ADAPTIVE, DYNAMIC SURFACE / WAVEFRONT METROLOGY AND ANALYSIS

by

Maham Aftab

Copyright © Maham Aftab 2019

A Dissertation Submitted to the Faculty of the

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

2019
THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Maham Aftab, titled Adaptive, Dynamic Surface / Wavefront Metrology and Analysis and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Daewook Kim
Assistant Professor
Optical Sciences/Astronomy

Date: 4/15/2019

Gregory Smith

Date: Apr. 5, 2019

Rongguang Liang

Date: 04/05/2019

Virendra Mahajan

Date: 5 Apr 2019

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.

Daewook Kim
Assistant Professor
Optical Sciences/Astronomy

Date: 4/15/2019
ACKNOWLEDGEMENTS

Thank you to my tiny tots, Minha and Dua for being the most precious little beings in my life.

I would like to thank Dr. Jim Burge for being an excellent advisor and mentor for the start of my dissertation work, Rick and Linda Hanson for being my second set of parents and fiercely protecting me, the wonderful Carmen Mestizo who is an inspiration and a great friend, Dr. Kelsey Miller for being an endless well of friendship and support, Sarah Komarabathuni for being a second sister, and Andrea Wheeler for being an amazing friend.

Thanks to the Large Optics Fabrication and Testing (LOFT) group for being a major part of this work, from scientific discussions to cheering me on at every stage of my Ph.D. My time at the College of Optical Sciences would not have been half as wonderful if it weren’t for the amazing ladies of Women in Optics, particularly Page King, Jhen Lumbres, and Emily Finan.

I extend a special thank you to the many amazing professors, students and staff at the College of Optical Sciences who taught, helped, inspired, and supported me endlessly.

Thank you to all my friends in Tucson who made my time there so much more beautiful, especially Parth, Amanda, Tasha, and Anahi.

My final and most enthusiastic thank you is for my advisor, Dr. Dae Wook Kim. I can never thank him enough for believing in me, teaching me about science as well as life, supporting me through the best and worst of times, being an inspiration, and being the most amazing mentor I could ever imagine. Dae Wook, I will always be indebted to you.
DEDICATION

To my parents, Shahina Aftab and Aftab Alam, my sister, Noor Aftab, and my mentor, Dr. Virendra N. Mahajan.
# TABLE OF CONTENTS

**LIST OF TABLES** ................................................................................................................................................. 7

**LIST OF FIGURES** ............................................................................................................................................... 8

**ABSTRACT** ......................................................................................................................................................... 10

Chapter 1. Introduction ........................................................................................................................................... 12

1.1 Gradient polynomials and their use in surface / wavefront reconstruction ................................................. 14
1.2 Curl polynomials and their use in modeling errors in metrology systems ...................................................... 16
1.3 Adaptive Shack–Hartmann Wavefront Sensor for increased dynamic range of measurement .................... 17
1.4 Dissertation contributions to scientific literature ............................................................................................. 19
1.5 Dissertation Outline ........................................................................................................................................... 20

Chapter 2. Gradient polynomials and their use in surface / wavefront reconstruction ...... 22

2.1 $G$ polynomials: background and derivation ..................................................................................................... 22
2.2 Data fitting methodology and its properties ....................................................................................................... 25
2.3 Use of $G$ polynomials in surface and wavefront reconstructions from simulated and real, measured deflectometry data .................................................................................................................................................. 28
  2.3.1 High fitting accuracy in the rectangular domain ......................................................................................... 29
  2.3.2 Good fidelity for representation of mid-to-high spatial frequencies .......................................................... 30
  2.3.3 Robustness against noise ............................................................................................................................ 33
  2.3.4 High reconstruction accuracy in the presence of obscurations ............................................................... 34
  2.3.5 Good reconstruction accuracy for unevenly sampled data ........................................................................ 39
2.4 Concluding discussion on $G$ polynomials ......................................................................................................... 41

Chapter 3. Curl polynomials and their use in modeling errors in metrology systems ...... 43

3.1 $C$ polynomials: background and derivation ...................................................................................................... 43
3.2 Applications of $C$ polynomials for modeling errors in deflectometry systems .................................................. 47
3.3 Modeling Imaging Distortion using $G$ and $C$ polynomials .............................................................................. 49

Chapter 4. Adaptive Shack–Hartmann Wavefront Sensor for increased dynamic range of measurement .......................................................................................................................... 53

4.1 Adaptive Shack–Hartmann Wavefront Sensor: basic concept and its implementation ................................... 53
4.2 System prototype ................................................................................................................................................. 58
4.3 Experimental results ......................................................................................................................................... 61
  4.3.1 Determining the A-SHWFS accuracy of measurement ............................................................................... 61
4.3.2 Displaying the adaptive wavefront sampling and dynamic detector subaperture reconfiguration abilities of the A-SHWFS

Chapter 5. Conclusion

REFERENCES

Appendix A. Modal Data Processing for High Resolution Deflectometry

Appendix B. Chebyshev gradient polynomials for high resolution surface and wavefront reconstruction

Appendix C. Rectangular domain curl polynomial set for optical vector data processing and analysis

Appendix D. Adaptive Shack-Hartmann Wavefront Sensor accommodating large wavefront variations
LIST OF TABLES

Table 2.1: Table of first ten $G$ polynomials [7]. .......................................................... 26
Table 2.2: RMS residual errors of the differences between the simulated and modal reconstructions [7]. .......................................................... 37
Table 3.1: Table of the first ten $C$ polynomials [13].................................................. 45
Table 3.2: $C$ polynomial-based mis-clocking estimation performance for the Alvarez lens deflectometry data analysis case [13]. .......................................................... 49
Table 3.3: Fitting performance comparison of the two vector polynomial set cases to represent/model the TMA camera’s distortion map [13]. .................................................. 51
LIST OF FIGURES

Figure 1.1: Schematic of a reflective deflectometry set-up [1]. ................................................... 13

Figure 1.2: Basic operating principle of a conventional Shack-Hartmann Wavefront Sensor: (a) Collimated reference wavefront case and (b) Aberrated wavefront case [15]. ......................... 18

Figure 2.1: (a) Surface plots of a simulated ideal surface, (b) $G$-polynomial-reconstructed surface, and c) Southwell integrated surface [7]. ………………………………………………. 30

Figure 2.2: (a) High-pass-filtered Southwell zonal reconstructed map and high-pass-filtered $G$-polynomial reconstructed maps generated using (b) 37, (c) 750, (d) 3000, and (e) 20,000 polynomial terms [7]. ………………………………………………………………………... 32

Figure 2.3: Residual errors from surface reconstruction in the presence of various amounts of noise [7]. ………………………………………………………………………………………. 34

Figure 2.4: (a) Simulated ideal data with a 90×2 mm defect. (b) Southwell and (c) modal residual error maps (differences from the simulated map) [7]. …………………………………………. 36

Figure 2.5: (a) Simulated ideal data. (b) Zernike gradient, (c) Southwell and (d) $G$ polynomial based modal residual error maps (differences from the simulated map) [7]. ………………… 38

Figure 2.6: Plot of change in reconstruction time as the number of polynomials used for data fitting is increased, for both modal methods. (Note: We acknowledge that the actual processing time could depend on the specific numerical implementation. This data is provided only for a baseline comparison.) [7]. ……………………………………………………………………… 39

Figure 2.7: (a) Simulated, ideal surface map (reference). The blue double ended arrows mark the approximate region on the map (1.25 × 1.25 cm) with a lower sampling (b) Error between the reference and Southwell methods generated surface (c) Error between the reference and $G$ polynomial generated surface. …………………………………………………………………. 40

Figure 3.1: Figure 3.1: (a) Schematic of the simulated TMA camera system in the Zemax library [27] (b) Distortion map for this system using 20×20 grid sampling with 15× arrow size magnification [27]. ………………………………………………………………………… 50
Figure 4.1: (a) Schematic layout of the A-SHWFS using actively modulated LCD lenslet array mask. (b) Depiction of how the LCD lenslet array mask (left) and the corresponding detector subapertures (right) are changed dynamically, from a fully unblocked (top) to partially blocked (bottom) situation. The squares on the left represent the pattern sent from the computer to the LCD screen [15].

Figure 4.2: A-SHWFS prototype configuration for proof-of-concept experiment. The green arrows represent the direction of laser wavefront propagation through the set-up [15].

Figure 4.3: (a) Picture of the three adjacent hexagonal mirror segments set-up. The green circle represents where the test beam hits the mirror set-up and the dotted orange circle shows the approximate location of the small mirror behind the active segment used for the reference autocollimator measurements. (b) Schematic of autocollimator set-up providing the golden standard for the A-SHWFS accuracy test. Red dotted arrows represent the autocollimator beam going to / from the small mirror [15].

Figure 4.4: (a) Small, and (b) Large magnitude of tilt measurements comparing the autocollimator and A-SHWFS values (blue circles). The error (red crosses), that is the difference between the autocollimator and A-SHWFS values, is shown with its own axis on the right side in both figures. A linear fit to the data is shown as the black line in both graphs [15].

Figure 4.5: Spots from the lenslets, on the CMOS detector. Red arrows show the spot motion relative to reference spots, on the active mirror portion of the image. (a) For small tilt, all three mirrors display the maximum number of spots (i.e., higher spatial sampling of the wavefront). The red arrows are shorter. (b) When the amount of tilt increases, the stationary mirror zones maintain the same number of spots while spots from the actively moving mirror have been selectively blocked by the LCD panel to increase the slope measurement dynamic range only in the optimal zone. Hence, the red arrows are longer [15].

Figure 4.6: Enhanced dynamic range demonstration of the adaptive wavefront sensing approach. The reconstructed wavefront time-lapse shows continuously increasing wavefront tilt as the active mirror was being tilted up-to and beyond the nominal 28.5 mrad dynamic range. The average slope magnitude for each map is shown as well. The dotted red line represents the start of adaptive gating i.e., when the blocking mask was applied to the LCD [15].
ABSTRACT

The demand for increasingly sophisticated optics continues to grow for a wide variety of applications, such as in astronomy, industrial manufacturing, medical imaging, and commercial photography. As more advanced fabrication methods are invented, especially for high-resolution or freeform designs, the tools and techniques for optical metrology and analysis must be made more precise, efficient, and robust. This study discusses various approaches for adaptive and dynamic surface or wavefront metrology and analysis which would aid in the ability to have more advanced and innovative optics.

Three techniques for improving optical testing and analysis are discussed in this work. The first two are mathematical frameworks, applied in software codes that provide new and improved solutions to challenges arising during optical metrology, e.g., deflectometry measurements and data analysis. Both are based on polynomial basis sets, and are optimized for systems with rectangular apertures. The first is used for reconstructing surfaces or wavefronts from measured slope data and the second uses the measured data to obtain information about possible misalignments or systematic errors in metrology systems. The third is the development of a sensor for measuring wavefront slope data, which allows solutions for optical testing and analysis problems that occur from a limited dynamic range of measurements. The dynamic
range of measurement is the range of wavefront slope values (largest and smallest values) that can be measured by a system. The aforementioned sensor uses the modal data fitting methodology described in this work.

Each of these topics has been researched, their main concepts tested, and software and (where applicable) hardware solutions developed for them. Simulations and real data analysis are used for verification of these tools and techniques.
Chapter 1

Introduction

Freeform optics have rapidly been gaining popularity and are being used for many systems including head mounted displays, segmented extremely large telescopes, camera systems and non-symmetric solar energy concentrators. Adding freeform surfaces can be beneficial for such systems in terms of compactness, higher throughput, and better image quality.

Fabrication and testing of freeform optics still poses a significant challenge. Besides the difficulty in finding a better mathematical description of freeform surfaces; requirements for optical form, mid-to-high spatial frequencies, and surface micro-roughness must be met simultaneously. Metrology tools for such surfaces must have high accuracy and excellent spatial resolution, and their parameters will be defined in chapters 2 and 3. A large dynamic range of measurement for the metrology system is generally also a requirement, particularly for precision freeform optics. These parameters will be defined in chapter 4.

Deflectometry is a type of optical metrology that measures the slopes of the surface under test. It is a simple yet powerful tool that primarily requires only a light source and a detector and can achieve measurement accuracies comparable to interferometry. It is a good solution for metrology of freeform systems, and has also proven valuable for the measurement of high-
resolution surfaces, such as telescope mirrors. Fig. 1.1 shows the layout of a general deflectometry system for reflective optics [1]. The source has a known pattern, for example, a sinusoidal wave. This light is reflected from the test optic and captured by the detector. By measuring the change in the recorded pattern versus the known source pattern, the local test surface slopes are calculated.

Figure 1.1: Schematic of a reflective deflectometry set-up [1].

For a long time, Shack–Hartmann wavefront sensors have been used for optical metrology. Despite having a higher dynamic range and greater operational stability than some other testing techniques such as interferometry, there is a constant push to increase the dynamic range of measurement of these sensors.
Traditionally, many optical systems had circular pupils but these days various systems are using non-circular pupils, of which a very common type are rectangular pupils. Some well-known systems that use rectangular apertures include the Gaia telescope system [2] and X-ray mirrors [3]. Imaging applications limited by size of the detection system ideally want to utilize the full detector, which often has a rectangular format. This was the case with the prototype Adaptive Shack–Hartmann Wavefront Sensor we designed and built, the details of which are discussed in chapter 4. Many solar panels and collectors also have rectangular shapes. The vector polynomial sets introduced in this dissertation are designed primarily for data from non-symmetric systems with rectangular apertures, but the methodology remains general enough to reconstruct other non-axially symmetric or freeform shapes.

1.1 Gradient polynomials and their use in surface / wavefront reconstruction

Deflectometry systems measure the local slopes of the surface of the unit under test (UUT). The methods to reconstruct measured surface from its slope or gradient data can be broadly classified into two categories [4]: first, modal methods such as polynomial fitting that decomposes data into a set of smoothly varying modes, for example Zernike polynomials [5], which can be used to fit the entire surface; second, zonal methods, such as the Southwell zonal method [6] that describe the surface through numerical integration in sections or ‘zones,’ for example, a uniform grid of rectangular elements. Both methods have their own unique strengths and may be used individually or in combination, depending on the application. For surface and wavefront
reconstruction, as well as modeling system errors, the applications discussed in this dissertation focus mainly on data from high-resolution or freeform optics, measured with deflectometry approaches. For these purposes, modal data fitting can be a very attractive option. As discussed in [7], modal reconstruction methods are generally less sensitive to measurement noise as compared to zonal methods, the number and types of modes used can be adapted to the problem [8], and various modes can be related to physical properties or parameters of the system [4]. Modal methods may also be useful in providing solutions or improvements to practical metrology problems, for instance modal reconstructions may reduce the error in surface reconstruction from deflectometry data when fiducials i.e., physical markers set-up as reference during measurements, are present on the optic being measured [7]. In this work, we present a new modal method, which uses our newly developed vector polynomial basis set that can reconstruct the surface or wavefront from its measured gradients [7]. This polynomial set, called $G$ polynomials, is derived from the gradients of two-dimensional Chebyshev polynomials of the first kind [9]. The polynomials for wavefront analysis of rotationally symmetric optical systems with a rectangular pupil have been given in [10]. The corresponding polynomials to analyze the wavefront slope data for such pupils have been discussed in [11]. An important class of such systems are those with high-power laser pulses that are rectangular or square in form but are transmitted by an optical system that is rotationally symmetric. An example of a high powered
laser system is the National Ignition Facility [12] which utilizes thousands of large optical components, most of which have rectangular apertures, as described in [11].

In this dissertation, the modal data processing methodology focuses on polynomials for systems that are not rotationally symmetric and have rectangular pupils. The polynomials used for wavefront analysis are based on the Chebyshev polynomials of the first kind and have closed-form solutions. The corresponding polynomials for analyzing the wavefront slope data are similarly based on the Chebyshev polynomials (of the first kind), for which we have obtained closed-form solutions. Both sets of polynomials are simple in form and easy to use, as demonstrated in this dissertation by applying them to deflectometry data [7, 1].

1.2 Curl polynomials and their use in modeling errors in metrology systems

We have derived another set of vector polynomials called \( C \) polynomials, related to the \( G \) polynomials [13]. The \( C \) polynomial set is based on the curl of two-dimensional Chebyshev polynomials of the first kind. This set only contains terms with zero divergence, that can be defined locally as a rotation or curl. The \( G \) and \( C \) sets together make a complete set of polynomials that can represent any vector data bounded by the rectangular domain. \( C \) polynomials on their own or the set of \( G \) and \( C \) polynomials together have very interesting applications for optical metrology and analysis. Two such applications explored in this work are (i) mapping imaging distortion from systems with rectangular pupils, which uses the set of \( G \) and
C polynomials to modally fit a simulated distortion measurement in a Three Mirror Anastigmat (TMA) system and (ii) measuring and analyzing systematic errors in deflectometry systems, which makes use of a numerical model based on C polynomials to represent and quantify a source misalignment in a typical deflectometry set-up. Both these applications are detailed in reference [13], as well as chapter 3.

1.3 Adaptive Shack–Hartmann Wavefront Sensor for increased dynamic range of measurement

A Shack–Hartmann wavefront sensor (SHWFS) measures the slope distribution of a wavefront [14]. It contains an array of lenses (called lenslets) that sample the wavefront to be tested. A collimated, unaberrated wavefront is used as the reference. When this wavefront is used as the input to the sensor, small focused spots of light are seen on the detector. As the wavefront starts to deviate from its reference condition, the focused spots of light move away from their reference locations. The deviation of a spot from its reference location is proportional to the local slope of the wavefront. The test wavefront can be reconstructed from its measured slope data using the zonal or modal techniques described in section 1.1. Fig.1.2 shows the basic operating principle of a SHWFS [15].
Figure 1.2: Basic operating principle of a conventional Shack-Hartmann Wavefront Sensor: (a) Collimated reference wavefront case and (b) Aberrated wavefront case [15].

We developed a modification to the SHWFS through an active approach, which reconfigures detection subaperture areas in the imaging plane by blocking or unblocking desired lenslets through a mask that can be generated and modified electronically. This method, called the Adaptive Shack–Hartmann Wavefront Sensor (A-SHWFS), can actively and dynamically reconfigure itself, based on the amount of aberration in the test beam. A liquid crystal display (LCD) panel placed in front of the lenslet array helps achieve the detection subaperture reconfiguration as the electronic signal (or mask) applied to the LCD can block or allow light to pass through each LCD pixels. By matching these pixels to the lenslets in the A-SHWFS lenslet array, each lenslet can be addressed and lenslets can be effectively blocked or unblocked. This blocking pattern can be symmetric (horizontal, vertical, diagonal etc.) or non-symmetric, thereby allowing the sensor to achieve variable dynamic range of measurement. There is a trade-
off between the resolution and dynamic range of the sensor, where a higher spatial resolution of the detected wavefront is obtained (i.e., the wavefront is sampled more finely) if the size of detection subapertures is small, but the maximum amount of aberration measurable for a fixed focal length also becomes smaller. One very useful application of the A-SHWFS can be for wavefronts that are severely aberrated in a limited region relative to the entire wavefront. This case, which is demonstrated with an experimental set-up as well as a simulation in [15] and chapter 4 uses regional blocking, where considerably different masks are applied to the section of the wavefront which is severely aberrated versus the rest of the wavefront. This ensures that the whole wavefront can be appropriately measured while utilizing the best possible spatial resolution for its various parts.

1.4 Dissertation contributions to scientific literature

The research in this dissertation has resulted in the following journal publications and proceedings:


1.5 Dissertation Outline

Chapters 2, 3, and 4 on Gradient polynomials and their use in surface / wavefront reconstruction, Curl polynomials and their application in modeling errors in metrology systems, and Adaptive Shack–Hartmann Wavefront Sensor for increased dynamic range of measurement, respectively, describe the main concepts, provide background information, and discuss the applications regarding the three main techniques for adaptive, dynamic surface / wavefront metrology and
analysis that we developed. Detailed work on these topics has been documented in research papers, which are attached as appendices.

We present concluding remarks in chapter 5.
Gradient polynomials and their use in surface / wavefront reconstruction

This section is a summary of Gradient polynomials and their use in surface / wavefront reconstruction, details of which can be found in appendices 1 and 2.

Chebyshev polynomials are a complete set of polynomials, orthogonal in the rectangular domain. They have many appealing properties that make them suitable for practical data fitting and analysis applications, for instance Chebyshev polynomials of the first kind hold discrete orthogonality in many cases [16]. Appendix A discusses many more advantages and properties of these polynomials.

2.1 $G$ polynomials: background and derivation

Chebyshev polynomials of the first kind can be defined as:

$$T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x)$$

where $T_0(x) = 1$, $T_1(x) = x$, for $-1 \leq x \leq 1$ \hspace{1cm} (2.1)

Two-dimensional Chebyshev polynomials were chosen as the scalar basis set upon which the vector polynomial sets ($G$ and $C$ polynomials) were based. This scalar set, named $F$ polynomials, can be defined as:
The polynomial order can be written in terms of the single index variable, $j$ or the double index variables, $n$ and $m$. Indexing details, including the conversion between $j$ and $n, m$ can be found in appendix A. In [17], the background, derivations, normalization factors, visualizations as well as examples of data fitting in the scalar domain, across rectangular apertures, for two-dimensional Chebyshev polynomials of the first kind (like the ones used in this work) are described. Fitting slope data directly in the measurement domain not only lets us exploit the advantages of orthonormality of the polynomial basis set, such as the fact that when fitting real data, noise propagation increases with the use of non-orthogonal basis functions [18] but also the advantage that specific modes can be related directly to the properties of the measurement system, which is necessary when $C$ polynomials are used for describing systematic errors in the measurement set-up.

In order to fit slope or gradient data directly in the slope domain, a set of vector polynomials is derived. These polynomials, that we named as the $G$ polynomials, are based on the gradients of $F$ polynomials, and can be written as:

$$
\begin{align*}
G_j(x, y) &= \tilde{G}_n^m(x, y) = \nabla F_n^m(x, y) = \frac{\partial}{\partial x} F_n^m(x, y) \hat{i} + \frac{\partial}{\partial y} F_n^m(x, y) \hat{j}, \\
\tilde{G}_j(x, y) &= T_n(y) T_m^\prime(x) \hat{i} + T_m(x) T_n^\prime(y) \hat{j}
\end{align*}
$$

(2.3)
The polynomial numbering scheme and relationship between the single \((j)\) and double \((n, m)\) index follows that of the \(F\) polynomials and is described in [7].

Our investigation (appendix A) shows the \(G\) polynomials to be orthogonal in the rectangular domain, and the normalization factor \((N_G)\) was derived to be:

\[
N_G = \begin{cases} 
\frac{\pi^2 n^2}{4}, & m = 0 \\
\frac{\pi^2 m^2}{4}, & n = 0 \\
\frac{\pi^2}{8}(m^2 + n^2), & \text{otherwise}
\end{cases}
\]  

(2.4)

Derivatives of the Chebyshev polynomial of the first kind \((T')\) can be related to Chebyshev polynomials of the second kind, named \(U\) polynomials [19], through Eqn. 2.5.

\[
U_{m-1}(x) = \frac{1}{m} T'_m(x) = \frac{\sin(m\theta)}{\sin(\theta)}, \quad x = \cos\theta
\]  

(2.5)

The \(G\) polynomial set can be expressed in terms of the \(T\) and \(U\) polynomials, as shown in Eqn. 2.6.

\[
\bar{G}_j(x, y) = mT_n(y)U_{m-1}(x)\hat{i} + nT_m(x)U_{n-1}(y)\hat{j}
\]  

(2.6)

This definition of the \(G\) polynomials allows them to be expressed in a simple, recursive way as recursive relations exist for both the \(T\) and \(U\) polynomial sets [9]. For example, in terms of the \(T\) polynomials, the recursive expression for \(G\) polynomials is given in Eqn. 2.7.
The first ten $G$ polynomials are listed in Table 2.1, where both single and double-indexing variables are shown as well.

\[ \hat{G}_j(x,y) = \frac{m}{2y(1-x^2)}(T_{n+1}(y) - T_{n-1}(y))(T_{m+1}(x) - xT_m(x))i \]
\[ + \frac{n}{2x(1-y^2)}(T_{m+1}(x) - T_{m-1}(x))(T_{n+1}(y) - yT_n(y))j \]  

(2.7)

2.2 Data fitting methodology and its properties

The data dealt with in this work, i.e., gradients of the surface / wavefront can be written in terms of the $G$ polynomials as:

\[ \nabla W(x,y) = \sum_{k=1}^{N} b_k \hat{G}_k \]  

(2.8)

where $\nabla W$ is the gradient of the surface or wavefront, $b_k$ are the expansion coefficients of the $G$ polynomials and $N$ is the total number of $G$ polynomial terms used in the expansion.
Table 2.1: Table of first ten $G$ polynomials [7].

