Distributed Wavefront Reconstruction for Adaptive Optics Systems

João Lopes e Silva

Master of Science Thesis

Delft University of Technology

Delft Center for Systems and Control

Distributed Wavefront Reconstruction for Adaptive Optics Systems

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

João Lopes e Silva

August 4, 2014

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

Abstract

We are currently facing an increasing amount of challenges in the area of photonics as more and more applications in need for active "photon control" sprout in different fields of science. Adaptive Optics (AO) is the subject which deals with measuring, reconstructing, and reshaping the phase of a photon wavefront in real-time and can, thus, provide the framework for controlling the photons in areas such as medicine, astronomy and telecommunications, among others.

The objective of this graduation project is to create a novel distributed method for wavefront reconstruction, integrate the method in an AO loop, and analyse its properties. This method will use the intensity distribution measured by the wavefront sensor instead of the classical slope approximation (obtained using a centroid algorithm). Using the complete intensity distribution gives us more information than the slope approximation and therefore, a more accurate reconstruction is expected. Moreover, we will estimate the wavefront using B-splines basis functions. These splines are defined locally which makes them suitable for the application of distributed reconstruction methods.

The content of this thesis is divided into two distinct parts. In the first part, we analyse the different components of an AO system with special emphasis on the state-of-the-art phase retrieval methods. Furthermore, the B-splines framework is presented alongside distributed optimization techniques with special emphasis on the Alternating Direction Method of Multipliers (ADMM). The second part of the thesis uses the theoretical information from the first chapters to support the development of one centralized and two distributed algorithms for solving phase-retrieval problems using pupil-plane sensors. The results from these methods, together with the results from a compressive sampling method which decreases the quantity of measurements used, are presented in the last chapters.

It was verified in simulation experiments that the average reconstruction error achieved by the novel centralized method surpasses the classical approaches which use slope measurements for aberrations with an RMS value smaller than λ. It is also shown that the variance of the reconstruction error using the novel method is reduced by two orders of magnitude. Regarding the two distributed methods (unstructured and structured ADMM applications), it is shown that the unstructured method has a very low convergence rate which renders this method

unpractical for real-time applications. The structured method showed much more promising results given that it was able to converge to within a 5% tolerance of the optimal centralized solution after $50 - 150$ iterations. This method can also be implemented in a completely decentralized manner which is suitable for a GPU/FPGA implementation.

Contents

List of Figures

List of Tables

Preface

This graduation project appeared as a natural continuation of a small project developed in the last quarters of the academic year of 2012/2013 under the supervision of Prof. dr. ir. Michel Verhaegen. The main purpose of that project was to develop a more accurate and distributable algorithm to reconstruct the phase distribution.

When the project was concluded both Prof. Verhaegen and myself showed interest in continuing the work already initiated. My motivation was two-fold: on the one hand, I enjoyed working in a multidisciplinary environment on a subject which was not very familiar to me and embedded in a team which was always both demanding and supportive. On the other hand, it would give me the opportunity of working on distributed optimization and signal processing, which are two subjects of great interest to me.

The most important technical acknowledgement is made to Prof. Verhaegen, for his vision in coming up with this project, for the motivation and excitement that he transmits, and for providing me with opportunities for learning, presenting, and teaching which would otherwise not be available for me. I would also like to thank Elisabeth Brunner for her constant support and willingness to discuss my problems. A special note of thanks for Dr. Tamás Keviczky, whose contributions were crucial in coming up with distributed solutions, and whose insight regarding possible career choices was truly enlightening. Furthermore, I would like to thank Dr. Alessandro Polo who has had the most extraordinary patience in explaining me the basic (albeit crucial) optics concepts.

Moreover, I would like to thank to all my friends with whom I revived fond memories of Portugal; with whom I travelled; with whom I shared an office; with whom I shared a house; with whom I shared a meal; with whom I shared laughter and joy; and with whom I always felt at home even when living 1200 km away from Lisbon.

Finally, I would like to deeply thank my family who have supported me [in all possible manners and] in all my endeavours and motivated me to aim higher.

An additional technical note is required: if any typo or incorrect information is detected please contact me via `jpedro.e.silva@gmail.com`.

Chapter 1

Introduction

We are currently facing an increasingly amount of challenges in the area of photonics as more and more applications in need for active “photon control” sprout in different fields of science. In this thesis, the main focus is on the field of Adaptive Optics (AO) which involves measuring, reconstructing and reshaping the phase of a photon wavefront in real-time. More specifically, the measurement phase involves a wavefront sensor which provides information regarding the wavefront. Then, using that data the phase of the wavefront is retrieved. Finally, an actuated device, such as a deformable mirror, is used to reshape the phase of the wavefront in real-time.

AO was first developed for military purposes. Using segmented mirrors to compensate for the effects of the atmosphere, scientists were able to obtain a larger signal intensity from laser beams that were used for telecommunication purposes [6]. Nowadays, there are a large number of real-time applications in the field of AO.

In astronomy, AO is used to counteract the effect of the turbulence in the atmosphere which blurs the images that arrive at ground-based telescopes. When it was applied to these telescopes (*e.g*, in Mauna Kea, Hawaii [7]) for the first time, the quality of the images rivalled that of the Hubble Space Telescope [6]. The current efforts are mainly directed to enhance the performance of the next-generation of large-scale ground-based telescopes. One of those telescopes is the European Extremely Large Telescope (E-ELT). This telescope will, by construction, have a very large collecting area which will, by itself, acquire a larger amount of astronomical data [8]. However, such vast amounts of data are only beneficial if an AO system is included in the telescope. Therefore, an effort has been made to integrate AO from the beginning of the design plans, which can lead to an improvement of one order of magnitude in the spatial resolution compared to the state-of-the art telescopes [9].

Adaptive optics also has its uses for Integrated Circuit manufacturing. Extreme Ultraviolet Lithography (EUVL), which is a promising next-generation method of circuit printing [10], must have a very high precision such that the doping of the semiconductor is done correctly. AO could be used to correct deformations that exist in the mirrors that perform part of the circuit printing which are very susceptible to heat deformation. Although there are, to the best of the authors' knowledge, no articles with a complete AO setup for this application, some authors have shown some promising results regarding the phase retrieval stage [11, 12, 13].

In confocal and multi-photon microscopes, which are able to produce three-dimensional images of volumetric objects, the resolution of the images is severely affected by the changes in the refractive index of the object in question. These aberrations can be minimized by using AO [14, 15] as the cause of the aberrations is similar to the one in astronomy.

In retinal imaging, it is worth mentioning two methods that have immensely improved the way scientists and physicians perceive the eye. Those methods are Scanning Laser Ophtalmoscopy and spectral domain Optical Coherence Tomography. AO has the same use as in microscopy given that the eye also has changes in its refractive index which jeopardizes getting a clear image [16, 17].

Given these numerous applications, it is clear that AO has a widespread use and plays an important role in many applications. In the context of this thesis, we will focus on improving a specific part of the AO loop: wavefront reconstruction (also known as phase reconstruction or phase retrieval methods). Developing more accurate phase-retrieval methods would greatly enhance the performance of the previously stated applications.

Besides accuracy, the speed at which these reconstruction algorithms run is crucial, specially for large-scale AO systems. Hence, we will try to take both accuracy and speed into account by developing distributed methods with good scalability properties in terms of computational complexity.

Taking the aforementioned premises into consideration, the goal of the thesis can now be presented.

1-1 Goals of the Thesis

The main objective of the thesis is to provide alternative methods for wavefront reconstruction, yielding better results than the standard slope-based methods and scalable to large-dimensional systems. This boils down to answer one question which is the fundamental problem of this thesis:

"Is there a method that allows us to achieve better wavefront reconstruction results (in terms of reconstruction error and its variance) and thus outperform the standard methods, and whose computational effort can be distributed?"

1-2 Research Approach

Based on the main goal of the thesis, the research approach can be divided into two parts containing several topics that must be addressed. The first part concerns the creation of a novel centralized wavefront reconstruction method that reconstructs the wavefront with high-accuracy. The second part is related to the way this method can be effectively distributed to make it suitable for real-time applications.

Regarding the centralized method, we predicted that the following topics needed to be addressed:

1. an estimator based on focal-plane information (intensity measurements) will be developed so as not to lose information when computing the wavefront slopes;
2. the estimator will be formulated in the B-splines framework to ensure locality and distributability of the model;
3. the performance in both open- and closed-loop of the new estimator needs to be compared with that of a standard slope measurement wavefront reconstruction algorithm;
4. the statistical properties of the estimators will be analysed;
5. the influence of the wavefront sensor characteristics, the splines parameters, and the simulation parameters must be analysed;

Concerning the creation of a distributed method, there are a number of issues that have to be tackled:

1. a distributed optimization method using only local measurements should be developed to solve the centralized problem in an iterative way;
2. the accuracy in both open- and closed-loop of the new estimator needs to be compared with that of the centralized method, and that of a standard slope measurement phase-retrieval;
3. the computational complexity and memory usage ought to be analysed in comparison to the centralized method.
4. the influence of the distributed optimization parameters has to be analysed;

1-3 Contributions

The research done during this thesis gave rise to peer-reviewed conference proceedings and possibly a journal publication together with a MATLAB implementation of all the methods developed and a user manual for documentation purposes. The publications already accepted for the proceedings are the following:

- J. Silva, E. Brunner, A. Polo, C. de Visser, and M. Verhaegen, “Wavefront Reconstruction Using Intensity Measurements for Real-time Adaptive Optics," in *European Control Conference '14*, 2014.
- E. Brunner, J. Silva, M. Verhaegen, and C. de Visser, “Compressive Sampling in Intensity Based Control for adaptive optics," in *IFAC World Congress '14*, 2014.

The following publication is still in preparation but its title and authors can be already disclosed:

- J. Silva, T. Keviczky, M. Verhaegen, E. Brunner, “Distributed Wavefront Reconstruction Using Intensity Measurements”, *to appear.*

The code is written in MATLAB and will be made public in `github.com/faitdivers` (provided that all the libraries used in the implementation can be made public) together with a user manual. The main features of the package are:

- Stand-alone Hartmann and Shack-Hartmann wavefront sensor simulator;
- Centralized intensity-based wavefront reconstruction method;
- Distributed versions of the centralized methods;
- Compressive sampling feature for the centralized method;
- Closed-loop implementation capable of:
  - arbitrary number of delay samples;
  - switching between perfect mirrors and linearised mirror models;
  - switching between the different wavefront reconstruction method;

1-4 Outline

The theoretical background of this thesis consists of three main chapters that can be read independently and that provide the theoretical support to the results presented afterwards. In Chapter 2, a small introduction to AO is presented, where each stage of the control loop is given individual in-depth treatment. In Chapter 3, the B-splines framework will be described. The B-splines will be used to parametrize the wavefront locally and to ensure smoothness of the global function up to a certain degree, by connecting the local pieces. In the last of the three introductory chapters, Chapter 4, a brief outline of some significant properties of a distributed optimization method called the Alternating Direction Method of Multipliers are presented.

After the first three background chapters, the results from the new methods are presented in the following four chapters where we will try to support our derivations and conclusions based on the background material. In Chapter 5, the fundamentals of the intensity-based wavefront reconstruction in the B-splines framework is presented. All the parameters that affect the phase-retrieval method are analysed extensively. The results are compared to a standard slope-based and modal wavefront reconstruction using Zernike basis polynomials. A compressive sampling is presented in Chapter 6 which allows a significant reduction in the complexity of the centralized method without significantly jeopardizing its performance. In Chapter 7, a distributed method based on ADMM, that does not take advantage of the underlying structure of the problem, is presented and compared to the centralized method. The last chapter of the results, Chapter 8, contains the description and results obtained via a distributed method that exploits the structure of the problem. This structured method is compared to the ones presented before. Finally, some conclusions and final remarks are presented in Chapter 9.

1-5 Nomenclature

Row or column vectors are represented using boldface lower-case symbols such as $\mathbf{x}$. Boldface and upper-case symbols, such as $\mathbf{A}$, are used for matrices. Regular font, x_1, denotes a scalar variable.

