each layer is then assigned a wind vector with a given speed and direction. Considering each layer as "frozen", they are then translated with respect to each other temporally, with the final result that is used being the sum of the layers at each time step. This does not fully capture the behavior of the atmosphere because it does not accurately represent the temporal moments, but in many cases it is an adequate description. If the effects of scintillation are to be included in the model, rather than summing the layers, a wave is propagated from layer to layer. A random

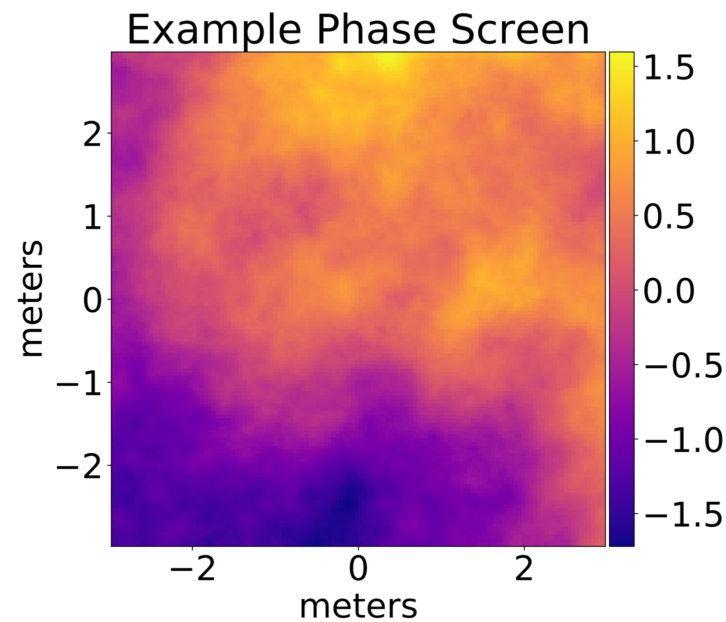
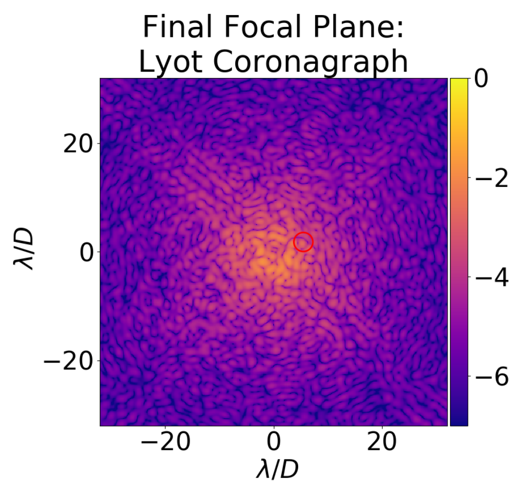


Figure 2.3: An example of the OPD induced into a wavefront from turbulence with an r_0 of 0.2m at $\lambda = 0.5\mu$m. The color bar is in unit of microns.

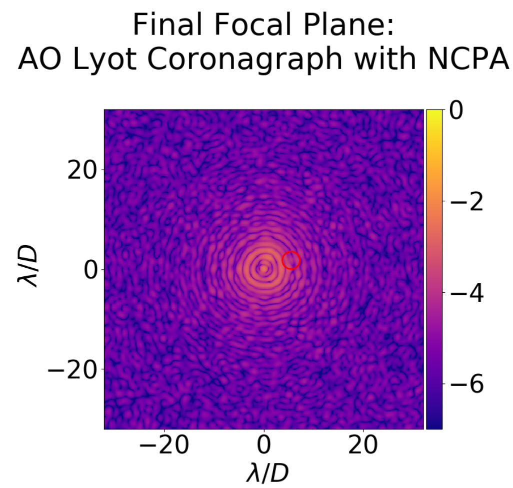
realization of the OPD induced by such a model, with $r_0 = 0.2$m can be seen in Figure 2.3. In this figure, we see variations across the spectrum of low and high spatial frequencies, which will be seen to distribute light throughout the focal plane. A more thorough explanation of the modeling of the atmosphere for the simulations is given in Chapter 5.

2.2.2 Degradation of the PSF

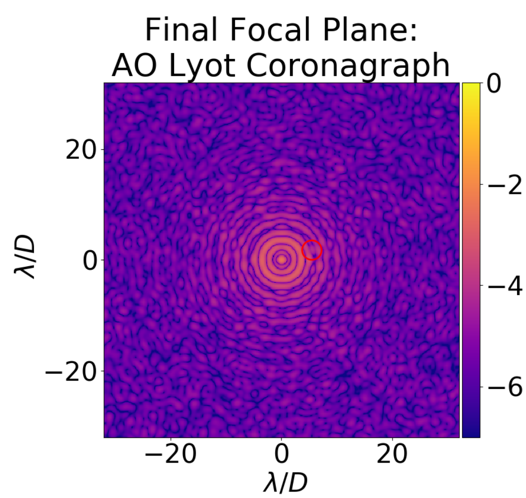
Exactly how the PSF is affected depends on several factors, but all fall from the four points enumerated above. The work in this dissertation will largely ignore the second and fourth points, dispersion and scintillation, but their inclusion in the methods described is straightforward, and thus their lack of appearance should not be considered as a shortcoming. With that in mind, we will continue to ease the burden on the reader and focus on the distribution of phase errors that lead to light being scattered about the focal plane when a wavefront that is distorted by the atmosphere propagates through an optical system and is brought to focus. If imaging is done through the atmosphere in a seeing–limited case, we no longer expect the nice, clean PSF that shows up in Chapter 1. Instead, the distortion in the wavefront will scatter light out of expected diffraction–based structure and place it into a swarm of *atmospheric speckles* of size proportional to λ/D. Because the temporal variations of the atmosphere happen on small (millisecond) time scales, these speckles are also changing very rapidly. This will prove to be extremely important in Chapter 4, and provide justification for the use of millisecond imaging techniques to freeze swarms of these speckles to be exploited. An example of such an image can be seen in the upper left frame of Figure 2.4. In this figure, we return to the Lyot coronagraph optical system, and examine what the "instantaneous" PSF looks like when imaging through the atmosphere. If we compare this to the right hand frame of Figure 1.3, we see a major difference. Instead of light being concentrated into a clearly defined set of bright rings surrounding an Airy disk core, it is distributed throughout the focal plane as speckles, with very little discernible structure. In addition, any planetary signal, located in the red circle, is completely indiscernible. We will ignore the upper right and lower frames of this figure for the time being, and continue to look at the effects of atmospheric turbulence by now considering a long-exposure image. Because the atmospheric speckles are rapidly changing with the dynamic nature of the atmosphere, and are distributed throughout the focal plane due to the continuum of spatial frequency content in the induced aberration, over a long-exposure, they will average into a spatially smooth halo of size approximately λ/r_0. An example of such a halo can be seen in Figure 2.5, as well as its cross-section to demonstrate the fact that the diffraction limited Airy Disk pattern is lost. In the case of this example, not enough time was integrated over in the observation for the halo to become spatially smooth, so some speckle features remain. However, it is easy to see that the effect of imaging through this atmosphere smears light out, greatly increasing the effective size of the PSF, meaning that any object imaged will be greatly blurred, and suffer an extreme reduction in resolution,



(a)



(b)



(c)

Figure 2.4: (a) The PSF when using the Lyot Coronagraph model to image through atmospheric turbulence. (b) The same PSF imaging through atmospheric turbulence using a Lyot Coronagraph, but also with an AO system running in closed-loop. Note that although light is returned to the diffraction rings at the center of the image, the outer region still displays the atmospheric speckles. This is because of the spatial frequency limited correction of the AO system, determined by the layout of the actuators of the DM. The AO correction reveals a low-order NCPA that is present, degrading the PSF quality. (c) The same simulation as (b) but without the NCPA. In all three cases, the red circle represents the location of a planetary signal that has been injected.

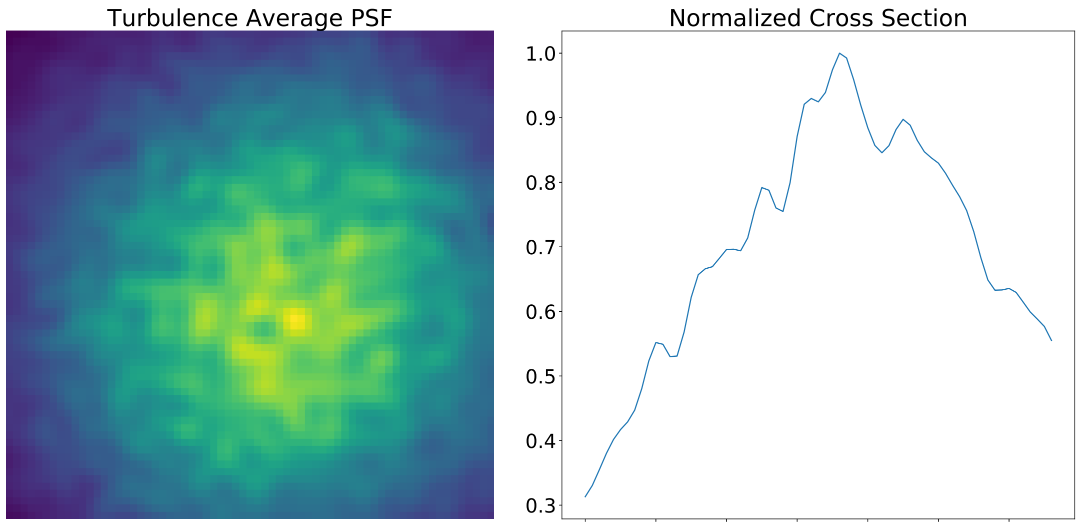


Figure 2.5: An example of a seeing–limited, long exposure PSF. In this case, $D/r_0 = 30$.

as predicted by the interpretation of r_0 above.

We can use a similar example to see the dependence on r_0. If r_0 is twice as a large, with respect to the same telescope diameter, the effect of the turbulence will be reduced, as given by Eq. (2.15). The long-exposure image and its cross-section for such a case is given in Figure 2.6. Here we clearly see that the light is not scattered as much, remaining more concentrated toward the center of the focal plane rather than in the halo. In fact, one can interpret this behavior as a means of thinking about the PSF averaged over turbulence. As the strength of the turbulence increases from nothing to very strong ($r_0 \gg D$ to $r_0 \ll D$), the diffraction limited PSF is present, but light from the core is progressively scattered into the halo until the halo becomes the dominant feature.

This drives us to define a useful figure of merit for the quality of a PSF: the Strehl ratio. The Strehl ratio is a measurement of the amount of light in the core of the PSF divided by the amount of light that would be in the core of the PSF if it were diffraction limited. In this figure of merit then, a value of 1 would mean that the PSF is diffraction limited and aberration free, as no light has been scattered out of the core, and any value in the range of $0 - 1$ will describe the severity of the effect

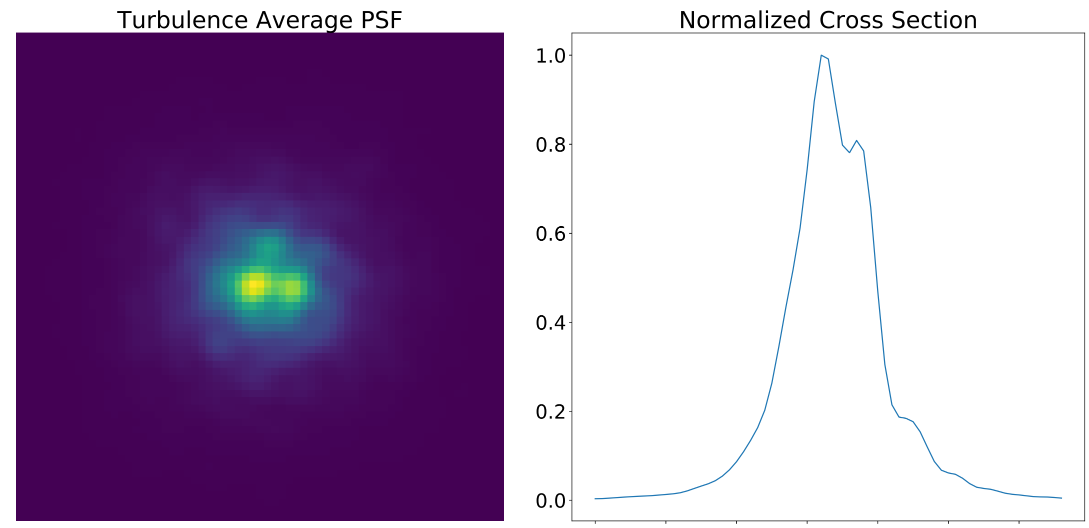


Figure 2.6: Another example of a seeing–limited, long exposure PSF, with $D/r_0 = 15$.

of the aberration on the PSF. The Maréchal Approximation to the Strehl is defined to directly relate the above quantity to the variance of the wavefront error, making it an easy way to quantify any methods trying to compensate wavefront error. The Maréchal Strehl is:

$$S = \exp\left(-\sigma_w^2\right) , \tag{2.17}$$

where σ^2 is the variance of the wavefront phase error.

Another way to examine the degradation of the PSF is to look at the corresponding OTF. As described in Section 2.1, the OTF is the spatial frequency spectrum of the PSF, which can tell us about how the optical system forming the PSF is functioning. Figure 2.7 shows the MTF for frame (a) of Figure 2.4 and the MTF for the same Lyot coronagraph if it were diffraction limited (no atmospheric turbulence). There is a stark difference between the two MTFs. The diffraction limited case shows structure that we are now familiar with, acting as a low–pass filter. The Lyot coronagraph does modify this from what was seen previous because it is designed to suppress starlight, so you would expect to see the dark rings rather than the smooth decrease from the center. However, with turbulence effecting the optical system, we see this structure is lost, and over the pictured range of spatial frequencies, the

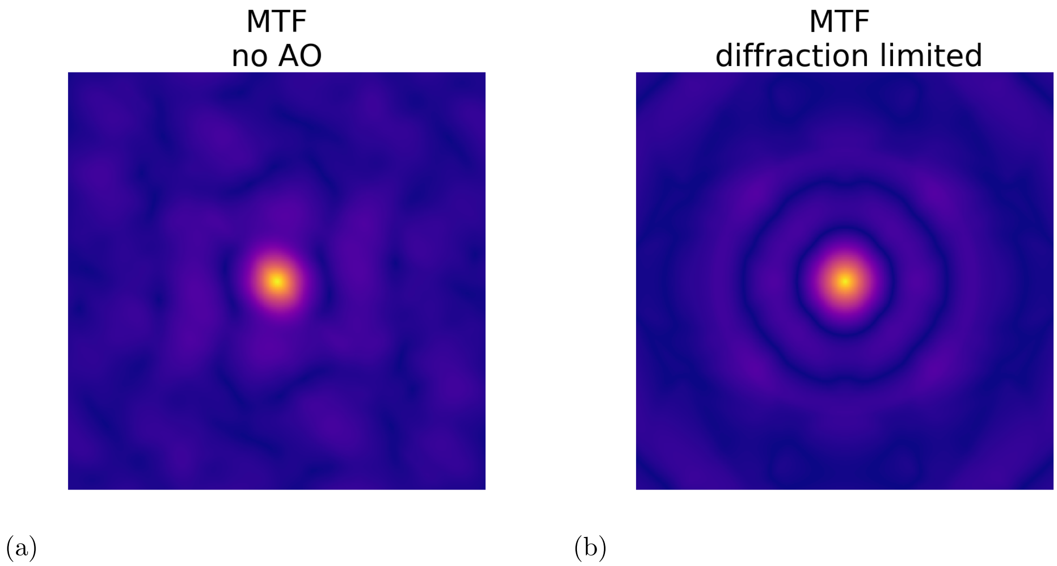


Figure 2.7: (a) The Modulation Transfer Function (or absolute value of the Optical Transfer Function) of a Lyot coronagraph in the presence of atmospheric turbulence (b) The Modulation Transfer Function of a diffraction limited Lyot coronagraph.

complex weighting factors applied are close to the same. This is another way of saying that the effect of the turbulence will cause the Lyot coronagraph, or any coronagraph for that matter, to no longer be able to suppress starlight as intended, disrupting hope of achieving the required contrast to directly image exoplanets.

2.2.3 Adaptive Optics

Now that we understand the effect of atmospheric turbulence on optical imaging through degradation of the PSF, we can talk about how to fight it: Adaptive Optics (AO). Simply put, an AO system is an optical system that's main purpose is to measure the wavefront passing through it to estimate any aberration that is present, and then compensate for it by way of a control system and some sort of active optical element. This means that a device called a wavefront sensor (WFS) is employed with an optical element that can change its shape in a predictable / controllable way such as a deformable mirror (DM) or spatial light modulator (SLM). The simulations of AO presented in this dissertation will stick to a WFS called the Pyramid

WFS [69], and largely to the use of a continuous face DM, so the discussion will be confined to these devices. The DM will be assumed to be like a Boston Micromachines (BMC) Kilo DM, or a DM with 1020 actuators placed in a 32×32 geometry. There are several key pieces to understand about the layout of the actuators: the interactuator spacing in combination with the number of actuators across the DM, and the actuator influence functions. The interactuator spacing is important because it defines the control radius of the AO system. This can be interpreted as such because adjacent actuators being pushed up and down respectively creates the maximum frequency sinusoid that can be represented by the DM surface, and thus is the highest correctable spatial frequency. Because the DM is conjugated to the pupil plane, the number of actuators can then be used to determine the control radius of the DM in the focal plane. For example, a DM that is 32×32 actuators results in a control radius of $16\lambda/D$ in both the x and y directions, resulting in a square region of the focal plane that can be compensated. The actuator influence function can be thought of as the shape that the actuator exerts on the membrane surface of the DM. In the case of the simulations presented, we represent the DM as a set of Bspline modes, with knots defining the actuators. In a real BMC-like DM, the influence functions can usually be well approximated by a Gaussian function. The extent to which the actuators couple with each other in fitting a surface is related to the overlap of the individual actuator influence functions. The effects of this will largely be assumed to be negligible. In Chapter 3, the simulations will deviate to the use of a segmented DM, in which the reflective surface is made up of 37 individual hexagons that are held by three points of contact, allowing them to piston, tip, and tilt individually from each other. This choice was made to demonstrate a specific use case of the method being examined, which does not include traditional AO for compensation of atmospheric wavefront errors. All simulations that do model AO in this manner assume the use of a BMC-like DM. The choice made to confine the AO simulations to a BMC-like DM and a Pyramid WFS is simply to narrow the focus of this dissertation, and in no way is stating that the work discussed is limited to this case.

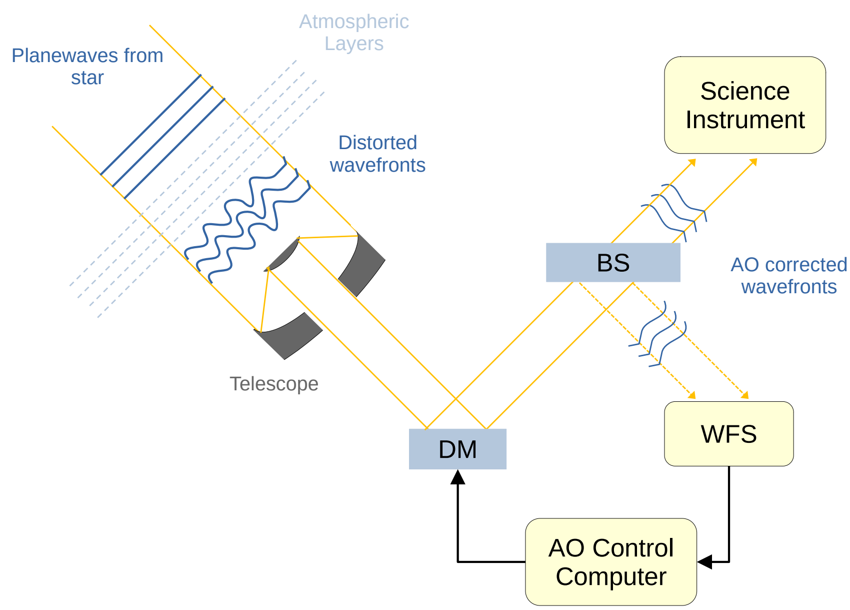


Figure 2.8: Schematic diagram of an astronomical telescope with a closed-loop AO system.

A ground–based AO enabled optical system essentially creates an instrument with two *arms*, or optical trains: the Wavefront Sensing and Control arm, and the Science arm, separated by a beam splitter or dichroic mirror. So, the light from the star travels through the atmosphere and becomes distorted, enters the telescope and propagates through the beam reducing optics, and then encounters the DM in a pupil plane conjugated to the telescope primary mirror. The light is then relayed to a beam splitting optic that separates the light in to the two arms, with some of the light going to the science instrument, and the rest going to the WFS. This layout can be seen in Figure 2.8. The AO control system is set up around the measurements (or estimates) of the wavefront made by the WFS. In many cases, this is done by calibrating a *reconstructor* matrix, in which known shapes are set on the surface of the DM (called modes), and their corresponding responses are measured by the WFS. A matrix is constructed as the known DM actuator commands for the modes against the WFS measurements (called the response matrix), and then is inverted so that it can be applied to a WFS measurement of an unknown wavefront to return an estimate of the DM commands that fit the modes to it. In this way, after calibration

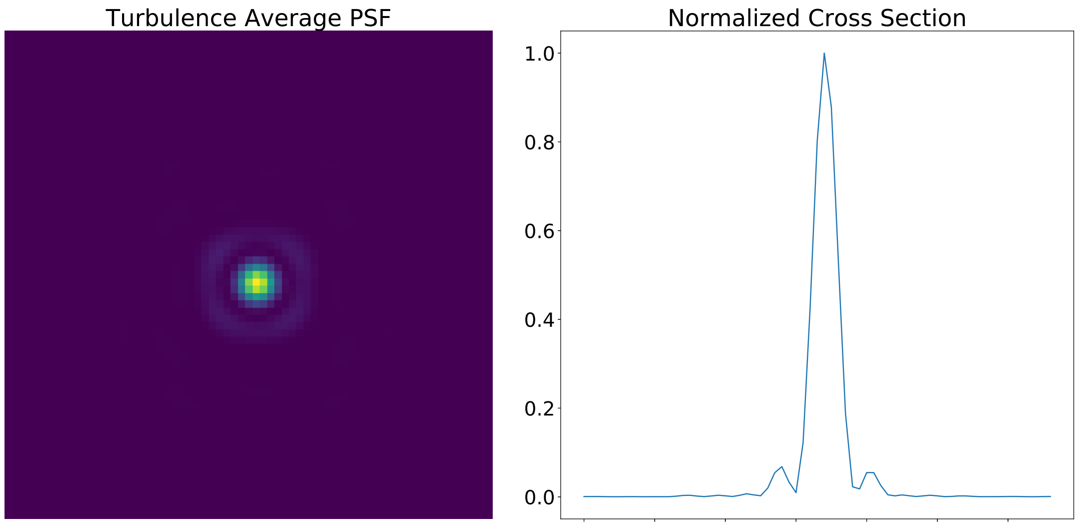


Figure 2.9: An example of imaging with the same optical system, but with AO turned on. In this case, the control radius is wider than the detector region, so no uncorrected halo can be seen.

of the reconstructor on known modes that the DM can control, a closed–loop can be set up on-sky to compensate the atmospheric turbulence by feeding the reconstructor WFS telemetry every 1ms (or faster), and using the estimated DM commands with a gain factor (called the AO gain) to continually reshape the DM surface to be the opposite of the measured aberration. The loop is considered closed because the WFS measures the wavefront downstream of the DM, meaning that it is sensing the wavefront after the compensation, and updating it to minimize the wavefront error in what is called the *AO residual* wavefront, which will be defined more carefully later in Chapter 4. Returning to the example used in this chapter, we turn on our simulated AO system with a Pyramid WFS and a 20×20 actuator DM, running in closed loop via a leaky integrator control system. In this case, we directly fit the DM to the estimated wavefront OPD from the Pyramid WFS rather than simulating a reconstructor matrix because of numerical expediency. The resulting PSF and its cross-section can be seen in Figure 2.9. Because the control radius of this simulated DM is larger than the detector region being shown, no residual uncorrected halo is visible, and instead it appears as a diffraction limited PSF.

Returning to Figure 2.4, we can see the improvement AO gives when used in combination with a Lyot coronagraph. This case will be the simulated telescope system for the work in Chapters 4 – 6; an AO system correcting for atmospheric turbulence using a Pyramid WFS and a 20×20 DM, with a Lyot coronagraph in the science arm. Because the frames in this figure represent an instantaneous PSF rather than averaging over turbulence, there is no spatially smooth halo, but we can still examine the resulting structures and behavior. With the 20×20 actuator DM, the control radius is $10\lambda/D$. This is visible in the lower frame of Figure 2.4, as we can see the light is pulled back into the central region and distributed as we expect for a nearly diffraction–limited Lyot coronagraph (seen in the right frame of Figure 1.3). In fact, we can even see the expected planetary signal in the red circle, so within the control radius, the achieved contrast of the Lyot coronagraph is restored. However, outside of $10\lambda/D$, we see the swarm of atmospheric speckles that will change rapidly with time, not following the expected structure from the Lyot coronagraph architecture. A simulated focal plane image using a similar (though not identical) optical system that is averaged over turbulence can be seen in the right hand frame of Figure 5.6. In this image, the control radius of the AO system is clearly visible as a square region in which the atmospheric speckles are largely compensated for, with a spatially smooth halo outside of it.

From all the figures presented showing the resulting focal plane following a working AO system, we can continue to understand the Strehl ratio metric defined in Eq. (2.17). As the AO system compensates for the atmospheric turbulence by fitting the DM surface to the measured / estimated wavefront by the WFS, more light is returned to the core, increasing the Strehl ratio closer to unity. Any aberration that is not corrected is considered to be the AO residual wavefront. The content of the AO residual is due to many possible factors (see Chapter 9 of Hardy (1998) for a thorough description), but the most common, and thus treated in the simulations that follow, are: content of spatial frequency greater than the DM can represent, error in the DM fit caused by the fact that the current wavefront at the DM is different than the one being represented by the DM surface because of the required time to measure the wavefront and update the DM commands, and error in the DM fit caused by the fact the WFS can only estimate the phase rather than perfectly measure it, meaning that what the DM is fitting to is subject to factors such as detector noise and measurement bandwidth limitations. All of these factors contribute to a residual, non-zero variance across the AO residual, which, from Eq. (2.17), tells us that we will likely never be able to achieve a Strehl of unity. In reality, for the observation wavelengths chosen for this work, reaching a Strehl between 0.7 and 0.9 will be considered realistic to excellent AO performance.

2.3 Non-common path aberrations

As we have discussed, and shown in Figure 2.8, the ground–based telescope instrument being used with an AO system has a common path up to a beam splitting optic, where two separate optical arms, the WFS and Control arm and the Science arm, are formed. The result of this setup is the so-called *quasi-static non-common path aberration* (NCPA) in the optical system hardware. The name non-common path derives from the fact that these aberrations occur in optical components that are beyond the beam splitting optic in the Science arm, meaning that the WFS is blind to them. The AO system itself is thus in charge of compensating the wavefront errors due to atmospheric turbulence. If designed well, the AO system will be able to flatten the wavefront across many spatial frequencies (largely up to the control radius of the DM), leaving behind only small errors in the AO residual (high spatial frequency content, DM fitting error caused by the fact that WFS measurements are estimates at best, etc). Because the WFS is estimating the wavefront phase in the WFS and Control arm of the instrument, any systematic aberrations (e.g. due to alignment, thermal flexure, gravity sag) within the common path and WFS will also be compensated for when the loop is closed, further optimizing the wavefront that is measured. However, the fact that the systematic aberrations of the WFS arm are compensated by the DM in the common path, their opposite influence will be shifted in to the science path. In addition to these systematic errors, NCPA from the optical elements in the science instrument will also build up. The temporal variability of the NCPA from any source results from thermal fluctuations and other variable conditions, such as the changes in gravitational stress as the telescope pointing varies. As a result of the ever-changing conditions, the NCPA on ground-based AO platforms change on a wide variety of time scales, ranging from minutes to months [30].

This is of particular interest to high-contrast imaging applications for several reasons. As we have discussed, on the ground, imaging exoplanets currently requires the combination of AO with a stellar coronagraph [31]. The combination of AO with a coronagraph compensates the swarms of rapidly changing atmospheric speckles in the coronagraph's science camera, and returns the performance back to the design capability of the coronagraph, allowing the suppression of starlight because it will no longer be scattered throughout the focal plane over the region within the control radius. The NCPA, though, will manifest themselves as *quasi-static speckles* in the science camera, and several authors have shown them to be a major limitation in high-contrast imaging [3; 54]. Simple, low-order NCPA redistribute the light, especially near a few λ/D. This reduces the ability of the coronagraph to reach the required contrast to directly image an exoplanet in regions around stars that are the most interesting for detecting habitable worlds (Figure 2.4 (b)). However, NCPA is not confined to low-order spatial frequencies. $\frac{1}{f^{\alpha}}$ type spectral error resulting from such things as beam walk can be much more complicated in how it distorts the wavefront. Furthermore, quasi-static speckles are particularly problematic because, unlike the speckles created by the atmospheric turbulence, they do not average to a spatially smooth halo that can be subtracted in post-processing in a relatively straightforward manner. The challenge of mitigating the quasi-static speckles is made further complicated by the fact that they are modulated at kHz rates by the residual atmospheric aberrations that the AO system is unable to correct. Some current methods with dealing with these problems will be presented in Section 2.4, with some discussion of their shortcomings that would be useful to overcome. Chapter 3 will summarize a technique developed in 2012 by Johanan Codona [8]called the *differential Optical Transfer Function wavefront sensor*, and how it can be employed to remove any systematic level, static (or seemingly static with lifetimes longer than observation times) NCPA. Finally, Chapters 4–6 will discuss a novel method now known as *Frazin's algorithm*, originally devised by Richard Frazin in 2013 [15], and evolved to become a technique that holds immense power in not only estimating quasi-static NCPA with lifetimes on the order of minutes, but also estimating the signal from exoplanets themselves, in real-time while an observation is taking place.

2.4 Summary of current techniques for NCPA estimation and exoplanet detection †

2.4.1 Differential Imaging

Currently quasi-static speckles are removed via background subtraction techniques that employ various types of *differential imaging*, the most common forms of which are *Angular differential imaging (ADI)* and *spectral differential imaging (SDI)*. SDI takes advantage of the fact that the point-spread function (PSF) stretches with wavelength (λ), while the location of a planet does not [77]. As this PSF stretching is proportional to the distance from the star, a weakness of this method is detecting targets at separations approaching the angular distance of λ/D, as well as targets that are extended in the radial direction, due to self-subtraction artifacts. Furthermore, SDI is also limited by chromatic optical path difference [49; 68; 70]. ADI takes advantage of the field rotation that occurs over the course of the night when the instrument rotator is turned off (Cassegrain focus) or adjusted to maintain fixed pupil orientation in a Nasmyth focus instrument, so that the planet appears to rotate relative to the PSF [50]. Thus, to the extent that aberrations do not evolve as the image rotates around the pointing center, correction can be achieved. Some known and characterized weaknesses of ADI are similar to SDI, in that star-planet separations close to λ/D are difficult to detect because the planet must travel an arc length of several resolution elements within the observing period [57]. Other

† This section has been published previously in Rodack et al. (2021)

effects caused by the time dependence of the planetary rotation rate around the star due to field rotation, as well as artifacts from self-subtraction, induce a host of biases that affect astrometry and photometry calculations [51; 21; 70]. ADI also will remove circularly symmetric components of the image, making the quality of the results dependent on the spatial arrangement of the planets and dust surrounding the target star. Furthermore, one of the largest issues with ADI is the fact that the quasi-static speckles change with time, which can greatly affect the subtraction accuracy.

2.4.2 Focal Plane Wavefront Sensing and NCPA Compensation

One way to eliminate the quasi-static speckles arising from the NCPA is to perform some type of focal-plane wavefront sensing with data obtained from the science camera, and then use a deformable mirror (DM) or other type of spatial-light modulator to apply a compensation to the wavefront [52; 53; 56; 39]. One way to do this is to apply two or more DM commands, called *probes*, and measure the intensity for each probe, thereby setting up a simple regression in which the complex-valued electric field at the locations of interest is estimated. Once this field has been estimated, the next step is to calculate a new DM command according to some merit function that creates a dark region. One such method is called *electric field conjugation (EFC)*, but there is now substantial literature on closely related methods [28; 52; 53; 56; 39; 67]. These methods are expected to be vastly more effective on space-based platforms than they have proven to be on ground-based platforms due to the Earth's atmospheric turbulence. Indeed, typical attempts to create an extended dark region on ground-based platforms only result in reducing the background starlight by a factor of about two. In addition, ground-based applications of such methods are quite slow, sometimes combating a few speckles at a time, and each DM probe must be remain in place long enough to obtain a turbulent average of the intensity [79; 53; 56]. The fundamental problem with ground-based probing methods and EFC (and other related methods) is that the focal plane electric field (which probing seeks to estimate and EFC seeks to cancel) is modulated by the atmosphere on the time-scale of a millisecond.

A new approach to estimating the NCPA is to replace the occulter (in the first focal plane of the coronagraph) with a phase mask, thereby turning the coronagraph into a Zernike wavefront sensor [82]. The Zernike wavefront sensor, combined with knowledge of the wavefront statistics and a coronagraph model, then provide sufficient information to estimate the NCPA.

Of the various approaches to determining the NCPA, the COFFEE algorithm of Refs. [62; 36], which is also referred to as "phase diversity," is most similar to the method in Ch. 4 – 6. In COFFEE, a series of DM commands (called "pokes") is applied and a long-exposure (of sufficient duration to average over the turbulence) image is acquired. Then, the NCPA coefficients are estimated via regression. The most important differences between method in this paper and COFFEE are:

- Our method method uses millisecond science camera telemetry, while COFFEE uses long (i.e., averaged over the turbulence) exposures.
- Our method uses the AO residuals (available from the millisecond WFS telemetry) as probes, not DM pokes to provide the information needed to determine the NCPA.
- Our method jointly estimates the NCPA and the exoplanet image, thereby treating their joint error statistics, while the COFFEE algorithm does not estimate the exoplanet image.

Note that both methods rely on coronagraph models and require knowledge of the 2nd order spatial statistics of the AO residual wavefronts.

2.4.3 Millisecond Exposures

Millisecond exposure times with the science camera are a new frontier in high-contrast imaging, and they are becoming observationally attractive due to a new generation of noiseless and nearly noiseless IR and near-IR detector arrays capable of millisecond read-out times [13; 59]. Millisecond exposures freeze the swarms of

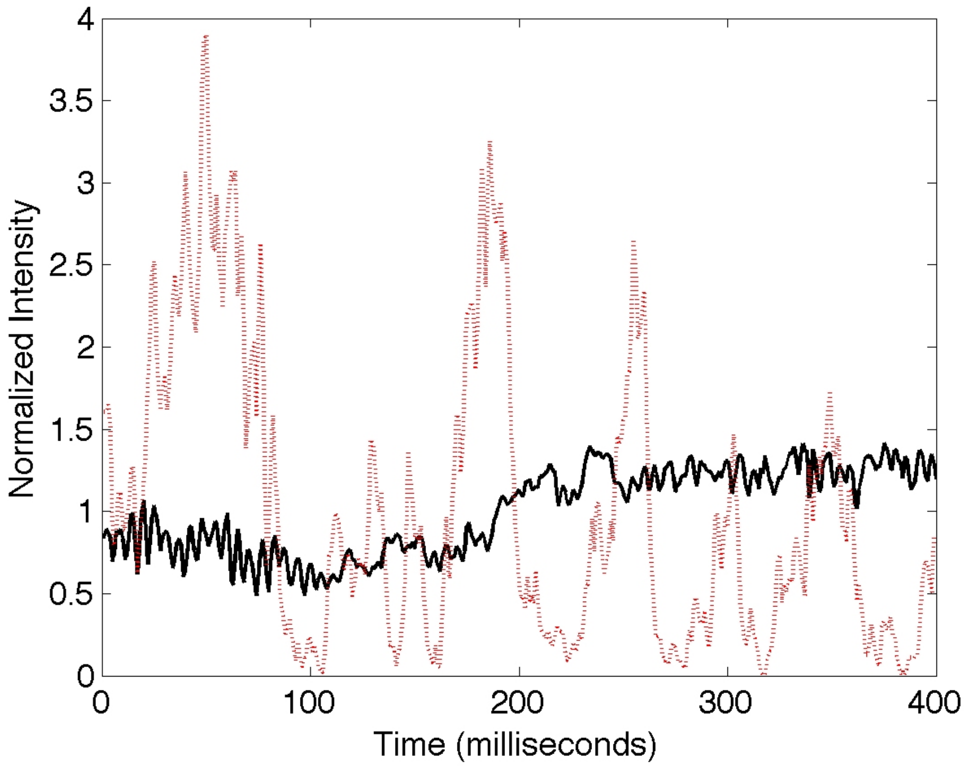


Figure 2.10: Simulation of turbulent modulation of intensities at a single pixel of the science camera from a stellar coronagraph. The black solid line shows the time-series of the temporal variation of the planetary intensity. The dotted red line shows the stellar intensity at the planet's location. Both the planetary and stellar intensity are normalized to have a mean of unity in this figure. From [15], used with permission.

speckles that arise due to atmospheric turbulence, whereas longer exposures average these speckles into a smooth halo. A planet and the stellar speckle exhibit radically different behavior at millisecond time-scales, as shown in Fig. 2.10 [15]. This is because at the center of a planet's speckle pattern, the AO system maintains the planet's intensity at a nearly constant level (given by the Strehl ratio), while the starlight at the planet's location exhibits much more volatility [14; 23]. Several methods have been proposed to exploit this difference in temporal behavior based on millisecond science camera time-series data alone [78; 83].

Simple physical optics arguments show that the speckles caused by the NCPA are modulated by the AO residual at the kHz time scale,Sauvage et al. (2010) and

Frazin showed that knowing the values of the AO residuals allows joint estimation of the NCPA and the exoplanet image from the millisecond science camera images [15]. Fig. 4.2 shows simulations of noise-free science camera images with the same NCPA being probed with different AO residual wavefronts demonstrating this phenomenon. In 2013, independent publications by Frazin and Codona & Kenworthy proposed to exploit the WFS telemetry in addition to the millisecond science camera images for focal-plane wavefront sensing [Codona and Kenworthy (2013); Frazin (2013)]. The fundamental shortcoming of the methods of Frazin and Codona & Kenworthy is that they were unable to account for wavefront measurement error (WME). This is to say, the WFS measurements only allow an imperfect estimate of the phase of the wavefront impinging on the WFS entrance pupil, not the actual phase. Specifically, any WFS exhibits spatial and temporal bandwidth limitations (the latter are less important due to the kHz frame-rate), nonlinearity in the phase of the wavefront, and noise. Outside of the high-contrast imaging problem, several authors have performed multi-frame deconvolution with millisecond focal plane and WFS telemetry to remove atmospheric image distortion. These works use the WFS-based estimates of the point-spread function (PSF) as an initial guess: Fusco et al. (1999); Chu et al. (2013). (For approaches that do not make use of WFS telemetry see, e.g., Refs. [76; 12]). The work of Frazin in 2013 will be expanded upon in Chapter 4, and serve as the launching point for the method described in this dissertation named after Frazin.

CHAPTER 3

The differential Optical Transfer Function wavefront sensor

3.1 Mathematical development, interpretation, and implementation

Now that we have obtained a basic understanding of the typical ground-based, adaptive optics (AO) system, and the idea of non-common path aberrations (NCPA), we can continue on the journey of this dissertation, and start to unwrap some methods that provide the ability to measure a wavefront disturbance, and thus inform us how to compensate for it via a deformable mirror (DM) to remove its influence on our ability to study exoplanets. The first such method that will be considered in this work is the so called *differential Optical Transfer Function wavefront sensor*, or dOTF for short, which is among the family of techniques known as phase retrieval [24]. There is a 10 year or so history now of dOTF, an idea first conceptualized and described by Johanan Codona [8] in 2012 , with further detail and potential application of the method expanded upon in works such as: Hart and Codona (2012), Codona (2013), Codona and Doble (2015), Rodack et al. (2015), Knight et al. (2015), Brooks et al. (2016), and Jiang et al. (2019). Rather than bog the reader down with all these details, we begin by summarizing the seminal work of Codona [8]. dOTF is a non-interferometric, non-iterative method to estimate the complex amplitude field (both amplitude and phase) in the pupil of an optical imaging system, without requiring any specialized hardware or sophisticated post-processing techniques. What makes it of particular use to us is its robustness to many types of aberrations that arise from misalignment and vignetting, while remaining in the same pixel scale as the optical system being tested. This allows for the method to be used via a camera in the science arm, downstream from the AO system WFS, to detect and estimate any static NCPA caused by these effects.

Returning to the description for the Optical Transfer Function (OTF) given in

Chapter 2.1, specifically to Eq. (2.9), we are reminded that the OTF is defined as the Fourier Transform of the PSF, with the PSF defined as the modulus squared of the field from a point source, $U(\xi,\eta)$, impinging upon a detector in a focal plane conjugated to the entrance pupil. Recalling that the PSF is real by definition, the resulting Fourier Transform into the OTF space will be complex, and have Hermitian symmetry. The importance of this point will come up in a moment. Furthermore, in this analysis, unless otherwise noted, the field $U(\xi,\eta)$ will also be taken to be monochromatic. If we examine the process of taking the Fourier Transform of the PSF, we can notice an alternative way to write the problem:

$$PSF = |U(\xi,\eta)|^2 = U(\xi,\eta)U^*(\xi,\eta)\,. \tag{3.1}$$

In this format, as the product of the field in the focal plane with its own conjugate, we can understand that the OTF can then also be defined, through application of the Convolution Theorem, as the convolution of Fourier Transforms of the field and its conjugate:

$$OTF = \mathcal{F}\left\{U(\xi,\eta)U^*(\xi,\eta)\right\} = \mathcal{F}\left\{U(\xi,\eta)\right\} \circledast \mathcal{F}\left\{U^*(\xi,\eta)\right\} = u(x,y) \circledast u^*(x,y)\,. \tag{3.2}$$

Now, with the OTF in these terms, we can see that it is in fact the autoconvolution of the field in the exit pupil of the optical system, or in other words, a quadratic representation of it. Given this, we can start to think about how we could retrieve information about the complex field itself. The simple solution here is to remove the quadratic nature of this quantity by constructing its functional derivative, leaving us able to estimate the complex field.

In order to construct the functional derivative of the OTF, a small change is introduced in a pupil plane conjugated to the exit pupil. Although there are several options for this change in practice, we will describe the underlying mathematics prior to discussing best practices. By introducing the modification in the pupil field, du:

$$u(x,y) \rightarrow u(x,y) + du(x,y)\,, \tag{3.3}$$

a corresponding change in the OTF happens, which can be written as:

$$\begin{aligned} d\mathscr{H}(k_x, k_y) &= \mathscr{H}_{0+d0}(k_k, k_y) - \mathscr{H}_0(k_x, k_y) \\ &= (u(x,y) + du(x,y)) \circledast (u(x,y) + du(x,y))^* - u(x,y) \circledast u^*(x,y) \\ &= u \circledast du^* + du \circledast u^* + du \circledast du^* \,, \end{aligned} \tag{3.4}$$

where the coordinate labels are dropped in the final line for clarity. Before we continue, it is important to notice here that each term in the dOTF contains a convolution with the pupil modification du that was chosen. Now, from the fact that the OTF has Hermitian symmetry, we can simplify this to:

$$d\mathscr{H}(k_x, k_y) = d\mathscr{H}_+(k_x, k_y) + d\mathscr{H}_+^*(-k_x, -k_y) + d\mathscr{H}_{\delta\delta}(k_x, k_y) \,, \tag{3.5}$$

which defines the dOTF as the sum of $d\mathscr{H}_+(k_x, k_y)$, the image of the pupil field we would like to be able to measure convolved with the pupil modification, its reflected conjugate, and a quadratic term $d\mathscr{H}_{\delta\delta}(k_x, k_y)$.

Eq. (3.5) quite clearly has three separate terms, each of which represent one of the three main regions within a dOTF signal:

1. The field in the pupil region (what we set out to be able to measure).

2. The conjugate of the field in the pupil region reflected about the modification made in the pupil (that includes a region of overlap between the two pupil regions that depends on the placement of the modification).

3. The localized contribution of the quadratic term related to the autoconvolution of the modification.

A cartoon example of the form of the dOTF is given for a pupil modification localized to the edge of the pupil in Fig. 3.1.

Now that we understand the definition of the dOTF and how it gives us access to measure the complex field in the exit pupil, we can describe how one would implement this technique in reality as a wavefront sensor, and why in an AO system specifically, it could be used to calibrate out certain NCPA prior to conducting

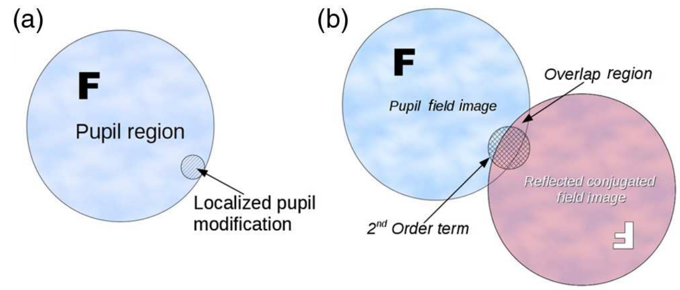


Figure 3.1: *Courtesy of Johanan Codona.* (a) Image of the field in the pupil conjugated to the exit pupil in which a modification is introduced. (b) Schematic of the dOTF, with three distinct regions: The pupil field image, the conjugate of the pupil field image reflected about the pupil modification leading to an overlap region between the them, and the location of the quadratic term related to the autoconvolution of the modification.

observations. It should be noted here that this technique is not a viable wavefront sensor for closing the AO loop on-sky largely due to the fact that it requires a large number of photons to obtain a useful signal. This, and other potential shortcomings of the method will be discussed in Section 3.4. Fig. 3.2 shows a schematic outline of how to measure the dOTF in practice. A monochromatic point source, typically a laser, is observed with the optical systems under test. The first step is simple; allowing the light to propagate through the optical system to the final focal plane, where the intensity is measured with a camera. This intensity, representing the "nominal" state of the optical system, will be labeled PSF_0. It is important to take in to consideration the measurement of this nominal PSF. In order to obtain a high quality dOTF measurement, as would be expected, the number of photons in each PSF exposure must be sufficient for the dOTF signal to dominate the sources of noise. This will mean, as this will often be used in a lab setting, that photons will not be scarce, and allow for multiple short exposures to measured. Care must be taken in this regard though, for several reasons. First, one must ensure the stability of the system is high, as large changes over the total exposure time measuring the nominal state PSF will corrupt the signal, as there will be a temporal component to the state. This limits the total exposure time that can be chosen to some maximum value before the PSF is unstable, and varies from system to system, and environment to environment. Factors that limit the stability can range from environmental factors

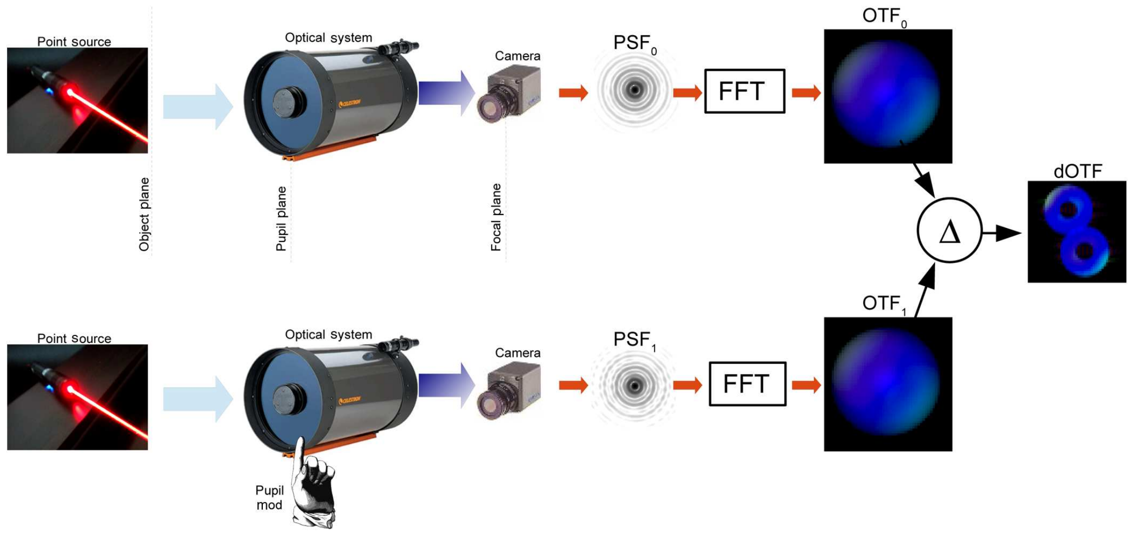


Figure 3.2: *Courtesy of Johanan Codona.* A schematic outline of implementing dOTF.

such as air currents, to vibrations in the optical bench. Second, while the system is stable, detector saturation must be strictly avoided. If the number of photons exceeds the well depth of the detector pixels, not only will the dynamic range of the measurement be corrupted, but electron bleeding to other pixels can occur, both of which will invalidate the required Fourier Transform of the measured PSF. Practically, this can be solved through the use of photon flux filters and shorter exposure times over the same total observation time. Additionally, one can add a small amount of defocus by physically shifting the detector at the end of the optical system out of the true focal plane. Although this will require a quadratic phase profile in the dOTF to be estimated and removed prior to analyzing the estimated complex amplitude field, it is an extremely effective solution because it reduces the Strehl ratio, meaning that it redistributes the photons out of the bright PSF core to other parts of the detector, allowing more light to be collected across the sensor prior to saturation. All of these factors are discussed more thoroughly in Codona (2013).

Next, a localized modification is made to the field in a pupil plane near the beginning of the optical system. Note the entrance pupil is the best choice for probing aberration content in the entire optical train, but a conjugate pupil slightly downstream of the entrance pupil is a fine choice. There are several trade-off parameters for this pupil modification including: its shape/size, location in the pupil, or whether it is done in amplitude or phase. This allows for tremendous freedom in the pupil modification implementation, but there are some general guidelines to follow:

1. As Eq. (3.4) tells us, the complex field quantity we would like to estimate is convolved with the pupil modification introduced. This means that the best course of action is to choose a modification that is both small and compact. This minimizes the blur it introduces when convolved with the pupil field, making for a more accurate estimate. If the application is going to require high spatial frequency estimates of the NCPA, understanding the form of the modification to a high degree will help with performing deconvolution techniques.

2. As we see in Fig. 3.1 and in Eq. (3.5), the complex amplitude field dOTF estimates, and the conjugate of it that is reflected about the pupil modification, overlap. This overlap region corrupts the estimate, making it unusable there. But its dependence on the pupil modification makes for a simple choice: placing the modification near the edge of the pupil. This serves to minimize the overlap region, and preserve the majority of the estimate of the pupil field. In addition, the entire pupil field can be estimated by using two separate modifications to measure the dOTF twice, and stitching together the results of the two estimates.

3. When choosing between an amplitude or phase modification, it's largely a consideration of what is most reliable to implement. An amplitude modification consists of an opaque material that can be slid in and out of the pupil to block light. If the mechanical requirement of this can be fit in to the optical system, and is repeatable and stable, it is a fine choice. A phase modification likely requires there to be an active optical element in a pupil plane conjugated to the exit pupil, such as a DM or Spatial Light Modulator (SLM). In the use case presented here, as a means of estimating NCPA in the science arm of an

AO system, this is the most practical implementation via bumping an actuator on a DM. This however has some consequences related to the first item in this list. Rather than being a simple, mathematically real modification, bumping an actuator induces a complex one, leading to much more complicated behavior with the convolution. If the maximum phase shift of the actuator poke exceeds π, the shape of the modification, and thus convolution kernel, will inherit an oscillatory behavior one would expect from an aberration with increasing OPD (as this is exactly what a strong actuator poke in the pupil is doing). This will not only harm the estimates at the edge of the pupil, but will also effect the mean value of the kernel, reducing the dOTF magnitude and lowering sensitivity. A trade study conducted by Codona [9] found that an actuator poke of $\lambda/4$ or less, with λ being the wavelength, is optimal.

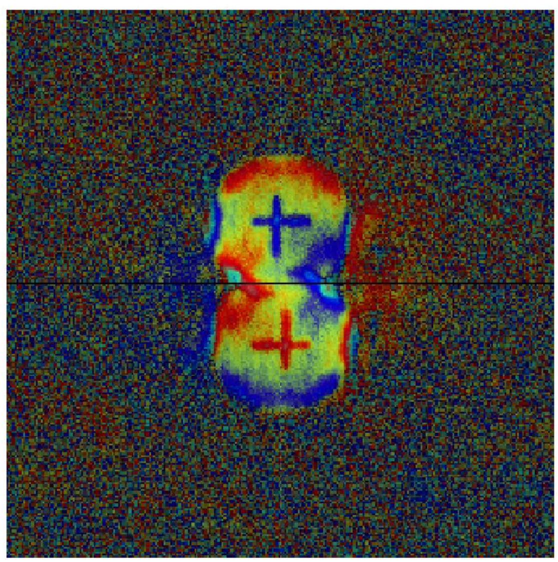


Figure 3.3: An example of a high quality dOTF signal measured in the lab

Once the form of the modification is chosen, it is applied to the pupil plane, and the intensity impinging on the detector in the final focal plane is again recorded, and labeled as the "modified PSF", or PSF_1. If there is a departure in stability, including that of a rapidly changing DM performing AO corrections, that leads to a broadly distributed modification over the pupil, the Fourier Transforms of PSF_0 and PSF_1 will not subtract properly leaving a residual that dominates the dOTF signal and renders it unusable. However, we assume that the system meets the stability requirements, and both the nominal and modified PSFs are measured with an adequate number of photons without saturation, getting to the dOTF signal is simple:

1. The Fourier Transform, likely through the Fast Fourier Transform (FFT), of the measured PSFs, PSF_0 and PSF_1, are performed, to obtain the nominal and modified OTFs, OTF_0 and OTF_1 respectively.