<table>
<thead>
<tr>
<th>$j$</th>
<th>$m$</th>
<th>$n$</th>
<th>Explicit form of $\hat{G}_j(x, y)$</th>
<th>$\hat{G}_j(x, y)$ in terms of $F_j(x, y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$0i + 0j$</td>
<td>$0$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$i$</td>
<td>$\frac{T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>$j$</td>
<td>$\frac{T_0(x)[T_0(y) - yT_i(y)]}{(1-y^2)} j$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>$T'_1(x)i$</td>
<td>$\frac{2T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>$yi + xj$</td>
<td>$\frac{T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i + \frac{T_0(x)[T_0(y) - yT_i(y)]}{(1-y^2)} j$</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>$T'_1(y)j$</td>
<td>$\frac{2T_0(x)[T_0(y) - yT_i(y)]}{(1-y^2)} j$</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>$T'_1(x)j$</td>
<td>$\frac{3T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i$</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>$yT'_1(x)i + T'_1(x)j$</td>
<td>$\frac{2T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i + \frac{T_0(x)[T_0(y) - yT_i(y)]}{(1-y^2)} j$</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>$T'_1(y)i + xT'_1(y)j$</td>
<td>$\frac{T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i + 2\frac{T_0(x)[T_0(y) - yT_i(y)]}{(1-y^2)} j$</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>3</td>
<td>$T'_1(y)j$</td>
<td>$\frac{3T_0(x)[T_0(y) - yT_i(y)]}{(1-x^2)} j$</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
<td>$T'_1(x)j$</td>
<td>$\frac{4T_0(y)[T_0(x) - xT_i(x)]}{(1-x^2)} i$</td>
</tr>
</tbody>
</table>
Since all data (measured and simulated) is discrete, we can describe this modal fitting process in terms of a matrix operation as:

\[ D = \tilde{G}b \]  \hspace{1cm} (2.9)

where \( D \) is a column vector containing the values of the measured or simulated data (i.e., slopes of the surface or wavefront), \( b \) is a column vector containing the \( N \) expansion coefficients, and \( G \) is a matrix that contains the values of the \( G \) polynomials at the locations corresponding to the data points. Building from Eqn. 2.9, the values of \( b \) can be calculated from the slope data. This process is described in appendix A and a similar modal fitting process is explained in [20]. The next step is to determine the values of the coefficients of the scalar \((F)\) polynomials. It is shown in [7] that the coefficients of the \( G \) polynomials are exactly equal to the coefficients of the \( F \) polynomials, for each mode. These coefficients can be plugged into the following equation to obtain the surface or wavefront under test:

\[ W(x,y) = \sum_{j=1}^{N} a_j F_j \]  \hspace{1cm} (2.10)

\( W \) is the surface of wavefront whose gradients are measured by the metrology system and \( a_j \) are the expansion coefficients of the \( F \) polynomial set.

As can be seen from the above analysis and previously listed information, this methodology is very efficient (for example in terms of numerical efficiency) for three main
reasons: 1) the coefficients of the scalar and vector polynomials do not require any additional conversion steps, 2) obtaining the values of the $G$ polynomials is much more straightforward and easier than most vector polynomial sets as closed-form and recursive solutions have been derived for the $G$ polynomial set, and 3) both the scalar and vector polynomials are orthogonal and hence the vector polynomial set does not need any additional orthogonalization steps that most other vector based polynomial methods, such as [11] require. This increases both the accuracy and efficiency of the $G$ polynomial based fitting. Based on these properties, this reconstruction method has many practical advantages, that are discussed in the following sub-section.

2.3 Use of $G$ polynomials in surface and wavefront reconstructions from simulated and real, measured deflectometry data

The modal fitting methodology employing $G$ polynomials has many advantages, that were investigated, through mathematical development and software coding and for proof as well as to demonstrate their applicability, several data sets were processed using simulated or real data. References 1 and 7 contain examples showing these properties, and they are summarized in this section. Readers must note that all numbers quoted in these examples, in terms of accuracy, percentage error, time of reconstruction etc. depend on the specific numerical implementation. We have used basic implementations as a baseline comparison. Of course, better and advanced numerical implementations of all these techniques can change the values shown here. However,
the purpose of this research is to provide a general sense of the scale of accuracies, errors and processing times, so a basic implementation is deemed sufficient. Additionally, we have used the same or similar approaches (such as dealing with missing data i.e., NaN values) across all modal and zonal techniques, for consistency.

### 2.3.1 High fitting accuracy in the rectangular domain

The orthogonality of both $G$ and $F$ polynomials as well as the accuracy and efficiency of the data fitting method ensures that the rectangular aperture data reconstructed using this method represents the surface or wavefront under test correctly (i.e., with very low reconstruction error). To prove this, both simulated and real data was processed using the $G$ polynomial based method and compared to conventional methods. Fig. 2.1 [7] shows a simulated surface (Fig. 2.1 a) and its reconstructions obtained with the $G$ polynomial method, using 100 polynomials (Fig. 2.1 b) and the Southwell zonal method (Fig. 2.1 c), along with the surface RMS (root mean squared) value and the amount of reconstruction error. Residual error from the modal fit is $4.9799 \times 10^{-4}$ $\mu$m (or 0.0108% error relative to the surface RMS of the simulated map), which is lower than the $5.1335 \times 10^{-4}$ $\mu$m (or 0.0112% error relative to the RMS of the simulated surface) residual error obtained with Southwell integration. As stated earlier, these numbers are based on a straightforward numerical implementation of the two reconstruction methods. The exact values will depend on the specifics of the implementation scheme.
Of course, in general the reconstruction accuracy of polynomial fitting improves as the number of polynomials used in the fit increases [7]. However, the disadvantage of using more polynomials is increased processing time. For the example from Fig. 2.1, the $G$ polynomial method (using 100 polynomials) took 0.830 seconds while the Southwell integration took 0.389 seconds to complete.

### 2.3.2 Good fidelity for representation of mid-to-high spatial frequencies

For optical applications, such as telescope mirrors, where the mid-to-high spatial frequency content of the surface under test must be preserved, zonal reconstruction approaches (such as the Southwell zonal method) are traditionally preferred over modal methods. This is because a very large number of polynomials or modes, of high order are required to represent the mid-to-high
spatial frequencies. It is often difficult to generate a very large number of polynomial terms, or polynomial terms of high orders, especially for vector polynomials, for reasons including: (a) the lack of closed-form or recursive solutions for polynomial generation, (b) the need to orthogonalize the polynomial set to be used as the gradient basis, (c) numerical inefficiency in obtaining the scalar polynomial coefficients from the vector polynomial coefficients.

$G$ polynomials avoid these problems for the reasons mentioned in section 2.1.2. As mentioned before, $G$ polynomials can be expressed as a closed form equation using the $T$ and $U$ polynomials, both of which have simple, recursive relationships. Hence, it is straightforward and computationally efficient to generate the $G$ polynomial set. This proves a significant benefit for data processing applications as a large numbers of $G$ polynomials can be generated and used efficiently and reliably, even with restricted computing resources. The difference between recursive and direct calculation of $T$ polynomials of order 20,000 is only $10^{-11}$, generated over a grid of 50,000 points [7].

One example [7] showing the $G$ polynomial-based method’s capability of accurately representing mid-to-high spatial frequencies uses deflectometry data from the primary mirror of the Daniel K. Inouye Solar Telescope (DKIST) [21]. A square section of the high-resolution DKIST data set, representing $1.2 \times 1.2$ m area on the mirror was used. To focus on the mid-to-high spatial resolution features from the mirror, a high-pass Gaussian filter was applied to both the modally and zonally reconstructed surfaces, using SAGUARO data processing software [22].
A set of surface reconstructions were obtained from the $G$ polynomial-based fit, each of which was generated using different numbers of $G$ polynomials. Figure 2.2 shows the different high-pass-filtered modally reconstructed surfaces, compared with the reference high-pass-filtered Southwell surface.

![Figure 2.2](image)

Figure 2.2: (a) High-pass-filtered Southwell zonal reconstructed map and high-pass-filtered $G$-polynomial reconstructed maps generated using (b) 37, (c) 750, (d) 3000, and (e) 20,000 polynomial terms [7].

As the number of $G$ polynomial terms increases, shown left to right in Figures 2.2 (b) - (e), the surface reconstruction maps can be seen as being sharper and more high-frequency features become resolvable. The figure on the extreme right closely resemble the Southwell zonal reconstructed surface in Fig. 2.2 (a).

Another option for reconstructing this data modally would be to use a Zernike polynomial based vector set, orthogonalized over a rectangular pupil [11]. However, this basis set doesn’t
have straightforward closed-form equations for its generation, making it difficult to generate the extreme orders required to accurately represent this surface. Another probable problem in this approach is obtaining Zernike polynomials gradients, orthogonalizing the gradients over a rectangular aperture, and generating the conversion matrices for acquiring the scalar coefficients from the vector polynomial coefficients is cumbersome, and has the likelihood of having numerical errors due to the several computation steps, approximations and numerical truncations. Also, since the deflectometry systems used in this work are neither rotationally symmetric nor anamorphic, so the balanced aberration representations of Zernike-type [11] or Legendre polynomials [23] are not applicable.

2.3.3 Robustness against noise

All metrology systems suffer from some amount of noise. When defining the performance of a surface or wavefront reconstruction method, it is also important to investigate the effect of noise on the performance of the method. For this purpose, a simulation study was set-up [7] where different amounts of Gaussian white noise were added to a simulated data set, which was then reconstructed using the $G$ polynomial method (using the first 100 polynomials) and comparing it against the Southwell zonal method based reconstruction. Fig. 2.3 shows the results, as a plot of the RMS percentage error (relative to RMS of the simulated, ideal, noiseless data) as a function of the standard deviation of the white noise.
It is immediately obvious that the modal fit outperforms the zonal fit, for all amounts of noise used for this simulation and that the error in the Southwell fit increases at a much higher rate than the error in the modal fit, as the amount of noise is increased.

Figure 2.3: Residual errors from surface reconstruction in the presence of various amounts of noise [7].

2.3.4 High reconstruction accuracy in the presence of obscurations

One of the most important advantages of the $G$ polynomial based reconstruction is its ability to accurately reconstruct surfaces from measured slope data when the surface under test has a scratch or an obscuration. One such situation encountered frequently in deflectometry
measurements is when physical markers, called fiducials, are placed on the surface of the optic as a part of distortion correction. When a measurement is made while the fiducials are present on the surface under test, the region of the optic covered by the fiducial will have missing data. Another well-known case is for telescope apertures that often have blockers, commonly called spiders, that also create regions in the aperture with no data. This absence of data is represented in software processing as NaN (non-number) values, which often cause surface reconstruction errors because areas with NaN data affect their surrounding regions as well during the integration process. Although some methods have been developed to reduce this problem, it is still a source of error, specially for zonal reconstruction as the nature of the zonal approach fundamentally limits the reconstruction accuracy when a sub-region is not well-defined. In contrast, modal reconstructions are impacted much lesser by obscurations since modal reconstructions are based on an overall view of the data. One example to show this [7] models a scratch or defect on the surface of the optic. The defect is modeled as a 90×2 pixels line of NaN values, representing an area of 90×2 mm, in a 250×250 pixel grid of synthetic data, representing a 250×250 mm aperture. The simulated slope data is reconstructed with the $G$ polynomial method employing 2000 terms, as well as the Southwell zonal method. Fig. 2.4 shows the simulated surface and the fitting results. The modal fit shows a smaller residual error compared to the Southwell fit, especially in the vicinity of the defect.
The simulated surface has an RMS of 95.1220 µm. The RMS of the Southwell residual error map (i.e., difference between the Southwell reconstructed and simulated surfaces) is 0.8387 µm (or 0.8817% relative to the RMS of the simulated surface) and the $G$ polynomials based residual error map has an RMS of 0.0627 µm (0.0659% error relative to the simulated surface), which is approximately 13 times lesser than that of the zonal method, in this case.

As the number of polynomials used for the fit increases, the residual fitting error decreases. Table 1 shows this for the case of the defected surface where the RMS errors of the difference maps (the differences of the modal fits from the ideal surface), obtained when using different number of polynomial terms in the fit is shown.
Table 2.2: RMS residual errors of the differences between the simulated and modal reconstructions [7].

<table>
<thead>
<tr>
<th>Number of $G$ Polynomials</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS error (µm)</td>
<td>1.4374</td>
<td>0.5032</td>
<td>0.0739</td>
<td>0.0627</td>
<td>0.0568</td>
</tr>
<tr>
<td>[RMS of difference map / RMS of ideal map] × 100 (%)</td>
<td>1.5111</td>
<td>0.5290</td>
<td>0.0777</td>
<td>0.0659</td>
<td>0.0597</td>
</tr>
</tbody>
</table>

Another example from [7] and appendix A models a telescope aperture with blockers (called ‘spiders’) as well as rounded edges of the rectangular aperture. A freeform wavefront is modeled over this aperture and the simulated gradient data is reconstructed by three methods: (a) Zernike based gradient polynomials [18] (b) Southwell zonal method (c) $G$ polynomials based method. Fig. 2.5 shows the simulated wavefront as well as the three reconstructions (both the modal reconstructions used 30 polynomials each). The RMS of the ideal (simulated) surface is 0.6792 µm. The residual (difference between the ideal surface and its reconstruction) RMS error for the Zernike based polynomial fit is 0.0745 µm (or 10.9625% relative to the RMS of the ideal surface). The residual RMS error for the Southwell zonal reconstruction is comparable, with 0.0775 µm of absolute error (or 11.4082% relative to the RMS of the ideal surface), whereas the $G$ polynomials fit is much better than both the other methods, as it has a residual RMS error of 0.0379 µm (or 2.5698 % error relative to the RMS of the ideal surface).
For both the modal basis sets, as the number of polynomials is increased, the fit becomes more and more accurate. However, it also takes longer for the data to be processed. Fig. 2.6 shows the comparison of reconstruction processing times for the two sets of polynomials, as the number of polynomials is increased. The $G$ polynomial fit is significantly faster, especially as the number of polynomials used increases. The maximum number of modes considered is 180 because for higher polynomial terms, the Zernike gradient polynomial based modal fit becomes numerically unstable (although the $G$ polynomial based fit remains numerically stable for up-to several thousands of polynomials).
Good reconstruction accuracy for unevenly sampled data

The $G$ polynomial method also shows good reconstruction accuracy when data is unevenly sampled. This could be due to variable sampling of data across the aperture, or a different sampling in the $x$ and $y$ directions etc. One example of this is the data from A-SHWFS, which is described later in chapter 4. One of the most useful and unique properties of the sensor is its ability to dynamically block lenslets locally, over a specific region that is severely aberrated compared to the rest of the beam. This results in different parts of the beam being spatially sampled at different rates (the more aberrated regions have lenslets blocked off periodically and
hence have a lower sampling). The prototype A-SHWFS we designed [15] is non-rotationally symmetric and has a rectangular pupil so the $G$ polynomial fitting works very well for it. One example, shown in [1] and appendix B, to demonstrate this modal fitting capability uses simulated data with the same sampling in the $x$ and $y$ directions but the sampling along the aperture (for both $x$ and $y$ directions) changes across the length of the aperture. The full dataset models a $2 \times 2$ cm aperture, the first 1.25 cm of which is sampled with 150 points, making the sampling $1/120$ cm in that region. 150 points are also used to sample the remaining 0.75 cm region of the wavefront, making the sampling in this section $1/200$ cm. The results of fitting the slope data from this simulation with both the zonal and modal methods are shown in Fig. 2.7. Data was simulated over the entire aperture, using Eqn. 2.11.

$$z = 0.15y^2 + 1.4x^4 - 0.2(x^3 + xy^2) + 0.3y^5 + 1 \times 10^{-3}x^{11} - (xy)^{18}$$

(2.11)

Figure 2.7: (a) Simulated, ideal surface map (reference). The blue double ended arrows mark the approximate region on the map ($1.25 \times 1.25$ cm) with a lower sampling (b) Error between the reference and Southwell methods generated surface (c) Error between the reference and $G$ polynomial generated surface.
The RMS of the simulated surface is 0.392 µm. The Southwell zonal reconstruction map has an RMS error of $8.68 \times 10^{-3}$ µm, which is 2.22% error relative to the simulated surface’s RMS. The $G$ polynomial method yields a surface with RMS error map of $2.66 \times 10^{-3}$ µm, or 0.68% relative to the ideal surface RMS, when 250 $G$ polynomials are used. Thus, the modal method achieves an accuracy (in terms of residual RMS error) of over 3 times the Southwell method.

### 2.4 Concluding discussion on $G$ polynomials

The ideal number of polynomials to be used for surface or wavefront reconstruction depends on the specifics on the application – primarily the required accuracy, requisite speed of data processing and importance of mid-to-high spatial frequency content. Of course, for cases with missing data (measurement with fiducials, telescope spiders etc.), more polynomials are required for the same level of accuracy than for the same dataset without any missing data.

It is also worthwhile to point out that there are many data processing and computing actions that can make the polynomial fitting process faster and / or more accurate. One of these include interpolation for missing data, instead of simply replacing it with zeros, which can make the fit more accurate and may also require less polynomials to achieve the desired level of accuracy. For improving the processing speed, one method is to generate and save the $G$ and $F$ polynomial values over a known grid, for a specific number of data points. As long as the aperture has the same grid size and sampling, the saved $G$ and $F$ polynomial sets can simply be
re-loaded and re-used for different measurements. This works even if a lesser number of polynomials are required for the fit (for which the saved polynomial data sets can easily be trimmed to the required size). This will save processing power and time, by eliminating the need to generate the data sets again.

Lastly, another interesting application that uses the $G$ polynomials has recently been published in scientific literature [24]. This work is an analytic methodology that guides the user as to what surface within an optical design should be selected to apply freeform optimization to. This approach has been shown to be effective even for an optical system with many freeform / non-freeform optical surfaces.
Chapter 3

Curl polynomials and their use in modeling errors in metrology systems

This chapter summarizes the discussion on curl polynomials and their applications in modeling errors in metrology systems. The in-depth work on this topic is presented in appendix C.

3.1 C polynomials: background and derivation

The $G$ polynomial set is a complete set of gradient polynomials, orthogonal in the rectangular domain and capable of representing data obtained as gradients of a surface or wavefront. To fully represent all vector data in the rectangular domain, another set of polynomials are needed, that can characterize parts of vector data not represented by the gradient polynomials. These polynomials, that we named as $C$ polynomials, are defined locally as a rotation or curl and have zero divergence. The $C$ polynomials themselves or the combined set of both $G$ and $C$ polynomials can be very useful in quantifying and providing corrections for mapping distortion and errors in the metrology system, such as alignment or systematic errors. The $C$ polynomial set, like its $G$ polynomial counterpart, is derived from the two-dimensional Chebyshev polynomials of the first kind and is orthogonal in the rectangular domain. Using the same definitions for the $T$ and $F$ polynomial sets as stated in chapter 2, we can define $C$ polynomials as:
The single index \((j)\) and double index \((n \text{ and } m)\) follow the same conventions and conversion procedure as for the \(G\) polynomials. More details on the curl polynomial set’s derivation and mathematical background is stated in appendix C. The normalization factor for \(C\) polynomials, called \(NC\) is given in Eqn. (3.2) and is the same as the normalization factor for \(G\) polynomials.

Table 3.1 lists the first ten \(C\) polynomials, with the single and double-indexing variables shown.

\[
\tilde{C}_j(x, y) = \tilde{C}_n^m(x, y) = \frac{\partial}{\partial y} F_n^m(x, y) \hat{i} - \frac{\partial}{\partial x} F_n^m(x, y) \hat{j},
\]

\[
\tilde{C}_n^m(x, y) = T_m(x) T_n^r(y) \hat{i} - T_n(y) T_m^r(x) \hat{j}
\]

(3.1)

Like the \(G\) polynomials, recursive relations were derived for the \(C\) polynomials as well. In terms of the \(T\) polynomials, \(C\) polynomials can be expressed recursively as:

\[
\tilde{C}_n^m(x, y) = \frac{n}{2x(1-y^2)}(T_{nv1}(x) - T_{m-1}(x))(T_{n-1}(y) - yT_n(y)) \hat{i}
\]

\[
- \frac{m}{2y(1-x^2)}(T_{nv1}(y) - T_{n-1}(y))(T_{m-1}(x) - xT_m(x)) \hat{j}
\]

(3.3)

\[
NC = \begin{cases} 
\frac{\pi^2 n^2}{4}, & m = 0 \\
\frac{\pi^2 m^2}{4}, & n = 0 \\
\pi^2 (m^2 + n^2), & \text{otherwise}
\end{cases}
\]

(3.2)

Like the \(G\) polynomials, recursive relations were derived for the \(C\) polynomials as well. In terms of the \(T\) polynomials, \(C\) polynomials can be expressed recursively as:
Table 3.1: Table of the first ten C polynomials [13].

<table>
<thead>
<tr>
<th>$m$</th>
<th>$n$</th>
<th>$\tilde{C}_j$</th>
<th>Simplified form</th>
<th>Expressed as Chebyshev polynomials of first kind</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>$\tilde{C}_0$</td>
<td>$0\hat{i} - 0\hat{j}$</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$\tilde{C}_1$</td>
<td>$-\hat{j}$</td>
<td>$-\frac{T_0(y)[T_0(x) - xT_1(x)]}{(1-x^2)} \hat{j}$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>$\tilde{C}_2$</td>
<td>$\hat{i}$</td>
<td>$\frac{T_0(x)[T_0(y) - yT_1(y)]}{(1-y^2)} \hat{i}$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$\tilde{C}_3$</td>
<td>$-T_2'(x)\hat{j}$</td>
<td>$-2\frac{T_0(y)[T_1(x) - xT_2(x)]}{(1-x^2)} \hat{j}$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>$\tilde{C}_4$</td>
<td>$-y\hat{j} + x\hat{i}$</td>
<td>$-\frac{T_1(y)[T_0(x) - xT_1(x)]}{(1-x^2)} \hat{j} + \frac{T_1(x)[T_0(y) - yT_1(y)]}{(1-y^2)} \hat{i}$</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>$\tilde{C}_5$</td>
<td>$T_2'(y)\hat{i}$</td>
<td>$2\frac{T_0(x)[T_1(y) - yT_2(y)]}{(1-y^2)} \hat{i}$</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>$\tilde{C}_6$</td>
<td>$-T_3'(x)\hat{j}$</td>
<td>$-3\frac{T_0(y)[T_2(x) - xT_3(x)]}{(1-x^2)} \hat{j}$</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>$\tilde{C}_7$</td>
<td>$-yT_2'(x)\hat{j} + T_2(x)\hat{i}$</td>
<td>$-2\frac{T_1(y)[T_1(x) - xT_2(x)]}{(1-x^2)} \hat{j} + \frac{T_2(x)[T_0(y) - yT_1(y)]}{(1-y^2)} \hat{i}$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>$\tilde{C}_8$</td>
<td>$-T_2(y)\hat{j} + xT_2'(y)\hat{i}$</td>
<td>$-\frac{T_2(y)[T_0(x) - xT_1(x)]}{(1-x^2)} \hat{j} + 2\frac{T_2(x)[T_1(y) - yT_2(y)]}{(1-y^2)} \hat{i}$</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>$\tilde{C}_9$</td>
<td>$T_3'(y)\hat{i}$</td>
<td>$3\frac{T_0(x)[T_1(y) - yT_1(y)]}{(1-x^2)} \hat{i}$</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>$\tilde{C}_{10}$</td>
<td>$-T_4'(x)\hat{j}$</td>
<td>$-4\frac{T_0(y)[T_3(x) - xT_4(x)]}{(1-x^2)} \hat{j}$</td>
</tr>
</tbody>
</table>
The $G$ polynomials can be thought of as irrotational vector fields, with zero curl everywhere and $C$ polynomials as solenoidal vector fields with zero divergence everywhere. These two fields overlap in the Laplacian vector domain, where both divergence and curl are zero. Consequently, both the $G$ and $C$ polynomials will contain common terms, for which the Laplacian is zero. If the scalar function used for generating both the gradient and curl polynomial sets, i.e., the $F$ polynomial set, represents wavefront, the overlapping terms correspond to a wavefront with zero net curvature at any point in the pupil. The Laplacian is a mathematical operator that can be defined as the divergence of the gradient of a function. It can be calculated both empirically, using the sine and cosine based definition of the $G$ polynomials [7] and numerically, from either the $G$ or $C$ polynomials.

For applications where only the $G$ or $C$ polynomials are used, the overlapping terms with Laplacian = 0 should be included in the set. However, if we want to use a combination of the two sets together, it must be ensured that the Laplacian = 0 terms aren’t counted twice. They should only be present in one of the two polynomial sets. One implementation would be remove the $G$ polynomials terms whose divergence is zero from the $G$ basis set. We implemented this in software codes, which makes it very efficient and accurate to get the combined $G$ and $C$ polynomial sets, even for very high polynomial orders.
3.2 Applications of $C$ polynomials for modeling errors in deflectometry systems

The $C$ polynomial set has zero divergence everywhere and represents local curl. Hence, it can be used to represent the effect of misalignments and system errors on measured or simulated metrology data. A well-known deflectometry system, called Scanning Long-wave Optical Test System, abbreviated as SLOTS [25] is an infrared (IR) deflectometry system, that measures the surface slopes of the UUT (unit under test) by capturing the UUT’s specular reflection from a scanning hot wire. This line source scans across the aperture first in the $x$ and then in the $y$ direction (or vice versa). Ideally, the $x$ and $y$ scans should be exactly orthogonal to each other. However, practically it is not always possible to get an exact right angle between the two scan directions and there exist limitations even in determining this angle precisely. The curl polynomials can be a powerful tool for this analysis. Using existing measurement data, an analysis involving $C$ polynomials can determine if there is any clocking (deviation from orthogonality) between the two scanning directions and how much of the clocking is present. Then, either the effects of the clocking can be calibrated out from the measurement or the system set-up and alignment can be improved based on this information. Reference [13] gives an example of a similar clocking correction analysis for the SCOTS (Software configurable optical test system) deflectometry system [26], which uses visible light, often from a monitor or another such LCD screen, as the source.
First, a numerical model based on $C$ polynomials was created. To obtain this model, we measured a freeform Alvarez lens (details of which can be found in [13]) using the SCOTS set-up. The first measurement was made with the lens set at a nominal, reference position (corresponding to 0° clocking). Then the lens was rotated every 0.5° till it reached a total rotation of 5°, and at each step a measurement was made. Each measurement consisted of $x$ and $y$ slope data. To replicate the effect of mis-clocking (such as non-orthogonality in the SLOTS scanning directions), the input for surface reconstruction was one slope map from the rotated measurement and the other at 0°. For instance, to represent a 3.5° mis-clocking, the $x$ slope from the 3.5° measurement and the $y$ slope from a 0° metrology were used. The eleven mis-clocking measurements (from 0.5° to 5.0°, taken every 0.5°) were fitted with the first 50 $C$ polynomials. When all eleven measurements were analyzed together, it was observed that specific $C$ polynomial terms showed a consistent increasing or decreasing trend as the mis-clocking was increased. These polynomials terms were $j = 1, 2, 4, 8, 11, 13$ and 17, where $j$ is the single index numbering for the $C$ polynomials, as defined in [13]. The selected polynomial terms and their calculated values were used to develop a regression type model, which could be used to calculate the mis-clocking value for any other measurements. To test the accuracy of the model, two additional mis-clocking measurements were made: one for 0.6° and the second for 2.3°. The $C$ polynomial coefficient values from the measured slope data were given to the predictive model and yielded results with accuracy greater than 0.1°. The results are summarized in Table 3.2.
Table 3.2: C polynomial-based mis-clocking estimation performance for the Alvarez lens deflectometry data analysis case [13].