The symbol $(\cdot)^*$ denotes the complex conjugate number of $(\cdot)$; it can also be used to represent the optimal result from an optimization problem. The symbol $(\cdot)^\top$ represents the transpose of $(\cdot)$. The hat symbol, $\hat{(\cdot)}$, is used to denote the estimate of $(\cdot)$.

The operator $\langle\cdot\rangle$ denotes the expected value of a random variable. The operator $\mathscr{F}$ denotes the Fourier transform. To denote the nullspace of a matrix $\mathbf{M}$ we will use the operator Null($\mathbf{M}$). The operator ∇ will be used to represent the gradient, the divergence and the curl operations. Given a certain scalar function $f : \mathbb{R}^N \rightarrow \mathbb{R}$, its gradient can be expressed as $\nabla f(\mathbf{x}) = [\partial f/\partial x_1, \cdots, \partial f/\partial x_N]$. The divergence of a certain vector field $\mathbf{f} = [f_x; f_y; f_z$ is computed as $\nabla \cdot \mathbf{f} = \partial f_x/x + \partial f_y/y + \partial f_z/z$. The curl of the same vector field $\mathbf{f}$ is given by $\nabla \times \mathbf{f} = [(\partial f_z/\partial y - \partial f_y/\partial z), (\partial f_x/\partial z - \partial f_z/\partial x), (\partial f_y/\partial x - \partial f_x/\partial y)]$.

Some matrices are characterized with the acronym (S)PD, which means that the matrix is (semi-)positive definite. To simplify notation, the following inequality may be used, $\mathbf{M} \geq 0$, to denote that $\mathbf{M}$ is positive definite. A symmetric matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ is said to be (semi-)positive definite if $\mathbf{z}^T\mathbf{M}\mathbf{z}(\geq) > 0$ for any non-zero vector $\mathbf{z} \in \mathbb{R}^{n\times n}$.

The pseudo-inverse of a tall matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ $(m > n)$ is denoted as $\mathbf{A}^+ \in \mathbb{R}^{n\times m}$, where $\mathbf{A}^+ = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}^\top$. For a fat matrix $(n > m)$ the pseudo-inverse is given by $\mathbf{A}^+ = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$.

The norm used during this thesis for both vectors and matrices will be, by default, the 2-norm, which is denoted by $||\cdot||_2$. For a vector $\mathbf{x} \in \mathbb{R}^N$ the 2-norm is defined as $||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + ... + x_N^2}$. For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, the 2-norm is defined as $||\mathbf{A}||_2 = \sqrt{\lambda_{\max}(\mathbf{A}^\top\mathbf{A})} = \sigma_{\max}(\mathbf{A})$, where $\lambda_{\max}(\cdot)$ represents the maximum eigenvalue and $\sigma_{\max}(\cdot)$ the maximum singular value.

The condition number of a certain matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ can be denoted as $\text{cond}(\mathbf{A}) = \sigma_{\max}(\mathbf{A})/\sigma_{\min}(\mathbf{A})$, where $\sigma_{\min}(\cdot)$ represents the minimum singular value. If the condition number of a matrix is close to one it implies that the matrix is well-conditioned and its inverse can be computed with good accuracy [18, § 8.2].

In order to assess the complexity and execution time of the algorithms the big-O notation will be used. For instance, if an algorithm would have quadratic complexity with respect to a variable b this would be denoted by $\mathcal{O}(b^2)$. This means that the algorithm will scale quadratically with respect to b: if the algorithm needs 100 floating-point operations (FLOPs) when $b = 1$, then for $b = 2$ the algorithm would require 400 FLOPs. Note that a single floating-point operation is defined as a multiplication followed by an addition. For more information regarding the complexity of linear algebra operations we refer the reader to [19, Lecture 5] and [20, Appendix C].

The symbols C^0, C^k, C^∞, etc., will be used to denote the continuity properties of a function. If a function is said to be of class C^0, the function is continuous. If the function belongs to class C^k, all its derivatives until the $k-$th order are continuous. A function of class C^∞ (or smooth) if it has continuous derivatives of all orders.

The calligraphic notation, $\mathcal{T}$ or $\mathcal{N}$, can represent sets or probability density functions. The tilde symbol $\sim$ is used to denote the probability density function of a random variable.

Chapter 2

Adaptive Optics: An overview

High resolution imaging with telescopes, microscopes or lithography machines is often hampered by the presence of aberrations in the wavefront. Such aberrations are induced in various ways. For example, in astronomical observations with ground based telescopes, the aberrations are due to atmospheric turbulence, temperature gradients, etc. Adaptive Optics (AO) was proposed more than half a century ago by Babcock [21] and Linnik [22] independently, and is now more and more being used to correct these aberrations in real-time. An AO system consists of a sensor measuring information from which the wavefront aberrations can be reconstructed and an actuator to correct these aberrations. For classical real-time AO control, where the bandwidth of the feedback controller is far below the first resonance frequency of the deformable mirror, the key problem is the reconstruction of the wavefront, which will be investigated the most in this thesis.

In Section 2-1, the standard AO control problem is presented in an astronomical context. The chapter proceeds with Section 2-2, where the inner workings of the wavefront sensors are explained with a special emphasis in the Shack-Hartmann sensor. A comprehensive survey of classical and contemporary wavefront reconstruction methods is presented in Section 2-3. The last section of this chapter, Section 2-4, concerns the classical and the distributed control of the deformable mirror.

2-1 The Adaptive Optics Control Problem

To explain the principle of AO and the role of wavefront reconstruction in the closed loop, we will briefly outline an AO application in an astronomical context. In order to support the astronomical example, consider the schematic in Figure 2-1, where the main components of an AO setup are depicted.

Before detailing the components of the system, some mathematical notation for describing the wavefront, φ (in phasor notation), must be established:

$$\varphi(\mathbf{r}, t) = A(\mathbf{r}, t)e^{\phi(\mathbf{r}, t)}, \tag{2-1}$$

where $\mathbf{r} \in \mathbb{R}^2$ specifies the spatial position in the telescope aperture and t denotes time. The quantity $A(\mathbf{r}, t)$ denotes the magnitude of the wavefront and $\phi(\mathbf{r}, t)$ its phase.

When light from a distant star arrives at the outer layers of the atmosphere, it has a flat wavefront. A flat wavefront is such that its phase profile $\phi(\mathbf{r}, t) = 0$ and, hence, if an image was captured from that wavefront it would not have any blurs. However, this flat wavefront will reach the telescope deformed as the turbulent atmosphere will introduce time and space varying optical path length differences. This gives rise to a turbulence induced phase profile $\phi(\mathbf{r}, t)$. The AO system tries to cancel out these wavefront distortions by actively introducing optical path length differences of opposite sign.

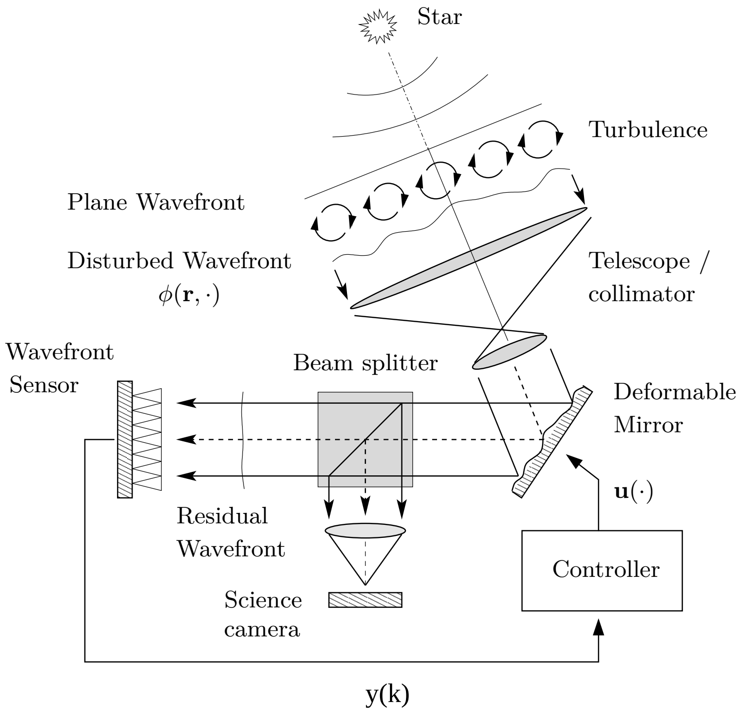

Figure 2-1: Schematic representation of an AO system and its main components (source: [1]).

An AO system is typically composed of the following components – a wavefront sensor (WFS), a wavefront corrector element to influence the phase and a feedback controller. In most systems, like the one depicted in Figure 2-1, the wavefront corrector element is a deformable mirror (DM).

Light that "enters" the AO system is first directed to the DM. By changing the mirror shape in real-time, the DM is able to apply a time-varying phase correction $\phi_{dm}(\mathbf{r}, t)$. The residual phase error ϵ is the difference between the turbulence induced wavefront and the applied correction, i.e. $\epsilon = \phi - \phi_{dm}$. After applying the wavefront correction, a beam splitter divides the reflected light beam in two parts. The first part of the corrected light beam leaves the AO system and is used by the science camera for object image formation. The remaining light is directed to the WFS, which provides quantitative information about the residual wavefront. Based on the WFS measurements $\mathbf{s}(\cdot)$, the controller has to determine the actuator inputs $\mathbf{u}(\cdot)$ to the DM. The controller should adapt the input signal in such a way that the DM cancels out most of the distortions. The latter degree of compensation is specified in terms of control criteria, like the “so-called” $\mathcal{H}_2$ criterium [23].

By counteracting the wavefront distortions, AO is able to reduce the degrading effect of atmospheric turbulence on the imaging process. The goal of an AO system is to make the wavefront of the light reaching the science camera as flat as possible. In this way, the image retrieved from the corrected wavefront can be recorded without being spread out when using long exposure times. By using AO, large ground-based based telescopes may reach close to diffraction limited performance (i.e. a performance close to the aberration free case only limited by the physical constituents of the telescope) in the near infrared light bandwidth [24, 25].

The next section will focus on one of the components of the AO system depicted in 2-1, the wavefront sensor.

2-2 Wavefront Sensors

Similarly to our eyes, the current types of image sensors can not measure the wavefront directly, neither magnitude nor phase, but only the intensity of the light. The reason behind this is that the sensors are only sensitive to brightness levels.

When the imaging system and/or the transmission medium disturbs the spherical (or planar in case of a source at infinity) wavefront, blurring of the image occurs. This corrupted wavefront is called an aberration. To deblur the image one can estimate the phase of the wavefront so as to try to eliminate the aberration in order to obtain a flat phased wavefront again.

We can divide imaging sensors in two main categories: pupil-plane sensors and focal-plane sensors. The pupil sensors split the incoming light into two beams and process one of the beams so that a linear relationship between the sensor reading and the unknown wavefront can be established. The focal plane sensors do not split the beams and thus work with the image of the science camera directly (hence, no photons are diverted from the image into the wavefront sensor). To solve the phase retrieval problem is a non-trivial task given the non-linear relation between the wavefront and the sensor readings.

The wavefront sensors which are relevant in the scope of this project are the pupil-plane ones, of which the Shack-Hartmann sensor will be described in Section 2-2-2. Beforehand, the principle that motivated the invention of such sensors is explained in Section 2-2-1. The following sections follow the structure and nomenclature of [1].

2-2-1 Focal Spot Deviation

In order to exemplify this principle, let us consider a point source, such as a star, to continue the astronomical example started in Section 2-1.

The image of such a point source in an ideal ground-based telescope without the presence of atmosphere is shaped by the diffraction inherent to the diameter of the telescope and the wavelength of the light that is being captured. Looking at Figure 2-2, the first dark ring (that is, the first negative concavity of the function) defines the resolution. This resolution represents the minimum distance between the maxima of two Airy disks such that they are distinguishable and it is given by $1.22\lambda/D$, where D represents the diameter of the telescope and λ the wavelength of the incident wavefront. The criterion that defines this resolution is called Rayleigh's criterion [26, Section 6.5.2] which characterizes our capability to distinguish two adjacent point sources.