2. OTF_1 is subtracted from OTF_0

The result of this calculation is the dOTF signal. An example of a high quality dOTF signal is provided in Fig. 3.3. This procedure can then be applied to a number of potential useful applications, a few of which will be outlined below.

3.2 dOTF self – calibration

The first application for utilizing dOTF is to use it as a wavefront sensor for a “self–calibration” control loop. This is to say that once an optical system is aligned, the clock starts for unwanted aberration to creep in. In the case of an AO system, closing the loop on the AO WFS will flatten the wavefront upstream of the beam splitter, but likely push some aberration into the science arm. In addition, slight misalignment of optics and/or the relentless tug of gravity can lead to effects in the science arm will lead to low–order aberrations like astigmatism, trefoil, coma, defocus, and others being introduced. Because they are downstream of the AO WFS, we can consider them to be NCPA, and seek to correct them. We will also assume that they NCPA that remain static over the period of time we want to do an experiment. One such way to accomplish this is to use dOTF. The advantage to this approach is that it costs no extra time in setting up hardware; it is simply identifying an actuator located at the edge of the pupil, and taking PSFs following the above prescription to compute the dOTF. Post–processing is minimal, with only requiring a calibration step to register DM actuator locations in dOTF space. This registration is also easy to do, placing a known test pattern on the DM such as the one pictured in Fig. 3.3, and backing out the actuator locations in a dOTF estimate. Because the actuators in the test pattern are known, their spacing in dOTF space can be determined, and that can then be used to extrapolate all the actuator locations in the pupil. With the registration complete, one need only compute the dOTF with measured PSFs on the science camera using a DM in a pupil plane in the science arm to add the modification, and then read OPL values directly from the dOTF estimate into DM commands, because what is estimated is the wavefront piston across the pupil. A similar method is possible with a segmented DM, placing a test pattern of segment pistons, and extrapolating the location of each segment. However, in this case, further steps must be taken to compute the segment tip and tilt values in addition to the piston values. Both of these cases are examined below in simulation, with a lab demonstration using a continuous face sheet DM.

3.2.1 Segmented mirror correction in simulation

An intriguing use of dOTF self–calibration is to cophase a segmented aperture. In this section, we will demonstrate cophasing segments of an IrisAO–like DM to calibrate the effects of a typical Kolmogorov–like wavefront aberration via simulation. It is important for the reader to note that although a Kolmogorov phase screen is being used to demonstrate the ability to calibrate out an aberration with a wide range of spatial frequency content, this method is strictly for analyzing static aberration, and not doing AO correction.

The first problem to tackle to accomplish this task is to devise a method to estimate the slopes of the wavefront local to each segment in the DM. Ideally this would be done as a numerical process on the dOTF itself, directly constructing the wavefront slopes with only knowledge of the segment pixels. To do this, we make two copies of the complex amplitude field estimated by dOTF, and shift them apart either left-to-right (to measure the x–slopes) or down-to-up (to measure the y–slopes). The copy of the signal that is shifted in the negative direction (left or down) is also complex conjugated. The number of pixels in the shifts is a free–

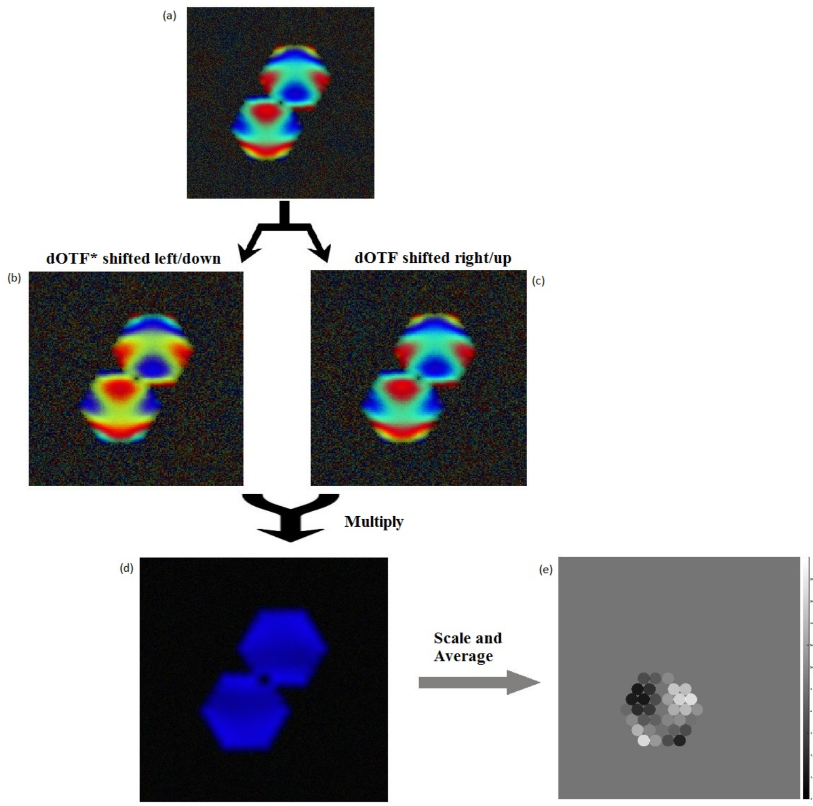


Figure 3.4: The method for constructing the wavefront slopes directly from a dOTF estimate. (a) An example dOTF estimate of a wavefront with a low–order aberration on a segmented pupil. (b) A copy of (a), shifted to the left by one pixel and conjugated. (c) A copy of (a) shifted to the right by one pixel. (d) The constructed wavefront gradient, scaled by the inverse of the wave number, the number of pixels in the shift, and the plate scale. (e) The argument of the complex amplitude gradients averaged over a region representing each segment.

parameter that can be optimized to get a better slope signal, as the length of the shift determines the constructed gradient in the wavefront, but the noise remains as the value of one pixel. However, the shift size should be limited to less than the pixel–width of the individual segments in dOTF space because the area of interest over which the slopes can be compensated for is limited to the size of an individual segment. It is also possible that, should the slopes be sufficiently steep, they may phase wrap. This can be mitigated somewhat by using a smaller pixel shift. Now, to construct the gradients themselves, the two shifted copies (left/down, conjugated and right/up) are multiplied together, and properly scaled. This scaling is done by dividing by the wave number, the total number of pixels in the shift, and the physical distance the pixel-to-pixel spacing represents (the real–world distance the pixel–based actuator spacing in dOTF space represents). With the wavefront gradients now constructed, the registration map discussed above is used to partition the dOTF estimate into the individual controllable regions (each segment), and read out values for the wavefront tip and tilt. This process is illustrated for the x–slopes in Fig. 3.4. The original dOTF estimate is then used with the registration map to read out segment pistons, and the typical piston–tip–tilt control of the segmented DM is fully specified.

With the estimate of the piston, tip, and tilt of the wavefront at each segment location, a simple integrator control loop can be set up to compensate the error. In doing this, the estimate is negated, and multiplied by a gain factor to ensure stable convergence. This gain value can be found experimentally by examining the convergence of the NCPA compensation, but simulations show that 0.3 is a conservative value that performs quite well. Fig. 3.5 shows frames from a simulation of this experiment. On the top left, the diffraction limited PSF is provided for comparison with the bottom left frame, which is the PSF after the conclusion of the self–calibration control loop. The image in the top row, second from the left, shows what the PSF would look like if no compensation of the NCPA injected (seen in the bottom right) is attained. The remaining two frames in the top row show the phase imparted to the field by the segmented DM, and the complex field after the DM, which shows clearly that field is flattened (the amplitude of the field is constant, shown via the brightness, and the phase is nearly constant aside from very high spatial frequencies that are not correctable, shown via the solid red color). The fact that there are no segments that are floating in piston demonstrate that the segments were in fact cophased using dOTF. Finally, in the bottom row are frames showing the current dOTF estimate, also demonstrating that the field has been flattened, and a plot of the Maréchal Strehl vs. Loop index, showing the self–calibration loop was stable, and recovered the Strehl from about 0.4 to 0.95.

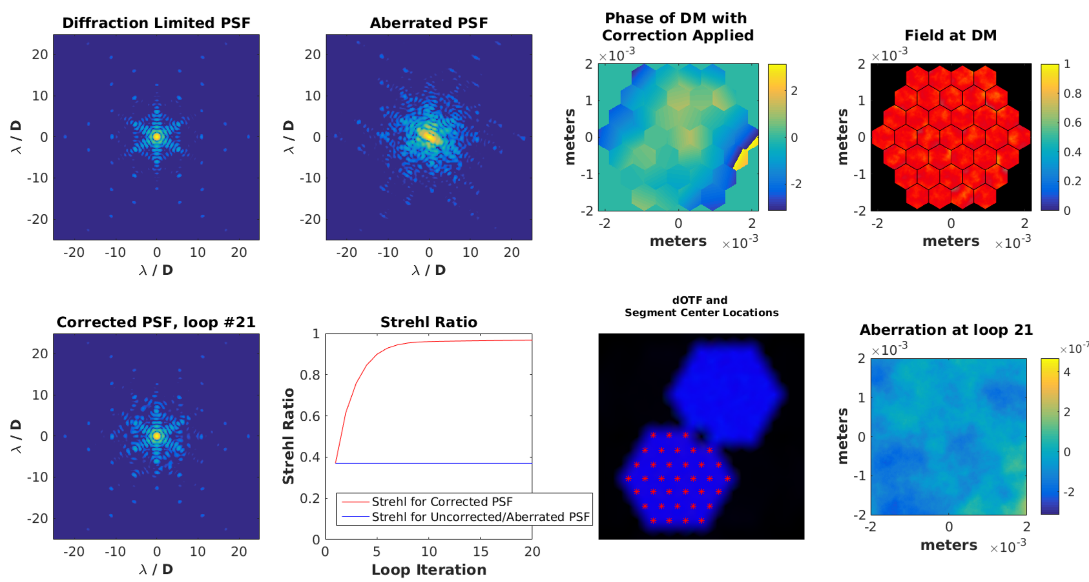


Figure 3.5: Simulated results using dOTF to flatten a static aberration with a segmented DM.

3.2.2 Continuous mirror correction in simulation

Next, we want take the same self–calibration control loop idea, and attempt to demonstrate it in the lab, where the optical system to use has a continuous face sheet, 1020 actuator DM rather than a segmented one. In order to proceed, we will first attempt to conduct this test in simulation, this time including noise in the detector to better understand the role it will play (note the prescription given in Section 3.1 to maximize the dOTF signal was born from these simulations and the lab work that followed). This is seemingly an easier problem than the segmented pupil because we do not need to worry about the wavefront slopes, and can instead solely rely on the piston estimates dOTF gives directly. There is a slight complication that arises in the registration of the actuators to dOTF space simply because there are 1000 of them (although practically it is fewer as some are not illuminated), but the method described above still performs adequately. In the same setup as used in the previous subsection, the segmented DM is replaced with a model of the continuous face sheet DM, and the simulation is rerun, except the measured PSFs are subject to photon and readout noise. To facilitate a meaningful photon noise component, the following statements are adopted for the simulated HeNe point source:

1. A detector quantum efficiency, η, of 50% is assumed
2. A bandpass of 0.01 microns is assumed
3. An exposure time, ΔT, of 0.200 seconds is chosen
4. A band flux of 9.97×10^{10} ph $\times$ μm^{-1} $\times$ m^{-2} $\times$ s^{-1}, corresponding to the V–band wavelength of 632nm, is adopted
5. A visual magnitude, m_v, of 6 is assumed for the source brightness
6. A telescope diameter of 6.5 m is chosen

These are then used to compute the total number of photons per exposure via the equation:

$$N_0 = \eta \times \text{Bandpass} \times \Delta T \times \text{Band Flux} \times (2.512^{-m_v}) \times ((\pi D^2_{Telescope})/4)\,. \quad (3.6)$$

Although a bandpass is given in this equation, it is strictly to give meaning to a number of photons being present. The calculations of the dOTF estimates that follow for this simulation are assumed to be monochromatic for reasons that will be discussed in Section 3.4. The noise in the intensity will assumed to follow a Normal distribution as usual.

As before, a Kolmogorov phase screen is injected into the entrance pupil of the modeled optical system to test flattening a complicated wavefront containing various spatial frequencies. An integrator control loop with a gain value of 0.3 is set up using the dOTF piston estimates as the updating DM commands. Figure 3.6 shows the results from the end of the simulation. Similarly to the segmented simulation, the

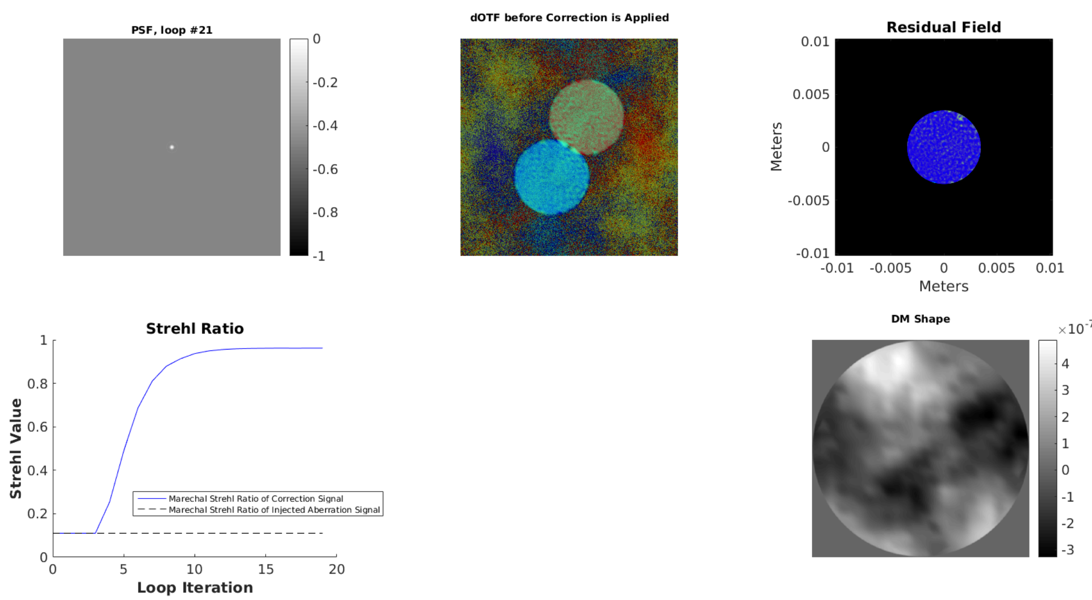


Figure 3.6: Simulated results using dOTF to flatten a static aberration with a continuous face DM.

top row contains the noisy PSF after compensation in log scale, the current dOTF estimate of the field, and the residual field in the exit pupil downstream of the DM. We see here that the self–calibration loop has again flattened the wavefront, achieving a nearly constant phase and amplitude across the pupil, with the departure being due to spatial frequencies higher than the interactuator spacing of the DM. The bottom row of Figure 3.6 contains an image of the current shape of the DM surface on the right, and a plot of the Maréchal Strehl vs. Control loop index on the left. From this plot, we see that the control loop for this calibration was also stable, and drove the Strehl from 0.11 to 0.96, even in the presence of photon and readout noise in the detector. Although not pictured, more simulations were conducted varying the parameters controlling the number of photons. From this work, the pictured dOTF estimates, which demonstrate roughly a dOTF signal of approximately three times higher than the noise floor, is a good rule of thumb to keep the control loop stable. This of course depends on the measurement of the intensity, which then couples into the dOTF signal through the calculation of the Fourier Transform and the difference between the OTFs. Given the guidelines above for how to increase the signal in the PSF measurements, and the fact that this method is largely confined to experiments in highly stable environments (also discussed above), getting enough photons in the PSFs that are Fourier Transformed should rarely be a limiting case.

3.2.3 Flattening a wavefront in lab

With all of this in mind, we now set out to demonstrate this on hardware, using the original optical system in the UA Wavefront Control Lab [60]. A HeNe laser source is used with a spatial filter / beam expander to achieve the point source, and an optical system containing eight lenses, three flat mirrors, three off-axis parabolic (OAP) mirrors, four glass plates, and a Boston Micromachines Kilo DM (32×32 actuator geometry). The optical system is aligned carefully to avoid any blatant transmission aberrations, but it is found that the major aberration that is imparted to the wavefront propagating through the field is from the uncalibrated (at the time), unpowered surface of the Kilo DM. Other contributions to the overall aberration content of the beam appear as astigmatism and coma from slight misalignment of the OAPs. In addition, a slight defocus is induced by shifting the imaging camera out of the focal plane using a micrometer.

The pattern in Figure 3.3 is the one applied to the Kilo DM to register the illuminated actuators to the dOTF estimate space. This is done by knowing the actuators on the DM that correspond to the asymmetric cross pattern. The number of pixels in each cross line is divided by the number of actuators the line represents to determine the pixel to actuator spacing. By utilizing both lines in the cross, and several dOTF estimates of the test pattern, the average pixel spacing is determined, and used to extrapolate to the locations of the illuminated actuators in the 32×32 DM geometry.

With the actuators registered, the next step was to experimentally determine how to get a high quality dOTF signal. This was done by setting a constant exposure time of 1/30 second, and averaging together a number of PSF frames ranging from $10 - 300$ for each of the nominal and modified. For each PSF pair, the dOTF was computed, and inspected by eye for quality of signal. After this process, it was determined that using averages of 90 frames of the PSF was sufficient to get enough photons that the dOTF would provide a high enough quality estimate for a control loop with a conservative gain to be stable. This number of frames, for the optical system under test, could have been further optimized to increase the servo speed, but as this test was simply to demonstrate a self–calibration loop, better signal was chosen to be more important than speed for calibrating a static aberration. The PSF at the start of the control loop can be seen in the left hand frame of Figure 3.7, displaying the nasty aberration from the unpowered DM, astigmatism, coma, and defocus mentioned above. The control loop using dOTF to read out estimates to send as DM commands is allowed to run for 9 steps, at roughly 30 seconds per step (6 seconds taking images and 24 seconds performing the numerical processing and sending the update to the DM). This numerical processing includes performing a shift and add scheme to the 90 PSFs measured per modality (nominal and modified) to get a high quality average PSF, cropping the detector region to a square centered on the PSF core, doing the FFT, and finally the difference between the numerical OTFs. The final result after the calibration loop is shown in the right hand frame of Figure 3.7. It is clear that the majority of the defocus is successfully removed, and the by eye quality of PSF is greatly improved (seen especially in the fact that the noise floor of the image is much darker because more of the light is concentrated in the core, suggesting a significant improvement in Strehl). There are two remaining spikes left, which come from effects that could largely be removed with more care. This includes a slight misregistration of the actuators to their pupil locations in the dOTF estimate, and the fact that we simply ignored the data from the overlap region without taking any steps to recover information from there (via a second dOTF estimate to stitch together). It so happened that the

Kilo DM had a large aberration in its surface corresponding to the location of the overlap region, so ignoring it left a fair amount of uncorrected OPD. However, with that the residual error that was not corrected being explainable, we consider this a successful demonstration of using dOTF to compensate for static aberration in an optical system.

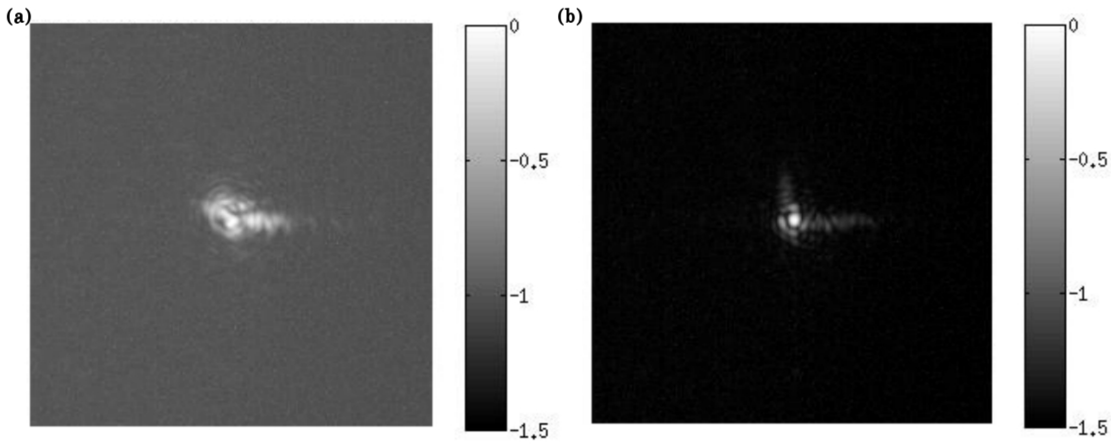


Figure 3.7: Before and after log10 scaled detector images of the final focal plane of a laboratory optical system.

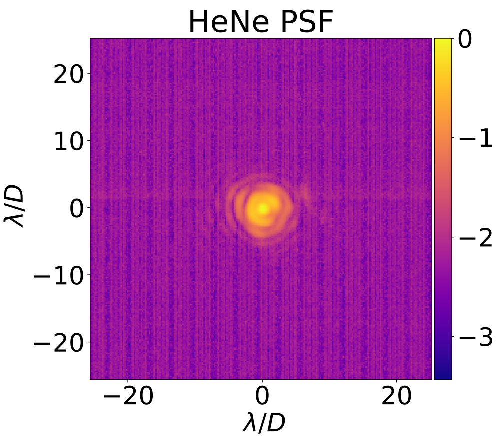
3.3 Measuring on-axis NCPA using an off-axis source

The next application of using dOTF is an experiment to determine if an on-axis NCPA at one wavelength (perhaps an observation of a HeNe laser) can be successfully recovered using dOTF estimates made using an off-axis beam at a separate wavelength (a green laser). To answer this question, the Comprehensive Adaptive Optics and Coronagraph Test Instrument (CACTI) at the University of Arizona is used, configured with a Kilo DM and a dichroic bypass mirror. In this configuration, a camera can be placed in a focal plane to detect two separate sources, an on-axis HeNe laser, and an off-axis green laser dubbed the Laser Guide Star (LGS) source. To facilitate the experiments, the beam footprint of each source on the Kilo DM is estimated to locate actuators near the edge of the illuminated pupil. This was done in a brute force manner of simply computing dOTF estimates by inducing the modification with targeted guessing of actuators until one at the edge of the illuminated pupil was located. Although this method works, it is far from efficient or the best means to locate such an actuator. Next, the exposure times were calibrated for each source so that the measured PSFs would not be saturated on the camera. For the HeNe source, 1ms exposure times was determined to be sufficient, and 3ms for the LGS source. Although these choices did not saturate the camera, the signal was still much too small to get a useful dOTF estimate. In order to gather more photons without saturating, 10 epochs of 250 exposures of the PSF were taken for each dOTF modality (nominal and modified), for each source, for each test case. This means that the PSF that is numerically processed is the average of 2500 individual frames. An example of such an average PSF can be seen in Figure 3.8(a). Four test cases are performed via introducing a known aberration on to the surface of the Kilo DM and trying to recover it via dOTF estimation:

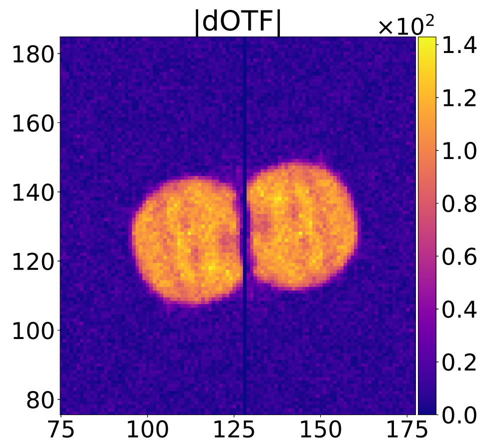
1. 0.05 micron RMS Defocus
2. 0.05 micron RMS Trefoil
3. 0.063 micron RMS Astigmatism
4. 0.063 micron RMS Astigmatism with both sources on the camera concurrently

Finally, a Python pipeline was built to numerically process the PSF data collected from the focal plane camera, which can be seen in Appendix C. This process consists of the following steps:

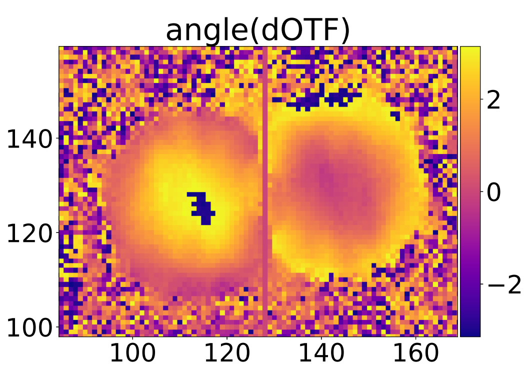
1. Assign a central pixel
2. Crop the average PSF image to a good FFT size (256×256), centered about the chosen pixel
3. FFT the cropped image to get the OTF
4. Repeat 2–3 for the modified pupil modality
5. Take the difference of the OTFs



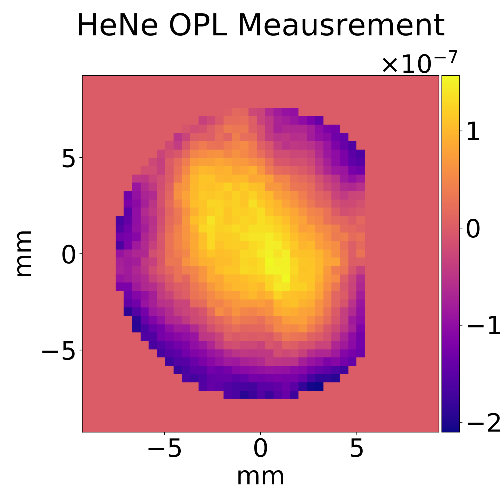
(a)



(b)



(c)



(d)

Figure 3.8: (a) A time averaged PSF with 0.05 micron RMS defocus added to the pupil via the Kilo DM. This is the average of 10 epochs of 250 images. The telltale noise pattern for the detector used in these experiments is visible in the background. (b) The absolute value of the dOTF measured for the defocus test case. No dark subtraction is done to lower the effect of the background noise. The zero frequency is zeroed out as a part of the numerical processing to remove the effects of the banded noise structure. Because it passes through the overlap region of the dOTF, this choice does not effect the estimate of the OPL. Instead, dark subtraction could be used to counter the effects of the noise pattern, and also serve to improve the signal to noise ratio. (c) The argument of the dOTF measured for the defocus test case. (d) The estimated OPL from the dOTF measurement, following the numerical processing.

6. If no dark subtraction was utilized on the PSFs, zero out the pixels corresponding to the zero frequency, as they will be effected the most by the noise. Removing their values does not effect the dOTF estimate because they pass only through the overlap region of the estimate, which is not used. This is the step taken in the experiments below, and can be seen in Fig. 3.8 (c) and (d).

7. Phase unwrap arg(dOTF) using methods of Ghiglia and Pritt (1998)

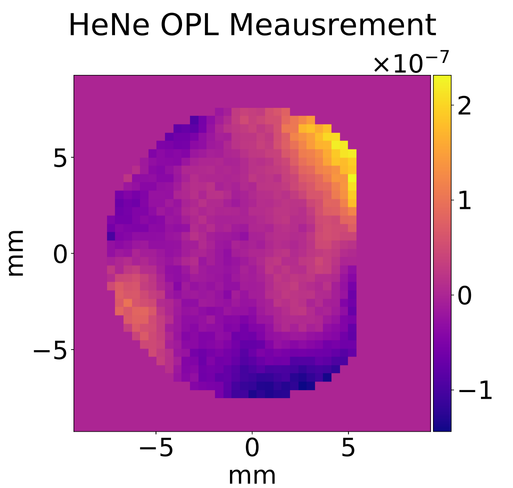
8. Mask off the conjugated pupil and overlap region, scale by the wave number corresponding to the source being used for the data to convert to units of OPL, and subtract the measured reference OPL.

The results of this process can be seen in the rest of Figure 3.8. To determine the quality of the OPL estimate dOTF gives using the off-axis beam, dOTF will also be performed on the on-axis beam to provide a means of comparison. Also, an initial dOTF will be computed for both sources without any injected, known aberration on the Kilo DM to establish reference measurements for each source to be subtracted. Finally, unless otherwise specified, the data is taken with only one source present on the detector at a time (the reasoning for this, and if it is necessary, will be discussed below). The results of the experiments can be found in Table 3.1. We note here that although all the values in this table are positive (and match the injected aberration's sign), as seen in the previous three sections using dOTF as a wavefront sensor in a self-calibration control loop, dOTF is capable of estimating the correct sign.

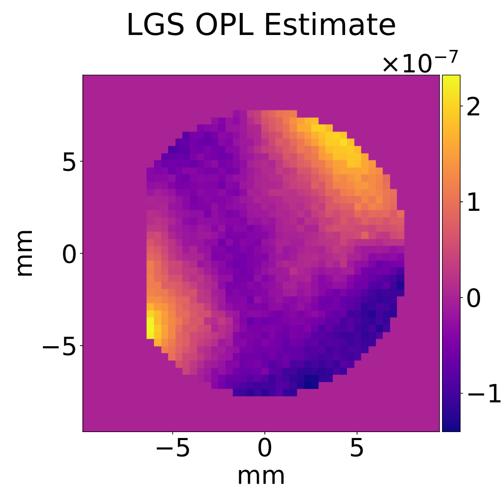
Injected Aberration RMS (μm)	**RMS Estimate HeNe (μm)**	**RMS Estimate LGS (μm)**
0.063 Astig	0.06298	0.06465
0.05 Trefoil	0.05937	0.04929
0.05 Defocus	0.07248	0.06814

Table 3.1: Estimated RMS aberration using the dOTF method using on and off axis sources.

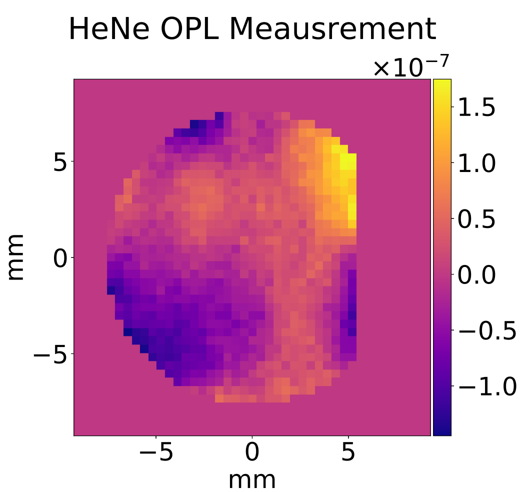
Taking a closer look at these results, we see excellent agreement in the RMS error detected for both the on– and off–axis sources for astigmatism compared to the true injected aberration on the DM, close agreement for trefoil, and although the estimates were close to each other for defocus, they were not particularly close to the input. There are several plausible explanations for this observed behavior. Looking at frames (a) and (b) in Figure 3.9, we see the estimated OPL for the astigmatism case. For both the HeNe and LGS sources, the form of astigmatism is clearly visible, and a very similar peak-to-valley error is seen. It should be noted that because of the fact that the dOTF was performed using different actuators for the modification on both sources, as their respective beam footprints on the DM are different, the estimates for each source are reflected with respect to each other. The estimated RMS of the injected trefoil is not as accurate for the HeNe source as it is for the LGS source. Frames (c) and (d) in Figure 3.9 show the estimated OPL for the trefoil case. We again see the form of the injected aberration clearly for both source estimates, accounting for the reflection explained previously. However, we note that the RMS of the HeNe estimate does not match as precisely to the true injected aberration as the LGS estimate. This can be seen in the OPL in frame (c) where the path length at the center of the pupil is not as well defined as would be expected for trefoil, and is seen in the LGS estimate. The reason for this likely lies in the signal of the dOTF for the HeNe not quite being as good as for the LGS source, leading to some more artifacts in the phase unwrapping algorithm employed. Although not pictured, estimating an injected defocus NCPA is not as accurate as would be anticipated. This is due to the fact that adding defocus on the DM effectively shifts the point the light comes to focus so that it is beyond where the detector is placed. This spreads out the photons as described above, in accordance with allowing a longer exposure being possible without saturating the detector pixels. However, to keep the parameters of each test the same, the exposure time was not increased to compensate the fact that the photons were more distributed, lowering the overall signal to noise ratio of this particular case compared to the cases estimating astigmatism and trefoil. Furthermore, this experiment took place about an hour after the references being subtracted so that we could only be estimating the injected aberration were measured, meaning that the system had



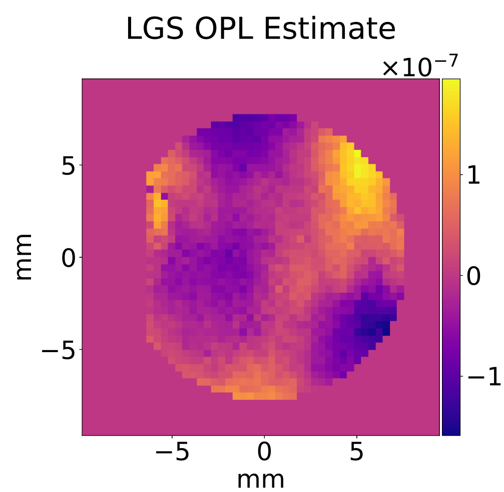
(a)



(b)



(c)



(d)

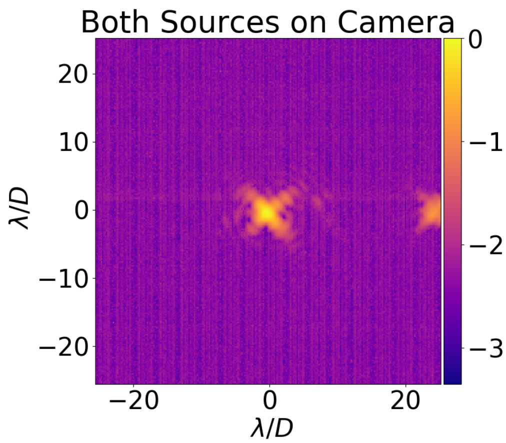
Figure 3.9: (a) The estimated OPL using the on-axis HeNe source for 0.063 micron RMS astigmatism injected by the DM. (b) The estimated OPL using the off-axis LGS source for 0.063 micron RMS astigmatism injected by the DM. (c) The estimated OPL using the on-axis HeNe source for 0.05 micron RMS trefoil injected by the DM. (d) The estimated OPL using the off-axis LGS source for 0.05 micron RMS trefoil injected by the DM.

likely drifted some, making the reference subtraction less accurate. This would be experienced as an increase in overall RMS error estimated, which is what is seen.

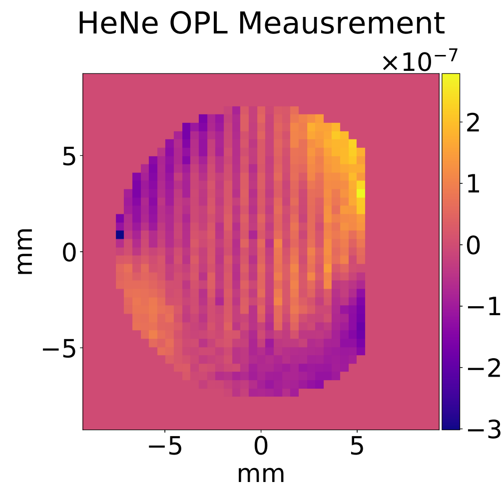
Finally, the test to probe how to deal with both sources being present on the detector at the same time is performed, keeping the same astigmatism as used previously. Figure 3.10(a) shows the detector image of the PSF after the centering operation in the code pipeline has been done. The HeNe is the source of the central intensity, with the LGS source being the dimmer image to the far right near pixel 250. What will happen when the sources are observed on the detector simultaneously is easy to predict. Because the detector sees both as it records the intensity, as far as the FFT operation is concerned, both PSFs will be taken to be “coherent”. This means in the FFT operation, they will interfere with each other, with the off–axis source having a very large phase tilt that will couple into a fringe–looking phenomenon across the pupil. Figure 3.10(b) shows that this is in fact what happens, meaning the estimate is corrupted. One potential way to allow for measuring dOTF estimates with multiple sources on the detector then becomes clear: masking. The second source can simply be masked away and replaced by a suitable estimate of the detector noise for those pixels prior to the FFT being done. It is also clear that this will only be possible so long as the PSFs are separated by enough that there is effectively no overlap between them (the diffraction rings are well below the noise floor at the location of the other PSF), as masking will not be able to disentangle the light should it overlap. An example of attempting to mask off the LGS source and compute the OPL estimate using dOTF for the HeNe source is shown in Figure 3.10(c). The phase tilt has been removed, as would be expected, and the estimate of astigmatism seen in Figure 3.9(a) is recovered.

3.4 Conclusions

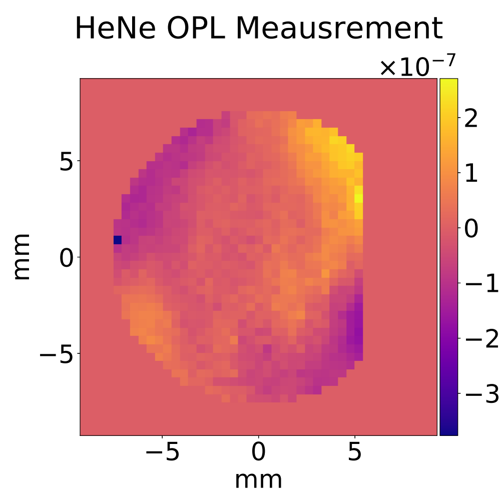
In this chapter, the differential Optical Transfer Function wavefront sensor (dOTF) was introduced, and several useful applications for estimating static NCPA were presented. dOTF can be applied very easily, without the need for expensive new



(a)



(b)



(c)

Figure 3.10: (a) Cropped average detector image with both the HeNe and LGS source PSFs present on the detector simultaneously. Note the vertical lines are again typical in the noise structure of the detector. (b) The estimated OPL if both PSFs are left present when the FFT to compute the OTF is taken. As expected, the off-axis source creates the appearances of a large phase tilt across the pupil. (c) The estimated OPL if the LGS source is masked off, and replaced by pixel values corresponding to a similar detector region, prior to the FFT. Because the PSFs do not overlap due to the noise floor, the effects of the LGS source being on the detector are largely mitigated.

hardware to be added to the optical system, and provide a means to use a DM in the system to compensate for any aberration, or cophase a segmented primary mirror, to improve the Strehl ratio. Furthermore, it can even be applied using an off-axis source, which could be important in some future missions. However, as alluded to, there are some drawbacks to using dOTF that do limit the scenarios in which it can be applied. Because the only major changes that can occur in the system over the time of measurement can be the intended modification via a phase change or amplitude blocker, the method can not be used for sensing anything other than aberrations that are static, at least over the observation interval. This limitation likely restricts use of the method to lab settings, or space-based observatories. In order to probe this restriction, a set of experiments or simulations could be run to compare the required number of photons in the signal of the two PSF measurements (or equivalently integration times at various source magnitudes) vs. a metric of precision of the resulting dOTF estimate (RMS error in OPD of the estimate compared to the true OPD). An examination of such a set of data would allow for the extraction of the required exposure times for a given source to achieve a high quality estimate, providing a window into the specific temporal variations in the optical system that could be successfully handled by dOTF. The author's intuition thinks that in a situation similar to the lab experiments reported, which required roughly 5 total seconds of PSF measurements to achieve the estimates, would be able to estimate aberrations with lifetimes much greater than 5 seconds. But without a future examination of this relationship, just how much greater the speckle lifetime for a given source brightness would need to be remains an open question. Similarly, because the method requires a large number of photons to compute quality estimates, its use in photon starved environments is not recommended. Another limiting factor for the use of dOTF are the effects of optical bandwidth. It may seem advantageous to use a broadband optical source to increase the number of photons, and thus increase the dOTF signal. The effects of bandwidth can be approximated by incoherently adding many narrow band PSFs, which results in a radially blurred total PSF (as the size of a PSF is directly proportional to the wavelength; longer wavelengths appear to stretch more). Because dOTF performs the Fourier Transform of this blurred PSF, this means the OTF sees the reverse effect, and with the difference operation, we can describe what will happen. The radial blurring in the dOTF will be centered on the pupil modification location (in the center of the overlap region), and progress outward with increasing strength. This effect is described more fully in Codona (2013), where Codona suggests possible mitigation by using multiple pupil modifications around the edge of the pupil, and stitching them together retaining only information in each measurement within the pupil radius from the modification. In addition to the blurring caused by bandwidth, there is the blurring due to the pupil modification itself. Knight et. al. [41] and Jiang et. al. [38] suggest methods for using knowledge of the form of the pupil modification to apply deconvolution techniques, and recover high spatial frequency information in the dOTF estimates. This would be especially useful for segmented aperture observatories because the segment boundaries in the pupil will be recovered. In spite of these limitations, the ease of implementation, and accuracy of the resulting estimates of the complex amplitude field in the exit pupil make dOTF an extremely powerful method worth consideration for optical system calibration.

CHAPTER 4

Frazin's algorithm †

Having just examined the dOTF method for calibrating static NCPA out of the science arm of an AO system, and the limitations of its application, it is clear that more work needs to be done in order to counteract the damaging effects of quasi-static speckles in real-time while observing. The next three chapters will cover the development, demonstration, and future of a novel statistical regression algorithm, called Frazin's algorithm, that will serve to fill this gap in NCPA control, and also provide a means of improving direct detection of exoplanet signals in high-contrast imaging settings.

4.1 The history of Frazin's algorithm

The journey of the regression methods, collectively known as Frazin's algorithm, presented here begins with Frazin (2013). To start, the observational setup in which the algorithm can be run must first be understood. A flow chart of the algorithm is provided in Fig. 4.1. An adaptive optics (AO) enabled ground-based telescope with a deformable mirror (DM) followed by a beam splitter or dichroic that separates the light in to two separate paths: the wavefront sensor (WFS) path and the science path which is home to a science instrument such as a coronagraph, which brings the light to focus on the science camera. As described in Ch. 2, an AO system in this configuration can be run in closed-loop, meaning that the DM provides compensation for the measured turbulent wavefront, and then the WFS measures the residual wavefront leftover (called the *AO residual*). Then, the control system converts the measurement of the wavefront in to a command to adjust the shape of

† Parts of this chapter have been published previously as Frazin and Rodack (2021).

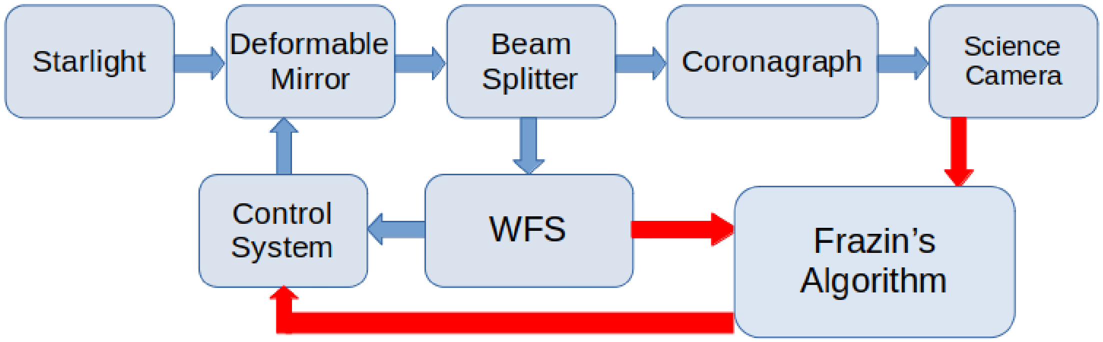


Figure 4.1: The observation model used with Frazin's algorithm. The starlight enters a ground-based telescope with a kHz AO system running in closed-loop. Synchronized telemetry from the WFS and science camera are fed into the algorithm at each millisecond. After T milliseconds, the algorithm gives an estimate of the NCPA and exoplanet image. The estimate of the NCPA can then be fed back into the control system to compensate for it.

the DM surface to the current state of the AO residual. To remind the reader, the two separate paths for wavefront sensing and science gives rise to the possibility of non-common path aberrations (NCPA) to be present in the science path that are downstream of the beam splitter and thus cannot be measured by the WFS. Because the NCPA change with time, as mechanical stresses (such as gravity's pull relative to the optic changes as the telescope tracks, thermal gradients change the optical surfaces, the starlight hits different parts of the telescope primary or secondary, etc.) on the telescope optics change, the lifetime of such aberrations can range widely from the span of minutes to longer. When using a telescope for high contrast imaging, these quasi-static NCPA create speckles in the science camera that can be quite difficult to distinguish from exoplanet signals. The primary goal of Frazin (2013) was to present an algorithm that would leverage measurements of the AO residual from the WFS simultaneously obtained with focal plane measurements from the science camera to estimate the underlying NCPA and exoplanet image in the data. It demonstrated via simulation a post-processing technique, henceforth referred to as *F13*, to exploit the fact that the AO residual phase provides a new phase in the pupil plane at each exposure, and provides a statistically independent phase screen which modulates a planetary or quasi-static speckle in a new way every atmospheric clearing time [44]. This allows for the rapidly changing AO residual speckles in millisecond or faster exposures to be used as probes to estimate both the exoplanet image and NCPA, because with each passing millisecond, more diversity in the observations is achieved. Fig. 4.2 shows simulations of noise-free science camera images with the same NCPA being probed with different AO residual wavefronts demonstrating this phenomenon. Though we will not dive too deeply in to the

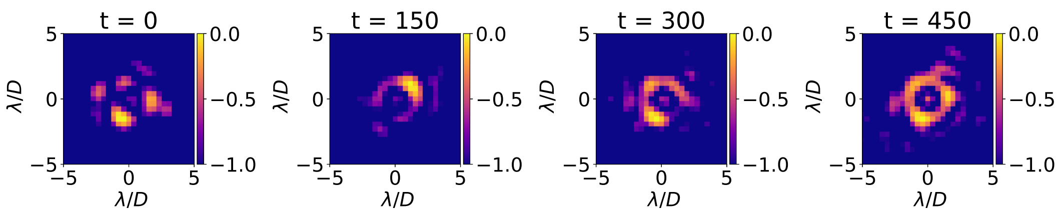


Figure 4.2: Log10 scale, noise-free focal plane images for different realizations of the instantaneous AO residual phases, showing the modulation of the signal. The range of the color scales are limited to a factor of 10 to make the modulation easy to see.

mechanics of F13, as the nitty gritty details will be saved for the following section and the derivations there, there is something to be gained from a brief introduction. The original notation of Frazin will be largely maintained, with some minor changes to attempt to alleviate some frustration comparing to the following sections. The (noise-free) intensity impinging on the science camera can be written as:

$$I(\boldsymbol{\rho}, t) = u_\bullet^2 i_p(\boldsymbol{\rho}, t) + \mathcal{A}(\boldsymbol{\rho}, t) + \boldsymbol{a}^\dagger \boldsymbol{b}(\boldsymbol{\rho}, t) + \boldsymbol{b}^\dagger(\boldsymbol{\rho}, t)\boldsymbol{a} + \boldsymbol{a}^\dagger \boldsymbol{C}(\boldsymbol{\rho}, t)\boldsymbol{a}\,, \tag{4.1}$$

where $\boldsymbol{\rho}$ is a vector of pixel locations, $\dagger$ is the complex conjugate transpose, $u_\bullet$ is the planet amplitude, $u_\bullet^2 i_p$ is the exoplanet point spread function (PSF), $\boldsymbol{a}$ is a vector of aberration coefficients, $\mathcal{A}$ is the intensity only depending on the AO residual ($\boldsymbol{w}$), $\boldsymbol{C}$ depends on the NCPA (ϕ_u) as modulated by $\boldsymbol{w}$, and $\boldsymbol{b}$ depends on the mixing of the AO residual and NCPA. If ϕ_u is decomposed into a "search" basis:

$$\phi_u(\boldsymbol{r}) = \sum_{k=1}^{K} a_k \theta_k(\boldsymbol{r})\,, \tag{4.2}$$

where a_k are the elements of $\boldsymbol{a}$, and θ_k are the functions in the search basis to which the method will fit the NCPA to. Considering N pixel locations $\boldsymbol{\rho} = \rho_0; \ldots; \rho_{N-1}$ in a single exposure, and given T total millisecond exposures synchronized with WFS measurements of $\boldsymbol{w}$, the following linear system model is constructed by Frazin:

$$I(\boldsymbol{\rho}, t) - \mathcal{A}(\boldsymbol{\rho}, t) = u_{\bullet}^2 i_p(\boldsymbol{\rho}, t) + \boldsymbol{a}^{\dagger}\boldsymbol{b}(\boldsymbol{\rho}, t) + \boldsymbol{b}^{\dagger}(\boldsymbol{\rho}, t)\boldsymbol{a} + \boldsymbol{a}^{\dagger}\boldsymbol{C}(\boldsymbol{\rho}, t)\boldsymbol{a}\,. \tag{4.3}$$

This is then rewritten to the form:

$$\begin{aligned}
&\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x}\,, \text{where} \qquad (4.4)\\
&\boldsymbol{y} = [\boldsymbol{y}_0; \ldots; \boldsymbol{y}_{T-1}], \boldsymbol{y}_i = I(\boldsymbol{\rho}_n, t_i) - \mathcal{A}(\boldsymbol{\rho}_n, t_i) + \boldsymbol{\nu}\,,\\
&\boldsymbol{H} = [\boldsymbol{H}_0; \ldots; \boldsymbol{H}_{T-1}], \boldsymbol{H}_i = [i_p(\boldsymbol{\rho}_n, t_i) \quad \boldsymbol{b}^T(\boldsymbol{\rho}_n, t_i) \quad \boldsymbol{b}^{\dagger}(\boldsymbol{\rho}_n, t_i) \quad \mathbf{c}^{\dagger}(\boldsymbol{\rho}_n, t_i)]\,,\\
&\boldsymbol{x} = [u_{\bullet}^2; \boldsymbol{a}; \boldsymbol{a}^*, \boldsymbol{a}_k\boldsymbol{a}_l^*]\,, \text{and}\\
&\boldsymbol{\nu} = [\boldsymbol{\nu}_0; \ldots; \boldsymbol{\nu}_{T-1}]\,,
\end{aligned}$$

where $\mathcal{A}(\boldsymbol{\rho}_n, t_i)$, $\boldsymbol{b}(\boldsymbol{\rho}_n, t_i)$, and $\boldsymbol{C}(\boldsymbol{\rho}_n, t_i)$ are computed using equations given in Frazin (2013) which are not terribly important to this discussion, $\boldsymbol{\nu}$ is the noise from the detector, and $\boldsymbol{x}$ is solved for to produce an estimate of the exoplanet brightness coefficients and the basis function coefficients for the NCPA fit.

The author then replicated and verified Frazin's work on F13, and expanded on the original premise by conducting simulations demonstrating the use of F13 as a real-time method for NCPA control instead of a post-processing method to simply estimate it. Rodack et al. (2018) presents a summary of the simulations conducted verifying the F13 method and probing its parameter space for the interested reader. We will instead focus on the simulations for real-time control of static NCPAs done in this proceedings article to set ourselves up for why further development was required to make Frazin's algorithm a powerful tool for astronomers. The simulator for the closed-loop NCPA control used Fourier Optics principles to perform plane-to-plane propagations between elements of a kHz AO system that includes an ideal WFS and ideal coronagraph (see [15] for the mathematical definition), and a noiseless science camera. Realistic turbulence modeling was done using AtmosphericTurbulenceSimul (`https://github.com/oguyon/AtmosphericTurbulenceSimul`), in order to produce phase only AO residuals ($\boldsymbol{w}$) that will be used in the model. In this case, the NCPA (ϕ_u) that was injected is static, and takes the form of a sinusoid modified by some high order Zernike polynomials (Z43-Z48), that produces a speckle pair in focal plane, with one of the speckles being coincident with an injected exoplanet PSF. The reader may recall from Eq. (4.1) the dependence on both $\boldsymbol{w}$ and ϕ_u, as well as the chosen search basis functions θ_k. Because we are controlling the simulations and choosing the static aberration, θ is chosen to be the same functions used to construct the NCPA. The AO system model was tuned to provide a Strehl of 0.95 in order to provide a high fidelity result from the ideal coronagraph, exposing to a viewer the 10^{-3} contrast exoplanet PSF should the NCPA be controlled by the real-time application of F13. A simple integrator control system is set up, taking estimates returned from the F13 regression on the NCPA coefficients every 150 milliseconds as a solution for a DM correction update. A modest gain is applied for loop stability. The results of this simulation, which can be seen in Fig. 4.3, showed that the method can be used in real-time to estimate and provide compensation for a static NCPA, as the injected speckle pair is removed and suppressed, revealing the exoplanet speckle beneath it. The section from Rodack et al. (2018) that deals with the so called "Real-time Frazin Algorithm" details some slight manipulations to the math to improve computational efficiency if the user only means to estimate NCPA rather than both the NCPA and exoplanet brightness coefficients. Although an interesting adjustment at the time, it turns out to be rendered unnecessary by the following sections, so it will not be discussed here.