<table>
<thead>
<tr>
<th>True (T) [Degree]</th>
<th>Computed (C) [Degree]</th>
<th>Error (T - C) [Degree]</th>
<th>% Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.6024</td>
<td>-0.0024</td>
<td>-0.4000</td>
</tr>
<tr>
<td>2.3</td>
<td>2.3917</td>
<td>-0.0917</td>
<td>-3.9870</td>
</tr>
</tbody>
</table>

This experiment proves the hypothesis that certain C polynomials can be used to characterize the errors in a measurement system. Further study on this topic could relate the specific C polynomial terms used for the regression model to other physical parameters or sources of error in the system. This thought process can be extended to other optical metrology systems as well.

### 3.3 Modeling Imaging Distortion using G and C polynomials

The combined set of G and C polynomials can be used to fit any vector data in the rectangular domain. One highly useful example of this is mapping imaging distortion in optical systems. This is shown through a simulated study in [13] and appendix C that models distortion in a simulated Three Mirror Anastigmat (TMA) with rectangular aperture, obtained from the Zemax standard library [27]. The optical layout of this design, obtained from Zemax files [27] is shown in Fig. 3.1(a) and its distortion map in Fig. 3.1(b) (also obtained from the Zemax file).
Figure 3.1: (a) Schematic of the simulated TMA camera system in the Zemax library [27] (b) Distortion map for this system using 20×20 grid sampling with 15× arrow size magnification [27].

This distortion data is essentially a position map, that tells the sampled image location values, in \( x \) and \( y \), for both the ideal and actual location. The difference between the two sets of values is proportional to the slope of wavefront distortion. This slope data is fitted with the combined \( G \) and \( C \) polynomial sets, as described in section 3.1. For this fitting, twelve polynomial terms were used (6 \( G \) and 6 \( C \)). For comparison, the same data was also fit with another vector polynomial set, that consists of gradient and curl polynomial terms as well. This set, called the S (gradient polynomials) [18] and T (curl polynomials) [28], is based on Zernike polynomials and is orthogonal across a circular aperture. The combined Zernike-based set used for fitting also
consisted of a total of 12 terms (6 $S$ and 6 $T$). For both the fits, the respective Laplacian = 0 terms were only counted once. Table 3.2 shows the results of the two fits, where the real error is the RMS error between the position predicted by the fitting and the actual position from the simulation, while % error is the actual error, expressed as a percentage of the RMS of actual position. As described in section 2.3, the specific numerical implementation of both these modal techniques will dictate the exact error values we obtain. The values given here are based on a simple numerical implementation, that is deemed sufficient for highlighting the capabilities of the $G$ and $C$ polynomial sets.

Table 3.3: Fitting performance comparison of the two vector polynomial set cases to represent/model the TMA camera’s distortion map [13].

<table>
<thead>
<tr>
<th></th>
<th>$S$ &amp; $T$ polynomial fitting</th>
<th>$G$ &amp; $C$ polynomial fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$x$-position</td>
<td>$y$-position</td>
</tr>
<tr>
<td>Real Error [µm]</td>
<td>0.593</td>
<td>0.654</td>
</tr>
<tr>
<td>% Error [%]</td>
<td>12.44</td>
<td>8.30</td>
</tr>
</tbody>
</table>

The $G$ and $C$ polynomial fitting outperforms the $S$ and $T$ polynomial fitting by having approximately 6% less average error (relative to the actual position based on values from the Zemax simulation). The average percentage error is defined as the average of the percentage errors in the $x$ and $y$ directions.
Distortion correction can be improved by adding more polynomial terms to the fit. One reason for choosing only 12 polynomials in this case was that the $S$ and $T$ polynomials do not have a straightforward way to calculate terms with $\text{Laplacian} = 0$ hence, determining the complete set needed reliance on overlap terms listed in [28], which were only given up to this order and to make a fair comparison between the two sets of polynomials, the same number of terms should be used in both cases. As mentioned before, one major advantage of the $G$ and $C$ polynomials is the ease and efficiency of various calculations, such as the calculation to determine the $\text{Laplacian} = 0$ terms. This ensures that the complete set of $G$ and $C$ polynomials, only including the overlapping Laplacian terms once, can easily be determined even for high polynomial orders.
Chapter 4

Adaptive Shack–Hartmann Wavefront Sensor for increased dynamic range of measurement

This chapter is a summary of the last major portion of this work, and its details are presented in appendix D.

4.1 Adaptive Shack–Hartmann Wavefront Sensor: basic concept and its implementation

As mentioned in the introduction, the Shack–Hartmann wavefront sensor (SHWFS) is a device that measures the local slopes of the wavefront under test. It uses an array of small lenses, called lenslets, to sample the wavefront. The detector is divided (virtually) into sub-sections, called detector subapertures, as shown in Fig. 1.2. In a nominal or unaberrated situation, a collimated beam of light passing through the lenslet array will result in small focused spots at the nominal, reference positions on the detector. For a typical SHWFS, each lenslet corresponds to a single detector subaperture; so a collimated, unaberrated beam of light will produce a focused spot at each detector subaperture, at its nominal position (e.g., the center of the subaperture). When an aberrated wavefront passes through the lenslet array, the focused spots of light move away from the nominal or reference situation. The amount of departure of a focused spot relative to its reference position is proportional to the local tip-tilt component of the aberration. Using the
measured wavefront slope values, the wavefront under test can be reconstructed with the help of a zonal or modal approach, as described earlier for the case of reconstructions from measured deflectometry data (section 1.1). While designing a generic SHWFS, a trade-off is made between the spatial resolution of measurement and the dynamic range of the sensor. A smaller detection subaperture size means the wavefront is sampled more finely and a higher spatial resolution is obtained for the measurement but the maximum amount of aberration measurable for a fixed focal length becomes limited. In other words, the sensor has a smaller dynamic range. Conventional SHWFSs have a fixed dynamic range of measurement. We introduce a novel idea ([15] and appendix D) to increase the dynamic range of measurement and allow the sensor to reconfigure adaptively, based on the amount and kind of aberration so that the best possible spatial resolution is achieved while being able to measure relatively large or small local wavefront slopes. This concept, named Adaptive Shack–Hartmann Wavefront Sensor (A-SHWFS), is an active approach that uses an electronically modulated mask to block or unblock lenslets in any desired pattern. If a lens or group of lenses is blocked, the corresponding detector subaperture size increases in that region and so larger slope measurements are possible. Of course, since fewer lenslets are now sampling the same area of the wavefront, the spatial resolution of measurement also decreases in that region.

To apply the electronically modulated masking that blocks desired lenslets, a liquid crystal display (LCD) panel, without its back-light illumination unit, is placed in front of the
lenslet array. Each pixel of the LCD can be electronically addressed through a computing device. By matching the pixels of the LCD to the lenslets, the amount of light passing through the lenslets can be controlled. This is shown schematically in Fig. 4.1. The electronic mask can be applied symmetrically e.g., we can block every other row of pixels in the vertical direction or it can be applied in an uneven, non-symmetric pattern. In addition to the type of mask, we can also choose which regions of the lenslet array to apply the mask to and the rest of the aperture can be left unmasked. This would be the situation when we want to measure wavefronts that have a large slope variation locally, only over a section of the entire wavefront. Then, the region of the lenslet array through which the highly aberrated part of the beam passes is strongly masked (i.e., more lenslets in a given region are blocked, making the wavefront sampling in that region lower) and the detected wavefront from that region will be coarsely sampled while the rest of the array either remains unblocked or has a smaller blocking pattern (i.e., the wavefront sampling remains finer in that region). In this way, we ensure that all the wavefront is reliable measured while maintaining the best possible spatial resolution.
Figure 4.1: (a) Schematic layout of the A-SHWFS using actively modulated LCD lenslet array mask. (b) Depiction of how the LCD lenslet array mask (left) and the corresponding detector subapertures (right) are changed dynamically, from a fully unblocked (top) to partially blocked (bottom) situation. The squares on the left represent the pattern sent from the computer to the LCD screen [15].

For testing our concept, we developed an in-house customized lenslet array [15] made on a Poly(methyl methacrylate), or PMMA, substrate by using the single point diamond turning (SPDT) method.

To properly and completely implement the A-SHWFS concept, we also did a significant amount of mathematical development and software coding. Among these, the two substantial topics were (a) a centroiding algorithm for determining the focused spots locations and (b) a modal method to reconstruct the wavefront being tested, from its measured slope values. Many SHWFSs apply some centroiding in image processing software as the focused spots at the detection plane usually have some spread but for the A-SHWFS this is especially important because the focal plane image spots are more degraded compared to the standard SHWFS
without an LCD. This is because of scattering from the LCD’s internal structures as well as
diffraction effects from the diamond-turned tool marks on the lenslet array. Hence, we developed
an intensity-weighted centroiding algorithm, that helped determine the statistical centroid of the
image spots. The implementation details of this algorithm are provided in appendix D. The
wavefront reconstruction algorithm uses the $G$ polynomials based fitting method described in
chapter 2. This method suits the A-SHWFS very well because it has a rectangular aperture, the
property measured is wavefront slope and the system is non-rotationally symmetric. For future
development, a $C$ polynomials-based analysis could be applied to this system for determining
and calibrating its systematic errors and misalignment. Also, the $G$ and $C$ polynomials could be
used to conduct a thorough investigation of various system parameters or a sensitivity analysis as
the various polynomial terms or modes can be correlated to physical properties or errors of the
system.

An increased dynamic range of measurement may be achieved by modifying other
system parameters, such as the focal length of the lenslets. However, it would be very difficult to
implement those changes in existing SHWFS systems because it would require customized and
fixed hardware and possibly software changes to the system such as switching out the lenslet
array. One of the most attractive qualities of our technique is that it can be assimilated into most
existing setups without a major system overhaul. All that is required is a one-time adjustment to
the system set-up, after which the electronic masking is implemented purely through software
control, and it can be completely turned off easily through software commands, without making any additional hardware or system set-up changes.

Another appealing property of this method is that it also does not need any complicated devices or heavy computing resources to function. One of the ways this technique is novel is that it allows for a highly efficient, even irregular, and greatly adaptive reconfiguration of the detection subapertures. This reconfiguration may even be done in real-time and a feedback loop can be set-up where the detector subapertures are reconfigured dynamically using the measured value of the wavefront aberration. This will ensure that the wavefront being measured has the best trade-off between sufficient dynamic range and spatial resolution, even if the wavefront keeps changing over time. Most other techniques that try to use a similar concept do not have the ability to dynamically change the sampling size and do not have different sampling sizes and configurations for different parts of the aperture, in a simple and computationally efficient way either.

4.2 System prototype

This work extended the concept and implementation scheme of the A-SHWFS into a prototype design, leveraging a fully matured economical solution, with an electronically modulated LCD panel and efficient software controls. The light source is a 532 nm laser, passed through a spatial filter. A 200 mm focal length lens collimates the filtered beam and this collimated light is the
reference wavefront for the sensor. A three-mirror system is used next, the purpose of which is to introduce systematic wavefront aberration. These mirrors are all flat and have a hexagonal shape. In order to generate localized high aberration in the system, two of the three hexagonal mirrors are kept stationary while the third mirror is tilted in a controlled and measured manner. The moving mirror has an electronically controlled precision tilt actuator connected to it, that allows the mirror to tilt continuously at a controlled pace. The portion of the beam reflected from this mirror provides the section of wavefront that creates a large, localized (relative to the entire beam reflected from the three mirrors) tilt, which is used to demonstrate the adaptive sampling competency of the A-SHWFS. The other two mirrors both remain stationary and the portion of beam reflected off these two mirrors corresponds to the section of wavefront that doesn’t change and hence the detection area which this light corresponds to will have no change in sampling. The collimating lens is the system stop. The beam overfills the detector and ensures that there is enough light to fill the detector even when the moving hexagonal mirror is tilted substantially. This also means that the system aperture is defined by the detector, which is a 20 Megapixel CMOS (Complementary metal–oxide–semiconductor) with a rectangular sensor of dimensions 32.77 × 24.58 mm. The segmented mirror configuration was chosen to symbolize a locally varying wavefront with large magnitude variations, the proper detection and best-possible spatial sampling of which is one of the most important target applications of the A-SHWFS concept. Fig. 4.2 shows an actual picture of this prototype A-SHWFS.
As mentioned in section 4.1, the lenslet array for the prototype A-SHWFS was custom designed and built in-house, at the University of Arizona. The lenslet array size is $40 \times 40$ mm and it consists of $35 \times 35$ lenslets (with missing corner regions). Each lenslet was designed to have a focal length of 20 mm and it covers $6 \times 6$ pixels of the LCD. This means that in order to fully block out one lenslet, the corresponding $6 \times 6$ LCD pixels will have to be electronically masked. The pixel size and pitch for the LCD as well as the lenslet size and pitch were both matched and verified using microscopes. Based on the system parameters, the native dynamic range of a single A-SHWFS lenslet in the $x$ and $y$ directions was calculated to be +/- 28.5 mrad. This means
that without the adaptive sampling and detector subaperture reconfiguration, this sensor would only be able to correctly measure wavefront slope values less than +/- 28.5 mrad.

4.3 Experimental results
Two experiments were performed with the prototype A-SHWFS, one to verify the basic performance of the sensor as well as to measure its accuracy and the second to demonstrate its novel adaptive sampling and re-configuration ability. A simulation study was also carried out to further express the capabilities of the A-SHWFS concept under various circumstances. A brief description of the experimental set-ups and their results are provided below. Further details can be found in appendix D.

4.3.1 Determining the A-SHWFS accuracy of measurement
To determine the prototype sensor’s measurement accuracy, the same wavefront slope values were measured from the A-SHWFS and from a commercial precision autocollimator. The autocollimator has an accuracy of ±1.21×10⁻³ mrad and resolution of 2×10⁻⁴ mrad. For this experiment, the three hexagonal mirrors set-up, introduced in section 4.2, and shown in Fig. 4.3 was used, where two mirrors remained stationary while the third is systematically moved so that a tilt is added to the reflected beam. A small, circular mirror was attached to the back-side of the moving hexagonal mirror (Fig. 4.3 (b)). The autocollimator is set-up facing the back-side of the three hexagonal mirrors set-up in such a way that light from the autocollimator reflects off the
small circular mirror and goes back to the autocollimator, which measures the orientation of the small mirror with respect to its optical axis (i.e., reference) as shown in Fig. 4.3 (b). As the moving hexagonal mirror is tilted, the autocollimator measures the co-mounted small mirror’s orientation which provides the comparison data for the A-SHWFS’s accuracy test.

Two types of measurements were made for this experiment: one with small motions of the active mirror’s precision actuator, which introduced small magnitudes of tilt in the beam and the second with a larger step sizes for the actuator motion, which introduced relatively larger amounts of tilt.
in the test wavefront. For both cases, the error is defined as the difference between the autocollimator value and the A-SHWFS value while the total percentage error is the average of all individual percentage error values, calculated as the actual error as a percentage of the reference (autocollimator) value.

The prototype A-SHWFS’s measurements have a 1.32 % average error for the small tilt values case and 1.13 % for the large tilt values case. The actual errors between the two sets of measurements is plotted in Fig. 4.4.

Figure 4.4: (a) Small, and (b) Large magnitude of tilt measurements comparing the autocollimator and A-SHWFS values (blue circles). The error (red crosses), that is the difference between the autocollimator and A-SHWFS values, is shown with its own axis on the right side in both figures. A linear fit to the data is shown as the black line in both graphs [15].
4.3.2 Displaying the adaptive wavefront sampling and dynamic detector subaperture reconfiguration abilities of the A-SHWFS

As outlined in section 4.1, one of the most important applications of the A-SHWFS is its ability to detect localized high-slope magnitude wavefronts while maintaining the best possible spatial resolution over the entire wavefront. To display this ability of the sensor, an experiment was set-up where the measured wavefront had a specific region over which it was aberrated while the rest of the wavefront had almost no aberration. This was achieved with the help of the three hexagonal mirrors set-up described before, and pictorially shown in Fig. 4.3. As the active mirror is tilted more and more, the tilt added to the reflected beam’s high aberrated region increases while the unaberrated region of the beam (part of the beam reflected from the two stationary mirrors) remains as-is throughout the experiment.

At the start of the experiment when the active mirror has no tilt or a very small amount of tilt, all the detection subapertures, over the entire detection area, have the same size. This is the smallest possible detector subaperture size, which corresponds to a situation of no or very small aberration. As the moving mirror adds a high enough local tilt where the wavefront slopes increase beyond the measurement capability of the smallest detection subaperture size (i.e., the focused image spot moves outside or very close to the boundary of the detection subaperture), electronic masking is activated. The mask is applied in such a way that every other lenslet is
blocked in both the horizontal and vertical directions. This decreases the sampling of the wavefront but increases the detector subapertures size and the wavefront can now be accurately measured by the sensor. It is important to note that the increase in detector subaperture size occurs only locally, in the region corresponding to the highly aberrated part of the wavefront. The rest of the detection area maintains the native (smallest possible detector subaperture size) configuration. So, the sampling of the wavefront doesn’t change in this nearly unaberrated region (some very small aberration may still be present because of measurement instability, noise etc.). These situations are shown in Fig. 4.5, which shows images from the CMOS detector of the focused spots on the imaging plane. The figure on the left (Fig. 4.5 (a)) shows the measurement near the start of the experiment, when a very small amount of tilt was added to the beam. All the regions of the wavefront have the same sampling size / frequency. The figure on the right (Fig. 4.5(b)) shows the sensor when the local tilt was large enough to cause the detection subapertures to reconfigure. Only the detection region from the highly aberrated wave has a reduced sampling, shown as much lesser (and farther spread out) image spots than for the low aberration case. Overlaid on both images in Fig. 4.5 are red arrows that plot the motion of the spots relative to the reference positions (when there was no tilt in the beam). The arrows in Fig. 4.5 (a) are much shorter, indicating a small distance traveled by the image spots while the ones in 4.5 (b) are much longer, indicating the larger distance traveled. Of course, the number of arrows also
reduces from Fig. 4.5(a) to Fig. 4.5(b) since many lenslets are blocked when the adaptive gating starts.

![Figure 4.5: Spots from the lenslets, on the CMOS detector. Red arrows show the spot motion relative to reference spots, on the active mirror portion of the image. (a) For small tilt, all three mirrors display the maximum number of spots (i.e., higher spatial sampling of the wavefront). The red arrows are shorter. (b) When the amount of tilt increases, the stationary mirror zones maintain the same number of spots while spots from the actively moving mirror have been selectively blocked by the LCD panel to increase the slope measurement dynamic range only in the optimal zone. Hence, the red arrows are longer [15].](image)

Tilt was added to a localized section of the beam in a controlled, steady way, by maintaining a constant speed of motion for the precision actuator that tilts the mirror. Images from the sensor were captured at regular intervals of time, as the mirror was being tilted and each of the images, or frames from the sensor, were used as the data for the image processing software we developed that detected the focused spots, centroidsed them (as explained in section 4.1), calculated their
difference from the reference location, and then used this slope data to reconstruct the wavefront using the $G$ polynomial modal method described in chapter 2. Fig. 4.6 shows the wavefront reconstructions from a portion of the titled part of the beam, for a few measurements. The average slope value for each measurement is also shown, along with the timestamps of the time that particular frame was captured by the CMOS sensor. From left to right, as the time increases, the slope of the wavefront also increases proportionally. The red dotted line in the center indicates the moment when the adaptive gating started for this measurement.

![Figure 4.6](image-url)

Figure 4.6: Enhanced dynamic range demonstration of the adaptive wavefront sensing approach. The reconstructed wavefront time-lapse shows continuously increasing wavefront tilt as the active mirror was being tilted up-to and beyond the nominal 28.5 mrad dynamic range. The average slope magnitude for each map is shown as well. The dotted red line represents the start of adaptive gating i.e., when the blocking mask was applied to the LCD [15].
We performed a simulation that modeled the response of the A-SHWFS for various aberrations. It was shown that for a combination of random aberrations, the sensor performs as expected. If a high localized aberration is present in some region of the wavefront, the sensor adapts itself to the most efficient sampling scheme, which gives a good trade-off between the ability to correctly measure the wavefront while maintaining the highest possible spatial resolution for the parts of the wavefront that aren’t severely aberrated. Details of this study can be found in appendix D.
Chapter 5

Conclusion

This dissertation has presented a set of novel tools and techniques for surface or wavefront metrology and analysis. The adaptive and dynamic nature of these techniques is necessary for fulfilling challenging optical metrology and analysis requirements that are critical for new and innovative advancements in a variety of optical systems and devices – such as precision freeform optics and high-resolution systems like optical telescopes.

The scope of the work presented can be divided into three main topics. The first topic is the development of a gradient polynomial basis set and an accompanying modal fitting method that is used for surface or wavefront reconstruction from measured slope data. This set, named the $G$ polynomials, is obtained from the gradients of two-dimensional Chebyshev polynomials of the first kind. Although it is orthogonal across rectangular apertures but it does not represent balanced aberrations across such pupils. However, the application of this polynomial set is primarily towards data from non-symmetric systems, such as measurements made from deflectometry systems. Closed-form and recursive solutions were found for $G$ polynomials and the development of an efficient data processing methodology using these polynomials means that tens of thousands of $G$ polynomial terms can be used in the surface / wavefront reconstruction process. This quality is necessary for modal fitting methodologies when preserving the mid-to-
high spatial frequencies of the surface is important (such as for astronomical telescopes), for precision freeform optics (the use of which includes systems for photography, Augmented / Virtual reality, medical devices etc.) or to solve practical metrology problems such poor surface reconstruction when measurements are made with blockers on the surface (e.g. fiducials placed on optics during distortion correction, telescope aperture blockers aka ‘spiders’ or a scratch on the surface of the optic). Analysis from simulated and real deflectometry data shows the $G$ polynomial method to have high accuracy of reconstruction and higher robustness to noise (compared to the traditional zonal method). It allows for a good representation of mid-to-high spatial frequencies (comparable to what is achieved by the standard Southwell zonal method) as well as higher accuracy of reconstruction in the presence of blockers (compared to both the zonal method and another modal method). The $G$ polynomial methodology also shows a better performance than the Southwell zonal method when data is sampled unevenly.

The second part of this work is the development of a curl polynomial set (called $C$ polynomials) and a vector data fitting methodology that uses either the $C$ polynomials or a combination of $G$ and $C$ polynomials to detect and measure errors in measurement systems, optimized for systems with rectangular apertures. $C$ polynomials are orthogonal across rectangular apertures and are obtained from the curl of the same scalar basis set used for generating $G$ polynomials i.e., two-dimensional Chebyshev polynomials of the first kind. A simulation example shows the technique of using the combined $G$ and $C$ polynomials set (where the common terms, which have
Laplacian $= 0$ are only counted once), for mapping imaging distortion. It also compared the results of this mapping with that from another gradient and curl polynomial basis set combination – that is based on the Zernike polynomials and shows that the $G$ and $C$ polynomial fit outperforms the other method by having a lower fitting residual error. Another application showing the usefulness of the $C$ polynomials was demonstrated by measuring a highly freeform Alvarez lens with a deflectometry system. Measurements from the system were intentionally mis-oriented to model a system misalignment by creating non-orthogonality (or mis-clocking error) between the gradients in the $x$ and $y$ directions. Values from the $C$ polynomial based model matched the actual mis-clocking values closely.

The last tool from the adaptive and dynamic metrology and analysis methods introduced in this work is a modified Shack-Hartmann wavefront sensor with an extended dynamic range of measurement. This sensor, called the Adaptive Shack–Hartmann wavefront sensor, abbreviated as A-SHWFS, has the ability to dynamically adapt to the wavefront under test and modify its detection subapertures in an appropriate way. The sensor uses an LCD screen placed in front of the lenslet array, whose pixels are matched with the lenslets, to control the amount of light passing through each lenslet. By stopping or allowing light through pixels of the LCD, the corresponding lenslets can be blocked or unblocked respectively. This gating or masking is electronically controlled. Any generic SHWFS needs to make a trade-off between the dynamic range of measurement and the spatial resolution of the detected wavefront. The A-SHWFS is
designed to provide the best possible compensation between the two. The image processing and control software for the A-SHWFS was designed such that if a wavefront with a highly localized (relative to the full wavefront) aberration is tested, only the region of the sensor measuring the localized high aberration would have its detection subapertures reconfigured and create a low spatial sampling for that part of the measured wavefront while the rest of the detection area would maintain smaller detection subapertures and consequently the wavefront measured in this region would maintain a high sampling. Measured slope data from the A-SHWFS is reconstructed using the $G$ polynomial based modal fitting.

A prototype of this sensor was developed and several tests were performed on it to verify its performance and capabilities. The prototype A-SHWFS uses a commercial LCD screen (with its backlight removed) and a customized in-house manufactured lenslet array. Performance verification test of the A-SHWFS revealed a close match to a commercial precision autocollimator, with excellent resolution and accuracy. The adaptive sampling and regional blocking competence of the A-SHWFS was demonstrated through an experiment that used three hexagonal mirrors (two stationary and one moving) to introduce a local tilt to a section of the beam, while the rest of the beam remained unaberrated. As expected, the sensor maintains its native detector subaperture sizes for a small amount of local tilt but when the tilt exceeds the dynamic range of measurement of the lenslets, the sensor applies an electronic mask to the LCD, which selectively blocks lenslets so that the detection subapertures area is increased in the region
corresponding to this high tilt while the rest of the detection subapertures maintain their native size. A simulation study extended this result further to include other aberrations and show a good response of the sensor to those (random) aberrations.

