On the Use of Classical MTF Measurements to Perform Wavefront Sensing

Brandon D. Dube

Supervised by:
Dr. James R. Fienup

The Institute of Optics
Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, NY

2018

Contents

1 Introduction

Imaging optical systems may be divided by their performance into two broad categories; those which operate at the diffraction limit, and those which operate below it. "Diffraction limited" may be determined by any number of rules, e.g. the Maréchal criteria [1] or Strehl ratio [2–5]. Systems near these limits typically have their performance analyzed in wavefront space. Systems below the diffraction limit are most typically analyzed by their Modulation Transfer Function (MTF) [6–9]. An unfortunate quality of MTF is that it offers no insight into what is wrong if a test returns a negative result. This causes there to be high cost or low precision associated with the correction of systems which had their quality control performed via MTF measurement. In contrast, wavefront metrology allows precise correction of alignment due to the more direct relationship between an optical system's parameters and its wavefront aberrations.

This thesis proposes to provide a bridge between the world of MTF and wavefront metrology via a phase retrieval algorithm. This will allow the same MTF benches which perform quality control measurements of many optics, in particular camera lenses, to also serve as alignment stations. This can be considered an enabling technology for the next generation of high resolution objective lenses suitable for ultra-high pixel density sensors.

2 Background

2.1 Wavefront Evaluation

It is often a requirement to evaluate the wavefront of an optical system. This can be directly measured via interferometry [10], Shack-Hartmann [11, 12] or other similar wavefront sensor technologies. For some systems, such as the Hubble Space Telescope [13] on-orbit, these tests are infeasible. A system could also operate in a spectral band where interferometers and wavefront sensors are not commonly available, such as long wave infra-red. In cases where these methods cannot or will not be used, an alternative is required.

In the mid 1970's, Muller and Buffington worked on correction of images with atmospheric turbulence [14], a task that is can be summarized as compensating for the phase errors of a system. Additionally, Gerchberg and Saxton [15] as well as Fienup [16] began developing algorithms used to reconstruct the phase of something after its destruction by a modulus operation. This task is known as phase retrieval. A comparison of techniques can be found in [17] and a history in [18].

Phase retrieval applied to the recovery of the phase errors of an optical pupil has become known as wavefront sensing. This technique was used to successfully diagnose the error in the Hubble Space Telescope [19] and is planned to perform the fine phasing of the primary mirrors of the James Webb Space Telescope (JWST) [20]. It generally requires that physical parameters be known, such as the shape of the pupil, the propagation distance (focal length), the wavelength of light and sample (pixel) spacing in the capture plane. The object is also usually a point source, and treated as known. These factors can all be retrieved in favorable conditions, see e.g. [21, 22], but the more work the algorithm has to do, the slower it will go and the lower the overall accuracy will be.

2.2 Anatomy and Theory of an MTF Bench

An MTF bench must accomplish four tasks: **(1)** the projection of an object with known spatial and spectral characteristics, **(2)** high speed mounting and pose of an objective under test (OUT) with respect to that object at any desired angle or distance, **(3)** achievement of oversampling to expand the applicable bandwidth of the system and avoid aliasing, and **(4)** processing of the resulting image to yield an MTF measurement.

The principal components of an MTF bench are laid out in Figure 1. This particular illustration indicates the geometry where the collimator is rotated and the objective platform is static. An alternative exists where the objective platform is rotated and the collimator is static. Optionally, a filter can be installed before the diffuser to adjust the spectral content of the light. The object is most commonly either a pinhole or slit; these provide the easiest task of extracting an MTF from the captured image and are known as the pinhole [23, 24] and slit [25–27] methods of MTF measurement.

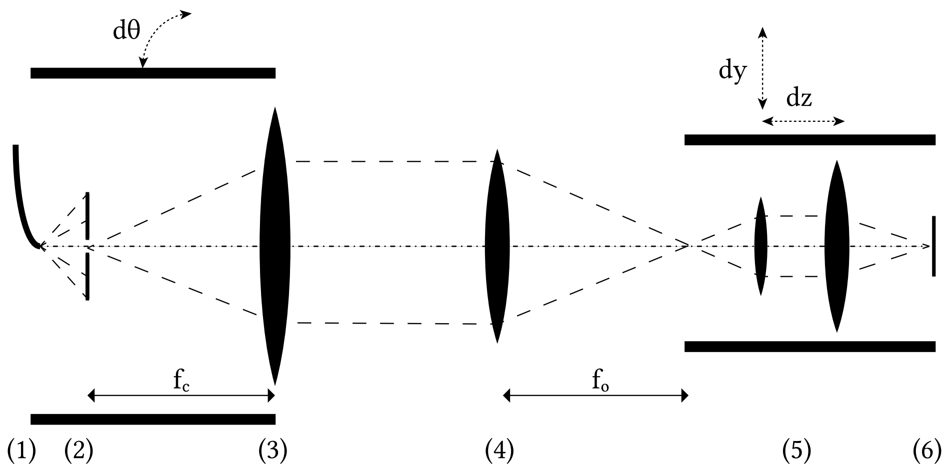

Fig. 1. Schematic of an MTF bench. (1) light source, (2) object (pinhole or slit(s)), (3) collimator, (4) objective under test, (5) microscope assembly (objective and tube lens), (6) detector. dθ, dy, and dz indicate axes of motion control. Elements 1-3 and 5-6 are subsystems that move as independent assemblies with (4) fixed. Note that alternative motion configurations exist.

An object $O(u, v)$ at the focal plane of the collimator will map into the OUT's image plane with spatial dimensions x and y via the left-hand change of variables in Equation 1. Similarly, the detector's Point Spread Function (PSF), $G_d(s, t)$ maps to x and y via the right-hand change of variables in Equation 1.

$$\begin{aligned} u &= m_1 x \qquad\qquad & s &= m_2 x \\ v &= m_1 y \qquad\qquad & t &= m_2 y \end{aligned} \tag{1}$$

These mappings are easily found via geometrical optics; the collimator and OUT form one telescope, while the microscope objective and its tube lens form another. Each relays an object from the focal plane of one component to the other. The value $m_1 = -f_o/f_c$ is the signed ratio of the OUT focal length to the collimator focal length. The value m_2 is a property of the microscope objective for an appropriately paired tube lens.

After applying these changes of variables, we may express the final image as a double convolution of the object O, OUT's PSF H, and the MTF bench detector's PSF G_d:

$$I(x,y) \approx O(x,y) * H(x,y) * G_d(x,y) \quad . \tag{2}$$

A convenient property of convolution is that it is simply a multiplication in the Fourier domain,

$$\mathcal{I}(\nu_x,\nu_y) \approx \mathcal{O}(\nu_x,\nu_y)\mathcal{H}(\nu_x,\nu_y)\mathcal{G}_d(\nu_x,\nu_y) \tag{3}$$

where $\mathcal{I}$, $\mathcal{O}$, $\mathcal{H}$, and $\mathcal{G}_d$ are the Fourier transforms of I, O, H, and G_d. $\mathcal{H}$ is known as the Optical Transfer Function or OTF.

The relationship is approximate, as any wavefront error in the microscope assembly or collimator will be combined with that of the OUT. This interaction occurs in a nontrivial way and is usually of sufficient difficulty to account for that it is only mitigated by use of high quality components [28].

An object pinhole of width α has a Fourier transform which is a jinc function, and the rectangular elements of the detector of widths (w_x,w_y) have a Fourier transform which is a separable product of sinc functions. Pixel crosstalk, response non-uniformity, dark current, read noise and other artifacts must also be accounted for but have more complicated models that are not presented in this text. By dividing $\mathcal{I}$ by the modeled components, we are able to estimate $\mathcal{H}$:

$$\mathcal{H}(\nu_x,\nu_y) \approx \frac{\mathcal{I}(\nu_x,\nu_y)}{\mathrm{jinc}\left(\frac{\nu_\rho}{\alpha/m_1}\right)\mathrm{sinc}\left(\frac{\nu_x}{w_x/m_2}\right)\mathrm{sinc}\left(\frac{\nu_y}{w_y/m_2}\right)}, \quad \nu_\rho = \sqrt{\nu_x^2+\nu_y^2} \quad . \tag{4}$$

Note that some residual estimation errors remain due to wavefront errors in the microscope assembly and collimator, but these are relatively small. In terms of system design, considering $w_x = w_y$ and fixed, the values of α, f_c, and m_1 must be well chosen for the type of lens to be measured. For a fixed f_c, reducing f_o will reduce the size of the image, using fewer pixels on the detector and driving the system towards an undersampled regime. m_1 must be balanced with f_c to avoid being unnecessarily oversampled or being undersampled. Table 1 provides two sample configurations for these parameters which are suitable for use with a wide variety of camera lenses.

Configuration	α [µm]	f_c [mm]	m_1 [x]
1	12	350	20
2	50	600	50

Table 1. Two possible configurations for an MTF bench.

A third technique, known as the Slanted-Edge method [29–33], involves no auxiliary optics and lacks these corrections in exchange for its own. Because there is no single algorithm for the Slanted-Edge method, the corrections are a feature of the implementation and not general. The technique is comprised of capturing an image of a white-black edge that is tilted by a small angle with respect to the pixel grid in order to achieve oversampling. A discrete derivative produces the Line Spread Function (LSF), and the result is processed as if it were acquired via the slit method.

Substantial material covering the theory of MTF and practice of its measurement can be found in [28, 34].

2.3 MTF Test Routines

In this section, both common and state-of-the-art MTF measurement schemes are discussed in addition to their relevance to wavefront sensing, and their output presented.

2.3.1 Through-Focus MTF

A through-focus MTF measurement is made by mounting an optic to the machine, doing a coarse focusing, and then using automated scanning software to measure the MTF as a function of focus position. The resulting data is two dimensional; a function of frequency and focus. This provides the focus diversity needed by the wavefront sensing algorithm to differentiate different Zernike modes' contributions to the wavefront. The graphical output is shown in Figure 2 with only one of the tangential and sagittal axes plotted for clarity.

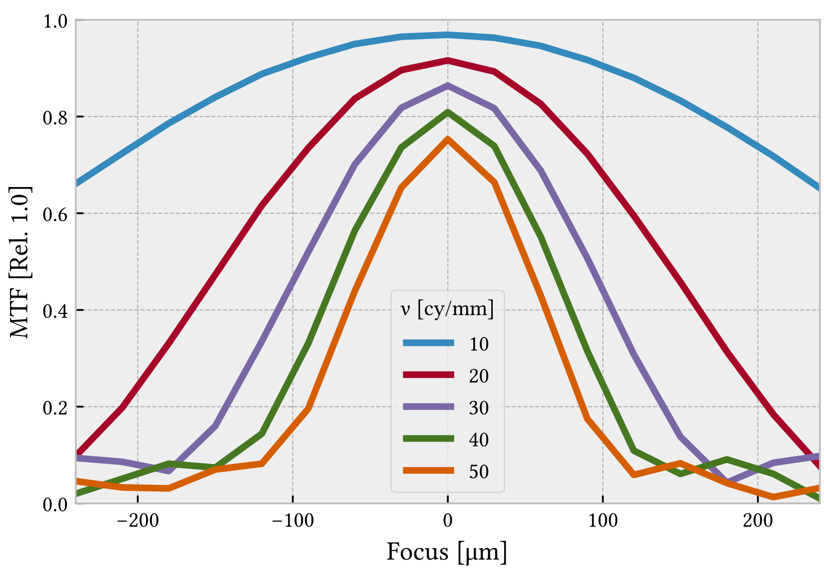

Fig. 2. A plot of through-focus MTF.

2.3.2 MTF vs Field

MTF vs Field is the most common MTF measurement routine. The lens is mounted to the machine and brought to best focus on-axis via a Through-Focus MTF measurement. Without adjusting focus, the MTF is measured on axis and across a linear field of view and reported at an ensemble of spatial frequencies for the tangential and sagittal azimuths.

The resulting data is two dimensional, a function of field and frequency. This form of measurement is not well suited to use as an input to a wavefront sensing algorithm due to its lack of focus diversity. The graphical output is shown in Figure 3.

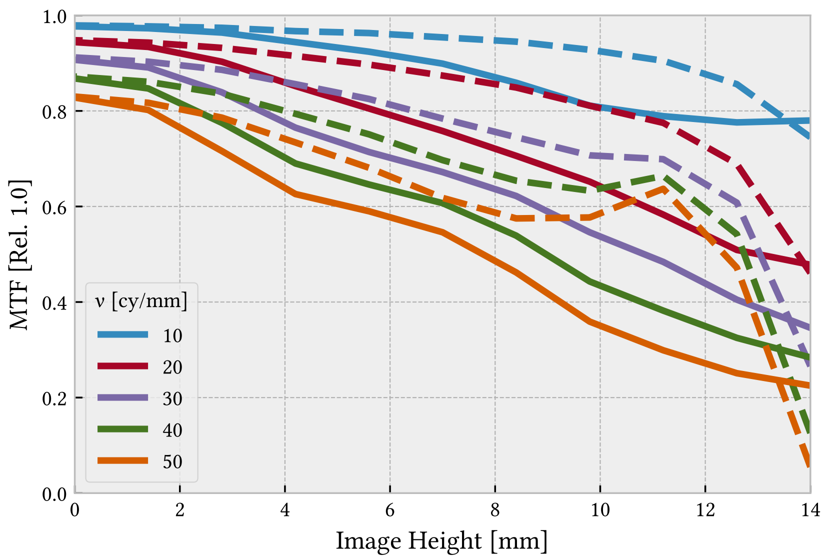

Fig. 3. A plot of MTF vs Field for a wide-angle camera lens at an azimuth of 0° with respect to the x axis of the image plane. Solid lines for the tangential orientation, dashed for sagittal.

2.3.3 MTF Full-Field Display

This is a newer technique described in [35]. It combines MTF vs Field measurements along several image plane azimuths to produce data over the full field of view of an optical system. This allows greater sensitivity to misalignment than an MTF vs Field measurement, which is severely azimuthally undersampled by comparison. The resulting data is of three dimensions: two axes of field and one of frequency. As this is only a 2 dimensional extension of MTF vs Field measurements, the lack of focus diversity makes it similarly unsuitable for wavefront sensing. Figure 4 shows the graphical output for the $\nu = 50$ cy/mm plane.

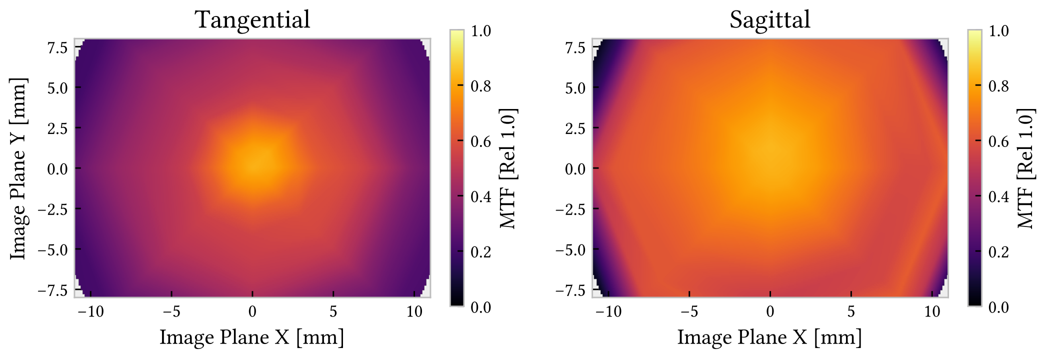

Fig. 4. A pair of MTF FFDs for the tangential and sagittal azimuths.

2.3.4 MTF vs Field vs Focus

MTF vs Field vs Focus is identical to MTF vs Field, except the focus position is re-centered and scanned through focus at each field point. It produces a 3D dataset of field, focus, and frequency. This is highly suitable for wavefront sensing over an extended field of view. If combined with the MTF FFD technique above, this would enable the production of any form of Full-Field Display, as is popular in the freeform optics community, from MTF data in only a few hours. The graphical output is presented in Figure 5.

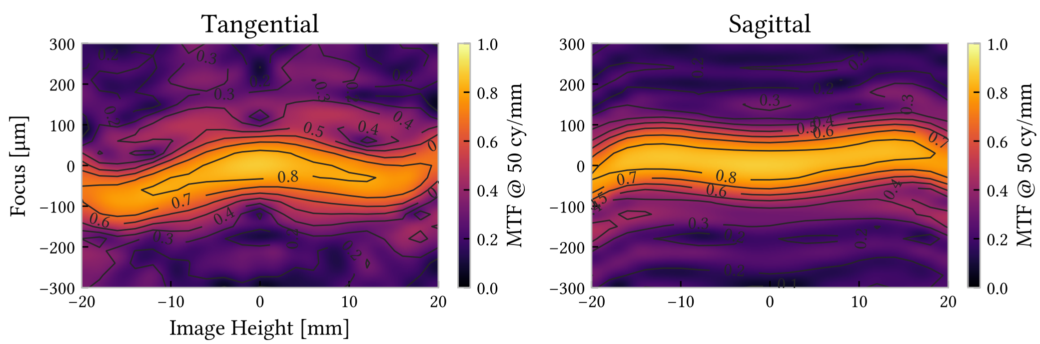

Fig. 5. A display of MTF vs Field vs Focus for each of the tangential and sagittal axes.