The most important takeaways from this brief look at the history of Frazin's algorithm, from its humble beginnings as F13 and its evolution through Rodack et al. (2018), however, are the discussions by Frazin and Rodack of the major shortcomings of the method that would prevent its deployment in the lab or on-sky. The lesser of the shortcomings uncovered is that as constructed, the calculation and concatenation of information at each exposure in an observation in to a large matrix is not computationally efficient enough to hope to run on-sky without improvement, or spending ridiculous sums of money on high performance computers with hundreds of gigabytes of RAM. The fact that the computational requirement involves heavy calculations of many Fast Fourier Transforms (FFTs) that must be done at

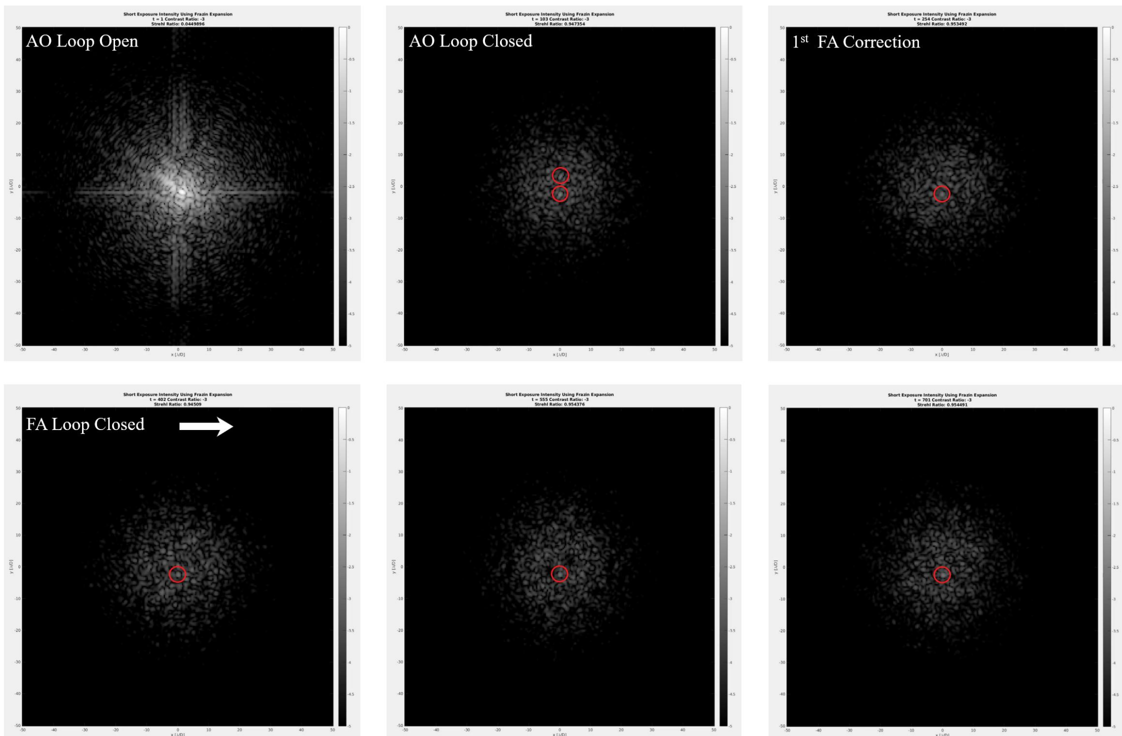


Figure 4.3: Frames from various points in the real-time simulation of the method showing the removal of the speckles created by a static NCPA. *Top left:* The science camera image of a simulated ideal coronagraph dominated by speckles caused by atmospheric turbulence. *Top middle:* The science camera after the AO loop is closed. *Top right:* The science camera after the first correction provided by the Frazin algorithm. Note that the top speckle in the pair is largely faded from view, as it was caused by an NCPA. *Bottom:* Science camera frames after the Frazin's algorithm loop is closed. The sinusoidal NCPA has been suppressed, leaving behind only the swarm of atmospheric speckles, and the exoplanet speckle.

each millisecond means the computer performing the calculations will struggle to keep up with the loop. In addition, the size of $\boldsymbol{y}$ and $\boldsymbol{H}$ scale by the number of millisecond exposures in the observation. This makes it easy to see how a memory problem quickly arises. Using a $\boldsymbol{\rho}$ that is comprised of even 100 pixel locations, and only 30 basis functions in the set θ, $\boldsymbol{H}_t$ is of size 100×961. Then $\boldsymbol{H}$ is constructed by concatenating all the $\boldsymbol{H}_t$ for $0 \leq t \leq T-1$, and $\boldsymbol{y}$ is then $100T$ elements. For a 1 minute observation, or $60,000$ milliseconds, $\boldsymbol{H}$ will be size 6000000×961, and y will be 6000000×1, which reaches approximately 46 GB of RAM assuming the data type is double. If we increase to 100 basis functions in θ instead, 489 GB of storage is now required.

However, the fundamental shortcoming is that the algorithm employed in these studies ignores the effects of wavefront measurement error (WME) in the regression equations. This is to say, the WFS measurements only allow an imperfect estimate of the phase of the wavefront being measured, not the actual phase. Specifically, any WFS exhibits spatial and temporal bandwidth limitations, nonlinearity in the phase of the wavefront, and noise. All these factors are swept under the rug in these examinations, which instead turn to the ideal wavefront sensor, and perfect knowledge of the wavefront to be used in the equations. Without accounting for the effects of the WME, an errors in variables type statistical problem arises, which will inform the choice of nomenclature below. The uncertainty in the measurement of $\boldsymbol{w}$ bleeds into the equations as $\boldsymbol{H}$ has deep dependency on $\boldsymbol{w}$, meaning that there are now errors in the independent variables, and, for a sufficient (and easy to fall in to due to the factors above) WME, the estimate of $\boldsymbol{x}$ will become hopelessly biased and unusable. Armed with the knowledge of these shortcomings, Frazin and Rodack set out to construct and demonstrate an improved method born out of this work.

4.2 Development of the improved Frazin's algorithm [1]

The following section delves in to Frazin's derivation of an improved method to accomplish the task of mitigating the shortcomings of the original methods described in the previous section, namely computational efficiency and WME.

4.2.1 Wavefronts and AO

We will start the discussion of the improved Frazin's algorithm statistical regression method by walking through the path of the light to and into the optical system. The light is taken to be quasi-monochromatic, centered at wavelength, λ, and a beam splitter will serve to divide the incoming photon flux between the WFS and the science instrument. This is not typical of the usual AO system, which often does the wavefront sensing in a wavelength band separate from the science band. However, we choose to use the same wavelength for ease of application/development of the methods below (which is not to say the methods are incompatible with separate wavelength bands). We also will adhere to the assumption that the star being observed is far enough away to be considered a planewave at its arrival at the atmosphere (which is most certainly the case).

The light from the star first arrives at the atmosphere, converting the planewave in to what we will call the *atmospheric wavefront*. The atmospheric wavefront represents the phase (and, if necessary, the amplitude perturbations caused by scintillation resulting from multi-layer turbulence) of the electric field that has arrived at the telescope optical system. The amplitude effects, though not excluded from being included in the method, will again be ignored for the simplification of the derivation. As our optical model does not include the telescopes beam reducing optics, and the AO system is being run in closed-loop, the first surface the atmospheric wavefront encounters is that of the DM. The pupil plane in which this DM sits is both the entrance pupil of the WFS and of the science instrument, with the possible

[1] The work in this and all following sections is based on work that has been published previously as Frazin and Rodack (2021). ©JOSAA [2021] Optica Publishing Group.

exception of the inclusion of NCPA in the latter. To continue, we will just refer to this plane as the entrance pupil, as it is the center of our initial discussion.

Finally, it is important to remember that the method requires both the science camera and the WFS to be operating at a 1kHz or faster, and be taking synchronized exposures, just as was the case in F13. The exposures from the cameras will thus have the same time indices, $\{t\}$. The total observation time is made up of T millisecond exposures, so the index t is an integer that runs between 0 and $T-1$.

AO residual wavefronts

We represent the entrance pupil as comprised of P pixels, with the phase at the $p\underline{\text{th}}$ pixel at time t represented as $\phi_p(\boldsymbol{w}_t)$. The function ϕ_p and the parameter vector $\boldsymbol{w}_t$ together specify the wavefront. There are many ways to do this. For example, $\phi_p(\boldsymbol{w}_t)$ could represent a sum over Zernike modes (evaluated at the location of pixel p), in which case the vector $\boldsymbol{w}_t$ would be a set of coefficients of the Zernike modes at time t. Here we make a simpler choice: the $p\underline{th}$ component of w_t is the phase at pixel p and the ϕ_p function just returns the value of the $p\underline{th}$ component of the input vector. This was called the *PhasePixel* representation in Frazin (2018), where some of its computational advantages are explained. With this representation, the complex-valued electric field in the post-DM pupil plane (ie just after the DM correction has been applied) is:

$$u_p(\boldsymbol{w}_t) = \exp[j\phi_p(\boldsymbol{w}_t)] \,, \tag{4.5}$$

where t is the time index in the range 0 to $T-1$, and the phase, $\phi_p(\boldsymbol{w}_t)$ will be complex valued if treating the amplitude effects is deemed necessary; if not, it is real-valued. Since the functions $\{\phi_p\}$ are fixed, the wavefront here is fully specified by the parameter vector $\boldsymbol{w}_t$. $\boldsymbol{w}_t$ will be called the *wavefront* or the *AO residual wavefront*, just as it was labeled in Sec. **??**.

The vector containing the complex electric field values at all P pixels in the entrance pupil is

$$\boldsymbol{u}'(\boldsymbol{w}_t) = (u'_0(\boldsymbol{w}_t),\, \ldots,\, u'_{P-1}(\boldsymbol{w}_t))^{\mathrm{T}}. \tag{4.6}$$

The prime is used to indicate that the field is specified at the location of the WFS entrance pupil, whereas $\boldsymbol{u}(\boldsymbol{w}_t)$ is the field entering the coronagraph.

Monte Carlo wavefront generation

The so-called bias-corrected estimator, which will prove to achieve treating the WME shortcoming of F13 in a realizable way and will be described below, requires the ability to generate Monte Carlo wavefronts with the same spatial statistical properties as the real wavefronts for the purpose of the required Monte Carlo calculations. The set of Monte Carlo wavefronts will be denoted as $\{\breve{\boldsymbol{w}}_l\}$, $0 \leq l \leq L-1$, where there are L samples in the set.

The true wavefronts, $\boldsymbol{w}_0, \ \ldots, \ \boldsymbol{w}_{T-1}$, form a time series with a characteristic correlation time. Unlike the true wavefronts, the L Monte Carlo wavefronts, $\{\breve{\boldsymbol{w}}_l\}$, form an unordered set. Under conditions of quasi-stationarity, the ability to perform Monte Carlo sampling allows approximation of the time average of any function (linear or otherwise) $f(\boldsymbol{w}_t)$ with the Monte Carlo mean of $f(\breve{\boldsymbol{w}}_l)$. In other words,

$$\langle f(\breve{\boldsymbol{w}}_l)\rangle_{\mathrm{mc}} \approx \langle f(\breve{\boldsymbol{w}}_l)\rangle_{\mathrm{E}} = \langle f(\boldsymbol{w}_t)\rangle_{\tau} \ , \tag{4.7}$$

which says that, given enough realizations in $\{\breve{\boldsymbol{w}}_l\}$, the Monte Carlo mean of $f(\breve{\boldsymbol{w}}_l)$ will closely approximate the ensemble mean $\langle f(\breve{\boldsymbol{w}}_l\rangle_{\mathrm{E}}$, which is equal to the temporal mean over the time interval of length T of the stochastic process $\langle f(w_l\rangle_{\tau}$. The Monte Carlo mean based on the L samples, $\{\breve{\boldsymbol{w}}_l\}$, is:

$$\langle f(\breve{\boldsymbol{w}}_l)\rangle_{\mathrm{mc}} \equiv \frac{1}{L}\sum_{l=0}^{L-1} f(\breve{\boldsymbol{w}}_l)\,, \tag{4.8}$$

and its importance will be further clarified in the discussion of the bias-corrected estimator in Sec. 4.2.3: Bias-Corrected Estimation.

Wavefront measurements

The WFS telemetry makes it possible to get an estimate (or measurement) of $\boldsymbol{w}_t$, which will be denoted as $\hat{\boldsymbol{w}}_t$, with the "hat" designating that it is an estimate of the vector of quantities that specify the true phases $\boldsymbol{w}_t$. A key assumption will be that an accurate computational model of the WFS can be constructed, making this an added prerequisite to on-sky implementation of this algorithm. The wavefront measurement (determined by analysis of the WFS signal) can be modeled by the equation

$$\hat{\boldsymbol{w}}_t = \mathcal{W}(\boldsymbol{w}_t) + \boldsymbol{n}_t \,, \tag{4.9}$$

where $\mathcal{W}$ is a nonlinear measurement operator, and $\boldsymbol{n}_t$ is the noise vector. $\mathcal{W}$ is necessarily nonlinear due to the fact that the intensity measured by the detector in the WFS is nonlinear in the phase values. Thus, this nonlinearity is inevitable, even if $\hat{\boldsymbol{w}}$ is based on an estimator that is linear in the WFS intensity values. $\mathcal{W}$ is the numerical model that simulates the effect the WFS has on an input wavefront (i.e. the propagation of the wavefront through the WFS optics and its following measurement in intensity on the WFS detector), and any further processing that takes place to produce an estimate of that wavefront from the detected intensity. Frazin (2018) provides linear and nonlinear estimation algorithms for the case of a pyramid WFS, but the formalism here is completely agnostic with respect to the wavefront sensing hardware, so any such numerical model is permissible.

For the purposes of the bias-corrected estimator, having $\mathcal{W}$ and a Monte Carlo sample $\breve{\boldsymbol{w}}_l$, it is straightforward to draw a sample from the distribution of the wavefront estimates:

$$\hat{\breve{\boldsymbol{w}}}_l = \mathcal{W}(\breve{\boldsymbol{w}}_l) + \boldsymbol{n}_l \,, \tag{4.10}$$

where $\boldsymbol{n}_l$ is a sample from the same stochastic processing governing the WFS detector noise, $\boldsymbol{n}_t$, in Eq. (4.9). With these wavefront terms defined, we can now step in to the science instrument path.

4.2.2 Coronagraph and NCPA model

In order to develop the regression equations that define Frazin's algorithm, we must now be able to describe the light in the science instrument. The regression algorithm simulated below makes no assumptions about the type of coronagraph architecture that is in the science instrument, and thus provides for the freedom of many choices, including Lyot, PIAACMC [32], APP [61; 65], and vector vortex [58] coronagraphs, to suppress host starlight. Furthermore, a coronagraph is not a mathematical necessity for running the algorithm, as will be demonstrated in Sec. 6.1. However, to proceed and make the verbiage concise, we will refer to the science instrument as the coronagraph.

Stellar intensity

For describing the NCPA in the coronagraph, the assumption that we can represent the entire effect of the NCPA via pupil plane manifestations is all that is treated. The estimation methods to follow are not limited to this assumption, and adding additional quantities, such as coronagraph alignment errors or NCPA that are not accurately described by pupil plane manifestations, is possible by augmenting the regression model appropriately. The electric field in the coronagraph entrance pupil, as mentioned above, is the same as electric field in the WFS entrance pupil, as given in Eq. (4.5), except that is it modified by the pupil plane NCPA. It is given by:

$$u_p(\boldsymbol{w}_t, \boldsymbol{a}) = u'_p(\boldsymbol{w}_t) \exp[j\theta_p(\boldsymbol{a})] = \exp j[\phi_p(\boldsymbol{w}_t) + \theta_p(\boldsymbol{a})] \,, \tag{4.11}$$

where $u'_p(\boldsymbol{w}_t)$ was given in Eq. (4.5), and $\boldsymbol{a}$ is a vector of length N_a containing real-valued parameters that specify the NCPA, just as it was in F13. The function $\theta_p(\boldsymbol{a})$ expresses the phase (and, if needed, the amplitude as well) of the NCPA at the pth pixel in the pupil as a function of the vector $\boldsymbol{a}$. The simplest way to do this to have a_l, which is the lth component of $\boldsymbol{a}$, be the phase of the NCPA at the pth pupil pixel so that $\theta_p(\boldsymbol{a}) = a_l$, but other representations, such as modal expansions along the lines of Zernikes, Fourier modes, or BSpline modes, may be useful and require fewer components in the $\boldsymbol{a}$ vector. Two such modal expansions will be demonstrated in Sec. 5.1 (Zernike polynomials) and Sec. 6.1 (BSplines). The vector containing the field values at all P pupil pixels is $\boldsymbol{u}_\star(\boldsymbol{w}_t, \boldsymbol{a}) = \big(u_0(\boldsymbol{w}_t, \boldsymbol{a}),\, \ldots,\, u_{P-1}(\boldsymbol{w}_t, \boldsymbol{a})\big)^\mathrm{T}$.

Since the coronagraph is a linear optical system, the field at the nth detector pixel in the science camera is a linear combination of the field values at the pupil pixels. The vector containing the detector field values at time t is then given by the matrix-vector multiplication:

$$\boldsymbol{v}_\star(\boldsymbol{w}_t, \boldsymbol{a}) = \boldsymbol{D}\, \boldsymbol{u}_\star(\boldsymbol{w}_t, \boldsymbol{a}) \tag{4.12}$$

where $\boldsymbol{v}_\star(\boldsymbol{w}_t, \boldsymbol{a})$ has L components, one for each of the L science camera pixels, and $\boldsymbol{D}$ is a complex-valued matrix that performs the optical propagation. The $\boldsymbol{D}$ matrix of size $L \times P$ is the result of pre-computed optical propagations (see Appendix B). Pre-computing the matrix $\boldsymbol{D}$ allows computationally expensive modeling to be employed at no cost to the execution time of Frazin's algorithm, an important step towards realizing adequate computational efficiency. An analogous procedure for pre-computing the propagation matrix for a pyramid WFS is given in Frazin (2018). Finally, the stellar component of the intensity in the science camera at time t is

$$\boldsymbol{i}_\star(\boldsymbol{w}_t, \boldsymbol{a}) = \boldsymbol{v}_\star(\boldsymbol{w}_t, \boldsymbol{a}) \circ \boldsymbol{v}_\star^*(\boldsymbol{w}_t, \boldsymbol{a})\,, \tag{4.13}$$

where the $\circ$ notation denotes element-wise multiplication.

Planetary intensity

To round out the discussion on the intensity in the science camera, an expression for the instantaneous planetary intensity impinging on the detector at time t is provided. This intensity will vary in time because the planetary light experiences the same atmospheric wavefront perturbations, $\boldsymbol{w}_t$, that the starlight does. Implicit in this assumption is that the effects of anisoplanatism are negligible, which will be the case since the separation between the target exoplanet and its host star is much less than an arcsecond [19]. In order to simplify the discussion, it will also be assumed that the effects of NCPA are negligible for the planetary light. This is probably a valid approximation, but including the NCPA in the regression formulation if not is simple (via a linearization in the NCPA coefficients exactly as described for the starlight in Sec. 4.2.2: Linearization). Under the usual assumption that astronomical sources are spatially incoherent, the planetary intensity impinging on the science camera is linear in the fluxes coming from the elementary patches of solid-angle on the sky.

This linearity makes the modeling task easier, as the resulting regression model is already linear in the parameter vector, called $\boldsymbol{p}$, that represents the planetary image. Sometimes the vector $\boldsymbol{p}$ will be referred to as the *exoplanet image*, even though it is really the set of parameters chosen to specify the image, not the image itself.

Let the sky-angle relative to the telescope pointing be represented by the two-component vector, $\boldsymbol{\alpha} = (\alpha_x, \alpha_y)$, where α_x and α_y correspond to local Cartesian coordinates on the sky, with units of radians, which will have values that range from about 0.5 to perhaps $10\lambda/D$. The image we wish to estimate is represented by the function $S(\boldsymbol{\alpha})$, which has units of energy flux per solid angle. Let $\{\boldsymbol{\alpha}_n\}$ correspond to a numerical grid of angles, with N_p points on the sky that will be used to represent the planetary image $S(\boldsymbol{\alpha})$, so that

$$S(\boldsymbol{\alpha}) = \gamma \sum_{n=0}^{N_p-1} p_n \delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_n) \tag{4.14}$$

where γ contains the necessary scaling factors accounting for the units of energy flux per solid angle, N_p is the number of points in the grid, δ is the Dirac delta function, and $\{p_n\}$, the brightness coefficients, are unitless. The expansion in Eq. (4.14) should be adequate if the grid spacing of the angles $\{\boldsymbol{\alpha}_n\}$ is roughly λ/D or less. Next, it is useful to define the vector $\boldsymbol{p} = (p_0, \ldots\ p_{N_p-1})^{\mathrm{T}}$, which has N_p components, one for each sky-angle on the grid. Mathematically, the essential property of the expansion in Eq. (4.14) is that it is linear in the coefficients $\{p_n\}$.

In the telescope beam above the atmosphere, at the transverse location $\boldsymbol{r}$, the electric field from the part of the planetary image at the angle $\boldsymbol{\alpha}_n$ is given by:

$$u_{\mathrm{pn}}(\boldsymbol{r}) = \sqrt{p_n} \exp\left[j \frac{2\pi \boldsymbol{\alpha}_n \cdot \boldsymbol{r}}{\lambda} \right] , \tag{4.15}$$

where λ is the wavelength and $\boldsymbol{\alpha}_n \cdot \boldsymbol{r}$ is a scalar (dot) product. The field arising from the sky-angle $\boldsymbol{\alpha}_n$ at time t and location $\boldsymbol{r}_l$ in the coronagraph entrance pupil is then given by:

$$u_{\mathrm{pt},n,l} = \sqrt{p_n} \exp\left[j\phi_l(\boldsymbol{w}_t) + \frac{2\pi \boldsymbol{\alpha}_n \cdot \boldsymbol{r}_l}{\lambda} \right] , \tag{4.16}$$

where the AO residual, $\phi_l(\boldsymbol{w}_t)$, has been included and the angle $\boldsymbol{\alpha}_n$ has been rescaled to account for the magnification of the telescope's beam reducing optics if needed.

The fields defined by Eq. (4.16) at all P pixels in the pupil can be collected in to a single vector (still coming from the one sky-angle $\boldsymbol{\alpha}_n$):

$$\boldsymbol{u}_{\bullet t,n} = \frac{\left(u_{\mathrm{p}t,n,0}\ ,\ \ldots\ ,\ u_{\mathrm{p}t,n,P-1}\right)^{\mathrm{T}}}{\sqrt{p_n}}\,, \tag{4.17}$$

where the amplitude, $\sqrt{p_n}$, has been normalized out. Similarly to Eq. (4.12), the $\boldsymbol{D}$ matrix can be used to propagate the planetary field to the detector plane as:

$$\boldsymbol{v}_{\bullet t,n} = \sqrt{p_n}\,\boldsymbol{D}\,\boldsymbol{u}_{\bullet t,n}\,, \tag{4.18}$$

where $\boldsymbol{v}_{\bullet,t,n}$ is a vector of length L (one component for each detector pixel) containing the field in the detector plane arising from the part of the planetary image at the sky-angle $\boldsymbol{\alpha}_n$. Next, this is converted to the intensity impinging on the L science camera pixels, and is thus the image of a planet at the sky angle $\boldsymbol{\alpha}_n$ as:

$$\boldsymbol{i}_{\bullet t,n} = \boldsymbol{v}_{\bullet t,n} \circ \boldsymbol{v}^*_{\bullet t,n} = p_n\left[(\boldsymbol{D}\boldsymbol{u}_{\bullet t,n}) \circ (\boldsymbol{D}^*\boldsymbol{u}^*_{\bullet t,n})\right]\,. \tag{4.19}$$

Because the planetary intensity can be considered to be the sum of incoherently radiating sources from the directions $\{\boldsymbol{\alpha}_k\}$, to get the entire planetary image at time t, we only need to sum $\boldsymbol{i}_{\bullet t,n}$ over the n sky-angles:

$$\begin{aligned}\boldsymbol{i}_\bullet(\boldsymbol{w}_t,\boldsymbol{p}) = \sum_{n=0}^{N_p-1}\boldsymbol{i}_{\bullet t,n} &= \sum_{n=0}^{N_p-1} p_n(\boldsymbol{D}\boldsymbol{u}_{\bullet t,n}) \circ (\boldsymbol{D}^*\boldsymbol{u}^*_{\bullet t,n}) \\ &\equiv \boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t)\boldsymbol{p}\,,\end{aligned} \tag{4.20}$$

where $\boldsymbol{p}$ is the vector of planetary image brightness coefficients and $\boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t)\boldsymbol{p}$ is a matrix-vector multiplication. Eq. (4.20) defines the real-valued, $L \times N_p$ planetary system matrix as $\boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t) \equiv (\boldsymbol{D}\boldsymbol{u}_{\bullet t,n}) \circ (\boldsymbol{D}^*\boldsymbol{u}^*_{\bullet t,n})$. The dependence of $\boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t)$ on the L detector pixels and the N_p sky-angles can be seen in Eq. (4.16). As can been seen in the dependence on $\boldsymbol{w}_t$, the planetary system matrix is not a constant matrix, but rather a matrix of nonlinear functions of the wavefront parameters $\boldsymbol{w}_t$ that provides for the modulation of the planetary image by the AO residual.

Linearization

The simulations to follow indicate there is no reason to believe that nonlinearity in the NCPA coefficients limits the accuracy of the method for any reasonable amplitude of the NCPA.

The regression method presented below relies on linearity in the parameters to be estimated, which include the NCPA coefficient vector $\boldsymbol{a}$. We will proceed on the assumption that linearizing the problem in the coefficients $\boldsymbol{a}$ is an acceptable procedure. The simulations to follow indicate that the accuracy of the method is not harmed by the nonlinearity in the NCPA coefficients for a reasonable amplitude of NCPA, given a procedure of successive relinearizations similar to gradient descent is employed. Importantly, the estimators presented here never linearize in the wavefront $\boldsymbol{w}_t$, the measured wavefront $\hat{\boldsymbol{w}}_t$, or even the wavefront measurement error $\delta\boldsymbol{w}_t \equiv \boldsymbol{w}_t - \hat{\boldsymbol{w}}_t$.

To get the total intensity impinging on the detector, the sum of the stellar intensity from Eq. (4.13) and the planetary intensity from Eq. (4.20) is taken:

$$\begin{aligned} \boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p}) &\equiv \boldsymbol{i}_\star(\boldsymbol{w}_t, \boldsymbol{a}) + \boldsymbol{i}_\bullet(\boldsymbol{w}_t, \boldsymbol{p}) \\ &= \boldsymbol{i}_\star(\boldsymbol{w}_t, \boldsymbol{a}) + \boldsymbol{A}_\mathrm{p}(\boldsymbol{w}_t)\boldsymbol{p} \end{aligned} \tag{4.21}$$

A Taylor expansion of $\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p})$ in Eq. (4.21) in the vectors $\boldsymbol{a}$ and $\boldsymbol{p}$ about the point $(\boldsymbol{a}_0, \boldsymbol{p}_0)$ can be written as:

$$\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p}) \approx \mathbf{c}(\boldsymbol{w}_t) + \boldsymbol{A}_\mathrm{a}(\boldsymbol{w}_t)(\boldsymbol{a} - \boldsymbol{a}_0) + \boldsymbol{A}_\mathrm{p}(\boldsymbol{w}_t)(\boldsymbol{p} - \boldsymbol{p}_0)\,, \tag{4.22}$$

where $\boldsymbol{A}_\mathrm{a}(\boldsymbol{w})$ is the real-valued $L \times N_\mathrm{a}$ *NCPA system matrix*. It is a matrix of functions of $\boldsymbol{w}_t$ defined as the Jacobian:

$$\boldsymbol{A}_\mathrm{a}(\boldsymbol{w}_t) \equiv \left.\frac{\partial\, \boldsymbol{i}_\star(\boldsymbol{w}_t, \boldsymbol{a})}{\partial \boldsymbol{a}}\right|_{\boldsymbol{a}_0}\,, \tag{4.23}$$

which can be calculated from Eq. (4.13). The zero point vector is defined as:

$$\mathbf{c}(\boldsymbol{w}_t) \equiv \boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}_0, \boldsymbol{p}_0)\,. \tag{4.24}$$

In order to avoid carrying excess notation, but without a loss of generality, the vectors $\boldsymbol{a}_0$ and $\boldsymbol{p}_0$ will be taken to be 0. In addition, since the iterative relinearization procedure has proved so successful in simulation, the approximation in Eq. (4.22) will be taken to be an equality, which will also help to make the following discussion less complicated. Finally, the expression for the total intensity impinging on the science camera is at time, t is:

$$\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p}) = \mathbf{c}(\boldsymbol{w}_t) + \boldsymbol{A}_{\mathrm{a}}(\boldsymbol{w}_t)\boldsymbol{a} + \boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t)\boldsymbol{p}\,, \tag{4.25}$$

which models all of the speckles seen on the science camera at time t. Note that $\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p})$ does not refer to the quantities measured by the science camera, since it does not include noise. Rather, even within the paradigm of this model (i.e., pretending the above model is exactly correct), it is a quantity that can never be known exactly since $\boldsymbol{w}_t$, $\boldsymbol{a}$ and $\boldsymbol{p}$ can only be estimated. A similar equation was given by Sauvage et al. (2010) early in the phase diversity literature, although they were not concerned with millisecond imaging.

Setting up the regression

The first step in setting up the regression is defining the vector $\boldsymbol{x}$ with $N = N_{\mathrm{a}} + N_{\mathrm{p}}$ components, and the $L \times N$ matrix $\boldsymbol{A}(\boldsymbol{w}_t)$. The $\boldsymbol{x}$ vector is thus defined as the concatenation of $\boldsymbol{a}$ and $\boldsymbol{p}$, while similarly, $\boldsymbol{A}(\boldsymbol{w}_t)$ is the concatenation of $\boldsymbol{A}_{\mathrm{a}}$ and $\boldsymbol{A}_{\mathrm{p}}$, as follows:

$$\boldsymbol{x} \equiv \begin{pmatrix} \boldsymbol{a} \\ \boldsymbol{p} \end{pmatrix} \quad \text{and} \tag{4.26}$$

$$\boldsymbol{A}(\boldsymbol{w}_t) \equiv \begin{pmatrix} \boldsymbol{A}_{\mathrm{a}}(\boldsymbol{w}_t) & \boldsymbol{A}_{\mathrm{p}}(\boldsymbol{w}_t) \end{pmatrix}. \tag{4.27}$$

With these definitions, Eq. (4.25) can be rewritten as:

$$\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{x}) = \mathbf{c}(\boldsymbol{w}_t) + \boldsymbol{A}(\boldsymbol{w}_t)\boldsymbol{x}\,. \tag{4.28}$$

The vector $\boldsymbol{y}_t$, which will have the length L, represents the fixed intensities measured by the L science camera pixels at time t. $\boldsymbol{y}_t$ is given by the equation:

$$\begin{aligned}\boldsymbol{y}_t &= \boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{x}) + \boldsymbol{\nu}_t \\ &= \mathbf{c}(\boldsymbol{w}_t) + \boldsymbol{A}(\boldsymbol{w}_t)\boldsymbol{x} + \boldsymbol{\nu}_t\,,\end{aligned} \tag{4.29}$$

where the vector $\boldsymbol{\nu}_t$ (also of length L) represents the noise in the measurements performed by the science camera, which are taken to be samples of a zero-mean stochastic process with a covariance matrix, $\boldsymbol{C}_t$, that is size $L \times L$. This can include any effect that models a physical detector, including but not limited to shot noise, readout noise, and thermal background. From Eq. 4.29, the goal of the regression is becoming clear, which is to estimate $\boldsymbol{x}$.

If the science camera and WFS are running at 1 kHz, the number of exposures, T, reaches 10^6 in under 17 minutes, and an astronomical observation of a single target could be hours long. This makes it necessary for the regression framework to allow for large numbers of exposures to be included practically. In order to proceed in this direction, the vectors $\boldsymbol{w}$ and $\hat{\boldsymbol{w}}$, as well as length LT vectors $\boldsymbol{y}$ and $\boldsymbol{\nu}$, are defined as concatenations of their individual exposure namesakes:

$$\boldsymbol{w} \equiv \begin{pmatrix} \boldsymbol{w}_0 \\ \vdots \\ \boldsymbol{w}_{T-1} \end{pmatrix}, \ \hat{\boldsymbol{w}} \equiv \begin{pmatrix} \hat{\boldsymbol{w}}_0 \\ \vdots \\ \hat{\boldsymbol{w}}_{T-1} \end{pmatrix}, \ \boldsymbol{y} \equiv \begin{pmatrix} \boldsymbol{y}_0 \\ \vdots \\ \boldsymbol{y}_{T-1} \end{pmatrix}, \ \boldsymbol{\nu} \equiv \begin{pmatrix} \boldsymbol{\nu}_0 \\ \vdots \\ \boldsymbol{\nu}_{T-1} \end{pmatrix}, \tag{4.30}$$

This concatenation of the noise vectors forces the assumptions that the total covariance matrix of $\boldsymbol{\nu}$, which will be similarly named $\boldsymbol{C}$, is made up of diagonal blocks of all the $\boldsymbol{C}_t$. Next, the concatenations for the set $\mathbf{c}(\boldsymbol{w}_t)$ and $\boldsymbol{A}(\boldsymbol{w}_t)$ are defined as:

$$\boldsymbol{d}(\boldsymbol{w}) \equiv \begin{pmatrix} \mathbf{c}(\boldsymbol{w}_0) \\ \vdots \\ \mathbf{c}(\boldsymbol{w}_{T-1}) \end{pmatrix}, \ \boldsymbol{Z}(\boldsymbol{w}) \equiv \begin{pmatrix} \boldsymbol{A}(\boldsymbol{w}_0) \\ \vdots \\ \boldsymbol{A}(\boldsymbol{w}_{T-1}) \end{pmatrix}, \tag{4.31}$$

where the $LT \times N$ matrix of functions $\boldsymbol{Z}(\boldsymbol{w})$ will be named the *grand system matrix*.

With the definitions in Eqs. (4.30) and (4.31), the concatenation of Eq. (4.29) over the T exposures can finally be written as:

$$\boldsymbol{y} = \boldsymbol{Z}(\boldsymbol{w})\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{w}) + \boldsymbol{\nu}\,. \tag{4.32}$$

Eq. (4.32) specifies the measured intensities $\boldsymbol{y}$ in terms of the true wavefronts $\boldsymbol{w}$, and the vector containing the parameters that represent the NCPA and exoplanet image $\boldsymbol{x}$. It is here that we have arrived at an equation that should remind the reader of Eq. (4.4) of the F13 method, as the derivations up to this point are nearly mathematically identical.

4.2.3 Regression

Given Eq. 4.32, three different scenarios can be determined for finding an estimator of $\boldsymbol{x}$ depending on what knowledge we have of the wavefronts $\boldsymbol{w}$ and their spatial statistical properties:

- Ideal estimation: the true wavefronts $\boldsymbol{w}$ are known exactly.
- Naïve estimation: only the measured wavefronts $\hat{\boldsymbol{w}}$ are known, but no statistical knowledge of the properties of $\boldsymbol{w}$ is given.
- bias-corrected estimation: the measured wavefronts $\hat{\boldsymbol{w}}$ and the statistical the properties of $\boldsymbol{w}$ are known.

The details of each estimators are given more thoroughly below.

Ideal estimation

In *ideal* estimation (a term from the errors in variables literature), the true AO residuals ($\boldsymbol{w}_t$) are used in the regression equations, resulting in an unbiased estimator (this is essentially the method used in Frazin (2013) and Rodack et al. (2018), though with large computational efficiency improvements that will be highlighted). Practically speaking, ideal estimation is not possible because only a measurement of the wavefront, ($\hat{\boldsymbol{w}}$), at least at the time of writing, can be known for optical and near-IR wavelengths, and it would instead require perfect knowledge of the independent variable, in this case $\boldsymbol{w}$. Ideal estimation is still important to understand because it will appear in the derivation of the bias-corrected estimate (which is realizable), and it will serve as a useful benchmark to evaluate the success or failure of the bias-corrected estimate in our simulations. The choice of regularized least-squares is a standard way to solve the linear regression problem, and is optimal in Gaussian noise if the measurement weights are chosen to be $\boldsymbol{C}^{-1}$. In terms of the vector being estimated, $\boldsymbol{x}$, the regularized least-squares cost function corresponding to Eq. (4.32) is:

$$\begin{aligned}\Phi(\boldsymbol{x}) = \frac{1}{2}\left[\boldsymbol{Z}(\boldsymbol{w})\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{w}) - \boldsymbol{y}\right]^{\mathrm{T}}\boldsymbol{S}\left[\boldsymbol{Z}(\boldsymbol{w})\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{w}) - \boldsymbol{y}\right] \\ + \frac{\beta}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}}\Xi(\boldsymbol{x} - \boldsymbol{x}_0)\end{aligned} \tag{4.33}$$

where $\boldsymbol{x}_0$ allows non-centered regularization, $\beta > 0$ is a regularization parameter, Ξ is an $N \times N$ symmetric (positive semi-definite) reglarization matrix, $\boldsymbol{S}$ is a $LT \times LT$ matrix of measurement weights with the $L \times L$ matrices $\{\boldsymbol{S}_t\}$ on the diagonal blocks, i.e.:

$$\boldsymbol{S} = \begin{pmatrix} \ddots & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \ddots \end{pmatrix}. \tag{4.34}$$

Note that $\boldsymbol{x}_0$ in Eq. (4.33) is unrelated to the linearization point $\boldsymbol{a}_0$ in Eq. (4.24). Common choices for the regularization matrix can be the identity matrix or finite-difference formulations that penalize gradients in the solution. Standard methods for choosing the regularization parameter include simulation and cross-validation. The quadratic form of the regularization term corresponds to a Gaussian prior density (centered on $\boldsymbol{x}_0$) in Bayesian inference. To maintain generality, the regularization terms are included in the derivation, but the simulation work below will assume $\beta = 0$.

The value of $\boldsymbol{x}$ that minimizes $\Phi(\boldsymbol{x})$ is called the ideal estimate, $\hat{\boldsymbol{x}}_{\mathrm{i}}$, and it can be shown (see Appendix A) that:

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{i}} \equiv \left[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w}) + \beta\Xi\right]^{-1} \\ \times \left\{\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\boldsymbol{w})] + \Xi\boldsymbol{x}_0\right\}.\end{aligned} \tag{4.35}$$

From this point forward, the assumption that $\boldsymbol{x}_0 = 0$ will be taken for convenience, but its inclusion is straightforward.

We also define the following quantities, which will prove useful in further cleaning up the notation:

$$\boldsymbol{Q}(\boldsymbol{w}) \equiv \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w}) \quad \text{and} \tag{4.36}$$

$$\boldsymbol{P}(\boldsymbol{w}) \equiv \left[\boldsymbol{Q}(\boldsymbol{w}) + \beta\boldsymbol{\Xi}\right]^{-1}. \tag{4.37}$$

Both $\boldsymbol{Q}(\boldsymbol{w})$ and $\boldsymbol{P}(\boldsymbol{w})$ are $N \times N$ matrices, and $\boldsymbol{P}(\boldsymbol{w})$ is the regularized inverse of $\boldsymbol{Q}(\boldsymbol{w})$.

Using Eqs. (4.36) and (4.37), the ideal estimate $\hat{\boldsymbol{x}}_{\mathrm{i}}$ in Eq. (4.35) can be rewritten as:

$$\hat{\boldsymbol{x}}_{\mathrm{i}} = \boldsymbol{P}(\boldsymbol{w})\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\boldsymbol{w})]\,. \tag{4.38}$$

It is important to remember that the presence of $\boldsymbol{y}$ in this equation means that $\hat{\boldsymbol{x}}_i$ will inherit the noise, $\boldsymbol{\nu}$ from the science camera measurement. This noise is the only stochastic process in the ideal estimator, allowing us to average over the statistics of $\boldsymbol{\nu}$:

$$\langle \boldsymbol{y} \rangle_{\boldsymbol{\nu}} = \boldsymbol{Z}(\boldsymbol{w})\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{w})\,, \tag{4.39}$$

because $\boldsymbol{\nu}$ is drawn from a zero-mean stochastic process. From Eq. (4.38), the mean of $\hat{\boldsymbol{x}}_{\mathrm{i}}$ is thus:

$$\langle \hat{\boldsymbol{x}}_{\mathrm{i}} \rangle_{\boldsymbol{\nu}} = \boldsymbol{P}(\boldsymbol{w})\boldsymbol{Q}(\boldsymbol{w})\boldsymbol{x} \tag{4.40}$$

$$\stackrel{\beta=0}{=} \boldsymbol{x}\,, \tag{4.41}$$

indicating the usual result that the ideal estimator is unbiased without regularization (i.e., $\beta = 0$). Similarly, the covariance of the ideal estimator is:

$$< \hat{\boldsymbol{x}}_{\mathrm{i}}\hat{\boldsymbol{x}}_{\mathrm{i}}^{\mathrm{T}} >_{\boldsymbol{\nu}} - < \hat{\boldsymbol{x}}_{\mathrm{i}} >_{\boldsymbol{\nu}} < \hat{\boldsymbol{x}}_{\mathrm{i}}^{\mathrm{T}} >_{\boldsymbol{\nu}}$$

$$= \boldsymbol{P}(\boldsymbol{w})\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{C}\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w})\boldsymbol{P}(\boldsymbol{w}) \tag{4.42}$$

$$= \boldsymbol{P}(\boldsymbol{w})\boldsymbol{H}(\boldsymbol{w})\boldsymbol{P}(\boldsymbol{w}) \tag{4.43}$$

$$\stackrel{\beta=0,\, \boldsymbol{S}=\boldsymbol{C}^{-1}}{=} \boldsymbol{P}(\boldsymbol{w}) \tag{4.44}$$

where Eq. (4.44) shows the simplification attained when the measurement weights $\boldsymbol{S}$ are chosen to be the inverse of the noise covariance (i.e., $\boldsymbol{S} = \boldsymbol{C}^{-1}$) and again without regularization. In Eq. (4.43), the $N \times N$ matrix $\boldsymbol{H}(\boldsymbol{w})$ is defined for reasons that become clear later as:

$$\boldsymbol{H}(\boldsymbol{w}) \equiv \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{C}\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w})\,. \tag{4.45}$$

If regularization is applied (i.e. $\beta \neq 0$), and/or the measurement weights $\boldsymbol{S}$ are chosen to be something other than $\boldsymbol{C}^{-1}$, the assumption is that they are chosen so the final result for $\hat{\boldsymbol{x}}_{\mathrm{i}}$ is "acceptably biased", which will not detract from the observation that ideal estimation is "unbiased".

Lastly, turning an eye towards an improvement in computational efficiency, the following observation, which follows from Eqs. (4.31) and (4.36) and writing out the matrix multiplication, proves quite useful:

$$\begin{aligned} \boldsymbol{Q}(\boldsymbol{w}) &= \sum_{t=0}^{T-1} \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t\boldsymbol{A}(\boldsymbol{w}_t) \\ &= T\left\langle \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t\boldsymbol{A}(\boldsymbol{w}_t)\right\rangle_\tau \\ &\equiv \boldsymbol{Q}_\tau \end{aligned} \tag{4.46}$$

which states the matrix $\boldsymbol{Q}(\boldsymbol{w})$ is a scaled time average over the interval defined by the T exposures that are measured, of a nonlinear function of the wavefronts $\{\boldsymbol{w}_t\}$. When T is much greater than the correlation time-scale of $\boldsymbol{w}_t$, $\boldsymbol{Q}_\tau$ depends largely only on the statistical properties of the wavefronts, and not their individual values. From Eqs. (4.37) and (4.46), we can easily then define:

$$\boldsymbol{P}_\tau \equiv \left[\beta\boldsymbol{\Xi} + \boldsymbol{Q}_\tau\right]^{-1}\,. \tag{4.47}$$

Next, using the definitions of $\boldsymbol{Q}_\tau$ and $\boldsymbol{P}_\tau$, Eq. (4.40), can be rewritten as:

$$\langle \boldsymbol{x}_{\mathrm{i}} \rangle_{\boldsymbol{\nu}} = \boldsymbol{P}_\tau\boldsymbol{Q}_\tau\boldsymbol{x} \overset{\beta=0}{=} \boldsymbol{x}\,. \tag{4.48}$$

This observation removes the need to store each individual millisecond in computer memory, as we now need only the sum of the computed quantities at each millisecond. The method can now be seen as computationally lightweight, solving one of the previous setbacks that was noted in Section **??**.

Naïve estimation

In *naïve* estimation (also a term from errors in variables literature, introduced at the end of Sec. **??**), the measurements of the AO residuals are plugged in to the equations wherever they call for $\boldsymbol{w}$, making it a technique that is realizable. The fact that measurements of the AO residuals differ from the true AO residuals is ignored, resulting in a biased estimate of $\boldsymbol{x}$. It is important to note that even if the wavefront measurements themselves are unbiased, the naïve estimator is still biased because of the uncertainty in the measurements. This is the result of using the methods of Frazin (2013) and Rodack et al. (2018) without the use of the ideal WFS, which spawned the discussions there of the major shortcoming of using the method. The naïve estimate, $\hat{\boldsymbol{x}}_{\mathrm{n}}$, is given by:

$$\hat{\boldsymbol{x}}_{\mathrm{n}} \equiv \boldsymbol{P}(\hat{\boldsymbol{w}})\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\hat{\boldsymbol{w}})]\,, \tag{4.49}$$

which of course is the same as Eq. (4.38), but with the wavefront measurements replacing the true wavefronts. The simulations shown below will use naïve estimation to good effect in estimating an NCPA with ~ 0.5 radian RMS based on only 1 minute of simulated sky time. However, when more precision is required, as when jointly estimating the exoplanet image with an NCPA of ~ 0.05 radian with 4 minutes of simulated sky time, the naïve estimate was too biased to be useful.

Biased estimation

The next step in the process is to define the *biased estimator* (even though the naïve estimator is biased, too). The only purpose of the biased estimator is to serve as a building block of the bias-corrected estimator. This biased estimate, $\hat{\boldsymbol{x}}_{\mathrm{b}}$, is defined as:

$$\hat{\boldsymbol{x}}_{\mathrm{b}} \equiv \boldsymbol{P}_{\tau}\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\hat{\boldsymbol{w}})]\,. \tag{4.50}$$

The biased estimate is nearly the same as the naïve estimate, except that $\boldsymbol{P}_{\tau}$ replaces $\boldsymbol{P}(\hat{\boldsymbol{w}})$, which removes some of the dependence on the measured wavefronts $\hat{\boldsymbol{w}}$. Now, the bias in the biased estimate is calculated so that we can later account for it with the Monte Carlo methods that follow. The wavefront measurement error is defined as:

$$\delta\boldsymbol{w} \equiv \hat{\boldsymbol{w}} - \boldsymbol{w}\,, \tag{4.51}$$

where the sign is chosen for convenience. Following the same sign convention, other errors that arise can be defined:

$$\delta\boldsymbol{Z}(\hat{\boldsymbol{w}}, \boldsymbol{w}) \equiv \boldsymbol{Z}(\hat{\boldsymbol{w}}) - \boldsymbol{Z}(\boldsymbol{w}) \tag{4.52}$$

$$\delta\boldsymbol{A}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t) \equiv \boldsymbol{A}(\hat{\boldsymbol{w}}_t) - \boldsymbol{A}(\boldsymbol{w}_t) \tag{4.53}$$

$$\delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w}) \equiv \boldsymbol{d}(\hat{\boldsymbol{w}}) - \boldsymbol{d}(\boldsymbol{w}) \tag{4.54}$$

$$\delta\mathbf{c}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t) \equiv \mathbf{c}(\hat{\boldsymbol{w}}_t) - \mathbf{c}(\boldsymbol{w}_t) \tag{4.55}$$

Using Eqs. (4.52) through (4.54), Eq. (4.50) can be rewritten as:

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{b}} &= \boldsymbol{P}_{\tau}\big[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w}) + \delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\big]\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\boldsymbol{w}) - \delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w})] \\ &= \hat{\boldsymbol{x}}_{\mathrm{i}} + \boldsymbol{P}_{\tau}\delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\boldsymbol{w})] \\ &\quad - \boldsymbol{P}_{\tau}\big[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w}) + \delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{S}\delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w})\big]\,,\end{aligned} \tag{4.56}$$

which makes use of Eq. (4.38), but replacing $\boldsymbol{P}(\boldsymbol{w})$ with $\boldsymbol{P}_{\tau}$. The next step is take the expectation of both sides of Eq. (4.56) with respect to the statistics of $\boldsymbol{\nu}$, following the same procedure that gave Eq. (4.40). The result is:

$$\begin{aligned}\langle\hat{\boldsymbol{x}}_{\mathrm{b}}\rangle_{\boldsymbol{\nu}} &= \langle\hat{\boldsymbol{x}}_{\mathrm{i}}\rangle_{\boldsymbol{\nu}} + \boldsymbol{P}_{\tau}\delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w})\boldsymbol{x} \\ &\quad - \boldsymbol{P}_{\tau}\big[\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w}) + \delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{S}\delta\boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w})\big]\,, \qquad (4.57) \\ &= \langle\hat{\boldsymbol{x}}_{\mathrm{i}}\rangle_{\boldsymbol{\nu}} + \boldsymbol{G}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{x} + \boldsymbol{g}_1(\hat{\boldsymbol{w}}, \boldsymbol{w}) + \boldsymbol{g}_2(\hat{\boldsymbol{w}}, \boldsymbol{w})\,, \qquad (4.58)\end{aligned}$$

which shows that $\hat{\boldsymbol{x}}_{\mathrm{b}}$ is biased relative to the ideal estimate $\hat{\boldsymbol{x}}_{\mathrm{i}}$. This bias has a term $\boldsymbol{G}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{x}$ that is linear in $\boldsymbol{x}$ as well as two additive terms, $\boldsymbol{g}_1(\hat{\boldsymbol{w}}, \boldsymbol{w})$ and $\boldsymbol{g}_2(\hat{\boldsymbol{w}}, \boldsymbol{w})$. The $N \times N$ matrix $\boldsymbol{G}(\hat{\boldsymbol{w}}, \boldsymbol{w})$ is given by:

$$\boldsymbol{G}(\hat{\boldsymbol{w}}, \boldsymbol{w}) \equiv \boldsymbol{P}_{\tau}\delta\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}}, \boldsymbol{w})\boldsymbol{S}\boldsymbol{Z}(\boldsymbol{w})\,, \tag{4.59}$$

which can be treated similarly to Eq. (4.46) to show $\boldsymbol{G}_\tau$ is a time average:

$$\begin{aligned}\boldsymbol{G}(\hat{\boldsymbol{w}}, \boldsymbol{w}) &= \boldsymbol{P}_\tau \sum_{t=0}^{T-1} \delta \boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t)\boldsymbol{S}_t \boldsymbol{A}(\boldsymbol{w}_t) \\ &= T\boldsymbol{P}_\tau \left\langle \delta\, \boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t)\boldsymbol{S}_t \boldsymbol{A}(\boldsymbol{w}_t)\right\rangle_\tau \\ &\equiv \boldsymbol{G}_\tau\,, \end{aligned} \tag{4.60}$$

However, unlike $\boldsymbol{Q}_\tau$, which can be approximated with knowledge of only the true wavefront statistics, $\boldsymbol{G}_\tau$ depends on both the statistics of the true wavefronts and their measurements. This joint dependence is also exhibited by the $N \times 1$ vectors $\boldsymbol{g}_1(\hat{\boldsymbol{w}}, \boldsymbol{w})$ and $\boldsymbol{g}_2(\hat{\boldsymbol{w}}, \boldsymbol{w})$, which also can be shown to be time averages:

$$\begin{aligned}\boldsymbol{g}_1(\hat{\boldsymbol{w}}, \boldsymbol{w}) &\equiv -\boldsymbol{P}_\tau\, \boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\delta \boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w}) \\ &= -\boldsymbol{P}_\tau \sum_{t=0}^{T-1} \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t \delta \mathbf{c}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t) \\ &= -T\boldsymbol{P}_\tau \left\langle \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t \delta \mathbf{c}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t)\right\rangle_\tau \\ &\equiv \boldsymbol{g}_{1\tau}\,, \end{aligned} \tag{4.61}$$

and

$$\begin{aligned}\boldsymbol{g}_2(\hat{\boldsymbol{w}}, \boldsymbol{w}) &\equiv -\boldsymbol{P}_\tau\, \delta\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{w})\boldsymbol{S}\delta \boldsymbol{d}(\hat{\boldsymbol{w}}, \boldsymbol{w}) \\ &= -\boldsymbol{P}_\tau \sum_{t=0}^{T-1} \delta\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t \delta \mathbf{c}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t) \\ &= -T\boldsymbol{P}_\tau \left\langle \delta\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t \delta \mathbf{c}(\hat{\boldsymbol{w}}_t, \boldsymbol{w}_t)\right\rangle_\tau \\ &\equiv \boldsymbol{g}_{2\tau}\,. \end{aligned} \tag{4.62}$$

Then, using Eqs. (4.60) through (4.62), the equation for the mean of the biased estimate in Eq. (4.58) is:

$$\langle \hat{\boldsymbol{x}}_{\mathrm{b}} \rangle_{\boldsymbol{\nu}} = \langle \hat{\boldsymbol{x}}_{\mathrm{i}} \rangle_{\boldsymbol{\nu}} + \boldsymbol{G}_\tau \boldsymbol{x} + \boldsymbol{g}_{1\tau} + \boldsymbol{g}_{2\tau}\,, \tag{4.63}$$

which is given in terms of the bias of the ideal estimator given in Eq. (4.48).

Using Eq. (4.50) and the definition of the noise covariance $\boldsymbol{C}$, we can arrive at the expression for the covariance of the biased estimator:

$$< \hat{\boldsymbol{x}}_{\mathrm{b}}\hat{\boldsymbol{x}}_{\mathrm{b}}^{\mathrm{T}} >_{\boldsymbol{\nu}} - < \hat{\boldsymbol{x}}_{\mathrm{b}} >_{\boldsymbol{\nu}} < \hat{\boldsymbol{x}}_{\mathrm{b}}^{\mathrm{T}} >_{\boldsymbol{\nu}} = \boldsymbol{P}_{\tau}\boldsymbol{H}(\hat{\boldsymbol{w}})\boldsymbol{P}_{\tau} \tag{4.64}$$

which is similar in appearance to covariance of the ideal estimator in Eq. (4.43), and where the $N \times N$ matrix $\boldsymbol{H}(\hat{\boldsymbol{w}})$ is defined as [see also Eq. (4.45)]:

$$\begin{aligned} \boldsymbol{H}(\hat{\boldsymbol{w}}) &\equiv \boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{SCSZ}(\hat{\boldsymbol{w}}) \\ &= \sum_{t=0}^{T} \boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t)\boldsymbol{S}_t\boldsymbol{C}_t\boldsymbol{S}_t\boldsymbol{A}(\hat{\boldsymbol{w}}_t)\,. \end{aligned} \tag{4.65}$$

This is an important result because $\boldsymbol{H}(\hat{\boldsymbol{w}})$ can be calculated with only the measured wavefronts. In this calculation, one can approximate $\boldsymbol{C}_t$ with an estimate $\hat{\boldsymbol{C}}_t$, to approximate the unknown true intensity (in photon units) with a measured number of photon counts when calculating the variance of the shot noise contribution to the science camera measurement noise covariance. Following from Goodman (2015), this is acceptable because the measured photon count rate is an unbiased estimator of the real photon count rate. That being said, $\boldsymbol{H}(\hat{\boldsymbol{w}})$ will only appear in the calculation of the error covariance matrix of the estimates, so any minor inaccuracy in this matter is not as critical, as it will not harm the estimates themselves.