We have introduced three important and novel tools / methodologies for improving optical metrology and analysis. The underlying basics, operational principles, applications and experimental verification have been discussed. We believe this work will help further the development of many cutting-edge technologies, across a wide range of disciplines and applications.
REFERENCES


Appendix A. Modal Data Processing for High Resolution Deflectometry

Maham Aftab, James H. Burge, Greg A. Smith, Logan Graves, Chang-jin Oh, and Dae Wook Kim


Permission to use this article in the dissertation was obtained from Springer Nature on March 03, 2019.
Modal Data Processing for High Resolution Deflectometry

Maham Aftab1 · James H. Burge1 · Greg A. Smith1 · Logan Graves1 · Chang-jin Oh1 · Dae Wook Kim1,2

Received: 31 August 2017 / Revised: 27 March 2018 / Accepted: 4 April 2018
© Korean Society for Precision Engineering 2019

Abstract
In this paper, we present a modal data processing methodology, for reconstructing high resolution surfaces from measured slope data, over rectangular apertures. One of the primary goals is the ability to effectively reconstruct deflectometry measurement data for high resolution and freeform surfaces, such as telescope mirrors. We start by developing a gradient polynomial basis set which can quickly generate a very high number of polynomial terms. This vector basis set, called the $G$ polynomials set, is based on gradients of the Chebyshev polynomials of the first kind. The proposed polynomials represent vector fields that are defined as the gradients of scalar functions. This method yields reconstructions that fit the measured data more closely than those obtained using conventional methods, especially in the presence of defects in the mirror surface and physical blockers/markers such as fiducials used during deflectometry measurements. We demonstrate the strengths of our method using simulations and real metrology data from the Daniel K. Inouye Solar Telescope (DKIST) primary mirror.

Keywords Surface measurements, numerical approximation and analysis · Instrumentation, measurement, and metrology · Information processing · Deflectometry · Testing

1 Introduction

The Daniel K. Inouye Solar Telescope (DKIST) [2] will be to-date, the largest optical solar telescope ever built. It aims to provide vastly enhanced spatial, spectral and image resolution. The telescope’s primary mirror is a 4.2 m aperture off-axis parabola. The surface quality specifications were quite challenging, particularly requirements for the surface figure, irregularity and BRDF. In fact, this may be the smoothest large mirror ever made [19]. One big challenge with fabrication of such large, freeform mirrors is that the metrology method requires high dynamic range coupled with excellent resolution and accuracy.

One such metrology technique that has proven effective for measurement of such freeform surfaces is deflectometry [7, 12, 27, 28, 32]. Deflectometry measures the slope of the test optic surface in a very simple package, requiring only a light source and a camera.

To obtain a surface sag map from the slope data, integration must be performed either with a zonal approach such as Southwell integration [24] or a modal approach such as the one proposed in this paper.

We show how our particular modal reconstruction method, using our gradient polynomials can work well for large, freeform and especially high-resolution surfaces (e.g., DKIST) and also deal with some common metrology requirements for such projects. One of these requirements is a large dynamic range of slopes and local obscurations from small fiducials commonly placed on the surface of the mirror. These fiducials are used as reference markers and can significantly impact the reconstruction data.

Modal reconstruction methods have several advantages over zonal approaches in the sense that modal techniques are less sensitive to measurement noise, and the number of modes considered can be adapted to the problem [21]. Li et al. [13] have proposed a zonal integration method for deflectometry that shows better results than the traditional Southwell zonal integration method, especially for an unequally sampled grid and circular apertures. However, even their improved zonal method does not reach the accuracy achieved by modal fitting. In their experiment [13], the modal reconstruction result is closer to the interferometric result in terms of surface figure and RMS value.
Numerous reports have detailed the advantages of fitting data directly in the slope domain. For instance, Zhao et al. [33] claim measured vector data (slopes) should be fit using vector polynomials. They also assert that it is preferable to fit data using vector functions that are orthogonal over the measurement domain for several reasons, including the fact that when fitting real data, noise propagation increases with the use of non-orthogonal basis functions. Our proposed polynomial set \((G)\) polynomials) are orthogonal vector polynomials and thus have many nice features.

It must be stressed that this methodology was not developed to be a representation of balanced aberrations over a rectangular aperture. The aim of this paper is not to provide a modal solution to optical fabrication problems where aberration analysis, based on the coefficients of the polynomial set, is critical. We use the modal fitting as an alternative to zonal integration, for the reasons discussed throughout this paper. Furthermore, the optical systems under consideration in this work are neither rotationally symmetric, nor anamorphic.

It is not uncommon to see optical systems utilizing rectangular apertures, as is the case for the Gaia telescope system [4, 20] or X-ray mirrors [29]. The high resolution deflectometry could easily be extended to metrology of solar panels and collectors, which have rectangular or non-circular aperture shapes, also. Although the proposed reconstruction is designed for rectangular datasets, the method is general enough to reconstruct other non-symmetric or freeform shapes. Freeform surfaces are gaining widespread attention because of their utility despite their manufacturing and metrology challenges [3, 8, 31].

This paper begins with a description of 2D Chebyshev polynomials, which can be obtained by taking the product of two Chebyshev polynomial sets representing the two Cartesian coordinate directions used to define a rectangle. It is important to define this set fully since it serves as the scalar basis for the proposed gradient \((G)\) polynomials, defined in Sect. 2. The application of these polynomials to surface reconstruction for deflectometry applications is then discussed in detail.

Section 3 describes the data processing methodology used in the modal fitting approach, which employs the \(G\) polynomials (especially for high-resolution data). Section 4 talks about infrared (IR) deflectometry, specifically for the DKIST. Section 5 provides detailed analysis of the reconstructions performed using the \(G\) polynomial approach as well as comparisons with the reconstructions obtained using the conventional Southwell zonal method. Concluding remarks are presented in Sect. 6.

2 Polynomial Basis

2.1 Scalar Polynomials

The Chebyshev polynomials of the first kind [35], in one dimension (1D) are defined by the recursion relation,

\[
T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x)
\]

where \(T_0(x) = 1, T_1(x) = x, \) for \(-1 \leq x \leq 1\),

or via analytical expressions:

\[
T_m(x) = \cos(m \theta), \quad x = \cos \theta. \tag{2}
\]

2-D scalar polynomials \((F)\) are then defined as [15]:

\[
F_j(x, y) = F_{m}^{n}(x, y) = T_m(x)T_n(y), \tag{3}
\]

where \(j\) is the polynomial number and \(m\) and \(n\) are indices representing the two dimensions of the polynomials.

The values of \(m\) and \(n\) are chosen such that for each order, \(m\) starts from the order number and goes all the way down to zero, while \(n\) starts from zero and goes all the way up to the order number. The conversion from single index \((j)\), to double indices \((m, n)\) has been made by using the following method:

\[
\text{No. of terms } (t) = \text{order} + 1, \quad m = a-1, \quad n = t-a.
\]

A visual representation and the mathematical expressions of the first 15 \(F\) polynomials are given in Fig. 1 and Table 1, respectively. The 2D polynomial set is defined in

![Fig. 1 Plots of 2D Chebyshev polynomials](image)
among the most valuable properties of Chebyshev polynomials are their derivatives being orthogonal to Chebyshev polynomials of the second kind [22], via the relation:

\[ U_{m-1}(x) = \frac{1}{m} T_m'(x) = \frac{\sin(m \theta)}{\sin(\theta)}, \quad x = \cos \theta. \] (5)

As for Chebyshev polynomials of the first kind, discrete orthogonality relations can be defined for the weighted Chebyshev polynomials of the second kind (where the weighting function is \( \sqrt{1-x^2} \)) [18].

By taking the derivatives of the \( F \) polynomials, the \( G \) polynomials are obtained as follows:

\[ \tilde{G}_j(x, y) = \tilde{G}^m_j(x, y) = \nabla F^m_j(x, y) \]
\[ = \frac{\partial}{\partial x} F^m_j(x, y) \hat{x} + \frac{\partial}{\partial y} F^m_j(x, y) \hat{y}. \] (6)

\[ \tilde{G}_j(x, y) = T_n(y) T_m'(x) \hat{y} + T_m(x) T'_n(y) \hat{x}. \] (7)

The \( G \) polynomials can be written in terms of the recursive \( T \) polynomials as follows,

\[ \tilde{G}_j(x, y) \]
\[ = \frac{1}{2x(1-x^2)} \left[ (T_{m+1}(y) - T_{-m-1}(y)) (T_{m-1}(x) - xT_m(x)) \right] \hat{i} \]
\[ + \frac{n}{2x(1-y^2)} \left[ (T_{m+1}(x) - T_{-m-1}(x)) (T_{m-1}(y) - yT_m(y)) \right] \hat{j}. \] (8)

Table 2 lists the first 15 \( G \) polynomials and quiver plots of the first few non-trivial \( G \) polynomials are provided in “Appendix 1”. Orthonormality of the \( G \) polynomials is provided in “Appendix 2”. Each gradient term is a combination of two terms, each containing one scalar Chebyshev polynomial (\( T_m \) or \( T_n \)) and one Chebyshev polynomial derivative term (\( T'_m \) or \( T'_n \)). This table was simplified using the following relations:

\[ T_0(x) = 1, \quad T_1(y) = 1; \quad T_1'(x) = x, \quad T_1'(y) = y; \]
\[ T_0'(x) = 0, \quad T_0'(y) = 0; \quad T_1'(x) = 1, \quad T_1'(y) = 1. \] (9)

Both \( T \) and \( U \) polynomials have simple, recursive relationships [35]. Since the \( G \) polynomials can be expressed as a closed form equation involving the \( T \) and \( U \) polynomials, it is very straightforward and computationally efficient to generate them. This is a significant advantage for data processing since it enables large numbers of terms of the polynomials to be evaluated efficiently and reliably even if computing resources are limited. For \( T \) polynomials, the difference between recursive and direct calculation is only

Table 1 2D Chebyshev polynomials

<table>
<thead>
<tr>
<th>( j )</th>
<th>( m )</th>
<th>( n )</th>
<th>( F_j(x, y) )</th>
<th>Explicit form of ( F_j(x, y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( T_0(x)T_0(y) )</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( T_1(x)T_0(y) )</td>
<td>( x )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>( T_0(x)T_1(y) )</td>
<td>( y )</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>( T_2(x)T_0(y) )</td>
<td>( 2x^2 - 1 )</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>( T_1(x)T_1(y) )</td>
<td>( xy )</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>( T_0(x)T_2(y) )</td>
<td>( 2y^2 - 1 )</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>( T_3(x)T_0(y) )</td>
<td>( 4x^3 - 3x )</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>( T_2(x)T_1(y) )</td>
<td>( (2x^2 - 1)y )</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>( T_1(x)T_2(y) )</td>
<td>( (2y^2 - 1)x )</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>3</td>
<td>( T_0(x)T_3(y) )</td>
<td>( 4y^3 - 3y )</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
<td>( T_4(x)T_0(y) )</td>
<td>( 8x^4 - 8x^2 + 1 )</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1</td>
<td>( T_3(x)T_1(y) )</td>
<td>( (4x^3 - 3x)y )</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>( T_2(x)T_2(y) )</td>
<td>( (2x^2 - 1)(2y^2 - 1) )</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>3</td>
<td>( T_1(x)T_3(y) )</td>
<td>( (4y^3 - 3y)x )</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>4</td>
<td>( T_0(x)T_4(y) )</td>
<td>( 8y^4 - 8y^2 + 1 )</td>
</tr>
</tbody>
</table>

both the \( x \) and \( y \) directions on the interval from \(-1\) to \(+1\). These definitions are similar to those in Ref. [15].

Chebyshev polynomials have several interesting properties that make them attractive for data fitting, especially for discrete data. In addition to their orthogonal properties as defined in Refs. [15, 35], there are many cases in which the discrete orthogonality of Chebyshev polynomials can be shown to hold exactly [18]. For example Chebyshev polynomials of the first kind, \( T_m(x) \), \( m = 0, 1, 2, \ldots, N \) are orthogonal over the discrete point set comprising the zeros \( x_{N+1,m} = 1, 2, \ldots, N + 1 \), of \( T_{m+1}(x) \) [1].

Additionally, Chebyshev polynomials can be used as an approximation to the minimax polynomial, for which the maximum value of the error between the function and its polynomial approximation, is a minimum within some specified range [5]. Also, of all expansions in terms of ultra-spherical polynomials, the Chebyshev series generally has the fastest rate of convergence [5].

### 2.2 Gradient Polynomials

Among the most valuable properties of Chebyshev polynomials are the facts that their derivatives are orthogonal and they can be generated recursively. In 1-D, the gradients are given by,

\[
T'_m(x) = \frac{m[T_{m-1}(x) - xT_m(x)]}{(1 - x^2)},
\]
\[T'_n(y) = \frac{n[T_{n-1}(y) - yT_n(y)]}{(1 - y^2)}.\] (4)
Table 2  Chebyshev gradient polynomials

<table>
<thead>
<tr>
<th>j</th>
<th>m</th>
<th>n</th>
<th>Explicit form of $\tilde{G}_j(x, y)$</th>
<th>$\tilde{G}_j(x, y)$ in terms of $F_j(x, y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\tilde{0}i + \tilde{0}j$</td>
<td>$0$</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>$\tilde{i}$</td>
<td>$\tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_1(0)]$</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>$\tilde{j}$</td>
<td>$\tilde{0}1[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>0</td>
<td>$T'_2(x)\tilde{i}$</td>
<td>$2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>$y\tilde{i} + x\tilde{j}$</td>
<td>$\tilde{1}1[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>2</td>
<td>$T'_2(x)\tilde{j}$</td>
<td>$2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0</td>
<td>$T'_3(x)\tilde{i}$</td>
<td>$3 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_3(0)]$</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>$yT'_2(x)\tilde{i} + T'_2(x)\tilde{j}$</td>
<td>$2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>$T'_2(x)\tilde{i} + xT'_2(x)\tilde{j}$</td>
<td>$\tilde{1}1[\tilde{r}_0(0)\cdots \tilde{r}_2(0)] + 2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>3</td>
<td>$T'_3(x)\tilde{j}$</td>
<td>$3 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_3(0)]$</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>0</td>
<td>$T'_4(x)\tilde{j}$</td>
<td>$4 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_4(0)]$</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>1</td>
<td>$yT'_3(x)\tilde{i} + T'_3(x)\tilde{j}$</td>
<td>$3 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_3(0)]$</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>2</td>
<td>$T'_2(x)T'_2(x)\tilde{i} + T'_2(x)T'_2(x)\tilde{j}$</td>
<td>$2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)] + 2 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_2(0)]$</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>3</td>
<td>$T'_3(x)\tilde{j} + xT'_3(x)\tilde{j}$</td>
<td>$\tilde{1}1[\tilde{r}_0(0)\cdots \tilde{r}_3(0)] + 3 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_3(0)]$</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>4</td>
<td>$T'_4(x)\tilde{j}$</td>
<td>$4 \tilde{1}0[\tilde{r}_0(0)\cdots \tilde{r}_4(0)]$</td>
</tr>
</tbody>
</table>

10^{-11}, for a polynomial order of 20,000, generated over a grid of 50,000 points.

3 High-Resolution Gradient Data Processing Using G Polynomials

Modal integration using the $G$ polynomials is discussed in detail by describing the reconstruction process step by step.

3.1 Modal Fit of Gradient Data

Assume the surface (or wavefront) can be represented as,

$$W(x, y) = \sum_{j=1}^{N} a_j F_j,$$

(10)

where $a_j$ are the expansion coefficients, which need to be determined for a given data set.

Similarly, to represent the measured gradient, the following relation can be employed:

$$\nabla W(x, y) = \sum_{k=1}^{N} b_k \tilde{G}_k,$$

(11)

where $b_k$ are the vector expansion coefficients of the $G$ polynomials.

For discrete data, we can represent the modal fitting as a matrix [17]

$$D = \tilde{G} b,$$

(12)

where $D$ is a column vector of $P$ data values, $b$ is a column vector containing the $N$ expansion coefficients, and $G$ is a $P \times N$ matrix representing the values of the $G$ polynomials at the locations of the data points.

The coefficients can then be found using a pseudo-inverse:

$$b = (\tilde{G}' \tilde{G})^{-1} \tilde{G}' D.$$

(13)

These $b$ coefficients and the associated $G$ polynomials represent a best-fit estimate of the surface we wish to reconstruct.
3.2 Obtaining Scalar Coefficients Directly from Gradient Data

Based on the method employed in the circular vector data papers by Zhao and Burge [33, 34], a simple conversion from $G$ polynomial coefficients to $F$ polynomial coefficients (see Eq. 3) is derived.

Denoting sets of vector and scalar data as $V$ and $S$, respectively, the data sets can be expanded in terms of vector and scalar polynomial sets as follows:

$$\vec{V} = \sum b_k \vec{G}_k; \quad S = \sum a_k F_k. \quad (14)$$

$$\vec{V} = \nabla S; \quad \vec{G} = \nabla F. \quad (15)$$

Following the above definitions, the vector data can be expressed as,

$$\vec{V} = \sum b_k \vec{G}_k = \sum \nabla \vec{S} = \sum a_k (\nabla F_k) = \sum a_k \vec{G}_k. \quad (16)$$

Then, the coefficients of the scalar and vector polynomials can be related as,

$$a_k = b_k. \quad (17)$$

Thus, there is a one-to-one correspondence between the coefficient sets, which makes the $G$ polynomials unique and powerful in terms of reconstructing surface/wavefront maps. Once the slope data is fit to $G$ polynomials, the surface uses the same coefficients applied to the $F$ polynomials defined in Eq. 3.

4 Deflectometry for DKIST

A deflectometry system is a metrology tool which directly measures surface slope of a test optic. Surface reconstruction therefore requires an integration routine such as the method discussed above. As we show in Sect. 5, $G$ polynomials are a particularly useful method for reconstructing surface shape from a deflectometry measurement.

Two in-house deflectometry systems have been developed and utilized for situations requiring high spatial resolution and large dynamic range freeform surface metrology [28, 30]. These systems are essentially computerized reverse Hartmann tests operating in the visible and long infrared wavelengths, and have been used to analyze high dynamic range tests of freeform surfaces [26]. A schematic of the infrared measurement scheme is shown in Fig. 2. A rectangular tungsten ribbon is heated to create a long wave infrared source, and is scanned in the $x$ and $y$ orthogonal directions. The camera focused on the mirror surface, records the reflections from the test mirror surface for both scans, and by knowing the system geometry, the local surface slopes in the $x$ and $y$ directions are calculated.

In these systems, the measured slopes are integrated to obtain the sag map of the surface under test. While historically the slopes were integrated using a zonal method [24], a polynomial fitting method has also been attempted. The older polynomial approach proved impractical at providing high spatial resolution due to the inability to compute sufficient number of terms.

4.1 DKIST Primary Mirror Metrology

The Daniel K. Inouye Solar Telescope (DKIST) is designed to provide greatly improved optical performance in imaging, and spatial and spectral resolution when measuring dynamic solar phenomenon.

The SLOTS system was used to guide fabrication of the DKIST primary mirror during the grinding phase. Over this period a 25 µm down to 12 µm loose grit abrasive was used to grind the mirror, and the root-mean-square shape error was reduced from 15 µm to less than 1 µm [10]. The DKIST primary presented a unique freeform shape to measure, being a 4.2 m off-axis parabolic mirror constructed from Zerodur®. The full mirror specifications are provided in Table 3.

The SLOTS system used for testing consisted of a 4 mm wide tungsten wire which, when current was applied, would emit as a black body radiator. This wire served as the source for the deflectometry system. A micro-bolometer camera was utilized with an aperture diameter of 14 mm, and a peak

![Fig. 2 Schematic diagram of the SLOTS deflectometry concept. LWIR stands for long-wave infrared [11]](image)
absorption band of 7–11 µm. The camera and source were placed near the center of curvature of the DKIST primary mirror, at 17.1 m from the surface in a test tower. The wire was scanned in orthogonal x and y directions, and as long-wave IR light was emitted from the wire it would reflect from the mirror surface and be recorded by the camera system. Figure 3 shows the camera used with the SLOTS system. The schematic of the entire system was shown in Fig. 2.

In order to obtain slope data the local slopes in both x and y were calculated by ray tracing from the wire, to the mirror surface, to a camera detector pixel. This process was performed for all camera pixels, which were mapped to the mirror surface and served as discrete ‘pixel’ areas on the mirror over which the local slopes in x and y were calculated. To achieve this, the Cartesian coordinates of the wire, camera, and mirror had to be known, as well as the camera mapping function. The Cartesian coordinates of all components were measured by using a laser tracker [6], which provided accuracy in position to within 0.1 µm. Additionally, the wire position during scans in both the x and y directions was recorded, again with the laser tracker. To determine the camera mapping function, reflective fiducials were placed on the mirror and the centroid position of the fiducials was calculated with the SLOTS system. This was a key step in the calibration process, and also was one of the most time consuming steps as the fiducials had to subsequently be removed, due to the integration method used at the time for surface reconstruction, which could not account for missing data regions that the fiducials represented for the mirror surface. The calculated coordinates were compared with a laser tracker measurement of the true centroids to determine the camera distortion. With this information, along with the full system parameters, listed in Table 4, a theoretical certainty of 8.5 µrad could be achieved for the calculated local slopes [25].

Fabrication of the DKIST primary mirror was performed by the College of Optical Sciences at the University of Arizona. During the fine grinding stage deflectometry systems provided sub-micron accuracy for the entire non-specular, ground, freeform surface metrology [9, 11]. Multiple measurements were performed over the course of the mirror fabrication with this SLOTS system, and the raw data was all preserved. This served as an ideal source for real world slope data, which previously had been reconstructed into surface sag via a traditional Southwell integration [24] method.

To evaluate the novel modal reconstruction method, this raw slope data was reconstructed using the new polynomials based method, and is presented along with the original surface reconstruction results. The data fitting methodology follows the procedure outlined in Sect. 3. Gradient data obtained from deflectometry is fit with the desired number of polynomials, using Eqs. 12 and 13. Vector polynomial coefficients (referred to as ‘b’ in Sect. 3) thus obtained are converted to coefficients of the scalar polynomials (called ‘a’ in Sect. 3). In the case of our basis sets, the scalar and vector coefficients have a one-to-one correspondence (as seen from Sect. 3.2). Lastly, the ‘a’ coefficients and scalar basis set (F) are used to obtain the measured surface, using Eq. 10.

### Table 4 Parameters in the DKIST SLOTS test

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power total consumed</td>
<td>44</td>
<td>W</td>
</tr>
<tr>
<td>Source surface area</td>
<td>6280</td>
<td>mm²</td>
</tr>
<tr>
<td>Source width</td>
<td>4</td>
<td>mm</td>
</tr>
<tr>
<td>Camera aperture diameter</td>
<td>14</td>
<td>mm</td>
</tr>
<tr>
<td>Camera focal length</td>
<td>35</td>
<td>mm</td>
</tr>
<tr>
<td>Detector pixel area</td>
<td>6.25×10⁻⁴</td>
<td>mm²</td>
</tr>
</tbody>
</table>

The surface is ground with 12 um grits (700 nm RMS roughness).

5 Performance Analysis

To evaluate the proposed modal method, we analyze its performance using several simulated and actual data sets. We discuss the results in terms of several figures of merit, such as the fit accuracy, performance in the presence of noise, processing time and the effectiveness of reconstruction obtained with local obscuration cases.

It should be noted that all simulated data had gradients generated by MATLAB’s ‘gradient’ function. This function uses correlations between pixels to determine the local slope, and therefore exact reconstruction is not possible.
It is possible unless the slope integration assumes the same correlations. Neither the Southwell nor $G$ polynomial method assumes these specific correlations and therefore we do not expect perfect reconstruction even though the data is simulated.

### 5.1 Fit Accuracy

We describe the modal method’s reconstruction accuracy and compare its results to ideal or Southwell reconstructed data.

#### 5.1.1 Fitting Accuracy Evaluation Using Synthetic (Ideal) Data

We start by demonstrating how a simulated data set can be reconstructed accurately using $G$ polynomials. A data set described by the following equation was generated in MATLAB in a $250 \times 250$ point grid:

$$z(x, y) = 3.57x + 6.53x^2 + 4.29y^2 - 4.75y^3 + 2.99(x^3 + xy^2) - 2.95(2x^4 + y^4 + 2x^2y^2) + 1.21(x^5 + xy^4) + 0.89y^5.$$  

(18)

The gradient of the data set was then reconstructed in two ways: modally using $G$ polynomials (with the first 100 polynomials) and zonally using Southwell integration. The reconstruction results are compared with the ideal (MATLAB-simulated) surface in Fig. 4. Both results are nearly identical in this ideal case. The modal fitting residual error is $4.9799 \times 10^{-4}$ $\mu$m (0.0108% error relative to the surface RMS of the ideal map) versus $5.1335 \times 10^{-4}$ $\mu$m (0.0112% error relative to the RMS of the ideal map) for Southwell integration.

For this reconstruction, Southwell integration was a little faster, taking $0.389$ s while $G$ polynomial processing took $0.830$ s.

#### 5.1.2 Fitting Accuracy Evaluation Using Actual (Measured) Data

To test practical applicability, we applied $G$ polynomials to actual deflectometry data ($2.4 \times 2.4$ m sub-area) [11, 27] which includes noise, uncertainties, and other measurement errors. The analyzed data had an area of $201 \times 201$ pixels. The surface reconstruction was performed using the $G$ polynomials as well as the standard Southwell integration process. The results of the two methods are compared in Fig. 5.

Even when only the first 37 terms of the $G$ polynomials are used, an RMS difference between the two methods of only $0.0183$ $\mu$m is obtained. This corresponds to about $0.2327\%$ of the RMS of the reconstructed surface or $0.0478\%$ of the $38.3154$ $\mu$m surface. It took $0.153$ s for the $G$ polynomial reconstruction versus $0.238$ s with Southwell reconstruction, making the modal method a bit faster in this case.

Predictably, increasing the number of $G$ polynomials terms, increases the reconstruction performance, as shown in Table 5, which lists the RMS values of the difference maps between the modal and Southwell surfaces. The Southwell surface was used as the reference surface.
5.1.3 Fidelity Check of Mid-to-high Spatial Frequency Reconstruction

To examine mid-to-high spatial frequencies, an arbitrary 1.2 × 1.2 m sub-section of the measured DKIST high-resolution data set was used. A high-pass Gaussian filter was applied to both the modally and zonally reconstructed surfaces, using SAGUARO data processing software [23], in order to focus on the high spatial resolution features in this case study. For the modal fit, a series of data sets were reconstructed, each of which was generated using a different number of $G$ polynomials. These different high-pass-filtered $G$-polynomial surfaces are compared with the reference high-pass-filtered Southwell surface in Fig. 6.