2.4 Why Not Use The Image?

The question may fairly be raised, "Why not use the image captured by the MTF bench?" The answer has two components. **(1)**, commercial MTF bench manufactures do not necessarily provide access at all or easily to the image itself. **(2)** use of MTF allows application with Slanted-Edge MTF measurements.

3 Overview

I have extended the wavefront sensing toolbox to include usage of tangential and sagittal MTF data as an input, aided by an above-average amount of focus diversity (21 focus planes vs. 3 or 5 for image-based wavefront sensing). This has involved the extension and performance optimization of existing Fourier optical modeling tools [36] in addition to the from-scratch creation of a wavefront sensing program based on the L-BFGS-B [37] implementation of the Broyden–Fletcher–Goldfarb–Shanno (BFGS) [38] nonlinear optimization routine. The Metropolis-Hastings or "basinhopping" algorithm [39, 40] is also used to facilitate pseudo-global parameter space search.

An exploration among 4 candidates for the best cost or objective function was done. With the optimum cost function, an investigation was done into the ability of this azimuthally limited data to correctly retrieve wavefronts with coma and astigmatism of arbitrary angle in the pupil plane, cases for which it was believed the algorithm would succeed. Additional trials were done with arbitrary combinations of spherical, coma, and astigmatism aberrations up to 8^{th} order.

More than 3,000 simulations were performed to statistically analyze the effectiveness of this method. These simulations include wavefronts that contain spherical aberration, coma, and astigmatism up to 8^{th} radial order but do not include higher azimuthal order aberrations. Additionally, an ensemble of experimental trials were performed. The results from each are presented.

4 Novelty

The novelty of this thesis lies in extending the wavefront sensing toolbox to work with azimuthally limited MTF data. This would open a new application space for wavefront sensing, utilizing commercial MTF instrumentation to collect data used to measure the wavefront of an optical system. This includes but is not limited to usage to diagnose misalignment of camera lenses based on their MTF test results and usage of slanted-edge test results for Earth-facing satellites [41] based on ground-based targets to verify their alignment.

While wavefront sensing has not been limited to the image-based techniques discussed in Section 2.1, with extension to usage of scenes [42], spatial light modulation [43, 44], use of surface plasmons [45], and use of holographic optics [46] among others, it has never been done based on either the full 2D or azimuthally limited MTF.

5 Methods

5.1 Fourier Modeling

The pupil of an optical system can be modeled as

$$P(\xi, \eta) = A(\xi, \eta) \exp[-i\tfrac{2\pi}{\lambda}\phi(\xi, \eta)] \tag{5}$$

where P is the generalized pupil function, A is the transmission function, and ϕ is the phase function [47, 48]. In our case, P models the exit pupil and we can paraxially approximate an optical propagation from an exit pupil plane to an image plane as a Fourier transform:

$$E(x, y) = \mathcal{F}\{P(\xi, \eta)\} \tag{6}$$

where E is the electromagnetic field in the image plane. This is the coherent PSF, whose mod square is the incoherent PSF, which is referred to only as the PSF throughout this document, as we are concerned with the incoherent case.

$$\mathrm{PSF}(x, y) = \left|E(x, y)\right|^2 = E(x, y)\, E^*(x, y) \tag{7}$$

The OTF is the Fourier transform of the PSF,

$$\mathrm{OTF}(\nu_x, \nu_y) = \mathcal{F}\{\mathrm{PSF}(x, y)\} \tag{8}$$

normalized to unity at the origin, where we do not explicitly write the normalization, and the MTF is the magnitude of the OTF:

$$\mathrm{MTF}(\nu_x, \nu_y) = \left|\mathrm{OTF}(\nu_x, \nu_y)\right| \quad . \tag{9}$$

The tangential and sagittal azimuths are two slices of the MTF, whose angles with respect to the cardinal axes are defined by the angle the object makes with respect to the cardinal axes. In this thesis, we will operate under the assumption that object is extended in y, and thus a slice through $x = 0$ yields the tangential, and a slice through $y = 0$ the sagittal MTF. The coordinate systems are shown in Figure 6 and sample of this process can be seen in Figure 7. Note that in cases where the object is extended along a different axis, a simple rotation of the coordinate frame is the only change needed.

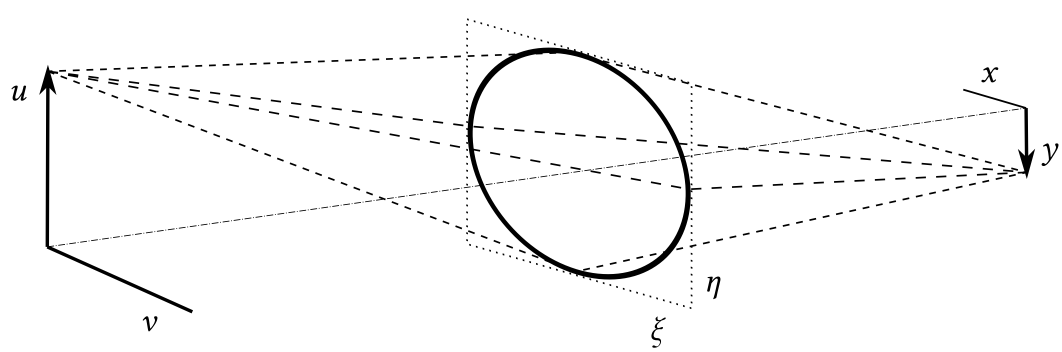

Fig. 6. Schematic of optical propagation. Left to right: object plane, pupil plane, image plane. A bounding box is drawn around the circular pupil to make more clear the Cartesian axes of ξ and η. The u and y axes have arrows denoting the axis the object is extended in. Optical systems are typically modeled as an object extended in y but not x unless otherwise necessary.

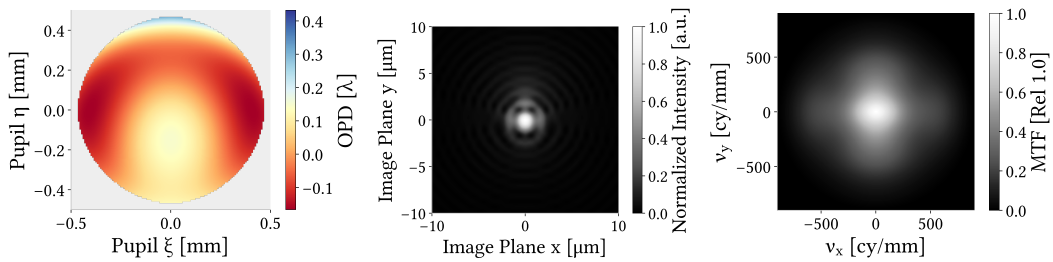

Fig. 7. Left to right: phase error in pupil, PSF, MTF. For our object orientation convention, the upper half of the $\nu_x = 0$ slice is the tangential MTF, and the right half of the $\nu_y = 0$ slice is the sagittal MTF.

5.2 Numerical Implementation

In order to do these calculations numerically, we must correctly sample our functions to avoid aliasing and FFT wrap around. We define Q, the oversampling factor, to be

$$Q = \frac{\text{width}_{\text{array}}}{\text{width}_{\text{pupil}}} \quad . \tag{10}$$

Q need not be an integer, and $Q > 2$ is required to avoid aliasing. We will operate at Q = 2, which gives us Nyquist sampling in the image plane and minimizes array sizes, improving speed. We will also not pad the PSF before computing the MTF, as it is assumed that there is sufficient "empty" space surrounding the PSF to satisfy the windowing requirements of FFTs. This is usually true with relatively low wavefront error.

The Fringe set of Zernike [49] polynomials will be used to model the wavefront in terms of Hopkins [50] wavefront aberration-like components that are orthonormal over the unit circle. Orthonormality gives each term unit variance.

We will focus on the monochromatic case, and choose $\lambda = 0.55$ μm as our wavelength, roughly the mean wavelength for the visible band. An aperture of F/2.8 will be used, though the aperture value is not of great importance, given the magnification discussed in Section 2.2.

[48, chapter 4] provides a detailed tutorial of this process. The python package prysm [36] is made freely available by the author and is used as the embedded modeling tool in this thesis.

5.3 Algorithm Architecture

Image-based and MTF-based wavefront sensing alike can be summarized as a nonlinear optimization problem with embedded parametric optical modeling. There is a rich body of prior and existing art in optimization theory, including numerous algorithms. In Fienup group, it has been found that the L-BFGS-B routine for local parameter space search is typically the best performing, both in terms of the accuracy of the result and speed of convergence.

BFGS optimization methods are a class of quasi-Newton methods that utilize curvature information to take a more direct path to local minima. They evaluate the cost function and its gradient at any given point in parameter space. A search direction is chosen, and the gradient and cost function simultaneously minimized along a line in order to reach a concave up critical point. Because both saddle points and minima are critical points, the process begins again if termination criteria are not met. These criteria include but are not limited to the production of a sufficiently small cost function, or exceeded time limit. The algorithm is shown schematically in Figure 9. The embedded task of comparing the optical model to the measured data is shown in Figure 8.

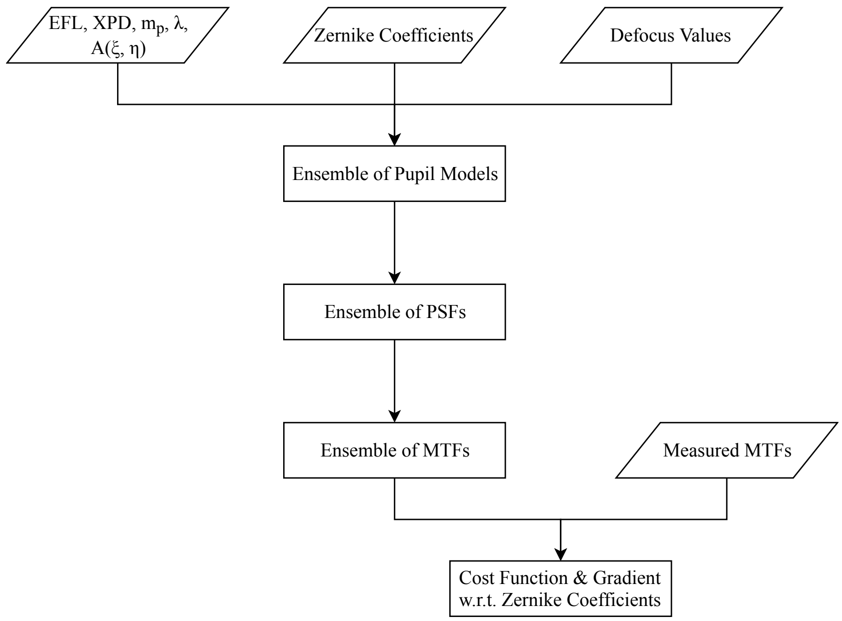

Fig. 8. Schematic summary of the optical model. The focal length or EFL of the system must be given, in addition to the exit pupil diameter or XPD, pupil magnification m_p, wavelength of light λ, and pupil transmission function A. Zernike coefficients are provided via user guess for the first iteration, and the optimizer for the remainder of the process. The defocus values are coupled to the through-focus and through-frequency MTF measurements provided to the algorithm.

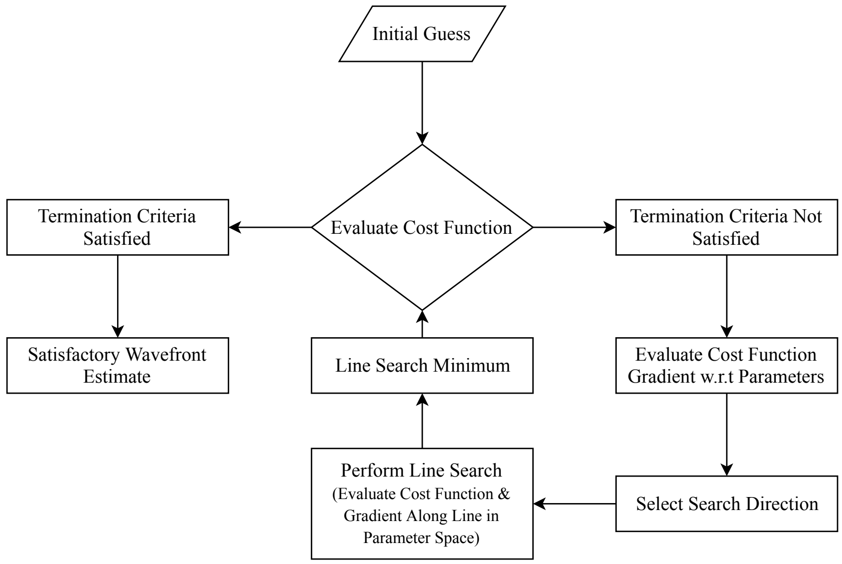

Fig. 9. Schematic of the L-BFGS-B optimization routine. The function effectively seeks critical points, at which it terminates or changes direction. L-BFGS-B is distinguished from BFGS by storing a reduced history of the gradient instead of requiring the full Hessian, which may not be available, or may not fit in memory for large numbers of parameters. Note that Figure 8 is embedded in the diamond that states "Evaluate Cost Function."

It is also well understood that when the input guess is far from the true parameters, a local search such as L-BFGS-B is susceptible to getting trapped in local minima. To avoid this problem, a pseudo-global method of optimization is required. For this task, I have selected to use the basinhopping [39, 40] algorithm. Briefly, this algorithm tracks the landscape of the cost function through parameter space during a local minimization attempt with another method, such as L-BFGS-B used here. After a minima is reached, a perturbation is made of the parameters. If all tests pass, including but not limited to a not-too-large cost function value, lack of violation of parameter space boundaries, and not lying very close to already explored regions of parameter space, a new local minimization is performed. Across an ensemble of local minimization attempts, the cost function landscape is mapped and the global minimum (hopefully) found. Basinhopping is also known as the Metropolis-Hastings algorithm, and simulated annealing is a variation on it where the temperature parameter, which controls the permissible uphill motion of a perturbation, is slowly reduced over time.

Three subsets of the Fringe Zernike polynomials with index starting at one are used in this thesis and will be annotated by the variable $\mathbf{W}$, the first two are:

$$\mathbf{W}_1 = \{\, Z4, Z9, Z15, Z25 \,\}, \tag{11}$$

$$\mathbf{W}_2 = \{\, Z4, Z5, Z6, Z7, Z8, Z9, Z12, Z13, Z14, Z15, Z16, Z21, Z22, Z23, Z24, Z25 \,\} \tag{12}$$

where the former, $\mathbf{W}_1$ includes only terms of azimuthal order zero, corresponding to focus and various kinds of spherical aberration up to 8$^{\text{th}}$ order. The latter, $\mathbf{W}_2$ includes additional comatic and astigmatic terms, likewise up to 7$^{\text{th}}$ and 6$^{\text{th}}$ order respectively. A third, $\mathbf{W}_3$, is equivalent to $\mathbf{W}_2$ with the addition of Z10 and Z11; trefoil. These polynomials are enumerated in Appendix A.

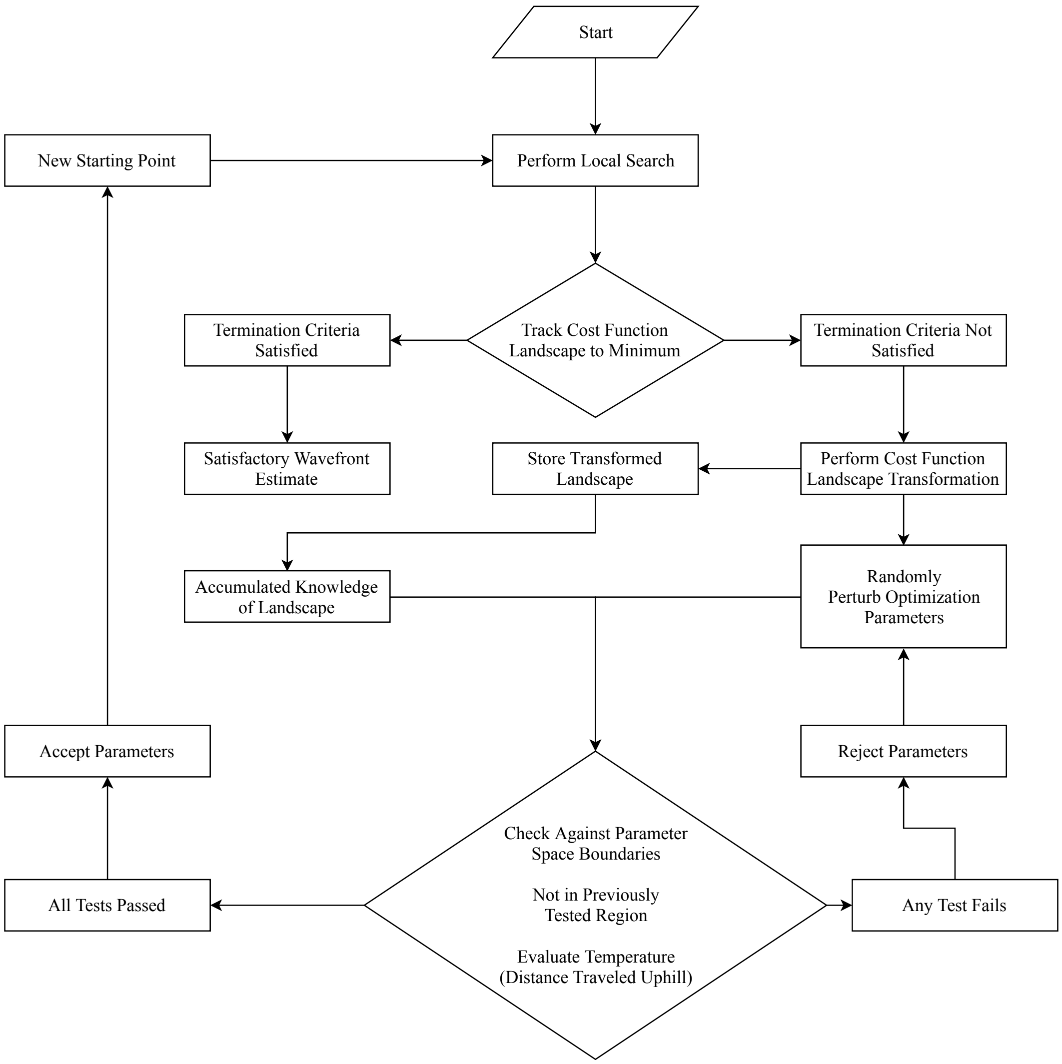

Fig. 10. Schematic of the basinhopping pseudo-global optimization routine. The algorithm has many moving parts, but is essentially a means of sampling a complicated function of many variables without wasting time evaluating regions of parameter space very far from the ideal (truth) values. Note that Figure 9 is embedded in the box that states "Perform Local Search."