Bias-corrected estimation

Lastly, the culmination of this work, *bias-corrected* estimation does treat the WME under the assumption that the spatial statistics of the true AO residual wavefronts are known. This statistical knowledge is used to create a set of Monte Carlo wavefronts (See Sec. 4.2.1) that form the basis of this technique. To the extent that the Monte Carlo wavefronts are representative of the random process that generates the true AO residual wavefronts, bias-corrected estimation converges to an unbiased estimate given a large enough number of samples. In other words, when the knowledge of the statistics is exact, the bias-corrected estimator is unbiased (and nearly as good as the ideal estimate). If the knowledge of the statistics is “good” but not exact, the bias-corrected estimator, as is the case for the simulations shown below, retains a small bias. Jumping in, the bias-corrected estimator, $\hat{\boldsymbol{x}}_\mathrm{c}$, is defined as:

$$\hat{\boldsymbol{x}}_\mathrm{c} \equiv \boldsymbol{P}_\tau \boldsymbol{Q}_\tau \big(\boldsymbol{G}_\tau + \boldsymbol{P}_\tau \boldsymbol{Q}_\tau\big)^{-1} \big(\hat{\boldsymbol{x}}_\mathrm{b} - \boldsymbol{g}_{1\tau} - \boldsymbol{g}_{2\tau}\big) \,, \tag{4.66}$$

which is expressed in terms of the biased estimator from Eq. (4.50). With the use of Eq. (4.48) and Eq. (4.63), we can verify that the bias-corrected estimate returns the same bias as the ideal estimate, namely that it is unbiased:

$$\langle \hat{\boldsymbol{x}}_\mathrm{c} \rangle_{\boldsymbol{\nu}} = \boldsymbol{P}_\tau \boldsymbol{Q}_\tau \boldsymbol{x} = \langle \hat{\boldsymbol{x}}_\mathrm{i} \rangle_{\boldsymbol{\nu}} \ . \tag{4.67}$$

Using Eqs. (4.64) and (4.66) , the covariance matrix of the bias-corrected estimator is:

$$\begin{aligned} < \hat{\boldsymbol{x}}_\mathrm{c} \hat{\boldsymbol{x}}_\mathrm{c}^\mathrm{T} >_{\boldsymbol{\nu}} - < \hat{\boldsymbol{x}}_\mathrm{c} >_{\boldsymbol{\nu}} < \hat{\boldsymbol{x}}_\mathrm{c}^\mathrm{T} >_{\boldsymbol{\nu}} &= \boldsymbol{P}_\tau \boldsymbol{Q}_\tau \big(\boldsymbol{G}_\tau + \boldsymbol{P}_\tau \boldsymbol{Q}_\tau\big)^{-1} \\ &\times \ \boldsymbol{P}_\tau \boldsymbol{H}(\hat{\boldsymbol{w}}) \boldsymbol{P}_\tau \left(\boldsymbol{G}_\tau + \boldsymbol{P}_\tau \boldsymbol{Q}_\tau\right)^{-\mathrm{T}} \boldsymbol{Q}_\tau \boldsymbol{P}_\tau \,, &(4.68) \\ &\overset{\beta=0}{=} \big(\boldsymbol{G}_\tau + \mathbb{1}\big)^{-1} \boldsymbol{P}_\tau \boldsymbol{H}(\hat{\boldsymbol{w}}) \boldsymbol{P}_\tau \big(\boldsymbol{G}_\tau + \mathbb{1}\big)^{-\mathrm{T}} &(4.69) \end{aligned}$$

where $^{-\mathrm{T}}$ indicates the transpose of the inverse of a matrix (unlike $\boldsymbol{Q}_\tau$, $\boldsymbol{G}_\tau$ is not symmetric), and $\mathbb{1}$ is the identity matrix (in this case, $N \times N$). Finally, we note that if there is no WME present, $\boldsymbol{G}_\tau = 0$, and the covariance of the bias-corrected estimator is exactly the same as the covariance of the ideal estimator shown in Eq. (4.43), as would be expected.

Monte Carlo approximations

In order to calculate the bias-corrected estimate and its error covariance matrix, the quantities: $\boldsymbol{Q}_\tau$, $\boldsymbol{P}_\tau$, $\boldsymbol{G}_\tau$, $\boldsymbol{g}_{1\tau}$ and $\boldsymbol{g}_{2\tau}$ must be obtained. This might seem impossible given the dependence on both the true and measured wavefronts, but alas, we can rely on the Monte Carlo sampling capabilities discussed in Sec. 4.2.1 to help.

Let us assume that the Monte Carlo simulator provides L samples of the AO residuals in the set $\{\breve{\boldsymbol{w}}_l\}$. If L and T are large enough we have:

$$\boldsymbol{Q}_{\mathrm{mc}} \equiv \frac{T}{L}\sum_{l=0}^{L-1} \boldsymbol{A}^{\mathrm{T}}(\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\boldsymbol{A}(\breve{\boldsymbol{w}}_l) \tag{4.70}$$

$$\approx T\left\langle \boldsymbol{A}^{\mathrm{T}}(\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\boldsymbol{A}(\breve{\boldsymbol{w}}_l)\right\rangle_{\mathrm{E}} \tag{4.71}$$

$$\approx T\left\langle \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{w}_t)\boldsymbol{S}_t\boldsymbol{A}(\boldsymbol{w}_t)\right\rangle_{\tau} = \boldsymbol{Q}_{\tau}\,, \tag{4.72}$$

where the weight matrix $\boldsymbol{S}_t$ is taken to be a constant that does not depend on t. Choosing the same matrix, $\boldsymbol{C}_t$ for all values of t prevents minimum mean-squared error estimation in the presence of noise that depends on t, which requires $\boldsymbol{S}_t = \boldsymbol{C}_t^{-1}$. Recalling Eq. (4.47), we define:

$$\boldsymbol{P}_{\mathrm{mc}} \equiv (\beta\boldsymbol{\Xi} + \boldsymbol{Q}_{\mathrm{mc}})^{-1}\;. \tag{4.73}$$

It is interesting to see that the regularization term, $\beta\boldsymbol{\Xi}$, that was included in Eq. (4.33) to improve the invertibility in the ideal estimator, could also be useful for the Monte Carlo approximations by mitigating Monte Carlo sampling errors.

The Monte Carlo approximations of the matrix $\boldsymbol{G}_\tau$ and the vectors $\boldsymbol{g}_{1\tau}$ and $\boldsymbol{g}_{2\tau}$ rely on both the ability to generate Monte Carlo wavefronts and their equivalent WFS measurements. Starting with Eq. (4.60), the matrix $\boldsymbol{G}_\tau$ can be approximated by the following Monte Carlo calculation:

$$\begin{aligned}\boldsymbol{G}_\tau &\equiv T\boldsymbol{P}_\tau\left\langle \delta\boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t,\boldsymbol{w}_t)\boldsymbol{S}_t\boldsymbol{A}(\boldsymbol{w}_t)\right\rangle_\tau\\ &\approx T\boldsymbol{P}_\tau\left\langle \delta\boldsymbol{A}^{\mathrm{T}}(\hat{\breve{\boldsymbol{w}}}_l,\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\boldsymbol{A}(\breve{\boldsymbol{w}}_l)\right\rangle_{\mathrm{E}}\\ &\approx \frac{T}{L}\boldsymbol{P}_{\mathrm{mc}}\sum_{l=0}^{L-1}\delta\boldsymbol{A}^{\mathrm{T}}(\hat{\breve{\boldsymbol{w}}}_l,\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\boldsymbol{A}(\breve{\boldsymbol{w}}_l)\\ &\equiv \boldsymbol{G}_{\mathrm{mc}}\end{aligned} \tag{4.74}$$

where $\hat{\breve{\boldsymbol{w}}}_l$ is a simulated measurement of a Monte Carlo wavefront $\breve{\boldsymbol{w}}_l$. Similarly, the vectors $\boldsymbol{g}_{1\tau}$ and $\boldsymbol{g}_{2\tau}$ in Eqs. (4.61) and (4.62) have the Monte Carlo approximations:

$$\boldsymbol{g}_{1\mathrm{mc}} \equiv -\frac{T}{L}\boldsymbol{P}_{\mathrm{mc}}\sum_{l=0}^{L-1}\boldsymbol{A}^{\mathrm{T}}(\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\delta\mathbf{c}(\hat{\breve{\boldsymbol{w}}}_l,\breve{\boldsymbol{w}}_l)\;,\ \text{and} \tag{4.75}$$

$$\boldsymbol{g}_{2\mathrm{mc}} \equiv -\frac{T}{L}\boldsymbol{P}_{\mathrm{mc}}\sum_{l=0}^{L-1}\delta\boldsymbol{A}^{\mathrm{T}}(\breve{\boldsymbol{w}}_l)\boldsymbol{S}_t\delta\mathbf{c}(\hat{\breve{\boldsymbol{w}}}_l,\breve{\boldsymbol{w}}_l)\,. \tag{4.76}$$

One thing to note about approximating the various $_\tau$ quantities with Monte Carlo methods is that all of the nonlinear dependencies on the wavefront are treated without approximation.

4.2.4 Numerical implementation of Frazin's algorithm

With the Monte Carlo approximations of the $_\tau$ quantities in hand, they can now be substituted in to Eq. (4.66) to arrive at the fully realizable, Monte Carlo bias-corrected estimate, which is given by [where $\boldsymbol{x}_{\mathrm{b}}$ from Eq. (4.50) is also substituted]:

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{cmc}} \equiv \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big(\boldsymbol{G}_{\mathrm{mc}} + \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big)^{-1} \times \\ \big\{\boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\hat{\boldsymbol{w}})] - \boldsymbol{g}_{1\mathrm{mc}} - \boldsymbol{g}_{2\mathrm{mc}}\big\}\,,\end{aligned} \tag{4.77}$$

with the covariance also computed by replacing the $_\tau$ quantities with their Monte Carlo approximations in Eq. (4.68).

At the end of this journey that has taken us through the three possible situations that can arise, we are left with a realizable, unbiased estimate even in the presence of WME, that is computationally light-weight, given that the assumptions on the knowledge of the wavefront statistical properties are met. As a reward to the reader for slogging through the math, the following summary of calculations required to fully implement the improved Frazin's algorithm (henceforth just referred to by the estimator names being applied) is provided:

1. Use the given equations for Monte Carlo sampling to calculate: $\boldsymbol{Q}_{\mathrm{mc}}$, $\boldsymbol{P}_{\mathrm{mc}}$, $\boldsymbol{G}_{\mathrm{mc}}$, $\boldsymbol{g}_{1\mathrm{mc}}$ and $\boldsymbol{g}_{2\mathrm{mc}}$.

2. As the observation takes place, gather the millisecond science camera intensities $\{\boldsymbol{y}_t\}$ and wavefront measurements $\{\hat{\boldsymbol{w}}_t\}$ (for a 1 kHz AO system), and accumulate the sums that depend on them:

   - $\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\hat{\boldsymbol{w}})] = \sum_{t=0}^{T-1} \boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t)\boldsymbol{S}_t[\boldsymbol{y}_t - \mathbf{c}(\hat{\boldsymbol{w}}_t)]$
     This quantity can be found in Eq. (4.77). This piece of the equation is accumulated during observation because it is the only piece that depends on the input data from the telescope.

- $\boldsymbol{H}(\hat{\boldsymbol{w}}) = \sum_{t=0}^{T} \boldsymbol{A}^{\mathrm{T}}(\hat{\boldsymbol{w}}_t)\boldsymbol{S}_t\hat{\boldsymbol{C}}_t\boldsymbol{S}_t\boldsymbol{A}(\hat{\boldsymbol{w}}_t)$ [Eq. (4.65)]

3. Calculate the Monte Carlo bias-corrected estimate from Eq. (4.77) (reproduced here for convenience) by plugging in the corresponding Monte Carlo approximations from 1 and the sum from the first bullet of 2.

$$\begin{aligned}\hat{\boldsymbol{x}}_{\mathrm{cmc}} \equiv \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big(\boldsymbol{G}_{\mathrm{mc}} + \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big)^{-1} \times \\ \big\{\boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Z}^{\mathrm{T}}(\hat{\boldsymbol{w}})\boldsymbol{S}[\boldsymbol{y} - \boldsymbol{d}(\hat{\boldsymbol{w}})] - \boldsymbol{g}_{1\mathrm{mc}} - \boldsymbol{g}_{2\mathrm{mc}}\big\}\end{aligned}$$

4. Calculate the Monte Carlo approximation to the error covariance matrix of $\hat{\boldsymbol{x}}_{\mathrm{c}}$ [see Eq. (4.68)] by plugging in the corresponding Monte Carlo approximations from 1 and the sum from the second bullet of 2:

$$\begin{aligned}\boldsymbol{C}_{\mathrm{cmc}} \equiv < \hat{\boldsymbol{x}}_{\mathrm{cmc}}\hat{\boldsymbol{x}}_{\mathrm{cmc}}^{\mathrm{T}} >_{\boldsymbol{\nu}} = \ \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big(\boldsymbol{G}_{\mathrm{mc}} + \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big)^{-1} \times \\ \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{H}(\hat{\boldsymbol{w}})\boldsymbol{P}_{\mathrm{mc}}\big(\boldsymbol{G}_{\mathrm{mc}} + \boldsymbol{P}_{\mathrm{mc}}\boldsymbol{Q}_{\mathrm{mc}}\big)^{-\mathrm{T}}\boldsymbol{Q}_{\mathrm{mc}}\boldsymbol{P}_{\mathrm{mc}}\end{aligned} \tag{4.78}$$

5. If the nonlinearity in the NCPA coefficients needs to be treated, re-linearize about the estimate, as per Eq. (4.22), and start over at 1.

Chapter 5 will follow this prescription in simulation to provide a demonstration and evaluation of the method.

CHAPTER 5

Frazin's algorithm: validation via simulation †

5.1 The simulator

In order to proceed with the verification of the regression model described in Sec. 4.2.3, a suite of simulation tools was developed in order to run numerical experiments using a millisecond imaging, adaptive optics coronagraph. An end-to-end simulator was constructed in order to create turbulent wavefronts, as well as models for the coronagraph and the AO system optical trains. The AO system includes a deformable mirror (DM) followed by a 4-sided pyramid WFS (PyWFS), and the coronagraph is a standard Lyot model, followed by a science camera. These simulations only require pupil planes and focal planes, and the propagation between the two planes is computed via fast Fourier transforms (see Goodman (2017)).

5.1.1 Simulated wavefronts and AO

The temporal evolution of the on-sky turbulent wavefronts must be specified in order to model the functionality of the AO system. We have taken the spatial statistics of the turbulence to be given by the von Kármán spectrum, with an outer scale parameter L_0, inner scale parameter l_0, and Fried parameter r_0 (see Goodman (2015)). To keep the model simple, and maintain realism in the temporal evolution, we implement the infinite phase screen method introduced by Assémat et al. (2006) via the hcipy package [66].

We simulate the DM surface with 2D cubic splines, with the knots placed in a 20×20 grid. In this way, the knots play the role of the actuators and the height

†This chapter has been published previously as Rodack et al. (2021).

of any point on the surface is given by the spline interpolator function. With the circular pupil inscribed on the DM surface, the resulting control radius is $10\,\lambda/D$.

Following the DM, a beam splitter separates the coronagraph and wavefront sensor optical trains. In our simulations, the beam splitter's only effect is to divide the available photon flux between the optical trains.

In the simulation experiments below, the PyWFS is chosen in large part because it is being used for many modern AO systems including MagAO-X [46] and SCExAO [43].

The input to the PyWFS regression model is the AO residual wavefront at time-step t, represented as $\boldsymbol{w}_t$, while the output is the estimate of the same, represented as $\hat{\boldsymbol{w}}_t$, as summarized in Eq. (4.9). The PyWFS model includes circular modulation of the beam, simulated as a discrete set of tips and tilts. This formulation allows for the effects of noise, spatial bandwidth, and aliasing to be included in the measurement of the wavefront. An example of an AO residual phase, $\phi(\boldsymbol{w}_t)$, its measurement error, $\phi(\boldsymbol{w}_t) - \phi(\hat{\boldsymbol{w}}_t)$, and the corresponding PyWFS intensity at t, are provided in Fig. 5.1.

Progressing to the coronagraph from the beam splitter, the wavefront that arrives at the entrance pupil of the coronagraph is the same as the wavefront at the WFS, but with the addition of the NCPA phase. For simplicity, the cumulative sum of any NCPA is treated as a single phase error in the coronagraph entrance pupil. This can be thought of as the total phase error induced by the individual optical elements in the coronagraph, aberrations upstream of the beam splitter that are not corrected by the AO system (e.g., island modes), as well as aberrations in the WFS optical train that are flattened by the AO system (for which the inverse of would appear in the coronagraph optical train). The electric field in the coronagraph entrance pupil is given by Eq. 4.11.

5.1.2 Numerical coronagraph model

For these simulations we choose a Lyot coronagraph (See Fig. 1.4) with a binary focal plane mask (FPM) occulter that is circular in shape with a radius of $1.5\,\lambda/D$,

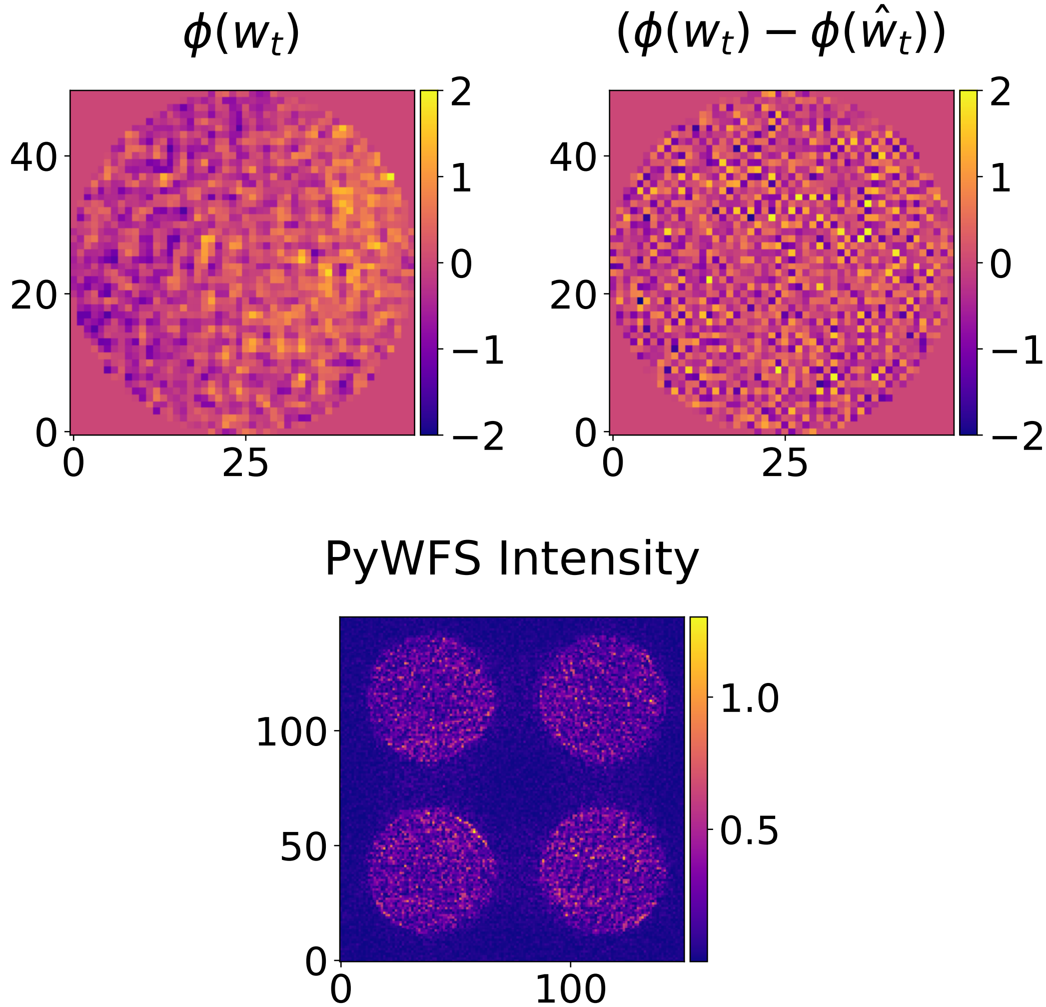


Figure 5.1: Example of wavefront measurement with the simulated PyWFS. *top left:* True AO residual (radian). *top right:* Error in measured AO residual (radian). *bottom:* Intensity at PyWFS detector, in normalized units.

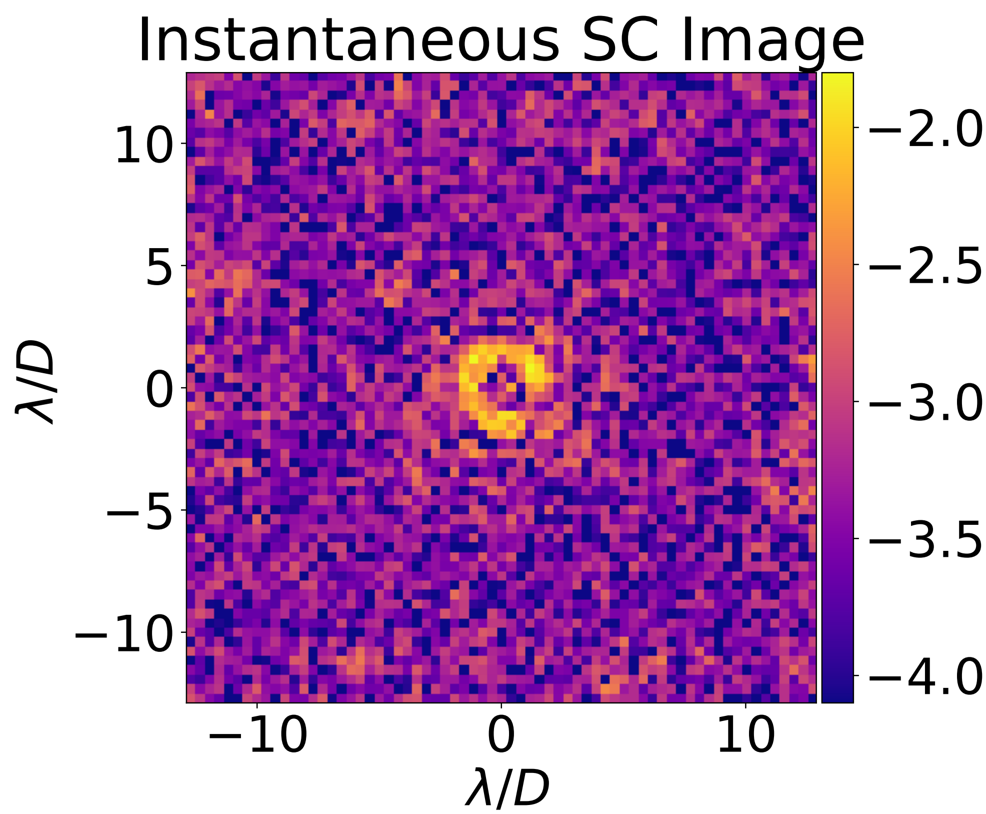


Figure 5.2: Log scale 1 ms exposure science camera image in contrast units of an 8 magnitude source at $\lambda = 1$ micron with 10% spectral bandwidth.

and a Lyot Stop with a diameter 90% of the pupil diameter. To remind the reader, the complex-valued electric field at the $l\underline{\text{th}}$ science camera detector pixel, v_l, can be represented in terms of the complex-valued electric field in the coronagraph entrance pupil, $u_p(\boldsymbol{w}_t, \boldsymbol{a})$, via the linear coronagraph operator, an $L \times P$ matrix, $\boldsymbol{D} = \{D_{lp}\}$ (L is the number of science camera pixels, P is the number of pixels in the discretization of the entrance pupil):

$$v_l(\boldsymbol{w}_t, \boldsymbol{a}) = \sum_{p=0}^{P-1} D_{lp} u_p(\boldsymbol{w}_t, \boldsymbol{a}) \,, \tag{5.1}$$

where $u_p(\boldsymbol{w}_t, \boldsymbol{a})$ is given by (4.11). Note that (5.1) is the form of a matrix-vector multiplication and is ideally suited to being carried out on a graphics processing unit (GPU). $\boldsymbol{D}$ comes from a computational model of the coronagraph optical train. For a Lyot coronagraph, such a model is provided by the operator:

$$\boldsymbol{D} = \{\mathcal{F}\{\mathcal{L} \times \ \mathcal{F}[\mathcal{M} \ \times \ \mathcal{F}(\ \ )]\}\} \,, \tag{5.2}$$

where $\mathcal{F}$ is the 2D discrete Fourier Transform operator, $\mathcal{L}$ is a mask representing the Lyot Stop, and $\mathcal{M}$ is a focal plane mask. The actual construction of $\boldsymbol{D}$ for several systems is given in Appendix B. Finally, the noise-free (i.e. true) representation of the intensity impinging on the science camera is given by Eq. 4.28 above.

This intensity is normalized such that a point source propagated to the final focal plane without a coronagraph has a maximum value of unity, centered at the origin of the science camera. This converts the effective units of the science camera measurements to be in "contrast" as defined as the ratio of coronagraphic intensity from our simulated source divided by the non-coronagraphic intensity of an on-axis point source. Scaling in this fashion allows for direct calculation of contrast values.

Finally, we simulate the science camera's measurement of the intensity by adding photon counting (shot) noise and detector readout noise. An example of such a measurement can be found in Fig. 5.2. The number of photons incident on the detector is determined by choice of target star magnitude, science wavelength, and spectral bandwidth, the optical system beam splitter ratio, and chosen values for system throughput, all of which are specified in Table (5.1). To simulate the noisy measurement, the intensity, $i(\boldsymbol{w}_t, \boldsymbol{a}, \boldsymbol{p})$, in units of contrast, is converted to units of photons. σ_{RN} is then chosen to represent the readout noise in photon counts per pixel, per read out (every 1 millisecond). The standard deviation of the intensity, in photon units, in the science camera measurement, is thus given as

$$\sigma_{l,ph} = \sqrt{i_{l,ph} + \sigma_{RN}^2}\,, \tag{5.3}$$

allowing for the calculation of the noise term in the $l\underline{th}$ pixel via sampling from a zero-mean, $\sigma_{l_{ph}}$ Normal distribution. This noise term, n_l , separate from the noise term in the WFS measurement in Eq. (4.9) because they come from different detectors, is then converted back to units of contrast and added to the result of Eq. (4.28), arriving at the simulated result of Eq. (4.29).

5.1.3 Simulation parameters

In the numerical experiments described below, the sensing and science wavelengths are both chosen to be Y band, centered around $1.036\,\mu$m. The Y band was essentially an arbitrary choice corresponding to a possible MagAO-X instrument observing case; the methods discussed would work at shorter or longer wavelengths. Although modern AO systems typically use separate wavelength bands for the WFS and coronagraph optical trains, we chose to use the same band to match the current expectation for a first on-sky implementation of this method. This mitigates chromatic effects that may arise when these functions are performed at different wavelengths. A 6.5 meter, circularly symmetric aperture with no central obscuration nor spiders is chosen to define the telescope pupil. The light is taken to be quasi-monochromatic. The photon fluxes, which are needed for noise statistics, calculated from apparent magnitude 6 or 8 sources with a 10% spectral bandwidth and a 50% overall throughput. The spectral bandwidth is only considered in the calculation of the number of photons/ms incident on the telescope, not for analysis on polychromatic simulation. This results in a total of 28,700 photons/ms for the magnitude 8 source and 182,200 photons/ms for the magnitude 6 source. 70% of the photon flux was apportioned to the WFS, while 30% was apportioned to the coronagraph. This non-standard way of splitting the photons was chosen so that changing from a magnitude 6 to magnitude 8 target in the simulations would not incur a penalty in the resulting Strehl ratio from the AO system. We did not run experiments with other ratios.

The AO system uses a leaky integrator control system with a loop gain of 0.35, with a two frame correction delay, and provides a closed-loop Strehl ratio of ~ 0.73 in Y band. The turbulence parameters are specified in Table 5.1. The control radius of the AO system (defined by the maximum correctable spatial frequency by the DM) is $10\lambda/D$ in both the x and y directions, leading to a square corrected region, and a residual bright halo outside. A 1 kHz frequency is chosen for both the wavefront and the science camera measurements, and they are assumed to collect simultaneous telemetry, meaning that $\boldsymbol{y}_t$, the vector of science camera intensities at all L pixels measured at time-step t, results from the wavefront $\boldsymbol{w}_t$, with measurement $\hat{\boldsymbol{w}}_t$. These choices are representative of modern ground-based AO systems equipped with high frame rate science cameras. This AO performance, and the resulting raw contrast provided by the Lyot Coronagraph without any NCPA present, is consistent with the current theoretical performance of such a simple integrator control system given by Males and Guyon Males and Guyon (2018).

The turbulent wavefront phase, $\boldsymbol{\phi}^(\boldsymbol{w}_t^-)$, is generated following Section 5.1.1 and corresponds to a 50×50 pixel grid inscribed with a circular aperture with a radius of 25 pixels. The total number of pixels in this inscribed circle is $P = 1976$. The turbulence model used in the simulation is a phase only atmosphere, with $L_0 = 25\,\mathrm{m}$ and $r_0 = 0.3\,\mathrm{m}$. The phase $\boldsymbol{\phi}(\boldsymbol{w}_t)$ is constructed as the sum of 6 layers (as scintillation effects are ignored to simplify the discussion in this work), each with their own wind direction vector, corresponding to a 10 m/s speed. The choice of 6 layers collapsed together in the entrance pupil helps ensure that each phase screen is unique.

The science camera (in the final focal plane of the coronagraph) consists of a 67×67 pixel square array, with 0.39 pixels per λ/D sampling. A value of 0.3

ph/pix/read of readout noise is assumed. This number of pixels in the detector frame, as well as the readout noise value, running at a 1kHz frame rate, is within specifications for modern, commercially available EMCCD cameras running at an acceptable gain setting. Table (5.1) summarizes the general parameters that are used in the numerical experiments.

The NCPA for the simulated experiments is chosen to be the linear combination of 6 radial orders of Zernike polynomials (Noll indices 4-36, so, $N_a = 32$), following Eq. (4.11). When needed for the joint estimate, a 13×8 object source grid corresponding to the $N_p = 104$ sky-angles for which the scene (exoplanet brightness coefficients) are to be estimated, is used. This source grid is a collection of point sources created such that their images propagated through the coronagraph are spaced by $1\lambda/D$, and have an inner edge at $3\lambda/D$ at the closest, and an outer edge at $10\lambda/D$. The choice of this scene is to demonstrate how the techniques presented can recover both point sources (exoplanets) and extended sources (such as circumstellar disks), located at a few times the classical optics resolution limit of λ/D. The brightest point is (1×10^{-4}) in the contrast units, while the faintest point is (1.8×10^{-6}), with an average brightness of 3.9×10^{-5}.

5.2 Phase A numerical experiment

The objective of the first experiment performed is to estimate and compensate for the NCPA. Compensating for the NCPA using only one minute of observation time provides a good starting position for the more demanding combined estimation of the residual NCPA and the exoplanet image in "Phase B," described in Section 5.3.

5.2.1 Estimating the NCPA

Phase A simulates a 1 minute observation, consisting of $60,000$ synchronized millisecond exposures in the science camera and WFS, the purpose of which is to provide a coarse estimate of large NCPA, which have an RMS phase of 0.52 radian. The Phase A estimate uses the naïve regression model, which has unwanted bias

Simulation Parameter	**Symbol**	**Value or Type**
Pixels in Entrance Pupil	P	1976
True AO Residual	$\boldsymbol{w}_t$	vector of length P
WFS Measurement	$\hat{\boldsymbol{w}}_t$	vector of length P
Pupil Diameter	D	6.5 m
Coronagraph Operator	$\boldsymbol{D}$	Lyot
Fourier Transform Operator	$\mathcal{F}$	
FPM	$\mathcal{M}$	binary; radius: $1.5\lambda/D$
Lyot Stop	$\mathcal{S}$	binary; diameter: $0.9D$
WFS Model	$\mathcal{W}$	Modulated PyWFS
PyWFS Modulation Radius		$3\,\lambda/D$
Science Camera	SC	EMCCD
Pixels in SC	L	4489
SC/WFS Cadence		1kHz, synchronized
SC Pixel Array Shape		67×67 pixels
SC Pixels per λ/D		0.39
SC Readout Noise	$\sigma_{I_{ph}}$	0.3 photon counts/px
Wavelength	λ	Y band ($1\mu m$)
Beam Splitter Ratio		70/30
Source Apparent Magnitude		8 (and 6)
Source Spectral Bandwidth		10%
Photons per ms in SC		$8,718$ $(55,008)$
Photons per ms in WFS		$20,342$ $(128,351)$
Fried Parameter	r_0	0.3 m
Inner Scale	l_0	0.01 m
Outer Scale	L_0	25 m
Turbulence Model		von Kármán; 6 layers; no scintillation; frozen flow
Strehl at $\lambda = 1\mu m$		0.73
AO Correction Delay		2 Frames
NCPA Zernike Modes		4–36 (6 radial orders)
NCPA RMS		0.52 radian
True Object Grid		13×8 points
Grid Locations		x: $[3\lambda/D, 10\lambda/D]$ y: $[-6\lambda/D, +6\lambda/D]$ Spacing of $1\lambda/D$
Relinearization Point	$x_{a,n}$	iteration n linearization point

Table 5.1: Summary of the defining simulation parameters for the performed numerical experiments.

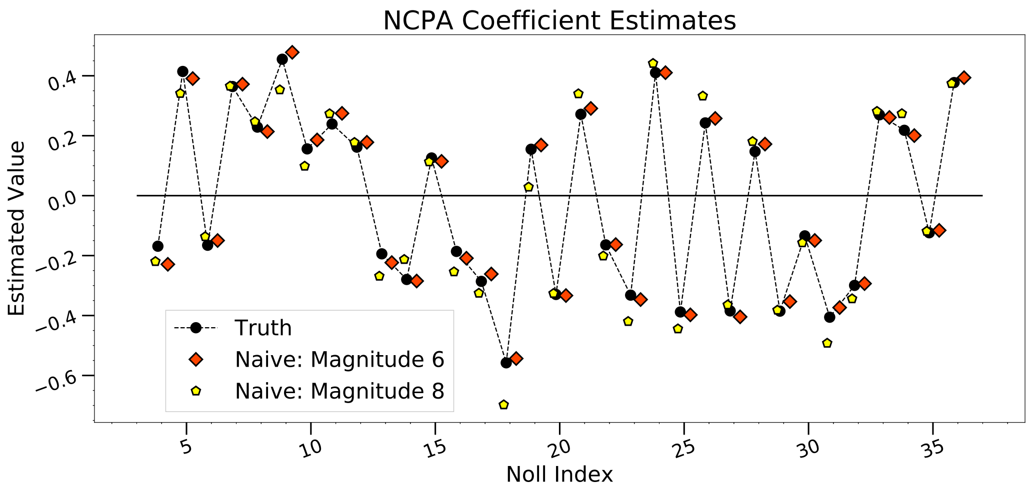


Figure 5.3: The estimated aberration Coefficients from the simulated experiments, following five relinearization iterations. Included are the true, starting NCPA coefficients, and their naïve estimates using both a magnitude 6 (RMS error of 2.14×10^{-2} radian) and magnitude 8 source (RMS error of 5.95×10^{-2} radian).

that increases increases as the wavefront measurement error gets larger. The initial linearization point was taken to the zero vector, ie., $\boldsymbol{a}_0 = 0$. The Phase A experiment was repeated for both a magnitude 6 and magnitude 8 target source. The AO system being simulated is not photon starved at magnitude 8, so although fewer photons are being observed compared to the magnitude 6 source, the Strehl ratio only decreases from 0.76 to 0.73 when going from the magnitude 6 source to the magnitude 8 source. The magnitude 8 source has a larger wavefront measurement error, which leads to larger bias in the naïve estimate.

Fig. 5.3 shows the results of the naïve estimate of the aberration coefficients composing the NCPA, following a 5 iteration treatment of the nonlinearity (see Section 5.2.5.2.2), for both the magnitude 6 and 8 sources. The root mean squared (RMS) error of the magnitude 6 source estimates of the aberration coefficients is 0.0214 radian, whereas for the magnitude 8 source the RMS error is 0.0595 radian, approximately a factor of 3 worse.

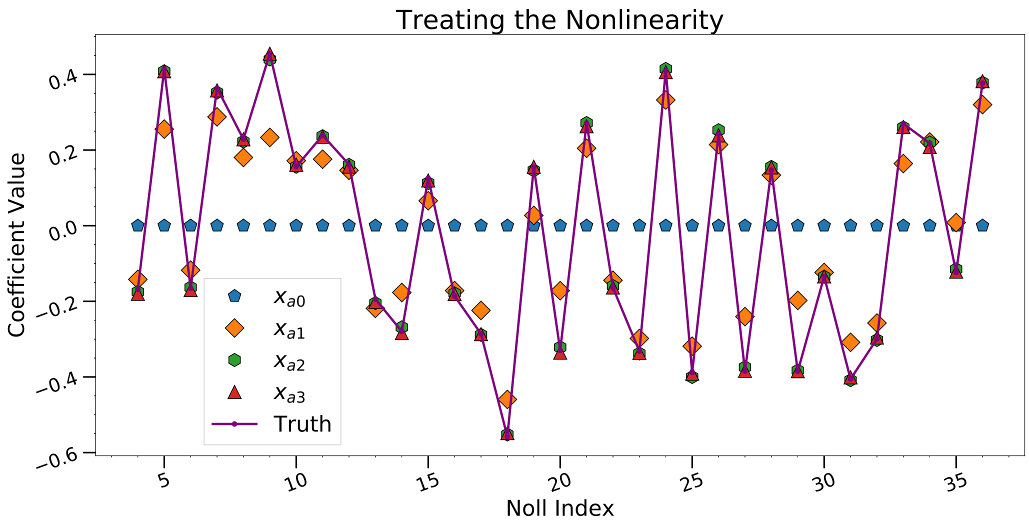


Figure 5.4: Successive relinearization points, $x_{a,n}$, for the Phase A experiment, starting from all zeros, and then using the previous iteration's estimate.

5.2.2 Treating the nonlinearity

In order to effectively treat the nonlinearity in the NCPA, the regression algorithm is run iteratively, successively relinearizing about the updated estimate values. Note that each iteration can be computed using the same set of observed data. The number of iterations that are needed can be determined by monitoring when the change in the updated linearization point from one iteration to the next becomes insignificant. This process is illustrated in Fig. 5.4, starting about the zero vector, and continuing using the previous iteration's estimate as the new linearization point. The only cost of adding relinearization steps is computation time since new observations are not needed. In these simulations, three iterations effectively estimated the aberration coefficients, but we applied five for good measure.

5.2.3 Compensating for the NCPA

Once we have estimated the NCPA, we can use a non-common path DM (such as the one in MagAO-X [7]) to compensate for them. A DM command that performs the compensation can be calculated by a least-squares fit of the NCPA optical path difference (OPD) map to the DM command modes. The resulting phase maps in the

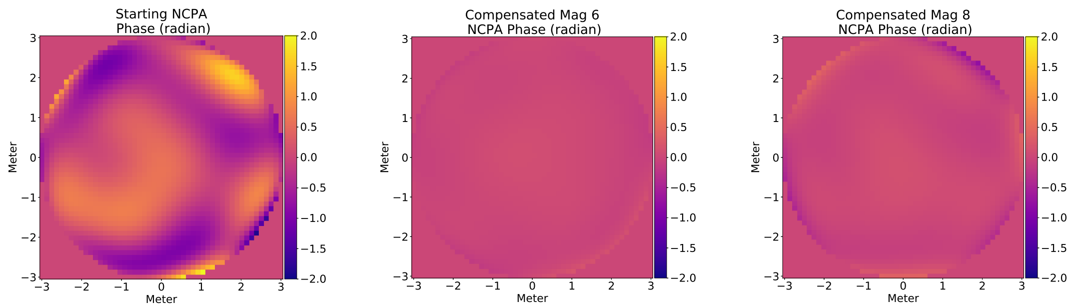


Figure 5.5: *left:* Starting NCPA phase (in radian), with RMS 0.52 radian, in the coronagraph entrance pupil. *middle:* The phase in the coronagraph entrance pupil with the magnitude 6 source naïve estimate used for compensation. The residual phase has RMS 0.0495 radian. *right:* The phase in the coronagraph entrance pupil with the magnitude 8 source naïve estimate used for compensation. The residual phase has RMS 0.11 radian.

coronagraph entrance pupil after DM compensation for our simulations are shown in Fig. 5.5 in the top right (magnitude 6 source) and bottom (magnitude 8 source) panels. As we expect from the the RMS error of the estimated aberration coefficients, compensating the NCPA via the magnitude 6 source estimate does a better job, achieving a reduction in RMS phase due to the NCPA from 0.52 radian to 0.0495 radian, a 10.5× improvement. Using the magnitude 8 source estimate achieves a reduction in RMS phase from 0.52 radian to 0.11 radian, a 4.74× improvement. In order to proceed in Phase B, we will adopt the better compensated NCPA while the magnitude 8 source in order to demonstrate being able to estimate a small NCPA jointly with the exoplanet image. This choice could represent an observation sequence where the NCPA is calibrated on a bright star, and then the telescope would slew to a fainter science target. If this is done, any change to the NCPA due to the slewing would likely remain small, especially if the calibration star is close to the target on sky. If additional NCPA is incurred through this slewing, it can be estimated while observing the target star, as will be discussed in the following section. The resulting time averaged images in the science camera in contrast units for the cases of pre- and post compensation, can be seen in Fig. 5.6.

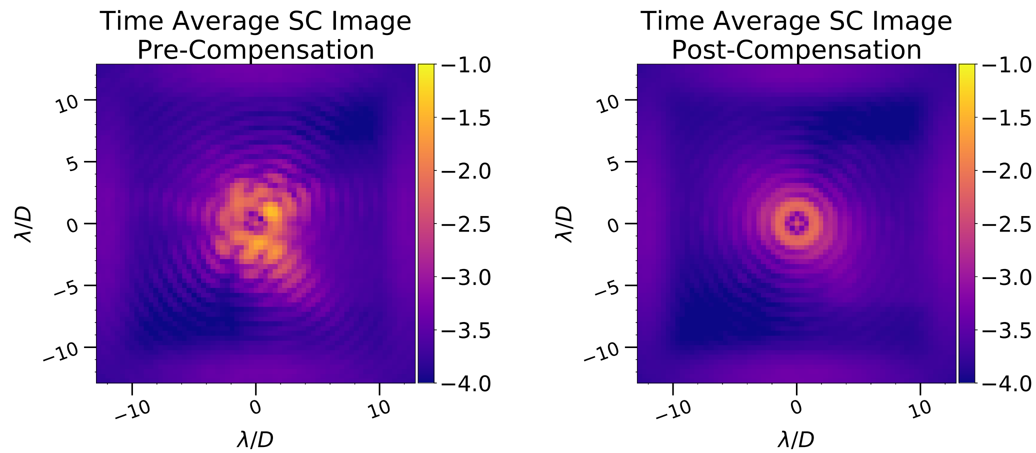


Figure 5.6: Log scale Time averaged images in the science camera for the Phase A and Phase B experiments in contrast. *left:* the manifestation of the 0.52 radian RMS ($\approx \boldsymbol{\lambda/13}$) NCPA is present, and shows a significant degradation to the coronagraphic image. *right:* the estimated NCPA has been applied to a DM to compensate, leaving behind the manifestation of a 0.0495 radian RMS residual NCPA.

Magnitude of Source	**RMS Error (radian)**	**Compenasted RMS Phase (radian)**
6	0.0214	0.0495
8	0.0595	0.11

Table 5.2: Summary of the resulting root mean square (RMS) error using the naïve estimate based on 1 minute of observations to solve for the NCPA, and the residual RMS phase left after using the estimates to compensate the NCPA. Note the starting RMS phase was 0.52 radian.

5.3 Phase B: joint estimation

The second experiment performed is to jointly monitor the residual NCPA left over after Phase A, while also estimating any present exoplanetary image. Recalling the results from Section 5.2, the NCPA present in the coronagraph optical train has an RMS phase error of 0.0495 radian, and the time average image in the science camera can be seen in the right frame of Fig. 5.6. Phase B is a 4 minute observation of a magnitude 8 star taking place directly after the 1 minute Phase A observation described in Section 5.2. Phase B gathers $240,000$ millisecond synchronized exposures in both the WFS and science camera telemetry streams. The same process of treating the nonlinearity described above is performed here, starting by linearizing the regression equations with the N_a NCPA coefficients and N_p exoplanet coefficients set to zero (see Eq. (4.32)), and successively updating with the previous iteration bias-corrected estimate, for a total of five iterations. Note that after compensating for the NCPA with the DM, the average level of stellar light in the science camera where the exoplanet image is being estimated is $\approx 3 \times 10^{-4}$ in contrast units.

5.3.1 The naïve estimate

As we have seen in the Phase A experiment, the naïve estimate can perform quite well. A combination of factors including but not limited to WFS measurements of reasonable quality and the large NCPA, provided conditions under which the naïve estimate recovers a useful estimate of the NCPA. However, we found that when using the naïve estimate to solve for the exoplanet image coefficients and smaller NCPA coefficients, the bias from the WME renders the estimate unacceptable, as can be seen in Fig.5.7 from the Phase B experiment. In this figure, the estimated imaging coefficients have been reshaped and plotted as a 13×8 image to allow for easy comparison to the object source grid. It is clear this is not a useful estimate of the exoplanet image, as it results in an RMS error in brightness (contrast units) of 1.63×10^{-4}, not to mention the fact that the morphology of the image is almost entirely lost.

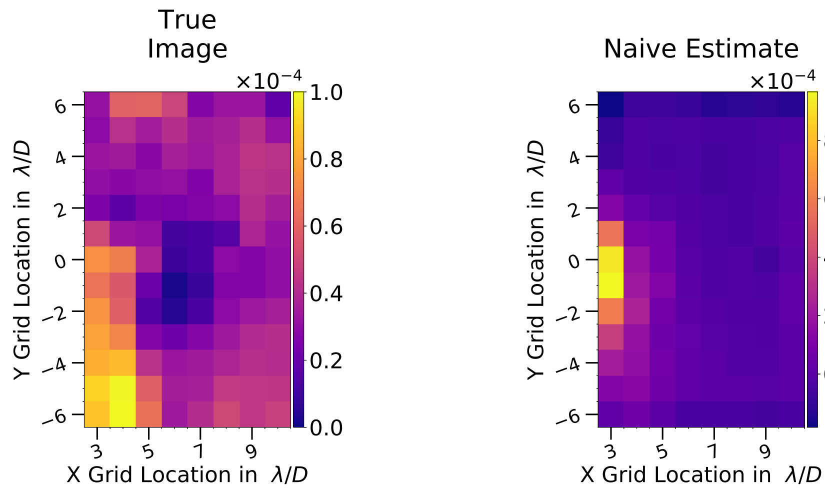


Figure 5.7: The estimated exoplanet imaging coefficients from the simulated Phase B experiment, following five relinearization iterations. *left:* the true object source grid. *right:* Naïve estimate of object source grid. All units are in contrast.

5.3.2 The bias-corrected estimate

Choosing Monte Carlo wavefronts

As described in Sec. 4.2.3, the bias-corrected estimator uses a set of Monte Carlo wavefronts created using the knowledge of the spatial statistics of the AO residual wavefronts. To generate the Monte Carlo wavefronts, we calculate the mean and covariance matrix of all of the T AO residual wavefront vectors, the set of which is represented as $\{\boldsymbol{w}_t\}$. While it is not possible to know the exact mean and covariance matrix of the AO residuals on a real system (since only the measurements $\{\hat{\boldsymbol{w}}_t\}$ are available), we ignore that complication for the purposes of this study. We then draw the Monte Carlo wavefronts from a P-dimensional multivariate normal with sample mean and covariance of the AO residual wavefronts. The Monte Carlo wavefronts are thus not temporally correlated like the true AO residual wavefronts, but there is no need for them to be, so long as their distribution has the same moments as the time-averaged moments of the stochastic process governing the true AO residual wavefronts.

Although our Monte Carlo wavefronts have the same 2nd order statistics as the

$T = 240,000$ AO residual wavefronts in the 4 minute on-sky data set, (apart from the error due to the finite number of Monte Carlo samples, which is mitigated by using $480,000$ separate Monte Carlo samples), the accuracy of the bias-corrected estimate is limited by the fact that the true AO residual wavefronts did not obey multivariate normal statistics. We know this because the univariate (i.e., single pixel) AO residual values failed the Anderson-Darling test for univariate normality. (Univariate normality of all of the individual variables is a necessary, but not sufficient, condition for multivariate normality.) These statistical details are discussed in Sec. 4.2.1.

Estimating the exoplanet image

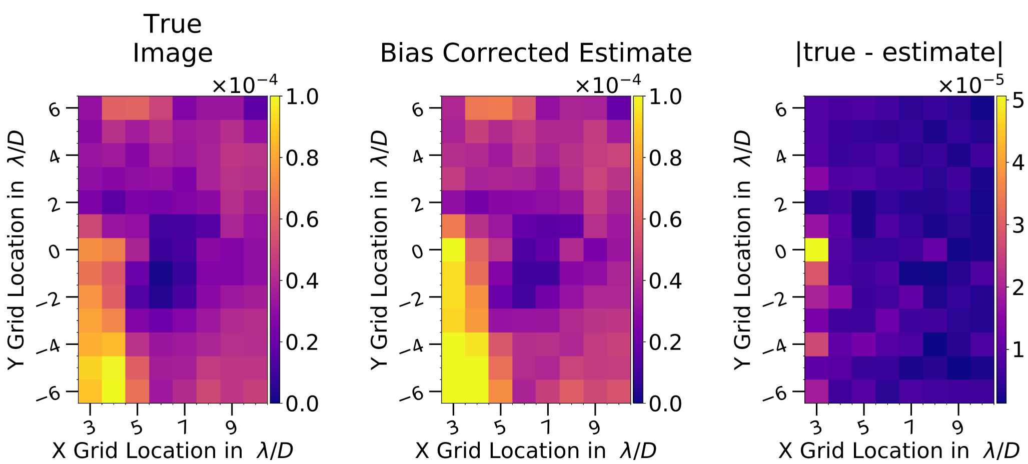


Figure 5.8: The estimated exoplanet imaging coefficients from the simulated Phase B experiment, following five relinearization iterations, reshaped in to the 13×8 grid. Left: the true object source grid. Middle: Bias-corrected estimate of object source grid. Right: The absolute value of the error in the estimate. All units are in contrast.

Performing the bias-corrected estimate shows a marked improvement over the naïve estimate, recovering much of the detail in the true signal, as can be seen in Fig. 5.8. The true image is in the left frame, the bias-corrected estimate is in the middle, and the error in the estimate is in the right frame. The RMS error for the bias-corrected estimate of the exoplanet brightness coefficients, averaged over all of the points is 9.5×10^{-6}, a $17.25\times$ improvement over the naïve estimate. Looking more closely at the error in the estimated coefficients, there are three grid points at a distance near $3\,\lambda/D$, where the PSF is bright, with an RMS error of about 3.5×10^{-5}. The 13 grid points with a distance near $10\,\lambda/D$, where the PSF is not a bright, have an RMS error of about 4.2×10^{-6}. Note these the RMS errors in the estimates are indicative of the $1\,\sigma$ contrast achieved, since the estimate error has very little dependence on the exoplanet brightness (assuming that it is much less bright than the star). The ideal estimates of the exoplanet coefficients are better by roughly a factor of 20. Note that we did not attempt to reduce the PSF brightness with dark hole techniques (in principle, dark hole techniques should be able to leverage the NCPA estimates).

The vector of exoplanet image coefficients is plotted in Fig. 5.9. The top frame shows the coefficients for each of the estimates as well as the truth values, and the bias of the naïve estimate is rather evident. In order to see the much smaller residual bias in the bias-corrected estimate, the middle frame in this figure does not contain the naïve estimate values. The bottom frame in this figure shows the estimated values subtracted from the true values. As explained above, this residual bias in the bias-corrected estimate is largely due to the fact that the true AO residuals do not obey multivariate normal statistics, while the Monte Carlo wavefronts do.

In principle (i.e., assuming that the Monte Carlo wavefronts have the same statistics as the true AO residual wavefronts), the bias-corrected estimate should be nearly as accurate as the ideal estimate, but here we see that the ideal estimate (which is unbiased) is very close to the true values, with an RMS error in contrast units of 8.83×10^{-7}. The plots in Fig. 5.8 and corresponding RMS error numbers show that the simple assumption that the statistics of $\boldsymbol{w}$ are governed by a multivariate normal, with mean and covariance matrix calculated from the 4 minute set of millisecond wavefronts, is a useful assumption (at least when compared to naïve estimation), but that it is not adequate to remove all of the bias.