From left to right in Fig. 6b–e (i.e., as the number of modal terms increases), the maps become less blurry and more high-frequency features become sharp and can be resolved. The graphs towards the right closely resemble the Southwell surface in Fig. 6a.

As expected, the representations of high frequencies improved as more terms were used for fitting. This finding emphasizes one of the critical advantages of the $G$ polynomial set over others, namely, the ability to generate many thousands of terms. The simple recursion relations ensure that it is easy and practical to generate all of these high-order terms.

In principle, this data set could have been reconstructed with other basis sets, such as a Zernike-basis vector set, orthogonalized over a rectangular pupil [14]. However, that does not have a straightforward closed-form relationship which makes it difficult to generate such extreme orders. For high-resolution and freeform optics metrology and data processing applications, it is essential to use a sufficient numbers of polynomials. Additionally, there are several potential problems in converting Zernike polynomials to their gradients, orthogonalizing the Zernike gradients over a rectangular aperture, and generating the conversion matrices for obtaining the scalar coefficients from the vector fitting. This process is cumbersome, and has the possibility to generate numerical errors, due to the several steps of computation, approximations and numerical truncations. As mentioned earlier, the metrology systems under consideration here are neither rotationally symmetric nor anamorphic, so the balanced aberration representations of Zernike-type [14] or Legendre polynomials [16] are not necessarily valid.

5.2 Robustness Against Noise

This subsection describes the reconstruction performance in the presence of noise. Varying amounts of Gaussian white noise were added to the simulated ideal data set defined by Eq. (18). This “noisy” data was reconstructed using both the Southwell and modal (using the first 100 polynomials) approaches. In Fig. 7, the error in each reconstructed data set (difference between the ideal and reconstructed maps) as a percentage of the RMS of the ideal (noiseless) surface is plotted against the standard deviation of the added noise.

The polynomial fit error is smaller than the Southwell integration error for all noise levels. Furthermore, the performance of the Southwell method decreases faster as the amount of noise increases. For example, when the standard deviation of the noise is 0.2 µm, the polynomial fit exhibits an error of about 0.1% (relative to the RMS of the ideal surface), while the Southwell fit exhibits an error of about 0.5% (relative to the RMS of the ideal surface). These errors increase to 0.4% for the polynomial fit and 2.5% for the Southwell fit when the standard deviation of the noise is 0.5 µm.

Fig. 7 Residual errors from surface reconstruction in the presence of various amounts of noise

Fig. 6 a Reference high-pass-filtered Southwell map and high-pass-filtered modal reconstructed maps generated using b 37, c 750, d 3000, and e 20,000 polynomial terms
When gradient data is processed using Southwell zonal integration, issues such as mirror/lens surface defects, distortion correction fiducials, non-uniform data point distribution due to imaging distortion, and the presence of multiple local clear apertures can cause local obscurations in the data. Any areas with missing data that are encountered during the integration process affect their surrounding regions. Although some numerical methods can be employed to reduce this effect (like replacing the missing data with zeros), the nature of the zonal approach fundamentally limits the reconstruction accuracy when a sub-region is not well-defined. In contrast, modal reconstructions are not significantly impacted by obscurations since they are based on overall views of the data.

For this analysis, a 250×250 pixel grid of synthetic data was simulated, and Fig. 8 shows the reconstruction results obtained when a small defect (a 90×2 pixel line of NaN (non-number) values, corresponding to an area of 90×2 mm) was added to this simulated data. After reconstructing the data set using both the Southwell and modal approaches (with 2000 polynomial terms), the differences between the results and the ideal scalar data are compared. Large differences are observed near the missing data.

The simulated ideal surface has an RMS of 95.1220 µm. The Southwell residual error map (difference between Southwell reconstructed and simulated surfaces) has an RMS of 0.8387 µm (or 0.8817% relative to the RMS of the simulated surface), whereas the modal error map has an RMS of 0.0627 µm (0.0659% error relative to the simulated surface). Thus, the residual error of the Southwell method is about 13 times higher than that of the modal method in this case.

Table 6 lists the RMS errors of the difference maps for the modal fits (i.e., the differences of the modal fits from the ideal surface) that were obtained using different number of polynomial terms. As expected, the RMS error decreases as the number of terms used in the expansion increases.

If an optic has multiple obscurations, Southwell zonal integration becomes more limited. Fiducials are one example, and are typically created by placing physical markers on the surface of the optic, essentially blocking parts of it. Ref. [19] provides an example of using fiducials for system calibration as a step in deflectometry. An actual picture of the DKIST mirror with fiducials placed on it during metrology is shown in Fig. 8d [2]. In this analysis, we added a few rectangular masks modelling fiducial areas to the same data that were used to generate Fig. 7. Each mask had dimensions of 24×30 pixels, corresponding to

![Figure 8](image1)

**Fig. 8** a Simulated ideal data with a 90×2 mm obscuration. b Southwell and c modal residual error maps (differences from the simulated map)

<table>
<thead>
<tr>
<th>Number of $G$ polynomials</th>
<th>100</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS error (µm)</td>
<td>1.4374</td>
<td>0.5032</td>
<td>0.0739</td>
<td>0.0627</td>
<td>0.0568</td>
</tr>
<tr>
<td>[RMS of difference map/RMS of ideal map]×100 (%)</td>
<td>1.5111</td>
<td>0.5290</td>
<td>0.0777</td>
<td>0.0659</td>
<td>0.0597</td>
</tr>
</tbody>
</table>

5.3 Reconstruction with Local Obscurations

When gradient data is processed using Southwell zonal integration, issues such as mirror/lens surface defects, distortion correction fiducials, non-uniform data point distribution due to imaging distortion, and the presence of multiple local clear apertures can cause local obscurations in the data. Any areas with missing data that are encountered during the integration process affect their surrounding regions. Although some numerical methods can be employed to reduce this effect (like replacing the missing data with zeros), the nature of the zonal approach fundamentally limits the reconstruction accuracy when a sub-region is not well-defined. In contrast, modal reconstructions are not significantly impacted by obscurations since they are based on overall views of the data.

For this analysis, a 250×250 pixel grid of synthetic data was simulated, and Fig. 8 shows the reconstruction results obtained when a small defect (a 90×2 pixel line of NaN (non-number) values, corresponding to an area of 90×2 mm) was added to this simulated data. After

![Figure 9](image2)

**Fig. 9** a Simulated ideal data with 24×30 mm fiducial masks. b Southwell and c modal residual error maps (differences from the simulated map). d Picture of fiducials on the DKIST mirror during a metrology run
The simulated data and residual errors of the Southwell and modal (with 2000 polynomial terms) reconstructions with these fiducials are presented in Fig. 9a–c.

The RMS of the ideal surface in Fig. 8 is 93.8532 µm. The residual RMS error for the Southwell map is 5.9627 µm (or 6.3532% relative to the RMS of the ideal surface), whereas that for the modal fit is 0.7684 µm (or 0.8187% error). In this case, the Southwell error is about eight times higher than the modal error.

### 5.4 Comparison of Reconstruction Accuracy and Processing Time

This section aims to compare reconstruction accuracy and processing time between three methods (a) Zernike gradient polynomial based modal method, from Refs. [33] and [34], (b) Southwell zonal method, and (c) G polynomial based modal method.

The simulated data is a freeform shaped wavefront, generated over a 200 x 200 point grid, representing a rectangular pupil with a mask pattern. The wavefront is of the form:

\[
W(x, y) = 0.14x^2 - 0.33y^5 + 1.89(3x^2y - y^3) + 0.57(x^6 - 10x^4y^2 + 13x^2y^4 - y^6).
\]  

Using 30 polynomials each for both modal methods, the RMS error between the ideal wavefront and its reconstructions is given in Table 7. The G polynomial based reconstruction performs better than the other two methods. Figure 10 shows the wavefront and its reconstructions from the three aforementioned methods.

It is also worthwhile to consider the time it takes for the three methods to process data. Figure 11 shows the reconstruction time as the number of polynomials used.

<table>
<thead>
<tr>
<th>Wavefront</th>
<th>Zernike gradient (modal)</th>
<th>Southwell (zonal)</th>
<th>G polynomial (modal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMS of difference map (µm)</td>
<td>0.0745</td>
<td>0.0775</td>
<td>0.0379</td>
</tr>
<tr>
<td>[RMS of difference map/RMS of ideal map] x 100 (%)</td>
<td>10.9625</td>
<td>11.4082</td>
<td>2.5698</td>
</tr>
</tbody>
</table>

Fig. 10 a Simulated ideal data. b Zernike gradient, c Southwell and d G Polynomial based modal residual error maps (differences from the simulated map)

Fig. 11 Plot of change in reconstruction time as the number of polynomials used for data fitting is increased, for both modal methods. We acknowledge that the actual processing time could depend on the specific numerical implementation. This data is provided only for a baseline comparison.
for the modal methods is increased. Data is shown only up to the 180th polynomial term because for higher polynomial terms, the Zernike gradient polynomial based modal fit becomes numerically unstable.

6 Conclusion

A complete and computationally efficient vector polynomial set, the $G$ polynomial set, was derived. Chebyshev polynomials, which form the basis for the $G$ polynomials, have several properties which make them an attractive choice to base our modal fitting method on.

One major advantage of the $G$ polynomial approach is that both the scalar and vector polynomials are easy to define and manipulate and have a very simple relationship that ensures that the coefficients of the polynomials in both sets are the same. Hence, conversion between the two sets is straightforward. In addition, the $G$ polynomials can be easily computed, so even very high polynomial orders can be generated with high numerical efficiency. Since hundreds of thousands of polynomials can be generated, mid-to-high spatial frequencies of surfaces can be reconstructed from high-resolution-metrology data, such as was demonstrated for the Daniel K. Inouye Solar Telescope (DKIST) mirror. Other advantages of this modal approach include better accuracy of reconstruction in the presence of scratches and fiducials.

Acknowledgements This material is partly based on work performed for the DKIST. DKIST is managed by the National Solar Observatory, which is operated by the Association of Universities for Research in Astronomy Inc. under a cooperative agreement with the National Science Foundation. Also, it is based in part upon work performed for the “Post-processing of Freeform Optics” project supported by the Korea Basic Science Institute. The deflectometry related software development is partially funded by the II–VI Foundation Block grant.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Appendix

Appendix 1

See Fig. 12.
Appendix 2

Orthonormality of the vector gradient polynomials

G polynomial orthogonality can be observed by noting that the dot product of two G polynomials is orthogonal:

\[
\int_{-1}^{1} \int_{-1}^{1} \left[ \tilde{G}_i(x,y) \cdot \tilde{G}_j(x,y) \right] \sqrt{1-x^2} \sqrt{1-y^2} \, dx \, dy = \begin{cases} 0, & i \neq j \\ N_{G,i} = 1, & i = j \end{cases}
\]  

(A.1)

Next, we demonstrate the G polynomials’ orthonormality using the expanded form of the polynomials (from Eq. (7)):

\[
\frac{1}{N_{G}} \int_{-1}^{1} \int_{-1}^{1} \left[ \{ T_{n}(y)T_{m}'(x) \hat{i} + T_{m}(x)T_{n}'(y) \hat{j} \} \cdot \left[ \{ T_{n'}(y)T_{m'}'(x) \hat{i} + T_{m'}(x)T_{n'}'(y) \hat{j} \} \right] \right] \, dx \, dy = \delta_{mm'}\delta_{nn'}
\]  

(A.2)

where \( m, m', n, \) and \( n' \) are integers used for indexing and \( N_{G} \) is the normalization factor, which is given by,
\[ N_G = \begin{cases} \frac{\pi^2 n^2}{4}, & m = 0 \\ \frac{\pi^2 m^2}{4}, & n = 0 \\ \frac{\pi^2}{8} (m^2 + n^2), & \text{otherwise} \end{cases} \tag{A.3} \]

The proof of this derivation and details on how to obtain the normalization factor are as follows:

\[ I = \int_{-1}^{1} (G_x \cdot G_y) \sqrt{1 - x^2} \sqrt{1 - y^2} dx dy \]
\[ = \int_{-1}^{1} T_n(y) T_n'(y) T_m(x) T_m'(x) \sqrt{1 - x^2} \sqrt{1 - y^2} dx dy \]
\[ + \int_{-1}^{1} T_m(x) T_m'(x) T_n(y) T_n'(y) \sqrt{1 - x^2} \sqrt{1 - y^2} dx dy. \tag{A.4} \]

The first integral \( I_1 \) can be expressed as:

\[ I_1 = \int_{-1}^{1} T_n(y) T_n'(y) T_m(x) T_m'(x) \sqrt{1 - x^2} \sqrt{1 - y^2} dx dy \]
\[ = \int_{-1}^{1} T_n(y) T_n'(y) \sqrt{1 - y^2} \int_{-1}^{1} T_m(x) T_m'(x) \sqrt{1 - x^2} dy dx \]
\[ T_m(x) = \cos (m \cos^{-1}(x)). \tag{A.5} \]

Then, the following substitutions and subsequent equations can be used:

\[ x = \cos (t), \quad dx = -\sin (t) dt, \quad \sin (t) = \sqrt{1 - x^2}, \]

\[ I_1 = \int_{-1}^{1} T_n(y) T_n'(y) \sqrt{1 - y^2} \]
\[ \left[ m \sin (m \cos^{-1}(x)) \right]_{\pi}^{0} \sin (t) \sin (m \cos^{-1}(x)) \sin (t) \left[ -\sin (t) \right] dy, \]

\[ I_1 = \int_{-1}^{1} \left( T_n(y) T_n'(y) \sqrt{1 - y^2} \right) \frac{m \cos (m \sin (m \sin(t)))}{m^2 - m'^2} \]
\[ \tag{A.6} \]

This integral goes to zero for all values of \( m \) and \( m' \) (except for when \( m = m' \)), since \( \sin(n \pi) = 0 \) for all integer values of \( m \). However, when \( m = m' \),

\[ I_1 = \int_{-1}^{1} T_n(y) T_n'(y) \sqrt{1 - y^2} \int_{\pi}^{0} \sin (mt) \sin (mt) dt dy. \tag{A.7} \]

Using the identity,

\[ \sin^2 (x) = \frac{1 - \cos (2x)}{2}, \]

\[ I_1 = \int_{-1}^{1} T_n(y) T_n'(y) \sqrt{1 - y^2} dy \left( \frac{m^2}{2} \frac{\sin (2 \pi m)}{2m - \pi} \right) \delta_{mm'}. \tag{A.8} \]

As before, \( \sin(2 \pi m) = 0 \) for all integer values of \( m \). Thus,

\[ I_1 = \frac{\pi \delta_{mm'}}{2} \int_{-1}^{1} \cos (\pi m) \cos (\pi m') \sqrt{1 - y^2} dy. \tag{A.9} \]

Using the same substitutions as before,

\[ y = \cos (t), \quad dy = -\sin (t) dt, \quad \sin (t) = \sqrt{1 - y^2}, \]

\[ I_1 = \frac{\pi \delta_{mm'}}{2} \int_{-1}^{1} \cos (\pi n) \cos (\pi n') \left( \frac{1 - \cos (2t)}{2} \right) dt. \tag{A.10} \]

Again using the trigonometric identity for the square of the sine function,

\[ I_1 = -\frac{\pi \delta_{mm'}}{2} \int_{-1}^{1} \cos (\pi n) \cos (\pi n') \left( \frac{1 - \cos (2t)}{2} \right) dt \]
\[ = -\frac{\pi \delta_{mm'}}{2} \int_{-1}^{1} \left[ \sin \left( \frac{\pi (n - n' - 2)}{n - n' + 2} \right) \right]_{n - n' - 2}^{n - n' + 2} \frac{n - n' + 2}{n - n' - 2} + \left[ \sin \left( \frac{\pi (n + n' - 2)}{n + n' + 2} \right) \right]_{n + n' - 2}^{n + n' + 2} \frac{n + n' + 2}{n + n' - 2} \]
\[ - \left[ \frac{n' \cos (\pi n) \sin (\pi n') - n \sin (\pi n) \cos (\pi n')}{n^2 - n'^2} \right] \]
\[ \tag{A.11} \]
This integral equals zero for all values of \( n \) and \( n' \), except when \( n = n' \), in which case the integral is,

\[
I_1 = -\frac{\pi m^2}{2} \delta_{mm'} \int_0^\infty \cos^2 (nt) \left( \frac{1 - \cos (2t)}{2} \right) dt
\]

\[
= \frac{\pi m^2}{4} \delta_{mm'} \left\{ \frac{2 \pi n + \sin (2\pi n)}{4n} \right\} + \left\{ \frac{n \sin (2\pi n)}{4(n^2 - 1)} \right\} \delta_{nn'}.
\]  
(A.12)

All of the sine terms become zero since \( n \) is an integer, so,

\[
I_1 = \frac{\pi m^2}{8} \delta_{mm'} \delta_{nn'}.
\]  
(A.13)

Similarly for \( I_2 \),

\[
I_2 = \frac{\pi n^2}{8} \delta_{mm'} \delta_{nn'}.
\]  
(A.14)

Now the original integral becomes:

\[
I = \frac{\pi m^2}{8} \delta_{mm'} \delta_{nn'} + \frac{\pi n^2}{8} \delta_{mm'} \delta_{nn'} = \frac{\pi^2}{8} \delta_{mm'} \delta_{nn'} (m^2 + n^2).
\]  
(A.15)

Next we address the cases in which either \( m = 0 \) or \( n = 0 \). When \( m = 0 \), \( I_1 = 0 \) since,

\[
T_{m'}(x) = T_m(x) = 0,
\]

so

\[
I = I_2 = \int_{-1}^{1} T_n'(y)T_m'(y) \sqrt{1 - x^2} \sqrt{1 - y^2} dxdy,
\]  
(A.16)

where we used the fact that

\[
T_n'(x) = T_m(x) = 1.
\]  
(A.17)

Following the same steps as before,

\[
I = \frac{\pi^2 n^2}{4} \delta_{mm'} \delta_{nn'}.
\]  
(A.18)

Similarly, for \( n = 0 \),

\[
I = \frac{\pi^2 m^2}{4} \delta_{mm'} \delta_{nn'}.
\]  
(A.19)

where \( m \) and \( n \) are in the original basis in the last two equations.

It can easily be seen that \( I = 0 \) when \( m = 0 \) and \( n = 0 \).

References


Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Dr. Chang-jin Oh is the PI of Optical Engineering and Fabrication Facility in College of Optical Sciences and he has provided leadership role to the group for the last 15 years. His expertise includes large scale optomechanical systems for space and ground based application, optical fabrication and precision metrology. He has been PI and Co-PI of various many challenging large optics projects for astronomy and

James H. Burge is Professor Emeritus of Optical Sciences at the University of Arizona.

Logan Graves is a PhD candidate in the College of Optical Sciences at the University of Arizona. His main research area covers precision optical metrology, focusing on deflectometry hardware and software development for improved freeform measurement capabilities. He also has a background in biomedical engineering, with research topics including early cancer detection utilizing autofluorescence and non-linear rod-cone interactions at the retinal level.

Greg A. Smith conducts research in large optics manufacturing and testing at the College of Optical Sciences, University of Arizona. He received his optics Ph.D. in 2006 and specializes in the areas of polarization optics and software algorithms for optical metrology.

Maham Aftab is a Senior Design Engineer at ASML, US. She is also a doctoral candidate in the College of Optical Sciences, University of Arizona. Her research focuses on hardware and software design and analysis; and mathematical study and development for optical metrology systems, particularly for high-resolution and freeform optics.

Dr. Chang-jin Oh is the PI of Optical Engineering and Fabrication Facility in College of Optical Sciences and he has provided leadership role to the group for the last 15 years. His expertise includes large scale optomechanical systems for space and ground base application, optical fabrication and precision metrology. He has been PI and Co-PI of various many challenging large optics projects for astronomy and...
optics community as well as national and international science organizations. He has successfully managed many large optics projects including HET wide field corrector, Daniel K. Inoue Solar Telescope Primary mirror fabrication, 6.5m Primary fabrication and large aperture vacuum compatible reference optical systems.

Dae Wook Kim is an assistant professor of optical sciences and astronomy at the University of Arizona. He has been working in the optical engineering field for more than 10 years, mainly focusing on very large astronomical optics, such as the 25 m diameter Giant Magellan Telescope primary mirrors. His main research area covers precision freeform optics fabrication and various metrology topics, such as interferometric test systems using computer generated holograms, direct curvature measurements, and dynamic deflectometry systems. He is currently a chair of the Optical Manufacturing and Testing conference (SPIE), Optical Fabrication and Testing conference (OSA), and Astronomical Optics: Design, Manufacture, and Test of Space and Ground Systems conference (SPIE). He is a senior member of OSA and SPIE and has been serving as an associate editor for the journal Optics Express.
Appendix B. Chebyshev gradient polynomials for high resolution surface and wavefront reconstruction

Maham Aftab, James H. Burge, Greg A. Smith, Logan Graves, Chang-jin Oh, and Dae Wook Kim.

Published as conference proceedings in SPIE 10742, Optical Manufacturing and Testing XII, 1074211 (2018).

Permission to use this article in the dissertation was obtained from SPIE on March 13, 2019.
Chebyshev gradient polynomials for high resolution surface and wavefront reconstruction

Chebyshev Gradient Polynomials for High Resolution Surface and Wavefront Reconstruction

Maham Aftab\textsuperscript{a}, James H. Burge\textsuperscript{a}, Greg A. Smith\textsuperscript{a}, Logan Graves\textsuperscript{a}, Chang-jin Oh\textsuperscript{a}, Dae Wook Kim\textsuperscript{a,b}

\textsuperscript{a}College of Optical Sciences, University of Arizona, 1603 E. University Blvd., Tucson, AZ 85721, USA
\textsuperscript{b}Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85719, USA

ABSTRACT

A new data processing method based on orthonormal rectangular gradient polynomials is introduced in this work. This methodology is capable of effectively reconstructing surfaces or wavefronts with data obtained from deflectometry systems, especially during fabrication and metrology of high resolution and freeform surfaces. First, we derived a complete and computationally efficient vector polynomial set, called \( G \) polynomials. These polynomials are obtained from gradients of Chebyshev polynomials of the first kind – a basis set with many qualities that are useful for modal fitting. In our approach both the scalar and vector polynomials, that are defined and manipulated easily, have a straightforward relationship due to which the polynomial coefficients of both sets are the same. This makes conversion between the two sets highly convenient. Another powerful attribute of this technique is the ability to quickly generate a very large number of polynomial terms, with high numerical efficiency. Since tens of thousands of polynomials can be generated, mid-to-high spatial frequencies of surfaces can be reconstructed from high-resolution metrology data. We will establish the strengths of our approach with examples involving simulations as well as real metrology data from the Daniel K. Inouye Solar Telescope (DKIST) primary mirror.

Keywords: Surface reconstruction, Surface measurements, Optical metrology, Information processing, Deflectometry, Testing, Modal fitting, Numerical approximation and analysis

1. INTRODUCTION

Deflectometry \cite{1-5} is an optical metrology technique particularly effective for the measurement of high-resolution and freeform surfaces. It measures the slope of the test optic surface in a modest package, requiring only a light source and a camera. To obtain the surface sag map from measured slope data, integration must be performed either with a zonal approach such as Southwell integration \cite{6} or a modal one such as proposed in this paper.

In this work we show how our modal reconstruction method, using the proposed Chebyshev gradient polynomials (called \( G \) polynomials) can work well for large, freeform and especially high-resolution surfaces, such as the Daniel K. Inouye Solar Telescope (DKIST) \cite{7}, while also dealing with some common metrology requirements for such projects. Some of these requirements are a large dynamic range of slopes and local obscurations from small fiducials commonly placed on the surface of the mirror during its testing. These fiducials are used as reference markers and can significantly impact the accuracy of surface reconstruction. The DKIST primary mirror is a 4.2 m aperture off-axis parabola. This telescope, which will be to-date the largest optical solar telescope ever built, aims to provide vastly enhanced spatial, spectral and image resolution for solar observation. The surface quality specifications for this project were quite challenging, particularly requirements for the surface figure, irregularity and BRDF (Bidirectional Reflectance Distribution Function). This may be the smoothest large mirror ever made \cite{8}. One big challenge with fabrication of such large, freeform mirrors is that the metrology method requires high dynamic range coupled with excellent resolution and accuracy. Deflectometry proved to be a great metrology tool for the DKIST primary and this paper includes examples of data processing from this mirror.

*letter2dwk@hotmail.com
Surface reconstruction techniques can be divided into two general categories: (a) Zonal methods, such as the Southwell zonal technique, which perform a numerical integration zone by zone and takes a localized approach to the measurement data and (b) Modal methods, such as the one with G polynomials, which reconstruct data using polynomial fitting and looks at the data over the entire aperture in one go. Both methods have their pros and cons. Generally, several factors such as processing speed, need for high-resolution data fitting and type of aperture etc. affect the choice of the surface / wavefront reconstruction technique. Modal methods are generally less sensitive to measurement noise and the number of modes or polynomials used can be adapted to the situation at hand [9]. Li et al. [10] discuss a zonal integration method for deflectometry that shows better results than the traditional Southwell zonal integration method, but it still underperforms compared to modal fitting in terms of reconstructed surface errors. Results from different modal methods may also vary as the type of polynomials used can have different effects. It is usually preferable to use orthogonal polynomial basis sets for many reasons [11].

We introduced the vector polynomials set (called G polynomials), orthogonal over a rectangular pupil and its associated modal fitting method in [12]. In this work, we expand on that method and show some of its other applications and advantages. It is important to note that this technique was not developed to be a representation of balanced aberrations over a rectangular aperture. We do not aim to provide a modal solution to optical design problems where aberration analysis, based on the coefficients of the polynomial set, is critical. Instead, we use this method as an alternative to zonal integration, for the reasons discussed in [12]. Furthermore, the optical systems under consideration in [12] and also in this work are neither rotationally symmetric nor anamorphic.