5.4 Cost Function Design

The cost function serves the purpose of turning a multivariate function into a scalar. In wavefront sensing, the term *data consistency* is often used to describe a family of cost functions that are designed to have a value of 0 when experimental data and the model match, and values strictly greater than zero when they do not. It is important that the cost function not be able to take on negative values as the optimizer will seek to find the smallest possible cost function value, and we wish for the minimum cost function value to have the best agreement between the model and data (and hopefully, the true and retrieved wavefronts).

In the case of image-based wavefront sensing, there are no general assumptions that can be made of the image data and the geometry of its intensity. In contrast, MTF is naturally constrained by the diffraction limit and thus smaller at high frequencies than low frequencies, though this guarantee is not necessarily monotonic. The difference of two small numbers at high frequency tends to be smaller than the difference of two potentially larger numbers at a lower frequency; division of the data by the diffraction limit will serve as an equalizer, undoing the deemphasis on high frequencies caused by the diffraction limit. The MTF data is numerical in nature, so we define it to be an array of N_ν elements indexed by n, with frequency spacing $\Delta\nu$, chosen arbitrarily to be 10 cy/mm, which gives 50 to 100 samples in the frequency axis for the F/#s of interest. This would be adjusted for very fast or very slow lenses. We also define the following symbols which will be used in the cost function:

$$
\begin{aligned}
\mathbf{D} &\equiv \mathbf{D}(n\Delta\nu) && \text{data} \\
\mathbf{M} &\equiv \mathbf{M}(n\Delta\nu) && \text{model} \\
\mathbf{L} &\equiv \mathbf{L}(n\Delta\nu) && \text{diffraction limit}
\end{aligned}
$$

each has a $_\mathrm{T}$ and $_\mathrm{S}$ variant to denote the tangential and sagittal azimuths, except for $\mathbf{L}$ which is rotationally invariant. We will use $\mathbf{D}$, $\mathbf{M}$, and $\mathbf{L}$ for shorthand of their fully explicit forms. We will also make explicit two ways to compute the difference, or distance between two datasets. The first is the Manhattan or cityblock distance, which is simply the sum of the magnitude of differences of the data:

$$d_\mathrm{M} = \sum_n^{N_\nu} |\mathbf{D} - \mathbf{M}| \quad . \tag{13}$$

The second is the sum of the square of differences, which is equivalent to a Euclidean distance, less the square root:

$$d_\mathrm{SSD} = \sum_n^{N_\nu} (\mathbf{D} - \mathbf{M})^2 \tag{14}$$

The removal of the square root eliminates a cusp at the origin that prevents the function from being everywhere differentiable.

It was initially proposed to use the same type of cost function as is commonly used in image-based wavefront sensing,

$$C_1 = \sum_\text{focus} \sum_n^{N_\nu} \underbrace{(\mathbf{D}_\mathrm{T} - \mathbf{M}_\mathrm{T})^2 + (\mathbf{D}_\mathrm{S} - \mathbf{M}_\mathrm{S})^2}_\text{core} \tag{15}$$

which is the unnormalized variant of d_SSD, with the core annotated. The four options analyzed have a core which is either normalized by diffraction or not,

$$\frac{\mathbf{D} - \mathbf{M}}{\mathbf{L}} \qquad\qquad (\mathbf{D} - \mathbf{M})$$

and utilize either d_{M} or d_{SSD}. All options considered share the coefficient

$$\frac{\Delta\nu}{\mathrm{N}_{\mathrm{focus}}\nu_{\mathrm{max}}}$$

whose numerator converts the summation over ν into a numerical integration, and whose denominator normalizes both the focus and frequency dimensions. With this normalization, L^1 norm-type cost functions are similar in magnitude to the residual RMS wavefront error (RMS WFE) which allows the cost function to be easily used to estimate the accuracy of the result.

500 sample wavefronts were synthesized by drawing from a uniform random probability density function spanning the range $[0, 0.25]$ and assigning this value to the coefficient of Z9. Z16 and Z25 were then set to negative one half of Z9, and positive one fourth of Z9 respectively, forming a geometric series. Two distinct random Gaussian distributions of mean 1 and standard deviation 0.25 were generated, and the values of Z16 and Z25 were multiplied by these values to add noise. This produces sets of { Z9, Z16, Z25 } variables that have a semi-fixed relationship that is realistic, but are allowed to vary to avoid absolutely enforcing a particular relationship on the Zernike modes.

Each sample wavefront was used to synthesize through-focus MTF data across 21 focus planes corresponding to $\pm 2\lambda$ PV of wavefront defocus and with ν ranging from 10 to ν_c in steps of 10 cy/mm. Where ν_c is not an integer multiple of 10, the highest value modulo 10 that is lower than ν_c is used. E.g. if $\nu_c = 584$ cy/mm, sampled values of ν would range from 10 to 580 cy/mm in steps of 10. Monochromatic light of wavelength 550 nm was used. The wavefront sensing algorithm was configured to use the relevant cost function, an initial guess of no wavefront errors, and up to 25 random starting guesses generated by the basinhopping algorithm. Optimization would terminate early if a sufficiently low cost function value, corresponding to about better than $\lambda/1000$ RMS, was achieved. No noise is included in these simulations. Stopping at this level leaves an order of magnitude budget for experimental uncertainty and noise before the 1 nm RMS level of accuracy is crossed.

Each of the four candidate cost functions were applied to all 500 of these simulations and the results compared. The full dataset for all four is shown in Figure 11. It is interesting that the sum of square of differences error metric has a looser grouping, spanning almost two orders of magnitude. It appears that this metric is easily able to dodge local minima far from the truth, but cannot escape local minima near the global minimum.

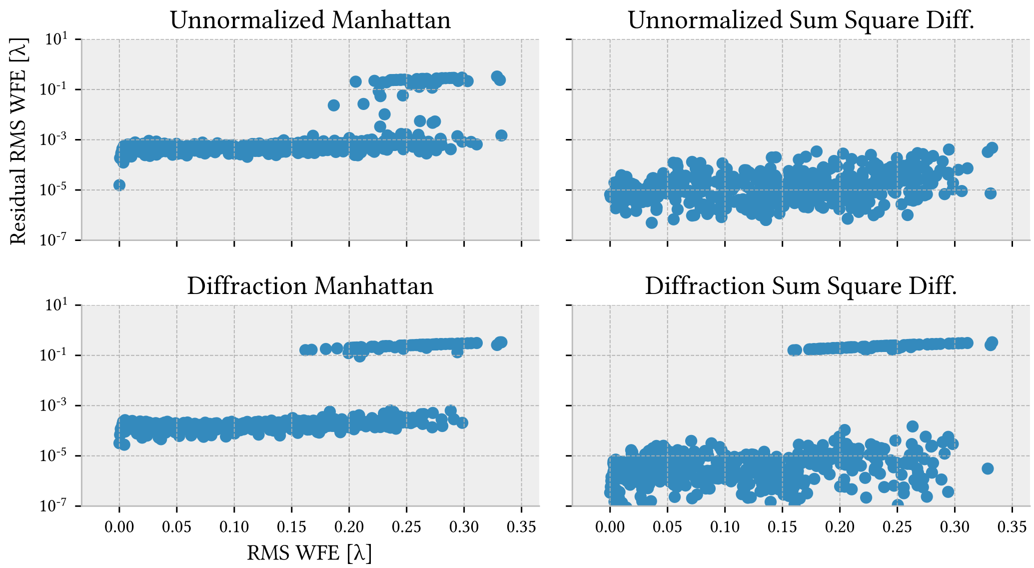

Fig. 11. Scatterplot of residual RMS WFE vs true RMS WFE on linear-log scale for each of the four cost function schemes. Attention should not be paid to the lower overall residual RMS WFE in some cases; most cases terminated optimization for finding a sufficiently low cost function, which was not finely tuned to cause termination at the same residual RMS WFE level.

The datasets were collapsed into just their residual RMS WFE values, and Gaussian Kernel Density Estimation (KDE) [51, 52] used to fit a probability density function. This is a more robust form of producing a histogram, and allows the process to be done with a logarithmic x axis. The results are presented in Figure 12 and show that the non diffraction normalized sum of square of differences cost function scheme performs best, having the largest area under the curve in the region over $[0, \lambda/1000]$ residual RMS WFE. The success rates of the various schemes are summarized in Table 2. Normalizing by the diffraction MTF did not improve the robustness of the optimizer, in both cases actually decreasing the success rate. The sum of squares of errors metric is more able to avoid local minima and retrieve the true wavefront.

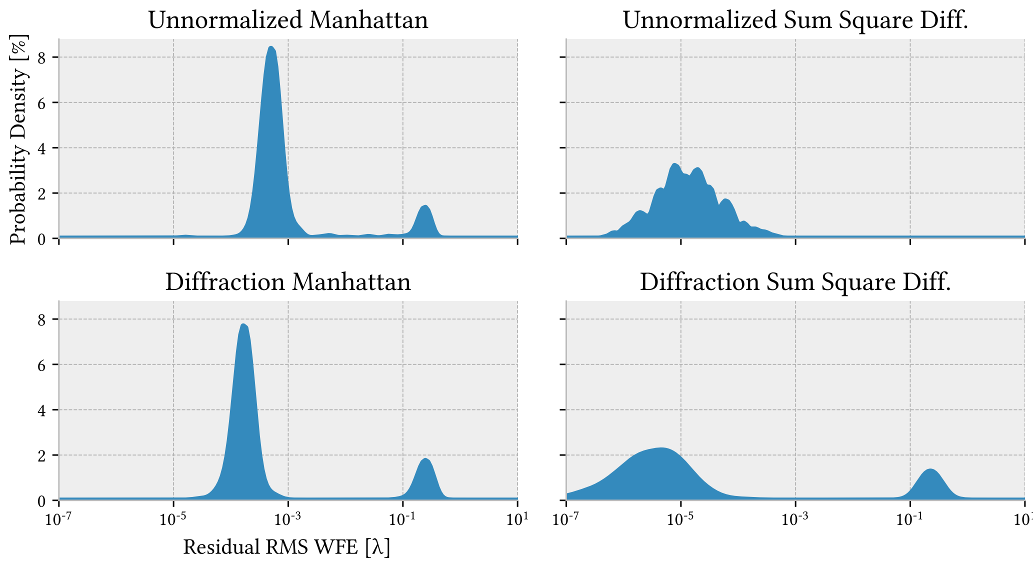

Fig. 12. Residual RMS WFE probability density for each cost function scheme. Attention should not be paid to the lower overall residual RMS WFE in some cases; most cases terminated optimization for finding a sufficiently low cost function, which was not finely tuned to cause termination at the same residual RMS WFE level.

Distance	Normalized	Success [%]
Manhattan	No	89.2
Manhattan	Yes	84.4
sum of squares	No	100
sum of squares	Yes	84.0

Table 2. Summary of success rate of various cost function designs. The non diffraction normalized sum of square of differences performs best.

As a result of this study, the following cost function was adopted for the remainder of the thesis:

$$C = \frac{\Delta\nu}{\mathrm{N_{focus}}\nu_{\mathrm{max}}} \sum_{\mathrm{focus}} \sum_{n}^{N_\nu} |\mathbf{D}_\mathrm{T} - \mathbf{M}_\mathrm{T}|^2 + |\mathbf{D}_\mathrm{S} - \mathbf{M}_\mathrm{S}|^2 \quad . \tag{16}$$

It can be summarized as the square of the normalized area between the measured and modeled MTF. Its square root will be used on most plots concerning the cost function, as that quantity is more proportional to the residual RMS WFE.

6 Results

6.1 Rotationally Invariant Aberrations

The wavefront generation routine described in subsection 5.4 was repeated, with the maximum RMS WFE limit increased to 0.35λ. The cost function shown in Eq. 16 was applied to all 500 trial wavefronts. A sample wavefront is shown in Figure 13 as well as the retrieved wavefront, their difference, and the cost function history.

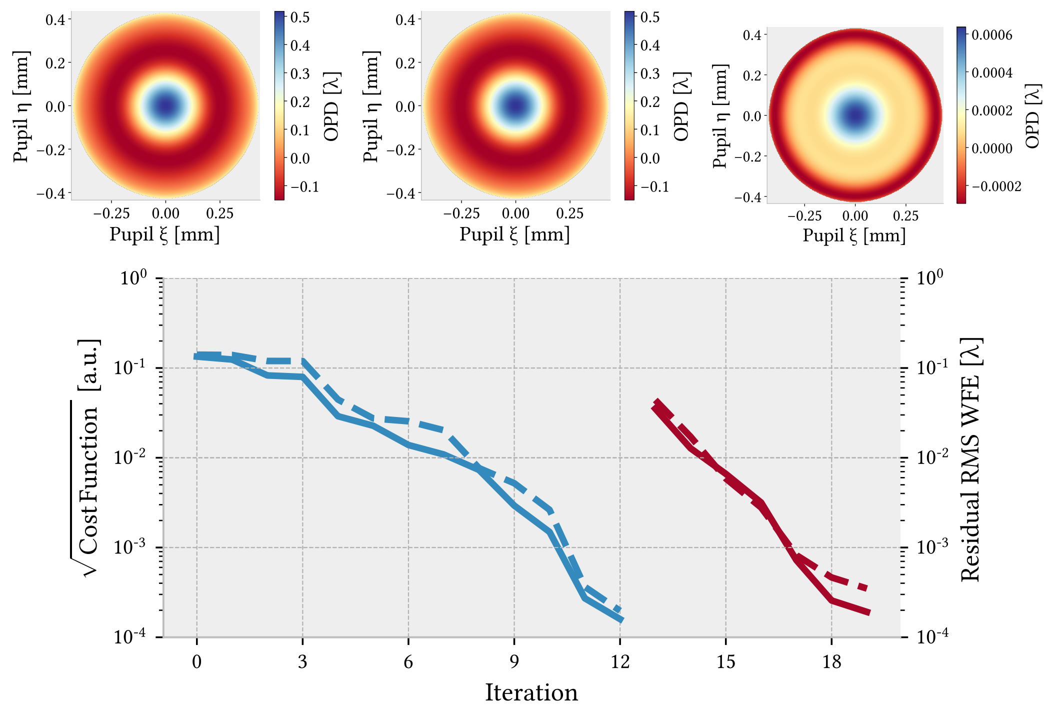

Fig. 13. Top, left to right: true wavefront, retrieved wavefront, difference. Bottom: cost function (lefthand y axis) and residual RMS WFE (righthand y axis) history. This result belongs to a class that converge extremely quickly, requiring only 2 iterations of basinhopping.

With the increased wavefront error of the simulations, the success rate dropped to 84.6%, with mixed results in the range [0.247,0.333] λ RMS. Some of these RMS WFE values were possible in the cost function design trials but did not occur. It seems that the algorithm is not guaranteed success in this range, unlike the range spanned by about [0, 0.25] λ RMS. In cases where success did not occur, failure was absolute, with the optimizer failing to converge at all and the cost function value staying very large.

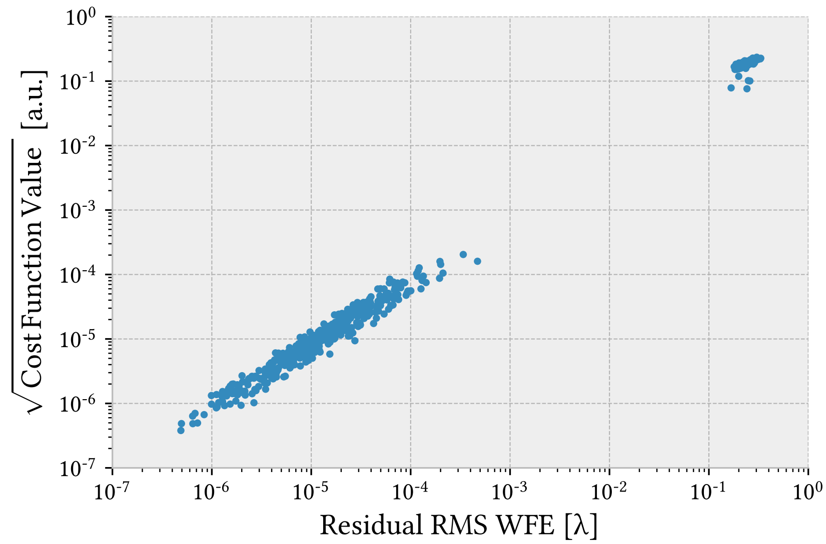

Fig. 14. Scatterplot of the square root of the cost function value vs residual RMS WFE. A nearly perfect proportionality can be observed between the two. Note as well that a large void exists between cases of failure and success.