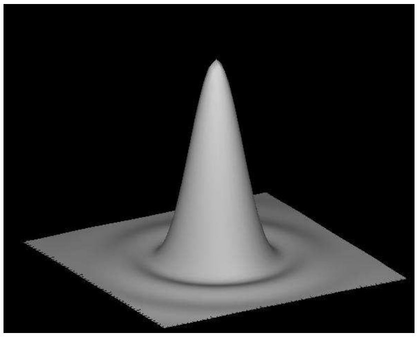

Figure 2-2: Airy Disk (source: [1]).

From a control engineering perspective, the image of a point source under ideal condition can be considered as the 2-D impulse response of the system.

Let us now consider that an aberration distorts the perfect Airy disk. In the case where the aberration is only a tilt, the effect on the deformed image can be computed analytically. When the deviation due to tilting the wavefront is denoted by $\phi(x, y)$ it has been shown [2] that the perturbed spot deviates from its center (obtained with the ideal flat wavefront) by the distance Δx given by

$$\Delta x = \kappa \frac{d\phi}{dx}, \tag{2-2}$$

where κ is a constant parameter determined by the optical parameters such as pupil size, distance between pupil and lens, etc. The same formula holds for the deviation in y direction, Δy.

2-2-2 Shack-Hartmann Sensor

In the previous section, we considered the whole wavefront phase distribution as being described by just a tilt. Normally, wavefront aberrations are more complex and so Hartmann, in the 1900, had the idea of constructing an array of apertures that would sample the wavefront (see Figure 2-3). The aberration "seen by" each aperture can then be approximated by a tilt. Thus, if we sample the wavefront with a finer aperture grid the approximation power is rendered higher as we can describe more complex wavefronts.

Roland Shack and Ben Platt [27] later proposed to replace the aperture array by a lenslet array. The idea stemmed from the fact that the Hartmann array was producing inaccurate and low intensity images which could be avoided by focusing the sampled rays with a set of lenses. This sensor is still used in numerous Adaptive Optics applications and the reader may refer to the introductory chapter for an overview of such applications.

In Figure 2-4, a simple one dimensional version of the principle behind the Shack-Hartmann sensor is shown. The array of lenses is parallel to a photon sensor, typically a Charge-Coupled Device (CCD) camera or a quad cell[1]. For each aperture (lens), the deviation of the focal

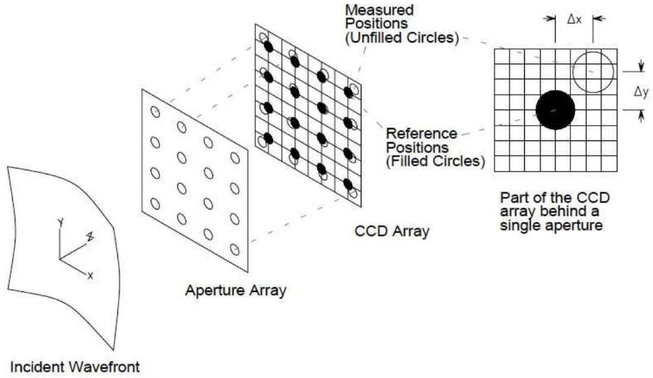
Measured Positions (Unfilled Circles)
Reference Positions (Filled Circles)
CCD Array
Aperture Array
Incident Wavefront
Part of the CCD array behind a single aperture
Δx
Δy
Y
Z
X

Figure 2-3: Schematic of an Hartmann array of pupils (or apertures), (source: [2]).

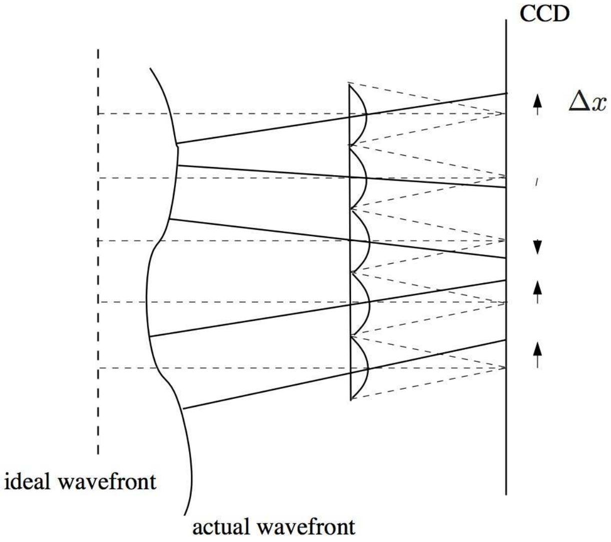

Figure 2-4: One dimensional representation of a Shack-Hartmann sensor. (source: [1])

spot from the center location is given by Equation 2-2 and denoted in 2D by a deviation in x and y direction as

$$\Delta x(i,j) = \kappa_x \frac{\partial \phi(x_i, y_j)}{\partial x} + \eta_x(i,j), \tag{2-3}$$

$$\Delta y(i,j) = \kappa_y \frac{\partial \phi(x_i, y_j)}{\partial y} + \eta_y(i,j), \tag{2-4}$$

where i, j represent the lenslet (aperture) in the i-th row, j-th column of the rectangular lenslet (aperture) array; x_i and y_i are the spatial coordinates of the wavefront and $\eta_x(i,j)$

[1] The CCD, concept which gave George Smith and Williard Boyle the Nobel Prize in Physics in 2009, converts the incoming photons into electron charges at a semiconductor-oxide interface; afterwards it is able to read out these charges by means of measuring the charge of the capacitor associated with each pixel [28]. The quad-cell bases its operating principles in a photodiode [29].

and $\eta_y(i,j)$ represent the measurement noise and the effect from higher order aberrations that differ from just the tilt of the wavefront that can be estimated in aperture (i,j). The κ_x and κ_y are a purely geometric quantity. Roughly, $\kappa = z$ through a series of trigonometric approximations, where z is the distance that separates the array from the camera. The full derivation can be consulted in the Appendix A-2.

Due to diffraction (as explained in Section 2-2-1) not a single pixel but an area will in general be highlighted by the projection of a point source on the CCD camera. In order to compute the slopes, a centroid algorithm is applied (*e.g.*, [30]), such that we can measure how much the centroid changed from its ideal position (flat wavefront). The slopes are then written as,

$$\begin{aligned} s_x(i,j) &= \frac{1}{z}\frac{\sum_{u,v} u\Delta_x I(u,v)}{\sum_{u,v} I(u,v)} = \frac{\partial \phi(x_i,y_j)}{\partial x} + \tilde{\eta}_x(i,j), \\ s_y(i,j) &= \frac{1}{z}\frac{\sum_{u,v} v\Delta_y I(u,v)}{\sum_{u,v} I(u,v)} = \frac{\partial \phi(x_i,y_j)}{\partial y} + \tilde{\eta}_y(i,j), \end{aligned} \tag{2-5}$$

where $I(u,v)$ is the intensity measured in the pixel in the u-th row and v-th column of the CCD camera measuring the wavefront seen by aperture (i,j). The quantities Δx and Δy are the spacing of the pixels along the x and the y axis, respectively.

Although we presented the general formula of the centroid algorithm, the pairs (u,v) where the summation in Eq. (2-5) is performed are not clearly specified. A straightforward way to define these pairs is, firstly, to select the maximum intensity value closest to where the maximum would ideally lie, if we had a flat wavefront. Then, we can select all the pixels around the maximum using a rectangular or circular window. Other methods include more involved thresholding mechanisms [31] or the inclusion of a weighting rectangular or Gaussian window [32]. Furthermore, there are also matched filter approaches such as the one presented in [33] and the references therein. In [34, Section 4] a very clear and succinct comparison between the different centroiding techniques is presented.

2-3 Wavefront Reconstruction

The information obtained from the wavefront sensors can be used to estimate the phase aberration. The process of estimating this quantity can be called wavefront/aberration reconstruction, phase-retrieval problem, or disturbance estimation. These terms will be used to denote the exact same procedure.

The seminal works from Fried [35], Hudgin [36] and Southwell [3], among others, created the classical wavefront reconstruction algorithms which influenced most of the methods used nowadays. Fried and Hudgin developed a zonal reconstruction method which, despite its simplicity, has been optimized and improved recently in [37, 38, 39, 40]. Southwell presented a phase-retrieval method which was coined as modal reconstruction. This method was later enhanced in terms of computational complexity in [41, 42].

This section starts by presenting the classical zonal and modal reconstruction algorithms, in Sections 2-3-1 and 2-3-2, respectively. After those, the state-of-the-art methods are presented in Section 2-3-3 and an analysis regarding their computational complexity and their performance is made.

2-3-1 Zonal Reconstruction

Once the slope measurements are available (via, e.g., the centroid method in Eq. (2-5)), a linear model can be created such that it relates the unknown phase values ϕ with the slopes of the phase s_x and s_y via a finite difference approach. Fried ([35]) and Hudgin ([36]) were the ones who firstly described this method, although each of them related the phase values and the slope measurements differently (see Figure 2-5).

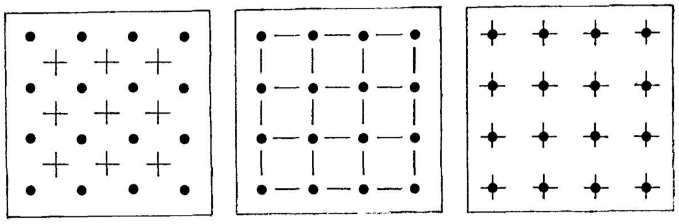

Figure 2-5: Phase points (denoted with a circle) and slopes (denoted with a dash) used by (Left) Southwell, (Center) Hudgin and (Right) Fried (source: [3]).

The equations that establish the relations in e.g. Fried's finite difference model are

$$\begin{aligned} s_x(i,j) &= \frac{(\phi(i+1,j)+\phi(i+1,j+1))/2-(\phi(i,j)+\phi(i,j+1))/2}{h}, \\ s_y(i,j) &= \frac{(\phi(i,j+1)+\phi(i+1,j+1))/2-(\phi(i,j)+\phi(i+1,j))/2}{h}, \end{aligned} \tag{2-6}$$

where h represents, for a uniformly arranged lenslet (aperture) array, the distance between the center of the lenslets (apertures) in the pupil plane.

Given Eq. (2-6) and Figure 2-5, the derivation of the equations in Hudgin's and Southwell's finite difference model is straightforward and can be found in [36, 3].

We can then stack the slopes in a vector $\mathbf{s}$ and the unknown phase values in another vector $\boldsymbol{\phi}$ such that their linear relationship can be compactly represented as

$$\mathbf{s} = \mathbf{A}\boldsymbol{\phi} + \boldsymbol{\eta}, \tag{2-7}$$

where $\mathbf{A}$ is the matrix that defines the finite differences, $\mathbf{s}$ are the stacked slope measurements, and $\boldsymbol{\phi}$ the stacked phase points. The noise that affects the slopes is denoted by $\boldsymbol{\eta}$.

A standard solution for the problem of finding $\boldsymbol{\phi}$ is to compute the pseudo-inverse matrix $\mathbf{A}^+$ and thus solving a least-squares problem. The estimate is given by $\hat{\boldsymbol{\phi}} = \mathbf{A}^+\mathbf{s}$.

However, due to the fact that any constant wavefront added to the obtained solution also satisfies the model, the problem is ill-posed which means that the matrix $\mathbf{A}^\top\mathbf{A}$ is singular, whichever geometry is chosen. This integration constant that can not be estimated via the slope measurements only is called the piston mode.

If we choose the Fried geometry, we have a second rank deficiency in $\mathbf{A}^\top\mathbf{A}$ due to the waffle mode. This mode corresponds to a checker-board-like pattern which that geometry can not take into account. The rank deficiencies associated with the different phase grids are analysed by Herrman in [43].

Also in [43], Herrman extended matrix $\mathbf{A}$ in such a way that $\mathbf{A}^\top\mathbf{A}$ is fully ranked. This extension is made by adding another set of equations to the system such that the piston and waffle mode are fixed at 0.