More and more optical systems are now seen to be employing rectangular apertures. Some examples are the Gaia telescope system [13, 14] and X-ray mirrors [15]. Many imaging applications may need to utilize the full detector size, which generally has a rectangular format. Although G polynomials are optimal for fitting rectangular datasets because of their orthogonality across them, the method is general enough to reconstruct other datasets. It may be particularly useful for non-symmetric or freeform shapes. This methodology also shows great results in the presence of markers such as fiducials used during deflectometry measurements or blockers such as telescope spiders. It also works well for cases where the wavefront is sampled non-traditionally such as uneven sampling in the x and y directions.

We start with a recap of the Chebyshev gradient polynomials basis set upon which the reconstruction technique is based, in Section 2, describing the development and advantages of these polynomials. Next, in Section 3, we provide an overview of deflectometry systems and specifically metrology of the DKIST primary mirror. Also explained are some properties of deflectometry and the DKIST mirror that could benefit from the G polynomials data processing technique. Section 4 provides specific cases of surface reconstruction and an evaluation of the G polynomials fitting technique for those examples. The cases with DKIST data are compared to the traditional Southwell zonal method reconstructions, which was used as the standard reconstruction technique for deflectometry measurements. The conclusion is presented in Section 5.

2. MODAL BASIS FOR SURFACE / WAVEFRONT RECONSTRUCTION

The foundation of this data fitting methodology lies in the modal basis set used. Two dimensional Chebyshev polynomials were chosen as the scalar basis set (called F polynomials), the gradient of which is used as the vector polynomial set (called G polynomials). F polynomials are obtained by the product of two one dimensional Chebyshev polynomials (called T polynomials). Eqn.1 is a mathematical description of the T polynomials and Eqn.2 of the F polynomials and Eqn.3 describes the derivation of the G polynomial set.

\[
T_{m+1}(x) = 2xT_m(x) - T_{m-1}(x) \quad \text{where} \quad T_0(x) = 1, \quad T_1(x) = x, \quad \text{for} \quad -1 \leq x \leq 1
\]

\[
F_n(x,y) = T_n(x)T_n(y)
\]

\[
G_n(x,y) = \nabla F_n(x,y) = \frac{\partial}{\partial x} F_n(x,y)i + \frac{\partial}{\partial y} F_n(x,y)j
\]

Chebyshev polynomials were used as the basis for modal fitting for a variety of reasons, one of the most important one being their orthogonality across rectangular pupils [16]. Ref [12] lists many other qualities of Chebyshev polynomials...
that make them an attractive option for data fitting such as the fact that in many cases their discrete orthogonality holds exactly.

We demonstrated that the gradients of the F polynomials set (i.e. G polynomials) are also orthogonal over rectangular apertures [12]. This is an exciting prospect as many vector polynomial sets, derived from orthogonal scalar polynomials, do not hold orthogonality over the specified aperture and must be orthogonalized, using a procedure such as the Gram-Schmidt orthogonalization process [17]. For example, this is the case for Zernike based rectangular gradient polynomials [18]. Although these polynomials have advantages for aberration balancing, we do not require the representation of balanced aberrations, as explained in Section 1. Hence, it is advantageous for our applications to use Chebyshev polynomials. The fact that G polynomials form an orthogonal set makes data processing and the surface / wavefront reconstruction process much simpler and computationally efficient. Additionally, this is one of the reasons why this methodology can make use of up to hundreds of thousands of polynomials.

The data fitting process is straightforward. Slope data (measured or simulated) can be expanded in terms of the gradient polynomial basis set, to obtain the coefficients of the expansion. These vector polynomial coefficients are then converted to coefficients of the scalar polynomial basis, using the previously derived relationship between the two [12]. Finally, using the scalar polynomial coefficients and the corresponding polynomial terms, the wavefront or surface is obtained. A great feature of the modal fitting method using G polynomials is that the relationship between scalar and vector polynomial coefficients is one-to-one. In other words, the scalar polynomial coefficients are the same as their vector polynomial coefficient counterparts. This further improves the accuracy and computational efficiency of the fit, especially when a large number of polynomial terms are involved. Lastly, it is worth mentioning that recursive relationships for generating both scalar and vector polynomial sets (F and G respectively) were derived and the error between the exact and recursive forms, even for very high orders, was negligible. Details of the polynomial sets, the mathematical background, data fitting method and properties of all aforementioned topics are detailed in Ref [12].

3. METROLOGY OF HIGH-RESOLUTION AND FREEFORM OPTICS

Freeform surfaces have rapidly been gaining popularity as they allow for additional degrees of freedom for optical design and instrumentation. Some advantages of such surfaces include system miniaturization / compactness by allowing for decreased system mass and volume, higher throughput due to reduced component count and innovative designs, better aberration correction and higher image quality and a larger field of view, with constant performance across it. Examples of systems that could greatly benefit from freeform optics include head mounted displays, very compact camera systems, asymmetric solar energy concentrators and segmented extremely large telescopes.

However, the additional degrees of freedom made possible by freeform optics fundamentally increases the positioning uncertainty and makes it challenging to simultaneous meet the requirements for optical form, mid-to-high spatial frequencies, and surface micro-roughness. Manufacturing and testing of precision freeform optics in particular requires metrology systems with a large dynamic range of measurement. Deflectometry is a good solution due to its high accuracy and large dynamic range.

Similarly, deflectometry provides a great option for measurement of high-resolution surfaces, such as telescope mirrors. An example we will use in particular for this work is the primary mirror of the Daniel K. Inouye Solar Telescope (DKIST), which is a 4.2 meter parabolic segment with more than 8 mm peak-to-valley aspheric departure [19]. This mirror was fabricated at the Optical Engineering and Fabrication Facility (O EFF) at the College of Optical Sciences, University of Arizona. Fig.1 is a schematic of a generic reflective deflectometry system, configured for testing a mirror like the DK IST primary. Light from a source, such as a computer monitor for visible wavelengths or a hot tungsten wire for infrared wavelengths, is incident on the test optic. Light reflected from the test mirror is then recorded by a camera or detector. The source is a known pattern e.g. a sinusoidal wave of a certain frequency. By measuring the change in the recorded pattern, which is the reflection of the source off of the test mirror, the local slopes of the test surface can be calculated. As mentioned in the introduction, metrology of this high-resolution, freeform mirror was a challenging task. At various stages during the fabrication process, different deflectometry systems were used for measurements and analysis [20].
Deflectometry systems measure the slope of the test surface and have traditionally used Southwell integration for reconstructing the surface being measured. Typically, this works well except in some cases such as when fiducials are placed on the mirror during a metrology run or when there is a scratch on the surface of the optic. Modal methods may work better for surface reconstruction in general under such circumstances, but they often fail to accurately represent high-resolution and freeform surfaces since such surfaces require a very large number of polynomials to be accurately characterized. Using simulated and real data, we show how the $G$ polynomial method can overcome this problem as it can use up-to hundreds of thousands of polynomials for surface reconstruction, as required by the specific situation.

4. SURFACE RECONSTRUCTION EXAMPLES

This section contains several cases of surface reconstruction that use the $G$ polynomials modal method, to highlight some of the situations where this technique can be useful. Also included, where appropriate, are comparisons against the traditional Southwell zonal method.

4.1 DKIST surface reconstruction

The Daniel K. Inouye Solar Telescope (DKIST) is a good reference to highlight some of the advantages of the $G$ polynomial method. Slope measurements of the DKIST primary mirror were made at the College of Optical Sciences, University of Arizona using deflectometry. This data represents the high-resolution, freeform surface of the mirror. Ref [12] contains reconstruction examples from DKIST data, showing the basic fitting accuracy as well as the accurate representation of the mid-to-high spatial frequencies by comparing high-pass filtered data from the modal reconstruction against the high-pass filtered Southwell reconstructed data, where the Southwell reconstruction is taken as the reference. Here, we will explore a few more examples related to reconstruction of the DKIST surface from its measured deflectometry data and why the $G$ polynomials are a good option for this.

First, we take data from the DKIST mirror, over a rectangular aperture, which measures $\sim 1.84 \times 3.06$ m and contains $150 \times 250$ pixels, giving it an aspect ratio of 0.6. The data set contains noise, measurement errors etc. that are expected from a real measurement. Fig.2 shows the reconstruction using Southwell zonal method, which is taken as the reference, the $G$ polynomial method using 50 polynomials and the difference map between the two surfaces. The difference or error map has a surface RMS of $8.308 \times 10^{-3}$ $\mu$m, which is 0.133% of the reference surface RMS. This shows a good agreement between the two methods, even for a rectangular data set with a relatively large difference between the length and width of the aperture (i.e., a small aspect ratio). From this and other examples [12] we can conclude that the $G$ polynomials method can accurately reconstruct data over rectangular apertures.
Figure 2. DKIST surface over a 1.84 × 3.06 m aperture, reconstructed using (a) Southwell zonal and (b) G polynomial methods. (c) Error map between the two reconstructions ((a) – (b)).

For the second example, we use measured DKIST data, similar to the first example and add fiducials to it in simulation. The analyzed data had an area of 225 × 225 pixels, which is roughly 2.76 × 2.76 m. Nine blockers (modeling fiducials), each having dimensions of 6 × 6 pixels, corresponding to ~7.35 × 7.35 cm were added to both the reference surface (that used the Southwell method) as well as the slope maps. The slope data with fiducials was reconstructed using both the Southwell and G polynomials method and compared. Using 500 G polynomials, the difference between the Southwell and modal reconstructions is 0.103 μm or 1.609% of the surface RMS of the Southwell surface. This is shown in Fig.3 with the map of the ideal surface as well as the error map, which is the difference between the Southwell method and G polynomial method.

Figure 3. (a) DKIST surface over a 2.76 × 2.76 m aperture, with 7.35 × 7.35 cm fiducials added (in simulation), representing the reference surface. (b) Error between the Southwell method and G polynomial reconstruction of the reference surface.

4.2 Reconstruction for unevenly sampled data

Although data is often simulated / measured over equidistance points in both x and y directions and the sampling for the two directions is also the same (i.e., the same sampling in y as in x), there may be several situations where this is not the case such as with non-square detector pixels or the presence of astigmatism in the camera lens. In these cases, non-uniform sampling in either or both directions or different sampling in the x and y directions may be required. We will investigate both types of situations using simulated data, and assess how well the reconstructions using the G polynomial modal method perform in such instances.

For the first example, we generate data over a square pupil, representing a 2 × 2 cm aperture, with a sampling of 500 × 300 points. The ideal, simulated, noise-free surface sag is described by Eqn.4.

\[
z = 0.15y^2 + 1.4x^4 - 0.2(x^3 + xy^2) + 0.3y^5 + 1 \times 10^{-3}x^{11} - (xy)^8
\]  

After obtaining x and y gradients of the data numerically, surface reconstruction is performed using the G polynomials. Fig.4 shows the ideal data, as well as the difference map between the ideal and reconstructed surfaces, when 250 G polynomials are used. RMS of the simulated surface is 0.403 μm while the RMS of the error map using is 1.35 ×10^{-3} μm or 0.33% relative to the ideal surface RMS. It must be noted that since we are using slope data, the mean (representing the piston) is subtracted from the ideal surface as well as its reconstruction.
In the second example, sampling was kept the same for the x and y directions but the sampling along the aperture was changed. Over a 2 × 2 cm aperture, the first 1.25 cm was sampled with 150 points, making the sampling 1/120 cm. The remaining 0.75 cm was also sampled with 150 points, making the sampling in this section 1/200 cm. Data was simulated over the entire aperture, using Eqn. 4. This gradient data was then reconstructed using the G polynomial modal method. The simulated surface RMS is 0.392 µm and the modal method error RMS, using 250 polynomials, is 2.66×10⁻³ µm, or 0.68% relative to ideal surface RMS. Once again, the mean is subtracted from both the ideal surface and its reconstruction. Fig.5 shows the ideal surface map, with an estimation of the regions of different sampling. The error maps between the ideal surface and its G polynomial based reconstruction is also shown.

Figure 5. (a) Simulated, ideal surface map (reference). The blue double ended arrows mark the approximate region on the map (1.25 × 1.25 cm) with a lower sampling (b) Error between the reference and G polynomial generated surface.

5. CONCLUSION

This paper builds upon the newly developed G polynomials modal method for surface / wavefront reconstruction. We provide a brief summary of this method that uses the gradients of Chebyshev polynomials, which are orthogonal over a rectangular aperture. We reiterate the properties of this polynomial set and its attractive ability of being able to generate hundreds of thousands of polynomial terms, which is necessary for the correct representation of high-resolution and freeform surfaces. Since this vector polynomial fitting methodology reconstructs wavefronts / surfaces from slope data and we place a heavy emphasis on high-resolution surfaces that do not have rotational or anamorphic symmetry, we use examples from deflectometry data. We have also provided an overview of deflectometry concepts and techniques and considered specifically the case of the Daniel K. Inouye Solar Telescope (DKIST) primary mirror metrology. Deflectometry provides high-resolution slope data of this freeform surface. Through examples of DKIST surface reconstruction in different cases, such as when fiducials are placed on it, we highlight some of the advantages of our modal data processing method. Also explored are examples from simulated data which highlight certain situations where this G polynomial method could prove to be useful, e.g. with uneven sampling in the two orthogonal Cartesian directions. This methodology can prove to be a very useful tool and along with traditional reconstruction techniques for our examples, such as the Southwell zonal method, it can provide a more accurate and effective tool for surface / wavefront reconstruction from slope data.

Figure 4. (a) Simulated, ideal surface map (reference) (b) Error between the reference and G polynomial generated surface.
ACKNOWLEDGMENTS

This material is partly based on work performed for the DKIST. DKIST is managed by the National Solar Observatory, which is operated by the Association of Universities for Research in Astronomy Inc. under a cooperative agreement with the National Science Foundation. Also, it is based in part upon work performed for the “Post-processing of Freeform Optics” project supported by the Korea Basic Science Institute. The deflectometry related software development is partially funded by the II-VI Foundation Block grant.

REFERENCES

Appendix C. Rectangular domain curl polynomial set for optical vector data processing and analysis


Submitted for publication in Optical Engineering (April, 2019).
Rectangular Domain Curl Polynomial Set for Optical Vector Data Processing and Analysis

Maham Aftaba, Logan R. Gravesa, James H. Burgea, Greg A. Smitha, Chang-jin Oha, and Dae Wook Kim\textsuperscript{a,b,*}

\textsuperscript{a}The University of Arizona, Large Optics Fabrication and Testing group, James C. Wyant College of Optical Sciences, 1630 E. University Blvd, Tucson, AZ, USA 85721
\textsuperscript{b}The University of Arizona, Steward Observatory, 933 N. Cherry Ave., Tucson, AZ, USA 85719

Abstract. A list of up to six keywords should immediately follow, with the keywords separated by commas and ending with a period. Rectangular pupils are employed in many optical applications such as lasers, anamorphic optics etc. as well as for detection and metrology systems such as some Shack–Hartmann wavefront sensors and deflectometry systems. For optical fabrication, testing and analysis in the rectangular domain, it is important to have a well-defined set of polynomials that are orthonormal over a rectangular pupil. Since we often measure the gradient of a wavefront or surface, it is necessary to have a polynomial set that is orthogonal over a rectangular pupil in the vector domain as well. In this paper, we derive curl (called $C$) polynomials based on two-dimensional versions of Chebyshev polynomials of the first kind. Previous work derived a set of polynomials (called $G$ polynomials) that are obtained from the gradients of the 2D Chebyshev polynomials. This work shows how the two sets together can be used as a complete representation of any vector data in the rectangular domain. The curl polynomials themselves or the complete set of $G$ and $C$ polynomials has many interesting applications. Two of those applications shown in this paper are systematic error analysis and correction in deflectometry systems and mapping imaging distortion.

Keywords: Measurement and metrology, Surface measurements, Numerical approximation and analysis, Optical instrumentation, Information processing, Testing.

* Dae Wook Kim, Email: letter2dwk@hotmail.com

1 Introduction

Previous works highlight the importance of fitting vector slope data in the domain of measurement as well as using orthonormal polynomials for the fitting [1 – 3]. To summarize: the fit to a non-orthogonal basis set can require many more terms than are necessary, and expansion coefficients themselves are not meaningful, because the value for any particular coefficient changes as the number of expansion terms changes. Also, when fitting to real data, the noise propagation is
increased with the use of non-orthogonal basis functions. It is simpler and more computationally efficient to use orthogonal polynomials, as they can easily be calculated using the inner product.

In reference [3] Chebyshev polynomials are chosen for the basis set from which we will derive the gradient and curl polynomials. This reference also defines gradient polynomials, denoted as the $G$ set, which are generated from the gradients of the Chebyshev polynomials of the first kind, and are orthogonal over a rectangle. As described in [3 – 6], these polynomials can be used for reconstructing high-resolution and freeform surfaces or fitting data from a Shack–Hartmann sensor etc.

In this paper, we will define an additional set of vector polynomials that only consist of terms with zero divergence – defined locally as a rotation or a curl. We explore the use of these curl polynomials and describe how we can combine them with the gradient polynomials to get a complete set of orthogonal vector polynomials.

Consideration of the curl calculation is used in many areas of optics ranging from vortex beams [7, 8] and lithography photomasks [9] to gravitational lensing [10]. In each of these cases, electromagnetic fields with non-zero curl components are used to create or detect significant effects. In other cases, expectation of zero curl in a vector field can improve signal estimation in the presence of noise [11].

The paper starts with background knowledge in Section 2 such as the gradient polynomials and an introduction to the two major applications covered in this work – deflectometry and imaging distortion. Section 3 is dedicated to details of the curl ($C$) polynomials, including their derivation and other details. Section 4 talks about the practical applications and uses of the $G$ and $C$ polynomials while concluding remarks are presented in Section 5.
2 Background

2.1 Chebyshev Polynomial and Gradient Polynomial Basis

As described in reference [3], one can use 2D Chebyshev polynomials of the first kind, as well as their derivatives to easily generate orthogonal polynomial sets.

The polynomial basis function is defined as:

$$F_n^m(x, y) = T_n(x)T_m(y)$$

(1)

where the 1D Chebyshev polynomials are:

$$T_m(x) = \begin{cases} 2xT_{m-1}(x) - T_{m-2}(x) & \text{if } m > 1 \\ 1 & \text{if } m = 0, 1 \\ 0 & \text{if } m = -1 \end{cases}$$

(2)

The 2D set is defined in both x and y on the interval -1 to +1 [12].

Another way to describe the Chebyshev polynomials is using their definitions in terms of sines and cosines. This definition, shown in Eq. 3, is useful if we want to prove the orthogonality of the polynomials and to determine their normalization factor [3].

$$T_m(x) = \cos(m\cos^{-1}(x))$$

(3)

The details of these polynomial, their orthonormality and plots can be found in reference [3].

The derivatives of the one-dimensional Chebyshev polynomials are:

$$T_m'(x) = \frac{m[T_{m-1}(x) - xT_{m}(x)]}{(1-x^2)}$$

$$T_m'(y) = \frac{n[T_{m-1}(y) - yT_{m}(y)]}{(1-y^2)}$$

(4)

The gradient polynomial set is defined as:

$$\bar{G}_j(x, y) = \bar{G}_n^m = \nabla F_n^m(x, y) = \frac{\partial}{\partial x} F_n^m(x, y) \hat{i} + \frac{\partial}{\partial y} F_n^m(x, y) \hat{j}$$

(5)
In terms of the 1D Chebyshev polynomials and their derivatives, the gradient polynomials can be written as:

\[ \tilde{G}_n^m(x, y) = T_n(y)T'_m(x)\hat{j} + T_m(x)T'_n(y)\hat{j} \]  
(6)

Where the conversion between the single index \((j)\) and double index \((n, m)\) can be found in [3]. The notation \(T'\) refers to the derivative of the \(T\) polynomial (as described in Eq. 4).

In terms of the recursive Chebyshev polynomials, the gradient polynomials can be written as:

\[ \tilde{G}_n^m(x, y) = \frac{m}{2y(1-x^2)}\left( T_{n+1}(y) - T_{n-1}(y) \right) \left( T_{m+1}(x) - xT_m(x) \right) \hat{i} \]
\[ + \frac{n}{2x(1-y^2)}\left( T_{m+1}(x) - T_{m-1}(x) \right) \left( T_{n+1}(y) - yT_n(y) \right) \hat{j} \]

(7)

The orthogonality and normalization factors are given as:

\[ \int \int_{-1}^{1} (\tilde{G}_j(x, y) \tilde{G}_j(x, y)) \sqrt{1-x^2} \sqrt{1-y^2} \, dx \, dy = \begin{cases} 0, & j \neq j' \\ N_G, & j = j' \end{cases} \]

(8)

\[ \frac{1}{N_G} \int \int_{-1}^{1} \left( T_n(y)T'_n(y)\hat{j} + T_m(x)T'_m(x)\hat{j} \right) \left( T_n(y)T'_n(y)\hat{j} + T_m(x)T'_m(x)\hat{j} \right) \sqrt{1-x^2} \sqrt{1-y^2} \, dx \, dy = \delta_{m,m'}\delta_{n,n'} \]

(9)

where \(m, m', n,\) and \(n'\) are integers and \(N_G\) is the normalization factor, given by:

\[ N_G \begin{cases} \pi^2n^2/4, & m = 0 \\ \pi^2m^2/4, & n = 0 \\ \pi^2/8(m^2 + n^2), & \text{otherwise} \end{cases} \]

(10)

2.2 Deflectometry

Section 4 describes experimental verifications of the calculations using a method known as deflectometry. This section provides a brief overview of that process.
Deflectometry is a non-null optical metrology method that measures the local slopes on a test surface. The method has been used extensively for precision metrology of symmetric and freeform optical surfaces [13 – 17]. Because it is a non-null test method, it is a desirable metrology tool for freeform and non-axisymmetric optics, as well as for metrology during a fabrication process. In both cases, a single deflectometry configuration can maintain testing of the unit under test without extensive changes. The measurable surface area of the test optic and slope range, known as the dynamic range, directly depend on the hardware configuration. A deflectometry system utilizes an illuminating source which emits light, sending some rays to the test optic, which are then reflected and recorded by a camera. If a clear line of sight can be made from a camera pixel to the test optic, and then after reflection to a point within the source area, obeying the law of reflection, then the deflectometry system will be able to test the local slope distribution of the test optic.

In a deflectometry test, the positions of all components in the system must be measured to high accuracy. For testing a concave optic, the camera and source will typically be placed close to the center of curvature of the optic, to achieve closer to a conjugate imaging condition to minimize the required source area. The camera is focused on the test optic surface, such that the camera pixels are mapped to the test optic surface, with each pixel representing a ‘mirror pixel’, or a finite local area over which the local slope will be measured. The source is then precisely modulated, such that for every camera pixel, the precise location on the source that satisfied peak illumination is known.

In visible deflectometry (~400 – 700 nm) a common source is a digital screen, which can display a phase stepped sinusoidal pattern in the horizontal and vertical directions. This method, known as phase-stepping deflectometry, has been explored extensively [18, 19]. Another option, which has been used for infrared deflectometry in the ~7 – 12 μm range as well as the visible range, is a
scanning line source, which is scanned in the horizontal and vertical directions [20, 21]. In either case, by precisely knowing the start location of a ray at the source, the local point of incidence on the test optic surface, and the final detection point at the camera, the local slope of the test optic can easily be calculated using the law of reflection. The standard setup for a deflectometry test of a concave test optic is shown in Figure 1.

**Fig. 1** A deflectometry test utilizes a source, which emits rays that strike a Unit Under Test (UUT) and are reflected, after which some are detected by a camera. For a concave test optic, the source and camera are placed near the center of curvature (C.C.) of the test optic. Then, by precisely measuring the location of all components, the local slopes of the test optic can be output by a deflectometry measurement.

After measuring the local slopes of the test optic using a deflectometry test, further processing is required to reconstruct the optical surface shape. Typically, this included further processing to reduce systematic errors, and an integration of the local slopes to generate a final reconstructed surface. The integration method can be a zonal [22] or modal [3] approach. It is important to note that any uncertainty in the location of all objects in a deflectometry test will reduce the slope estimation accuracy.
2.3 Imaging distortion

Imaging distortion [23] is a very common problem in optics. It is a deviation from rectilinear projection. As a consequence of distortion, the image magnification does not remain constant. It either increases or decreases with distance from the optical axis. Although most imaging systems suffer from some amount of distortion, one case where it is very significant is interferometric null testing of steep aspheric surfaces [24, 25]. Figure 2 shows an example of what the beam footprint looks like for an interferometric null testing system. It is relatively easy to quantify and find a mapping correction for the distortion. The most common approach is to place a set of fiducial markers on the optic being tested, which have known, measured locations. The measured image will contain the distorted positions of the fiducials. By constructing a mapping model between the real and measured fiducial locations, distortion can be calibrated out of the system. One of the ways to construct the mapping distortion model is to use a polynomial fit. Reference [24] gives an example of imaging distortion and introduces a modal fitting method for it, based on the $S$ & $T$ polynomials [1, 2], which are a complete vector domain set, based on the gradient and curl of Zernike polynomials respectively, and orthogonalized across a circular aperture. They can be used to represent any vector data in the circular domain. In Section 4.2, we provide an example of system distortion for a telescope and propose its fitting with two vector polynomial based methods – one using the $S$ & $T$ and another, the $G$ & $C$ polynomials.
Fig. 2 Example of a mapping distortion from interferometric null test, showing the footprint of the beam through the system when it is (a) undistorted and (b) distorted.

3 The curl polynomial set

It is known that vector fields can be uniquely specified by giving their divergence and curl within a region and the normal component over the boundary. This leads to Helmholtz's theorem which allows us to express the vector field \( \mathbf{V} \) as the sum of an irrotational \( \mathbf{\phi} \) and a solenoidal part \( \mathbf{P} \), as follows:

\[
\mathbf{V} = \nabla \phi + \nabla \times \mathbf{P}
\]

The divergence of the vector field is:

\[
\nabla \cdot \mathbf{V} = \nabla \cdot \phi
\]

(11)

Its curl is:

\[
\nabla \times \mathbf{V} = \nabla \times (\nabla \times \mathbf{P})
\]

(13)

Since the \( G \) polynomials are defined as the gradients of scalar functions they have no curl component.