Figure 14 shows a scatterplot of the square root of the cost function vs the residual RMS WFE for the lowest cost function value found in all 500 trials. Near perfect proportionality is observed, implying $\sqrt{C}$ is a very good predictor for the residual RMS WFE.

In the example shown, basinhopping has not yet distinguished itself in any meaningful way from random starting guesses, a practice that has some heritage in image-based wavefront sensing. With only two iterations of the algorithm, each including many iterations of the local search, it is guaranteed that the transformed cost function landscape is too incomplete to be used to filter out new guesses. The step size chosen is 0.02 λ RMS for the perturbation of each parameter, which is fairly well tuned to maneuver around local minima near the global minimum.

When the starting guess is far enough from the global minimum that L-BFGS-B will become trapped in a distant local minima, this step size may be inappropriate; Figure 15 shows the same type of plot as Figure 13, but for where the guess is very far from the truth. Here, basinhopping makes 21 random starts and accumulates enough knowledge of the cost function landscape to move into a region within L-BFGS-B's capture range, and optimization is successful. All of the 500 trial wavefronts fall into one of two cases: one where less than three random starts guesses were needed, in which case basinhopping is neither helpful nor harmful, and one where more than fifteen random starts were needed, and basinhopping's accumulated knowledge is likely to have been helpful.

The results are summarized in Figure 16, which shows the clear demarcation between success and failure (left or right of 10^{-2}) while also showing the overwhelming success rate of MTF-based wavefront sensing for these types of wavefronts.

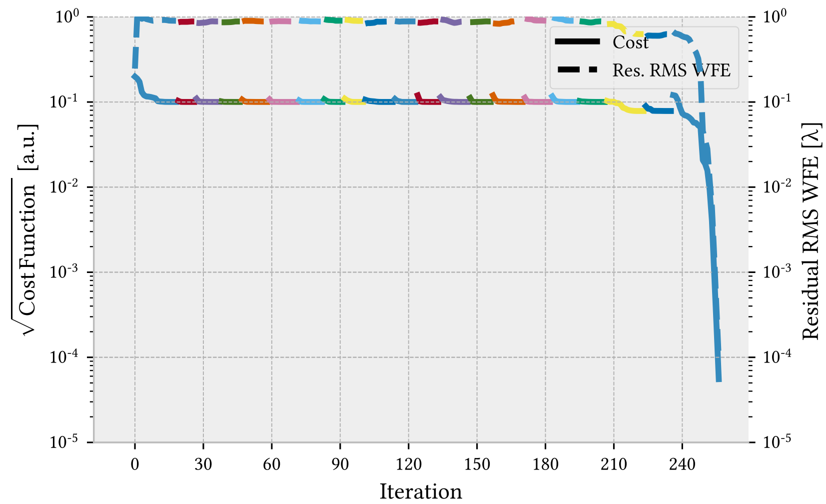

Fig. 15. Cost function and residual RMS WFE history, as in Figure 13, but for a wavefront where L-BFGS-B would become trapped in a local minima severely far from the global minimum. Note that $\sqrt{C}$ and the residual RMS WFE differ when very far from the truth, but converge to similar values near truth.

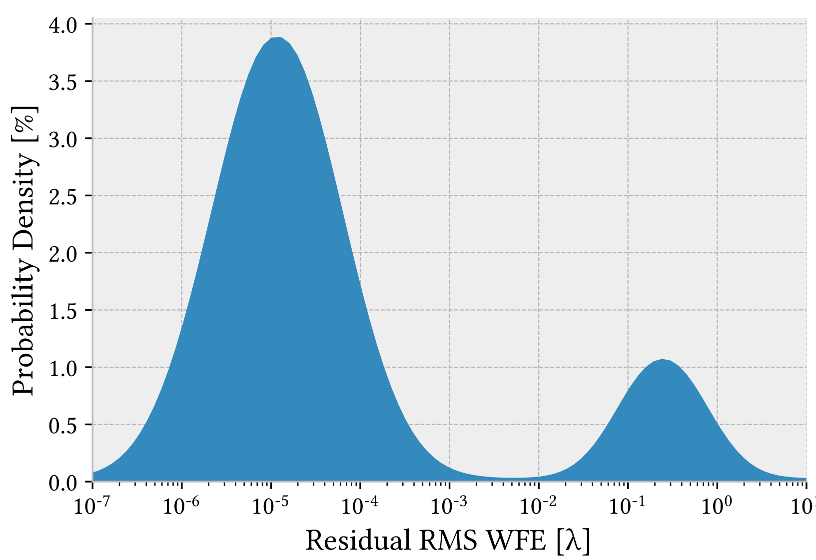

Fig. 16. Residual RMS WFE probability density of the 500 trials of mixed spherical aberration over a range of [0, 0.35] λ true RMS WFE.

6.2 Comatic Aberrations

Fringe Zernike terms with azimuthal order greater than zero occur in pairs, e.g. Z7 and Z8 are primary coma. Each pair can alternatively be characterized by a magnitude $|\mathbf{Z}|$ and angle $\Theta(\mathbf{Z})$:

$$|\mathbf{Z}| = \sqrt{Z_x^2 + Z_y^2} \qquad\qquad \Theta(\mathbf{Z}) = \arctan\left(\frac{Z_y}{Z_x}\right) \quad . \tag{17, 18}$$

Given the magnitude and angle, we can easily express the x and y components:

$$Z_x = |\mathbf{Z}| \cos(\Theta) \qquad\qquad Z_y = |\mathbf{Z}| \sin(\Theta) \quad . \tag{19, 20}$$

Coma with magnitude 0.05, 0.1, or 0.2 λ RMS and an azimuth between 0 and 90 degrees in 5° steps was combined with primary spherical aberration (Z9) having magnitude 0, 0.05, 0.1, and 0.2 λ RMS. This produced a total of 228 unique wavefronts. Due to the symmetry of these terms, angles from 90 to 180 degrees are anti-symmetric to angles from 0 to 90 degrees, and need not be evaluated. Three samples are shown in Figure 17. This allows us to probe if coma and spherical aberration can be distinguished as a function of the orientation of the non rotationally invariant aberration in the pupil, as well as a function of their relative magnitudes. Figures 19-21 show the residual RMS WFE as a function of the angle of coma in the pupil and a fixed amount of primary spherical aberration. On each plot, success may loosely be defined as dropping off of the y axis, whose minimum is $\lambda/1000$ RMS. This is an exceptionally good result in practice, but due to noise not being included in these simulations, a margin is left before reaching the 1 nm RMS level of accuracy. $\mathbf{W}_3$ is used during optimization, which allows both astigmatism and trefoil to be estimated as well as higher radial order terms, that are not present in the data.

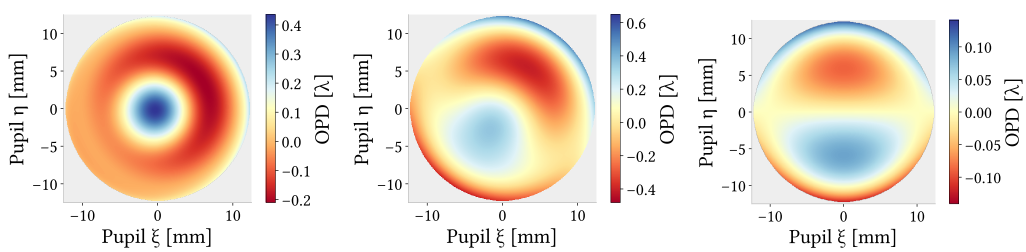

Fig. 17. Three sample wavefronts from the batch of 228. Note that the orientation of coma is free to vary, there are cases where spherical aberration dominates (left), where coma and spherical aberration are roughly equal (center), and where coma dominates (right).

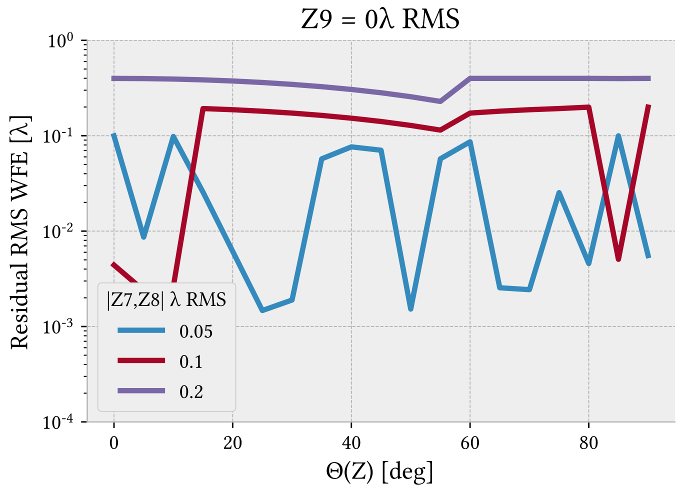

Fig. 18. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is no spherical aberration.

The results with no spherical aberration are somewhat chaotic. There is a substantial dependence on the angle of coma in the pupil plane that determines whether optimization will be successful or not. The angles of 0, 45, and 22.5 degrees have some measure of specialness, as they are when the difference between the tangential and sagittal MTF caused by coma are maximized, globally minimized, and locally minimized respectively.

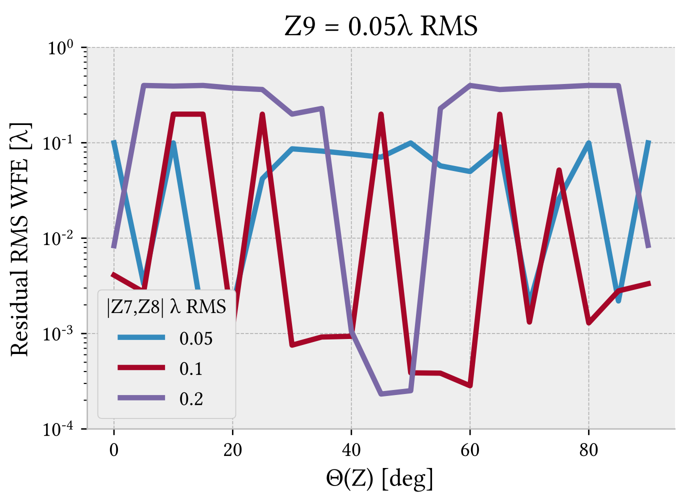

Fig. 19. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/20 λ RMS of spherical aberration.

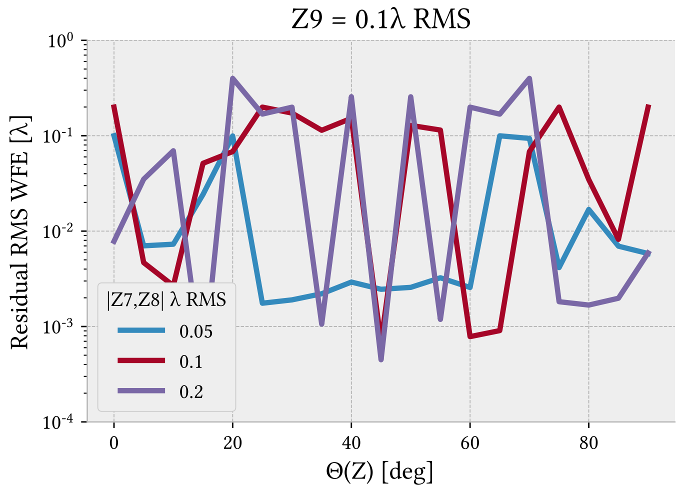

Fig. 20. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/10 λ RMS of spherical aberration.

When spherical aberration is reduced, the results are increasingly chaotic. When there is only $\lambda/20$ coma RMS, the optimizer appears to be able to accurately estimate the underlying spherical aberration very well. When the amount of coma is raised to $\lambda/10$ or $\lambda/5$, success or failure is very chaotic, and there appear to be a tremendous number of local minima due to mis-estimation of higher order coma terms.

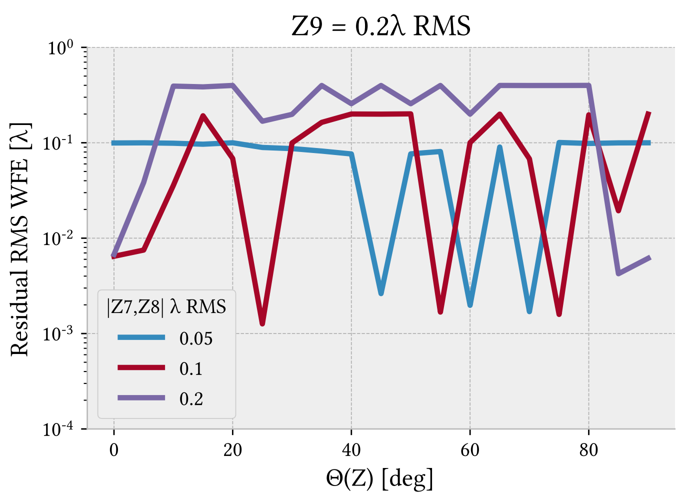

Fig. 21. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/5 λ RMS of spherical aberration.

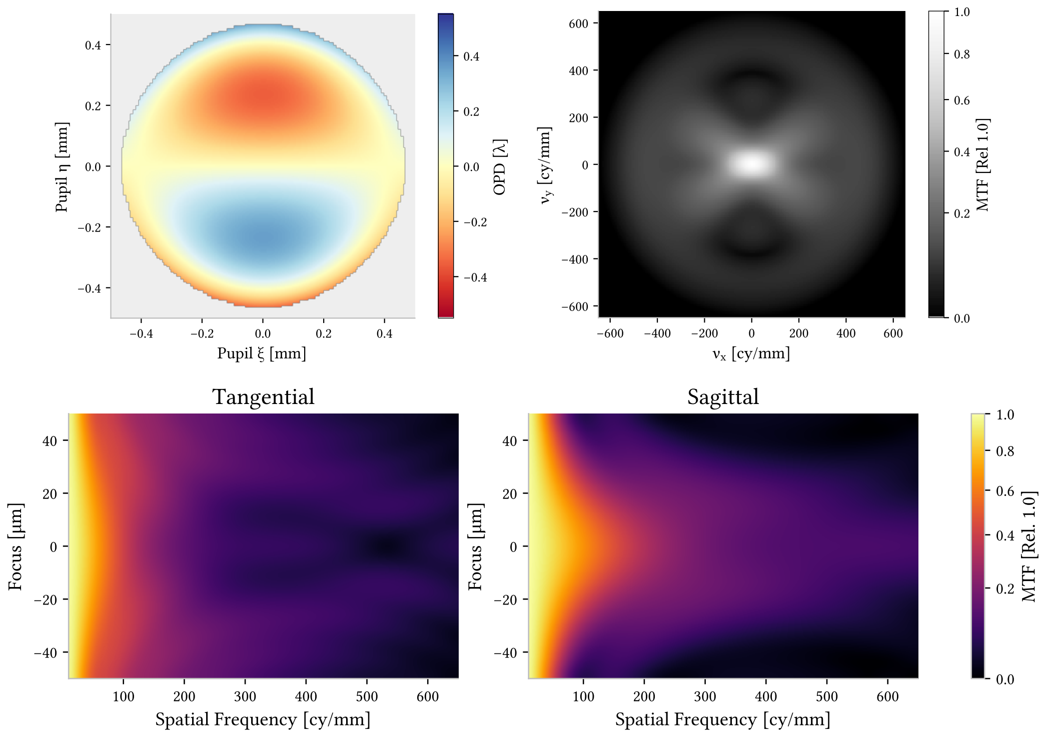

Fig. 22. Top left: wavefront, top left: 2D MTF. Bottom: through-focus tangential and sagittal MTF. Data for 0.2 λ RMS of primary coma aligned with the cardinal axes. Note the x-like pattern in the 2D MTF, and inverse-like relationship between the tangential and sagittal MTF through focus.