An improvement on the estimation can be done if we know more about the statistical properties of the phase or the noise. The phase can be statistically described by $\boldsymbol{\phi} \sim \mathcal{N}(0, \mathbf{C}_{\boldsymbol{\phi}\boldsymbol{\phi}})$, where the symbol $\mathcal{N}(0, \mathbf{C}_{\boldsymbol{\phi}\boldsymbol{\phi}})$ denotes a zero-mean Gaussian probability density function with covariance matrix $\mathbf{C}_{\boldsymbol{\phi}\boldsymbol{\phi}} = \langle \boldsymbol{\phi}\boldsymbol{\phi}^\top \rangle$. Additionally, we may be able to have access to the statistical properties of the noise, $\boldsymbol{\eta}$, namely to its covariance matrix $\mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}} = \langle \boldsymbol{\eta}\boldsymbol{\eta} \rangle$. Considering that the noise can be described by $\boldsymbol{\eta} \sim \mathcal{N}(0, \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}})$, a stochastic and weighted least squares problem must be solved. A possible expression for the estimator needed ([44, Theorem 4.3] and[45, Section 10.6]) to solve this problem is given by

$$\hat{\boldsymbol{\phi}} = (\mathbf{A}^\top \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \mathbf{A} + \mathbf{C}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1})^{-1} \mathbf{A}^\top \mathbf{C}_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1} \mathbf{s}. \tag{2-8}$$

In applications for astronomy, this covariance matrix can be defined using the Kolmogorov power spectrum [46, see Eq. 11] which describes the atmospheric turbulence.

2-3-2 Modal Reconstruction

This type of reconstruction is also formulated as a least-squares problem, although the underlying principle is different. The seminal work was done by Southwell in [3] some years after Fried presented the zonal reconstruction method.

The first step is to parametrize the entire phase distribution $\phi(x, y)$ in terms of K arbitrary orthogonal and normalized basis functions F_k as

$$\phi(x, y) = \sum_{k=0}^{K} a_k F_k(x, y), \tag{2-9}$$

where a_k are the coefficients to be determined.

Differentiating the phase expression in (2-9) the slope model is obtained as

$$\begin{aligned} s_x(x, y) &= \sum_{k=1}^{K} a_k \frac{\partial F_k(x, y)}{\partial x}, \\ s_y(x, y) &= \sum_{k=1}^{K} a_k \frac{\partial F_k(x, y)}{\partial y}. \end{aligned} \tag{2-10}$$

The slope measurements in Eq. (2-10) are defined in a continuous domain by the (x, y) pair. However, the domain needs to be discretized for implementation purposes. We can, thus, sample the domain according to the number of elements in the grid of our wavefront sensor and replace the continuous pair (x, y) by a discrete one, (i, j), corresponding to a certain grid element.

After defining a suitable set of basis functions (Legendre [3], Zernike [47] or complex exponentials [41]), Equation (2-10) can be reformulated in matrix form

$$\mathbf{s} = \mathbf{A}\mathbf{a} + \boldsymbol{\eta}, \tag{2-11}$$

and solved for the coefficients $\mathbf{a}$ by inverting the matrix $\mathbf{A}$.

At this point one may wonder about the absence of the a_0 coefficient in the slope model in Eq. (2-10). This coefficient can not be determined as it represents the piston mode. However, if one uses basis functions that have zero-mean (as it is done in [3]), then the estimated phase will also be zero-mean if the piston mode is set to zero ($a_0 = 0$). Due to this absence, we guarantee that the normal equations are not ill conditioned.

2-3-3 Novel methods for Wavefront Reconstruction

After the breakthrough in wavefront reconstruction in the late 70s and early 80s, the technological developments that led to the construction of optical devices using more and more sensors and actuators have driven researchers to look for faster reconstruction techniques, implementable in real-time environments. Some of these methods are described below.

Fast Fourier Transform

The work of Poyneer et al. in [42] uses the Fast Fourier Transform (FFT) to reconstruct the wavefront improving the earlier works by Freischlad et al. [41]. The algorithm Freischlad developed can be explained using the framework of modal reconstruction algorithms, where the basis functions are defined as complex exponentials in such way that the wavefront can be described as a Discrete Fourier Transform (DFT) of the basis functions weighting coefficients. After some manipulation, those coefficients can be estimated by taking the DFT of the slope measurements together with other scaling factors. Poyneer et al. improve this idea and apply it to circular sensor configurations and other grid geometries.

Multigrid methods

Gilles et al. [48] formulate the reconstruction in a least-squares sense using a multigrid phase retrieval method to perform a zonal reconstruction. However, it restricts the problem formulation for astronomical applications only. Therefore it can enhance the quality of the estimate in terms of its variance if the characteristics of the atmospheric phase covariance matrix $\mathbf{C}_{\phi\phi}$ and the noise affecting the slope measurements $\mathbf{C}_{\eta\eta}$ are known. In this paper, it is assumed that $\mathbf{C}_{\eta\eta} = \sigma^2\mathbf{I}$. As for the atmospheric turbulence, it is proved that $\mathbf{C}_{\phi\phi}$ can be closely approximated by a Semi-Positive Definite (SPD) matrix, $\tilde{\mathbf{C}}_{\phi\phi}$.

The expression in (2-8) can be adapted into the following format (without the inverse and for a diagonal noise covariance matrix):

$$\mathbf{B}\boldsymbol{\phi} = \mathbf{A}^\top \mathbf{s}, \quad \text{with } \mathbf{B} = \mathbf{A}^\top \mathbf{A} + \sigma^2 \tilde{\mathbf{C}}_{\boldsymbol{\phi}\boldsymbol{\phi}}^{-1}. \tag{2-12}$$

Due to the fact that $\tilde{\mathbf{C}}_{\phi\phi}$ is SPD, its inverse will also have that property. Hence, matrix $\mathbf{B}$ will also be SPD because it is the sum of two SPD matrices given that $\mathbf{A}^\top\mathbf{A}$ is symmetric SPD and sparse.

Equation (2-12) can then be solved iteratively using the Gauss-Seidel (GS) method [49], which guarantees convergence if matrix $\mathbf{B}$ is SPD. However this procedure is not very effective in eliminating the low frequency-error. For this reason, the GS method is called a smoother. Different smoothers can be devised to be more effective at lower or higher frequencies. After a solution is obtained with the classical GS, the residual is projected onto a coarser phase grid and a smoother more prone to correct lower frequency errors is applied. This can be done recursively until we achieve the coarsest resolution we defined. The method, when applied recursively is called multigrid.

Vogel and Yang [50] refine the work by Gilles et al. by generalizing it to other sensor-actuator geometries.

Wavelet-based Methods

In [51], a reconstruction method using wavelets is presented which also has linear complexity ($13N$, where N represents the number of unknown phase points). This method is highly scalable and uses a modified 2-D Haar wavelet filterbank which processes the wavefront gradients directly. This method has the advantage of enabling denoising the image via standard wavelet denoising techniques.

Fractal Iterative Method

In [38], Thiébaut and Tallon present an iterative method called FrIM (Fractal Iterative Method) with linear complexity ($\mathcal{O}(N)$, where N represents the number of actuators). They make use of the fractal (self-similar) structure of the Kolmogorov phase screens [46] and are able to avoid inverting the covariance matrix of a turbulence screen.

Cumulative reconstructor

Another method was proposed by [39] (and improved in [40]) called CuRe (Cumulative Reconstructor) which achieves a very low computational effort ($\mathcal{O}(N)$). The results were comparable in terms of accuracy with the standard zonal reconstruction methods. This algorithm works by integrating cumulatively the horizontal and the vertical lines of the phase point grid. Using a grid from Hudgins (Figure 2-5, center) the phase profile can be created by joining two adjacent points in a given direction by drawing a line whose steepness will be given by the slope measurements. After doing this (parallelizable) procedure for both horizontal and vertical lines an average horizontal and vertical profile is computed and the lines are shifted according to this profile.

Spline-based methods

There is a recent method which parametrizes the wavefront using splines which is one of the seminal works that gave rise to this graduation project. The algorithm is called SABRE and was presented by de Visser and Verhaegen in [52]. It parametrizes locally the wavefront using splines defined in non-overlapping triangular regions. It claims to have a bigger approximation power than the common modal modes due to the partitioning of the wavefront. Although it is implemented as a Matrix-Vector Multiplication (MVM) method, SABRE has a very high potential for parallelization and it is suitable for any type of sensor-actuator arrangement. The results were comparable to the classical zonal reconstruction methods using a square uniform phase point grid. A distributed version of this algorithm can be found in [53].

Non-linear Intensity Phase Retrieval

Another novel method for WFR is the one presented by Polo et al. in [12]. In this paper, a phase-retrieval method is proposed for EUVL applications using the Hartmann wavefront sensor as the SH sensor blocks the wavelengths used in EUVL technology. The main difference in this method is that the centroid algorithm is not used which means that the wavefront reconstruction is performed using directly the intensity distribution, without ever computing the slope measurements.

In order to retrieve the phase information, a non-linear model is created that relates the unknown wavefront at the pupil plane and the intensity distribution at the propagation plane[2]. A non-linear least squares algorithm is employed afterwards in order to minimize the discrepancies between the experimental and the modelled intensity distribution. The minimization is made with respect to the coefficients of the Zernike basis functions that parametrize the unknown wavefront.

When dealing with static aberrations, these wavefront reconstruction methods are enough to compensate them by inputting the necessary corrections to a deformable mirror. However, if the aberrations change in time, then an active control mechanism is required. This topic is discussed in the next section.

2-4 Closed loop Control with Deformable Mirror

The active element in an AO setup is the deformable mirror. When actuated, the DM is able to effectively minimize in real-time the blurring caused by phase aberrations. In Section 2-4-1, a classical control approach is presented and in the next section, Section 2-4-2, we refer to some references on distributed control methods.

[2] The fact that we are using, in the context of [12], a Hartmann sensor, which has no lenses, is why the intensity distribution is said to be in the *propagation* plane. If we were referring to the same plane for the SH sensor the term *focal* plane would be the most appropriate.

2-4-1 Classical Approach

The classical approach for controlling a deformable mirror in an AO setup can be seen in the block diagram of 2-6.

The process begins when the incident phase of the wavefront $\phi(k)$ at time instant k is reflected by the mirror. The DM modifies the incident phase as it introduces a phase change of $-\phi_m(k)$. The resulting reflected wavefront is given by $\epsilon_\phi(k) = \phi(k) - \phi_m(k)$ which enters the WFS and is corrupted by noise denoted by $\eta(k)$. This WFS measurement is denoted as $\epsilon_y(k)$. The WFS block has some dynamics which are not explicit in the block diagram: it needs to integrate the photons as they arrive and should contain a delay as to model the communication overhead and the processing time.

The corrupted measurement is then used to reconstruct the wavefront. In this block diagram the reconstruction is depicted as a linear operation by means of the reconstruction matrix **R**. Although the great majority of the algorithms described in this chapter are linear, that need not be so as a non-linear algorithm such as the one in [12] may be used in the reconstruction. Regardless of what method is employed, the reconstructed wavefront is given by $\hat{\epsilon}_\phi$.

After the reconstruction, matrix **F** is responsible for projecting the reconstructed wavefront into the actuator space. The construction of matrix **F** requires knowledge of the model of the DM as it provides an inverse mapping between a desired phase and the actuator commands. The actuator commands generated by matrix **F** are then filtered by the controller and can be provided to the deformable mirror.

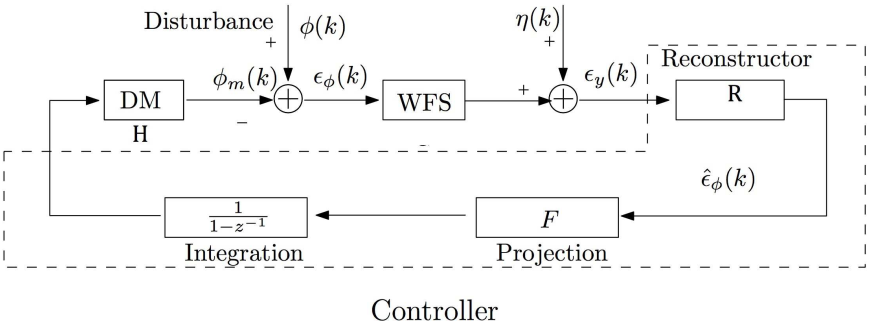

Figure 2-6: Block diagram of a classical feedback loop used in Adaptive Optics (source [4]).