We want to derive another vector set (\( C \) polynomials) that have zero divergence. So, \( C \) polynomial set has zero \( \phi \), but non-zero \( P \). Hence it is defined as:
Like the $G$ polynomials, the $C$ polynomials are defined in the \(x-y\) plane only. One way to do that is to use a scalar \(\phi\), which is a function of \(x\) and \(y\), to represent \(P\) as follows:
\[
\vec{P} = \phi \hat{k}
\] (15)

Then the curl polynomials are then:
\[
\vec{C} = \nabla \times \vec{P} = \begin{vmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P_x & P_y & P_z \end{vmatrix}
\] (14)

For the $C$ polynomials, we can use the same basis functions \((F)\) that were used for generating the $G$ polynomials. Then, by Eq. (8), we know that the $C$ polynomials will be mutually orthogonal. Hence, we can also define them as:
\[
\vec{C}_j(x,y) = \vec{C}_m(x,y) = \frac{\partial}{\partial y} F^n_m(x,y) \hat{i} - \frac{\partial}{\partial x} F^n_m(x,y) \hat{j},
\]
\[
\vec{C}_n(x,y) = T_n(x) T'_m(y) \hat{i} - T_n(y) T'_m(x) \hat{j}
\] (17)

The conversion between the double index \((n, m)\) and single index \((j)\) is the same as for $G$ polynomials [3].

\(\vec{C}_i(x,y)\) and \(\vec{C}_j(x,y)\) are orthogonal to each other at any point in a unit rectangle for terms where the Laplacian is non-zero.

Using Eq. (2) and Eq. (4), we can define the $C$ polynomials in terms of the recursive Chebyshev polynomials of the first kind as:
\[
\vec{C}_n(x,y) = \frac{n}{2x(1-y^2)} \left( T_{n+1}(x)-T_{n-1}(x) \right) \left( T_{n+1}(y)-y T_{n}(y) \right) \hat{i} \\
- \frac{m}{2y(1-x^2)} \left( T_{m+1}(y)-T_{m-1}(y) \right) \left( T_{m+1}(x)-x T_{m}(x) \right) \hat{j}
\] (18)
The orthonormality of the basis set is defined the same way as with the gradient set:

\[
\int_{-1}^{1} \int_{-1}^{1} (\hat{C}_j(x,y) \hat{C}_{j'}(x,y)) \sqrt{1-x^2} \sqrt{1-y^2} \, dx \, dy = \begin{cases} 
0, & j \neq j' \\
N_c, & j = j'
\end{cases}
\] (19)

This can also be described as:

\[
\frac{1}{N_c} \int_{-1}^{1} \int_{-1}^{1} \left( -(T_n(x)T_n'(x) \hat{j} + T_n'(x)T_n(y) \hat{i}) - (T_{n'}(y)T_{n'}'(x) \hat{i} + T_{n'}'(y)T_{n'}(x) \hat{j}) \right) \sqrt{1-x^2} \sqrt{1-y^2} \, dx \, dy = \delta_{nm} \delta_{n'n'}
\] (20)

The normalization factor is the same as for \( G \) polynomials as given in Eq. (10).

\[
N_c \begin{cases} 
\frac{\pi^2 n^2}{4}, & m = 0 \\
\frac{\pi^2 m^2}{4}, & n = 0 \\
\frac{\pi^2}{8} (m^2 + n^2), & \text{otherwise}
\end{cases}
\] (21)

The derivation of orthonormality and how to get the normalization factor is nearly identical to reference [3]. Table 1 lists the first 15 curl polynomials and Appendix 1 illustrates several terms graphically.

<table>
<thead>
<tr>
<th>( m )</th>
<th>( n )</th>
<th>( \hat{C}_j )</th>
<th>Simplified form</th>
<th>Expressed as Chebyshev polynomials of first kind</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>( \hat{C}_0 )</td>
<td>( \hat{0i} - \hat{0j} )</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>( \hat{C}_1 )</td>
<td>( -\hat{j} )</td>
<td>( -\frac{T_0(y)\left[T_0(x) - xT_1(x)\right]}{(1-x^2)} \hat{j} )</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>( \hat{C}_2 )</td>
<td>( \hat{i} )</td>
<td>( \frac{T_0(x)\left[T_0(y) - yT_1(y)\right]}{(1-y^2)} \hat{i} )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>( \hat{C}_3 )</td>
<td>( -T_2'(x) \hat{j} )</td>
<td>( -\frac{T_0(y)\left[T_0(x) - xT_2(x)\right]}{(1-x^2)} \hat{j} )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>( \hat{C}_4 )</td>
<td>( -y \hat{j} + x \hat{i} )</td>
<td>( -\frac{T_1(y)\left[T_0(x) - xT_1(x)\right]}{(1-x^2)} \hat{j} + \frac{T_1(x)\left[T_0(y) - yT_1(y)\right]}{(1-y^2)} \hat{i} )</td>
</tr>
</tbody>
</table>
Table 1 is simplified using the following relations:

\[
\begin{align*}
T_o(x) &= 1, \quad T_o(y) = 1, \quad T_i(x) = x, \quad T_i(y) = y \\
T'_o(x) &= 0, \quad T'_o(y) = 0, \quad T'_i(x) = 1, \quad T'_i(y) = 1
\end{align*}
\]

(22)

3.1 Two sets to completely define a vector field

Similar to the circular case in reference [2], we can visualize \( G \) polynomials as irrotational vector fields which have zero curl everywhere and \( C \) polynomials as solenoidal vector fields which have zero divergence everywhere. These two fields have some overlap where both divergence and curl

<table>
<thead>
<tr>
<th>0</th>
<th>2</th>
<th>( \hat{C}_5 )</th>
<th>( T'_2(y)\hat{i} )</th>
<th>( 2 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>( \hat{C}_6 )</td>
<td>( -T'_3(x)\hat{j} )</td>
<td>( -3 \frac{T_o(y)\left[T_i(x)-xT'_i(x)\right]}{(1-x^2)}\hat{j} )</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>( \hat{C}_7 )</td>
<td>( -yT'_2(x)\hat{j}+T'_1(x)\hat{i} )</td>
<td>( -2 \frac{T_o(y)\left[T_i(x)-xT'_i(x)\right]}{(1-x^2)}\hat{j}+2 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>( \hat{C}_8 )</td>
<td>( -T'_1(y)\hat{j}+xT'_i(y)\hat{i} )</td>
<td>( -T_o(y)\left[T_i(x)-xT'_i(x)\right] \hat{j}+3 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
<tr>
<td>0</td>
<td>3</td>
<td>( \hat{C}_9 )</td>
<td>( T'_3(y)\hat{i} )</td>
<td>( \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-x^2)}\hat{i} )</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>( \hat{C}_{10} )</td>
<td>( -T'_4(x)\hat{j} )</td>
<td>( -4 \frac{T_o(y)\left[T_i(x)-xT'_i(x)\right]}{(1-x^2)}\hat{j} )</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>( \hat{C}_{11} )</td>
<td>( -yT'_3(x)\hat{j}+T'_2(x)\hat{i} )</td>
<td>( -3 \frac{T_o(y)\left[T_i(x)-xT'_i(x)\right]}{(1-x^2)}\hat{j}+2 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>( \hat{C}_{12} )</td>
<td>( -T'_2(y)T'_2(x)\hat{j}+T'_1(x)T'_i(y)\hat{i} )</td>
<td>( -2 \frac{T_o(y)\left[T_i(x)-xT'_i(x)\right]}{(1-x^2)}\hat{j}+3 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>( \hat{C}_{13} )</td>
<td>( -T'_1(y)\hat{j}+xT'_i(y)\hat{i} )</td>
<td>( -T_o(y)\left[T_i(x)-xT'_i(x)\right] \hat{j}+3 \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>( \hat{C}_{14} )</td>
<td>( T'_4(y)\hat{i} )</td>
<td>( \frac{T_o(x)\left[T_i(y)-yT'_i(y)\right]}{(1-y^2)}\hat{i} )</td>
</tr>
</tbody>
</table>

Table 1 Curl polynomials in terms of Chebyshev polynomials of the first kind and associated derivatives.
are zero, which is known as the Laplacian vector field. The overlapped area contains terms derived from the corresponding scalar $\nabla$ polynomials (in our case, the 2D Chebyshev polynomials) whose Laplacian is 0. If $\nabla$ (in our case, $F$) represents wavefront, these terms correspond to a wavefront with zero net curvature at any point in the pupil.

In order to completely specify the orthogonal rectangular domain, we can combine the $G$ and $C$ polynomials and to ensure that the Laplacian is not counted twice, the $C$ polynomial set should be modified to only include the independent terms (i.e., all $C$ terms except the ones that are common with the Laplacians). The Laplacian operator is defined as the divergence of the gradient of a function. It can be calculated empirically using the sine and cosine based definition of the $G$ polynomials [3] and it can be programmed numerically using the recursive definitions. A simple implementation would be to take the divergence of the $G$ polynomials and for the terms that equal zero, remove them from the complete set.

As with the gradient polynomials, each term is a combination of two terms, both containing one Chebyshev polynomial and one Chebyshev polynomial derivative terms each. Appendix 1 gives the quiver plots of the first few non-trivial curl polynomials.

4. Employing the vector polynomial set in practical applications

4.1 Deflectometry data analysis using $C$ polynomials

To demonstrate the vector polynomial’s fitting ability for a misaligned optical testing case, a highly freeform Alvarez lens, shown in Figure 3 is measured using a phase-shifting deflectometry setup as described in Section 2.2. The Alvarez lens, made from a 1-inch diameter substrate, was created using a diamond turned computer-numeric control process. The process was set to generate a 6 mm central aperture inside of the disk, having 17 $\mu$m of horizontal coma, and -17 $\mu$m of 45° trefoil
described as Zernike polynomials. A liquid crystal display was utilized as the source and was mounted in place above the Alvarez lens. A camera manufactured by Point Grey was additionally mounted in place above the Alvarez lens. The Alvarez lens itself was placed on a precision manual rotation stage, with rotation accuracy of ±0.05 degrees. The clear aperture of the Alvarez lens for this measurement set-up was a square of length 5.6 mm.

To test the optic, a deflectometry measurement was taken at a 0° clocking position for the Alvarez lens. Next, the Alvarez lens was clocked every 0.5° and repeating measurements were taken.

![Fig. 3](image)

**Fig. 3** The Alvarez lens (left) with a clear aperture of 5.6 × 5.6 mm square (middle) was designed to have only horizontal coma and 45° trefoil. Phase-shifting deflectometry was used to measure the clear aperture lens area (right) at various clocking positions.

The goal is to be able to successfully check for non-orthogonality error of the deflectometry setup and be able to determine the amount of error (i.e., degree of mis-clocking) so it can be corrected / accounted for.

First we construct a ‘clocking error’ model based on known clocking errors from the set-up. We first measure the gradients when the Alvarez lens is set at 0 degree. Then, we start to rotate the optic by known amounts and collect the slope maps for all the rotations. Next, we only use the \(x\)-gradient map from the rotated sets and all the \(y\)-gradients are the unrotated \(y\)-gradient i.e., the \(y\) slope at 0°. This simulates the situation when the system is set-up in such a way that the \(x\) and \(y\)
slopes aren’t taken at a 90° angle, but have some mis-clocking between them. A practical example of this could be the non-orthogonality between the position of the scanning hot wire (line source) in the two scanning directions (x and y) for the case of IR deflectometry [20].

For this experiment, measurements were made every 0.5° from 0° to 2.5°. These data sets were fitted with the first 50 \( C \) polynomials.

Based on the polynomial terms that stood out from all the processed measurements and showed a steady change (increase or decrease) as the mis-clocking between the two slopes increased, we constructed a model that predicted the slope clocking mismatch between the x and y slopes when an unknown mis-clocking was introduced.

The \( C \) polynomials used in the analysis were: 1, 2, 4, 8, 11, 13 and 17. Figure 4 shows the change in polynomial values, as mis-clocking is increased. The absolute value of the polynomials at each mis-clocking angle is not important. What matters is their relative values. These results were input to a least squares fit model, where a linear combination of each of these polynomial terms was used to calculate an overall transformation matrix. When an unknown deflectometry measurement is taken, it can be decomposed into the first 50 (or fewer) \( C \) polynomial terms. Values for \( C \) polynomials 1, 2, 4, 8, 11, 13 and 17 are used to calculate the coefficient of the transformation matrix, which is the predicted clocking for the input measurement.
To test the model, we used two measurements – one at 0.6° and the other at 2.3°. Table 2 shows the error between the predicted and actual mis-clocking. Both predictions showed good agreement with the actual values. Figure 5 shows the value of select $C$ polynomials for the data used to construct the model, as well as the two trial clocked data sets (0.6° and 2.3°).

Fig. 5 Selected $C$ polynomial terms (C2, C4 and C8) used in Alvarez lens deflectometry mis-clocking modeling, shown as plots with lines. Overlaid are corresponding $C$ polynomial values from trial data sets (0.6° and 2.3°),
shown as large cross shaped markers. This is written in the legend as (for example): “R 0.6, C4” meaning the C4 polynomial coefficient for the trial data at 0.6 degrees.

### Table 2

<table>
<thead>
<tr>
<th>True (T) [Degree]</th>
<th>Computed (C) [Degree]</th>
<th>Error (T - C) [Degree]</th>
<th>% Error [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>0.6024</td>
<td>-0.0024</td>
<td>-0.4000</td>
</tr>
<tr>
<td>2.3</td>
<td>2.3917</td>
<td>-0.0917</td>
<td>-3.9870</td>
</tr>
</tbody>
</table>

#### 4.2 Imaging distortion modeling using G and C polynomials

For this case study shown in Figure 6(a), we use a retraceable distortion map, shown in Figure 6(b), from a standard Zemax library model of a Three Mirror Anastigmat (TMA) with rectangular aperture. The TMA camera system parameters are listed in Table 3 while its optical layout can be seen in Figure 6(a).

### Table 3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Focal Length (in air at system temperature and pressure)</td>
<td>216.2162 mm</td>
</tr>
<tr>
<td>Back Focal Length</td>
<td>-383.7838 mm</td>
</tr>
<tr>
<td>Total Track</td>
<td>383.5603 mm</td>
</tr>
<tr>
<td>Image Space F/#</td>
<td>1.528678</td>
</tr>
<tr>
<td>Working F/#</td>
<td>1.564337</td>
</tr>
<tr>
<td>Image Space NA</td>
<td>0.3108736</td>
</tr>
<tr>
<td>Stop Radius</td>
<td>35.36 mm</td>
</tr>
<tr>
<td>Entrance Pupil Diameter</td>
<td>141.44 mm</td>
</tr>
</tbody>
</table>
The distortion map is fit in two ways and compared: one using the $G$ & $C$ polynomials and the other using Zernike-based $S$ & $T$ polynomials [1, 2]. Since Zernike polynomials, as well as both $S$ and $T$ basis sets are orthogonal over a unit circle, and in this example we have rectangular data, the use of Zernike-based polynomial set is not ideal to represent/model the distortion vector distribution. This is not a problem for the $G$ and $C$ polynomial sets because of their orthogonality over rectangular apertures. For simple distortion which is simulated without noise, the corner points are not likely to have a significant impact on fitting performance. However, in practical systems with noise, the removal of points for Zernike-based fitting can lead to inaccurate results.

**Fig. 6** (a) Schematic of the simulated TMA camera system in the Zemax library [26] (b) Distortion map for this system using $20 \times 20$ grid sampling with $15 \times$ arrow size magnification.
For both polynomial sets, the first twelve (6 gradient and 6 curl) polynomials were chosen. It was ensured that the common terms in both sets (i.e., for which the Laplacian are zero) were only counted once. Reference [2] lists the first 12 non-trivial $T$ polynomials, as well as the Laplacian = 0 terms for those twelve terms, so these were selected. For the Zernike-based fit, we used terms 2 – 7 of the $S$ polynomials and terms 4, 7, 8, 11, 12 and 13 of the $T$ polynomial set. For the Chebyshev-based fit, we used terms 1 - 6 for the $C$ polynomials and terms 3, 5, 6, 7, 8 and 9 for the $G$ polynomials.

The residual error from both fits is calculated as the RMS (root mean squared) error in $x$ and $y$ position as well as percentage errors in both positions, calculated as follows:

$$\text{% Error} = 100\% \times \frac{\text{RMS (Real Error)}}{\text{RMS (Reference position))},$$

where reference position is the ideal (simulated) distortion, i.e., predicted - real position from the distortion map.

**Fig. 7** (a) Quiver plots from the TMA camera’s distortion fitting. Red arrows are the ideal (i.e., simulated values), green arrows are results from $S$ & $T$ fitting, and blue arrows are the results from $G$ & $C$ fitting. The quiver plot is downsampled and arrows are magnified 1.5×. The $x$ and $y$ axes are normalized pupil coordinates (the pupil from Fig. 6(b) is normalized from -1 to +1).
As can be seen from the results in Figure 7 and Table 4, the G & C fitting provides more accurate representation of the ideal distortion map, with lower error in both x and y positions.

**Table 4** Fitting performance comparison of the two vector polynomial set cases to represent/model the TMA camera’s distortion map.

<table>
<thead>
<tr>
<th></th>
<th>S &amp; T polynomial fitting</th>
<th>G &amp; C polynomial fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>x-position</td>
<td>0.593</td>
<td>0.255</td>
</tr>
<tr>
<td>y-position</td>
<td>0.654</td>
<td>0.285</td>
</tr>
<tr>
<td>Real Error [µm]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% Error [%]</td>
<td>12.44</td>
<td>5.35</td>
</tr>
</tbody>
</table>

Current literature search did not show an easy way to find the overlapping (Laplacian = 0) terms for the S and T polynomials beyond the 12 terms used in this example. To keep consistency in the comparison, only 12 terms for the G & C polynomials were also used, although we can easily use many more terms and easily obtain the overlap terms (through numerical programming for example), hence reducing the error even further than shown in Table 4. Part of this ease is due to the fact that the G & C polynomials are obtained from direct derivatives of the scalar function (F) while the S & T polynomials require additional manipulation of their scalar (Zernike) basis set. Table 5 shows the improvement in distortion correction as more terms of the G & C polynomials are used, as well as the time taken for the fitting (including polynomial generation). Note that the number of terms in this table refers to the total number of G & C polynomial terms, i.e., 20 terms would mean 10 G and 10 C polynomial terms. The error reduction begins to saturate around 50 terms (25 G and 25 C terms), as we are possibly reaching the limit of numerical accuracy.

**Table 5** Residual error and processing time comparison as a function of the total number of G & C polynomials used to represent/model the TMA camera’s distortion map.
<table>
<thead>
<tr>
<th>No. of terms</th>
<th>% Error [%]</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x-position</td>
<td>y-position</td>
</tr>
<tr>
<td>12</td>
<td>5.35</td>
<td>3.62</td>
</tr>
<tr>
<td>20</td>
<td>0.82</td>
<td>3.12</td>
</tr>
<tr>
<td>50</td>
<td>0.09</td>
<td>0.05</td>
</tr>
</tbody>
</table>

As a comparison, the Zernike based ($S$ & $T$ polynomial) fitting took about 0.11s when the first 12 polynomials were used (results from Table 4). We would like to acknowledge that the processing time depends on the specific numerical implementation of the fitting method and may improve with better algorithms. Although our implementation of either the $G$ & $C$ or $S$ & $T$ polynomial sets was not targeted for optimum processing time, this example serves as a baseline comparison for the two fitting methods.

5. Concluding Remarks

We have extended the previously defined set of vector polynomials that are the gradients of a scalar function and included an orthonormal set that has zero divergence everywhere and can be considered the rotation or curl. The scalar function to generate the curl is the two-dimensional Chebyshev polynomial of the first kind.

The combined set of gradient and curl polynomials (considering only the independent terms, i.e., where the common terms having Laplacian = 0 are only counted once) is useful for fitting any continuously differentiable vector functions in the rectangular domain. In particular, it is useful in modeling or fitting mapping distortions. Numerical examples prove that the combined set works
well to define a measured vector field and can be used to easily calibrate out the misalignment state of a deflectometry test configuration.

The orthonormality of the polynomials give us advantages in terms of noise reduction, computational efficiency and meaningful expansion coefficients for vector fields in rectangular apertures.
Appendix A: Quiver plots of a few non-trivial $C$ polynomials

$C_3$

$C_4$

$C_5$

$C_6$

$C_7$

$C_8$

$C_9$

$C_{10}$
Acknowledgment

This research was made possible in part by the II-VI Foundation Block-Gift Program, the Korea Basic Science Institute Foundation, and the Friends of Tucson Optics Endowed Scholarships in Optical Sciences.

References


doi:10.1117/1.3569692.


Maham Aftab is a senior design engineer at ASML US. She received her BS in Engineering Sciences from GIK Institute, Pakistan, MS in applied physics from TU Delft, The Netherlands, MS in Optics, Matter and Plasma from École Polytechnique, France, and MS and Ph.D. in Optical Sciences from The University of Arizona. Her current research interests include hardware and software design and analysis; and mathematical study and development for optical metrology systems, particularly for high-resolution and freeform optics.

Dae Wook Kim is an assistant professor of optical sciences and astronomy at the University of Arizona. His research area covers precision optical engineering, including interferometry and deflectometry. He is the chair of the Optical Manufacturing & Testing (SPIE) and Optical Fabrication & Testing (OSA) conferences. He is a senior member of OSA and SPIE and has been serving as an associate editor for Optics Express.

Biographies and photographs for the other authors are not available.
Appendix D. Adaptive Shack-Hartmann Wavefront Sensor accommodating large wavefront variations

Maham Aftab, Heejoo Choi, Rongguang Liang, and Dae Wook Kim


Permission to use this article in the dissertation was obtained from The Optical Society (OSA) on March 05, 2019.
Adaptive Shack-Hartmann wavefront sensor accommodating large wavefront variations

Maham Aftab,1 Heejoo Choi,1 Rongguang Liang,1 and Dae Wook Kim1, 2,*

1College of Optical Sciences, University of Arizona, 1603 E. University Blvd., Tucson, AZ 85721, USA
2Department of Astronomy and Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85719, USA
*letter2dwk@hotmail.com

Abstract: Shack-Hartmann wavefront sensors (SHWFSs) usually have fixed subaperture areas on the detector, in order to fix the minimum and maximum amounts of wavefront departure, or the dynamic range of measurement. We introduce an active approach, named Adaptive Shack Hartmann Wavefront Sensor (A-SHWFS). A-SHWFS is used to reconfigure detection subaperture areas by either blocking or unblocking desired lenslets by using an electronically modulated mask. This mask either increases or decreases the measurable aberration magnitude by placing a liquid crystal display (LCD) panel in front of the lenslet array. Depending on which control signal that is sent to the LCD, the variable, application-dependent blocking pattern (horizontal, vertical, diagonal, uneven) makes this an adaptive and efficient sensor with a variable dynamic range of measurement. This scheme is also useful for regional blocking, which occurs when the wavefront is severely aberrated in a limited region.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

1. Introduction

The Shack-Hartmann wavefront sensor (SHWFS) is a well-known device for measuring a wavefront by determining its local slope distribution, or the first spatial derivative [1]. It is used for a wide range of applications, including astronomy [2], ophthalmology and commercial optical testing, to name a few [1,3,4]. The principle of operation for a SHWFS is that an array of small lenses (called lenslets) samples the wavefront and creates focused small spots of light in the focal plane of the lenslet array. At the nominal (i.e., unaberrated) situation in a collimated beam of light, each spot from each lenslet will be centered at the nominal or reference position on the detector (Fig. 1(a)). If the measured wavefront is aberrated, the position of the focused spots moves according to the magnitude of the local tip-tilt component of aberration (Fig. 1(b)). By measuring the new positions of the spots relative to their reference positions, we can reconstruct the slope distribution of the wavefront.

Fig. 1. Basic operating principle of a conventional Shack-Hartmann Wavefront Sensor: (a) Collimated reference wavefront case and (b) Aberrated wavefront case.

The minimum and maximum amount of aberration or spatial frequency of the aberrated wavefront measurable by a SHWFS depends primarily on the focal length and size of the
lenslets and the detector subapertures. A fixed optimal design of SHWFS can be dictated by a variety of factors, such as atmospheric turbulence, in the case of astronomical applications. Generally, a trade-off needs to be made in terms of resolution and dynamic range of the SHWFS, i.e., if the detection subaperture size is small, the wavefront is sampled more finely and a higher spatial resolution is produced but the maximum amount of aberration measurable for a fixed focal length becomes limited. We define dynamic range as the largest and smallest wavefront slope values that can be measured by the system. A conventional SHWFS has a fixed dynamic range of measurement but several concepts for adaptable/reconfigurable SHWFS systems have been proposed over the years.

One approach to improving the dynamic range of a traditional SHWFS is to track the movement of Hartmann spots along an optical axis by measurements at additional planes between the lenslet array and the detection plane and use a predictive algorithm to match the spots to their correct measurement locations [5]. The method in [5] also makes use of a movable relay lens and camera assembly to switch between Hartmann spot imaging and pupil imaging so that the mapping of the lenslets onto the pupil and hence an optimal size and distribution for the detector subarray can be established. Another novel approach uses an algorithm that unwraps spot dislocations and assigns spots to their correct subapertures, thereby eliminating discontinuities in the patterns of the local positions, which are wrapped modulo P, where P is the lenslet pitch [6]. The technique presented in [7] uses two different measurements, a conventional image of the wavefront and the image from the classical Shack–Hartmann sensor, to estimate a parametrized description of the measured aberration. In [8], an estimate of the positions of the focal spots of neighboring lenslets is proposed by extrapolating an iterative two-dimensional spline function that assigns the spots to their respective reference points.