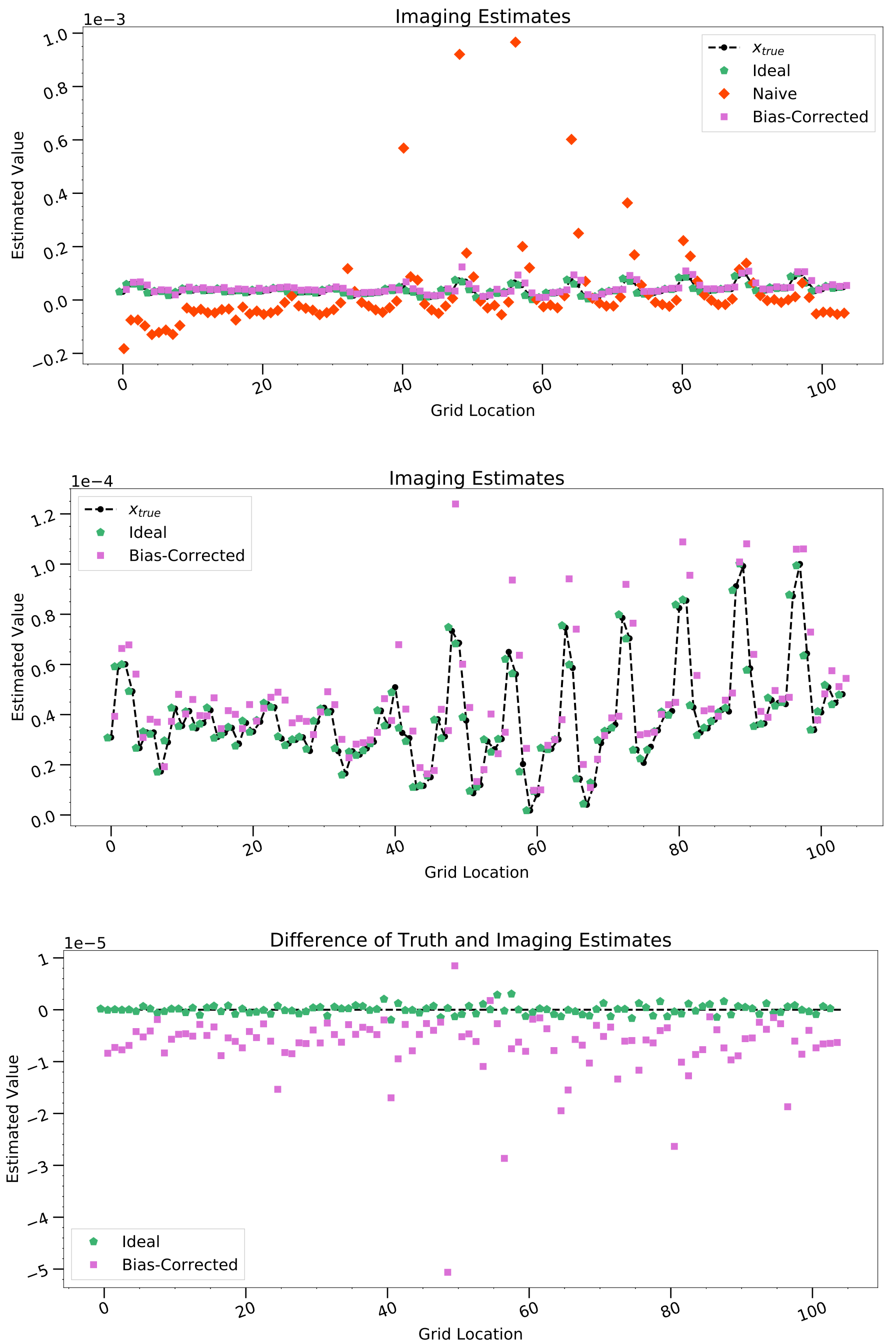


Figure 5.9: *top:* The exoplanet brightness coefficients for each of the three estimators for the Phase B experiment, following five relinearization iterations. *middle:* The same as the top, except the naïve estimates are not displayed. The error bars provided by the ECM are too small to be seen. The partial χ^2 value for the ideal estimate is 0.76, and for the bias-corrected estimate it is 225. *bottom:* The difference of the estimated coefficients and the true values for both the ideal and bias-corrected estimators

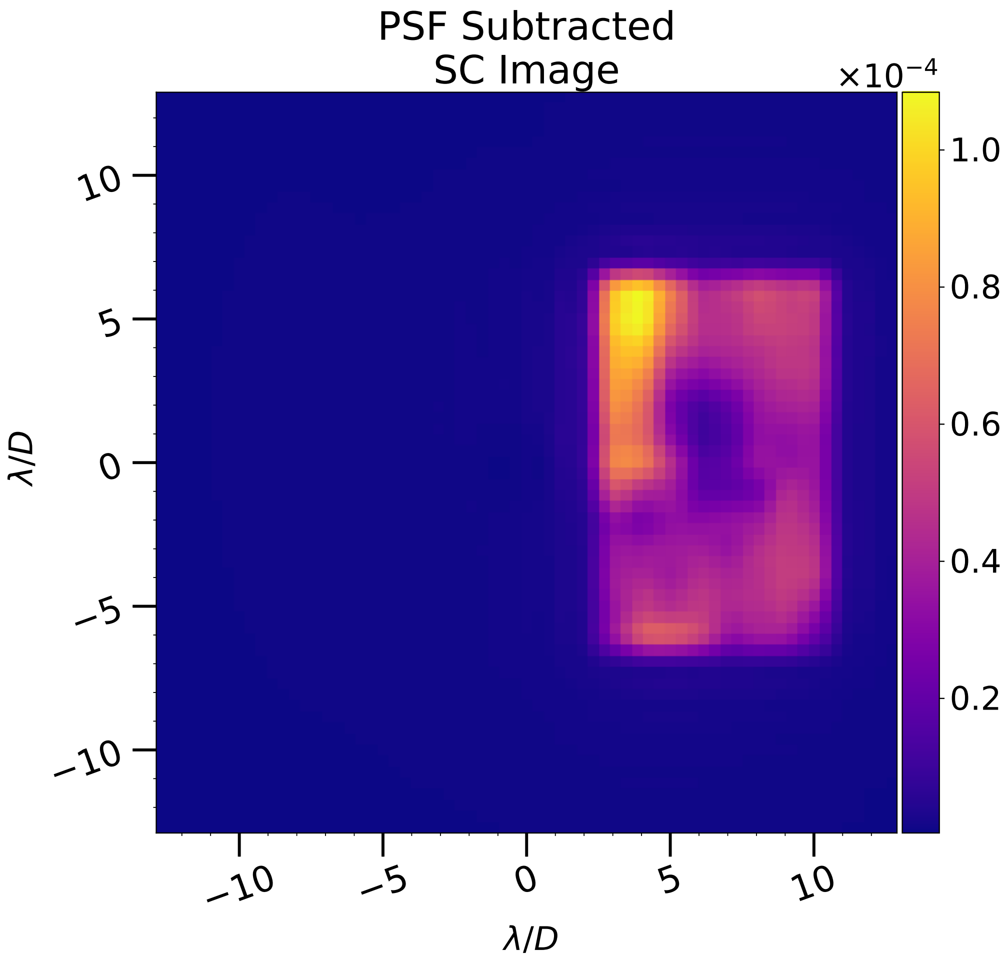


Figure 5.10: Focal plane image in contrast units after subtracting a perfect PSF from the science image. The perfect PSF is created by using the same sequence of wavefronts as the science image.

Comparing to perfect PSF subtraction

In this section, we compare the bias-corrected estimate with a more common method, PSF subtraction. To proceed with this discussion, we first define some terms that will be used. PSF stands for ”point-spread function,” which for our purposes amounts to the time-average image in the science camera that would be observed without any planets. It carries the effects of diffraction, NCPA and the AO residual wavefronts. The *science image* is the time-average of the science camera telemetry, and it is shown in the right-hand frame of Fig. 5.6. We further define the *perfect PSF*:

- it is created using the same sequence of wavefronts as the science image
- it has the exact same NCPA as the science image
- it has the same noise realizations as the science image
- it excludes all light coming from the circumstellar material (“the exoplanet scene).

Thus, the perfect PSF is the ultimate image to subtract from the science image, as it perfectly removes starlight and noise. In particular, subtracting the perfect PSF does not suffer from self-subtraction artifacts associated with differential imaging (see Section 2.4.1), and which arise from having to measure a PSF to subtract on-sky. Of course, the perfect PSF cannot be known outside of simulation studies, but it is a useful benchmark. Fig. 5.10 shows the resulting difference image of subtracting the perfect PSF from the science image following the convention described in Section 5.1.2.

Despite the perfect PSF subtraction, the image in Fig. 5.10 is missing much of the detail present in true image seen in left frame of Fig. 5.8. This is due to the fact that the image is blurred by the off-axis PSF. While one could attempt deconvolution to mitigate the blurring, instead, we used our knowledge from the simulation that the exoplanet image consists of point sources located on a grid. This allowed us

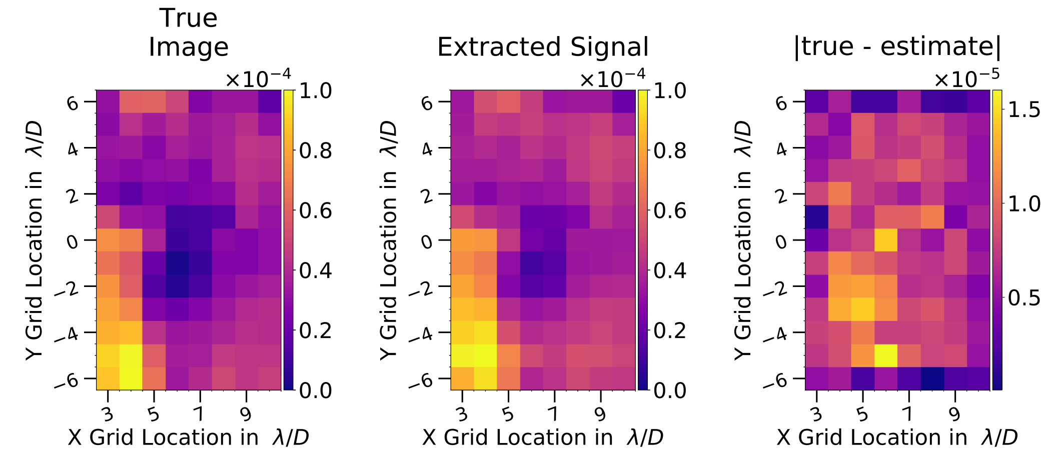


Figure 5.11: Extracting the exoplanet image signal from the PSF subtracted focal plane. Left: The true object source grid. Middle: Extracted signal from doing perfect PSF subtraction. Right: The absolute value of the error in the extracted signal. All units are in contrast.

to sidestep deconvolution simply by extracting the values of the image in Fig. 5.10 at the grid locations. The result of that extraction is seen in the middle frame of Fig. 5.11, with the true source grid pictured for convenience in the left frame, and the corresponding error in the right frame. The subsequent measurement of the RMS error for this extraction is 7.63×10^{-6}. Referring to Table (5.3), this result can be directly compared to the RMS error for the various estimate techniques presented, showing that the perfect PSF subtraction result quite is comparable to that of the bias-corrected estimate, with the RMS error of the bias corrected estimate being about 25% larger. The RMS error of the ideal estimate turned out to be 8.6× better than the perfect PSF subtraction.

While we were able to use knowledge of point-source locations to sidestep the deconvolution issue after performing the PSF subtraction, this would not be possible on-sky, and it is important to understand the our estimation methods automatically include the blurring effects that off-axis PSF has on the planetary image.

Estimate Type	**NCPA Estimate RMS error (radian)**	**Exoplanet Image Estimate RMS error (brightness)**
Naïve	1.13×10^{-2}	1.63×10^{-4}
Bias-Corrected	4.39×10^{-3}	9.45×10^{-6}
Ideal	1.82×10^{-4}	8.83×10^{-7}
Perfect PSF subtraction	N/A	7.63×10^{-6}

Table 5.3: Error metrics for a simulated 4 minute observation. Summary of the resulting root mean square (RMS) error using the naïve, bias-corrected, and ideal estimates to jointly solve for the NCPA and exoplanet image. The results for using ideal PSF subtraction for estimating the exoplanet image are also included

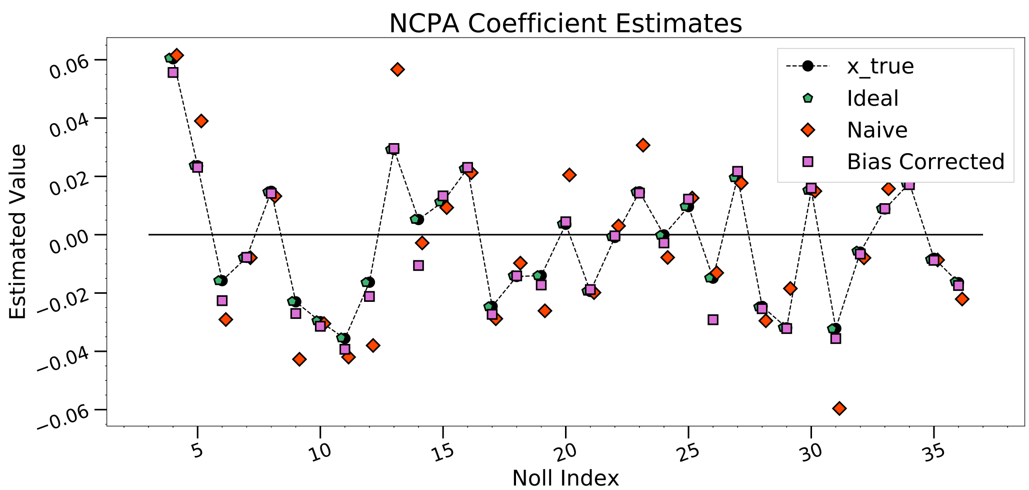


Figure 5.12: The joint estimates of the NCPA coefficients found while estimating the image in the 4 minute Phase B experiment. The error bars provided by the ECM are too small to see. The partial χ^2 value for the ideal estimate is 1.25, and for the bias-corrected estimate it is 203.

5.3.3 Jointly monitoring the NCPA

The regression equations for the Phase B experiment were set up to jointly estimate the exoplanet image and the NCPA. The NCPA had an RMS of 0.0495 radian, which is the uncorrected portion from Phase A. The estimated coefficients from each estimator are plotted in Fig. 5.12. Starting our analysis on the naïve estimate, the bias due to the WME is just as noticeable as it was in the exoplanet brightness coefficients, leading to an RMS error in the aberration coefficients of 0.011 radian, which places the error on the same order of magnitude as the NCPA itself. With the bias-corrected estimate, we see that where some small residual bias is left even after the correction. This is also largely due to the fact that the Monte Carlo wavefronts did not reflect the non-normal character of the AO residual wavefronts. In fact, we see that the bias-corrected estimates are actually worse than the naïve estimates for several of the coefficients, although the bias-corrected estimates generally are better, as evidenced by an RMS error that is smaller by a factor of about 2.5. Looking now to the ideal estimate, which tells us the level of performance that can be reached in an unbiased estimate, the RMS error is smaller than the naïve one by a factor of about 62.

5.3.4 The joint error covariance matrix

We now turn our attention to the interaction between the exoplanet image estimation and the NCPA estimation. This can be examined by looking at the error covariance matrix (ECM), $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$, calculated as part of the regression technique. The ideal estimator is simply classical linear regression and its ECM takes a particularly simple form when the weighting matrix of the measurements, is equal to $\boldsymbol{C}_{\boldsymbol{y}}^{-1}$, which is the inverse of error covariance matrix of the noise in the measurements. Under this assumption (without this assumption the more complicated formula given in Eq. (4.42) is needed), the ECM, is given by:

$$\boldsymbol{C}_{\hat{\boldsymbol{x}}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{C}_{\boldsymbol{y}}^{-1}\boldsymbol{A}\right)^{-1} , \tag{5.4}$$

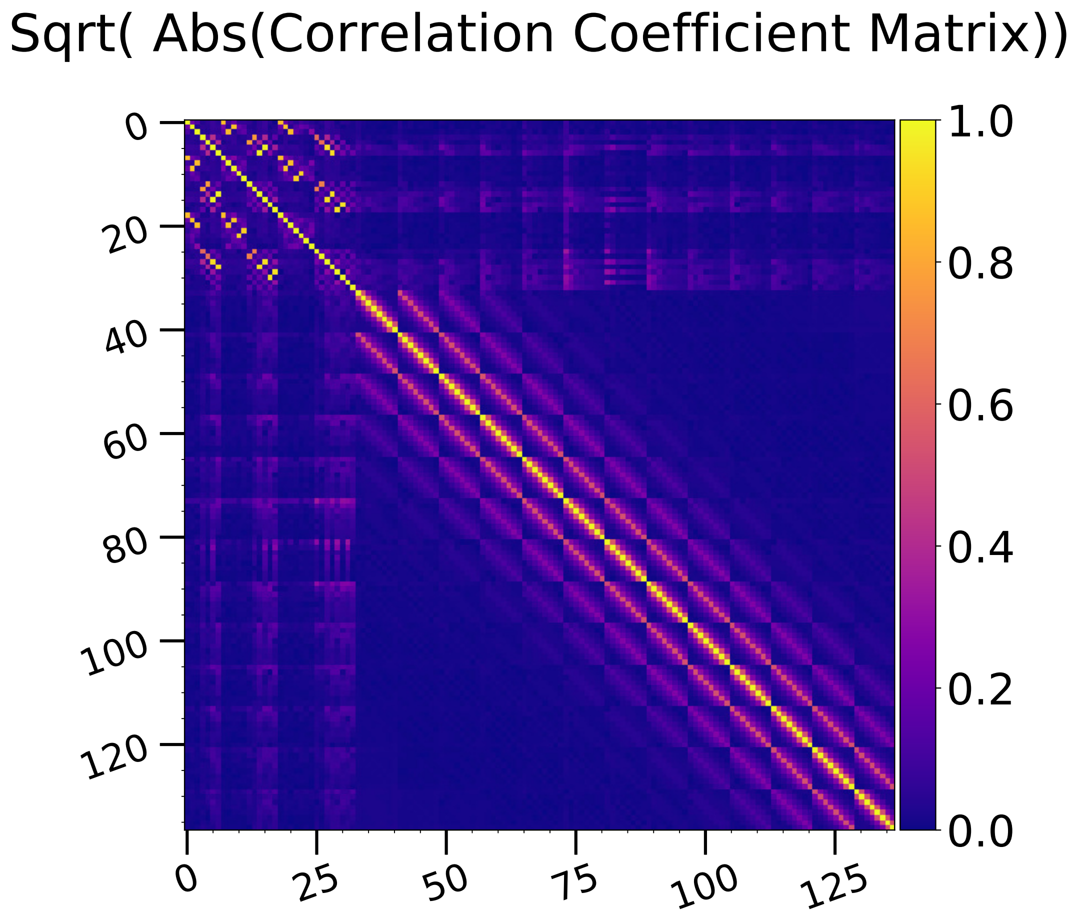


Figure 5.13: The square root of the unsigned correlation coefficient matrix of the ideal regression method for the Phase B experiment. The aberration coefficient coupling reaches a maximum of 0.96. The coupling for neighboring points on the exoplanet grid is a maximum of about 0.3. The coupling between the aberration and exoplanet coefficients and reaches a maximum of 0.1.

where $^{\mathrm{T}}$ indicates matrix transposition and $\boldsymbol{A}$ is the matrix obtained when one vertically stacks all T instances of the matrices $\{\boldsymbol{A}(\boldsymbol{w}_t)\}$ in (4.32). The ECM in (5.4) has two obvious asymptotic (i.e., $T \to \infty$) behaviors: 1) it is proportional to $1/T$ and 2) it is proportional to $\boldsymbol{C}_{\boldsymbol{y}}$, which is the covariance of the noise in the science camera measurements. For small values of T, as T increases $\boldsymbol{C}_{\boldsymbol{x}}$ decreases more quickly than the asymptotic rate because the condition of $\boldsymbol{A}$ improves as more rows of $\boldsymbol{A}(\boldsymbol{w}_t)$ are added. Indeed, when T is small enough, $\boldsymbol{A}$ is likely to be singular, but the random nature of each wavefront, $\boldsymbol{w}_t$, is rather helpful in this respect. For the bias-corrected estimate, the ECM takes a form that is similar to the classical one, but inflation matrices (calculated with the Monte Carlo wavefronts) must be applied, as shown in Eq. (4.69).

One may wonder what happens if the NCPA contains a component that is poorly observed or unobserved by the coronagraph. Firstly, the random modulation of the NCPA by AO residual tends to make it difficult to find modes that are completely unobserved. If a mode can be represented by the parameterization of the NCPA in the regression model (i.e., $\theta_p(\boldsymbol{a})$ in (4.11)), then the error in its estimate can be found from the ECM. On the other hand, if the NCPA parameterization cannot fully account for a given mode, one would expect aliasing in the sense that the mode leaks into the parameterized mode, creating an unwanted bias in the estimates of those modes.

The unsigned correlation coefficient matrix, shown in Fig. 5.13, is defined as $\boldsymbol{C}'_{\hat{\boldsymbol{x}}} \equiv ||\boldsymbol{B}\,\boldsymbol{C}_{\hat{\boldsymbol{x}}}\,\boldsymbol{B}||$, where $|| \bullet |||$ indicates the absolute value of all matrix elements and $\boldsymbol{B}$ is a diagonal matrix chosen so that every element on the diagonal of $\boldsymbol{C}'_{\hat{\boldsymbol{x}}}$ is unity. Without the absolute value operation, $\boldsymbol{C}'_{\hat{\boldsymbol{x}}}$ would be matrix of correlation coefficients. As would be expected for a joint estimate of two types of quantities, we see a four block matrix, with the upper left being the NCPA-NCPA block, the lower right being the exoplanet-exoplanet block, and the other two being NCPA-Exoplanet blocks which are related by a transpose operator since the matrix is real and symmetric.

Starting with the NCPA-NCPA block (upper left) of the correlation coefficient matrix of the ideal estimate shown in Fig. 5.13, the off diagonal behavior shows of coupling in the estimates of different modes, which is to be expected from the Zernike polynomials used to represent the NCPA in this study. There are twenty off-diagonal elements in the upper triangle of this block with values greater than about 0.8, meaning some the NCPA coefficient estimate errors show strong coupling. The exoplanet-exoplanet block (lower right) shows significant power on the diagonal, with the power generally decreasing with distance (the tiling is due to the two-dimensional grid being flattened to a one-dimensional vector). The maximum off-diagonal value occurs here for neighboring points on the exoplanet grid, and is about 0.3. The off-diagonal power is largely due to the blurring by the off-axis PSF, which would be worse if the sky-angles in the grid were separated by less than λ/D, as they are in these simulations. It is of course possible to have a more densely sampled grid for the purposes of maximizing the spatial resolution, but we have not yet explored this option in detail.

Next, we come to the NCPA-exoplanet blocks. These blocks describe the coupling of the estimate errors in the NCPA and exoplanet coefficients, and show that there is significant coupling. The maximum value of the NCPA-exoplanet correlation coefficients is about 0.1, and we can see that many of the exoplanet coefficient errors have correlations to one or more NCPA coefficients of roughly similar value. This tells us that using the joint regression is critical to achieving high contrast in the estimates.

For an unbiased estimator, the validity of the ECM can then be examined by evaluating the χ^2 test, given by

$$\chi^2 = \frac{1}{N_a + N_p} (\hat{\boldsymbol{x}} - \boldsymbol{x})^{\mathrm{T}} \boldsymbol{C}_{\hat{\boldsymbol{x}}}^{-1} (\hat{\boldsymbol{x}} - \boldsymbol{x}) \,, \tag{5.5}$$

where $\boldsymbol{x}$ is the true value of the coefficient vector (so, this metric is only possible to evaluate in simulations), $N_a + N_p$ is the length of $\boldsymbol{x}$, $\hat{\boldsymbol{x}}$ is a regression estimate of $\boldsymbol{x}$, and $\boldsymbol{C}_{\hat{\boldsymbol{x}}}$ is the ECM from the regression model. With χ^2 so defined, we expect its value to be close to unity if the ECM does indeed correctly characterize the estimate errors. The partial and total χ^2 values for the bias-corrected and ideal models can be found in Table (5.4). The partial χ^2 values come from evaluating the components of $\hat{\boldsymbol{x}}$ (aberration coefficients or exoplanet brightness coefficients) with the corresponding diagonal block of the ECM. The total χ^2 is evaluated using the full $\hat{\boldsymbol{x}}$ vector, meaning it also includes the entire ECM (including the off-diagonal blocks). As we expect, the total χ^2 value for the ideal estimate is very close to 1. The χ^2 value of about 200 obtained for the bias-corrected estimate tells us that its ECM is too small. We believe that much of this discrepancy is the result of the non-normality of the AO residual wavefronts that was not captured in the Monte Carlo wavefronts used by the bias-corrected estimator.

The square root of the main diagonal of the ECM can be treated as the error bars (which are too small to be seen in Figs. 5.9 and 5.12). While the error bars of the ideal estimate are reasonable, as the χ^2 test tells us, the error bars of the bias-corrected estimate are barely any larger and do not come close to describing the errors. Again, we attribute this unfortunate circumstance to the fact that our Monte Carlo wavefronts did not account for the non-normality of the true wavefronts. It is not unfair to say when the Monte Carlo wavefronts are missing some key elements of realism that the resulting ECM does not describe the errors in the estimates, even when the estimates themselves may be quite accurate and useful. Nevertheless, the ECM still tells how the errors in the estimates of various quantities are likely to be coupled.

Estimate Type	χ^2 **NCPA**	χ^2 **Imaging**	**Total** χ^2
Bias-Corrected	203	225	232
Ideal	1.25	0.76	0.97

Table 5.4: χ^2 test values for the Phase B experiment (see (5.5)). he partial χ^2 for the NCPA and Imaging are computed by only examining the corresponding components of the $\hat{\boldsymbol{x}}$ vector and block from the ECM. The total χ^2 is computed using the entire $\hat{\boldsymbol{x}}$ vector and ECM.

5.4 Frazin's algorithm conclusions

This chapter presents realistic simulations of regressions based on simultaneous millisecond telemetry from a WFS and science camera behind a stellar coronagraph. The simulations include self-consistent treatment of an AO system with a pyramid wavefront sensor, a Lyot coronagraph, and photon counting and readout noise in the detectors. The objective of the regressions is simultaneous estimation of the non-common path aberrations (NCPA) and the exoplanet image. We presented two realizable regression models as well as a non-realizable one, which we used as a benchmark. The two realizable regression models are called the *naïve estimator* and the *bias-corrected estimator*, and the non-realizable one is called the *ideal estimator*. The fact that just a few minutes of simulated sky time allowed us to make estimates of the NCPA with RMS errors that were much smaller than the NCPA themselves suggests that our methods could be implemented inside a control loop that compensates for the NCPA in real time. This is under investigation, and preliminary results are reported in Chapter 6. Furthermore, analysis of the error covariance matrix of the estimators demonstrate that the errors in estimating the NCPA and the exoplanet intensity are correlated, suggesting that exoplanet imaging and determination of the NCPA must be done self-consistently to achieve high contrast.

We illustrated the utility of the naïve estimator by estimating NCPA of 0.52 radian ($\sim \lambda/12$) RMS with an accuracy of 0.06 radian ($\sim \lambda/100$) RMS error using only 1 minute of simulated observation time of an 8th magnitude star. Then, in a followup observation sequence consisting of 4 minutes of simulated observation time, we assumed an NCPA of 0.05 radian RMS. We found that the error of the bias-corrected estimate of the NCPA was 0.004 radian ($\sim \lambda/1600$) RMS, while jointly estimating planetary image on a 13×8 object grid (focal plane distances from center ranging from 3-10 λ/D). The bias-corrected estimate of the exoplanet scene was nearly identical to the image that would be obtained by PSF subtraction if the PSF were *exactly* known. The bias-corrected estimate obtained a $5\,\sigma$ contrast at $3\,\lambda/D$

of $\sim 1.7 \times 10^{-4}$, while at $10\,\lambda/D$ it was $\sim 2.1 \times 10^{-5}$. The contrast achieved by the bias-corrected estimator was limited by our inability to draw Monte Carlo samples from non-normal probability density governing the statistics of the AO residual wavefronts. Additional simulation results for experiments in which the AO residuals were drawn directly from a multivariate normal distribution instead of a simulated AO system were also obtained. This simplification allowed us to ensure the Monte Carlo samples were drawn from the same distribution as the AO residuals exactly. In these experiments, the $5\,\sigma$ contrast achieved by the bias-corrected estimator from ~ 17 m of simulated sky time ($T = 10^6$) at $3\,\lambda/D$ was 5.5×10^{-6} and at $10\,\lambda/D$ it was 2.9×10^{-6}; these values were almost the same as those from the ideal estimate. The contrasts reported here should not be interpreted as a fundamental limits to what can be obtained by the regression methods employed, rather they should be seen as illustrative exampled within the context of our simulations.

For comparison, recent efforts using SCExAO with the VAMPIRES module in a bandpass containing the Hα line achieved a $5\,\sigma$ contrast of $\sim 10^{-3}$ at a distance of $\sim 17\,\lambda/D$ (the Strehl ratio was about 0.45) [80]. Achieving this contrast required both angular differential imaging (ADI) and spectral differential imaging (SDI) to be applied in post-processing. As far as we know, the highest contrast ever achieved on-sky was reported in 2015 by Vigan et al. on the SPHERE AO system [81]. In this case, SPHERE was looking for a second companion to Sirius A (magnitude -1.46), which allowed the AO system to operate at a Strehl in excess of 0.9 (at $\lambda = 1.6\,\mu$m). The 59 minutes of observations were collected over a 2.5 hour period with the coronagraph feeding an integral field spectrograph (IFS) covering the range $\lambda = 0.95$ to $2.3\,\mu$m. The field rotation over the 2.5 hour period combined with imaging spectrograph data allowed the authors to apply both ADI and SDI in a post-processing step. The $5\,\sigma$ contrast finally achieved was $\sim 5 \times 10^{-5}$ at a distance of 0.2 arcsec, which corresponds to 3.5 and $8.5\,\lambda/D$ at $\lambda = 2.3$ and $0.95\,\mu$m, respectively. At a distance of 0.4 arcsec, they reported a contrast of $\sim 3.5 \times 10^{-6}$. Thus, SPHERE's reported contrasts for this observation are quite comparable to contrasts achieved in our simulation. While simulations cannot carry the same weight as results from physical experiments, it is instructive to point out that our simulations in some sense address an imaging problem that is much more difficult than this particular SPHERE experiment:

- The SPHERE experiment collected spectral data over more than factor of 2 in wavelength, which allowed SDI, while our data were monochromatic. (We note that extending our method to multi-wavelength data is relatively straightforward.)
- The field rotation during the SPHERE observation allowed ADI, while we had no field rotation. (Including field rotation into our regression equations is easy.)
- SPHERE's Strehl ratio was over 0.9, while ours was about 0.75.
- Sirius A is brighter than our magnitude 8 source by a factor of over 7000, and SPHERE collected photons for about an hour, whereas we simulated only 4 minutes of sky time.

The fact that we obtained such good results without the benefit of multi-wavelength data or field rotation with only a few minutes of simulated sky time speaks to the richness of the millisecond data sets we seek to exploit. Perhaps it even suggests that our development of the regression models is on a useful track.

Both the naïve and bias-corrected estimators require models of the WFS and coronagraph optical trains, but the bias-corrected estimator also needs knowledge of the spatial statistics of the AO residual wavefronts. Model errors and inaccurate characterization of the AO residual statistics will cause unwanted biases in the estimates. A closed-form expression for the size of the bias resulting from violating a given assumption is generally not easy to obtain. Finding the biases is most easily done for specific cases via simulation studies. In these simulations, we have seen an example of one of the assumptions not holding true: the AO residual wavefronts did not obey multivariate normal statistics, but bias-corrected estimator employed the assumption of a multivariate normal when creating the Monte Carlo wavefronts that it utilizes as part of its machinery (explained in Chapter 4). It should be remembered any ground-based exoplanet imaging method must contend with NCPA and the statistical properties of the AO residual wavefronts, so these issues are not unique to this approach.

In order to implement this type of regression method on sky, there are number of technical challenges that must be overcome:

- The regression equations require accurate numerical models of the WFS and coronagraph optical trains. These models may contain free parameters representing such things as alignment drifts that are determined from the data, much as the NCPA coefficients are in this study. Removing high spatial frequencies from the beam via applying a stop in a focal plane at one or more locations in the apparatus may prove critical for limiting the degrees of freedom that must be taken into account in the regression models.

- While an initial application may work with a bandpass narrow enough to ignore chromatic effects in the AO residual, moving to larger bandpasses may require taking chromatic considerations into account.

- This study ignored amplitude effects (scintillation). Ideally, the wavefront sensing scheme would measure amplitudes as well as phase.

- The Monte Carlo calculations needed by the bias-corrected estimator require the statistics of the AO residual wavefronts.

Chapter 6 will examine some first steps taken to demonstrate the naïve estimator in the lab, using the MagAO-X science instrument and a basic numerical model of it.

CHAPTER 6

Future work on Frazin's algorithm

6.1 Estimating and compensating QSA in real-time: realizing the goal from 2018

In order to come full circle to 2018, and to prepare for setting up a laboratory demonstration of the naïve estimator, we made some adjustments to the framework of the simulation to allow for the use of estimates made through Frazin's algorithm as real-time commands to a non-common path correcting (NCPC) DM. This process was envisioned to take all the lessons learned from the simulations done and reported in Chapter 5, where only a static aberration had been estimated, and test if the the estimators could provide quality predictions of what an NCPA is over the integration time to be able to compensate for it, in a slow 0.1Hz integrator control loop. Because this is to also be used as a stepping stone towards demonstrating the method in the lab, several new steps needed to be taken, including attempting to make a high quality computational model of MagAO-X (and some steps toward both understanding what that means, and how to make it easier), choosing a realistic NCPA to model, and adjusting simulation parameters to better fit in lab values rather than on-sky ones.

6.1.1 Modeling MagAO-X for simulations

To begin, we choose to make a numerical model of MagAO-X to use both in the regression equations and the propagation simulations. Although not realistic for real-world application of Frazin's Algorithm, it allows us to test the ability of the estimators to provide for DM commands to control the NCPA as it evolves in time with one fewer variable to worry about. The consequences of approaching this more realistically will be touched on in Section 6.2. To facilitate the best effort without requiring extensive background work rehashing the simulation code, the extensive modeling work of Jennifer Lumbres [37] (in prep) building an end-to-end Fresnel model including realistic surface error profiles using Poppy [63] was adopted. This work, alongside the work of Kyle Van Gorkom characterizing the MagAO-X hardware itself [27] and flattening the system, combine to validate the accuracy of the Poppy model, with both the model and hardware achieving a 0.94 Strehl ratio at Hα, with consistent values for WFE between the two. Although we will not delve into the specifics of Lumbres' work and code, Appendix C includes the script used to interface with it, computing a propagation matrix that can easily be slotted in to the simulation framework built for the work done in Chapter 5. As a result of having to interface with the POPPY code, the number of pixels in the entrance pupil is amended to 1986, slightly more than the previous simulations. An example of the PSF averaged over 4 minutes of AO residual wavefronts for the constructed model can be seen in Figure 6.1. This image represents a 67×67 pixel science camera region, and displays the expected diffractive behavior for the MagAO-X pupil, which has a central obscuration and four spider arms at multiples of 45 degrees. It should also be noted that because we are assuming this simulation to be more held to lab standards than on-sky ones, the visual magnitude of the source for the purpose of evaluating photon noise calculations is chosen to be 0, which is much brighter than the source used in previous simulations. This has the added effect of helping the AO system to perform better, as the WFS signal is more precise, so the Strehl ratio observed in the science camera is higher than previous simulations. We also choose for the following simulations to drop the joint estimate on the planetary image. From the standpoint of these simulations, the purpose is to try to demonstrate controlling quasi-static NCPA in real time, so extending the regression equations to include the planetary image estimation is not necessary, especially when we are simulating a lab setting where we know there is no exoplanet image to self-consistently estimate. This will make the calculations faster numerically (fewer elements to estimate), and eliminate any concern of error cross-talk from the planetary image estimator that was discussed in Section 5.3.4. Finally,

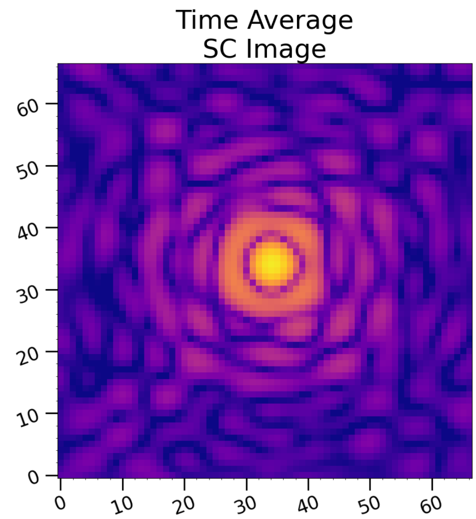


Figure 6.1: A simulated time average PSF using the MagAO-X model over 4 minutes in closed loop.

the only other change made to the simulation setup and parameter list specified in Chapter 5 is that instead of estimating coefficients fitting the NCPA to low-order Zernike polynomials, coefficients for a set of Bspline functions that represent the actuators that control the surface of an NCPC DM are chosen. Although this vastly increases the number of estimated parameters for the NCPA (from 32 modes to 711), the output will already be in the form of a DM command (as the spline knots estimated match the DM model actuators exactly), and better be able to fit to the higher spatial frequency content of the NCPA being modeled.

In order to have a convincing argument that Frazin's Algorithm can control quasi-static aberration (QSA) via closing the loop on its estimates, including a realistic QSA is a key component of these simulations. We decide to model a beam walk aberration with a 5 minute lifetime. Beam walk is the change in phase in the wavefront over time as the light hits different parts of the surface of the optics in the telescope and AO system over time, especially on the secondary mirror. As the telescope tracks a star, the beam of light "walks" along the surface of the secondary mirror, changing the content of optical surface quality as it does. The kind of NCPA this induces thus contains a range of spatial frequencies corresponding to a

$\frac{1}{f^\alpha}$ type power spectral density. By selecting the value for the spatial α as 2, and the temporal α as 4, we can tune a realistic beam walk-like aberration that behaves as a frozen-flow process. The simulations use a total of $300,000$ millisecond frames of an NCPA tuned to have roughly 5 minutes of stability, placed in the entrance pupil of the science instrument. An example frame from this set is shown in Figure 6.2.

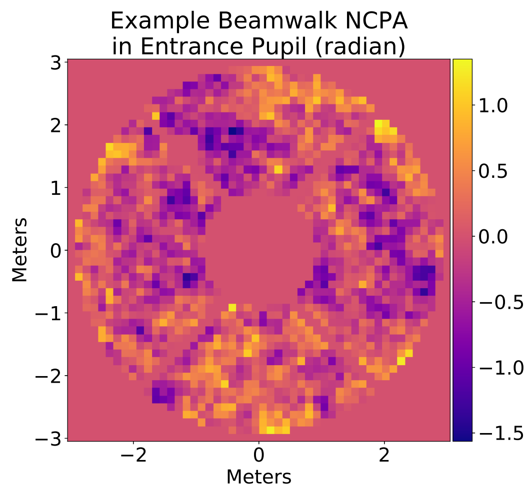


Figure 6.2: An example frame of a beam walk like NCPA injected into the entrance pupil of the MagAO-X model. It is produced using a spatial α of 2 and a temporal α of 4 in the $\frac{1}{f^\alpha}$ PSD.

6.1.2 Results of the simulation

The paradigm of the experiment to control a realistic beam walk-like QSA using the bias-corrected estimator is to simply demonstrate that it is possible. With that in mind, very little effort was placed in to optimizing the integrator loop gain, nor how often updates from Frazin's Algorithm could be demanded and still work. This analysis is left for future work. For the simulations performed, the update rate of the NCPC DM using estimated Bspline coefficients was once every 10 seconds, corresponding to $10,000$ ms time steps. This means that the regression equations are updated once a millisecond for ten seconds, and then solved for the estimates of the NCPA over that time span. The QSA correction loop gain is set as 0.125, chosen experimentally repeating the simulation at various gain values until the largest one that remained stable over the 5 minute simulated time span was found. To determine if the QSA is adequately being controlled by Frazin's Algorithm in real time, the residual RMS of the NCPA pupil plane just after the NCPC DM will be measured each time step of the simulation. Although this is not possible to do in a real experiment, where the correction would have to be judged via quality of the science camera image, it provides a simple and clear means of judgment on the validity of the method. In order to provide more clarity on this judgment, the RMS of the NCPA at the given time step should there be no attempt to compensate it is also computed and reported, alongside the RMS of the residual NCPA that exceeds the maximum spatial frequency of the NCPC DM should it have a perfect shape. Figure 6.3 plots the tracking of these three different RMS phase values. The

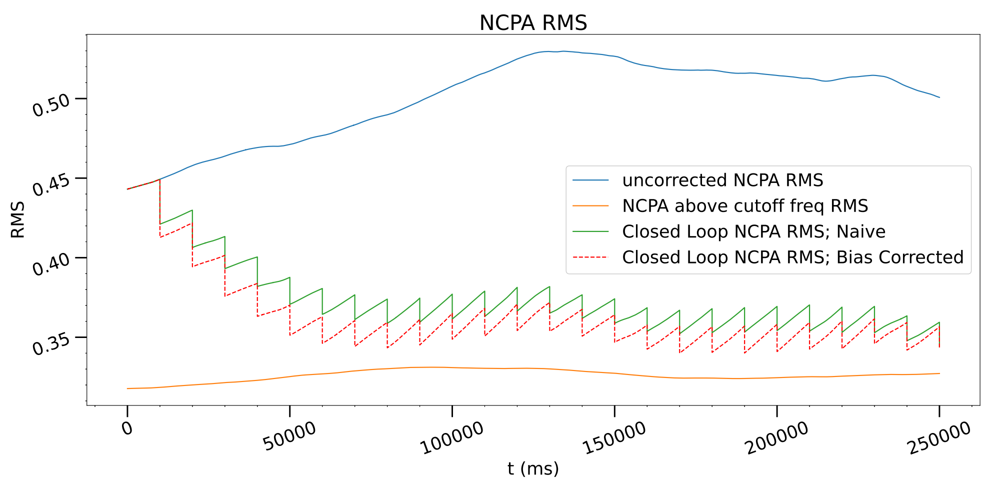


Figure 6.3: Computing the RMS of the NCPA at each time step of a simulation running Frazin's algorithm in real time to compensate beam walk every 10s. The blue line represents the RMS of the beam walk induced aberration if no compensation were done. The yellow line represents the RMS error that is left after a perfect DM compensation (in other words, the residual aberration at spatial frequencies untouchable by the DM). The green line is the RMS of the NCPA after the DM is updated to compensate it using the naïve estimator. The dashed red line is the RMS of the NCPA after the DM is updated to compensate it using the bias–corrected estimator.

blue line, representing the uncorrected NCPA RMS, combined with the yellow line, representing the residual NCPA RMS left that the DM could never correct, provide the control to compare the green and red lines, representing the RMS of the NCPA post correction by DM commands estimated by the naïve and bias corrected estimators respectively. If Frazin's Algorithm is successfully controlling the QSA, the green and red lines will start coincident with the blue line for 10 seconds, and the drop lower than it, converging towards the yellow line (but likely never reaching it). We can see this is exactly what happens. In the 10s that the regression equations are being evaluated in order to be able to compute an estimate, the RMS phase increases because the shape of the NCPC DM is becoming outdated. Once the 10 second interval is complete, an estimate is made via the bias corrected estimator. This estimate is then multiplied by the loop gain, and sent as an update to the shape of the NCPC DM, and shows in the plot as a drop in the NCPA RMS. Once the control loop has closed, at around 1 minute, the QSA is held at roughly 1.5$\times$ lower than if nothing was done to compensate it. It is notable that the control loop using the naïve estimator achieves close to the same performance in compensating the QSA as the control loop using the bias–corrected estimator does. The consequence of this result is that we know, at least for the size of aberration and the WME under test here, that both estimators appear to be monotonic, meaning that it is estimating the sign of the residual QSA at each step correctly. This allows the control system to move in the correct direction at each step, driving the residual NCPA RMS down and stabilized. It is possible that for smaller wavefront errors being estimated that the estimator will not get the sign correct every step, meaning that although it would be correcting large errors, the controller would oscillate about the small ones.

Examining the limits of Frazin's algorithm being implemented as a controller is an exercise for future study. This preliminary study does demonstrates that the initial hope from 2018, the real-time control of QSA using Frazin's Algorithm, is now realizable. But further examination of several factors is needed to fully flush out this use of Frazin's Algorithm:

1. A parameter sweep to better understand the limits and capabilities of the

closed-loop implementation. Parameters that are especially pertinent to examine are:

- The minimum number of exposures per estimate (the rate at which the control loop is run) for which reasonable correction can be obtained.
- The optimal gain for the control loop.
- The best pixel sampling of the pupil (to allow for potentially more modes to be estimated to fit to the surface of a more actuator dense DM than the one simulated above, pushing the cutoff spatial frequency that can be compensated closer to that of the aberration content due to beam walk) that still allows for the algorithm math to keep up with real-time data.
- The optimal number of pixels kept in the science camera images to capture enough behavior at higher spatial frequencies potentially now represented in the pupil sampling that still allows for the algorithm math to keep up with real-time data.

2. Vary the WFS performance in the simulator to vary the WME. This will push the naïve estimator to determine at what point, if any, that its use in closed-loop breaks down, and the more complicated bias-corrected estimator would become necessary to implement.

3. With the parameter sweep to determine a potentially minimum number of time steps included in the algorithm, test against QSA of different lifetimes to understand the temporal limits of the real-time controller.

4. Determine at what point, if any (which surely there is), an off-axis source such as an exoplanet will corrupt the estimates too much to avoid implementing the joint estimator discussed in Chapters 4 and 5 to treat it self-consistently.

6.2 Examining error in the science instrument numerical model

One question that is begged to be asked when discussing Frazin's Algorithm is: how good do the computational models of the instrument need to be? This is a complex question that will require much more study going forward in the future. To provide some insight however, we have done some simple controlled numerical experiments to start to probe this space of algorithm performance. As mentioned, the simulations above all assumed a perfect computational model for the regression equations, as in the simulations the same matrix was used in the regression as was used in the calculation of the propagation of the field impinging on the science camera. The likelihood of achieving such a match in real life at the moment may be quite small. However, given the state of optical design software like Zemax and LightTrans, as well as modern interferometers capable of making extremely precise measurements of optical surface qualities, this is a problem that may improve as time marches forward. Along these lines, we decide to construct two new propagation matrices for MagAO-X, each differing from the model used in the previous section closing the loop on the QSA, in a known way.

6.2.1 The flawed models

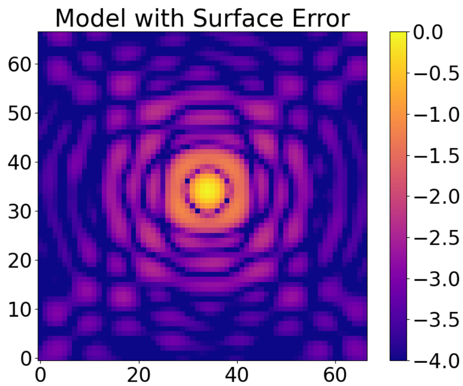
For the first newly made MagAO-X propagation matrix that will be used in the regression equations, but not in the science camera intensity calculations, we return to Jennifer Lumbres' code, and turn off the surface quality errors on each optical surface in the model. This is an interesting model to look at because it essentially represents the as designed performance of the system, and could be a model that is achievable in real life through the use of software like Zemax and LightTrans. Although it seems odd to consider this a "flawed" model given what it represents, the flaws are in that it does not match the model matrix that will be used to compute the science camera intensity. The true PSFs (no turbulence averaging) in log10 scale for both the model including surface error (which was used in the previous sub-section's simulations for all the calculations) and this new model without surface errors, along with the difference between them, can be seen in Figure 6.4. We see that the difference is small, largely contained to the core region and about $1-2$ decades in magnitude.

The second flawed model that will be tested is another one chosen to be something that one may think could be realizable in real life without the use of a lot of sophisticated interferometric surface measurements being required. This is to know the summary statistics of the surface errors for the optical elements used to build the system, and use Monte Carlo methods to generate statistically similar (in terms of spatial correlation and distribution) surfaces to what the true ones are. In a sense, it already seems that this model should perform poorly just from understanding that it includes optical surface errors that have the same statistics, but entirely different realizations than the true system, but it is interesting to look at nevertheless.

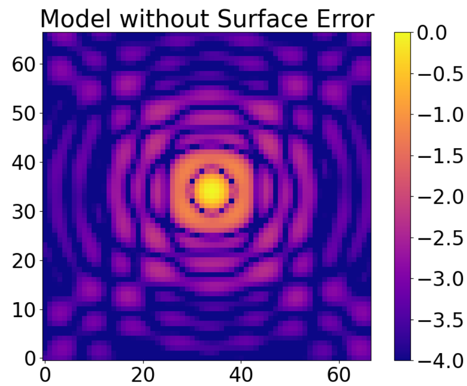
It is also important to note that although we will be able to draw conclusions about how well the computational model needs to be when implementing Frazin's algorithm, there is a flaw in discussing the experiment on only these terms. Because Frazin's algorithm is essentially fitting the difference in the measured science camera intensity to the modeled intensity given a measurement of the wavefront, what will appear as bias when comparing the final estimates to a known true value for the NCPA will actually be a change in the estimated coefficients trying to account for the missing/wrong surface errors in the model. In other words, the estimate of the known, injected NCPA is "corrupted" by the surface errors unknown to the model, as the estimator treats the WFE induced by the surface errors as another source of NCPA. This discussion will be picked up in the next sub-section, as it will inform the conclusions that we can reach.

6.2.2 Implementation and analysis

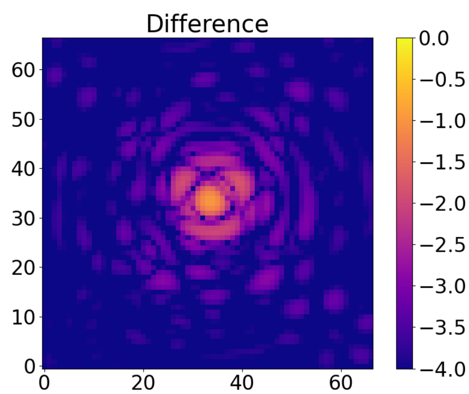
To simplify this implementation of Frazin's algorithm, we return to the NCPA setup of the simulations performed in Chapter 5, estimating 32 Zernike coefficients fitting to the same low order aberration shown in the left frame of Figure 5.5. The same simulations parameters that are used in the Phase A experiment in that chapter are



(a)



(b)



(c)

Figure 6.4: (a) The PSF going through a true model of MagAO-X, including surface errors on the optics. (b) The PSF going through an ideal model of MagAO-X, which is constructed of perfect optical elements. (c) The log10 scale of the absolute value of the difference between the two PSFs.

repeated for these tests, other than the changing models in the regression equation calculations. Three tests will be performed in which all the estimators (ideal, naïve, and bias corrected) are evaluated, again only estimating the NCPA (no joint estimate of the planetary image):

1. The propagation matrix used in both the regression equations and the calculations of the science camera intensity is the "true" model, corresponding to the one used in the closed-loop experiments of Section 6.1. This is referred to as the *Matching Model* case.

2. The propagation matrix used in the regression equations is swapped to be the model that was made without including surface errors. The matrix used to compute the science camera intensities is the "true" model. This is referred to as the *Mismatch Model* case.

3. The propagation matrix used in the regression equations is swapped to be the model made with new realizations of the surface errors. The matrix used to compute the science camera intensities is again the "true" model. This is referred to as the *Mismatching surface WFE Model* case.