If we do not have access to an experimental setup and want to test an algorithm, the loop could be closed by having the actuator commands transformed into the phase modification that the mirror will provide. For that we need a model of the mirror (usually a linear model, represented by a matrix **H**). With that final link the full feedback loop is completed.

Usually, in this application there are delays due to communication or processing. The effect of these delays is minimized by using an integral controller. If we denote the actuator commands by $u(k)$, the filtering made by the integral controller can be written as

$$u(k+1) = u(k) + \mu \Delta u(k+1), \tag{2-13}$$

where μ denotes the integral gain and $\Delta u(k+1)$ is the incremental controller action calculated after the reconstruction of the residual wavefront. This stems from the fact that we are only measuring the residual wavefront in the WFS and hence only an incremental input needs to be applied.

2-4-2 Distributed Approach

The need for a distributed control approach in AO is justified by the fact that the next generation of large-scale ground-based telescopes will involve a very high number ($\approx 10^4$) of actuators and sensors [8]. Besides that, the reconstruction of the wavefront and the computation of the control signal should be done in less than 1 ms [8].

Massioni in [54] and [55] applies a decomposition method to the complete DM model where he assumes that each of its actuators is an identical subsystem. The structure of the decomposed system is given by a “pattern matrix” which is sparse due to the fact that each subsystem only influences its nearest neighbours. It is possible to take advantage of the sparsity of the “pattern matrix” and devise an $\mathcal{H}_2$ optimal controller that is scalable with respect to the number of actuators.

Chapter 3

B-Splines Framework

The nomenclature used in this chapter will follow closely the one presented by de Visser in [52, 5]. For a more in-depth treatment of the multivariate simplex B-Splines the reader is advised to read [56] or the seminal work [57]. In order to maintain the nomenclature by de Visser, the quantity $\mathbf{B}$ *will in fact be a vector and not a matrix.*

The recent work [52, 5] in the field of B-splines has led to the development of a solid framework which can be (and has been) used to the benefit of certain stages of Adaptive Optics. These geometrical entities allow a local parametrization of the wavefront or the deformable mirror which can lead to the development of distributed estimation or control methods. Besides this important locality property, the splines framework inherently yields smoothness equality constraints, which are useful to provide reliable and smooth estimates for the afore mentioned distributed approaches.

This chapter is organized as follows. In Section 3-1 a general introduction is made regarding B-splines and their main properties. The types of triangulations that will be used in the following chapters will be presented in Section 3-2. Afterwards, the spatial position of the splines coefficients is analysed in Section 3-3. Lastly, in Section 3-4, the importance and influence of having continuity constraints is analysed.

3-1 Introduction to Simplex Splines

A spline is a potentially multivariate, piecewise function which is described by a polynomial. It has a certain degree of continuity between its individual segments. The usefulness of this piecewise nature is that it can be used to fit data that is too complex to be approximated with a single functional piece.

Among the several types of existent splines, this work will focus on two dimensional simplex splines given their usefulness to model, e.g. the phase distributions in a wavefront or the profile of a deformable mirror. These splines were chosen mainly for their inherent suitability to be incorporated in distributed algorithms.

The simplex splines have their domain defined by several geometric structures called simplices. A simplex is a polyhedral formed by the convex hull of $n+1$ points, when the spline is defined in $\mathbb{R}^n$. In the two dimensional case relevant to this work, the simplices are triangles which can be defined using three non-degenerate vertices $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^2$.

The basis polynomials of the simplex B-Splines are defined in local barycentric coordinates defined only inside a particular simplex. The reason for this barycentric representation is that it is easier to evaluate the function when working with these coordinages. Given a certain point $\mathbf{x} = (x, y) \in \mathbb{R}^2$ belonging to the Cartesian plane a barycentric coordinate in $\mathbb{R}^3$ can be determined as follows:

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \mathbf{V}^{-1} \begin{bmatrix} x \\ y \end{bmatrix}, \tag{3-1}$$

$$b_0 = 1 - b_1 - b_2, \tag{3-2}$$

where the transformation matrix $\mathbf{V}$ is defined as

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 - \mathbf{v}_0 & \mathbf{v}_2 - \mathbf{v}_0 . \end{bmatrix} \tag{3-3}$$

To simplify the notation, the transformation from Cartesian $\mathbf{x} \in \mathbb{R}^2$ to Barycentric coordinates $\mathbf{b} \in \mathbb{R}^3$ will be denoted as

$$\mathbf{b}(\mathbf{x}) = (b_0, b_1, b_2). \tag{3-4}$$

A distinguishing feature of all splines functions is their locality. This property can be used in order to achieve a higher approximation power and thus, model more complex functions than we could if we used a global basis function, *e.g.*, Zernike polynomials. We start by having a local parametrization that describes the function in only one simplex and then join several simplices together and create a partition of the domain which forms a triangulation. This triangulation is usually represented by $\mathcal{T}$ and consists of non-overlapping simplices.

Now that the barycentric coordinates and the simplices have been clearly defined, we can introduce the splines' basis functions. Using the fact that any polynomial of degree d can be expanded into a sum of monomials, we can write the following expression in the barycentric coordinate system:

$$(b_0 + b_1 + b_2)^d = \sum_{|\kappa|=d} \frac{d!}{\kappa_0!\kappa_1!\kappa_2!} b_0^{\kappa_0} b_1^{\kappa_1} b_2^{\kappa_2}, \tag{3-5}$$

where

$$|\kappa| = \kappa_0 + \kappa_1 + \kappa_2 = d, \quad \kappa_0, \kappa_1, \kappa_2 \in \mathbb{N}_0 \geq 0. \tag{3-6}$$

Each of the monomials in the right-hand side of (3-5) are an individual basis function. This function takes non-zero values only inside a certain simplex t, by definition, and can be written as

$$B_\kappa^d(\mathbf{b}(\mathbf{x})) = \begin{cases} \frac{d!}{\kappa_0!\kappa_1!\kappa_2!} b_0^{\kappa_0} b_1^{\kappa_1} b_2^{\kappa_2}, & \mathbf{x} \in t \\ 0, & \mathbf{x} \notin t. \end{cases} \tag{3-7}$$

The total number of these basis functions is given by the different combinations of κ_j that satisfy $\kappa_0 + \kappa_1 + \kappa_2 = d$. The expression that yields the total number of basis functions in a certain simplex for $n = 2$ and arbitrary d is given by $\hat{d}$ as follows:

$$\hat{d} = \frac{(d+2)!}{2d!}. \tag{3-8}$$

Weighting each of the $\hat{d}$ basis functions with their corresponding coefficient c^t_κ yields a polynomial $p(\mathbf{b}(\mathbf{x}))$ defined as

$$p(\mathbf{b}(\mathbf{x})) = \begin{cases} \sum_{|\kappa|=d} c^t_\kappa B^d_\kappa(\mathbf{b}(\mathbf{x})) & , \mathbf{x} \in t \\ 0 & , \mathbf{x} \notin t. \end{cases} \tag{3-9}$$

Equation (3-9) can easily be translated into vector form (also denoted as B-form) firstly by placing each of the basis functions B^d_κ for all κ such that $|\kappa| = d$ evaluated at a certain point $\mathbf{x}$ in the columns of the row vector $\mathbf{B}^d_t(\mathbf{b}(\mathbf{x}))$. The coefficients c^t_κ, for all κ such that $|\kappa| = d$ that weigh the basis functions should be placed in a column vector $\mathbf{c}_t$, sorted in accordance with the order of $\mathbf{B}^d_t(\mathbf{b}(\mathbf{x}))$. It is crucial to bear in mind, however, that the vector $\mathbf{x}$ is only defined in one single triangle t:

$$p(\mathbf{x}) = \begin{cases} \mathbf{B}^d_t(\mathbf{b}(\mathbf{x}))\mathbf{c}_t & , \mathbf{x} \in t \\ 0 & , \mathbf{x} \notin t \end{cases} \tag{3-10}$$

So far, we have only defined polynomials within a single triangle t. If we generalize Eq. (3-10) for the J triangles pertaining to an arbitrary triangulation $\mathcal{T}$, we will obtain the following global function s_d:

$$s_d(\mathbf{x}) = \mathbf{B}^d\mathbf{c}. \tag{3-11}$$

To clarify the way Eq. (3-11) was constructed, let us consider first the global vector $\mathbf{B}^d$ which groups the information pertaining to all the simplices t_1 until t_J. This vector is constructed by concatenating horizontally the row vector $\mathbf{B}^d_t(\mathbf{b}(\mathbf{x}))$, presented in (3-10), as

$$\mathbf{B}^d = \begin{bmatrix} \mathbf{B}^d_{t_1}(\mathbf{b}(\mathbf{x})) & \mathbf{B}^d_{t_2}(\mathbf{b}(\mathbf{x})) & \cdots & \mathbf{B}^d_{t_J}(\mathbf{b}(\mathbf{x})) \end{bmatrix} \in \mathbb{R}^{1\times J\hat{d}}. \tag{3-12}$$

One must note, however, that when $\mathbf{x}$ is defined outside the simplex t_j we will have $\mathbf{B}^d_{t_j}(\mathbf{b}(\mathbf{x})) = 0$ and, thus, the global $\mathbf{B}^d$ will be sparse.

The global vector $\mathbf{c}$ of B-coefficients in Eq. (3-11) follows the same construction guidelines of $\mathbf{B}^d$ and is given by

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}^\top_{t_1} & \mathbf{c}^\top_{t_1} & \cdots & \mathbf{c}^\top_{t_1} \end{bmatrix}^\top \in \mathbb{R}^{J\hat{d}\times 1}. \tag{3-13}$$

3-2 Triangulations

There are numerous ways in which triangulations can be defined from a set of points. The relevant triangulations in the scope of this thesis are Type I and Type II triangulations which will be presented next.

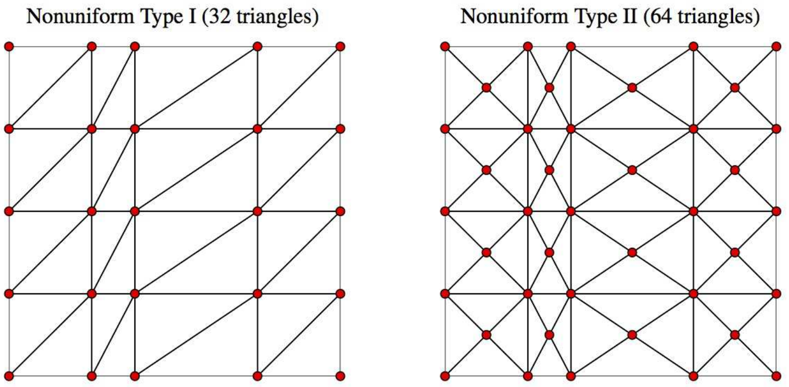
Nonuniform Type I (32 triangles)
Nonuniform Type II (64 triangles)

Figure 3-1: Nonuniform rectangular Type I (left) and Type II (right) triangulations in two dimensions. (source: [5].)

The Type I and Type II triangulations are the simplest forms of triangulation. A Type I/II triangulation is constructed by filling in the cells of a grid with a single symmetric triangulation (see Figure 3-1).

The only geometrical difference between the Type I and Type II triangulations is that the Type II uses a single extra vertex at the center of the square cell.

3-3 B-net

The B-coefficients are ordered in a unique spatial structure called the B-coefficient net (see Figure 3-2, or B-net [5, § 2.2.6]. The B-net enables a number of features that are unique to multivariate simplex splines such as:

- The simplification of the formulation of the continuity equations that govern continuity between polynomials on neighbouring simplices (see Section 3-4).
- The ability to perform local model modification without disrupting the global model structure by modifying B-coefficients close to a specific region of interest.

Some interesting observations can be made regarding the B-net. These observations are not necessarily relevant within the theoretical scope of this thesis, but were extremely useful when programming and testing functions which used B-splines.

In general, the value of the polynomial differs from the value of the B-coefficient at the same location. The only exceptions are the vertices, where the B-coefficients have the exact same value as the polynomial. The barycentric coordinate at a vertex contains only one nonzero component which is equal to 1. A second observation is that the B-form polynomial is bounded by the B-coefficients.