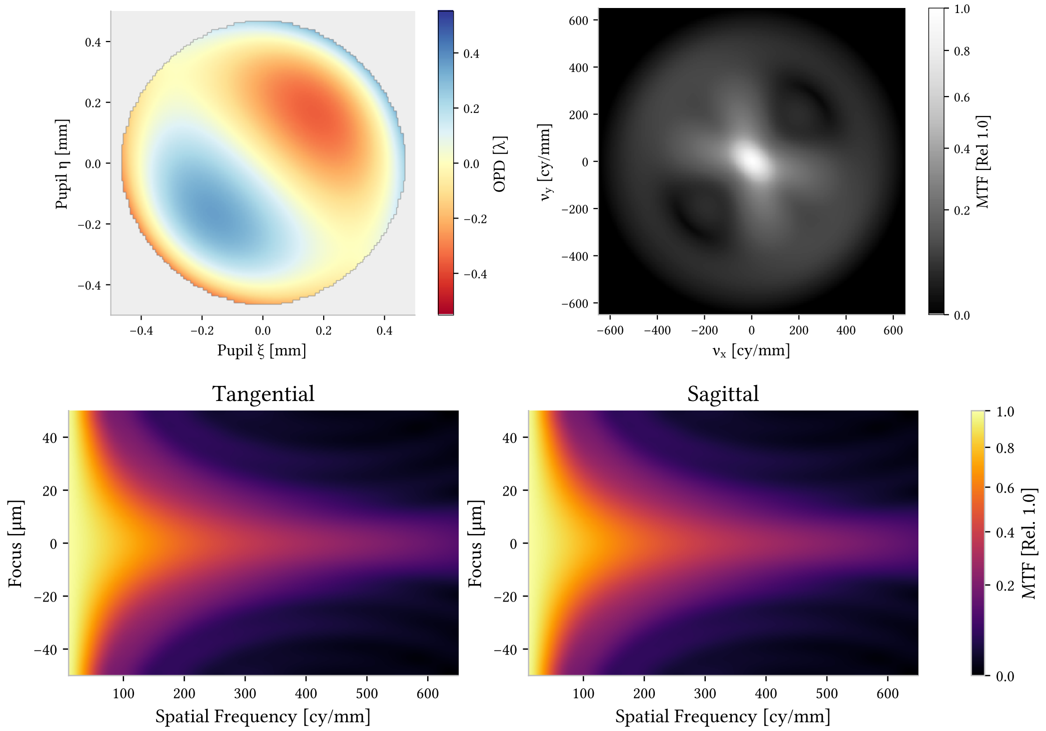

Fig. 23. Equivalent to Figure 22 for 0.2 λ RMS of primary coma aligned at 45° to the cardinal axes. The through focus MTF lacks curvature that would be associated with a term of even radial order, but is identical for T & S. The optimizer struggles to reconcile this with the Zernike terms, since secondary coma at the same orientation appears quite similar.

In the case where the coma is aligned to the cardinal axes, the inverse-like relationship between the tangential and sagittal MTF is reminiscent to the twin through-focus peaks of astigmatism, and the optimizer at times has difficulty distinguishing the two aberrations. In the case where the coma is 45° off of the cardinal axes, the MTF is more similar to the diffraction MTF, with enhanced depth of field at lower spatial frequencies. Because the x and y elements of the coma must be covariables to improve the match between the model and data, but adjusting either alone would worsen the match, the optimizer has difficulty finding the solution. There are also many local minima caused by mis-estimation with higher order terms.

In the case where coma and spherical are both present, and their magnitudes take a reasonable ratio of e.g. 4:1 or less, the characteristic features of spherical aberration – the curvature of the MTF vs frequency vs focus – and extra features related to coma (extended depth of field at lower frequencies, asymmetric "fringes" at extremes of focus) are visible. More specifically, these features are related to the odd radial order of coma, which is shared by trefoil. In this case, mis-estimation occurs due to trefoil in addition to higher order comatic terms. There is no apparent difficulty estimating astigmatism, as it is always near zero in the results.

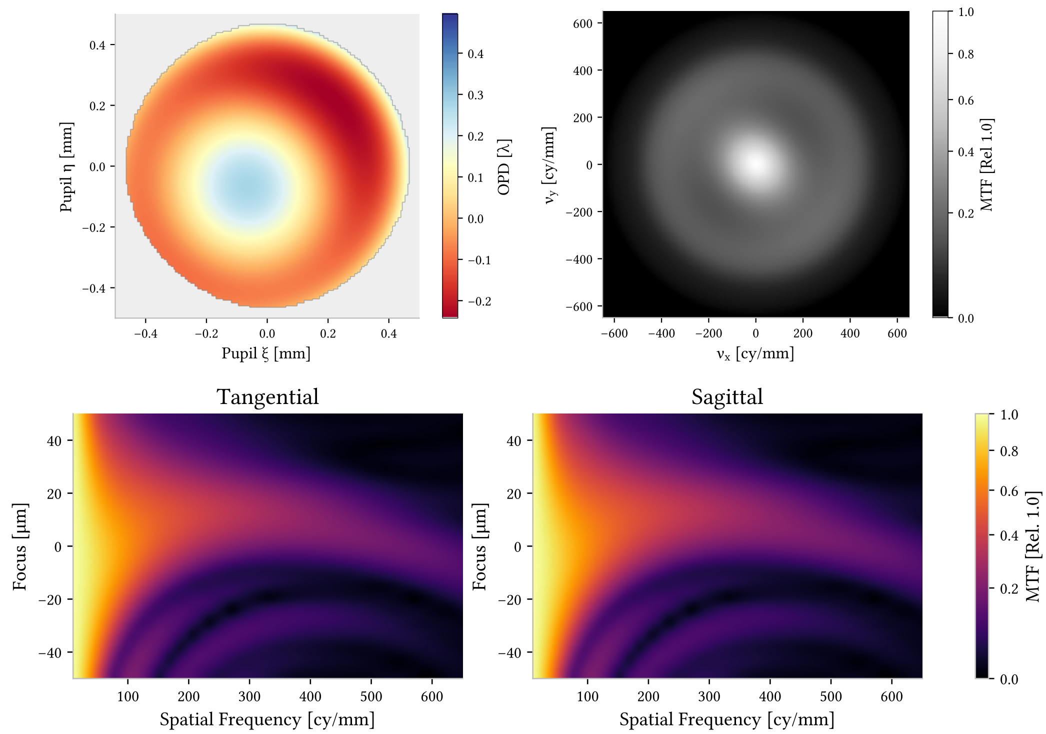

Fig. 24. Equivalent to Figure 22 for 0.1 λ RMS of primary coma aligned at 45° to the cardinal axes with 0.1 λ RMS of primary spherical. The MTF has the curvature characteristic of a Zernike mode of even radial order, but the ripples at the extremes of focus are different to if there were no coma. These ripples allow the optimizer to distinguish the aberrations.

Figure 25 shows the probability density of the residual RMS WFE. Figure 26 shows the probability density of the final cost function value. Because the dip in the distribution is not as sharply defined as the residual RMS WFE case, it can be inferred that there are some uniqueness issues with this limited Tangential and Sagittal data.

Figure 27 shows an example of the input, output, difference, and trajectory of the optimizer for only coma of relatively high magnitude at a small angle to the cardinal axes. This is one of the relatively successful cases, converging to $\lambda/195$ RMS. While it is categorically a success, the cost function and residual wavefront error can be observed in the history plot, as well as the ambiguity between low and higher order coma. Note that from the 0^{th} to the 225^{th} iteration, the optimizer repeatedly converges from different basinhopping starting points, but the residual RMS WFE does not change meaningfully. It is only from iteration 225-420 that the optimizer and residual RMS WFE simultaneously converge, but the match is still imperfect. Note that in the final four iterations of basinhopping (x-axis about 440-500^{th} iteration) the optimizer is exploring a different region of parameter space and returns to the initial condition of cost function convergence but no convergence in terms of residual RMS WFE. The overall success in these trials was 32.5%, defined as when the residual RMS WFE is less than $\lambda/100$. This is a reduced standard compared to that used with spherical aberration. The final cost function values imply a success rate of 78.9%. This demonstrates the severe ambiguity issues associated with coma.

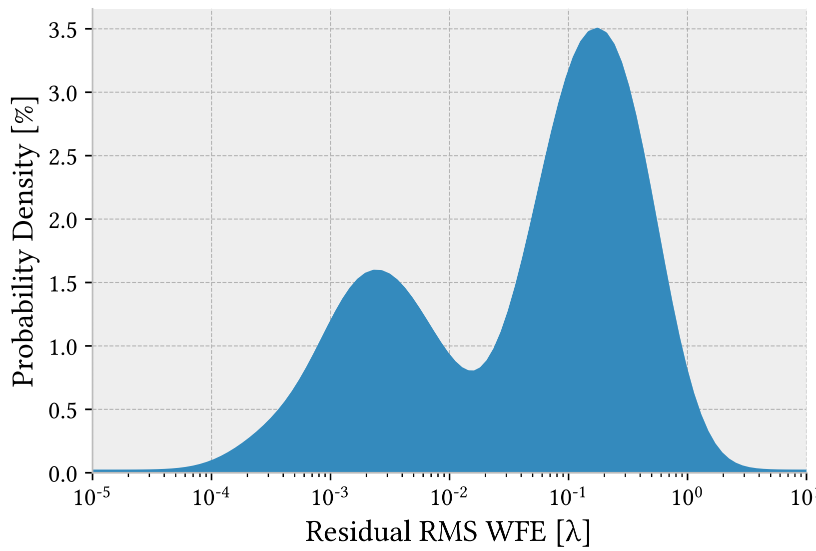

Fig. 25. Probability density of the residual RMS WFE for the 228 total trials. There is a relatively good ability to distinguish success (left mode) from failure (right mode).

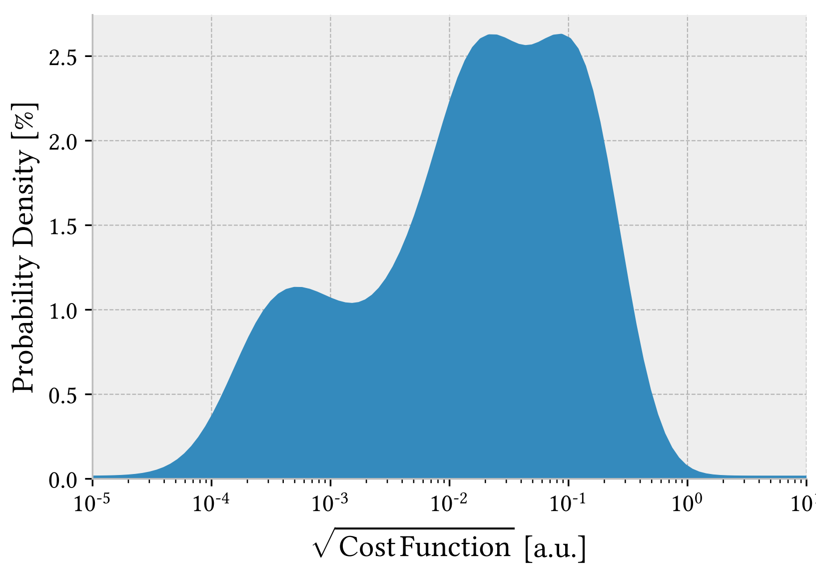

Fig. 26. Probability density of the final cost for the 228 total trials. There is reduced ability to distinguish success and failure compared to the residual RMS WFE, implying some uniqueness problems.

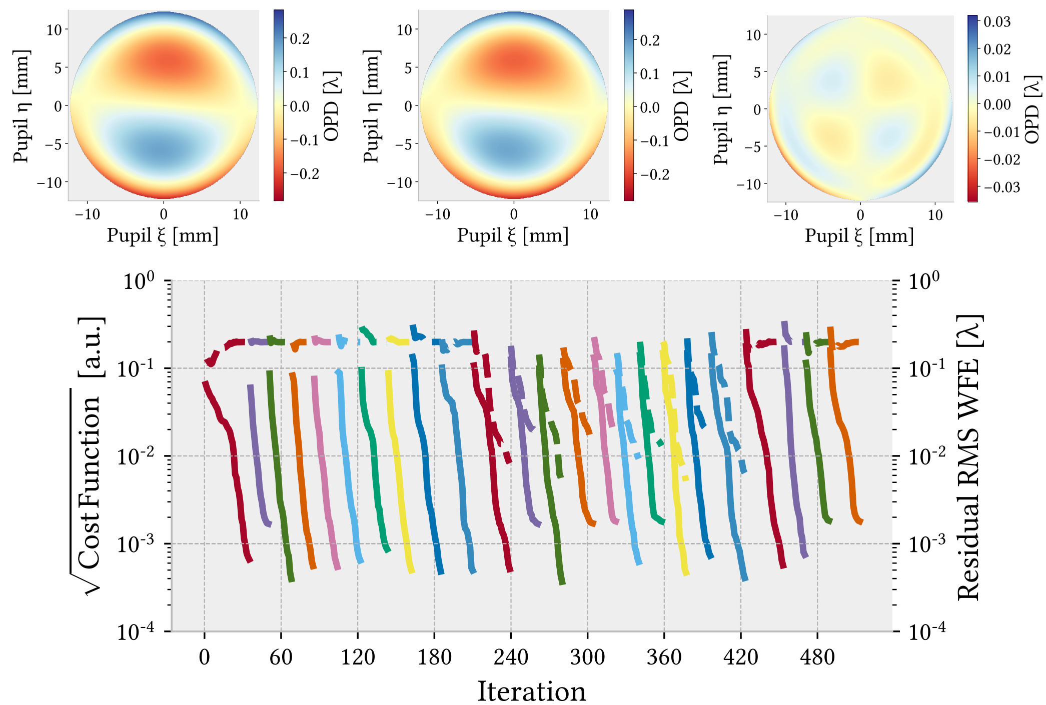

Fig. 27. Top, left to right: true wavefront, retrieved wavefront, difference. Bottom: optimizer history, with $\sqrt{C}$ on the lefthand y axis, solid lines, and residual RMS WFE on the righthand y axis, dashed lines. The estimate is quite high quality, and has only residual higher order coma. This is an example of a relatively successful convergence, and has a difference of 0.00513 λ RMS.

6.3 Astigmatic Aberrations

The same process was used as in the comatic case to generate 228 astigmatic wavefronts, which have varying combinations of angle of astigmatism in the pupil plane, magnitude of astigmatism, and magnitude of spherical aberration. The results are similarly compiled and presented in Figs. 28-31. More general success is observed. This is due to the unique property of astigmatism to simply shift the MTF through focus along its axis. This effect is highly distinct to that of other aberrations. The optimizer appears to struggle when overall wavefront error is high. This may because the region of high MTF is sufficiently shifted to not at all overlap with the starting guess, and thus adjusting astigmatism by a small delta will not improve the cost function in a significant way. Figure 32 shows the distribution of residual RMS WFE and $\sqrt{C}$ values. Significant disagreement between them can be seen; this is because the solutions are highly ambiguous when far from the truth.

The success rate, matching the requirement of $\lambda/100$ residual RMS WFE or better, is 64.5%. Evaluated via the cost function, this only rises to 71.1%. This implies astigmatism does not have the same ambiguity issues coma does. It is likely that these values would be highly improved by an embedded algorithm to produce better starting guesses.

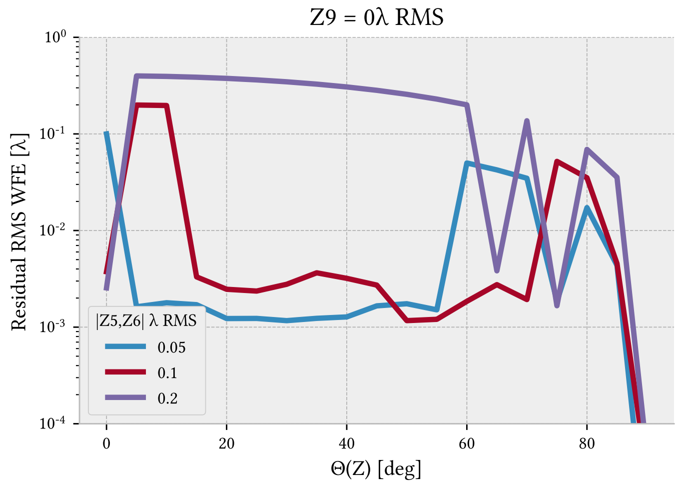

Fig. 28. Residual RMS WFE as a function of the orientation of primary astigmatism in the pupil plane for the case where there is no spherical aberration.

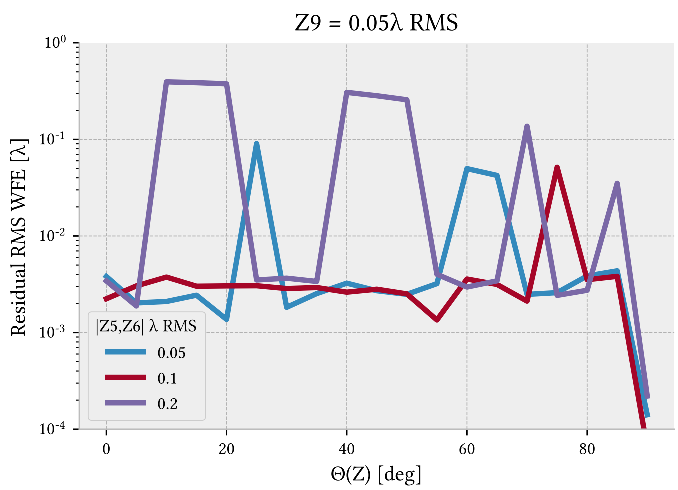

Fig. 29. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/20 λ RMS of spherical aberration.