Figures 6.5 – 6.7 show the estimated Zernike coefficients for each of these cases respectively, for each estimator type, as well as the true values of the coefficients. As expected, the Matching Model case performs very well, as it has no error in the regression equation model matrix, and is essentially a repeat of Phase A in the previous chapter. Applying the bias corrected estimate as a compensating factor to the static NCPA with an RMS of 0.5 radian results in a residual RMS of 0.0042 radian. For the Mismatch Model case, seen in Figure 6.6, now all the results for all the estimators display a noticeable bias, and are no longer near the true coefficients. This is because of the error in the model used in the regression equations showing up as a new kind of error in the independent variables. This can be thought of in quite an easy way. Because the regression equations are using a propagation matrix that differs from the matrix governing the measured science camera intensity, it will try

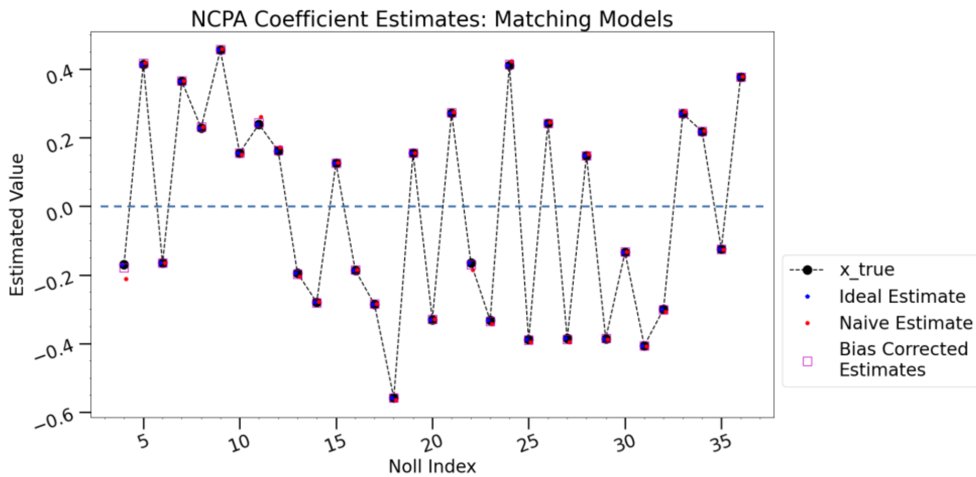


Figure 6.5: A plot of estimated coefficients for the same Zernike coefficients as estimated in Chapter 5, with the regression equation and science camera intensity calculations using the "true" propagation matrix model. This is essentially repeating the same simulations as in that Chapter, to be used as a means of comparison to the other two test cases.

to fit the additional differences that it finds in its calculation of the intensity from the WFS measurement compared to the measured intensity from the science camera it is given into the NCPA coefficients, causing them to be biased. But remembering the discussion above, this bias is actually different than the bias that is directly due to the WME that led to the derivation of the bias-corrected estimator. The error in the estimates we see here is largely due to the fact that the estimator is fitting the coefficients to the surface errors the model is missing in addition to the injected NCPA because of the fact that the surface errors manifest as speckles in the measured science camera intensity that are not explained by the WFS measurement nor the computational model. This can be seen in the effectiveness of the estimates at still compensating the injected NCPA. If the bias corrected estimate is used to compensate the initial NCPA that has a static RMS of 0.5 radian, the residual RMS is reduced to 0.078 radian. So although the error in the estimated coefficients is increased, reducing the compensation by a factor of $\approx 18\times$, the overall performance for assuming an "as-designed" model for the regression equations is still quite good. Finally, for the Mismatching surface WFE Model case, the estimated coefficients degrade even further because the difference in the regression equation model to the

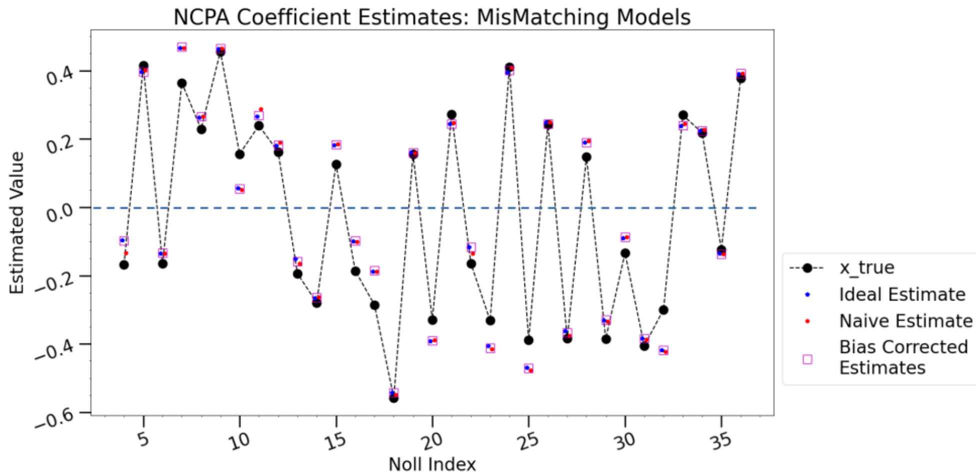


Figure 6.6: A plot of estimated coefficients for the same Zernike coefficients as estimated in Chapter 5, with the regression equation calculations using the surface error free model matrix, and science camera intensity calculations using the "true" model.

true optical system is even greater. This can be seen in Figure 6.7. If the bias corrected estimate is used to compensate the initial NCPA that has a static RMS of 0.5 radian, the residual RMS is only reduced to 0.143 radian. This is twice as bad as the Mismatch Model case, and aligns with what we thought would be the result of doing this, as the estimator will see a large discrepancy in the manifestation of the modeled surface error speckles and the true surface error speckles in the intensity.

From this initial peak into model error, which is summarized in Table 6.1, we get a few interesting insights. First, it is easy to see that the model error is the dominant error term in these experiments. This is evident because in the Matching Model case, the difference in the achieved NCPA compensation using the estimates varies widely between the estimator types. This is expected in this case because, as described in the previous chapter, the ideal estimate is completely unbiased, the naïve is biased by wavefront measurement error, and the bias-corrected has a small residual bias due to mismatched spatial statistics in the Monte Carlo wavefronts. All three of these levels of bias is quite different. In the cases of the Mismatch Model and the Mismatching surface WFE Model, all three estimates compensate the starting NCPA by nearly the same amount, degraded by the model error when

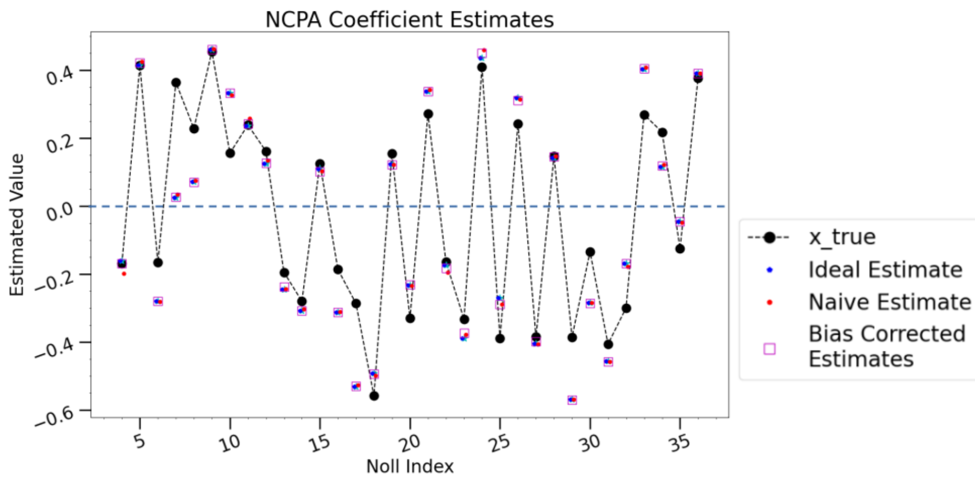
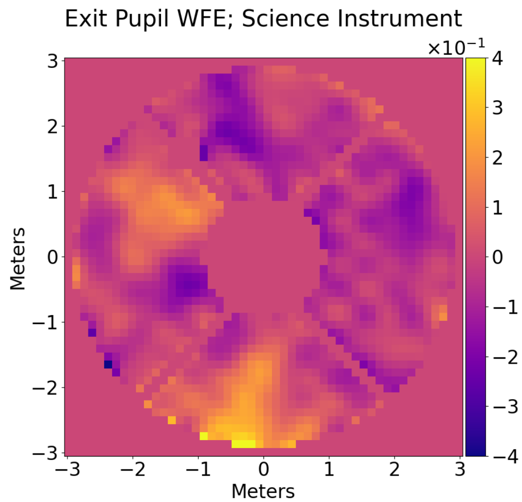


Figure 6.7: A plot of estimated coefficients for the same Zernike coefficients as estimated in Chapter 5, with the regression equation calculations using the model matrix with incorrect realizations of the surface errors, and science camera intensity calculations using the "true" model.

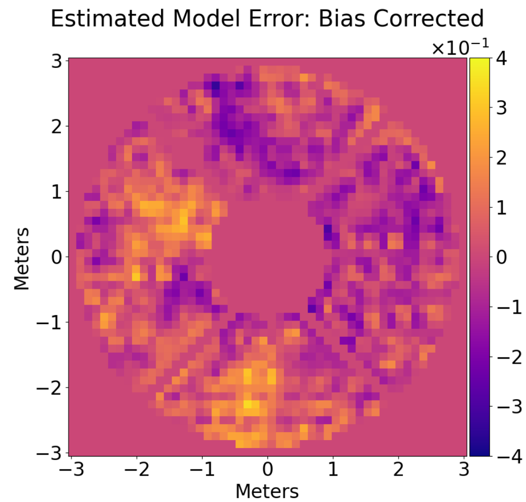
comparing only to the injected NCPA coefficients. With that being said though, the fact that the ideal and bias-corrected remain nearly identical, and slightly better than the naïve, means that the degradation in overall estimate accuracy is directly related to the error in the regression equation matrix model.

Next, it might be better to assume no surface errors on the optical elements in the computational model than to insert guesses of the wrong ones. This cannot be definitively stated from the experiments conducted, as the Strehl of the "true" model used to compute the measured intensities for the regression is about 0.89, meaning that the departure from the as designed performance isn't particularly significant (as can be seen in Figure 6.4. This means that if true system performance is greatly degraded by surface quality and cannot be compensated via a DM to return to a high Strehl ratio, this statement may not be true. However, the large drop in estimate quality across the board for using surface errors that are statistically similar to the truth, but different realizations, combined with the fact that some means of calibrating the static error caused by alignment or surface quality (such as dOTF) to return the system performance to a high (0.9+) Strehl ratio can be used, the perfect, as-design model should be what is started with.

If it proves to not be a good enough model to get good estimates of a known aberration in a real application of Frazin's Algorithm, then several possible actions can be taken. The first, although expensive and time consuming, is that an interferometer can be used to measure all the surfaces in the optical system to include them in the computational model in addition to a thorough examination of where the beam actually hits the surface of each optic. This is not a particularly practical solution because it is likely that the beam will move on the surfaces of each optic due to dynamic processes like vibrations and gravity, not to mention this calibration needing to be repeated after every realignment or change to the optical system hardware. The second option, which is vastly more practical, is to exploit the fact that what is driving these estimates away from the known coefficient values is the fact that the surface errors the model does not know about are themselves an NCPA. This means that instead of injected a known NCPA to test, Frazin's algorithm could be run on the "flattened" optical system to obtain an estimate. If the optical system performance is well calibrated to a high Strehl, like MagAO-X [27], the resulting estimates will be turn out to be dominated by the combined OPD effects of the model's missing surface errors. This estimate could then be used to update the computational model used in the regression equations, reducing the bias-like affects on estimating a known (or unknown) aberration in the next implementation of Frazin's algorithm. Although this process will need future study to implement, an attempt was made to demonstrate the fact that the combined effect of the surface errors can be estimated simply by applying Frazin's algorithm using a computational model with assumed perfect surfaces. To do this in simulation, we return to the Mismatch Model case, and apply the estimators without injecting the low-order NCPA. The results of this effort can be seen in Fig. 6.8. The left frame shows the phase (in radians) of the Exit Pupil for the "true" model that includes the surface errors, which is used in the simulations to propagate the light and compute the measured science camera intensity. The right frame shows the bias–corrected estimate achieved applying Frazin's algorithm, which can easily be seen to be a reasonable estimate of the Exit Pupil phase, as we expected. Although this was



(a)



(b)

Figure 6.8: (a) The phase (in radians) in the Exit Pupil of the simulated MagAO-X optical system. This is due to the optical surface errors only. (b) The bias–corrected estimate for the Mismatch Model case without any injected NCPA. This is an estimate of the error in the as designed model because it describes the difference in the science camera intensity predicted by the model with perfectly flat surfaces to the slightly aberrated intensity of the "true" model.

not then fed into the model to test if it improves the results obtained just above (removing the error of the unknown surfaces from the estimate), in principle that should be possible, and as mentioned, requires further study.

Keeping in line with the conclusion that the Mismatching surface WFE Model is the wrong way to go, we again return to the Mismatching Model case, and go one step further to try to determine if model error of this type might be manageable, even without trying to improve the computational model in the suggested ways. To test this, we perform the Mismatching Model simulation again, but in the framework of the real-time QSA correction of Section 6.1.2. In doing this, we hope to be able to show that if the bias-corrected estimator is used in an integrator control loop, even with error in the regression equation model, that we can converge to similar performance as if the computational model is perfect. We run the same simulation as the previous section, of estimating the beam walk-like QSA every 10 seconds

Case	Ideal Comp RMS (rad)	Naïve Comp RMS (rad)	Bias-Corrected Comp RMS (rad)
Matching Model	5.04×10^{-6}	0.016	0.0042
Mismatching Model	0.078	0.066	0.078
Mismatching surface WFE Model	0.144	0.123	0.143

Table 6.1: Summary of the residual RMS error when using the labeled estimator type to compensate a static NCPA for each of the three model error cases. Note the starting RMS phase was 0.5 radian.

before computing an estimate and applying it to the NCPC DM, but with the propagation matrix in the regression equations replaced by the surface error free model matrix. Again, the figure of merit analyzed in this simulation is the RMS of the pupil plane just after the DM compensation is applied. Figure 6.9 plots the RMS phase for the uncompensated QSA (blue line), the residual RMS outside the range of correction of the NCPC DM (orange line), the RMS of the compensated QSA using the estimates of the Mismatching Model case (green solid line), and finally the RMS of the compensated QSA using the Matching Model case, for comparison (red dashed line). The results of this simulations are quite encouraging, because the error in the estimates due to the model error is largely swept under the rug by the control system, achieving nearly the same results as if the regression equation computational model was perfect. This result should provide some comfort as it demonstrates the fact that the effort of modeling the science instrument likely need not involve high precision measurements from interferometers and calibrated knowledge of exactly where the beam hits each surface, as running in a closed-loop integrator control system can still obtain the desired real-time compensation of the QSA, even without trying to push the model to be more precise.

This, of course, is only one source of potential modeling error however. Further

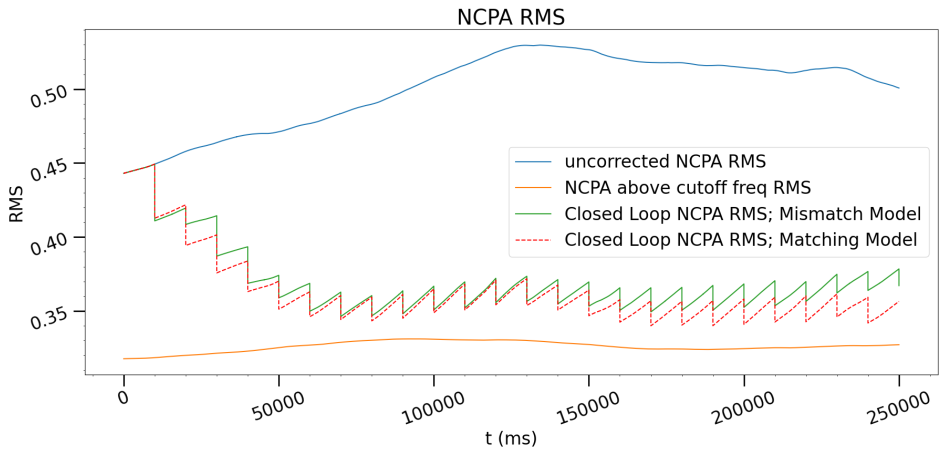


Figure 6.9: A plot returning the real-time simulation compensating the beam walk QSA, but comparing to if the surface error free model is used in the regression equations instead of the "true" model. We see that closing the loop using Frazin's Algorithm largely overcomes the differences in the models to return to the performance of the ideal case where you have perfectly modeled the real optical system.

examination of errors in the numerical model of the WFS used in the bias-corrected estimator, as well as in the measurement and implementation of the AO residual statistics used in modeling the Monte Carlo wavefronts, is required to reveal the full story of the biases introduced due to computational modeling. This is beyond the scope of this dissertation, but is an excellent space for further developing the method by increasing the understanding of an error budget allowed in the computational modeling effort.

6.3 Doing the naïve estimate in the lab

The next big step in the advancement of Frazin's algorithm is to demonstrate its viability in a lab setting. The most tangible starting place for this effort, without having gone through the process of conducting studies on the modeling error budget, is to simply attempt to use the naïve estimator described in Section 4.2.3 to estimate a "known" NCPA, as this requires only the ability to acquire measurements in both a WFS and the science camera, and a good numerical model of the instrument being used (essentially keeping us in the space of errors that are examined in the previous section). As we have constructed a numerical model of the MagAO-X science instrument following the empirical work of Jennifer Lumbres[37] (in prep), in which the model is verified when compared to the hardware testing and calibration done by Van Gorkom [Gorkom et al. (2021)], first discussed in Section 6.1.1, it makes the most sense to attempt to tackle this demonstration using the MagAO-X instrument itself. The experimental setups tested, results, and discussion is the subject of the rest of this section. Although the intricacies of using MagAO-X are important for the completion of the following experiments, the details will be left to the reader to better follow if they are going to attempt to replicate or improve upon what is presented here. For a brief reference, the online MagAO-X handbook includes a beginner's guide to using the instrument can be referenced.

6.3.1 Experimental attempt 1

Assuming that MagAO-X is properly aligned, with streams open to the two important cameras for our purposes: the Pyramid WFS intensity measurements on camwfs, and the science camera, camsci1, the first experiment can be attempted. In this test of the naïve estimator, a closed-loop control loop will be run on the WFS, with turbulence injected via the woofer (11×11 actuators) DM, corrected by the tweeter (64×64 actuators) DM. The Pyramid WFS measures light at 850nm with a 40nm bandpass filter, and is run at 2kHz, for an exposure time of 0.0005 seconds. This is the rate at which the loop is closed, updating the tweeter DM surface through use of a calibrated *reconstructor* matrix to go from camwfs intensity measurements directly to DM actuator offsets representing estimated wavefront piston values at the actuator locations. We will make use of this same reconstructor matrix to assist us in getting OPD measurements of the wavefront to use in the regression equations. The science wavelength impinging on camsci1 is at Hα: 668nm, with an exposure time of 0.0250 seconds. The consequences of the difference in exposure times between the two cameras will be discussed, but in the case of this first attempt at demonstrating Frazin's algorithm in the lab, the science camera is not capable of running at the high frame rate that the WFS is running at.

The observing sequence conducted in this experiment is:

1. take dark images on camwfs and camsci1

2. take images on both cameras with the loop closed on the WFS, without injecting turbulence yet. These images will serve the purpose of allowing us to adjust our model as needed to better replicate the aberration free, flatten wavefront performance of MagAO-X.

3. inject 0.025 micron RMS Astig-V on dmncpc, the DM in a pupil plane in the science instrument downstream of the WFS. This serves the purpose of being our known aberration to estimate, as its presence will only be in the science camera measurements. Furthermore, because of the angled beam path incident on dmncpc, the astigmatism we introduce will actually be a slight projection of Astig-V.

4. take images on both cameras, again with no injected turbulence.

5. configure the turbulence with 0.1 micron amplitude, a wind speed of 20 m/s, and an interval of 1000 microseconds.

6. Inject turbulence on the woofer DM. We note here that that MagAO-X Pyramid WFS is not spatial frequency limited by the tweeter DM, but we choose the woofer to inject the turbulence to keep it much lower order. The merits of this choice will be discussed.

7. take images on both cameras.

With this sequence complete, we move on to the potentially complicated chore of organizing the captured data so that it can be used. Sparing the reader the details of this process, we will simply describe the end goal of the data organization: using the time stamps in the fits file headers for acquisition and write times, all the WFS exposures are grouped to their corresponding science camera exposure times. In other words, for each time interval exposure of the science camera intensity, there are approximately 50 WFS camera intensities that were taken, and the time stamps are used to identify and group them. However, because the AO system was running the WFS at 2kHz, and the science camera at 40Hz, the DM shape also changes 50 times within each science camera exposure. This removes the ability to simply sum the WFS camera intensities together because at each 0.5ms interval, the wavefront reaching the science camera changes (albeit slightly in this experiment because the turbulence is weak and of low spatial frequency content, meaning it is extremely well corrected by the AO system). Instead, we must be able to account for this in our regression equations by taking advantage of the assumption that a "long" exposure image can be approximated by the sum of "short" exposures over the same time span. This essentially introduces a second time index to the equations, with the first, t, tracking the science camera intensities, and the second, t', tracking the WFS measurements. Adding T' as the total number of WFS measurements that fit within a single science camera exposure to the nomenclature, we can adjust Eq. (4.25) to:

$$\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a}) = \sum_{t'=0}^{T'} \left[\mathbf{c}(\boldsymbol{w}_{t'}) + \boldsymbol{A}_{\mathrm{a}}(\boldsymbol{w}_{t'})\boldsymbol{a} \right] , \tag{6.1}$$

where we understand that $\boldsymbol{i}(\boldsymbol{w}_t, \boldsymbol{a})$ on the left-hand side of the equation is measured directly by the science camera, $w_{t'}$ on the right-hand side is measured by the WFS, and summing from $t' = 0 \to T'$ gives a response that approximates the intensity due to w_t. We must also adjust Eq. (4.23):

$$\boldsymbol{A}_{\mathrm{a}}(\boldsymbol{w}_{t'}) \equiv \left. \frac{\partial\, \boldsymbol{i}_\star(\boldsymbol{w}_{t'}, \boldsymbol{a})}{\partial \boldsymbol{a}} \right|_{\boldsymbol{a}_0} . \tag{6.2}$$

We can change the index here only because the derivative being taken is with respect to the aberration parameter vector, $\boldsymbol{a}$, not the wavefront itself. Although we can fix this in the equations and presumably be in the clear for using the data we have, there is a potential red flag with approaching the experimental setup in this manner: each science camera measurement is the average of intensity due to 50 different AO residual wavefronts. As we will see, this will greatly reduce the amount of information that we can gather in the regression, as this average over AO residuals will not change all that much throughout the experiment, especially with the fact that our AO system is doing such a complete effort in compensating the injected turbulence. Attempting to reduce this effect will be the main motivation in the second lab experiment described below.

The next task is to set up converting the camwfs intensity measurements into measurements of the wavefront phase. Luckily, because we closed the loop on MagAO-X, we have access to several pieces of information that can allow us to do this very simply. As part of the process to set up MagAO-X for use, a response matrix is generated, along with a WFS pupil mask, a WFS reference image, and a WFS dark image (measured with the WFS camera shutter closed). Because of how the software is written, there is a simple recipe for going from the WFS intensity directly to an OPD measurement that is well calibrated, and can be converted into phase through knowledge of the operating wavelength of the WFS. Code for how to apply this recipe can be found in Appendix C.3.

Now, a proper S matrix must be computed. This is done very simply, exactly how it was done in the simulations of Chapter 5. The average science camera intensity for the length of the full observation is found by averaging together all the measured camsci1 frames. If all these values were placed on the diagonal of an $L \times L$ matrix, they would form an estimate of the noise covariance matrix, $\boldsymbol{C_y}$. So to get the proper scaling matrix, each value in the average science camera intensity is inverted, and then placed along the diagonal of an $L \times L$ matrix, providing us the proper weighting for our estimated scaling matrix to function as described in Chapter 4.

Finally, we return to the model matrix, D. We have specifically chosen the dimensions of D to be $L \times P$, where L is again the number of pixels in our science camera measurements, in this case 4096, and P is the number of pixels in our WFS measurements after applications of our recipe to go from intensity to phase estimate, in this case 2040. This matching of the model dimension to the physical outputs of our hardware is required to do the calculations in the regression. Following the results of Section 6.2, we also opt for using the computational model that does not include any surface errors, as we know with the AO loop closed, and the previous calibration of the MagAO-X system [27], the penalty on the Strehl ratio due to the surface errors is largely compensated already. However, there is one further step that must be taken: scaling and interpolating the model matrix so that the intensity that is obtained using it to propagate a field is appropriately matched to the camsci1 measurements. This is why our experimental scheme included taking data that was free of aberration. This scenario is the most closely matching to the numerical MagAO-X model that we have used to construct D, allowing for a comparison in features and scaling. Because the loop was closed when taking this aberration free data, the PSF in each measurement does not change substantially aside from the photon and readout noise. This means that we can average together this data set to get a better signal in the hardware measured, aberration free, high Strehl science camera image. With this averaged science camera image in hand, we can choose how to normalize D such that the model outputs a similarly scaled intensity. In this work, we choose to normalize D by the square root of the ratio of the maximum value of the averaged hardware image to the maximum value of the intensity predicted by the model without any scaling. The square root is necessary because D propagates the field, not the intensity. Lastly, in a departure from the simulations presented in Chapter 5, in this case, the WFS and science camera are operating at different wavelengths. We account for this by adjusting D through the use of mild interpolation to resize the plate scale of the modeled intensity to match what we measured in the average image mentioned just above. Although this is an inexact method that is hard to get right, it is the easiest way to try ensure the scaling of the wavefront by wavelength between the measured WFS phase and the science camera is not ignored without a more significant modeling effort. The results of this effort can be found in Fig. 6.10 in the form of comparing intensity measured using MagAO-X with a “flat” wavefront to the intensity the model predicts for a flat wavefront.

After performing all of the above, we arrive at evaluating the regression equations using the MagAO-X matrix model, D, and the data taken in lab on camwfs and

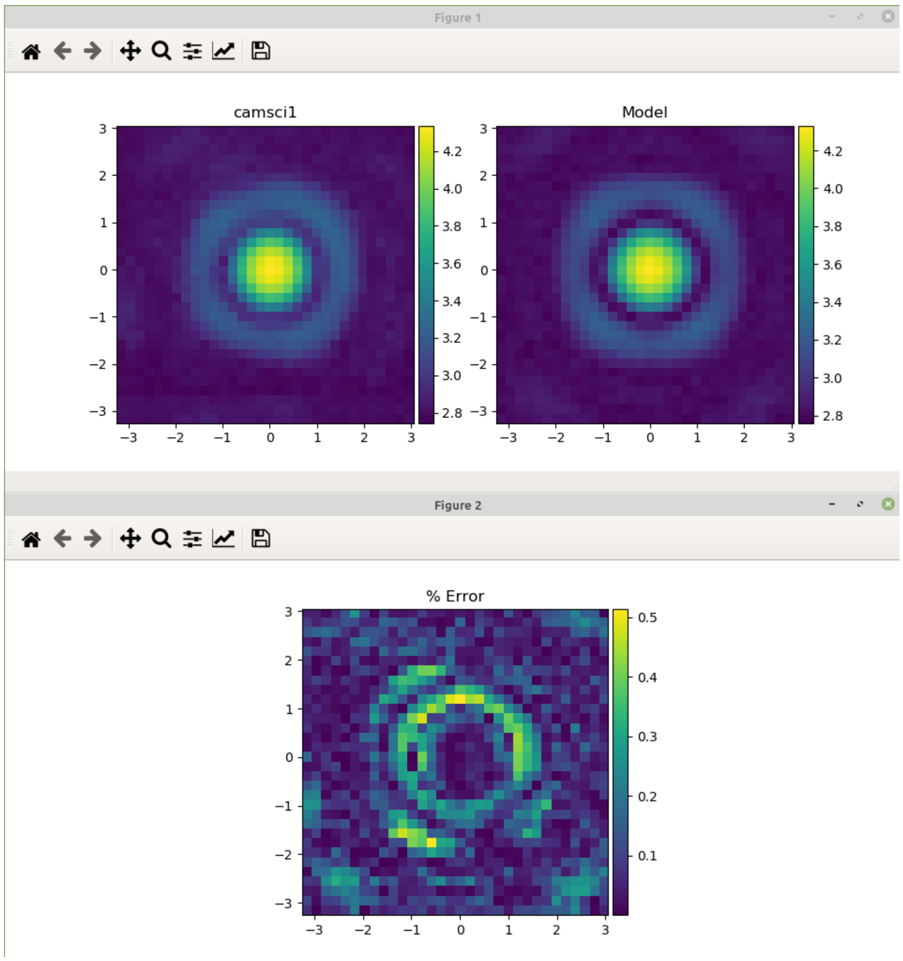


Figure 6.10: Top Left: A single exposure intensity measured by camsci1 without any injected aberration using MagAO-X shown in log10 scale. Axes are in units of λ/D. Top Right: The science camera intensity our model predicts for a flat wavefront shown in log10 scale. Bottom: The percent error of the model predicted intensity image.

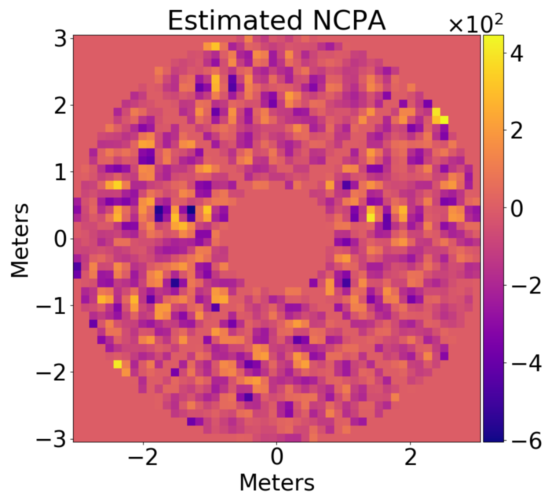


Figure 6.11: Naïve estimate of the NCPA in the first lab experiment, using 622 Bsplines as the estimation basis set, represented in radian units.

camsci1. We first attempt to run the naïve estimator using 2500 science camera measurements and their corresponding WFS measurements to estimate 622 Bspline functions in the pupil. The resulting estimate of the aberration, which to remind the reader, is intended to be the projected version of 0.025 micron rms Astig-V, can be seen in Figure 6.11. As is easy to see, this is a particularly "noisy" estimate of the aberration. There are several likely candidates for the cause of this, but the main one is the problem discussed above, that each science camera image was the average of 50 different AO residuals. We can examine if this is the case by looking at the eigenvalues of the $\boldsymbol{Q}$ matrix that is built up in the evaluation of the regression equations (see Figure 6.12). In this figure, we see that the eigenvalues of $\boldsymbol{Q}$ drop by two orders of magnitude after only about 10 modes of the 622, meaning that for the vast majority of the modes, the estimates are amplified by noise, leading to the very random looking phase values we see, including a few very large displacements in absolute value. This means that we did not get nearly enough information in the data that we took to try to estimate this many modes, with the likely culprit being the averaging problem that occurred in every science camera exposure. This emphasizes the importance of synchronizing the measurements in both telemetry streams, so that the amount of information obtained in several thousand exposures is powerful enough to do the kinds of regressions seen in Chapter 5, where our simulations guaranteed this to occur. Other contributing factors to the lack of precise result include:

1. the slight mismatch between the numerical model of the science instrument and the true hardware,
2. mistakes in the data pipeline, both taking the requested number of frames, and then in organizing the data by time stamp,
3. a lack of knowledge of what the form of the true injected NCPA by the DM is.

As this was a quick first attempt at lab results, future work would be wise to focus on making further improvements to the plate scale of the model (although it is close, it is still affecting the estimation results) by better calibrating the WFS measurements to the focal plane to ensure the model is precise, improving the data stream pipeline to function without errors in the organization and collecting of WFS measurements to their corresponding science camera measurements to ensure that the regression model is comparing the right measurements at the right time step, and verifying the form of the aberration that is injected so that it is truly a "known" quantity to compare estimates to (perhaps through use of a known phase plate in the science instrument entrance pupil or using an interferometer to measure the exact phase off of the DM being used to inject it). Because we don't have a good, calibrated sense of the actual form of the aberration phase, the only way to evaluate the quality of this estimate is to look at the intensity the science instrument model predicts with the estimate as the NCPA compared to the intensity measured by camsci1. Even though the model has the aforementioned minor errors in it, Fig. 6.10 tells us that it should be close enough to draw conclusions on the results of our naïve estimated NCPA. This comparison can be made by looking at Fig. 6.13. In spite of the fact that the estimate was very noisy due to lack of information contained in the measured

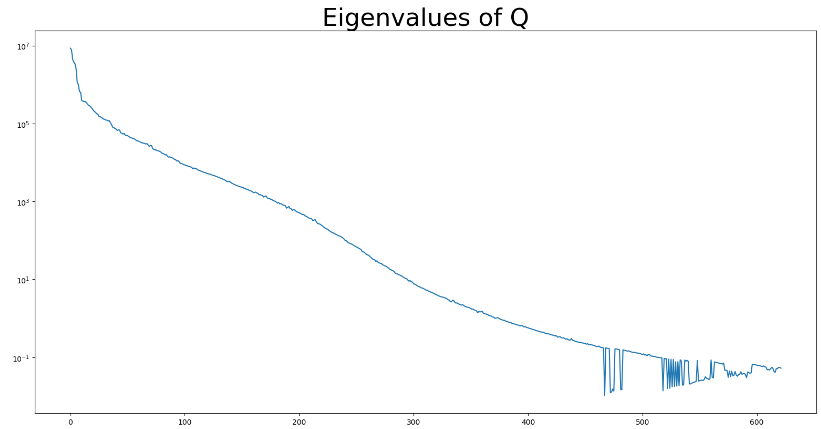
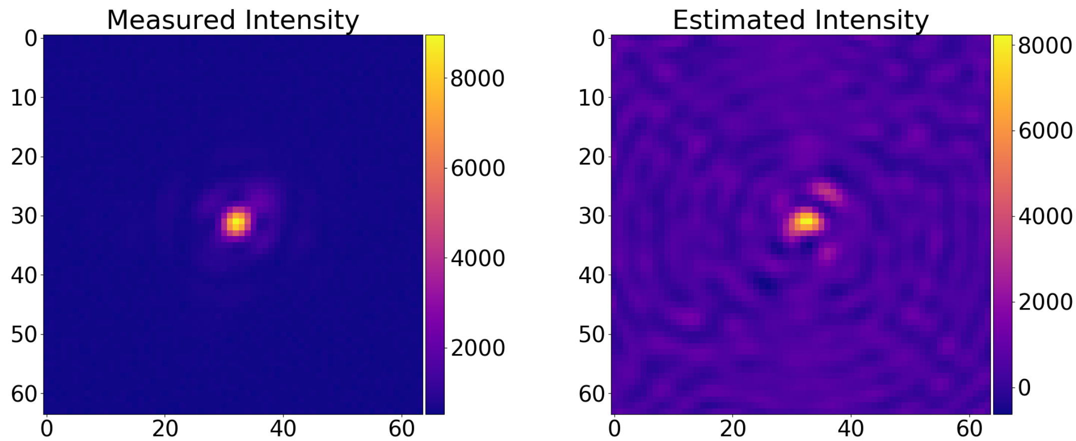


Figure 6.12: Eigenvalues of the $\boldsymbol{Q}$ matrix that is accumulated in the evaluation of the naïve regression equations using the MagAO-X data from experiment 1, using 622 Bsplines as the estimation basis set.

data, we see this is not quite as poor an estimate as it may have first appeared. The noisy, high spatial frequency content of the estimate is evident in the predicted intensity, as there are is significantly higher amplitude content away from the PSF core and first ring. That being said, the peak number of counts in the core, and the general structure of the first diffraction ring are similar enough to make us believe that given a better experimental setup, the naïve estimate could be demonstrated well.

Armed with the knowledge that we only have enough information to estimate a handful of modes using the naïve estimator, we can easily swap the estimation basis to include a small number of Zernike modes instead of 622 Bsplines, and rerun the experiment. This accounts for only a few lines of changed code. The chosen Zernike modes for the estimation basis are Tip, Tilt, Astig-O, Astig-V, Coma-H, Coma-V, Trefoil-O, and Trefoil-V. Defocus and spherical modes are eliminated from the basis set because they are the most likely to be impacted by the slight mismatch in plate scale in the numerical model compared to the hardware; the slight change in PSF core width between the model and the hardware will automatically show up as an increase of defocus to explain it. Figure 6.14 shows the resulting naïve estimate



(a) (b)

Figure 6.13: (a) The measured intensity at one exposure on camsci1 with the injected astigmatism. (b) The model predicted intensity for the estimated aberration using 622 Bspline functions in Frazin's algorithm.

of the phase in this case, which returns an RMS error of 0.153 microns. We see that astigmatism is the dominant aberration found, but that it is more oblique than vertical,differing from the expected form our choice of input aberration on the DM would suggest. However, we have not verified that we truly understand what the form of the aberration actually injected into the hardware via the DM is what we expect it to be, so we cannot draw many conclusions from just looking at this result. Looking at Figure 6.15, we can instead compare the measured intensity on camsci1 and the intensity the model predicts with this estimated NCPA to judge the outcome. Although the peak intensity remains close to the measured value, the form of the estimated intensity shows in the first diffraction ring confirms that the estimate is not fully representative of the true aberration. Four speckles can be seen with symmetry on either side of the core as making up the first ring, whereas the measured intensity shows three main speckles, with the majority of the light skewed toward the lower right corner. The core of the estimated intensity is also stretched along the lower left to upper right diagonal, whereas the measured intensity shows

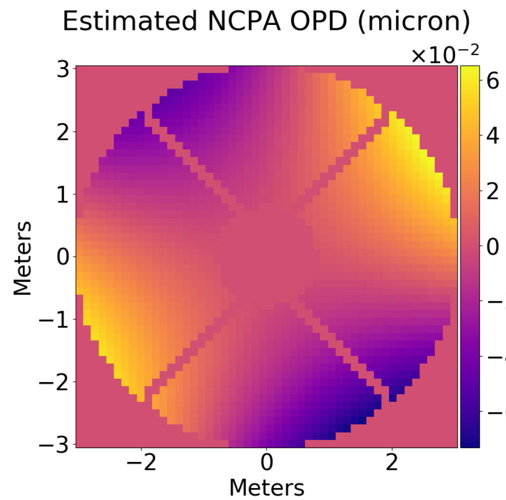
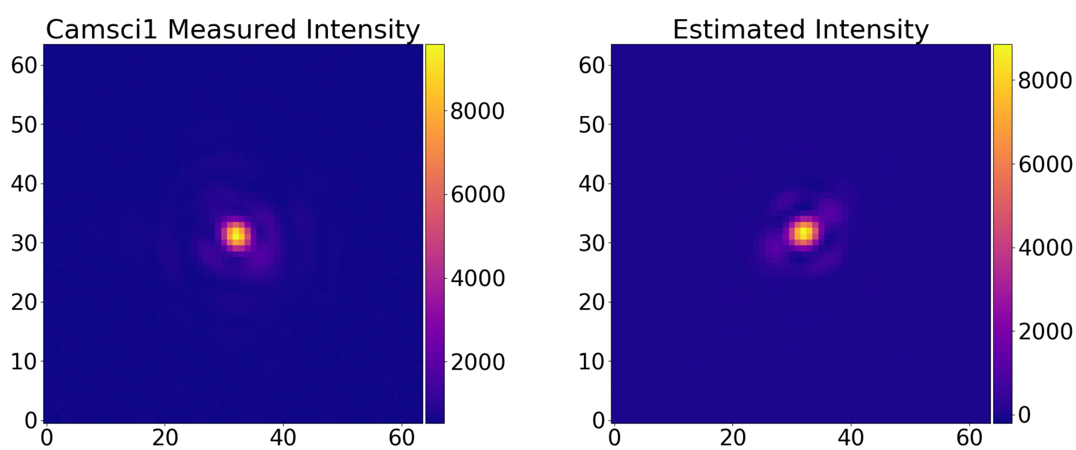


Figure 6.14: Naïve estimate of the NCPA in the first lab experiment, using 8 Zernike modes as the estimation basis set, presented in OPD. The resulting rms is 0.153 micron, 6.1× larger than the injected amount of astigmatism.



(a) (b)

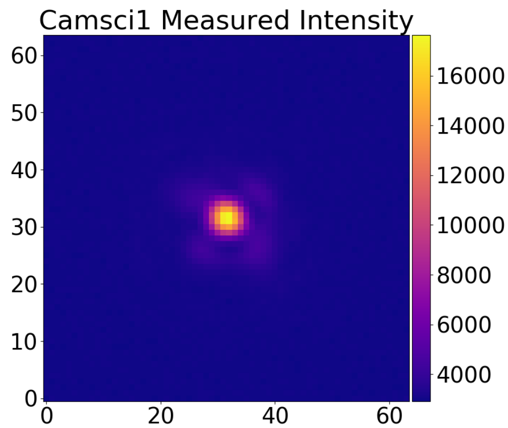
Figure 6.15: (a) The measured intensity at one exposure on camsci1 with the injected astigmatism. (b) The model predicted intensity for the estimated aberration using 8 low order Zernikes in Frazin's algorithm.

a more circularly symmetric core.

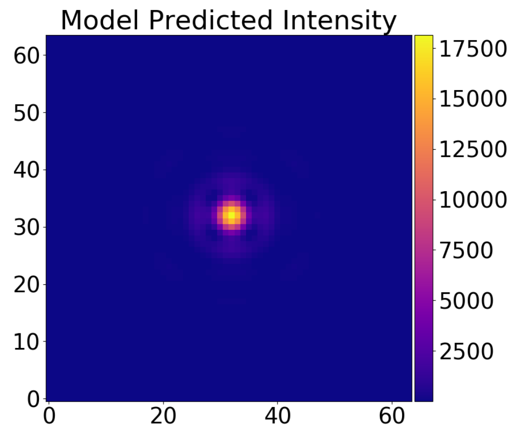
6.3.2 Experimental attempt 2

In an attempt to get better results from a preliminary lab demonstration, a second experiment is designed with hopes of mitigating some of the issues we learned in the first attempt. As we are simply trying to achieve a lab based result that validates the simulations, rather than implementing the full regression method in real time, the measurements on the science camera and wavefront sensor can be slowed down some. In this vein, we decide not to close the loop on the WFS using injected turbulence on the woofer at all. Instead, we will be more deliberate, and after calibrating a response matrix to allow us to convert our measured WFS intensities to wavefront OPD, an AO residual will be approximated by setting the woofer surface using draws from a Gaussian random variable for each actuator. This is to say each actuator piston will be drawn from a Gaussian distribution at each time step, attempting to ensure that each overall DM shape will be statistically independent from the rest. Furthermore, the woofer shape is changed at a much slower pace, allowing the problem that plagued the first experiment of our science camera exposing over multiple DM shapes to be eliminated. For the stability of the software running MagAO-X, the WFS is run at 1kHz this time, and the science camera is run at 32Hz (for no particular reason other than 0.03125 second exposure time is easily within the possible readout capability for camsci1). With the fact that this new experiment is not changing any DM shape while the science camera will be exposing, multiple WFS measurements can be taken and averaged together in intensity to improve the signal. We choose the number of WFS measurements to take to equalize to the exposure time of the science camera, essentially giving us "synchronized" performance. However, to do this experiment, the Python interface to MagAO-X must be used, rather than the default MagAO-X software. This also means that each camera must be exposed sequentially rather than in parallel. We choose first to take the WFS exposures, and then the science camera exposure. The assumption being made here is that the wavefront going through MagAO-X remains stable throughout this longer total time interval for each step (consisting of changing the woofer shape to a simulated AO residual and allowing it to settle, then measuring the WFS intensity, then measuring the science camera intensity, and then writing both to disk) of the experiment. As long as the wavefront does not change due to drift much in this time interval and the assumption holds, constructing the experiment in this manner solves two of the encountered issues in the first experiment: averaging over multiple AO residuals in each science camera exposure and the data organization pipeline having inconsistencies in it. The last step before taking data is then to flatten the wavefront by setting the calibrated flats on all three DMs, and then apply a 0.030 micron RMS Astig-V on dmncpc to serve as the NCPA to be estimated.

To begin this experiment, more care was taken to try to understand if dmncpc was setting the form of aberration we anticipated. To accomplish this, the first step was to take an exposure on camsci1 with the aberration in place, and compare it to what the computational model of MagAO-X we are using predicts the NCPA should be for a 0.030 micron RMS Astig-V mode placed in the entrance pupil. The results of this comparison can be seen in Figure 6.16. It is clear from this figure that the aberration being supplied by dmncpc is not as simple as a regular Zernike mode for Astig-V. This is somewhat expected due to the fact that the input beam to the DM has an angle of incidence of 30 degrees. However, it is much clearer that the actual NCPA appears to be much closer to an oblique astigmatism with a small degree of defocus, as we note that light is shifted more to the lower right hand side of the first diffraction ring than in the expected PSF (similarly to the measured PSF from the previous experiment). Furthermore, the core of the PSF in the measured by the hardware is 8 pixels across and appears slightly elongated from top left to bottom right, whereas the model predicted is only 7 pixels across and displays circular symmetry. Given that the model being used has been verified as shown and discussed above to be free of large errors that could manifest in this way, we are confident that this difference is best explained by the lack of knowing exactly what NCPA we have put in place via the DM compared to the form of the



(a)



(b)

Figure 6.16: (a) The measured intensity at one exposure on camsci1 with the injected astigmatism for the second experimental attempt. Note the longer exposure time used has increased the number of counts compared to the previous experiment. This is expected. (b) The intensity the model predicts for injecting a 0.030 micron RMS Astig-V Zernike mode as the aberration.

modes being used in the regression.

The described scheme is implemented on MagAO-X using the code in Appendix C.4, giving us 250 measurements of WFS and science camera telemetry that are much better paired. These are processed through the naïve regression equations, estimating the coefficients for the same 8 Zernike modes as used in the previous experiment. The resulting estimate of the NCPA is given in Figure 6.17, in units of microns. The RMS of the estimated OPD is 0.0972 micron, 3.89× larger than the injected amount of aberration. However, as discussed comparing the measured intensity to the expected intensity given by the model, it does not make much sense to compare our result to what we thought the input aberration on the hardware is. Instead, we will again turn to comparing the intensity the model predicts given our estimate to the measured intensity, as this will be a more apples-to-apples comparison. This comparison can be seen in Figure 6.18. We see that this estimated phase is in fact a poor representation of the NCPA in the system, as we

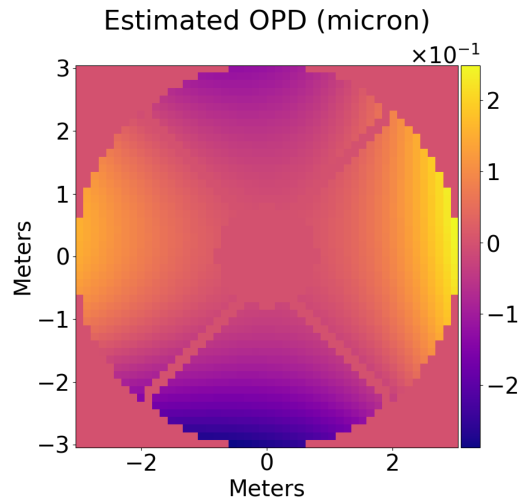
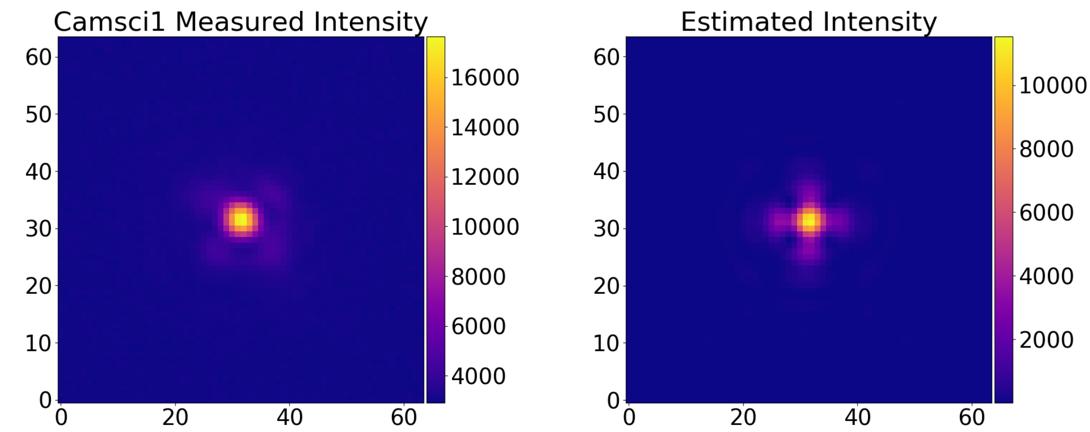


Figure 6.17: Naïve estimate of the NCPA in the second lab experiment, using 8 Zernike modes as the estimation basis set, presented in OPD. The resulting RMS is 0.0972 micron, 3.89× larger than the injected amount of astigmatism.

would have expected from the analysis above that the true NCPA does not take the form of Astig-V even though that is what the DM was commanded to be (because of uncertainty in the shape of the DM, the 30 degree angle of incidence projecting the shape, and any unaccounted for other sources of WFE in the hardware). The estimated intensity displays a much more blatant, stronger astigmatism than the measured, so much so that 6000 counts are spread from the peak pixel out into the canonical astigmatism intensity cross. This tells us that our improved experimental design still has flaws, that appear to be even more damaging than the mistakes made in the first experiment.

Given these results, we present suggestions to further improve the setup in a future attempt. Although not shown, estimates on the 622 Bspline functions were repeated, and suffered the same fate of not having enough information to reliably trust estimating more than 10 modes. This is due to the fact that an abundance of caution was taken commanding the woofer shape to avoid the Gaussian random draws from forcing neighboring actuators from being set with a large surface gradient between them. This essentially means the mean of the Gaussian was chosen to be



(a) (b)

Figure 6.18: (a) The measured intensity at one exposure on camsci1 with the injected astigmatism for the second experimental attempt. (b) The model predicted intensity for the estimated aberration using 8 low order Zernikes in Frazin's algorithm.

very small, making the Strehl ratio remain exceedingly high, and the change in the "atmospheric" speckles hardly noticeable from exposure to exposure. Furthermore, using the 11×11 actuator woofer DM to inject our Gaussian phase screens severely limits the spatial frequency content in the screens, meaning we aren't throwing speckles throughout the full detector region. These affects combine to produce nearly the same problem that experiment one had; the information we were able to gather in 250 measurements on the hardware was not adequate for precise estimation of the NCPA because there was simply not adequate modulation of the NCPA for the regression to work. So we suggest that future work on a lab demonstration can make several improvements to this second framework, and the naïve estimate can be fully verified. These improvements include:

1. Switch to using the 64×64 actuator tweeter DM for injecting the Gaussian random phase screen in to the system. This will give access to high spatial frequencies, allowing the speckles in subsequent exposures to change more freely and better modulate the NCPA.

2. Increase the amplitude of the Gaussian phase screen serving as the AO residual to lower the Strehl to ≈ 0.7. This also ensures more light is scattered in to speckles throughout the focal plane, allowing more modulation of the NCPA and thus more information to be gathered.

3. Run the science camera a bit faster. This is a hardware limitation of the EMCCD currently in use as camsci1. However, any increase in readout speed reduces the dependence on the assumption that the system is free of any unaccounted for drift over the experimental steps.

4. Adjust the sleep times in the Python code between each link in the chain occurring. As performed in the second experiment, the WFS likely took some of its 32 frame sample while the DM shape was still changing, corrupting the wavefront measurement when compared to the measured science camera intensity.

5. Take several thousand exposures, rather than only 250, if system stability allows.

6. Spend more time carefully developing a numerical model of the MagAO-X science instrument, using LightTrans software.

7. Construct an accurate computational model of the MagAO-X Pyramid WFS. This will reduce dependence on the calibration of a reconstructor matrix, for which the calibration starts to degrade immediately as the system drifts with time.

8. Calibrate the form of the aberration being injected into the hardware by dm-ncpc. This can also include changing the NCPA modes included in the regression to better be able to fit the aberration, which could be checked numerically prior to conducting the experiment, ensuring a good estimate is even possible with the chosen modes.

9. Generate the Gaussian phase screens to be used in the experiment beforehand, and conduct simulations along the lines of those presented in Chapter 5 with them as the AO residuals using the naïve estimator, the computational MagAO-X model (for both the propagation and regression; as in the Matching Model case from above), the calibrated NCPA form, and the new set of modes demonstrated to fit the NCPA well. These simulations will inform how the lab experiment should be expected to perform in terms of quality of estimate, and how many time steps are required to be measured. Any deviation from these results in the lab experiment would be easier to pin down the cause of because more parameters would already be verified as functional.

The most difficult and time consuming of these suggestions are the final four, as they require more effort towards empirically modeling as built hardware and implementing that work into simulation to verify the expectations of the lab demonstration. The potential investment is worth this effort though, because this will further allow for the validation of the bias-corrected estimator without much further change to the experiment. Because the AO residual itself is being created via a known multivariate Gaussian distribution, the statistics of the AO residual are very well known (to within uncertainties actually moving the DM surface, which could be calibrated and included in the wavefront modeling). Thus, a set of Monte Carlo wavefronts would be quite easy to construct and utilize with the computation model of the WFS, allowing all the Monte Carlo proxies to be found. Taking these steps will provide the most concrete demonstration that validates the equations and simulation work presented in this dissertation, and bringing Frazin's algorithm closer to the point of on-sky deployment.

6.4 Future work

Throughout this chapter, many suggestions for the continued development of Frazin's algorithm are given in the context of work that is presented. This includes:

- Exploring the limits of implementing Frazin's algorithm as an "NCPA WFS" in a closed-loop control system (see Sec. 6.1).

- Exploring the effects of science instrument computational model errors on the implementation of the method, and ways of trying to mitigate them (see Sec. 6.2.

- Exploring the effects of errors in the computational model of the WFS, as well in the measurement of the spatial statistics of the AO residual and the generation of the Monte Carlo wavefronts

- Ways of improving the presented experimental setups to achieve a much more accuate and precise estimate using the naïve and/or bias-corrected estimators to validate the equations and simulations of this dissertation.

In addition to these topics, there are several other intriguing places to develop the method that were not within the scope of this dissertation, and would make an excellent place for future work to take place. These include:

- Utilizing a spatial filter in the optical system. The optimal way of doing this would appear to be downstream of a simple, low-order AO system, and prior to a high-order Extreme AO system (for example, after AO188 but before SCExAO at the Subaru telescope). In this way, the light would be partially compensated by the lower-order AO system, making for a cleaner implementation of a spatial filter prior to the high-order AO system that will be responsible for providing the telemetry streams to Frazin's algorithm from its WFS and associated science instrument. The reason a spatial filter would be useful is that it would provide a known spatial frequency cutoff to the light entering the system we want to use. This not only assists with the modeling effort, but also reduces the spatial frequencies in the turbulent wavefront, helping to minimize the WME due to bandwidth.

- Generalize the regression to take in to account field rotation and spectral information. The work presented ignored field rotation, and was performed in quasi-monochromatic light. Integrating field rotation and multiple wavelengths into the optical system model will greatly increase the diversity of the information that can be fed in to the regression at each time step, and increase the power of estimators. This also reduces some of the headache of requiring quasi-monochromatic light to be used in observation, as well as any processing of telemetry to remove field rotation effects.

- Examine the effects of both pupil sampling in the WFS and PSF sampling in the science camera on the number and spatial frequency content of modes that can be included in the regression equations. The work presented in this dissertation used approximately the same sampling for both planes in all the simulations. This leaves the relationship between this sampling and the NCPA estimation unknown, and may reveal an optimal choice that achieves estimates containing the necessary spatial frequencies without slowing down the calculations.