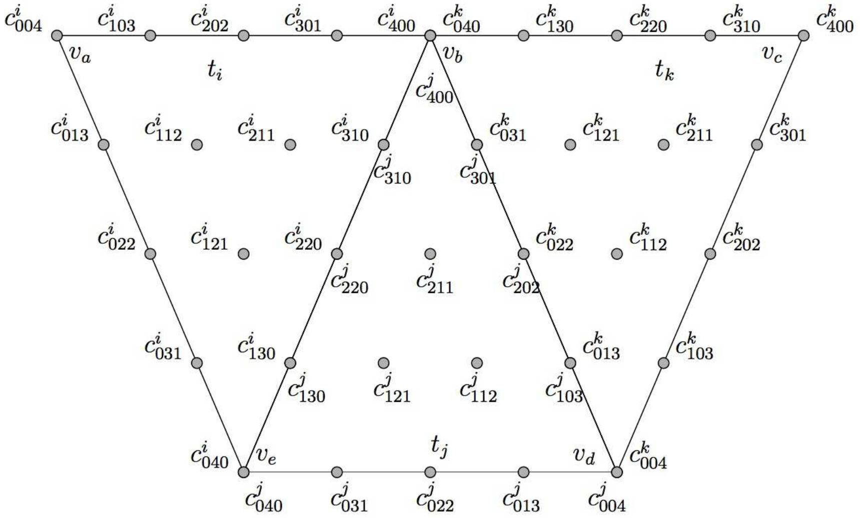

Figure 3-2: The B-coefficients have a unique spatial location inside their supporting simplex, the B-net. In this figure, we represent a B-net for a 4th degree spline function on a triangulation consisting of the 3 triangles t_i, t_j and t_k (source: [5])

.

3-4 Continuity Constraints

One of the main advantages of using B-splines is the fact one can force continuity of order r such that all m-th order derivatives, with $0 \leq m \leq r$, of two B-form polynomials defined on two neighbouring simplices are equal on the edge between the simplices. For C^r continuity there are a total of Q continuity conditions per simplex edge. Given that there are E edges, the total number of continuity conditions will be $Q \times E$, where Q is given by

$$Q = \sum_{m=0}^{r} (d - m + 1), \tag{3-14}$$

for a B-splines defined in $\mathbb{R}^2$.

The effects of applying different degrees of continuity are depicted in Figure 3-3. The effect of increasing the continuity degree can be summarized as follows:

- increases the global smoothness of a spline function,
- minimizes the possibility of overfitting and thus, reduces the sensitivity of a B-Splines estimator to noise,
- increases the numerical stability of the estimator for the B-coefficients of a spline function when data is scarce,
- reduces the approximation power of a spline function,
- leads to a higher propagation factor.

The last item has to do with the effect that a perturbation in a coefficient subject to continuity constraints has in the neighbouring simplices. It was shown in [5, § 3] that a type II triangulation is much less susceptible to perturbations than a triangulation of type I.

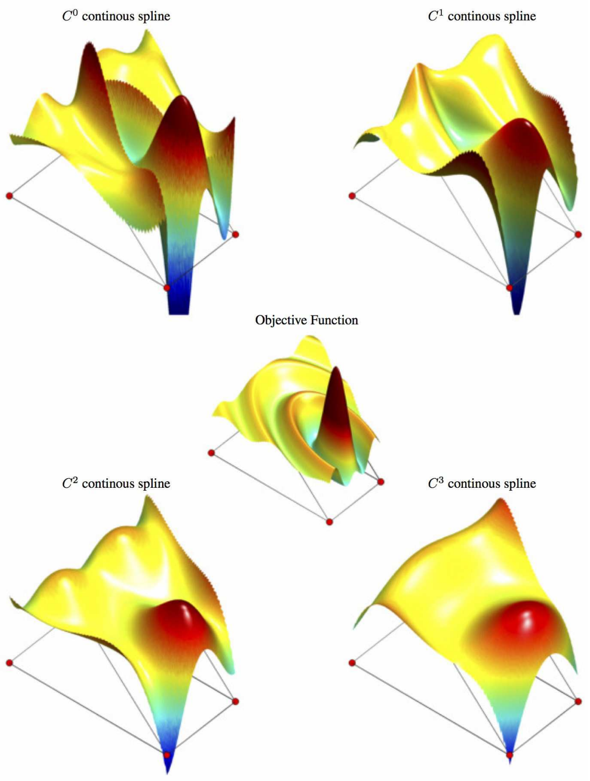

Figure 3-3: Four 4th degree spline functions with continuity orders C^0, C^1, C^2 and C^3 approximating the same objective function plotted in the center. (source: [5])

Smoothness Matrix. In [5, § 3], a general method was created to guarantee continuity between all the common edges of neighbouring B-form polynomials. The method yields a smoothness matrix $\mathbf{H}$ which is used in the following equality constraint

$$\mathbf{Hc} = 0, \tag{3-15}$$

that establishes a relation between the coefficients of the neighbouring polynomials and ensures continuity of the splines of neighbouring simplices up to a certain degree.

This global smoothness matrix $\mathbf{H} \in \mathbb{R}^{QE \times J\hat{d}}$ in Eq. (3-15) has in each of its rows a single continuity condition imposing an equality constraint on 2 B-coefficients belonging to neighbouring simplices. The vector $\mathbf{c}$ is the one containing all the coefficients as presented in Eq. (3-11).

Another property of the smoothness matrix is that it has a very high sparseness factor which increases with the size of the triangulation and the polynomial degree. Besides that, the matrix $\mathbf{H}$ is only fully ranked for very simple triangulations. In order to construct a full rank smoothness matrix, it is proposed, in [5], that the construction is performed row by row followed by a rank check of the recently created matrix $\mathbf{H}_{\text{current}}$ (via the efficient computation of the condition number of $\mathbf{H}_{\text{current}}\mathbf{H}_{\text{current}}^{\top}$, which gives us an indication on the singularity of matrix $\mathbf{H}_{\text{current}}$).

Chapter 4

Distributed Optimization

Finding the solution to optimization problems which involve a great number of variables can be very expensive computationally and in terms of memory usage, if done in a centralized manner. Decomposition methods can minimize these costs and achieve results similar to the centralized methods by dividing the problem into many subproblems and thus, working on a much smaller set of data and parallelizing the computational effort. In the context of AO, where the imaging systems have an increasing number of sensors and actuators, these methods can be used to find solutions in a distributed way to the wavefront reconstruction problem or to the control of the deformable mirror.

These decomposition methods are mainly divided into two categories: primal decomposition and dual decomposition. A combination of these two methods is denoted as alternative decomposition but it falls out of the scope of this thesis and will not be presented. The reader is encouraged to consult [58] for more information about alternative decomposition methods, and to refer to [59] and [60, Chapter 6] for other decomposition procedures.

The understanding of this chapter implies some background knowledge on general optimization techniques. The interested reader may refer to [61] or [20] for more information regarding subgradient methods and duality.

This chapter is organized as follows. We will start by explaining primal decomposition in Section 4-1. Afterwards, we will present the dual ascent and dual decomposition methods in Section 4-3-1 and 4-2, respectively. Then the method of multipliers in Section 4-3-2 will be discussed before finally presenting ADMM in Section 4-4.

4-1 Primal Decomposition

Using primal decomposition involves the direct manipulation of the primal variables. These primal variables are the variables with respect to which we want to minimize the cost function, and are usually denoted as $\mathbf{x}$. The applicability of this method mainly depends on the convexity of the cost function which makes this method suitable for a wide range of problems.

The principle behind primal decomposition can be illustrated using a very simple problem. Let us consider the following unconstrained optimization problem of minimizing the convex function f:

$$\text{minimize } f(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) = f_1(\mathbf{x}_1, \mathbf{y}) + f_2(\mathbf{x}_2, \mathbf{y}), \tag{4-1}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are the "private" variables of the convex functions f_1 and f_2, respectively, and $\mathbf{y}$ is the complicating variable that binds both the sub-problems together. If $\mathbf{y}$ was not present, the problem could be easily solved by minimizing f_1 and f_2 separately with respect to their non-overlapping private variables. If $\mathbf{y}$ is present, however, we need to perform the following minimizations

$$\begin{aligned} \phi_1(\mathbf{y}) &:= \underset{\mathbf{x}_1}{\text{minimize}}\ f_1(\mathbf{x}_1, \mathbf{y}), \\ \phi_2(\mathbf{y}) &:= \underset{\mathbf{x}_2}{\text{minimize}}\ f_2(\mathbf{x}_2, \mathbf{y}). \end{aligned} \tag{4-2}$$

These problems are respectively called subproblem 1 and subproblem 2. After obtaining the optima $\phi_1(\mathbf{y})$ and $\phi_2(\mathbf{y})$, we can recast the problem in (4-1) into the following equivalent formulation:

$$\underset{\mathbf{y}}{\text{minimize}}\ \phi_1(\mathbf{y}) + \phi_2(\mathbf{y}). \tag{4-3}$$

This last minimization problem is called the *master problem*. Notice that, if the original problem is convex, the master problem is also convex.

A subgradient method can be used to solve alternately the problems in Eq. (4-2) and the master problem in Eq. (4-3).

This decomposition method can also integrate constraints of the form $h(\mathbf{x}_1) + h(\mathbf{x}_2) \leq 0$. The problem needs to be reformulated and an extra variable per each constraint must be added. Finally, a subgradient method can also be used to solve such a problem. For a more comprehensive explanation regarding the integration of the constraints as well as some examples, we refer the reader to [62, Section 1].

4-2 Dual Decomposition

While in primal decomposition the optimization only occurs with respect to the primal variables, in dual decomposition, the dual variables are also manipulated. The dual variables are created, in general, using Lagrangian duality. Additional information regarding duality can be found extensively in literature, *e.g.*, in [20, Chapter 5] or [61, Chapter 5-6].

The seminal idea to decompose the dual problem was presented by Dantzig and Wolfe [63] and was applied to large-scale linear programming problems. Everett [64] presented a more general decomposition approach using the Lagrange multipliers, afterwards.

Using the dual of a problem is a common way to solve a constrained optimization problem, even if the problem is not decomposable. A general problem such as

$$\begin{aligned} &\text{minimize} \;\; f(\mathbf{x}), \\ &\text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned} \tag{4-4}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^{m \times 1}$ and the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, can be solved, for instance, by firstly forming the Lagrangian as

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + \mathbf{y}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}). \tag{4-5}$$

The Lagrangian in (4-5) was formed by introducing the dual variables $\mathbf{y} \in \mathbb{R}^{m \times 1}$, one dual variable per constraint in the $\mathbf{A}$ matrix. The dual variables introduced in the Lagrangian are also called Lagrange multipliers.

Then, we minimize the Lagrangian with respect to the primal and the dual variables and find its infimum as follows:

$$g(\mathbf{y}) := \inf_{\mathbf{x}} \Big(f(\mathbf{x}) + \mathbf{y}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) \Big), \tag{4-6}$$

where the dual function $g : \mathbb{R}^m \to \mathbb{R}$ is concave [20, Section 5.1.2] (regardless the convexity of f).

Given the concavity of the dual function g, we find the optimal Lagrange multipliers by solving the dual problem which maximizes the dual function. The dual problem is

$$\underset{\mathbf{y}}{\text{maximize}} \; g(\mathbf{y}), \tag{4-7}$$

and can be solved by a subgradient method.

In general, solving the dual problem yields an optimum $g^\star$ which is smaller than the minimum of the problem in (4-4), denoted as $u^\star$. When $g^\star = u^\star$, we say that *strong duality* holds, and the optimum value of the original problem can be obtained via solving the dual problem. A condition that guarantees strong duality is Slater's condition [61, Proposition 5.4.1] which is satisfied for the problem posed in (4-4) given that the condition $\mathbf{A}\mathbf{x} = \mathbf{b}$ specifies that $\mathbf{x}$ is defined in a hyperplane which is a closed (contains its boundary), convex set.

When strong duality holds, extract the optimal primal $\mathbf{x}^\star$ from the optimal dual $\mathbf{y}^\star$ as follows:

$$\mathbf{x}^\star := \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{y}^\star), \tag{4-8}$$

provided that there is only one minimizer $\mathbf{y}^\star$.