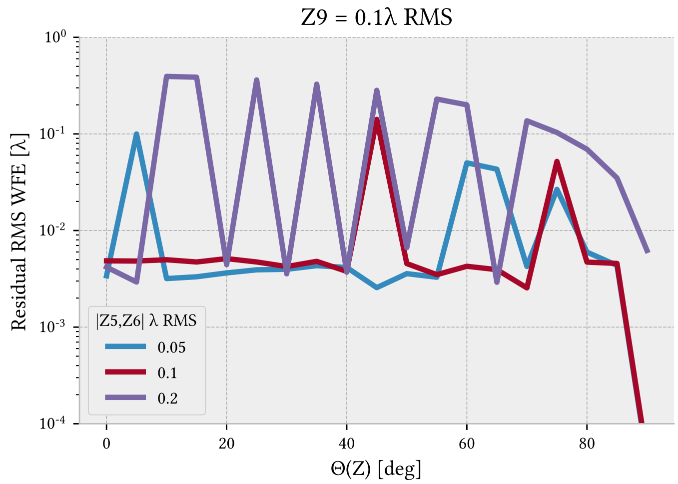

Fig. 30. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/10 λ RMS of spherical aberration.

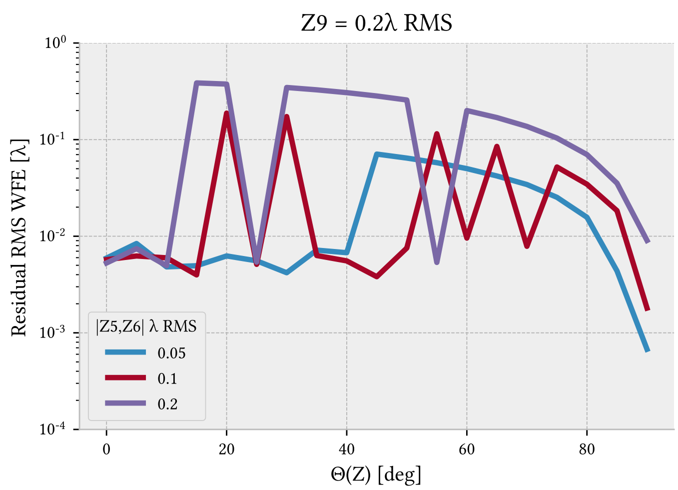

Fig. 31. Residual RMS WFE as a function of the orientation of primary coma in the pupil plane for the case where there is 1/5 λ RMS of spherical aberration.

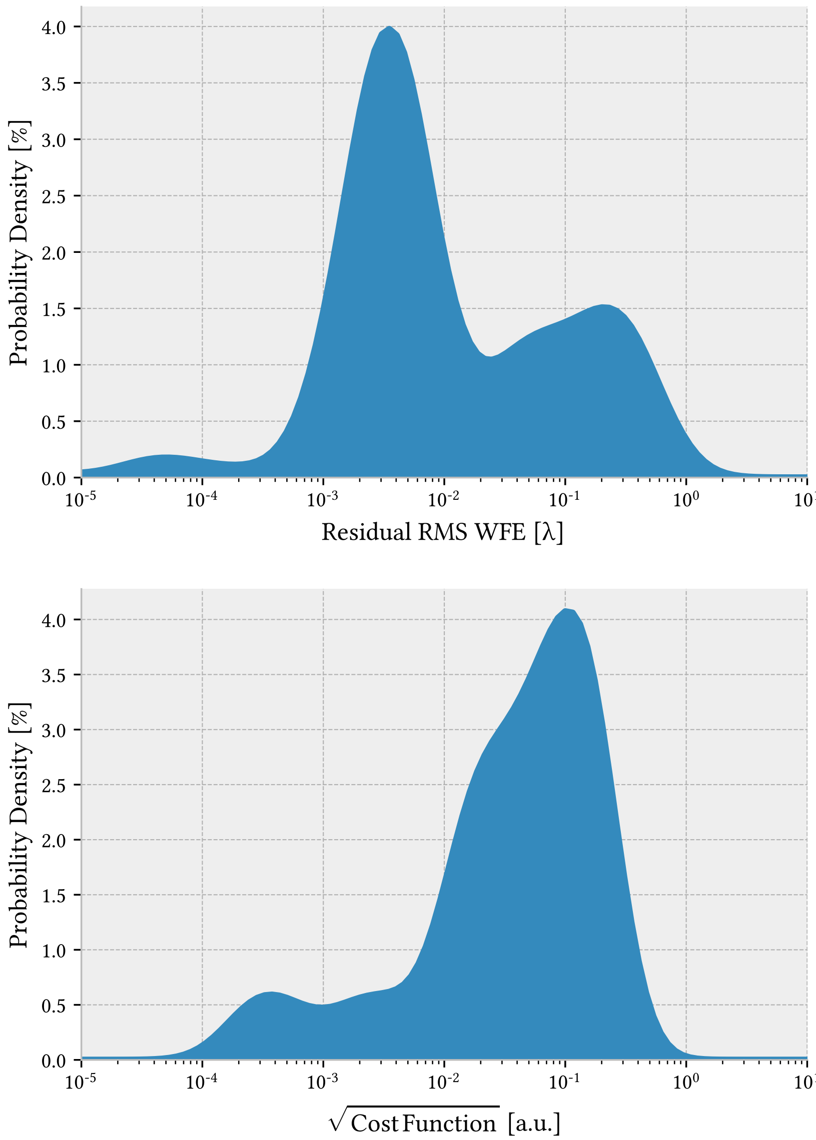

Fig. 32. Top: distribution of residual RMS WFE values; bottom: distribution of final $\sqrt{C}$ values. The two do not agree, unlike the spherical case (near perfect proportionality) or coma case (worse, but still reasonable proportionality). This is likely to do with the higher azimuthal order creating worse ambiguity at larger wavefront error levels.

6.4 Arbitrary Low Azimuthal Order Aberrations

500 Additional trials were performed with arbitrary combinations of coma, spherical aberration, and astigmatism. Each had the noisy geometric series among its various radial orders discussed in subsection 5.4 applied, with independent noise or fudge factor distributions for each. The higher radial order variants of each aberration were enforced

to be at an equal angle in the pupil plane to their lower radial order counterparts. While this may not be completely realistic, it is more realistic than if these modes were free to have any azimuthal orientation. The angle of coma and astigmatism were independent, and drawn from uniform random distributions spanning [0,180]°. The level of success in these trials was very low, and as such the presentation is relatively minimal.

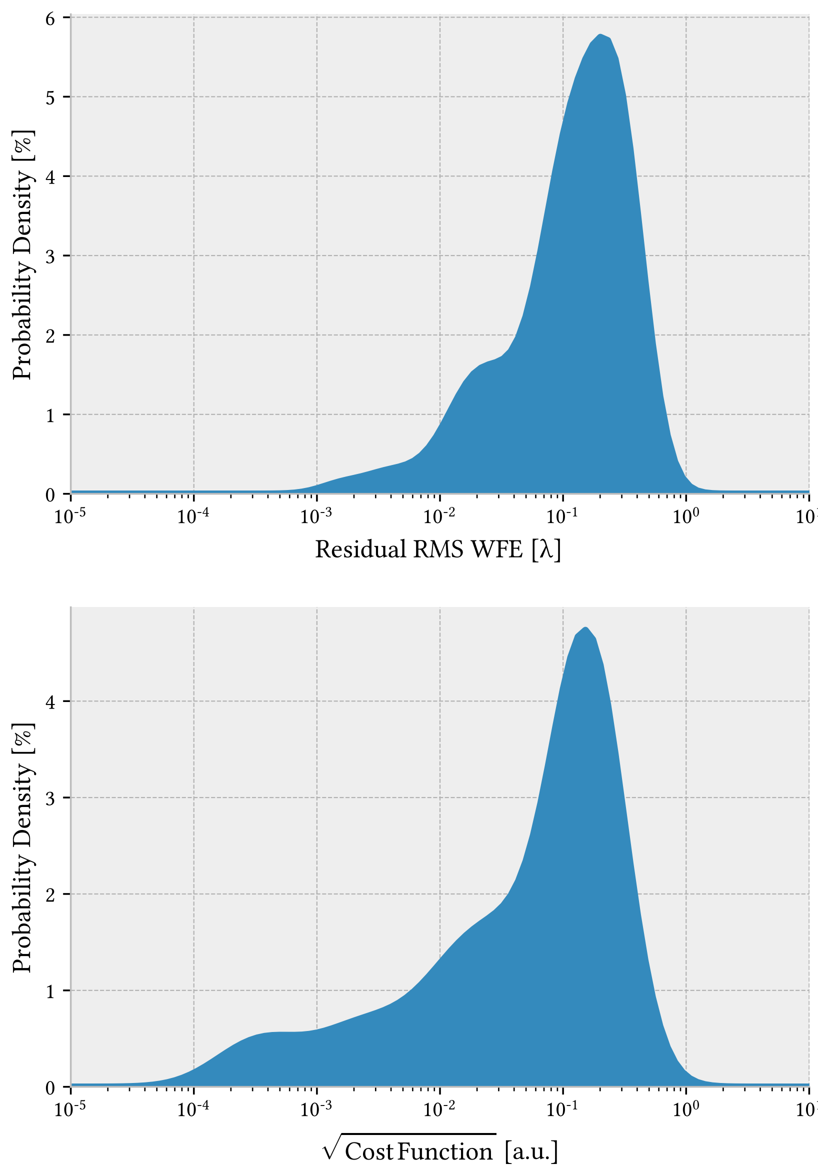

Fig. 33. Top: probability density of residual RMS WFE values, bottom: probability density of $\sqrt{C}$ values. The two appear reasonably well correlated, however the overall success is low with very few falling near or to the left of the 10^{-3} level. The value of $\sqrt{C}$ sometimes reached low values on the order of 10^{-4} while this never happened for the residual RMS WFE; this strongly implies the existence of ambiguity near the global minimum.

6.5 Experimental Trials

Five experimental trials were performed with a small set of Canon 35mm f/2 IS USM lenses and a Trioptics Imagemaster MTF bench. A sample of the five is reported here. The experimental details are provided in Appendix B.

The MTF was measured through-focus across a range of ±50 μm. It is shown in Figure 34. Additionally, the wavelength of light, pupil amplitude, and illumination profile from the collimator were also measured. The function A from Equation 5 was modeled as the product of a circle of appropriate diameter extracted from the pupil amplitude measurement, and a 2D Gaussian fit to the square root of the irradiance pattern measured from the collimator. The factor of square root converts irradiance to electromagnetic field amplitude.

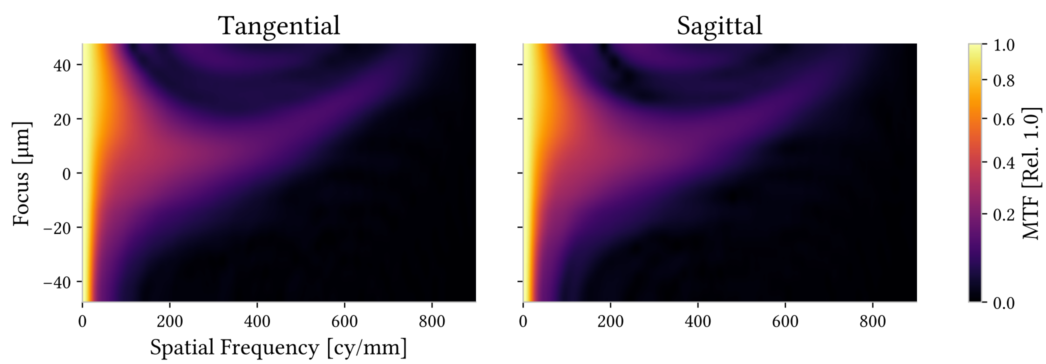

Fig. 34. Experimental MTF data as a function of frequency and focus for the tested lens. A Lanczos interpolation function is used to smooth the data for display, but inter-focus interpolation is not done during optimization.

The complete basinhopping algorithm described in Subsection 5.3 was applied to the data with up to 25 starting points for each segment of optimization. Several "bootstrapping" strategies were tried, where a subset of terms are optimized for and additional terms added after some had settled close to their correct value. The final path taken involved optimization for Z4, Z9, and Z16 – defocus and spherical aberration up to sixth order. The optimizer was then allowed to optimize for these terms in addition to trefoil, and finally the simultaneous addition of the coma and astigmatism terms. Figure 35 shows the wavefront estimate at the end of each of these segments of optimization, each of which entails up to 25 starts at basinhopping. Figure 36 shows the convergence of the optimizer's cost function over time, with the square root taken. As a reminder, the square root of this type of cost function is suggestive of the residual RMS WFE. Here, it reached a value suggestive of about $\lambda/43$.

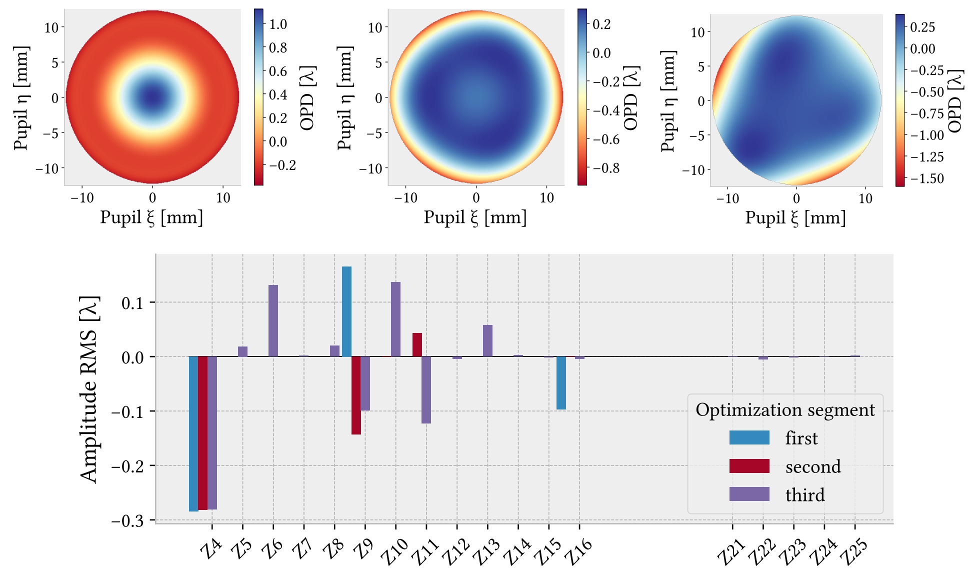

Fig. 35. Top left to right: predicted wavefront after the first, second, and final segment of optimization. Bottom: orthonormal Fringe Zernike coefficients, labeled first second third set of iterations in the same order as above. It can be seen that when the comatic and astigmatic terms are introduced, the value of Z4 remains stable, but all other terms included change. The stability of the focus term even in error of higher order terms is important to some applications where precision measurement of focus must resort to exotic means, such as wavefront sensing, and can only use MTF data.

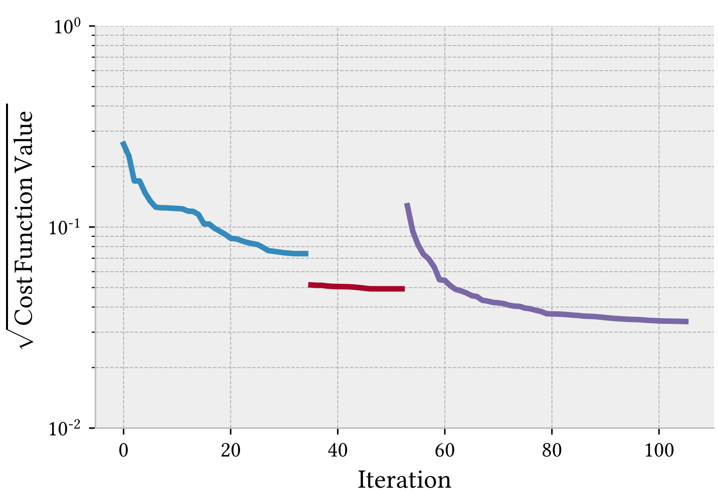

Fig. 36. Convergence of the optimizer. Each line segment is the **final** iteration of the basinhopping algorithm, so there is no expected continuity between the three. It can be seen that with the addition of each new set of polynomials to work with, the optimizer is better able to produce a model that matches the data. No further improvement was possible after the final point reached on the graph.

The final model and the measured data are compared in Figure 37. The focus rolls off in the negative focus direction more slowly in the model, and some finer features at the positive focus direction are not present in the model. Using a similar script to the one described in Appendix B, the through-focus PSF was captured to provide a more robust comparison, as wavefront data for the OUT is unavailable. The notionally in focus measured PSF is compared to its equivalent modeled one in Figure 38. In the model, Q has been manipulated such that the sample spacing is equal to the measured PSF. Both are better than Nyquist sampled.