- Conduct a study on the required frame rate of the WFS and science camera. Although 1ms has been discussed in this work, and shown to be sufficient for use alongside an AO system in closed-loop in the simulations presented (where 1ms was shown to even be too long to see statistically independent information in each exposure) and through examination of speckle lifetimes against various parameters for a simple integrator control system (among many other parameters) by Males et al. [47], this may not be the case if the algorithm is instead applied in open-loop. Because the RMS phase error due to turbulence would be ≈ 1 rad in open-loop, with a coherence time on the order of $5 - 10$ms, the WFE would likely be changing fast enough that ms exposures would not effectively freeze the swarm of speckles, and instead would be averaging over changing speckles. This would greatly damage the effectiveness of the algorithm, similarly to what happened in the first lab experiment that was discussed. Understanding the relationship of frame rate vs. temporal atmospheric effects with regard to implementing Frazin's algorithm is an unexplored area that could better extend the ability to use the algorithm beyond closed-loop AO systems.

CHAPTER 7

Conclusions

In the work presented in this dissertation, we have derived and explored two wavefront control techniques that can be implemented to estimate non-common path aberrations: dOTF and Frazin's algorithm. To remind the reader, dOTF is a simple, non-interferometric, non-iterative method used to estimate any static wavefront error due to misalignment, flexure, or lack of segment cophasing, prior to the system being used in a dynamic environment like astronomical observation. Furthermore, we demonstrated its use as the WFS in a closed-loop, self-calibration control system within an AO enabled optical system as a means of compensating this static systematic error without incurring any additional costs for new hardware. This is important to use in high-contrast imaging applications because a cleaner, more flat wavefront propagating through the coronagraph improves the starlight suppression performance, meaning a better achieved raw contrast, and a better chance to directly observe light from an exoplanet. With the static, systematic WFE corrected prior to going on sky, the starting point for a second algorithm to be implemented while observation is being taken place is significantly improved.

Frazin's algorithm is a statistical regression framework that takes ms telemetry from the WFS and science camera, combined with computational models of the WFS and science instrument and knowledge of the spatial statistics of the AO residual wavefront, to output estimates of quasi-static NCPA and the exoplanet image. The mathematical steps of the framework were rigorously derived, and then demonstrated via a quasi-realistic end-to-end simulation of a closed-loop extreme AO system with a Lyot coronagraph in the science instrument. In these simulations, it was shown that implementing Frazin's algorithm in a real-time control loop was able to reduce the RMS phase of a realistic quasi-static NCPA with a lifetime of approximately 5 minutes by a factor of $\approx 1.5\times$, and maintain that level of compensation over the duration of the lifetime. This reduces the number of stellar photon that fall into quasi-static and pinned speckles, effectively removing them from the science camera measurements of the intensity. Because of this, and the fact that the exoplanet image can be jointly estimated with the NCPA, post-processing differential imaging techniques become less necessary to be able to directly image exoplanets. Along those lines, the simulations presented demonstrated the ability to jointly estimate the exoplanet image alongside the NCPA when implemented in a high-contrast imaging system, and nn just 4 minutes of simulated observation, a detectable contrast of between 10^{-5} and 10^{-6} at the IWA was achieved. This is comparable to the best contrast ever achieve on the SPHERE AO system, of $\approx 5 \times 10^{-5}$ at $3\lambda/D$ [81]. But the SPHERE data took 2.5 hours of observation with an IFS, and utilized both ADI and SDI to achieve, with a much higher Strehl ratio than our simulations assumed.

Finally, we presented preliminary studies conducted on directions the future development of Frazin's algorithm can be expanded, and provided a road map for intriguing ideas that in principle are all possible to add to the method to make it even more powerful. Achieving the implementation of these suggestions, as well as increasing the observation length and quality of the coronagraph (from a basic Lyot to something more modern like a vAPP or PIAACMC), combined with the fact that Frazin's algorithm can self-consistently estimate the NCPA and control it over that increased observation time to keep the coronagraph performing closer to optimally, means we can push this detectable contrast toward being able to directly image an Earth-like exoplanet in a habitable zone around another star, and finally start to answer the question of if we are alone in the universe.

APPENDIX A

The least squares regression solution

Here we step through some of the math left out of Chapter 4. As is traditional in regression theory, we define our starting equation as:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\nu}\,, \tag{A.1}$$

where $\boldsymbol{y}$ is an $M \times 1$ vector representing the dependent variable, $\boldsymbol{A}$ is an $M \times N$ matrix of independent variables that are exactly known, $\boldsymbol{x}$ is an $N \times 1$ vector we eventually want to estimate, and $\boldsymbol{\nu}$ is an $N \times 1$ vector representing an unknown, additive noise that is assumed to be drawn from a zero-mean Normal distribution with a known $M \times M$ covariance matrix, $\boldsymbol{C}$. In an optical problem, the general interpretation of this equation is that $\boldsymbol{y}$ is the image measured by an optical system modeled by $\boldsymbol{A}$ of an object $\boldsymbol{x}$, with measurement noise $\boldsymbol{\nu}$. This or similar interpretation is applied to the work in this dissertation, often both in terms of performing the AO corrections via wavefront reconstruction, and in the regression method described. With this in mind, we will proceed in this appendix only treating the mathematics themselves, without any thought to what the variables themselves represent.

With Eq. (A.1) setting up our system, we turn our attention to how we can recover what $\boldsymbol{x}$ is given measurements of $\boldsymbol{y}$ and the model $\boldsymbol{A}$. In order to do this, we will employ the use of statistical inference. First, we define the likelihood function as:

$$P(y|x) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{C}|}} \exp\left[-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x})\right] = \mathcal{N}(\boldsymbol{y}; \boldsymbol{A}\boldsymbol{x}, \boldsymbol{C})\,. \tag{A.2}$$

Now, we wish to maximize the log-likelihood to get what is called the ML estimate of $\boldsymbol{x}$. In other words:

$$\begin{aligned}
\hat{\boldsymbol{x}}_{ML} &= \operatorname{argmax}_{\boldsymbol{x}} \ln\left(P(\boldsymbol{y}|\boldsymbol{x})\right) \\
&= \operatorname{argmax}_{\boldsymbol{x}} \left(-\ln\left(\sqrt{(2\pi)^N |\boldsymbol{C}|}\right) \left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \right] \right) \\
&= \operatorname{argmax}_{\boldsymbol{x}} \left(\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) \right) \\
&= (\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{y}\,. \qquad \text{(A.3)}
\end{aligned}$$

The keen reader will recognize this as the least-squares, minimum norm solution, which will produce an unbiased, minimum variance estimate $\hat{\boldsymbol{x}}_{ML}$. However, $\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}$ is often nearly singular, which will make $\hat{\boldsymbol{x}}_{ML}$ unacceptable. In order to proceed, a Prior, $P(\boldsymbol{x})$ is introduced to bring us into the realm of Bayesian Regularization. With the Prior defined as:

$$P(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x}; \boldsymbol{x}_0; (\beta\,\boldsymbol{\Xi})^{-1}) \propto \exp\left[-\frac{\beta}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T \boldsymbol{\Xi}(\boldsymbol{x}-\boldsymbol{x}_0) \right], \qquad \text{(A.4)}$$

where β is the regularization parameter and is greater than or equal to 0, and $\boldsymbol{\Xi}$ is the $N \times N$ regularization matrix.

Now, we continue by finding the joint probability density on $\boldsymbol{y}$ and $\boldsymbol{x}$:

$$P(\boldsymbol{y},\boldsymbol{x}) = P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x}) \propto \exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x})^T \boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}) - \frac{\beta}{2}(\boldsymbol{x}-\boldsymbol{x}_0)^T \boldsymbol{\Xi}(\boldsymbol{x}-\boldsymbol{x}_0) \right]. \qquad \text{(A.5)}$$

But what we would like to have is knowledge of $\boldsymbol{x}$ given a measurement of $\boldsymbol{y}$, so we apply Bayes Rule:

$$P(\boldsymbol{x},\boldsymbol{y}) = P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x}) = P(\boldsymbol{x}|\boldsymbol{y})P(\boldsymbol{y})\,, \qquad \text{(A.6)}$$

to find the posterior distribution:

$$P(\boldsymbol{x}|\boldsymbol{y}) = \frac{P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x})}{P(\boldsymbol{y})}\,. \qquad \text{(A.7)}$$

Finally, we define the maximum a posteriori estimate of $\boldsymbol{x}$:

$$\begin{aligned}\hat{\boldsymbol{x}}_{MAP} =& \text{argmax}_{\boldsymbol{x}}\left[P(\boldsymbol{x}|\boldsymbol{y})\right]\\ &= \text{argmax}_{\boldsymbol{x}}\left[P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x})\right]\\ &= \text{argmax}_{\boldsymbol{x}}\left[P(\boldsymbol{y},\boldsymbol{x})\right]\\ &= \left(\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A} + \beta\,\boldsymbol{\Xi}\right)^{-1}\left[\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{y} + \beta\,\boldsymbol{\Xi}\boldsymbol{x}_0\right]\big)\ .\end{aligned} \tag{A.8}$$

With some deft renaming of variables, it is easy to see this is the process that gives us Eq. (4.35).

Now, to show that this is an unbiased estimate, we simply recall that $\langle y\rangle_\nu = \boldsymbol{A}\boldsymbol{x}$, and compute the expectation value of Eq. (A.8).

$$\langle\hat{\boldsymbol{x}}_{MAP}\rangle = \left(\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A} + \beta\,\boldsymbol{\Xi}\right)^{-1}\left[\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{x} + \beta\,\boldsymbol{\Xi}\boldsymbol{x}_0\right]\big)\ .$$

When $\beta = 0$, this simplifies to:

$$\begin{aligned}\langle\hat{\boldsymbol{x}}_{MAP}\rangle &= \left(\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}\right)^{-1}\left[\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}\boldsymbol{x}\right]\big)\\ &= \left(\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}\right)^{-1}\left[\boldsymbol{A}^T\boldsymbol{C}^{-1}\boldsymbol{A}\right]\boldsymbol{x}\big)\\ &= \mathbb{1}\boldsymbol{x} = \boldsymbol{x}\,,\end{aligned} \tag{A.9}$$

meaning that when unregularized, the estimate is unbiased. If β is chosen to be nonzero, the result is taken to be acceptably biased, as is described above, which is a standard trade-off in regularization theory.

APPENDIX B

Creating a matrix representation of an optical system

With the need for a model matrix of an optical system to perform Frazin's Algorithm, we take a moment to describe a simple and easy way to construct it. This is to take advantage of the linearity of the optical system. This requires that the optical system being modeled to be adequately discretized such that the field in the entrance pupil, $\boldsymbol{u}_0(\boldsymbol{r})$, is represented by P points, or pixels, and the field in the science detector, $\boldsymbol{u}_1(\boldsymbol{r'})$, is represented by L pixels. Given a numerical operator that describes the propagation of the light through the optical system, $\mathcal{O}$, that is taken to have a discrete representation, which is a typical numerical implementation using Fresnel or Fraunhofer integrals transferring the field from the entrance pupil to the science camera, requiring models of optical elements as necessary. Defining a basis set on the points in the pupil made of vectors $\{\boldsymbol{b}_p\},\ 0 \leq p < P$, such that $\boldsymbol{b}_p$'s only nonzero element is the pth one. Exploiting the linearity of the optical system being modeled, we can write the final field in the science camera as:

$$\boldsymbol{u}_1(\boldsymbol{r'}) = \sum_{p=0}^{P} \boldsymbol{u}_0(\boldsymbol{r}_p)\mathcal{O}(\boldsymbol{b}_p; \boldsymbol{r'})\,. \tag{B.1}$$

This equation shows that $\mathcal{O}(\boldsymbol{b}_p; \boldsymbol{r'})$ can thus be stored as values in a $L \times P$ matrix, D, that collects the field response of the optical system in the science camera for each individual input pixel in each row. The benefit of taking such an approach to model the optical system as D is two-fold. First, this matrix can be inserted into the equations given in Chapter 4 directly, allowing easy application of the method. Second, application of the matrix on an input field vector via one complex-valued matrix-vector multiplication (MVM) returns a vector representing the field propagated to the science camera. This means, at the cost of a potentially large upfront cost in computational time, depending on the complexity of the *mathcalO* operator, to compute D, all future numerical propagations can be done with this one MVM, avoiding any need for real-time Fourier transforms in the numerical processing. This process is clearly defined for a Pyramid WFS in Frazin (2018). Listing C.2 provides code using Jennifer Lumbres' Fresnel propagation work to model MagAO-X in this manner. Listing C.7 includes code for creating D for several optical systems, including vAPP and Lyot coronagraphs, and a system without a coronagraph.

APPENDIX C

Selected code implementations †

Listing C.1: Code for running dOTF experiments on CACTI

```
%reload_ext autoreload
%autoreload 2
%matplotlib inline
import datetime

#load modules
import numpy as np
import matplotlib.pyplot as plt
from astropy import units as u
from astropy.io import fits

# accessing the cameras
from magpyx.utils import ImageStream

import os
import struct
import pickle

```

†All code used can be found on GitHub upon request.

```python


def colorbar(mappable):
    from mpl_toolkits.axes_grid1 import make_axes_locatable
    import matplotlib.pyplot as plt
    last_axes = plt.gca()
    ax = mappable.axes
    fig = ax.figure
    divider = make_axes_locatable(ax)
    cax = divider.append_axes("right", size="5%", pad=0.05)
    cbar = fig.colorbar(mappable, cax=cax)
    plt.sca(last_axes)
    return cbar


# initialize the camera
cam_lgsfp = ImageStream('camlgsfp')
# settings: 1.45 ms framerate, HeNe ND at 2.0, no extra ND


# initialize the DM channel 2 (not where the flat is loaded)
dm = ImageStream('dm00disp02') # channel 2
with fits.open('/opt/MagAOX/calib/dm/bmc_1k/bmc_2k_actuator_mask.fits') as f:
    dm_mask = f[0].data
with fits.open('/opt/MagAOX/calib/dm/bmc_1k/bmc_2k_actuator_mapping.fits') as f:
    dm_map = f[0].data

dm_mask_filled = np.ones((32,32), dtype=bool)


# set the camera semaphore
if cam_lgsfp.semindex is None:
    cam_lgsfp.semindex = cam_lgsfp.getsemwaitindex(1)



# take measurements with poked DM
# HeNe = 0.001 s exposure
# LGSsrc = 0.003 s exposure with OD in
val = 0.075
dm.write(np.zeros((32,32)).astype(dm.buffer.dtype))
flat = np.zeros((32,32)).astype(dm.buffer.dtype)
# HeNe
act2poke = np.array([15,27])
# LGSsrc
```

```python
#act2poke = np.array([15,6])

poke = flat.copy()
poke[act2poke[0],act2poke[1]] = val
cam_lgsfp.semflush(cam_lgsfp.semindex)



# Flatten the DM
dm.write(np.zeros((32,32)).astype(dm.buffer.dtype))
# take measurements with the flat DM
nimages = 250
nrepeats = 10
#dm.write(np.zeros((32,32)).astype(dm.buffer.dtype))
flat = np.zeros((32,32)).astype(dm.buffer.dtype)
images = []
poke_images = []
cam_lgsfp.semflush(cam_lgsfp.semindex)

for n in range(nrepeats):
    print(n)
    # positive values
    cam_lgsfp.semflush(cam_lgsfp.semindex)
    dm.write(flat.astype(dm.buffer.dtype))
    #sleep(0.05)
    cam_lgsfp.semwait(cam_lgsfp.semindex)
    #images.append(np.mean(cam_lgsfp.grab_many(nimages), axis=0))
    images.append(cam_lgsfp.grab_many(nimages))
    #dm.write(np.zeros((32,32)).astype(dm.buffer.dtype))

    cam_lgsfp.semflush(cam_lgsfp.semindex)
    dm.write(poke.astype(dm.buffer.dtype))
    cam_lgsfp.semwait(cam_lgsfp.semindex)
    #poke_images.append(np.mean(cam_lgsfp.grab_many(nimages), axis=0))
    poke_images.append(cam_lgsfp.grab_many(nimages))

images = np.asarray(images)
poke_images = np.asarray(poke_images)




images_ = images.sum(0)
poke_images_ = poke_images.sum(0)

```

```
centerpoint_flat = [337,195] # HeNe
#centerpoint_flat = [338,319] # LGS
images_ = images_.sum(0) / images_.shape[0]
images_ = images_[(centerpoint_flat[0]-128):(centerpoint_flat[0]+127), (centerpoint_flat[1]-128):(
    centerpoint_flat[1]+127)]

centerpoint_poke = [337,195] # HeNe
#centerpoint_poke = [338,319] # LGS
poked_images_ = poke_images_.sum(0) / poke_images_.shape[0]
poked_images_ = poked_images_[(centerpoint_poke[0]-128):(centerpoint_poke[0]+127), (centerpoint_flat
    [1]-128):(centerpoint_flat[1]+127)]

OTF_images = np.fft.fftshift(np.fft.fft2(np.fft.fftshift(images_)))
OTF_poked_images = np.fft.fftshift(np.fft.fft2(np.fft.fftshift(poked_images_)))
OTF_images[127,127] = 0
OTF_poked_images[127,127]=0
dOTF_HeNe = (OTF_images - OTF_poked_images)
dOTF_HeNe = -1j*dOTF_HeNe.conj()

mag_HeNe = np.abs(dOTF_HeNe)
pha_HeNe = np.angle(dOTF_HeNe)

#uw_pha_HeNe = uwrap_v1(pha_HeNe,'unwt')
uw_pha_HeNe = pha_HeNe.copy()
k_HeNe = -(2*np.pi) / (632.8e-9)
k_LGSsrc = -(2*np.pi) / (531e-9)
OPL_HeNe_ref = uw_pha_HeNe / k_HeNe


plt.figure(1);plt.clf()
cbax = plt.imshow(mag_HeNe,origin='lower',cmap='plasma')
cb = colorbar(cbax); cb.formatter.set_powerlimits((0,0)); cb.formatter.set_useMathText(True);
plt.title('HeNe')

plt.figure(2);plt.clf()
cbax = plt.imshow(OPL_HeNe_ref,origin='lower',cmap='plasma')
cb = colorbar(cbax); cb.formatter.set_powerlimits((0,0)); cb.formatter.set_useMathText(True);
plt.title('HeNe')



# Write the data to FITS
now = datetime.datetime.now().isoformat()
images.shape
hdr = fits.Header()
```

```
hdr.set('laser', 'hene', "source type (hene or lgs)")
#hdr.set('laser', 'lgssrc', "source type (hene or lgs)")
hdr.set('actPoke', False, "bool if this is poked data")
#hdr.set('poked_actuator', act2poke, "Actuator array location")
hdr.set('exptime', 0.0015, "camera exposure time in sec")
hdr.set('nd', 1.5, "ND at the laser")
hdr.set('cam_nd', True, "bool if ND at the detector")
hdr.set('dichroic', 'none', "dichroic trans/refl/none")
hdr.set('eyedr', True, "bool if eye doctor performed")
fits_filename = 'Data/HeNe_Data_9_12_2021/Images_HeNe_flat_noDMflat_'+now[0:19]+'.fits'
#fits_filename = 'Data/LGS_Data_9_12_2021/Images_LGS_flat_0_05Tilt'+now[0:19]+'.fits'
fits.writeto(fits_filename, images, overwrite=True)


# Write the data to FITS
poke_images.shape
tmp = act2poke[0]*32 + act2poke[1]

hdr = fits.Header()
hdr.set('laser', 'hene', "source type (hene or lgs)")
#hdr.set('laser', 'lgssrc', "source type (hene or lgs)")
hdr.set('actPoke', True, "bool if this is poked data")
hdr.set('pokedAct', tmp, "Actuator array location")
hdr.set('exptime', 0.0015, "camera exposure time in sec")
hdr.set('nd', 1.5, "ND at the laser")
hdr.set('cam_nd', True, "bool if ND at the detector")
hdr.set('dichroic', 'none', "dichroic trans/refl/none")
hdr.set('eyedr', True, "bool if eye doctor performed")
fits_filename = 'Data/HeNe_Data_9_12_2021/Images_HeNe_poked_noDMflat_'+now[0:19]+'.fits'
#fits_filename = 'Data/LGS_Data_9_12_2021/Images_LGS_poked_0_05Tilt'+now[0:19]+'.fits'
fits.writeto(fits_filename, poke_images, hdr, overwrite=True)
```

Listing C.2: Code for running Jennifer Lumbres' Fresnel model of MagAO–X to make a propagation matrix

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Thu Jul  1 20:42:53 2021

@author: Jennifer Lumbres and Alexander Rodack
"""

import sys
sys.path.insert(0,'/home/archdaemon/Research/Github/poppy/')
```

```
sys.path.insert(0,'/home/archdaemon/Research/Github/magaox_poppy')
sys.path.insert(0,'/home/archdaemon/Research/Github/model_kit')

import numpy as np
from astropy import units as u
from astropy.io import fits
import copy
import time

import matplotlib.pyplot as plt
from matplotlib.colors import LogNorm, Normalize
import matplotlib

# need to look up which version
import pickle

# the only thing you need from my code
from model_kit import magaoxFunctions as mf

import poppy
poppy.__version__



mag_pupil = mf.mag_pupil_mask(samp=538, entrance_radius=3.25*u.m, wavelength=656e-9*u.m, bump=False)


# set up file locations
home_dir = '/home/archdaemon/Research/GitHub/magaox_poppy/' # change for your MagAO-X Fresnel directory
data_dir = home_dir + 'data/'
rx_dir = data_dir + 'rxCSV/'

# Set up some stuff
FP_crop = 64
poppy_pupil_space = np.load(home_dir + 'poppy_pupil.npy')
pup_norm_val = poppy_pupil_space.max()
pupil_mask = poppy_pupil_space.copy() / pup_norm_val
pupil_mask[pupil_mask>0.] = 1.0
numPoints = int(pupil_mask.sum())




binWidth = 11

```

```python

# Load in 550x550 preconstructed mask
with fits.open('/home/archdaemon/Research/GitHub/magaox_poppy/demo/LabDemo_mask.fits') as hdul:
    tmp_mask = np.float32(hdul[0].data)

pupil_size_binningy = tmp_mask.shape[0]
pupil_size_binningx = tmp_mask.shape[1]



# declare MagAO-X variables
fr_parm = {'wavelength': 656e-9 * u.m,
           'npix': 538, # sample size
           'beam_ratio': 0.25, # oversample
           'leak_mult': 0.01, # vAPP leakage multiplier
           'bump': True, # T/F to use the MagAO-X pupil with tweeter bump masked
           'wfe_data': 'common'} # other options: common

# set up prescription details, this is important for labeling later.
wavelen_str = str(np.round(fr_parm['wavelength'].to(u.nm).value).astype(int))
br = int(1/fr_parm['beam_ratio'])
parm_name = '{0:3}_{1:1}x_{2}nm'.format(fr_parm['npix'], br, wavelen_str)


# load the CSV prescription values
#rx_loc = rx_dir+'rx_magaox_NCPDM_sci_{0}_noap_{1}_openfits.csv'.format(parm_name, fr_parm['wfe_data'])

# Change to this one if no wfe on surfaces!
rx_loc = rx_dir+'rx_magaox_NCPDM_sci_538_4x_656nm_noap_nowfe.csv'
rx_sys = mf.makeRxCSV(rx_loc, print_names=False)


# Commented out to remove surface qualities (Lines 93-109), also need to set rx_loc to the nowfe csv (see above)
# We need to rename all the folder location for the PSD WFE surfaces
# In theory, you can skip this if your rxCSV file has this fixed. (it's not in this file)
# This left here for being able to switch with different PSD WFE sets.

# Setting the folder name
#n_set = 0
#s_set = 0
#psd_wfe_dir = 'data/PSD_WFE/n{0}/s{1}/'.format(n_set, s_set) # Change this for what you want, and make the folder too.
#
## Quick renaming of the folders
```

```
#for t_optic, test_opt in enumerate(rx_sys):
#    if test_opt['surf_PSD_filename'][0:7] == 'wfe_psd':
#        test_opt['surf_PSD_folder'] = psd_wfe_dir
#
#
## Tweeter index number in the rxCSV (particularly the Optical_Element_Number value)
#j_tweeter = 16
#
## Rename the folder and filename for the pre-solved Tweeter surface
#rx_sys[j_tweeter]['surf_PSD_folder'] = 'data/wfemap/tweeter_opd_map/allopd/n{0}/'.format(n_set)
#rx_sys[j_tweeter]['surf_PSD_filename'] = 'dm_opd_s{0}_i4'.format(s_set)


# Run a test case stepping through the binning process to double check there are
# no pixel gaps over overlaps. This also precalculates the number of rows
# Dmat will have
ts = time.time()
summer = np.zeros_like(tmp_mask)
minipupil = np.zeros((50,50),dtype=np.float32)
bincounter = 0
for px in range(int(pupil_size_binningy / binWidth)):
    for py in range(int(pupil_size_binningx / binWidth)):
        if np.mod(bincounter,10) == 0:
            print('Current bin number is: ',bincounter)

        # load in the pupil mask Poppy looks for (2152x2152)
        tmp_pupil_Alex = np.zeros_like(pupil_mask)

        # To bin, we have to get more creative:
        # pull out 540x540 region in the center of the pupil region
        tmp_pupil = np.zeros_like(tmp_mask)
        # set the bin pixels equal to the normalized pupil value
        tmp_pupil[px*binWidth:(px+1)*binWidth,py*binWidth:(py+1)*binWidth] = 1.0
        # Make sure pixels are within active area
        tmp_pupil *= tmp_mask.copy()
        minipupil[px,py] = np.sum(tmp_pupil)
        if np.sum(tmp_pupil) >= 1.0:
            summer+=tmp_pupil.copy()
            bincounter+=1

minipupil /= minipupil.max()
Dmat = np.zeros((bincounter,int(FP_crop*FP_crop)),dtype=np.complex64)


# Build the MagAO-X System and double check things... This will propagate
```

```
# through the Fresnel model to verify the surfaces. Can be skipped if you trust
# your setup of Poppy and Jhen's code

# build the FresnelOpticalSystem
#magaox = mf.csvFresnel(rx_csv=rx_sys,
#                        samp=fr_parm['npix'],
#                        oversamp=fr_parm['beam_ratio'],
#                        home_folder=home_dir,
#                        break_plane='F69Sci',
#                        bump=fr_parm['bump'])
#
#sci_psf, sci_wf = magaox.calc_psf(wavelength=fr_parm['wavelength'].value, return_final = True)
#
#psf_crop = poppy.utils.pad_or_crop_to_shape(sci_psf[0].data, (67,67))
#plt.figure(dpi=100)
#plt.imshow(np.log10(psf_crop), origin='lower')
#plt.colorbar()

# Test the binning process and it's interactions with Poppy. Can skip

#for p in range(numPoints):
## load in the pupil mask Poppy looks for (2152x2152)
#    # If we want to scan 1 pixel at a time, just jump through listx,listy
#    tmp_pupil_Alex = np.zeros_like(pupil_mask)
#    tmp_pupil_Alex[listx[p],listy[p]] = pup_norm_val
#    # Write the pupil mask to a fits file for Poppy to use
#    fhdr = fits.Header()
#    fhdr.set('puplscal', 0.012081784386617101,
#                  'pupil scale m/pix')
#    fits.writeto(home_dir+'singlepixpup.fits', tmp_pupil_Alex, fhdr, overwrite=True)
#
#
#    # DO THIS TO EDIT PUPIL TO A SINGLE PIXEL
#    # For each process, make a copy of the csv file, and change this fileneame
#    # in the loop, rewrite this fits file for each pixel
#    rx_sys[0]['surf_PSD_folder'] = ''
#    rx_sys[0]['surf_PSD_filename'] = 'singlepixpup'
#
#
#    # build the FresnelOpticalSystem
#    magaox = mf.csvFresnel(rx_csv=rx_sys,
#                            samp=fr_parm['npix'],
#                            oversamp=fr_parm['beam_ratio'],
#                            home_folder=home_dir,
#                            break_plane='F69Sci',
```

```
#                             bump=fr_parm['bump'])
#
#     sci_psf, sci_wf = magaox.calc_psf(wavelength=fr_parm['wavelength'].value, return_final = True)
#     wf_crop = poppy.utils.pad_or_crop_to_shape(sci_wf[0].wavefront, (FP_crop, FP_crop))
#     Dmat[p,:] = wf_crop.flatten()
#print('Time to run code is: ', time.time()-ts, ' seconds')


bincounter = 0
summer = np.zeros_like(tmp_mask)
for px in range(int(pupil_size_binningy / binWidth)):
    for py in range(int(pupil_size_binningx / binWidth)):
        if np.mod(bincounter,10) == 0:
            print('Current bin number is: ',bincounter)

        # load in the pupil mask Poppy looks for (2152x2152)
        tmp_pupil_Alex = np.zeros_like(pupil_mask)

        # To bin, we have to get more creative:
        # pull out 540x540 region in the center of the pupil region
        tmp_pupil = np.zeros_like(tmp_mask)
        # set the bin pixels equal to the normalized pupil value
        tmp_pupil[px*binWidth:(px+1)*binWidth,py*binWidth:(py+1)*binWidth] = 1.0
        # Make sure pixels are within active area
        tmp_pupil *= tmp_mask.copy()

        if np.sum(tmp_pupil) >= 1.0:
            summer+=tmp_pupil.copy()

            # place this back in to the full size pupil
#             tmp_pupil_Alex[(1076-270):(1076+270),(1076-270):(1076+270)] = tmp_pupil
            tmp_pupil_Alex[(1076-275):(1076+275),(1076-275):(1076+275)] = tmp_pupil
#             tmp_pupil_Alex[(1076-270):(1076+270),(1076-275):(1076+275)] = tmp_pupil

            # Write the pupil mask to a fits file for Poppy to use
            fhdr = fits.Header()
            fhdr.set('puplscal', 0.012081784386617101,
                             'pupil scale m/pix')
            fits.writeto(home_dir+'singlepixpup.fits', tmp_pupil_Alex, fhdr, overwrite=True)


            # DO THIS TO EDIT PUPIL TO A SINGLE PIXEL
            # For each process, make a copy of the csv file, and change this fileneame
            # in the loop, rewrite this fits file for each pixel
            rx_sys[0]['surf_PSD_folder'] = ''
```

```
            rx_sys[0]['surf_PSD_filename'] = 'singlepixpup'


            # build the FresnelOpticalSystem
            magaox = mf.csvFresnel(rx_csv=rx_sys,
                                   samp=fr_parm['npix'],
                                   oversamp=fr_parm['beam_ratio'],
                                   home_folder=home_dir,
                                   break_plane='F69Sci',
                                   bump=fr_parm['bump'])

            # Calculate the PSF for the given input bin
            sci_psf, sci_wf = magaox.calc_psf(wavelength=fr_parm['wavelength'].value, return_final = True)

            # Crop this to the size detector we want (no change in resolution, only cropping)
            wf_crop = poppy.utils.pad_or_crop_to_shape(sci_wf[0].wavefront, (FP_crop, FP_crop))

            # Store in the Propagation Matrix
            Dmat[bincounter,:] = wf_crop.flatten()
            bincounter+=1

print('Time to run code is: ', time.time()-ts, ' seconds')
#Dmat *= (1 / pup_norm_val)

# Transpose so it is in the right shape
Dmat = Dmat.T

# Save the Matrix
np.save('MagAOX_labdemo_4096x2040_11x11binreduction_Dmat.npy',Dmat)
```

Listing C.3: Python code for doing reconstructions of MagAO-X WFS intensity into OPD measurements

```

#!/usr/bin/env python3
# -*- coding: utf-8 -*-

recon_dir = 'path/to/reconstruction/files/'

# Load files into memory
zrespM = np.float32(PPU.simple_fitsread(recon_dir + 'zrespM.fits'))
wfsref0 = np.float32(PPU.simple_fitsread(recon_dir + 'wfsref0.fits'))
wfsdark = np.float32(PPU.simple_fitsread(recon_dir + 'wfsdark_2022-02-04_02:32:57.fits'))

# Initialize an empty array for making the response matrix
```

```
M = np.zeros((zrespM.shape[0], zrespM.shape[1]*zrespM.shape[2]),np.float32)

# Take the data of the measured responses and organize into a matrix
for n in range(zrespM.shape[0]):
    M[n,:] = zrespM[n,:,:].flatten()

# Transpose the constructed matrix to have the correct shape
M = M.T

# Compute the SVD efficiently
M_tmp = M.T.dot(M)
u,s,vh = np.linalg.svd(M_tmp)

# Truncate at the number of illuminated actuators
cutoff = 1600
s_inv = np.zeros_like(s)
for n in range(len(s)):
    if n < cutoff:
        s_inv[n] = 1/(s[n])
    else:
        s_inv[n] = 0
S = np.diag(s_inv)

# Construct the Truncated Inverse to get the reconstructor matrix
M_inv = (vh.T.dot(S).dot(u.T)).dot(M.T)


# For a loaded measurement of WFS intensity called wfs_image:

# Get the frame of WFS Intensity
wfs_image = xp.array(MagAOX.WFSInt.copy()) # A loaded camwfs intensity file

# Prepare the Intensity for reconstruction from CACAO
wfs_image -= MagAOX.wfsdark
wfs_image /= wfs_image.sum()
ref_sub = wfs_image - MagAOX.wfsref0

# Reconstruct the OPD using the Reconstructor
reconst_out = MagAOX.WFS_Reconstructor.dot(ref_sub.flatten().reshape(14400,1)).reshape(50,50)

# Convert to phase
phase_map = ((2*np.pi) / (WFS_wvl)) * reconst_out * 1e-6

# Pull out pixel values within the pupil
phase_vec = phase_map[pup_pixy,pup_pixx].reshape(MagAOX.D.shape[1],1) * MagAOX.PupilMask
```

Continued

Listing C.4: Python code for taking data on MagAO-X for the improved experimental setup

```python
%reload_ext autoreload
%autoreload 2
%matplotlib inline
import datetime

#load modules
import numpy as np
import matplotlib.pyplot as plt
from astropy import units as u
from astropy.io import fits

# accessing the cameras
from magpyx.utils import ImageStream

import os
import struct
import pickle



def colorbar(mappable):
    from mpl_toolkits.axes_grid1 import make_axes_locatable
    import matplotlib.pyplot as plt
    last_axes = plt.gca()
    ax = mappable.axes
    fig = ax.figure
    divider = make_axes_locatable(ax)
    cax = divider.append_axes("right", size="5%", pad=0.05)
    cbar = fig.colorbar(mappable, cax=cax)
    plt.sca(last_axes)
    return cbar




# initialize the cameras
cam_wfs = ImageStream('camwfs')
cam_sci1 = ImageStream('camsci1')



# set the camera semaphore
```

```
if cam_wfs.semindex is None:
    cam_wfs.semindex = cam_wfs.getsemwaitindex(1)



# set the camera semaphore
if cam_sci1.semindex is None:
    cam_sci1.semindex = cam_sci1.getsemwaitindex(1)




# camera info (Update with real exposure times)
camsci1_Exptime = np.float64(0.03125)
camwfs_Exptime = np.float64(0.001000)






# initialize the woofer DM in channel 2 (not where the flat is loaded)
dm = ImageStream('dm00disp02') # channel 2
cmd = np.zeros_like(dm.buffer)
n,m = cmd.shape



# Flatten the DM
dm.write(np.zeros((n,m)).astype(dm.buffer.dtype))
sleep(0.5)

# Loop parameters
nimages_wfs = 32
nimages_camsci = 1
nrepeats = 5
nexposures = 250

flat = np.zeros((n,m)).astype(dm.buffer.dtype)
cam_wfs.semflush(cam_wfs.semindex)
cam_sci1.semflush(cam_sci1.semindex)

for exposure in range(nexposures):
    #initialize lists for data
    wfs_images = []
```

```
    camsci_images = []
    print(exposure)
    # Loop over number of repeats
    for im in range(nrepeats):
        #print(im)
        # command DM to a gaussian random shape with 0.7 variance, 0 mean
        dm_shape = (np.sqrt(0.7)*np.random.randn(n,m))*0.1
        dm_shape -= dm_shape.mean()
        dm_shape = dm_shape.astype(dm.buffer.dtype)

        # Take WFS intensity images
        cam_wfs.semflush(cam_wfs.semindex)
        dm.write(flat.astype(dm.buffer.dtype))
        sleep(0.05)
        cam_wfs.semwait(cam_wfs.semindex)
        wfs_images.append(cam_wfs.grab_many(nimages_wfs))

        # Take science camera intensity
        cam_sci1.semflush(cam_sci1.semindex)
        cam_sci1.semwait(cam_sci1.semindex)
        camsci_images.append(cam_sci1.grab_many(nimages_camsci))

        # Return DM to flat
        sleep(0.05)

    dm.write(np.zeros((n,m)).astype(dm.buffer.dtype))

    # Process taken data into numpy arrays and sum over nrepeats dim
    wfs_images = np.asarray(wfs_images)
    wfs_images_ = wfs_images.sum(0)
    camsci_images = np.asarray(camsci_images)
    camsci_images_ = camsci_images.sum(0)

    # Convert images into a single intensity frame
    camwfs_output = wfs_images_.sum(0) / wfs_images_.shape[0]
    camsci_output = camsci_images_.sum(0) / camsci_images_.shape[0]


    # Write the data to FITS
    now = datetime.datetime.now().isoformat()

    hdr = fits.Header()
    hdr.set('camera', 'wfs', "camera taking data")
    hdr.set('nframe', exposure, "frame number")
    hdr.set('exptime', camwfs_Exptime, "camera exposure time in sec")
```

```
    fits_filename = 'AlexData/no_ab/Camwfs/Camwfs_'+now[0:19]+'.fits'
    fits.writeto(fits_filename, camwfs_output, overwrite=True)


    hdr = fits.Header()
    hdr.set('camera', 'camsci', "camera taking data")
    hdr.set('nframe', exposure, "frame number")
    hdr.set('exptime', camsci1_Exptime, "camera exposure time in sec")
    fits_filename = 'AlexData/no_ab/Camsci/Camsci_'+now[0:19]+'.fits'
    fits.writeto(fits_filename, camsci_output, overwrite=True)

    sleep(0.1)

    # Clean up a bit
    #del(camwfs_output)
    #del(camsci_output)
    del(wfs_images_)
    del(wfs_images)
    del(camsci_images_)
    del(camsci_images)


# Close cameras and DM
cam_wfs.close()
cam_sci1.close()
dm.close()
```

Listing C.5: Code for running Frazin's Algorithm using WFS and science camera data from MagAO-X

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Tue Feb 15 15:13:36 2022

@author: archdaemon
"""



# Set up the path for running script on remote computer (update as needed)
import sys
sys.path.insert(0,'/home/atrodack/Research/Frazin-algo-Py/lib')
sys.path.insert(0,'/home/atrodack/Research/Optics')
sys.path.insert(0,'/home/atr0dack/Research/Github/Frazin-algo-Py/lib')
```

```
sys.path.insert(0,'/home/atr0dack/Research/Github/Frazin-algo-Py/Examples')
sys.path.insert(0,'/home/atr0dack/Research/Github/Optics')

# Imports
import numpy as np
import cupy as cp
import pickle

from tqdm import tqdm
import time
import datetime
import matplotlib.pyplot as plt
from astropy.io import fits
from scipy.interpolate import LSQBivariateSpline as SPL
from matplotlib.ticker import (MultipleLocator, FormatStrFormatter, AutoMinorLocator)

# Code Base Imports
import AlgorithmSim_dissertation as ALGOSIM
from PyPropUtils import PyPropUtils as PPU

# Define new colorbar code
def colorbar(mappable):
    from mpl_toolkits.axes_grid1 import make_axes_locatable
    import matplotlib.pyplot as plt
    last_axes = plt.gca()
    ax = mappable.axes
    fig = ax.figure
    divider = make_axes_locatable(ax)
    cax = divider.append_axes("right", size="5%", pad=0.05)
    cbar = fig.colorbar(mappable, cax=cax)
    plt.sca(last_axes)
    return cbar


# Plot marker size and font size
markerSz = 28
plt.rcParams.update({'font.size': 28})


# Open Text File for logging
file1 = open("Benchmark_Results_labdata.txt","a+")
file1.write("**************************************************************\n")
now = datetime.datetime.now().isoformat()
file1.write("Benchmark Date/Time: " + now + "\n" )

```

```
## Print number of GPUs
#nGPUs = cp.cuda.runtime.getDeviceCount()
#print('The number of Detected GPUs is: ', nGPUs)
#
## Choose GPU (MagAOX needs to be removed from memory to change GPUs)
#if nGPUs == 1:
#    GPU0 = cp.cuda.Device(0)
#elif nGPUs > 1:
#    GPU0 = cp.cuda.Device(0)
#    PPU.DumpInfo(GPU0)
#    print('\n\n')
#    GPU1 = cp.cuda.Device(1)
#    PPU.DumpInfo(GPU1)
#
#GPU2Use = 0
#if GPU2Use == 0:
#    GPU0 = cp.cuda.Device(0)
#    GPU0.use()
#elif GPU2Use == 1:
#    GPU1 = cp.cuda.Device(1)
#    GPU1.use()
#elif GPU2Use == 2:
#    GPU2 = cp.cuda.Device(2)
#    GPU2.use()
#elif GPU2Use == 3:
#    GPU3 = cp.cuda.Device(3)
#    GPU3.use()
#print('\nGPU ', str(GPU2Use), ' is selected')
#



RON = 1

# MagAO-X Model
Dfile = '/home/archdaemon/Research/GitHub/Frazin-algo-Py/lib/MagAOX_labdemo_4096x2040_11x11binreduction_Dmat.npy'
pupilmask = np.float32(PPU.simple_fitsread('/home/archdaemon/Research/GitHub/Frazin-algo-Py/Examples/MagAOX_pupil_labdata.fits'))
mask_allpix = np.float64(PPU.simple_fitsread('/home/archdaemon/Research/GitHub/Frazin-algo-Py/Examples/MagAO-X_pupil_labdata_allpix.fits'))
pup_pixy,pup_pixx = np.where(mask_allpix==1)

WFS_wvl = np.float64(850 * 1e-9) # 40nm bandpass
WFS_Exptime = np.float64(0.0005) # seconds
```

```
camsci1_wvl = np.float64(668 * 1e-9) # Ha-cont
camsci1_Exptime = np.float64(0.025000) #seconds

# Initialize Simulation class object
MagAOX = ALGOSIM.WorkingCoronagraphModels(DmatFilename=Dfile, RON=RON)

# Get array module for CPU/GPU calculations
xp = cp.get_array_module(MagAOX.gpu_ary)

MagAOX_pup = pupilmask[pup_pixy,pup_pixx].reshape(MagAOX.D.shape[1],1)
MagAOX_pup[MagAOX_pup!=1] = 0
PupilMask = MagAOX_pup.copy()
# PupilMask = 1


MagAOX.PupilMask = xp.array(PupilMask)
PupilMask = xp.array(PupilMask)



# Set some important simulation parameters outside of object
blurb = 'BSpline_NaiveLabDemo_'



# Set Flags for Experiment Regression Methods
Ideal = False # Flag for running Ideal and Ideal-noise estimates
Naive = True # Flag for running Naive estimate
CSN = False # Flag for allowing Monte Carlo loop, and running SN and CSN estimates
UseWeightingMat = True
scaleElements = False


# Set switches in code to False
MagAOX.compute_xb_flag = False
MagAOX.filterWFs = False
MagAOX.FilterFileLoaded = False



textlines = ["\nIdeal Flag: " + str(Ideal), "\nNaive Flag: " + str(Naive), "\nCSN Flag: " + str(CSN)]
file1.writelines(textlines)
```

```
# Set flags for estimate inclusion
NCPA=True # Flag for including NCPA in the joint estimator
exoplanet=False # Flag for including exoplanet image in the joint estimator
noNCPA = False # Flag for also returning Intensity computed without the NCPA [Probably not needed anymore]

textlines = ["\nNCPA Flag: " + str(NCPA), "\nexoplanet Flag: " + str(exoplanet), "\nnoNCPA Flag: " + str(
    noNCPA)]
file1.writelines(textlines)

# Choose Atmospheric Time Series
# MagAOX.params['WFGtype'] = 'useWhiteNoiseAtmo'
MagAOX.params['WFGtype'] = 'RealData'
MagAOX.params['camsci_dir'] = '/usr/local/Lab_Experiments/Camsci1_lc_astigv_turb_organized/'
MagAOX.params['camwfs_dir'] = '/usr/local/Lab_Experiments/Camwfs_lc_astigv_turb_organized/'
MagAOX.params['recon_dir'] = '/home/archdaemon/Research/Lab_Experiments/Friday_Feb_4_2022/
    Reconstruction_Files/'
textlines = ["\nWFGtype: "+(MagAOX.params['WFGtype']), "\ncamsci_dir: "+MagAOX.params['camsci_dir'], "\
    ncamwfs_dir: "+MagAOX.params['camwfs_dir'], "\nrecon_dir: "+MagAOX.params['recon_dir']+"\n\n"]
file1.writelines(textlines)

# Set up NCPA Regression
#MagAOX.params['BasisType'] = 'BSpline'
#MagAOX.params['BasisModes'] = 622
#MagAOX.params['TList'] = np.arange(2,MagAOX.params['BasisModes']).astype(np.int32)

MagAOX.params['BasisType'] = 'Zernike'
MagAOX.params['ZList'] = np.array([2,3,5,6,7,8,9,10,12,13,14,15,16])
MagAOX.params['TList'] = np.array([2,3,5,6,7,8,9,10,12,13,14,15,16])
MagAOX.params['BasisModes'] = 13


# List for Regression Basis functions
MagAOX.params['abco'] = np.zeros((MagAOX.params['BasisModes'],1),dtype=MagAOX.params['dataType'])



print('Initializing Simulation')
amp = xp.array(np.array(1.).astype(MagAOX.params['dataType']))
Offpoint = xp.zeros(shape=(MagAOX.K,1),dtype=MagAOX.params['dataType'])

x_a0 = xp.zeros((MagAOX.params['BasisModes'],1),MagAOX.params['dataType'])

#MagAOX.init_sim(alphas_true=alphas_true, alphas=alphas, contrast=contrast, x_a0 = x_a0)
#MagAOX.init_sim_v2(NCPA, exoplanet, alphas_true=alphas_true, alphas=alphas, contrast=contrast, x_a0=x_a0)
MagAOX.init_real_data(NCPA, exoplanet, x_a0 , wfsdark_filename='wfsdark_2022-02-04_02:32:57.fits')
```

```
Nexposures = len(MagAOX.camwfs_files)
Nexposures = 250
nexposures = (np.array([Nexposures,Nexposures,Nexposures])).astype('int')
nLoop = len(nexposures) - 1


#print('Initializing the Exoplanet Phasors')
#MagAOX.init_exoImage_phasors()
#MagAOX.init_exoImageAlgoPhasors()


print('Starting Simulation at ' + datetime.datetime.now().strftime("%b %d, %Y at %H:%M:%S") +'!\n\n')


# Set up time extension
ms_per_file = 250 # ms time steps in each Atmo file
files_per_timestep = np.int(nexposures[0] / ms_per_file) # how many files are in each timestep
AtmoTstep_total = 0
with fits.open(MagAOX.params['camsci_dir']+MagAOX.camsci_files[MagAOX.camscifilenumber]) as hdul_sc:
    sc_frames = (hdul_sc[0].data).astype(xp.float32)
    # Update the counters
    MagAOX.camscifilenumber += 1
    MagAOX.camsciframenumber = 0
num_sci_frames = sc_frames.shape[0]


# Relinearization Loop
for n in range(len(nexposures)):
    print("\n***************************************\n")
    print("Starting Relinearization Step "+str(n+1)+" of "+str(len(nexposures))+"\n\n")
    # print(tmp_filenumber)

    # Initialize the On-Sky Loop
    MagAOX.init_OnSky(NCPA=NCPA, exoplanet=exoplanet, Ideal=Ideal, Naive=Naive, CSN=CSN, scaleElements=
    scaleElements)
    MagAOX.atmoFlag=None # Makes Object reseed / reload phasescreens
    MagAOX.atmoTstep = 0 # Reset steps through phasescreen for tracking t > T_AO_start
    T = nexposures[n].copy() # Set On-Sky loop for loop range

    file1.write("\nStarting On-Sky Sim for " + str(T) + " Exopsures!\n")
    ts = time.time()
    print('\nStarting On-Sky Loop')

    # On Sky Data Loop
```

```
    for t in tqdm(range(T)):
#         if np.mod(t,25)  == 0:
#             print('Exposure number: ', t)



        # Do the regression step
        # Load the WFS camera data fits file for the current science camera frame
        MagAOX.WFG()        # This puts the current PyWFS intensity frames into MagAOX.WFSInt

        # Loop over each frame computing c and A_a
        num_WFS_exposures = MagAOX.WFSInt.shape[0]
        A_a_t = xp.zeros((MagAOX.L, MagAOX.N),xp.float32)
        c_t = xp.zeros((MagAOX.L,1),xp.float32)
        for wfs_exp in range(num_WFS_exposures):
            # Get the frame of WFS Intensity
            wfs_image = xp.array(MagAOX.WFSInt[wfs_exp,:,:].copy())
            # Prepare the Intensity for reconstruction from CACAO
            wfs_image -= MagAOX.wfsdark
            #wfs_image *= wfsmask # this may not be needed
            wfs_image /= wfs_image.sum()
            ref_sub = wfs_image - MagAOX.wfsref0
            # Reconstruct the OPD using the Reconstructor
            reconst_out = MagAOX.WFS_Reconstructor.dot(ref_sub.flatten().reshape(14400,1)).reshape(50,50)
            # Convert to phase
            phase_map = ((2*np.pi) / (WFS_wvl)) * reconst_out * 1e-6
            # Pull out pixel values within the pupil
            phase_vec = phase_map[pup_pixy,pup_pixx].reshape(MagAOX.D.shape[1],1) * MagAOX.PupilMask

            # Make the phasor for the algorithm
            MagAOX.naive_phasor = MagAOX.PupilMask * xp.exp(xp.multiply(1j,phase_vec + MagAOX.phi_u_estim)).astype(xp.complex64)
            # Init variables for the calculation of A_a and c
            u_s_m_ideal = None
            u_s_m_naive = (xp.multiply(amp,MagAOX.naive_phasor.copy()))

            # compute A_a and c for the naive estimate
            A_t, u_t, A_tprime_naive, u_tprime_naive = MagAOX.compute_A_and_ut_v3(g1=u_s_m_ideal, g2=u_s_m_naive, NCPA=NCPA, exoplanet=exoplanet, Ideal=Ideal, Naive=Naive, CSN=CSN, scaleElements=scaleElements)

            # Add them to the sum to approximate the longer exposure in the science camera
            A_a_t += A_tprime_naive
            c_t += u_tprime_naive