Now that we have used duality to solve a simple problem, we consider that the objective function is separable, which means that we can write the convex function f as

$$f(\mathbf{x}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i), \tag{4-9}$$

where $\mathbf{x} \in \mathbb{R}^n$ can be segmented into several subvectors $\mathbf{x}_i \in \mathbb{R}^{n_i}$ (with no components in common). If splitting f is possible then the partitioning of $\mathbf{A}$ can be done swiftly in conformity with the partitioning of $\mathbf{x}$, such that $\mathbf{A}\mathbf{x} = \sum_{i=1}^{N} \mathbf{A}_i\mathbf{x}_i$, with $\mathbf{A}_i \in \mathbb{R}^{m\times n_i}$

The partitioning of both $\mathbf{x}$ and $\mathbf{A}$ make it possible to define the Lagrangian $\mathcal{L}$ as follows:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) := \sum_{i=1}^{N} \mathcal{L}_i(\mathbf{x}_i, \mathbf{y}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i) + \mathbf{y}^\top \mathbf{A}_i\mathbf{x}_i - (1/N)\mathbf{y}^\top \mathbf{b}, \tag{4-10}$$

which is also separable in $\mathbf{x}$. This separability means that Eq. (4-15) can be solved independently for each $\mathbf{x}_i$, such that

$$\mathbf{x}_i^{k+1} := \arg\min_{\mathbf{x}_i}\ \mathcal{L}_i(\mathbf{x}_i, \mathbf{y}^k), \tag{4-11}$$

$$\mathbf{y}^{k+1} := \mathbf{y}^k + \alpha^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \tag{4-12}$$

After each $\mathbf{x}_i$ is calculated, the results are concatenated to form again $\mathbf{x}$ and the dual variable $\mathbf{y}$ can then be updated. The 2-steps in each iteration are thus called *gather* and *broadcast*. The residual $\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}$ is gathered from each individual processing unit and after the dual update the dual variable is broadcasted back.

4-3 Precursors of ADMM

In this section, we will first present preliminary information to provide the background for the understanding of the Alternating Direction Method of Multipliers (ADMM). This method allows a more robust performance than simple dual decomposition in terms of the type of problems that it can tackle together with the possibility of distributing the computational effort.

4-3-1 Dual Ascent

Let us consider the canonical optimization problem with equality constraints presented in (4-4). The Lagrangian for that problem is

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) := f(\mathbf{x}) + \mathbf{y}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) \tag{4-13}$$

and the dual function is

$$g(\mathbf{y}) := \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mathbf{y}). \tag{4-14}$$

With the dual ascent method the dual problem is solved by using the gradient ascent which translates into the following procedure:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}}\ \mathcal{L}(\mathbf{x}, \mathbf{y}^k), \tag{4-15}$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \alpha^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \tag{4-16}$$

The updates of the dual variable are performed using a step size $\alpha^k > 0$ which, if chosen appropriately and other assumptions hold [1], guarantees the convergence of $\mathbf{x}^{k+1}$ and $\mathbf{y}^{k+1}$ to the primal and dual optima, respectively. The strength of these assumptions prevent the usage of dual ascent in a wider range of problems.

Notice that in the expression for the update of the dual variable, we must follow the direction of the positive gradient, where $\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}$ is the gradient of $\mathcal{L}(\mathbf{x}, \mathbf{y})$ with respect to $\mathbf{y}$, and thus maximizing the dual function. Hence, the name of the method - dual ascent.

4-3-2 Method of Multipliers

The method of multipliers appears as an extension of the dual ascent that can provide a more robust performance under weaker assumptions than the ones needed to ensure the convergence of dual ascent. This method uses the augmented Lagrangian ([65] and [66]) $\mathcal{L}_\rho = \mathcal{L} + (\rho/2)||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$, with $\mathcal{L}$ defined in (4-5), which can be integrated in a problem formulation equivalent to the one in Eq. (4-4), giving

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) + (\rho/2)||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2 \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{array} \tag{4-17}$$

where $\rho > 0$ denotes the penalty parameter.

Notice that for a feasible $\mathbf{x}$ the extra term does not affect the objective function as the residual would be zero.

We can solve this problem by applying the dual ascent method presented in Eqs. (4-15) and (4-16), which yields the following update expressions

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \ \mathcal{L}_\rho(\mathbf{x}, \mathbf{y}^k), \tag{4-18}$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho^k(\mathbf{A}\mathbf{x}^{k+1} - \mathbf{b}). \tag{4-19}$$

The method of multipliers converges under far more general conditions than dual ascent, including cases when f takes the value $+\infty$ or is not strictly convex. However, by introducing the augmented Lagrangian, the problem is no longer separable which jeopardizes the application of a such a method in a distributed optimization context.

4-4 ADMM

The Alternating Direction Method of Multipliers (ADMM) method enables us to join the decomposability of the dual ascent with the superior convergence properties of the method of multipliers. This method was firstly proposed by Glowinski and Marocco[2] [67, Eqs. 9.7-9.9] for a very specific finite-element approach to solve the non-linear Dirichlet problem. Gabay

[1] For example, if f is not a proper function (*e.g.*, non-zero affine) the update in (4-15) fails as $\mathcal{L}$ is unbounded below in $\mathbf{x}$ for most $\mathbf{y}$.

and Mercier's [68] problem formulation is more similar to the one presented in this section as they try to minimize expressions of the form $f(Av) + g(v)$, where A is a linear operator.

The canonical form of the problems that can be solved by ADMM is

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) + g(\mathbf{z}), \\ &\text{subject to} && \mathbf{Ax} + \mathbf{Bz} = \mathbf{c}, \end{aligned} \tag{4-20}$$

with primal variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^m$, and where $\mathbf{A} \in \mathbb{R}^{p \times n}$, $\mathbf{B} \in \mathbb{R}^{p \times m}$ and $\mathbf{c} \in \mathbb{R}^p$. The functions f and g are assumed to be convex.

The augmented Lagrangian is then formed as

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}) := f(\mathbf{x}) + g(\mathbf{z}) + \mathbf{y}^\top(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2}||\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}||_2^2, \tag{4-21}$$

with the dual variable $\mathbf{y} \in \mathbb{R}^p$, and where $\rho > 0$ represents the penalty parameter already introduced in the method of multipliers.

If we were to apply the method of multipliers to solve the problem in Eq. (4-20) the two-step procedure would be described as follows:

$$\begin{aligned} (\mathbf{x}^{k+1}, \mathbf{z}^{k+1}) &= \arg\min_{\mathbf{x},\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \mathbf{y}^k) \\ \mathbf{y}^{k+1} &= \mathbf{y}^k + \rho(\mathbf{Ax}^{k+1} + \mathbf{Bz}^{k+1} - \mathbf{c}). \end{aligned} \tag{4-22}$$

In Eq. (4-22) the minimization is performed jointly in both $\mathbf{x}$ and $\mathbf{z}$. In ADMM however, it is proposed that the variable update alternates between the two variables, such that

$$\begin{aligned} \mathbf{x}^{k+1} &:= \arg\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^k, \mathbf{y}^k) \\ \mathbf{z}^{k+1} &:= \arg\min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{k+1}, \mathbf{z}, \mathbf{y}^k) \\ \mathbf{y}^{k+1} &:= \mathbf{y}^k + \rho(\mathbf{Ax}^{k+1} + \mathbf{Bz}^{k+1} - \mathbf{c}). \end{aligned} \tag{4-23}$$

ADMM decomposes the method of multipliers by using Gauss-Seidel iterations (see [70, Section 10.1]) instead of a joint minimization.

Scaled form. We can write ADMM in a different form by combining the augmented Lagrangian and scaling the dual variable. If we define the residual $\mathbf{r}$ as $\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}$ we have

$$\mathbf{y}^\top\mathbf{r} + (\rho/2)||\mathbf{r}||_2^2 = (\rho/2)||\mathbf{r} + \mathbf{u}||_2^2 - (\rho/2)||\mathbf{u}||_2^2,$$

where $\mathbf{u} = (1/\rho)\mathbf{y}$ is the scaled dual variable. The update equations can then be written as

[2] A less comprehensive English version of [67] can be found in [69].

$$\begin{aligned}\mathbf{x}^{k+1} &:= \arg\min_{\mathbf{x}} \Big(f(\mathbf{x}) + ||\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^k - \mathbf{c} + \mathbf{u}^k||_2^2 \Big) \\ \mathbf{z}^{k+1} &:= \arg\min_{\mathbf{z}} \Big(g(\mathbf{z}) + (\rho/2)||\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z} - \mathbf{c} + \mathbf{u}^k||_2^2 \Big) \\ \mathbf{u}^{k+1} &:= \mathbf{u}^k + (\mathbf{A}\mathbf{x}^{k+1} + \mathbf{B}\mathbf{z}^{k+1} - \mathbf{c}).\end{aligned} \tag{4-24}$$

The update laws in (4-24) represent the *scaled* version of ADMM, since it is expressed in terms of a scaled version of the dual variable. The first version presented in (4-23) is called, by opposition the *unscaled* form.

Convergence. The ADMM converges under the assumptions that the unaugmented Lagrangian has a saddle point and that the function f and g are closed, proper, and convex. If the assumptions hold, the following properties are satisfied:

1. *Residual convergence*: $\mathbf{r}^k$ as $k \to \infty$, where $\mathbf{r}^k = \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}$, *i.e.*, the iterates approach feasibility.

2. *Objective convergence*: $f(\mathbf{x}^k) + g(\mathbf{z}^k) \to p^\star$ s $k \to \infty$, where $p^\star$ denotes the optimal value of the problem in (4-20).

3. *Dual variable convergence*: $\mathbf{y}^k \to \mathbf{y}^\star$ as $k \to \infty$, where $\mathbf{y}^\star$ is a dual optimal point.

The convergence proofs and optimality conditions are presented in [71, Appendix A].

Stopping criteria. In [71, Section 3.3.1], it was shown that an indicator on the optimality of the system is given by bounding the norms of the primal and dual residuals at iteration k, which are denoted as $\mathbf{r}^k = \mathbf{A}\mathbf{x}^k + \mathbf{B}\mathbf{z}^k - \mathbf{c}$ and $\mathbf{s}^k = \rho\mathbf{A}^\top\mathbf{B}(\mathbf{z}^{k+1} - \mathbf{z}^k)$, respectively. The iterative method should stop when both the residuals satisfy the following bounds:

$$||\mathbf{r}^k||_2^2 \leq \epsilon^{\text{pri}}, \quad ||\mathbf{s}^k||_2^2 \leq \epsilon^{\text{dual}}, \tag{4-25}$$

where the constants $\epsilon^{\text{pri}} > 0$ and $\epsilon^{\text{dual}} > 0$ need to be defined *a priori*.

Varying Penalty Parameter. The penalty parameter affects dramatically the speed of convergence of the algorithm. A common extension to the classical ADMM implementation is to enable a varying penalty parameter ρ^k. Although it is difficult to prove convergence when ρ is varying at each time step, we can achieve significantly better results in practice [71, Section 3.4.1].

The ADMM update equations suggest that large values of ρ tend to produce smaller primal residuals $\mathbf{r}$, while smaller values tend to reduce the value of the dual residual $\mathbf{s}$.

Some results and heuristics regarding this extension have been presented in [72, 73].

4-4-1 Exchange Problem

Now that the ADMM method has been presented for two sets of variables connected through their constraints, we can generalize this method for N sets of variables. One particular example of such a generalization is the exchange problem [71, Section 7.3.2]. Assuming that the objective function $f(\mathbf{x})$ is separable in N functions $f_i(x_i)$, each of them using only a disjoint part of vector $\mathbf{x}$, we can formulate the problem as follows,

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^{N} f_i(x_i) && \text{(4-26)} \\ &\text{subject to} && \sum_{i=1}^{N} x_i = 0. && \text{(4-27)} \end{aligned}$$

The solution to this problem is given by

$$x_i^{k+1} := \arg\min_{x_i} \Big(f_i(x_i) + y^k x_i + (\rho/2)||x_i - (x_i^k - \bar{x}^k)||_2^2 \Big) \tag{4-28}$$

$$y^{k+1} := y^k + \rho \bar{x}^{k+1}, \tag{4-29}$$

where $\bar{x}^k = \sum_{i=0}^{N} x_i^k$. This means that in each iteration one must minimize the augmented Lagrangian for each of the x_i, $\forall i = 1, ..., N$, and then update the duals.