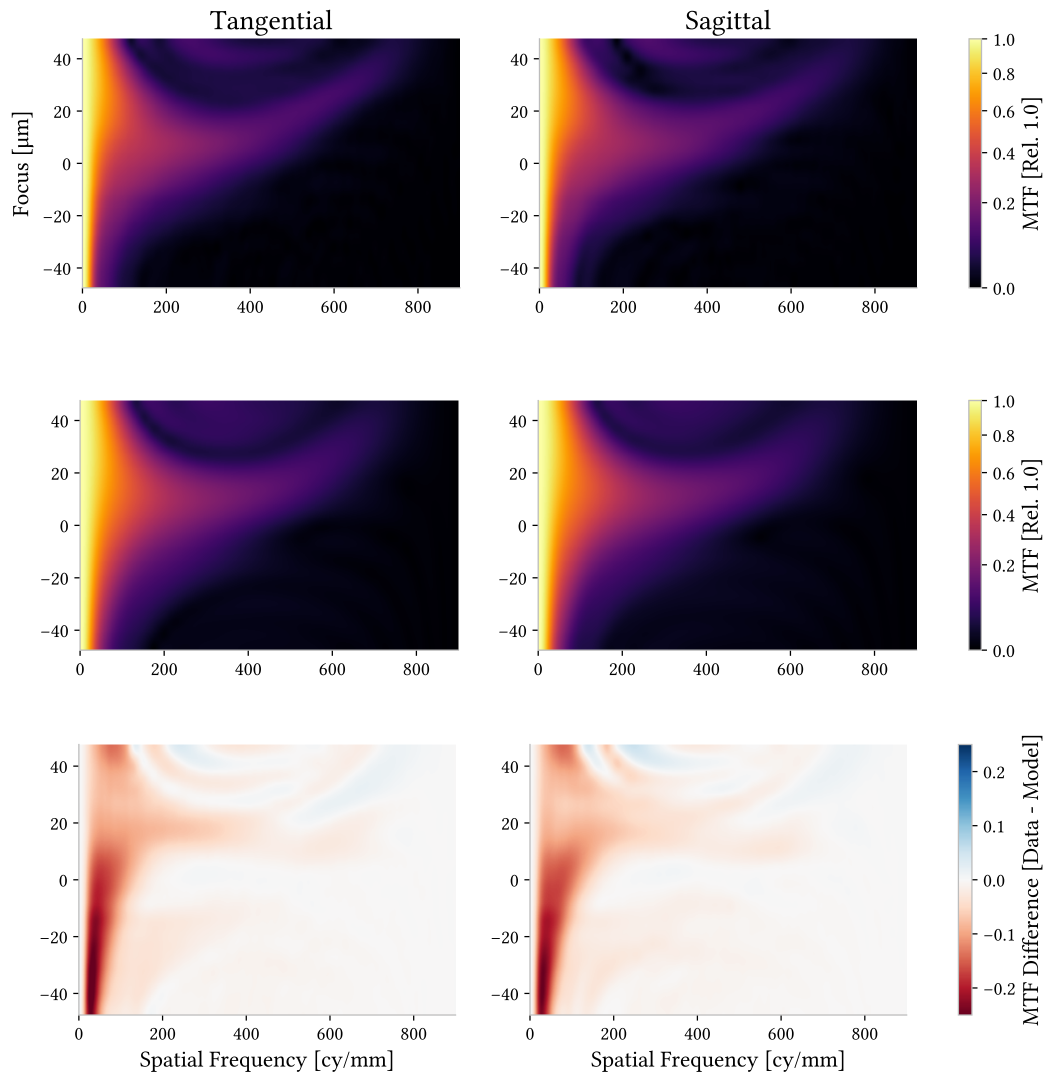

Fig. 37. Top: measured data, middle: retrieved MTF model, bottom: difference. Note that the data and model are presented on a power-scaled axis to enhance “dark” features in the data, while the difference is shown in a linear scale symmetric about 0 (perfect agreement).

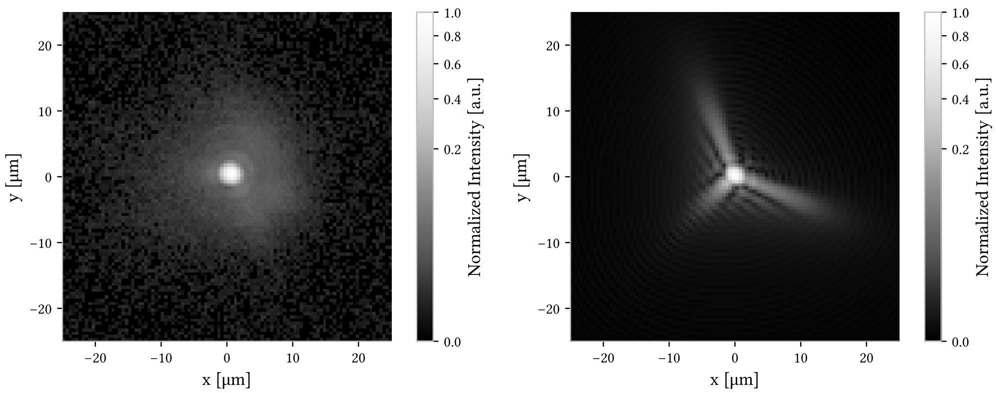

Fig. 38. Left: measured PSF, right: modeled PSF. The agreement between the measurement and model is poor; this implies a so-called uniqueness problem. Note that the trefoil in the retrieved PSF features "flares" at small angles to the Cartesian axes. Note as well that these PSFs are presented on nonlinear color scale to make more clear dim features.

Relatively severe ambiguity between coma and trefoil are likely the reason this experiment failed. The defocused PSFs at the extremes of focus (not presented here) show trefoil present in the measured PSFs, but it is much smaller than that which appears in the model.

7 Conclusions and Future Work

In this thesis, it was proposed to use "classical" MTF measurements with only tangential and sagittal data to perform wavefront sensing. MTF-based wavefront sensing is a type of phase retrieval algorithm and as such its success is sensitive to the design of the program. Four candidate metrics that perform the comparison between the data and model were compared, and the best – the sum of squares of differences – was selected. A normalization scheme was developed such that the value of these cost functions is nearly exactly proportional to the residual RMS error of the estimate. Critically, this allows the user to set an appropriate cost function tolerance which is synonymous with the tolerance on the quality of wavefront estimate produced. This property is only observed in this thesis in the case of rotationally invariant Zernike modes, as there is insufficient azimuthal information in the case of just tangential and sagittal MTF for non rotationally invariant wavefronts to be properly distinguished.

In excess of 3,000 simulations were performed to statistically evaluate the effectiveness of this method. It was shown that it is overwhelmingly successful in the case of rotationally invariant aberrations (spherical and defocus), with 100% success when the RMS WFE of the wavefront optimized for is less than 0.2 λ. Success over the larger range [0, 0.35] λ RMS remained high at 84.6%. Cases where the algorithm was not successful have complete failure, with no convergence in cost function or residual RMS WFE.

In the case of non rotationally invariant aberrations, namely coma and astigmatism, success was greatly reduced. Results including coma were only successful 32.5% of the time, Judged by the cost function value instead of the residual RMS WFE, results including coma would be said to be successful 78.9% of the time, but this is not true when the estimated and true wavefronts are compared, owing to ambiguity between primary coma (Z7 and Z8 in the Fringe Zernike notation) and its higher order variants, both in radial and azimuthal order. Results were better with astigmatism, with 64.5% success. The cost function value would suggest 71.1% success, indicating there are substantially fewer ambiguities with astigmatism than with coma. It is possible that a relatively simple preprocessing of the data to produce a better estimate of astigmatism would lead to a much higher success rate. When spherical aberration, coma, and astigmatism were combined in arbitrary orientations and magnitudes, the success rate was only 4.6%. The cost function suggests a success rate of 50%, indicating severe ambiguity of the wavefront for a given cost function value.

An ensemble of experiments were performed with a commercial MTF bench and small set of lenses of the same model. The appropriate MTF measurements were made, as well as captures of the lens' through-focus PSF to compare to the predicted PSFs. The PSFs are a substitute for measurements of the true wavefront of the lens. In the example trial shown, the cost function predicted $\lambda/43$ residual RMS WFE, however looking at the PSFs, the error is clearly substantially larger. This is due to the ambiguities previously discussed. The other experimental trials did not fare better.

From a physics perspective, the most important next step is the evaluation of how many azimuths of the MTF are needed for the data to become sufficiently sampled to resolve ambiguities. The translation from a PSF to MTF incurs the loss of Fourier phase; this appears to be very well compensated by an extended number of focus planes. The azimuthal under-sampling appears much more damaging. Of additional importance is the more rigorous study of how many distinct frequencies or what frequency spacing is required; the pitch of 10 cy/mm used in this thesis was arbitrary. Additional study of the number of focus planes and focus range used should also be done.

From a computational perspective, the Zernike coefficients could be re-expressed in terms of their magnitude and angle. This will eliminate pairwise redundancy in the terms when the magnitude is correct, but the orientation is wrong or vice-versa. In the x-y or 0-45° notation typically used, two terms must change, when there is a transformation that requires only one to change. The temperature and step size parameters of the basinhopping algorithm may also be tuned as these too were chosen by hand and no systematic study was performed to optimize them.

Study should be done with reduced spatial frequency bandwidth, e.g. only 0 to 0.5 ν_c and a reduced number of focus planes to see if these break the algorithm. Additional work should be performed to tune the through-focus range used, as the value of about 2 λ PV of defocus was chosen by intuition, not a rigorous study. Simulation with higher azimuthal order aberrations should be done, but this is likely wasted effort unless better azimuthal sampling of the MTF Is provided. Simulations should also be done that include noise, though this becomes specific to the type of MTF measurement done (pinhole based, slit based, slanted-edge, etc.).

The initial implementation of this algorithm typically works very quickly, returning a result within five minutes on a laptop computer when estimating spherical aberration, coma, and astigmatism. It is easily able to leverage the resources of a more powerful machine to work more quickly, fully saturating one node per simulation on the University of Rochester BlueHive cluster computer. Implementation of algorithmic differentiation can improve performance by approximately 4x when spherical aberration, coma, and astigmatism are estimated. Gains from this increase with the number of Zernike terms estimated. The program never returned an error and exited prematurely throughout all simulations and experimental trials. As such, it can be recommended that this program is ready for scientific and industrial use alike. Modification would be required to include more azimuths of MTF data, but these changes are not overwhelming in scale.

8 Acknowledgments

I would like to acknowledge the support of many people towards this thesis. First and foremost, I would like to thank Dr. Jannick Rolland, who accepted me into her research group studying aberration theory very early in my academic career. I would also like to give thanks to the late Dr. Kevin Thompson, who sparked in me a deep passion for aberration theory and the aberration fields of optical systems. I would also like to thank Dr. Joe Green, Andy Kee, Shannon Zareh-Kian, Dr. Norbert Sigrist, and Dr. Erkin Sidick of Caltech/NASA JPL for inspiring in me an interest in wavefront sensing. Finally, I would like to thank Dr. Fienup and several students within his group (Aaron Michalko, Scott Paine, Alex Iacchetta, Wes Farriss, and Joseph Tang) for their support and guidance as I've embarked on this thesis. I would also like to extend my deep thanks to Dr. Roger Cicala and Aaron Closz of Olaf Optical Testing, who have provided significant support to this project and allowed me the great privilege of exploring my interests in metrology without restriction for the past several years. Thanks are deserved as well to Ali Hashim, who produced Figure 1.

References

[1] A. Maréchal, *Etude des effets combinés de la diffraction et des aberrations géométriques sur l'image d'un point lumineux* (Editions de la Revue d'Optique Theorique et Instrumentale, 1948).

[2] K. Strehl, *Die Beugungstheories des Femrohrs* (1894).

[3] K. Strehl, "Ueber luftschlieren und zonenfehler," Zeitschrift für Instrumentenkunde **22**, 213–217 (1902).

[4] V. N. Mahajan, "Strehl ratio for primary aberrations: some analytical results for circular and annular pupils," J. Opt. Soc. Am. **72**, 1258–1266 (1982).

[5] V. N. Mahajan, "Strehl ratio for primary aberrations in terms of their aberration variance," J. Opt. Soc. Am. **73**, 860–861 (1983).

[6] O. Schade, "Electro-optical characteristics of television systems. introduction and part I: Characteristics of vision and visual systems," RCA Rev. (1948).

[7] O. Schade, "Electro-optical characteristics of television systems. part II: Electro-optical specifications for television systems," RCA Rev. (1948).

[8] O. Schade, "Electro-optical characteristics of television systems. part III: Electro-optical characteristics of camera systems," RCA Rev. (1948).

[9] O. Schade, "Electro-optical characteristics of television systems. part IV: Correlation and evaluation of electro-optical characteristics of imaging systems," RCA Rev. (1948).

[10] D. Malacara, *Optical Shop Testing* (John Wiley & Sons, Inc, Hoboken, NJ, 2007), 3rd ed.

[11] J. F. Hartmann, *Bemerkungen über den Bau und die Justirung von Spektrographen* (Springer, 1900), pp. 17–27, 47–58.

[12] R. V. Shack and F. D. Smith, "Production and use of a lenticular Hartmann Screen," (1971).

[13] R. S. Capers and E. Lipton, "Hubble error: Time, money and millionths of an inch," The Acad. Manag. Exec. (1993-2005) pp. 41–57 (1993).

[14] R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," J. Opt. Soc. Am. **64**, 1200–1210 (1974).

[15] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik. **35**, 237–246 (1971).

[16] J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," Opt. Lett. **3**, 27–29 (1978).

[17] J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).

[18] J. R. Fienup, "Phase retrieval algorithms: a personal tour," Appl. Opt. **52**, 45–56 (2013).

[19] D. Redding, P. Dumont, and J. Yu, "Hubble Space Telescope prescription retrieval," Appl. Opt. **32**, 1728–1736 (1993).

[20] B. H. Dean, D. L. Aronstein, S. J. Smith, R. Shiri, and S. D. Acton, "Phase retrieval algorithm for JWST flight and testbed telescope," in *Proc. SPIE. Space Telescopes and Instrumentation I: Optical, Infrared, and Millimeter,* vol. 6265 (2006), p. 17.

[21] G. R. Brady and J. R. Fienup, "Nonlinear optimization algorithm for retrieving the full complex pupil function," Opt. Express **14**, 474–486 (2006).

[22] A. S. Jurling and J. R. Fienup, "Phase retrieval with unknown sampling factors via the two-dimensional chirp z-transform," J. Opt. Soc. Am. A **31**, 1904–1911 (2014).

[23] F. D. Smith, "Optical image evaluation and the transfer function," Appl. Opt. **2**, 335–350 (1963).

[24] "ISO 9335:2012 optics and photonics – optical transfer function – principles and procedures of measurement," Standard, International Organization for Standardization, Geneva, CH (2012).

[25] R. Barakat and A. Houston, "Line spread function and cumulative line spread function for systems with rotational symmetry," J. Opt. Soc. Am. **54**, 768–773 (1964).

[26] E. W. Marchand, "Derivation of the point spread function from the line spread function," J. Opt. Soc. Am. **54**, 915–919 (1964).

[27] E. W. Marchand, "From line to point spread function: The general case," J. Opt. Soc. Am. **55**, 352–354 (1965).

[28] L. R. Baker, *Selected Papers on Optical Transfer Function: Measurement*, SPIE milestone series (SPIE Optical Engineering Press, 1992).

[29] F. Scott, R. M. Scott, and R. V. Shack, "The use of edge gradients in determining modulation transfer functions," Photogr. Sci. Eng. **7**, 64–68 (1963).

[30] B. Tatian, "Method for obtaining the transfer function from the edge response function," J. Opt. Soc. Am. **55**, 1014–1019 (1965).

[31] R. Barakat, "Determination of the optical transfer function directly from the edge spread function," J. Opt. Soc. Am. **55**, 1217–1221 (1965).

[32] "ISO 12233:2017 photography – electronic still picture imaging – resolution and spatial frequency responses," Standard, International Organization for Standardization, Geneva, CH (2017).

[33] F. v. d. Bergh, "Deferred slanted-edge analysis: a unified approach to spatial frequency response measurement on distorted images and color filter array subsets," J. Opt. Soc. Am. A **35**, 442–451 (2018).

[34] L. R. Baker, *Selected Papers on Optical Transfer Function: Foundation and Theory*, SPIE milestone series (SPIE Optical Engineering Press, 1992).

[35] B. Dube, R. Cicala, A. Closz, and J. P. Rolland, "How good is your lens? assessing performance with MTF full-field displays," Appl. Opt. **56**, 5661–5667 (2017).

[36] B. Dube, "prysm, github.com/brandondube/prysm," (2017).

[37] C. Zhu, R. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: FORTRAN subroutines for large-scale bound-constrained optimization," ACM Trans. Math. Softw. **23**, 550–560 (1997).

[38] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering (Springer-Verlag New York, 2006).

[39] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," The J. Phys. Chem. A **101**, 5111–5116 (1997).

[40] D. Wales, *Energy Landscapes: Applications to Clusters, Biomolecules and Glasses*, Cambridge Molecular Science (Cambridge University Press, 2004).

[41] K. Kohm, "Modulation transfer function measurement method and results for the orbview-3 high resolution imaging satellite," in *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 35,* (2004).

[42] A. Wirth, A. Jankevics, and F. Landers, "Scene based wavefront sensing for figure control of airborne and space optics," in *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest,* (Optical Society of America, 2009), p. AOThC4.

[43] S. Wang, P. Yang, M. Ao, L. Dong, and B. Xu, "Wavefront sensing by means of binary intensity modulation," Appl. Opt. **53**, 8342–8349 (2014).

[44] D. T. Ralph, P. A. Altin, D. S. Rabeling, D. E. McClelland, and D. A. Shaddock, "Interferometric wavefront sensing with a single diode using spatial light modulation," Appl. Opt. **56**, 2353–2358 (2017).

[45] B. Vohnsen and D. Valente, "Surface-plasmon-based wavefront sensing," Optica. **2**, 1024–1027 (2015).