There are other approaches that use specially designed devices. For example, an astigmatic lenslet array that gives a characteristic mark to each spot, allowing a definite recognition of the spot even if it moves beyond its detection subaperture [9]. Another approach describes a coding algorithm with a minimum number of measurement cycles to allow definitive assignment of spots if a spatial light modulation array, placed in front of the microlens array, is used to switch subapertures on and off [10]. The MEMS technique in [11] is used to improve dynamic range through individual address. This method varies the mechanical resonant frequencies of individual lens-support carriages and identifies the focal spot from a particular lenslet by detecting the line image resulting from the motion of that spot. The sensor proposed in [12] replaces traditional lenslets with a microhologram array which gives a discriminable pattern to each focal spot and employs a pattern matching technique that uses cross correlation between the reconstructed images and template images. Rha et al. [13] use a reconfigurable array of Fresnel lenslets written on a phase-modulated LCD module, Zhao et al. [14] examine the customization of a digital SHWFS that uses a diffractive optical lens pattern, encoded on a spatial light modulator, as the microlens array while Yoon et al. [15] employ a translatable plate with subapertures placed conjugated to the lenslet array. By moving the plate, desired lenslets can be blocked and by taking several measurements where the plate is translated between the measurements, all focal spots can be correctly associated to their respective subapertures.

We propose an adaptive SHWFS (A-SHWFS) with the ability to change the LCD lenslet mask dynamically. This allows the sensor to change the subaperture and distribution of spatial sampling based on the aberration present in the system. While previous works on this topic may have achieved similar objectives, our method provides a simple yet powerful technique to make an adaptable SHWFS system that can be integrated into most existing SHWFS-based setups without a major system overhaul. A similar objective may be achieved by changing system parameters, such as increasing the lenslet size but that would mean making customized-and-fixed hardware (and possibly software) changes to the system such as switching out the lenslet array. Our approach applies a one-time modification to the system.
set-up, after which the adaptive sampling is implemented only through software control. It also does not use any complicated devices or heavy computing. Furthermore, it allows for a highly efficient, even irregular, and adaptive reconfiguration of the detection subapertures. For instance, if the measured wavefront is highly aberrated in a small area relative to the entire wavefront, only a small portion of the detection area can be used for reconfiguration to accommodate for this localized high wavefront slope change. The rest of the detection area can still maintain a high sampling. Also, it provides great flexibility in its implementation, with easy-to-apply changes. For example, the criteria for switching to a larger detection subaperture or the type of blocking / mask applied to areas with high wavefront slope change can be modified by the user per the objectives and implementation details of the application. This paper discusses the idea and implementation of our technique leveraging a fully matured economical solution, electronically modulated LCD panel, and software controls.

2. Adaptive spatial sampling and modal wavefront reconstruction

2.1 Actively modulated lenslet array using a LCD panel

The key to dynamically blocking and unblocking lenslets is by placing a LCD screen (without its back-light illumination unit) in front of the lenslet array as shown in Fig. 2(a). Transmission of light through the LCD can be controlled by addressing an opaque (i.e., black colored) and transparent (i.e., white colored) pattern through the computer connected to the LCD screen. The pixels where the pattern (e.g., boxes, lines, irregular zones, etc.) is black will have no light passing through them and the corresponding lenslets effectively become inactive. By matching up the lines or squares of the blocking pattern on the LCD to the lenslets, we can control each lenslet or groups of lenslets in any desired, adaptable pattern. The details of the matching and alignment process applied to the real prototype system is given in Section 3.2. The areas of the LCD where the addressed pattern is white will let light through and the corresponding lenslet or lenslets become active. Essentially, we are dynamically changing the active pattern of the lenslet array, either over the entire area of the wavefront, or only in certain sections. Then, as we block certain lenslets, the corresponding detector subaperture areas expect no focused spots, which now can be used for the neighboring spots’ extended detector subaperture zones. This enables originally immeasurable highly aberrated wavefronts to be detected without being limited by mixed spots between neighboring lenslets at the expense of spatial sampling resolution as depicted in Fig. 2(b). Of course, this will be a tradeoff between the dynamic range and the spatial resolution of the wavefront mapping and can be actively optimized for a given situation. However, importantly, the detection sensitivity remains the same.

![Fig. 2. (a) Schematic layout of the A-SHWS using actively modulated LCD lenslet array mask. (b) Depiction of how the LCD lenslet array mask (left) and the corresponding detector subapertures (right) are changed dynamically, from a fully unblocked (top) to partially blocked (bottom) situation. The squares on the left represent the pattern sent from the computer to the LCD screen.](image)
2.2 Focused spot centroiding algorithm

Due to scattering from the LCD’s internal structures as well as the diffraction effect from the diamond-turned tool marks on the lenslet array, the focal plane image spots are degraded compared to the standard SHWFS without an LCD mask. Hence, an intensity-weighted centroiding algorithm was applied to determine the statistical centroid of the spots. The following equations are used to calculate image centroids:

\[
\begin{align*}
    x_c &= \frac{\sum y \sum x I(x,y)}{\sum y \sum I(x,y)}, \\
    y_c &= \frac{\sum y \sum x I(x,y)}{\sum y \sum I(x,y)}
\end{align*}
\]  

(1)

where \(x_c\) and \(y_c\) are coordinates of the image centroid and \(I\) is the spot intensity on the detector subaperture.

To improve the process of determining the correct spot locations, the following technique was applied. The algorithm finds the brightest spot in each subaperture region. This is implemented in our code by scanning the subaperture region, applying MATLAB’s inbuilt findpeaks function [16] and comparing the values in the region to find the location of the brightest peak in that subaperture. In order to mitigate the effect from the noise background, a thresholding, which is specific to the as-built system, is applied to the findpeaks function. Once the peak location is detected, certain area around the peak position is selected and the weighted centroiding algorithm (Eq. (1) is applied over this area of interest. This positional information is utilized to calculate the differential spot motion indicating the local wavefront slope change.

The centroiding methodology used in this work is generic but sufficient for our applications. Many techniques, found in existing literature, may be applied to the data processing pipeline in this method and can further improve the performance of the sensor. In the meantime, the data reconstruction process, described in Section 2.3, is a novel approach and can improve or enhance the wavefront reconstruction process in many situations.

2.3 Modal wavefront reconstruction

Once the local slope data is obtained (i.e., \(x\) and \(y\) slope distribution), it can be integrated to reconstruct the wavefront. Measured slope data is processed by a modal reconstruction algorithm, based on the newly developed gradient polynomials [17], called the \(\tilde{G}\) polynomial set. These polynomials are obtained from the gradients of two-dimensional Chebyshev polynomials as shown in Eq. (2). The scalar and vector polynomial sets are both orthogonal across a rectangular aperture, which optimally matches the format of the A-SHWFS detector.

\[
\tilde{G}_m^n (x,y) = \nabla F^n_m (x,y) = \frac{\partial}{\partial x} F^n_m (x,y) \hat{i} + \frac{\partial}{\partial y} F^n_m (x,y) \hat{j}
\]  

(2)

The scalar \((F)\) polynomial set is a two-dimensional Chebyshev basis set, constructed from two one-dimensional Chebyshev polynomials of the first kind, as shown in Eq. (3).

\[
\begin{align*}
    F^n_m (x,y) &= T_n(x)T_m(y), \\
    T_{m+1}^n (x) &= 2xT_m^n(x) - T_{m-1}^n(x) \quad \text{where} \quad T_0^0 (x) = 1, T_1^0 (x) = x, \quad \text{for} \quad -1 \leq x \leq 1, \\
    T_{n+1}^m (y) &= 2yT_n^m(y) - T_{n-1}^m(y) \quad \text{where} \quad T_0^0 (y) = 1, T_1^0 (y) = y, \quad \text{for} \quad -1 \leq y \leq 1
\end{align*}
\]  

(3)

For Eq. (2) and Eq. (3), \(\tilde{G}_m^n\) are the gradient polynomials, \(F^n_m\) are the scalar polynomials, and \(T_n^m\) are the one-dimensional Chebyshev polynomial sets used for the construction of the scalar basis. The double index variables \(n\) and \(m\) are related to the order of the
polynomials [17] while \( \hat{i} \) and \( \hat{j} \) are unit vectors representing the axes of a Cartesian coordinate system. Figure 3 shows the quiver plots for the first three non-trivial \( \bar{G} \) polynomials.

![Quiver plots for three low-order \( \bar{G} \) polynomials in a normalized rectangular domain.](image)

The virtue of this modal vector-based methodology is to fit the data in the measurement (i.e., slope) domain and directly obtain the gradient polynomial coefficients. These coefficients are used to obtain the coefficients of the scalar polynomials, which can then be used along with the scalar polynomial basis set to acquire the reconstructed wavefront. The scalar and vector polynomial coefficients have a one-to-one relationship. An attractive feature of this reconstruction methodology for the A-SHWFS is that it can easily deal with different kinds of non-uniform samplings. As introduced in Section 2.1 and expanded with an example in Section 4.2, the A-SHWFS can actively and efficiently vary the sampling distribution of the wavefront. The modal reconstruction method used here can handle all these situations and other variations of samplings effectively [18].

The \( \bar{G} \) polynomial modal set allows for efficient and accurate generation of up to tens of thousands of polynomial terms, employing recursive relationships for both the scalar and vector polynomial basis sets. This method gives a lower error compared to several traditional zonal and modal methods when the slope of the wavefront changes sharply or when the aperture is blocked in certain regions. Several examples of comparisons between this modal method and a traditional Southwell zonal method have been reported [17], for surface reconstructions from simulated and real data. Also included is a comparison with Zernike gradient polynomial fitting [17]. From these examples, it is seen that the \( \bar{G} \) polynomials perform better (e.g., in terms of accuracy), compared to the zonal or Zernike gradient polynomial methods for various cases. As a comparative reference, Neal et al. [19] provided a detailed account of data fitting and reconstruction methods for a conventional SHWFS, as well as an in-depth investigative result of SHWFS precision and accuracy.

### 3. A-SHWFS system prototype

To demonstrate the concept and quantitatively investigate different aspects of the A-SHWFS system, a prototype was designed and built using an off-the-shelf LCD panel. This system also includes a combination of customized components such as a diamond-turned lenslet array.

#### 3.1 Overall system layout and configuration

Light from a laser source goes through a spatial filter and a lens is used for collimating the beam, which sets the reference wavefront. Then, light is reflected off the three hexagonal segmented flat mirrors system, which is used for introducing and controlling systematic wavefront aberrations in the system. Two of the three mirrors are kept stationary while the third mirror’s tilt actuator is motorized as discussed in Section 4.1. The moving mirror has a
precision actuator mounted on it. This mirror provides the section of wavefront that creates a large, localized (relative to the entire beam reflected from the three mirrors) tilt, which will be used to demonstrate the adaptive sampling capability of the sensor. The other two mirrors are both kept stationary and together they provide the section of wavefront that will not change and hence the detection area corresponding to light reflected from these two mirrors will show no change in sampling. The actuator allows the mirror to be moved in a certain direction. By connecting it to an actuator controller, we can electronically control the amount the actuator moves, which allows for a reliable and precise motion of the mirror (in this case, it allows the mirror to be tilted). The collimating lens acts as the system stop, so the beam hitting the detector has a diameter of ~50.8 mm. The beam overfills the detector and ensures that there is enough light to fill the detector even when the hexagonal mirror is tilted considerably. The segmented mirror configuration was specially designed and built in order to represent and examine a locally varying wavefront with large magnitude variations, which is the key target situation of the A-SHWFS technology. Figure 4 is the prototype set-up and Table 1 summarizes the specifications of all non-trivial system components in order to provide a retrace-ability of the presented system performance.

In principal, the A-SHWFS works using the relative change of the measured wavefront from a reference state. To obtain the reference state, the sensor can be calibrated using a known wavefront input, such as a collimated beam. Unlike a typical SHWFS, the A-SHWFS can accommodate large wavefront departures, so the reference wavefront can even be a powered wave. In our case, the reference (or calibration) wavefront was a collimated plane wave and the initial spots were recorded based on the measurement of this reference wavefront. This helps account for the residual spatial error introduced by the LCD. In practice, the A-SHWFS should first be calibrated using the target (or reference) wavefront and the reference spot positions recorded. Then, the sensor reports the wavefront deviation from the calibrated reference wavefront, as a slope measurement.

Table 1. Key components of the A-SHWFS prototype system

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
<th>Model Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collimating lens</td>
<td>Focal length: 200 mm, Diameter: 50.8 mm</td>
<td></td>
</tr>
<tr>
<td>LCD panel</td>
<td>Resolution: 800 × 480, Screen size: 152.5 × 91.3 mm, Pixel pitch: 0.19 × 0.19 mm</td>
<td></td>
</tr>
</tbody>
</table>
3.2 Customized LCD matching lenslet array

Design of the lenslet array was determined by the dimensions of the LCD panel and the CMOS sensor. The lenslet array size is 40 × 40 mm and contains 35 × 35 lenslets (with missing corner regions). This ensures that the lenslet array is slightly larger than the detection area, so as to make use of the full detector size. The active area of the lenslet array, also consistent with the measured data in Section 4, is ~32 × 24 mm, corresponding to the CMOS sensor dimensions and consists of 29 × 21 lenslets. This results in a detector subaperture size of ~1.1 × 1.1 mm. Each lenslet was designed to cover 6 × 6 pixels of the LCD and have a focal length of 20 mm. The native dynamic range of a single A-SHWFS lenslet in the x and y directions is defined as

\[ \pm \tan^{-1} \left( \frac{3 \times 0.19 \text{[mm/pixel]}}{20 \text{[mm]}} \right) = \pm 28.5 \text{ mrad} \]  

(4)

The custom designed and diamond turned lenslet array is shown in Fig. 5. The lenslet array was made on a Poly(methyl methacrylate), also known as PMMA, substrate by using single point diamond turning (SPDT) method. The periodic tool marks due to the SPDT process cause diffraction orders at the focal image plane, so the centroiding algorithm with a thresholding (Section 2.2) is applied to recognize and process the zero order spots. The tool mark effect can be minimized by controlling and optimizing the diamond turning process parameters, but we have used the as-manufactured lenslet array for this prototype system development.

![Fig. 5. The custom designed and diamond turned lenslet array of dimensions 40 × 40 mm and containing 35 × 35 lenslets (with missing corner regions) using PMMA, used for the A-SHWFS prototype system.](image)

The commercial LCD’s pixel size / pitch and the in-house manufactured lenslet array’s lenslet size / pitch were both matched and verified using microscopes. After a coarse alignment between the LCD screen and lenslet array, the LCD / lenslet array placement was fine-tuned and adjusted using the measured output on the CMOS detector. The blocking pattern matching was also done by checking the measured signal from the CMOS detector after the pattern was applied and comparing it to the reference (i.e., no blocking) situation.
4. Experimental performance verification

4.1 Wavefront slope measurement accuracy

The experiment described in this section aims to quantify the accuracy of the A-SHWFS by comparing its measurement values against those obtained by a commercial precision autocollimator, which directly measures the mirror surface slope change. The MÖLLER-WEDEL ELCOMAT 3000 electrical autocollimator was used, which has a superb accuracy and resolution of $\pm 1.21 \times 10^{-3}$ mrad and $2 \times 10^{-4}$ mrad respectively. The maximum possible measurement range of this autocollimator is 9.70 mrad.

The wavefront measurement area was carefully chosen as the boundary region of the three hexagonal mirrors, so that two remain stationary while one is systematically moved to introduce tilt in the reflected beam wavefront as shown in Fig. 6(a). Also, a small mirror was attached to the backside of the moving hexagonal mirror. Light from the autocollimator reflects off the small mirror and goes back to the autocollimator which measures the orientation of the mirror with respect to its optical axis (i.e., reference) as depicted in Fig. 6(b). As the active hexagonal mirror is tilted using the Piezo inertia actuator, the autocollimator measures the co-mounted small mirror’s orientation which provides the golden standard values for the A-SHWFS test. (Note: The factor of 2 between the direct surface orientation change and the reflected beam’s wavefront slope change due to the double-path was considered and accounted for in this comparison.)

The A-SHWFS measures the tilt by calculating centroids of the focused light spots from the moving mirror (as described in Section 2.2) and modally fitting the slope data using the $\hat{G}$ vector polynomials. This experiment was done for two cases: one with small actuator motions corresponding to small changes in the beam’s tilt (Fig. 7(a)) and the second with a larger actuator motion corresponding to large tilt angle range (Fig. 7(b)). Results, including the error are summarized in Fig. 7. The percentage error between the two sets of measurements was calculated using Eq. (5) and then averaged.

\[
\text{% Error} = \frac{\text{Autocollimator value} - \text{A-SHWFS value}}{\text{Autocollimator value}} \times 100\% \tag{5}
\]

The final results from the A-SHWFS show agreement of 1.32% average error for the small range case and 1.13% error for the large range case compared to the autocollimator values, which confirms the fidelity of overall data processing pipeline and the actual prototype system performance.
Fig. 7. (a) Small, and (b) Large magnitude of tilt measurements comparing the autocollimator and A-SHWFS values (blue circles). The error (red crosses), which is the difference between the autocollimator and A-SHWFS values, is shown with its own axis on the right side in both figures. A linear fit to the data is shown as the black line.

4.2 Adaptive wavefront sampling

One of the biggest applications of the A-SHWFS is reliably measuring wavefronts that have high regional aberrations or sharp local slope change (i.e., not over the entire wavefront, but only in a certain region). Since the subaperture blocking using LCD panel does not have to be applied over the entire wavefront sampling area, this enables a unique tradeoff between dynamic range and spatial resolution over a selective localized area where the wavefront is aberrated beyond its nominal detection dynamic range.

To demonstrate this adaptive capability, the three hexagonal mirrors set-up was employed (Fig. 6(a)) in a way that two mirrors remained stationary while the third one was moved by the Piezo inertia actuator continuously. Portion of the beam reflected off the moving mirror had increasing tilt (i.e., localized excessive tilt) as the actuator continued tilting the active hexagonal mirror. The rest of the beam (reflected off the two stationary mirrors) had almost no or a fixed, small amount of tilt as a reference.

When the amount of tilt is small, the entire detection area has the nominal and uniform spatial sampling. This sampling is determined by the total number of lenslets and corresponds to the smallest possible detector subaperture area for this particular lenslet array. This situation corresponds to the image in Fig. 8(a). It is also represented in Fig. 9(a), where the length of the quiver plot arrows (corresponding to the amount of wavefront slope) is small. As the active mirror’s tilt increases to the point where it can no longer be measured by this small subaperture area on the detector, the blocking LCD mask pattern is activated and the detector subaperture area increases (as described in Section 2.1) to adapt to this steeper wavefront to be within the measurable dynamic range. However, since the steep wavefront is only over a portion of the beam reflecting off the active hexagonal mirror, only the localized section of the detection area adapts its spatial sampling by controlling the matching LCD panel’s blocking on / off signals.

The result is selectively optimized detector subapertures as shown in Fig. 9(b). The higher spatial sampling was achieved over the small dynamic range portion of the wavefront from the stationary mirror zones. In contrast, the lower spatial sampling with enhanced dynamic range was applied in order to monitor the wavefront from the highly tilted active mirror. The measured wavefront slope data in Fig. 9(b) clearly highlights this adaptive concept and capability, where the quiver arrows are much longer in the localized region corresponding to the high tilt active mirror.
Fig. 8. Spots from the lenslets, on the CMOS detector. Red arrows show the spot motion relative to reference spots, on the active mirror portion of the image. (a) For small tilt, all three mirrors display the maximum number of spots (i.e., higher spatial sampling of the wavefront). The red arrows are shorter. (b) When the amount of tilt increases, the stationary mirror zones maintain the same number of spots while spots from the actively moving mirror have been selectively blocked by the LCD panel to increase the slope measurement dynamic range only in the optimal zone. Hence, the red arrows are longer.

Fig. 9. Quiver plot arrows for centroids taken (a) before and (b) after the LCD adaptive gating starts. The actively tilted segment part on the left side shows longer arrows while the other two stationary segments on the right side shows very short or no arrows. The area enclosed by the red dotted line corresponds to the zoomed-in regions shown in Fig. 10.

Various image captures were extracted and investigated at regular time intervals during the adaptive wavefront sensing experiment, with the active test mirror segment tilting (i.e., adding tilt to the reflected beam wavefront) continuously so that the magnitude of tilt at any instant of time is proportional to the timestamp of that frame. Figure 10 displays the \( \tilde{G} \) polynomial-based wavefront reconstructions from several of these zoomed-in frames (corresponding to the red dotted regions of Fig. 9). It successfully confirms the adaptable wavefront measurement capability accommodating to the change of the wavefront slope magnitude. As demonstrated in the sequential tilt change, wavefront slope that started within the nominal dynamic range of 28.5 mrad from Eq. (4) goes well beyond the limit, up to 53.1 mrad (the last wavefront map on the right in Fig. 10) via the extended detector subapertures.
Fig. 10. Enhanced dynamic range demonstration of the adaptive wavefront sensing approach. The reconstructed wavefront time-lapse (zoomed-in portion of data corresponding to red dotted areas in Fig. 9) shows continuously increasing wavefront tilt as the active mirror was being tilted up-to and beyond the nominal 28.5 mrad dynamic range. The average slope magnitude for each map is shown as well. The dotted red line represents the start of adaptive gating i.e., when the blocking mask was applied to the LCD.

Here, only a few screenshots are shown but this is one way of measuring aberration using the A-SHWFS where either images can be taken at different intervals (depending on the required frequency of measurement) and these images can be processed in nearly real-time or a video can be recorded once and then screenshots can be extracted at any interval required for a post-analysis need. The frequency of image processing depends primarily on the computational resources connected to the sensor. It also depends on the amount of data obtained through the detector.

4.3 Simulation study for A-SHWFS

As mentioned previously, the A-SHWFS concept allows the sensor to be very efficient and highly customizable. It can measure various types of wavefronts, under different circumstances. To show the possibility and capability of the sensor to work for other situations, a simulation study was performed, the details of which are described in this section. Random wavefronts are generated using combinations of various low and higher order Zernike terms in combination with pupil masks. A detection area corresponding to 22 × 22 lenslets is used for this study.

The adaptive masking algorithm is applied in the following manner: The size of each detection subaperture is 130 × 130 pixels. If the absolute wavefront slope measurement (i.e., the absolute difference between the measured and reference spot locations) at any location exceeds 80% of the native detection subaperture half-width, the algorithm automatically switches to a lower sampling resolution but larger detection subaperture area. This new subaperture area corresponds to 3 × 3 lenslets (or 390 × 390 detector pixels). The measurement location (“unblocked lenslet”) is now the central lenslet and 8 lenslets around it are blocked. If the next measurement keeps the higher magnitude of wavefront slope, the subaperture stays at the larger size. If, however, the magnitude of aberration at that location is reduced with the next measurement so that the measured slope is within the 80% threshold, the sensor switches that location back to the native sampling subaperture size (i.e., 130 × 130 detector pixels).

Visualization 1 is a video showing the adaptable detection process, as various wavefronts with low and high local slope changes are applied in the simulation. The images are quiver plots, where the length of the arrows corresponds to the magnitude of the simulated wavefront
slope using A-SHWFS concept. Figure 11 shows four snapshots from the video, with different wavefronts being simulated.

Fig. 11. Quiver plots of simulated data, representing measurements (Visualization 1) from different wavefronts. The blue arrows are measurements from native detection subapertures while the red arrows are measurements from optimized, larger detection subapertures, where regional A-SHWFS blocking was applied. (Note: a magnification factor of 4 is applied to the length of all arrows, to make them easier to see.)

Details and implementation of the blocking algorithm can be modified and customized according to the situation and specific applications. For example, the switching criterion can be changed to 90% or 75% etc. of the native subaperture dynamic range. The user may also choose to apply some completely different criteria, such as applying blocking when multiple spots are detected. It can be iterated to make the subaperture larger and larger till no multiple spots are detected in a single detection subaperture. Other modifications can include a different size of the larger subaperture as soon as blocking is applied, e.g., instead of switching to a $3 \times 3$ lenslet subaperture, one may switch to $2 \times 2$, $5 \times 5$, or even $2 \times 3$ lenslets.

5. Summary

The Adaptive Shack-Hartmann Wavefront Sensor introduced in this work was designed to have an adaptable dynamic range of wavefront measurement accommodating dynamic local (or global) wavefront changes at the expense of spatial sampling, as necessary. This concept does not require a mechanical motion or other such mechanisms of the lenslet array to deliver this capability. No complex optics or heavy computation is necessary either. With only an
electrically controlled LCD panel, it is able to dynamically change the detection subaperture distribution on the detector. This paper presents and demonstrates the adaptive wavefront sensing concept, application and basic implementation as an actually built system.

We discussed the details of hardware and software implementation of the prototyped A-SHWFS so the concept can be understood, verified, and improved upon in future modifications and cross-checks. Additionally, by taking measurements from the sensor and simultaneously comparing them against a standard testing method, we prove the merit of this method, as well as make an estimate of its accuracy. Finally, measurement for a locally-and-highly aberrated dynamic wavefront was simulated using a three hexagonal mirrors segment and tested by the A-SHWFS to provide a realistic case example of some of the unique capabilities and advantages of this system. However, we acknowledge that some of the avenues of improvement include LCD and SPDT lenslet array with lesser scattering, better matching between pixels of the LCD and lenslet array, and better calibrations. If scattering or diffraction is a problem for certain applications, a customized micro-shutter system replacing the LCD may be considered in the future. Also, computational power for quicker and more efficient data processing will be a critical factor for some high-speed application cases requiring kHz-level operations.

The A-SHWFS concept showed a unique strength and its versatile applications to meet adaptive wavefront sensing and analysis needs. The adaptability ensures that while a more and more aberrated wavefront can be reliably measured, the system does not have to sacrifice with low spatial resolution when the aberration magnitude is not strong. This is of particular importance for the case where a wavefront is highly aberrated locally or in a certain region only. The new tradeoff space enabled by this solution can benefit future optical systems requiring multi-purpose adaptive functionalities.

Acknowledgments

This material is partly based on work performed for the “Post-processing of Freeform Optics” project, which is supported by the Korea Basic Science Institute. The modal reconstruction related algorithm and software development is partially funded by the II-VI Foundation Block grant. Also, we thank Sukmock Lee for his help in the laser source creation for the experimental setup.

References