```

```
        A_a_t /= num_WFS_exposures
        c_t /= num_WFS_exposures
        # Get the corresponding Science Camera Frame        MagAOX.camscifilenumber is the file, MagAOX.camsciframenumber is the frame
        # Load a new fits file if we need to
        if xp.mod(t,num_sci_frames) == 0 and t !=0:
#            print('Opening new SC file\n')
            with fits.open(MagAOX.params['camsci_dir']+MagAOX.camsci_files[MagAOX.camscifilenumber]) as hdul_sc:
                sc_frames = (hdul_sc[0].data.astype(xp.float32))
                # Update the counters
                MagAOX.camscifilenumber += 1
                MagAOX.camsciframenumber = 0
                num_sci_frames = sc_frames.shape[0]


        # Store the science camera measurement for this time step
        y = xp.array(sc_frames[MagAOX.camsciframenumber,:,:].reshape(MagAOX.L,1).copy())
        # Advance the frame number counter
        MagAOX.camsciframenumber += 1
        # Estimate the noise covariance matrix
        C_y_est = xp.diag(y.reshape(MagAOX.L,)).astype(xp.float32)


        # Update the Algorithm Quantities
    #        tt = time.time()
        if UseWeightingMat is False:
            MagAOX.update_OnSky_Naive(A_a_t, c_t, y, C_y_est, NCPA=NCPA, exoplanet=exoplanet)
        else:
            MagAOX.update_OnSky_Naive_S(A_a_t, c_t, y, C_y_est, NCPA=NCPA, exoplanet=exoplanet)


        # Advance the counter to load the next file next step through the for loop
        MagAOX.WFSfilenumber += 1

        # count the total exposures done
        if n==nLoop:
            AtmoTstep_total+=1
    # End of On-Sky for loop

    tf = time.time()
    print('On-Sky Loop Time: %.20fs' % (tf - ts)) # This includes the time for loading in screens printed to the workspace

    exposurespersecond = (T / (tf-ts))
```

```
    file1.write("On-Sky Loop Time: "+str(tf-ts)+"\n")
    file1.write("On-Sky Loop Exposures per second: " + str(exposurespersecond) + "\n\n")

    # Compute Ideal / Naive Estimates
    MagAOX.computeEstimates_OnSky(NCPA=NCPA, exoplanet=exoplanet, Ideal=Ideal, Naive=Naive, CSN=CSN, scaleElements=scaleElements)
    # Do relinearization if required
    if n<nLoop:
        if (Ideal is True and CSN is False):
            print('Relinearizing using Ideal\n')
            x_a0 = cp.asnumpy(x_a0)
            x_a0 += MagAOX.x_ideal_noise_NCPA[0:MagAOX.N].copy()
            for linenum in range(len(x_a0)):
                if linenum == 0:
                    file1.write("x_a0 = xp.array(["+str(x_a0[linenum])+",\n")
                elif linenum == len(x_a0)-1:
                    file1.write("\t"+str(x_a0[linenum])+"]).astype(MagAOX.params['dataType'])")
                else:
                    file1.write("\t"+str(x_a0[linenum])+",\n")
            MagAOX.computeLinearizationPoint((xp.array(x_a0[0:MagAOX.N_a])))

        if( (Naive is True) and (Ideal is False) and (CSN is False)):
            print('Relinearizing using Naive\n')
            x_a0 = cp.asnumpy(x_a0)
            x_a0 += MagAOX.x_naive_NCPA[0:MagAOX.N].copy()
            for linenum in range(len(x_a0)):
                    if linenum == 0:
                        file1.write("x_a0 = xp.array(["+str(x_a0[linenum])+",\n")
                    elif linenum == len(x_a0)-1:
                        file1.write("\t"+str(x_a0[linenum])+"]).astype(MagAOX.params['dataType'])")
                    else:
                        file1.write("\t"+str(x_a0[linenum])+",\n")
            MagAOX.computeLinearizationPoint((xp.array(x_a0[0:MagAOX.N_a])))
            MagAOX.WFSfilenumber = 0
            MagAOX.camscifilenumber = 0
            with fits.open(MagAOX.params['camsci_dir']+MagAOX.camsci_files[MagAOX.camscifilenumber]) as hdul_sc:
                sc_frames = (hdul_sc[0].data).astype(xp.float32)
                # Update the counters
                MagAOX.camscifilenumber += 1
                MagAOX.camsciframenumber = 0
            num_sci_frames = sc_frames.shape[0]
        if (CSN is True):
            print('Relinearizing using CSN\n')
            x_a0 = cp.asnumpy(x_a0)
```

```
            x_a0 += MagAOX.x_csn_mc[0:MagAOX.N].copy()
            for linenum in range(len(x_a0)):
                    if linenum == 0:
                        file1.write("x_a0 = xp.array(["+str(x_a0[linenum])+",\n")
                    elif linenum == len(x_a0)-1:
                        file1.write("\t"+str(x_a0[linenum])+"]).astype(MagAOX.params['dataType'])")
                    else:
                        file1.write("\t"+str(x_a0[linenum])+",\n")
            MagAOX.computeLinearizationPoint((xp.array(x_a0[0:MagAOX.N_a])))
            # meandifftotrue = np.mean(np.abs(x_a0.copy() - MagAOX.params['abco']))
            # file1.write("\n"+"Mean absolute difference from true linearization point: "+str(meandifftotrue))

        if xp.__name__ == 'cupy':
            MagAOX.x_a0 = cp.array(x_a0)
            MagAOX.GPUify_OnSky(Ideal=Ideal, Naive=Naive, CSN=CSN, scaleElements=scaleElements, NCPAonly=NCPA)
            MagAOX.GPUify_MC(scaleElements=scaleElements)

        # Reset Atmo to use same files
#        MagAOX.Atmofilenumber = tmp_filenumber
        MagAOX.atmoFlag = None
    elif n==nLoop:
        print('Final Step...Need to add x_a0 to estimates!\n')


    # Store some stuff in the Object for saving later
    # MagAOX.x_a0 = x_a0.copy()




x_a0 = cp.asnumpy(x_a0)

if Ideal is True:
     MagAOX.x_ideal = MagAOX.x_ideal_NCPA.copy()
     MagAOX.x_ideal_noise = MagAOX.x_ideal_noise_NCPA.copy()
     MagAOX.C_ideal_noise = MagAOX.C_ideal_noise_NCPA.copy()
     MagAOX.P_true = MagAOX.P_true_NCPA
     MagAOX.Q_true = MagAOX.Q_true_NCPA
     MagAOX.V_true = MagAOX.V_true_NCPA
     MagAOX.V_ideal = MagAOX.V_ideal_NCPA
     MagAOX.H_ideal = MagAOX.H_ideal_NCPA
     MagAOX.x_true = MagAOX.x_true[0:MagAOX.N_a]

```

```
if Naive is True:
    MagAOX.x_naive = MagAOX.x_naive_NCPA.copy()
    MagAOX.H_naive = MagAOX.H_naive_NCPA.copy()
    MagAOX.V_naive = MagAOX.V_naive_NCPA.copy()
    MagAOX.C_naive = MagAOX.C_naive_NCPA.copy()
    MagAOX.Q_naive = MagAOX.Q_naive_NCPA.copy()
    MagAOX.P_naive = MagAOX.P_naive_NCPA.copy()
#     MagAOX.x_naive = MagAOX.x_naive_planet.copy()
#     MagAOX.C_naive = MagAOX.C_naive_planet.copy()
#     MagAOX.H_naive = MagAOX.H_naive_planet.copy()
#     MagAOX.P_naive = MagAOX.P_naive_planet.copy()
#     MagAOX.Q_naive = MagAOX.Q_naive_planet.copy()
#     MagAOX.H_naive = MagAOX.H_naive_planet.copy()
#     MagAOX.V_naive = MagAOX.V_naive_planet.copy()

if CSN is True and Naive is False:
    MagAOX.H_naive = MagAOX.H_naive_NCPA.copy()
    MagAOX.V_naive = MagAOX.V_naive_NCPA.copy()
    # MagAOX.H_naive = MagAOX.H_naive_planet.copy()
    # MagAOX.V_naive = MagAOX.V_naive_planet.copy()

# MagAOX.x_true = MagAOX.x_true[0:MagAOX.N_a]


if nLoop == n:
    # Add the linearization point to the estimates from the last step
    if Ideal is True:
        MagAOX.x_ideal += x_a0
        MagAOX.x_ideal_noise += x_a0

    if Naive is True:
        MagAOX.x_naive += x_a0

    if CSN is True:
        MagAOX.x_csn_mc += x_a0
        MagAOX.x_sn += x_a0

    if MagAOX.compute_xb_flag is True:
        MagAOX.x_b += x_a0
        MagAOX.x_c += x_a0



# Compute the error bar values for each estimate and store
if Ideal is True:
```

```
    var_ideal_noise = np.diag(MagAOX.C_ideal_noise)
    MagAOX.sigma_in = 1.*np.sqrt(var_ideal_noise)

if Naive is True:
    var_naive = np.diag(MagAOX.C_naive)
    MagAOX.sigma_naive =1. * np.sqrt(var_naive)

if CSN is True:
    var_csn = np.diag(MagAOX.C_csn)
    MagAOX.sigma_csn = 1.*np.sqrt(np.abs(var_csn))


# Close file
file1.write("\n\n\n")
file1.close()


estimate = cp.asnumpy(MagAOX.T).dot(MagAOX.x_naive) * cp.asnumpy(MagAOX.PupilMask)
estimated_intensity = cp.asnumpy(A_a_t).dot(MagAOX.x_naive) + cp.asnumpy(c_t)
MagAOX.plotPP(xp.array(estimate) * MagAOX.PupilMask * WFS_wvl / 2 / np.pi * 1e6, colormap='plasma'); plt.
    title('Estimated NCPA')
plt.figure();cbax = plt.imshow((estimated_intensity).reshape(64,64), cmap='plasma'); colorbar(cbax); plt.
    title('Estimated Intensity')
plt.figure(); cbax = plt.imshow(cp.asnumpy(y).reshape(64,64), cmap='plasma'); colorbar(cbax); plt.title('
    Measured Intensity')

"""
Save the class object
"""

MagAOX.x_a0 = cp.asnumpy(x_a0)
MagAOX.phi_r_true = cp.asnumpy(phase_vec)
MagAOX.phi_r_measured = cp.asnumpy(phase_vec)
MagAOX.I_sim_LE = 0
MagAOX.I_SNR = 0
Nexposures = nexposures
Lsamples = np.array([0])
now_ = datetime.datetime.now()
now = now_.strftime("%Y-%b-%d_%H-%M-%S")
if Ideal is True:
    Idealstr = 'IT'
else:
    Idealstr = 'IF'

if Naive is True:
```

```
    Naivestr = 'NT'
else:
    Naivestr = 'NF'

if CSN is True:
    CSNstr = 'CSNT'
else:
    CSNstr = 'CSNF'
filename = 'LabDemo/Dict_T'+str(Nexposures[-1])+'_L'+str(Lsamples[-1])+'-'+str(len(Nexposures) - 1)+'x'+str
    (int(Nexposures[0]/1000))+'kRL_'+Idealstr+'_'+Naivestr+'_'+CSNstr+'-'+blurb+now

if Ideal is True:
    var_ideal_noise = np.diag(cp.asnumpy(MagAOX.C_ideal_noise))
    sigma_in = 1.*np.sqrt(var_ideal_noise)

if Naive is True:
    var_naive = np.diag(cp.asnumpy(MagAOX.C_naive))
    sigma_naive =1. * np.sqrt(var_naive)

if CSN is True:
    var_csn = np.diag(cp.asnumpy(MagAOX.C_csn))
    sigma_csn = 1.*np.sqrt(np.abs(var_csn))

numpysavefile = dict()
MagAOX.params['Integrated_Off_Axis_PSF'] = cp.asnumpy(MagAOX.params['Integrated_Off_Axis_PSF'])
MagAOX.params['Integrated_Off_Axis_PSF_WFS'] = cp.asnumpy(MagAOX.params['Integrated_Off_Axis_PSF_WFS'])
numpysavefile['params'] = MagAOX.params


numpysavefile['x_true'] = cp.asnumpy(MagAOX.x_true)
numpysavefile['x_a0'] = cp.asnumpy(MagAOX.x_a0)

if Ideal is True:
    numpysavefile['x_ideal'] = cp.asnumpy(MagAOX.x_ideal)
    numpysavefile['x_ideal_noise'] = cp.asnumpy(MagAOX.x_ideal_noise)
    numpysavefile['C_ideal_noise'] = cp.asnumpy(MagAOX.C_ideal_noise)
    numpysavefile['sigma_in'] = cp.asnumpy(sigma_in)
    numpysavefile['P_true'] = cp.asnumpy(MagAOX.P_true)
    numpysavefile['Q_ideal'] = cp.asnumpy(MagAOX.Q_true)
    numpysavefile['V_true'] = cp.asnumpy(MagAOX.V_true)
    numpysavefile['V_ideal'] = cp.asnumpy(MagAOX.V_ideal)
    numpysavefile['H_ideal'] = cp.asnumpy(MagAOX.H_ideal)
```

```
if Naive is True:
    numpysavefile['x_naive'] = cp.asnumpy(MagAOX.x_naive)
    numpysavefile['C_naive'] = cp.asnumpy(MagAOX.C_naive)
    numpysavefile['sigma_naive'] = cp.asnumpy(sigma_naive)
    numpysavefile['P_naive'] = cp.asnumpy(MagAOX.P_naive)
    numpysavefile['Q_naive'] = cp.asnumpy(MagAOX.Q_naive)
    numpysavefile['V_naive'] = cp.asnumpy(MagAOX.V_naive)
    numpysavefile['H_naive'] = cp.asnumpy(MagAOX.H_naive)



if CSN is True:
    numpysavefile['x_sn'] = cp.asnumpy(MagAOX.x_sn)
    numpysavefile['x_csn_mc'] = cp.asnumpy(MagAOX.x_csn_mc)
    numpysavefile['C_csn'] = cp.asnumpy(MagAOX.C_csn)
    numpysavefile['sigma_csn'] = cp.asnumpy(sigma_csn)
    numpysavefile['P_mc'] = cp.asnumpy(MagAOX.P_mc)
    numpysavefile['g1_mc'] = cp.asnumpy(MagAOX.g1_mc)
    numpysavefile['g2_mc'] = cp.asnumpy(MagAOX.g2_mc)
    numpysavefile['G_mc'] = cp.asnumpy(MagAOX.G_mc)
    # numpysavefile['B_mc'] = cp.asnumpy(self.B_mc)
    # numpysavefile['b_mc'] = cp.asnumpy(self.b_mc)
    numpysavefile['Q_mc'] = cp.asnumpy(MagAOX.Q_mc)
    numpysavefile['V_naive'] = cp.asnumpy(MagAOX.V_naive)
    numpysavefile['H_naive'] = cp.asnumpy(MagAOX.H_naive)

if MagAOX.compute_xb_flag is True:
    numpysavefile['x_b'] = cp.asnumpy(MagAOX.x_b)
    numpysavefile['x_c'] = cp.asnumpy(MagAOX.x_c)
    numpysavefile['G'] = cp.asnumpy(MagAOX.G)
    numpysavefile['g1'] = cp.asnumpy(MagAOX.g1)
    numpysavefile['g2'] = cp.asnumpy(MagAOX.g2)


numpysavefile['BasisArray'] = cp.asnumpy(MagAOX.T)
numpysavefile['N_a'] = cp.asnumpy(MagAOX.N_a)
numpysavefile['N'] = cp.asnumpy(MagAOX.N)
numpysavefile['pltx_NCPA'] = np.arange(0,MagAOX.N_a)
numpysavefile['pltx_exoplanet'] = np.arange(0, MagAOX.N - MagAOX.N_a)


numpysavefile['phi_r_true'] = cp.asnumpy(MagAOX.phi_r_true)
numpysavefile['phi_r_measured'] = cp.asnumpy(MagAOX.phi_r_measured)
numpysavefile['phi_u_true'] = cp.asnumpy(MagAOX.phi_u_true)
numpysavefile['I_sim_LE'] = cp.asnumpy(MagAOX.I_sim_LE)
```

```
numpysavefile['I_SNR'] = cp.asnumpy(MagAOX.I_SNR)

numpysavefile['OnSky_exposures'] = cp.asnumpy(Nexposures)
numpysavefile['MC_samples'] = cp.asnumpy(Lsamples)



with open(filename,'wb') as prev_SLC_file:
    pickle.dump(numpysavefile, prev_SLC_file, protocol=4)


print('Save Complete!\n')
```

Listing C.6: Code for computing $\mathbf{A}(\mathbf{w}_t)$ and $\mathbf{c}(\mathbf{w}_t)$

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Mon Feb 28 13:18:46 2022

@author: archdaemon
"""

import numpy as np
import cupy as cp

# Set xp as np or cp depending on desire for GPU or CPU use
xp = cp


# D is the computational model of the science instrument, used to propogate the field from entrance pupil
    to science camera
# T is a matrix containing the modes of the estimation basis set, shaped to match the dimension of D
# g1 is a vector respresenting the ideal estimate representation of the field:
#       g1 = PupilMask * xp.exp(xp.multiply(1j,phi_r_true + phi_u_estim)).astype(xp.complex64)
# g2 is a vector representing the naive estimate representation of the field:
#       g2 = PupilMask * xp.exp(xp.multiply(1j,phi_r_measured + phi_u_estim)).astype(xp.complex64)
# NCPA and exoplanet are boolean flags that determine if their namesake is included in the regression
    framework
# Ideal, Naive, and CSN are boolean flags that determine if their namesake estimator is computed

# phi_u_estim is a vector representing the current estimate of the NCPA
#        phi_u_estim = T.dot(x_a0)
# phi_r_true is a vector of the true AO residual, in shape [P,1]
# phi_r_measured is a vector of the measured AO residual from the WFS, in shape [P,1]
# PupilMask is a vector of the pupil geometry, in shape [P,1]
```

```
# atmophasor_true = PupilMask * xp.exp(xp.multiply(1j,phi_r_true)).astype(xp.complex64)
# atmophasor_measured = PupilMask * xp.exp(xp.multiply(1j,phi_r_measured)).astype(xp.complex64)
# exophasors is an array of phasors corresponding to sky angles being estimated
# ideal_exophasor =  xp.multiply(atmophasor_true, exophasors).astype(xp.complex64)
# naive_exophasor =  xp.multiply(atmophasor_measured, exophasors).astype(xp.complex64)
# x_a0 is the linearization point
# N_a is the number of NCPA coefficients being estimated
# N_p is the number of Sky angle coeffiecients being estimated
# N is the total number of terms being estimated
# L is the number of science camera pixels
def compute_A_and_ct(D,T,g1,g2, NCPA=True, exoplanet=True, Ideal=False, Naive=False, CSN=True):

        if Ideal is True:
            # Compute u_t(w_t)
    #        tt = time.time()
            c_t_Ideal = D.dot(g1)
            c_t_Ideal_conj = c_t_Ideal.conj()
            c_tIdeal = xp.real(xp.multiply(c_t_Ideal, c_t_Ideal_conj))
            if xp.__name__ == 'cupy':
                cp.cuda.Stream.null.synchronize()

    #        print("Time to propagate g: % 5.8f seconds" %(time.time() - tt))
        else:
            c_tIdeal = None


        if(Naive is True or CSN is True):
            #        tt = time.time()
            c_t_Naive = D.dot(g2)
            c_t_Naive_conj = c_t_Naive.conj()
            c_tNaive = xp.real(xp.multiply(c_t_Naive, c_t_Naive_conj))
            if xp.__name__ == 'cupy':
                cp.cuda.Stream.null.synchronize()

    #        print("Time to propagate g: % 5.8f seconds" %(time.time() - tt))
        else:
            c_tNaive = None


        # Compute Star component of A
        if NCPA == True:
            if Ideal is True:
    #            tt = time.time()
                V1 = g1*T
    #            print("Time to make V: % 5.8f seconds" %(time.time() -tt))
```

```
    #            tt = time.time()
                tmp1 = D.dot(V1)
                tmp0 = xp.multiply(1j*tmp1, c_t_Ideal_conj)
                A_a_t_Ideal = xp.real(tmp0 + tmp0.conj())

                if xp.__name__ == 'cupy':
                    cp.cuda.Stream.null.synchronize()

    #            print("Time to compute A_a: % 5.8f seconds" %(time.time() - tt))

            if(Naive is True or CSN is True):
    #            tt = time.time()
                V2 = g2*T
    #            print("Time to make V: % 5.8f seconds" %(time.time() -tt))
    #            tt = time.time()
                tmp3 = D.dot(V2)
                tmp2 = xp.multiply(1j*tmp3, c_t_Naive_conj)
                A_a_t_Naive = xp.real(tmp2 + tmp2.conj())

                if xp.__name__ == 'cupy':
                    cp.cuda.Stream.null.synchronize()

    #            print("Time to compute A_a: % 5.8f seconds" %(time.time() - tt))
    #        print('A_a_t dtype: ',A_a_t.dtype)


        # Compute planet component of A
        if exoplanet == True:
#            tt = time.time()
            if Ideal is True:
                detField_planet1 = D.dot(ideal_exophasor)
                A_p_t_Ideal = xp.real(xp.multiply(detField_planet1, detField_planet1.conj())).astype(xp.float32)

                k0_tIdeal = xp.multiply(A_p_t_Ideal.copy(),x_a0[N_a:N].reshape(N_p,)).sum(1)
                c_tIdeal += k0_tIdeal.reshape(L,1)

            if(Naive is True or CSN is True):
                detField_planet2 = D.dot(naive_exophasor)
                A_p_t_Naive = xp.real(xp.multiply(detField_planet2, detField_planet2.conj())).astype(xp.float32)

                k0_tNaive = xp.multiply(A_p_t_Naive.copy(), x_a0[N_a:N].reshape(N_p,)).sum(1)
                c_tNaive += k0_tNaive.reshape(L,1)
#            print("Time to compute A_p: % 5.8f seconds" %(time.time() - tt))
```

```
        if xp.__name__ == 'cupy':
            cp.cuda.Stream.null.synchronize()

#         tt = time.time()
        if(NCPA==True and exoplanet==True):
            if Ideal is True:
                A_t_Ideal = xp.hstack((A_a_t_Ideal, A_p_t_Ideal))
            else:
                A_t_Ideal = None

            if(Naive is True or CSN is True):
                A_t_Naive = xp.hstack((A_a_t_Naive, A_p_t_Naive))
            else:
                A_t_Naive = None
        elif(NCPA==True and exoplanet==False):
            if Ideal is True:
                A_t_Ideal = A_a_t_Ideal
            else:
                A_t_Ideal = None

            if(Naive is True or CSN is True):
                A_t_Naive = A_a_t_Naive
            else:
                A_t_Naive = None

        elif(NCPA==False and exoplanet==True):
            if Ideal is True:
                A_t_Ideal = A_p_t_Ideal
            else:
                A_t_Ideal = None
            if(Naive is True or CSN is True):
                A_t_Naive = A_p_t_Naive
            else:
                A_t_Naive = None
#         print("Time to hstack: % 5.8f seconds" %(time.time() - tt))


        return A_t_Ideal, c_tIdeal, A_t_Naive, c_tNaive
    # End of compute__A_and_ut
```

Listing C.7: Code for computing the computationl propagation model of an optical system **D**

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
```

```python
"""
Created on Mon Feb 28 14:43:08 2022

@author: archdaemon
"""
import numy as np
import cupy as cp
from PyPropUtils import PyPropUtils as PPU


# PUPIL_zp is an array containing a zero padded version of the entrance pupil geometry
# params is a list containing system parameters. See code on GitHub for more details
#       Detx and Dety are vectors of integers containing pixel indices in the detector plane (likely 1-64)
# FPM is an array containing the geometry of a binary or gray scale amplitude focal plane mask
# Lyot is an array containing the geometry of a Lyot Stop

def construct_System_Operator_lyot_efficient(DataType=np.complex64):
    # Operator needs to be correctly normalized to produce a PSF with intensity max of 1
    sz_k = int(PUPIL_zp.shape[0])
    row, col = np.where(PUPIL_zp==1.)
#        norm = 1./(sz_k*np.sqrt(2.))
#        cnorm = np.complex(norm,norm)

    D = np.zeros(shape=(row.size, params['Detx'].size * params['Dety'].size),dtype=DataType)
    for k in range(row.size):
        e_k = np.zeros(shape=(sz_k,sz_k), dtype=DataType)
        e_k[row[k],col[k]] = 1.

        IFP = PPU.FraunhoferPropWF_GPU(e_k, 1., 1, 1.)
        IFP *= FPM
        LP = PPU.FraunhoferPropWF_GPU(IFP,(1.), -1, 1.)
        LP *= Lyot
        tmp = PPU.FraunhoferPropWF_GPU(LP,1.,1,1.)

        det_crop = tmp[params['Detx'].min():(params['Detx'].max()+1), params['Dety'].min():(params['Dety'].max()+1)]

        D[k,:] = np.hstack(det_crop.flatten())
        # if pscale is 1, the field max is 1.
        # if pscale in Fraunhofer props is cnorm, the field max is scaled by np.abs(cnorm) * Photon_Density * row.size

    return D.transpose()
# End of construct_System_Operator_lyot_efficient
```

```
def construct_System_Operator_APP_efficient(self,APPfilename,DataType=np.complex64):
    # Operator needs to be correctly normalized to produce a PSF with intensity max of 1
    sz_k = int(PUPIL_zp.shape[0])
    row, col = np.where(PUPIL_zp==1.)
#        norm = 1./(sz_k*np.sqrt(2.))
#        cnorm = np.complex(norm,norm)
    app = PPU.simple_fitsread(APPfilename)
    D = np.zeros(shape=(row.size, params['Detx'].size * params['Dety'].size),dtype=DataType)
    for k in range(row.size):
        e_k = np.zeros(shape=(sz_k,sz_k), dtype=DataType)
        e_k[row[k],col[k]] = 1.

        pup_k = e_k * np.exp(1j*app[k])
        tmp = PPU.FraunhoferPropWF_GPU(pup_k, 1., 1, 1.)


        det_crop = tmp[params['Detx'].min():(params['Detx'].max()+1), params['Dety'].min():(params['Dety'].max()+1)]

        D[k,:] = np.hstack(det_crop.flatten())
        # if pscale is 1, the field max is 1.
        # if pscale in Fraunhofer props is cnorm, the field max is scaled by np.abs(cnorm) * Photon_Density * row.size

    return D.transpose()
# End of construct_System_Operator_lyot_efficient

def construct_System_Operator_NoC_efficient(self,DataType=np.complex64):
    # Operator needs to be correctly normalized to produce a PSF with intensity max of 1
    sz_k = int(PUPIL_zp.shape[0])
    row, col = np.where(PUPIL_zp==1.)
#        norm = 1./(sz_k*np.sqrt(2.))
#        cnorm = np.complex(norm,norm)

    D = np.zeros(shape=(row.size, params['Detx'].size * params['Dety'].size),dtype=DataType)
    for k in range(row.size):
        e_k = np.zeros(shape=(sz_k,sz_k), dtype=DataType)
        e_k[row[k],col[k]] = 1.

        IFP = PPU.FraunhoferPropWF_GPU(e_k, 1., 1, 1.)
        LP = PPU.FraunhoferPropWF_GPU(IFP,1., -1, 1.)
        tmp = PPU.FraunhoferPropWF_GPU(LP,1.,1,1.)

        det_crop = tmp[params['Detx'].min():(params['Detx'].max()+1), params['Dety'].min():(params['Dety'].max()+1)]
```

```
        D[k,:] = np.hstack(det_crop.flatten())

    return D.transpose()
# End of construct_System_Operator_none_efficient
```

REFERENCES

Assémat, F., R. W. Wilson, and E. Gendron (2006). Method for simulating infinitely long and non stationary phase screens with optimized memory storage. *Opt. Express*, **14**(3), pp. 988–999. doi:10.1364/OE.14.000988.

Beuzit, J.-L., M. Feldt, K. Dohlen, D. Mouillet, P. Puget, F. Wildi, L. Abe, J. Antichi, A. Baruffolo, P. Baudoz, A. Boccaletti, M. Carbillet, J. Charton, R. Claudi, M. Downing, C. Fabron, P. Feautrier, E. Fedrigo, T. Fusco, J.-L. Gach, R. Gratton, T. Henning, N. Hubin, F. Joos, M. Kasper, M. Langlois, R. Lenzen, C. Moutou, A. Pavlov, C. Petit, J. Pragt, P. Rabou, F. Rigal, R. Roelfsema, G. Rousset, M. Saisse, H.-M. Schmid, E. Stadler, C. Thalmann, M. Turatto, S. Udry, F. Vakili, and R. Waters (2008). SPHERE: a planet finder instrument for the VLT. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7014 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. doi:10.1117/12.790120.

Boccaletti, A., P. Riaud, P. Baudoz, J. Baudrand, D. Rouan, D. Gratadour, F. Lacombe, and A.-M. Lagrange (2004). The Four-Quadrant Phase Mask Coronagraph. IV. First Light at the Very Large Telescope. *Pub. Astron. Soc. Pacific*, **116**, pp. 1061–1071. doi:10.1086/425735.

Borucki, W. J., D. Koch, G. Basri, N. Batalha, T. Brown, D. Caldwell, J. Caldwell, J. Christensen-Dalsgaard, W. D. Cochran, E. DeVore, E. W. Dunham, A. K. Dupree, T. N. Gautier, J. C. Geary, R. Gilliland, A. Gould, S. B. Howell, J. M. Jenkins, Y. Kondo, D. W. Latham, G. W. Marcy, S. Meibom, H. Kjeldsen, J. J. Lissauer, D. G. Monet, D. Morrison, D. Sasselov, J. Tarter, A. Boss, D. Brownlee, T. Owen, D. Buzasi, D. Charbonneau, L. Doyle, J. Fortney, E. B. Ford, M. J. Holman, S. Seager, J. H. Steffen, W. F. Welsh, J. Rowe, H. Anderson, L. Buchhave, D. Ciardi, L. Walkowicz, W. Sherry, E. Horch, H. Isaacson, M. E. Everett, D. Fischer, G. Torres, J. A. Johnson, M. Endl, P. MacQueen, S. T. Bryson, J. Dotson, M. Haas, J. Kolodziejczak, J. Van Cleve, H. Chandrasekaran, J. D. Twicken, E. V. Quintana, B. D. Clarke, C. Allen, J. Li, H. Wu, P. Tenenbaum, E. Verner, F. Bruhweiler, J. Barnes, and A. Prsa (2010). Kepler Planet-Detection Mission: Introduction and First Results. *Science*, **327**(5968), p. 977. doi:10.1126/science.1185402.

Brooks, K. J., L. Catala, M. A. Kenworthy, S. M. Crawford, and J. L. Codona (2016). Polarization dOTF: on-sky focal plane wavefront sensing. In Navarro, R. and J. H. Burge (eds.) *Advances in Optical and Mechanical Technologies for*

Telescopes and Instrumentation II, volume 9912, pp. 13 – 23. International Society for Optics and Photonics, SPIE.

Chu, Q., S. Jefferies, and J. G. Nagy (2013). Iterative Wavefront Reconstruction for Astronomical Imaging. *SIAM Journal on Scientific Computing*, **35**(5), pp. S84–S103. doi:10.1137/120882603.

Close, L. M., J. R. Males, O. Durney, C. Sauve, M. Kautz, A. Hedglen, L. Schatz, J. Lumbres, K. Miller, K. V. Gorkom, M. Jean, and V. Gasho (2018). Optical and mechanical design of the extreme AO coronagraphic instrument MagAO-X. In Close, L. M., L. Schreiber, and D. Schmidt (eds.) *Adaptive Optics Systems VI*, volume 10703, pp. 1227 – 1236. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2312280.

Codona, J. L. (2012). Theory and application of differential OTF (dOTF) wavefront sensing. In Ellerbroek, B. L., E. Marchetti, and J.-P. Vran (eds.) *Adaptive Optics Systems III*, volume 8447, pp. 2188 – 2199. International Society for Optics and Photonics, SPIE. doi:10.1117/12.926963.

Codona, J. L. (2013). Differential optical transfer function wavefront sensing. *Optical Engineering*, **52**(9), pp. 1 – 13. doi:10.1117/1.OE.52.9.097105.

Codona, J. L. and N. Doble (2015). James Webb Space Telescope segment phasing using differential optical transfer functions. *Journal of Astronomical Telescopes, Instruments, and Systems*, **1**(2), pp. 1 – 13. doi:10.1117/1.JATIS.1.2.029001.

Codona, J. L. and M. Kenworthy (2013). Focal Plane Wavefront Sensing Using Residual Adaptive Optics Speckles. *Astrophysical Journal*, **767**, 100. doi:10.1088/0004-637X/767/2/100.

Fétick, R., L. Mugnier, T. Fusco, and B. Neichel (2020). Blind deconvolution in astronomy with adaptive optics: the parametric marginal approach. *arXiv e-prints*, arXiv:2006.11160.

Finger, G., I. Baker, D. Alvarez, D. Ives, L. Mehrgan, M. Meyer, J. Stegmeier, and H. J. Weller (2014). SAPHIRA detector for infrared wavefront sensing. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9148 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 17. doi:10.1117/12.2057078.

Fitzgerald, M. P. and J. R. Graham (2006). Speckle Statistics in Adaptively Corrected Images. *Astrophysical Journal*, **637**(1), pp. 541–547. doi:10.1086/498339.

Frazin, R. A. (2013). Utilization of the Wavefront Sensor and Short-exposure Images for Simultaneous Estimation of Quasi-static Aberration and Exoplanet Intensity. *Astrophysical Journal*, **767**, 21. doi:10.1088/0004-637X/767/1/21.

Frazin, R. A. (2018). Efficient, nonlinear phase estimation with the nonmodulated pyramid wavefront sensor. *Journal of the Optical Society of America A*, **35**, p. 594. doi:10.1364/JOSAA.35.000594.

Frazin, R. A. and A. T. Rodack (2021). Millisecond exoplanet imaging: II. regression equations and technical discussion. *J. Opt. Soc. Am. A*, **38**(10), pp. 1557–1569. doi:10.1364/JOSAA.426339.

Fried, D. L. (1966). Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures. *J. Opt. Soc. Am.*, **56**(10), pp. 1372–1379. doi:10.1364/JOSA.56.001372.

Fried, D. L. (1982). Anisoplanatism in adaptive optics. *J. Opt. Soc. Am.*, **72**(1), pp. 52–61. doi:10.1364/JOSA.72.000052.

Fusco, T., J. P. Véran, J. M. Conan, and L. M. Mugnier (1999). Myopic deconvolution method for adaptive optics images of stellar fields. *Astron. and Astrophys. Supp.*, **134**, pp. 193–200. doi:10.1051/aas:1999133.

Galicher, R. and C. Marois (2011). Astrometry and photometry in high contrast imaging. In *Second International Conference on Adaptive Optics for Extremely Large Telescopes, id.P25*, p. 25P.

Ghiglia, D. C. and M. D. Pritt (1998). *Two Dimensional Phase Unwrapping*. Wiley & Sons, Inc.

Gladysz, S., N. Yaitskova, and J. C. Christou (2010). Statistics of intensity in adaptive-optics images and their usefulness for detection and photometry of exoplanets. *Journal of the Optical Society of America A*, **27**(26), pp. A260000–A75. doi:10.1364/JOSAA.27.000A64.

Gonsalves, R. A. and R. Chidlaw (1979). Wavefront Sensing By Phase Retrieval. In Tescher, A. G. (ed.) *Applications of Digital Image Processing III*, volume 0207, pp. 32 – 39. International Society for Optics and Photonics, SPIE. doi:10.1117/12.958223.

Goodman, J. W. (2015). *Statistical Optics, 2nd Edition*. John Wiley and Sons, Inc.

Goodman, J. W. (2017). *Introduction to Fourier Optics, fourth edition*. The McGraw-Hill Companies, Inc.

Gorkom, K. V., J. R. Males, L. M. Close, J. Lumbres, A. D. Hedglen, J. D. Long, S. Y. Haffert, O. Guyon, M. Y. Kautz, L. Schatz, K. L. Miller, A. T. Rodack, J. M. Knight, and K. M. Morzinski (2021). Characterizing deformable mirrors for the MagAO-X instrument. *Journal of Astronomical Telescopes, Instruments, and Systems*, **7**(3), pp. 1 – 18. doi:10.1117/1.JATIS.7.3.039001.

Groff, T. D. and N. J. Kasdin (2013). Kalman filtering techniques for focal plane electric field estimation. *J. Opt. Soc. Am. A*, **30**(1), pp. 128–139. doi:10.1364/JOSAA.30.000128.

Guyon, O. (2003). Phase-induced amplitude apodization of telescope pupils for extrasolar terrestrial planet imaging. *Astronomy & Astrophysics*, **404**(1), p. 379. ISSN 1432-0746. doi:10.1051/0004-6361:20030457.

Guyon, O. (2005). Limits of Adaptive Optics for High-Contrast Imaging. *Astrophysical Journal*, **629**, pp. 592–614. doi:10.1086/431209.

Guyon, O. (2018). Extreme Adaptive Optics. *Ann. Rev. Astron. Astrophys.*, **56**, pp. 315–355. doi:10.1146/annurev-astro-081817-052000.

Guyon, O., P. M. Hinz, E. Cady, R. Belikov, and F. Martinache (2014). High Performance Lyot and PIAA Coronagraphy for Arbitrarily Shaped Telescope Apertures. *Astrophysical Journal*, **780**, 171. doi:10.1088/0004-637X/780/2/171.

Hardy, J. W. (1998). *Adaptive Optics for Astronomical Telescopes*. Oxford University Press.

Hart, M. and J. L. Codona (2012). Wavefront sensing with the differential optical transfer function. In Dolne, J. J., T. J. Karr, and V. L. Gamiz (eds.) *Unconventional Imaging and Wavefront Sensing 2012*, volume 8520, pp. 122 – 132. International Society for Optics and Photonics, SPIE.

Hecht, E. (2012). *Optics*. Pearson. ISBN 9788131718070.

Herscovici-Schiller, O., J.-F. Sauvage, L. M. Mugnier, K. Dohlen, and A. Vigan (2019). Coronagraphic phase diversity through residual turbulence: performance study and experimental validation. *Monthly Notices Royal Astron. Soc.*, **488**(3), pp. 4307–4316. doi:10.1093/mnras/stz1986.

Jennifer Lumbres, e. a. (in prep).

Jiang, F., G. Ju, X. Qi, and S. Xu (2019). Cross-iteration deconvolution strategy for differential optical transfer function (dOTF) wavefront sensing. *Opt. Lett.*, **44**(17), pp. 4283–4286. doi:10.1364/OL.44.004283.

Jovanovic, N., O. Absil, P. Baudoz, M. Beaulieu, M. Bottom, E. Cady, B. Carlomagno, A. Carlotti, D. Doelman, K. Fogarty, R. Galicher, O. Guyon, S. Haffert, E. Huby, J. Jewell, C. Keller, M. A. Kenworthy, J. Knight, J. Kuhn, K. Miller, J. Mazoyer, M. N'Diaye, E. Por, L. Pueyo, A. J. E. Riggs, G. Ruane, D. Sirbu, F. Snik, J. K. Wallace, M. Wilby, and M. Ygouf (2018). Review of high-contrast imaging systems for current and future ground-based and space-based telescopes II. Common path wavefront sensing/control and Coherent Differential Imaging.

Kaltenegger, L., W. A. Traub, and K. W. Jucks (2007). Spectral Evolution of an Earth-like Planet. *ApJ*, **658**(1), pp. 598–616. doi:10.1086/510996.

Knight, J. M., A. T. Rodack, J. L. Codona, K. L. Miller, and O. Guyon (2015). Deconvolution of differential OTF (dOTF) to measure high-resolution wavefront structure. In Shaklan, S. (ed.) *Techniques and Instrumentation for Detection of Exoplanets VII*, volume 9605, pp. 659 – 666. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2189575.

Krishnamoorthy, P., J. Walawender, W. T. Gee, and O. Guyon (2020). PANOPTES: A citizen science project to discover exoplanets from your backyard using off-the-shelf hardware. In Marshall, H. K., J. Spyromilio, and T. Usuda (eds.) *Ground-based and Airborne Telescopes VIII*, volume 11445, pp. 577 – 583. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2563188.

Lozi, J. and others (2018). SCExAO, an instrument with a dual purpose: perform cutting-edge science and develop new technologies. In *Adaptive Optics System VI*, volume 10703-270 of *Proc. SPIE*.

Macintosh, B., L. Poyneer, A. Sivaramakrishnan, and C. Marois (2005). Speckle lifetimes in high-contrast adaptive optics. In Tyson, R. K. and M. Lloyd-Hart (eds.) *Astronomical Adaptive Optics Systems and Applications II*, volume 5903, pp. 170 – 177. International Society for Optics and Photonics, SPIE.

Macintosh, B. A., J. R. Graham, D. W. Palmer, R. Doyon, J. Dunn, D. T. Gavel, J. Larkin, B. Oppenheimer, L. Saddlemyer, A. Sivaramakrishnan, J. K. Wallace, B. Bauman, D. A. Erickson, C. Marois, L. A. Poyneer, and R. Soummer (2008). The Gemini Planet Imager: from science to design to construction. In Hubin, N., C. E. Max, and P. L. Wizinowich (eds.) *Adaptive Optics Systems*, volume 7015, pp. 315 – 327. International Society for Optics and Photonics, SPIE. doi:10.1117/12.788083.

Males, J. R., L. M. Close, K. Miller, L. Schatz, D. Doelman, J. Lumbres, F. Snik, A. Rodack, J. Knight, K. Van Gorkom, J. D. Long, A. Hedglen, M. Kautz, N. Jovanovic, K. Morzinski, O. Guyon, E. Douglas, K. B. Follette, J. Lozi, C. Bohlman, O. Durney, V. Gasho, P. Hinz, M. Ireland, M. Jean, C. Keller, M. Kenworthy, B. Mazin, J. Noenickx, D. Alfred, K. Perez, A. Sanchez, C. Sauve, A. Weinberger, and A. Conrad (2018). MagAO-X: project status and first laboratory results. In *Proc. SPIE*, volume 10703, p. 1070309. doi:10.1117/12.2312992.

Males, J. R., M. P. Fitzgerald, R. Belikov, and O. Guyon (2021). The Mysterious Lives of Speckles. I. Residual Atmospheric Speckle Lifetimes in Ground-based Coronagraphs. *Publications of the Astronomical Society of the Pacific*, **133**(1028), p. 104504. doi:10.1088/1538-3873/ac0f0c.

Males, J. R. and O. Guyon (2018). Ground-based adaptive optics coronagraphic performance under closed-loop predictive control. *Journal of Astronomical Telescopes, Instruments, and Systems*, **4**(1), pp. 1 – 21. doi:10.1117/1.JATIS.4.1.019001.

Marois, C., R. Doyon, R. Racine, and D. Nadeau (2000). Efficient Speckle Noise Attenuation in Faint Companion Imaging. *Pub. Astron. Soc. Pacific*, **112**, pp. 91–96. doi:10.1086/316492.

Marois, C., D. Lafrenière, R. Doyon, B. Macintosh, and D. Nadeau (2006). Angular Differential Imaging: A Powerful High-Contrast Imaging Technique. *Astrophysical Journal*, **641**, pp. 556–564. doi:10.1086/500401.

Marois, C., B. Macintosh, and J.-P. Véran (2010). Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7736 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. doi:10.1117/12.857225.

Martinache, F., O. Guyon, N. Jovanovic, C. Clergeon, G. Singh, T. Kudo, T. Currie, C. Thalmann, M. McElwain, and M. Tamura (2014). On-Sky Speckle Nulling Demonstration at Small Angular Separation with SCExAO. *Pub. Astron. Soc. Pacific*, **126**, pp. 565–572. doi:10.1086/677141.

Martinache, F., N. Jovanovic, and O. Guyon (2016). Closed-loop focal plane wavefront control with the SCExAO instrument. *Astronomy and Astrophysics*, **593**, A33. doi:10.1051/0004-6361/201628496.

Martinez, P., M. Kasper, A. Costille, J. F. Sauvage, K. Dohlen, P. Puget, and J. L. Beuzit (2013). Speckle temporal stability in XAO coronagraphic images. II. Refine model for quasi-static speckle temporal evolution for VLT/SPHERE. *Astronomy and Astrophysics*, **554**, A41. doi:10.1051/0004-6361/201220820.

Mascareo, A. S., J. P. Faria, P. Figueira, C. Lovis, M. Damasso, J. I. G. Hernndez, R. Rebolo, S. Cristiano, F. Pepe, N. C. Santos, M. R. Z. Osorio, V. Adibekyan, S. Hojjatpanah, A. Sozzetti, F. Murgas, M. Abreo, M. Affolter, Y. Alibert, M. Aliverti, R. Allart, C. A. Prieto, D. Alves, M. Amate, G. Avila, V. Baldini, T. Bandi, S. C. C. Barros, A. Bianco, W. Benz, F. Bouchy, C. Broeng, A. Cabral, G. Calderone, R. Cirami, J. Coelho, P. Conconi, I. Coretti, C. Cumani, G. Cupani, V. D'Odorico, S. Deiries, B. Delabre, P. D. Marcantonio, X. Dumusque, D. Ehrenreich, A. Fragoso, L. Genolet, M. Genoni, R. G. Santos, I. Hughes, O. Iwert, K. Ferber, J. Knusdrtrup, M. Landoni, B. Lavie, J. Lillo-Box, J. Lizon, G. L. Curto, C. Maire, A. Manescau, C. J. A. P. Martins, D. Mgevand, A. Mehner, G. Micela, A. Modigliani, P. Molaro, M. A. Monteiro, M. J. P. F. G. Monteiro,

M. Moschetti, E. Mueller, N. J. Nunes, L. Oggioni, A. Oliveira, E. Pall, G. Pariani, L. Pasquini, E. Poretti, J. L. Rasilla, E. Redaelli, M. Riva, S. S. Tschudi, P. Santin, P. Santos, A. Segovia, D. Sosnoswska, S. Sousa, P. Span, F. Tenegi, S. Udry, A. Zanutta, and F. Zerbi (2020). Revisiting Proxima with ESPRESSO.

Matthews, C. T., J. R. Crepp, G. Vasisht, and E. Cady (2017). Electric field conjugation for ground-based high-contrast imaging: robustness study and tests with the Project 1640 coronagraph. *Journal of Astronomical Telescopes, Instruments, and Systems*, **3**(4), pp. 1 – 12 – 12. doi:10.1117/1.JATIS.3.4.045001.

Mawet, D., L. Pueyo, P. Lawson, L. Mugnier, W. Traub, A. Boccaletti, J. T. Trauger, S. Gladysz, E. Serabyn, J. Milli, R. Belikov, M. Kasper, P. Baudoz, B. Macintosh, C. Marois, B. Oppenheimer, H. Barrett, J.-L. Beuzit, N. Devaney, J. Girard, O. Guyon, J. Krist, B. Mennesson, D. Mouillet, N. Murakami, L. Poyneer, D. Savransky, C. Vérinaud, and J. K. Wallace (2012). Review of small-angle coronagraphic techniques in the wake of ground-based second-generation adaptive optics systems. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8442 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. doi:10.1117/12.927245.

Mawet, D., J. T. Trauger, E. Serabyn, D. C. Moody, K. M. Liewer, J. E. Krist, D. M. Shemo, and N. A. O'Brien (2009). Vector vortex coronagraph: first results in the visible. In Shaklan, S. B. (ed.) *Techniques and Instrumentation for Detection of Exoplanets IV*, volume 7440, pp. 301 – 310. International Society for Optics and Photonics, SPIE. doi:10.1117/12.826569.

Meeker, S. R., B. A. Mazin, A. B. Walter, P. Strader, N. Fruitwala, C. Bockstiegel, P. Szypryt, G. Ulbricht, G. Coiffard, B. Bumble, G. Cancelo, T. Zmuda, K. Treptow, N. Wilcer, G. Collura, R. Dodkins, I. Lipartito, N. Zobrist, M. Bottom, J. C. Shelton, D. Mawet, J. C. van Eyken, G. Vasisht, and E. Serabyn (2018). DARKNESS: A Microwave Kinetic Inductance Detector Integral Field Spectrograph for High-contrast Astronomy. *Pub. Astron. Soc. Pacific*, **130**(988), p. 065001. doi:10.1088/1538-3873/aab5e7.

Miller, K., O. Guyon, J. Codona, J. Knight, and A. Rodack (2015). UA wavefront control lab: design overview and implementation of new wavefront sensing techniques. In Shaklan, S. (ed.) *Techniques and Instrumentation for Detection of Exoplanets VII*, volume 9605, pp. 659 – 668. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2189915.

Otten, G. P. P. L., F. Snik, M. A. Kenworthy, M. N. Miskiewicz, and M. J. Escuti (2014). Performance characterization of a broadband vector Apodizing Phase Plate coronagraph. *Opt. Express*, **22**(24), pp. 30287–30314. doi:10.1364/OE.22.030287.

Paul, B., J.-F. Sauvage, and L. M. Mugnier (2013). Coronagraphic phase diversity: performance study and laboratory demonstration. *Astronomy and Astrophysics*, **552**, A48. doi:10.1051/0004-6361/201220940.

Perrin, M., J. Long, E. Douglas, A. Sivaramakrishnan, C. Slocum, and others (2016). POPPY: Physical Optics Propagation in PYthon.

Perrin, M. D., L. Pueyo, K. V. Gorkom, K. Brooks, A. Rajan, J. Girard, and C.-P. Lajoie (2018). Updated optical modeling of JWST coronagraph performance contrast, stability, and strategies. In Lystrup, M., H. A. MacEwen, G. G. Fazio, N. Batalha, N. Siegler, and E. C. Tong (eds.) *Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave*, volume 10698, pp. 98 – 113. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2313552.

Por, E. H. (2017). Optimal design of apodizing phase plate coronagraphs. In Shaklan, S. (ed.) *Techniques and Instrumentation for Detection of Exoplanets VIII*, volume 10400, pp. 236 – 247. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2274219.

Por, E. H., S. Y. Haffert, V. M. Radhakrishnan, D. S. Doelman, M. van Kooten, and S. P. Bos (2018). High Contrast Imaging for Python (HCIPy): an open-source adaptive optics and coronagraph simulator. In Close, L. M., L. Schreiber, and D. Schmidt (eds.) *Adaptive Optics Systems VI*, volume 10703, pp. 1112 – 1125. International Society for Optics and Photonics, SPIE. doi:10.1117/12.2314407.

Potier, A., P. Baudoz, R. Galicher, G. Singh, and A. Boccaletti (2020). Comparing focal plane wavefront control techniques: Numerical simulations and laboratory experiments. *Astronomy and Astrophysics*, **635**, p. A192. ISSN 1432-0746. doi:10.1051/0004-6361/201937015.

Pueyo, L., J. R. Crepp, G. Vasisht, D. Brenner, B. R. Oppenheimer, N. Zimmerman, S. Hinkley, I. Parry, C. Beichman, L. Hillenbrand, L. C. Roberts, R. Dekany, M. Shao, R. Burruss, A. Bouchez, J. Roberts, and R. Soummer (2012). Application of a Damped Locally Optimized Combination of Images Method to the Spectral Characterization of Faint Companions Using an Integral Field Spectrograph. *Astrophysical Journals*, **199**, 6. doi:10.1088/0067-0049/199/1/6.

Ragazzoni, R. (1996). Pupil plane wavefront sensing with an oscillating prism. *Journal of modern optics*, **43**(2), pp. 289–293.

Rameau, J., G. Chauvin, A.-M. Lagrange, A.-L. Maire, A. Boccaletti, and M. Bonnefoy (2015). Detection limits with spectral differential imaging data. *Astronomy and Astrophysics*, **581**, A80. doi:10.1051/0004-6361/201525879.

Ricker, G. R., J. N. Winn, R. Vanderspek, D. W. Latham, G. . Bakos, J. L. Bean, Z. K. Berta-Thompson, T. M. Brown, L. Buchhave, N. R. Butler, R. P. Butler, W. J. Chaplin, D. B. Charbonneau, J. Christensen-Dalsgaard, M. Clampin, D. Deming, J. P. Doty, N. D. Lee, C. Dressing, E. W. Dunham, M. Endl, F. Fressin, J. Ge, T. Henning, M. J. Holman, A. W. Howard, S. Ida, J. M. Jenkins, G. Jernigan, J. A. Johnson, L. Kaltenegger, N. Kawai, H. Kjeldsen, G. Laughlin, A. M. Levine, D. Lin, J. J. Lissauer, P. MacQueen, G. Marcy, P. R. McCullough, T. D. Morton, N. Narita, M. Paegert, E. Palle, F. Pepe, J. Pepper, A. Quirrenbach, S. A. Rinehart, D. Sasselov, B. Sato, S. Seager, A. Sozzetti, K. G. Stassun, P. Sullivan, A. Szentgyorgyi, G. Torres, S. Udry, and J. Villasenor (2014). Transiting Exoplanet Survey Satellite. *Journal of Astronomical Telescopes, Instruments, and Systems*, **1**(1), pp. 1 – 10. doi:10.1117/1.JATIS.1.1.014003.

Rodack, A. T., R. A. Frazin, J. R. Males, and O. Guyon (2021). Millisecond exoplanet imaging: I. method and simulation results. *J. Opt. Soc. Am. A*, **38**(10), pp. 1541–1556. doi:10.1364/JOSAA.426046.

Rodack, A. T., J. M. Knight, J. L. Codona, K. L. Miller, and O. Guyon (2015). Adaptive optics self-calibration using differential OTF (dOTF). In Shaklan, S. (ed.) *Techniques and Instrumentation for Detection of Exoplanets VII*, volume 9605, pp. 669 – 677. International Society for Optics and Photonics, SPIE.

Rodack, A. T., J. R. Males, O. Guyon, B. A. Mazin, M. P. Fitzgerald, and D. Mawet (2018). Real-time estimation and correction of quasi-static aberrations in ground-based high contrast imaging systems with high frame-rates. In Close, L. M., L. Schreiber, and D. Schmidt (eds.) *Adaptive Optics Systems VI*, volume 10703, pp. 859 – 868. International Society for Optics and Photonics, SPIE.

Sauvage, J.-F., L. M. Mugnier, G. Rousset, and T. Fusco (2010). Analytical expression of long-exposure adaptive-optics-corrected coronagraphic image. First application to exoplanet detection. *J. Opt. Soc. Am. A*, **27**(11), pp. A157–A170.

Schulz, T. J., B. E. Stribling, and J. J. Miller (1997). Multiframe blind deconvolution with real data: imagery of the Hubble Space Telescope. *Optics Express*, **1**, p. 355. doi:10.1364/OE.1.000355.

Sparks, W. B. and H. C. Ford (2002). Imaging Spectroscopy for Extrasolar Planet Detection. *Astrophysical Journal*, **578**, pp. 543–564. doi:10.1086/342401.

Stangalini, M., G. Li Causi, F. Pedichini, S. Antoniucci, M. Mattioli, J. Christou, G. Consolini, D. Hope, S. M. Jefferies, and R. Piazzesi (2018). Recurrence Quantification Analysis as a Post-processing Technique in Adaptive Optics High-contrast Imaging. *Astrophysical Journal*, **868**(1), 6. doi:10.3847/1538-4357/aae58e.

Thomas, S. J., A. A. Give'On, D. Dillon, B. Macintosh, D. Gavel, and R. Soummer (2010). Laboratory test of application of electric field conjugation image-sharpening to ground-based adaptive optics. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7736 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. doi: 10.1117/12.857511.

Uyama, T., B. Norris, N. Jovanovic, J. Lozi, P. G. Tuthill, O. Guyon, T. Kudo, J. Hashimoto, M. Tamura, and F. Martinache (2020). High-contrast Hα imaging with Subaru/SCExAO + VAMPIRES. *Journal of Astronomical Telescopes, Instruments, and Systems*, **6**(4), pp. 1 – 17. doi:10.1117/1.JATIS.6.4.045004.

Vigan, A., C. Gry, G. Salter, D. Mesa, D. Homeier, C. Moutou, and F. Allard (2015). High-contrast imaging of Sirius A with VLT/SPHERE: looking for giant planets down to one astronomical unit. *Monthly Notices of the Royal Astronomical Society*, **454**(1), pp. 129–143. ISSN 0035-8711. doi:10.1093/mnras/stv1928.

Vigan, A., M. N'Diaye, K. Dohlen, J. F. Sauvage, J. Milli, G. Zins, C. Petit, Z. Wahhaj, F. Cantalloube, A. Caillat, A. Costille, J. Le Merrer, A. Carlotti, J. L. Beuzit, and D. Mouillet (2019). Calibration of quasi-static aberrations in exoplanet direct-imaging instruments with a Zernike phase-mask sensor. III. On-sky validation in VLT/SPHERE. *Astronomy and Astrophysics*, **629**, A11. doi:10.1051/0004-6361/201935889.

Walter, A. B., C. Bockstiegel, T. D. Brandt, and B. A. Mazin (2019). Stochastic Speckle Discrimination with Time-tagged Photon Lists: Digging below the Speckle Noise Floor. *Pub. Astron. Soc. Pacific*, **131**(1005), p. 114506. doi: 10.1088/1538-3873/ab389a.