[46] F. Ghebremichael, G. P. Andersen, and K. S. Gurley, "Holography-based wavefront sensing," Appl. Opt. **47**, A62–A69 (2008).

[47] M. Born and E. Wolf, *Principles of Optics* (Cambridge University Press, 1999), 7th ed.

[48] J. W. Goodman, *Introduction to Fourier Optics* (W. H. Freeman, 2017), 4th ed.

[49] F. Zernike, "Beugungstheorie des schneidenver-fahrens und seiner verbesserten form, der phasenkontrastmethode," Physica. **1**, 689–704 (1934).

[50] H. H. Hopkins, *Wave Theory of Aberrations* (Oxford University Press, 1950).

[51] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization*, Wiley Series in Probability and Statistics (Wiley, 2015).

[52] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, 1986).

[53] "ANSI Z80.28-2017, ophthalmics – methods of reporting optical aberrations of eyes," Standard, American National Standards Institute (2017).

[54] "ISO 517:2008 photography – apertures and related properties pertaining to photographic lenses – designations and measurements," Standard, International Organization for Standardization, Geneva, CH (2017).

[55] D. Coffin, "DCRaw, `https://www.cybercom.net/~dcoffin/dcraw/`," (2018).

[56] Microsoft, "Image composite editor, `https://www.microsoft.com/en-us/research/product/computational-photography-applications/image-composite-editor/`," (2018).

Appendices

A Fringe Zernike Polynomials

The Zernike polynomials are widely used in optical metrology and design as a convenient radial basis set. The original and ANSI standard sets [53] require two indices, which has led the development of alternative sets, among them the Noll, Zemax Standard, Fringe, Born & Wolf, and Malacara sets. The Fringe set is the most popular in optical metrology, and is used and reproduced in this thesis.

Index	Name	Norm	Equation
1	Piston	1	1
2	Tip	2	$\rho\cos(\theta)$
3	Tilt	2	$\rho\sin(\theta)$
4	Power	$\sqrt{3}$	$2\rho^2 - 1$
5	Primary Astigmatism 0°	$\sqrt{6}$	$\rho^2\cos(2\theta)$
6	Primary Astigmatism 45°	$\sqrt{6}$	$\rho^2\sin(2\theta)$
7	Primary Coma X	$2\sqrt{2}$	$(3\rho^3 - 2\rho)\cos(\theta)$
8	Primary Coma y	$2\sqrt{2}$	$(3\rho^3 - 2\rho)\sin(\theta)$
9	Primary Spherical Aberration	$\sqrt{5}$	$6\rho^4 - 6\rho^2 + 1$
10	Primary Trefoil X	$2\sqrt{2}$	$\rho^3\cos(3\theta)$
11	Primary Trefoil Y	$2\sqrt{2}$	$\rho^3\sin(3\theta)$
12	Secondary Astigmatism 0°	$\sqrt{10}$	$(4\rho^4 - 3\rho^2)\cos(2\theta)$
13	Secondary Astigmatism 45°	$\sqrt{10}$	$(4\rho^4 - 3\rho^2)\sin(2\theta)$
14	Secondary Coma X	$2\sqrt{3}$	$(10\rho^5 - 12\rho^3 + 3\rho)\cos(\theta)$
15	Secondary Coma Y	$2\sqrt{3}$	$(10\rho^5 - 12\rho^3 + 3\rho)\sin(\theta)$
16	Secondary Spherical Aberration	$\sqrt{7}$	$20\rho^6 - 30\rho^4 + 12\rho^2 - 1$
17	Primary Tetrafoil X	$\sqrt{10}$	$\rho^4\cos(4\theta)$
18	Primary Tetrafoil Y	$\sqrt{10}$	$\rho^4\sin(4\theta)$
19	Secondary Trefoil X	$2\sqrt{3}$	$(5\rho^5 - 4\rho^3)\cos(3\theta)$
20	Secondary Trefoil Y	$2\sqrt{3}$	$(5\rho^5 - 4\rho^3)\sin(3\theta)$
21	Tertiary Astigmatism 0°	$\sqrt{14}$	$(15\rho^6 - 20\rho^4 + 6\rho^2)\cos(2\theta)$
22	Tertiary Astigmatism 45°	$\sqrt{14}$	$(15\rho^6 - 20\rho^4 + 6\rho^2)\sin(2\theta)$
23	Tertiary Coma X	4	$(35\rho^7 - 60\rho^5 + 30\rho^3 - 4\rho)\cos(\theta)$
24	Tertiary Coma Y	4	$(35\rho^7 - 60\rho^5 + 30\rho^3 - 4\rho)\sin(\theta)$
25	Tertiary Spherical Aberration	3	$70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1$
26	Pentafoil X	$2\sqrt{3}$	$\rho^5\cos(5\theta)$
27	Pentafoil Y	$2\sqrt{3}$	$\rho^5\sin(5\theta)$
28	Secondary Tetrafoil X	$\sqrt{14}$	$(6\rho^6 - 5\rho^4)\cos(4\theta)$
29	Secondary Tetrafoil Y	$\sqrt{14}$	$(6\rho^6 - 5\rho^4)\sin(4\theta)$
30	Tertiary Trefoil X	4	$(21\rho^7 - 30\rho^5 + 10\rho^3)\cos(3\theta)$
31	Tertiary Trefoil Y	4	$(21\rho^7 - 30\rho^5 + 10\rho^3)\sin(3\theta)$
32	Quaternary Astigmatism 0°	$3\sqrt{2}$	$(21\rho^6 - 30\rho^4 + 10\rho^2)\cos(2\theta)$
33	Quaternary Astigmatism 45°	$3\sqrt{2}$	$(21\rho^6 - 30\rho^4 + 10\rho^2)\sin(2\theta)$
34	Quaternary Coma X	$2\sqrt{5}$	$(126\rho^9 - 280\rho^7 + 210\rho^5 - 60\rho^3 + 5\rho)\cos(\theta)$
35	Quaternary Coma Y	$2\sqrt{5}$	$(126\rho^9 - 280\rho^7 + 210\rho^5 - 60\rho^3 + 5\rho)\sin(\theta)$
36	Quaternary Spherical Aberration	$\sqrt{11}$	$252\rho^{10} - 630\rho^8 + 560\rho^6 - 210\rho^4 + 30\rho^2 - 1$

Table 3. The 36 Fringe Zernike polynomials and their norms for unity RMS value over the unit circle. ρ and θ are the normalized radial and azimuthal pupil coordinates. A line is drawn under Z25, the highest order term used in this thesis. Terms 26-36 are reproduced for completeness.

B Details of Experiment

Here we present the details of the experiment done to validate this method. It is our goal that anyone be able to replicate the result, in part or in full, depending on their level of access to instrumentation to perform these measurements. The selected lens is a Canon 35 mm f/2 IS USM, S/N 4430000379. Each sample of this lens model, like any other, will have somewhat different residual aberrations at any point in the field of view due to the reality of as-built performance.

1 Setup

The lens was mounted to a Trioptics ImageMaster HR MTF bench, with the 50mm diameter refractive collimator. The 546 nm narrowband spectral filter was used to ensure monochromatic test conditions. The `Center Sample Rotation` macro was run in a feedback loop with adjustment of the decenter controls of the mounting platform to ensure the boresight axis of the OUT was matched to that defined by the collimator and microscope. The boresight axis is a substitute for the optical axis in otherwise rotationally symmetric optical systems that contain small tilts and decenters.

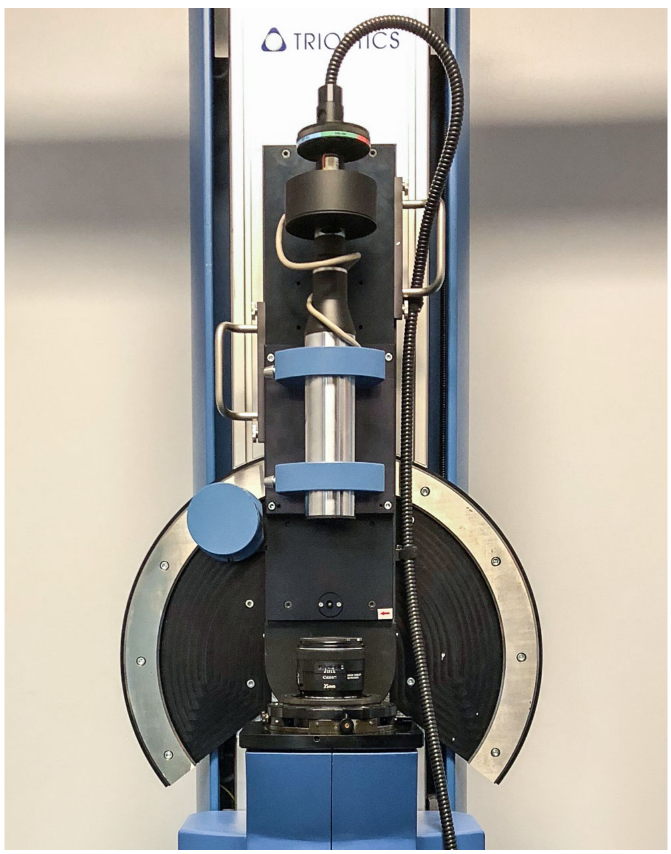

Fig. 39. The MTF bench used to make the measurement, with the OUT mounted. Motion can be controlled as shown in Figure 1.

The `EFLMag` test routine was used to measure the EFL 10 times with the 10..90 mm reticle selected, and the average value taken.

The entrance pupil diameter (EPD) and pupil magnification were measured via a modified variant of the method described in ISO 517 [54]. A Zeiss 100 mm f/2 Makro-Planar lens mounted to a Canon 6D camera and set to F/5.6 was used to image the entrance pupil. This lens has a small chief ray angle, excellent optical performance in terms of resolution and distortion, and is a suitable replacement for a bi-telecentric machine vision lens. The latter is preferred, but the usage here of a normal lens is more broadly reproducible as a typical optics lab does not have the preferred objective to perform dimensional measurements. Without adjusting its focus, a ruler was placed at the best focus object plane and used to calibrate the plate scale of the image. The diameter of the limiting aperture was computed using subpixel coordinates, and this process repeated with the 35 mm lens set to F/5.6 for both the entrance and exit pupils to ascertain the pupil magnification. The entire exit pupil may not be viewable from behind the lens' image plane, so a reduced aperture of the OUT or short working distance method is required to image the exit pupil and deduce the pupil magnification. Figure 40 shows the source image used to compute the entrance pupil diameter of the OUT.

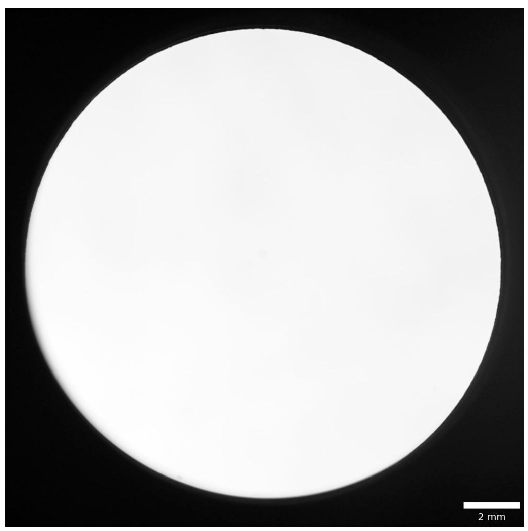

Fig. 40. Image of the entrance pupil of the OUT used to measure the entrance pupil diameter. The plate scale in the original is 25.4 μm per pixel. The small gradient visible in the left-hand side of the pupil is related to small angular alignment between the test setup and the OUT. The slight texture visible is due to imperfection in the diffuser used.

A Hasselblad X1D-50c camera was used to measure the intensity profile of the beam produced by the collimator by creating a 2x2 stitched panorama. This camera's large, 44x33 mm sensor allows the capture of a minimal number of frames to cover the entire 50 mm diameter beam. The raw images were developed in DCRaw [55] using the `-D -T -4` command line flags. This converts the manufacturer's proprietary raw format to a 16-bit linear TIFF file with no other processing. A python script was used to extract the G1 green color plane of each file and export the results to an intermediate set of TIFF files. These were fed to Microsoft Image Composite Editor [56] which rendered a complete mosaic with the default options. The illumination was found to be non-uniform, likely to do with the $\cos^4$ law and the obliquity of the edge of the collimator aperture to the pinhole in its focal plane. A Gaussian of the form in Equation 21 was fit to the data to an analytical amplitude mask in the pupil plane, where σ is a normalized width parameter, found to be 1.075. Note that in the argument of exp, ξ and η are normalized to span $[-1, 1]$.

$$f(\xi, \eta) = \exp\left(\frac{\xi^2 + \eta^2}{\sigma^2}\right) \tag{21}$$

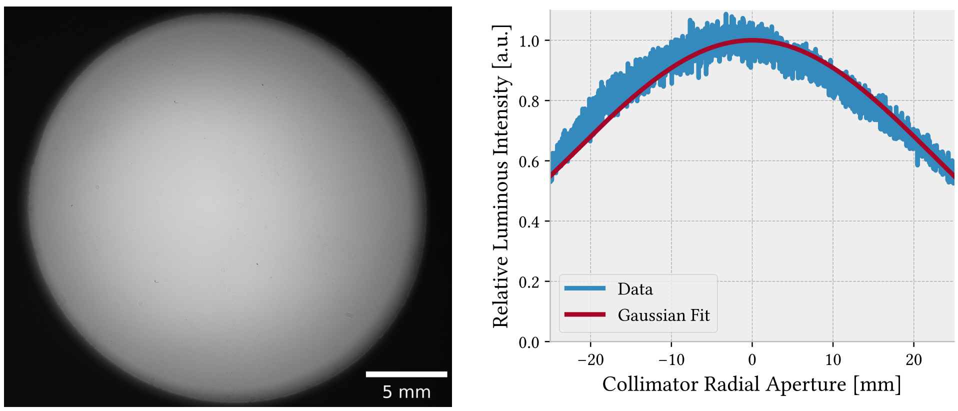

Fig. 41. left: image of the beam from the collimator used to calibrate the intensity profile. Right: 1D slice through the x axis and analytic fit. The small shift in the $-x$ direction and tilt in the data is likely due to a small physical tilt in the test setup and attenuation from obliquity.

A custom script was used to measure the MTF at 21 focus points spanning a range of ±0.1mm. The script is reproduced below. The original .mht files are available from the primary author upon request, and can be read with the functions `read_trioptics_mtf` and `read_trioptics_efl` from `prysm.io`.

The EFL measurement may be repeated in a handful of seconds. the EPD measurement takes a small number of minutes, but the result will vary no less than low double digit microns (machining tolerances on a single surface) between samples of the same lens model. Less than five minutes of time are required to produce the EFL, pupil, and MTF dataset. The data is summarized in Table 4.

Complete script to ingest this data and feed it to the wavefront sensing algorithm is also available upon request.

Parameter	Value	±	Unit
EFL	34.449	0.0013	mm
EPD	16.955	0.0128	mm
λ	0.546	0.0001	µm
Illumination Uniformity	See Figure 41		

Table 4. Non-MTF parameters measured

2 MTFLab Script

Here the script used to automate measurement is reproduced. Trioptics' software includes its own scripting language which allows automation of various measurements, including the MTF FFD described in subsubsection 2.3.3. In this case, the Through-Focus MTF routine built into the software only allows the use of a single frequency at a time. This script performs motion control and uses the regular MTF measurement routine to generate an ensemble of MTF measurements. The focus range, number of steps, frequency pitch, maximum frequency, and output filenames are hard-code here. Readers can modify the script with a `ParamBegin..ParamEnd` block to introduce flexibility of these parameters. The run time is approximately 10 seconds. Note that the number of points is specified twice; once as an integer for iterating, and again as a floating point value due to issues with mixed precision arithmetic as of version 4.8.0.9 of MTFLab. Both of these values must be changed for correct behavior to be observed.

```
Proc Main
    Dim nPoints, Integer, 21
    Dim nPointsDbl, Double, 21.0
    Dim focusRange, Double, 0.1
    Dim maxFreq, Integer, 900
    Dim freqPitch, Integer, 10

    Dim focusStep, Double, 0.0
    Dim refFocus, Double, 0
    Dim currentFocus, Double, 0.0
    Dim loopCounter, Integer, 0

    Assign focusStep, focusRange
    Mul focusStep, 2.0
    Div focusStep, nPointsDbl

    GetPosition "FocusPosition", refFocus
    Add currentFocus, refFocus
    Sub currentFocus, focusRange
    MoveAbs "FocusPosition", currentFocus, 50
    Report ClearAll
    MTF 100, 2.0, maxFreq, freqPitch, "Both"
    Report Save, "C:\Users\TRIOPTICS\Desktop\scriptreport", loopCounter, 1
    Sub nPoints, 1

    Loop nPoints
        Add loopCounter, 1
        Add currentFocus, focusStep
        MoveAbs "FocusPosition", currentFocus, 50
        Report ClearAll
        MTF 100, 2.0, maxFreq, freqPitch, "Both"
        Report Save, "C:\Users\TRIOPTICS\Desktop\scriptreport", loopCounter, 1
    LoopEnd

    MoveAbs "FocusPosition", refFocus, 50.0
    Report ClearAll
ProcEnd
```