Note that, in the canonical problem, all variables are connected with one another through the constraint $\sum_{i=1}^{N} x_i = 0$. This constraint can be generalized to one of the form $\mathbf{Ax} = \mathbf{b}$. Moreover, the variable x_i is scalar and can also be generalized to a vector, $\mathbf{x}_i$. A specific realization of this generalization can be found in the last chapter, Chapter 8.

Chapter 5

Intensity-based Wavefront Reconstruction Using B-Splines

In this chapter, a novel wavefront reconstruction method will be presented. This method uses the intensity measurements collected by a detector from a (Shack-)Hartmann wavefront sensor and tries to locally reconstruct the phase sampled by a subaperture based on the diffraction pattern collected in the detection plane. The B-splines presented in Chapter 3 will be used to parametrize the wavefront. We will make use of their inherent local structure and of the smoothness matrix presented in Section 3-4.

The problem is first formulated in Section 5-2-1. Then, in Section 5-2-3, we define it for a local subaperture, and finally in Section 5-2-3 the global problem is presented. After the problem is completely formulated, numerical simulations emphasizing the effect that different parameters have on the quality of the reconstructed wavefront will be presented in Sections 5-4-1 to 5-4-4. A comparison with the modal reconstruction method [47] using a simple slope calculation technique [30] is presented in Section 5-4-6 in an open-loop scenario. The results in closed-loop using a classical AO feedback loop [74, 23] and a perfect deformable mirror are presented separately in Section 5-4-7.

5-1 Introduction

The Hartmann (resp. Shack-Hartmann) wavefront sensor consists of an array of apertures (resp. lenslets[1]) that sample the incoming wavefront. A centroiding algorithm (Sec. 2-2-2) then provides an approximation of the spatial slope of the wavefront for each aperture. This algorithm computes the center of mass of the intensity measurements collected by a detector (*e.g.*, CCD camera), thus neglecting information present in the intensity patterns. This will, in general, cause a loss in accuracy of the wavefront reconstruction results.

[1] Given the applicability of this method both to the Hartmann and the Shack-Hartmann wavefront sensors, we will designate them both by referring to them abstractly as wavefront sensors, for the sake of brevity and simplicity. Both the lenslets and the apertures will be designated as apertures, also for the sake of brevity.

In order to preserve the main advantage of the centroid based wavefront reconstruction, that is, the linearity of the wavefront reconstruction problem, but to make direct use of the measured intensities, as Polo did in [12], without first approximating the spatial slopes, a new wavefront reconstruction method is presented in this section. The method is based on the integration of two principles. The first is a physical principle where we perform a local linearisation of the relationship between the local wavefront aberrations and the intensity measurements of the aperture that "sees" this local wavefront. The second is a numerical principle on the use of B-splines to parametrize and reconstruct the unknown wavefront, given that they are specially suited to reconstruct smooth wavefronts. The methodology behind the parametrization is based on our recent work in [5, 52] .

5-2 Operating Principle

The classical approaches mentioned in Sections 2-3-1 and 2-3-2 use the recorded intensities behind each aperture to extract an approximation of the local spatial gradient of the wavefront. In this section, we propose that all the recorded intensity measurements are used in the wavefront reconstruction as to use all the information possible to estimate the phase of the wavefront. In the following subsection, we derive the model that relates the wavefront and the intensity measurements as well as the local linearisation of that model. Furthermore, we parametrize the wavefront using B-splines and define the phase-retrieval problem as a least-squares problem subject to linear constraints.

5-2-1 General Intensity-based Phase Retrieval Algorithm

In order to obtain a relation between the phase distribution of the wavefront and the intensity distribution measured by the wavefront sensor, we must start by providing some definitions. Let the complex field $U(x, y, z)$ define the wavefront at a certain distance z from the sampling sensor array. The complex field of the wavefront immediately after being transmitted $U(x, y, 0)$ by the sampling array can be described by its amplitude $A(x, y)$ and its unknown phase distribution $\Phi(x, y)$ as

$$U(x, y, 0) = A(x, y)\Phi(x, y),$$

where the spatial coordinates (x, y) represent the pupil plane where the wavefront sensor is located. Considering that the sampling array and the detection plane are separated by a distance L, we can define the complex field at $z = L$ by $U(u, v, L)$. Notice that in this plane we have a different pair of coordinates, (u, v). The following expressions describe these two quantities:

Incoming Wavefront

$$U(x,y,0) = A(x,y)\exp(i\nu\Phi(x,y)), \qquad (5-1)$$

where ν represents the wavenumber.

Wavefront after propagation (Hartmann sensor)

$$U(u,v,L) = \mathscr{F}^{-1}\left[\mathscr{F}[U(x,y,0)]H(f_x,f_y)\right], \qquad (5-2)$$

where $H(f_x,f_y)$ represents the Rayleigh-Sommerfeld transfer function[2]and f_x and f_y are the spatial frequencies in x and y directions respectively.

Wavefront at the focal plane (SH sensor)

$$U(u,v,L) = \frac{e^{ikz}}{i\lambda z}e^{\frac{ik}{2z}(u^2+v^2)}\mathscr{F}\left[U(x,y,0)\right]. \qquad (5-3)$$

The derivations of these formulas can be consulted in Appendix A-1. More specifically one should note that Eqs. (5-2) and (5-3) correspond to Eqs. (A-14) and (A-18), respectively. The Fourier transform and its inverse were introduced to ease the implementation of the simulation.

The phase ϕ can be parametrized as a linear combination of K general basis functions f_k weighted by a vector of weighting coefficients $\boldsymbol{\alpha}$ as

$$\phi(x,y) = \sum_{k=1}^{K}\alpha_k f_k(x,y) = \begin{bmatrix} f_1(x,y) & \cdots & f_K(x,y)\end{bmatrix}\boldsymbol{\alpha}. \qquad (5\text{-}4)$$

These coefficients α_k can be estimated by minimizing the error between the intensity of the field given by the model and the measured intensity. Let $I_{\text{meas}}(u_i,v_j)$ represent the measured intensities at each discrete pixel (i,j) at the detection plane and let

$$I(u_i,v_j,L) := |U(u_i,v_j,L)|^2 = \left|\mathscr{F}^{-1}\left[\mathscr{F}[U(x,y,0)]H(f_x,f_y)\right]\right|^2_{(i,j)} \qquad (5\text{-}5)$$

be the intensity given by the physical model also evaluated in pixel (i,j). With these two quantities one is able to define a simple cost function $J(\boldsymbol{\alpha})$. When minimizing $J(\boldsymbol{\alpha})$ we try to match the intensities given by the model we constructed against the experimental data. The cost function is defined as follows:

$$J(\boldsymbol{\alpha}) = \sum_{i,j}\left[I_{\text{meas}}(u_i,v_j) - I(\boldsymbol{\alpha},u_i,v_j,L)\right]^2. \qquad (5\text{-}6)$$

In [12] a nonlinear least squares optimization procedure was performed to estimate the coefficients α_k that describe the phase distribution ϕ in (5-4). The non-linear optimization method presented has no guarantees of arriving at a global minimum and, besides that, it is very computationally demanding. In the aforementioned article the convergence time of the algorithm was 10 s which is not real-time implementable.

In the next section, we linearise the model so that the phase-retrieval problem can be solved using a least-squares approach instead of a non-linear method. With this linearisation we guarantee that the global optimum is achieved and that the running time is reduced, given the linearity of the method.

[2] Regarding the Rayleigh-Sommerfeld transfer function, the reader should refer to [26, § 3.5.2] for the full theoretical derivation, and consult [75, § 4.4.1] for implementation details. In the Appendix A-1 a concise summary of the most important results from both references is presented, as well.

5-2-2 Formulation of the Linearised Problem

Firstly, both $I_{\text{meas}}(u_i, v_j)$ and $I(u_i, v_j, L)$ in (5-6) are vectorized such that, given a total of M pixels, two vectors $\mathbf{i}_{\text{meas}}$ and $\mathbf{i}_L \in \mathbb{R}^{M\times 1}$ are created. Thus, a new cost function J_{vec} can then be defined as

$$\begin{aligned} J_{\text{vec}}(\boldsymbol{\alpha}) &= \sum_{m=1}^{M} \left(I_{\text{meas}}(m) - I(\boldsymbol{\alpha}, m, L)\right)^2 && \text{(5-7)} \\ &= ||\mathbf{i}_{\text{meas}} - \mathbf{i}_L||_2^2. && \text{(5-8)} \end{aligned}$$

Note that each element m of the newly created vectors has a direct mapping to a point (u_i, v_j) in the Cartesian plane such that we can define $(\mathbf{i}_{\text{meas}})_m = I_{\text{meas}}(m)$ and $(\mathbf{i}_L)_m = I(\boldsymbol{\alpha}, m, L)$.

In order to find the coefficients $\boldsymbol{\alpha}$ that minimize J_{vec} using a linear method, we linearise the intensity term $I_L(m) = I(\boldsymbol{\alpha}, m, L)$ defined at a certain pixel m using a first-order Taylor series around an arbitrary vector of phase coefficients $\tilde{\boldsymbol{\alpha}}$. For this application, we will use $\tilde{\boldsymbol{\alpha}} = 0$.

Computing the derivative of the intensities $I_L(m) = U_L^*(m)U_L(m)$ by

$$\frac{\partial I_L(m)}{\partial \boldsymbol{\alpha}} = \left[\frac{\partial U_L^*(m)}{\partial \boldsymbol{\alpha}} U_L(m) + U_L^*(m)\left(\frac{\partial U_L(m)}{\partial \boldsymbol{\alpha}}\right)\right], \tag{5-9}$$

the truncated first-order Taylor expansion of $I_L(m)$ can then be expressed as follows:

$$\begin{aligned} I_L(m) &\approx I_L(m)_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}} + \left[\frac{\partial U_L^*(m)}{\partial \boldsymbol{\alpha}} U_L(m) + U_L^*(m)\frac{\partial U_L}{\partial \boldsymbol{\alpha}}(m)\right]_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}} (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}) \\ &:= c_{0m} + \mathbf{c}_{1m}^{\top}(\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}), \end{aligned} \tag{5-10}$$

where

$$\frac{\partial U_L(m)}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} \frac{\partial U_L(m)}{\partial \alpha_1} & \cdots & \frac{\partial U_L(m)}{\partial \alpha_K} \end{bmatrix} \in \mathbb{R}^{1\times K}, \tag{5-11}$$

with

$$\frac{\partial U_L(m)}{\partial \alpha_k} = \mathscr{F}^{-1}\left[\mathscr{F}[U_0(x_i, y_j) i\nu f_k(m)]H(m)\right]. \tag{5-12}$$

The terms $c_{0m} \in \mathbb{R}$ and $\mathbf{c}_{1m} \in \mathbb{R}^{K\times 1}$ in Eq. (5-10) are computed for each $m = 1, ..., M$ in order to define $\mathbf{c_0} = [c_{01}, ..., c_{0M}]^{\top} \in \mathbb{R}^{M\times 1}$ and $\mathbf{C_1} = [\mathbf{c}_{11}^{\top}, ..., \mathbf{c}_{1M}^{\top}] \in \mathbb{R}^{M\times K}$.This new formulation allows for the conversion of the non-linear problem in (5-6) into a linear least-squares optimization:

$$\underset{\boldsymbol{\alpha}}{\text{minimize}}\, J_{\text{lin}}(\boldsymbol{\alpha}), \tag{5-13}$$

with $J_{\text{lin}}(\boldsymbol{\alpha})$ defined as

$$J_{\text{lin}}(\boldsymbol{\alpha}) = ||\mathbf{g} - \mathbf{C}_1\boldsymbol{\alpha}||_2^2. \tag{5-14}$$

where $\mathbf{g} = \mathbf{i}_{\text{meas}} - \mathbf{c}_0$. If the phase coefficients $\boldsymbol{\alpha}$ have unknown statistical properties, the optimal solution for this problem is then given by the phase coefficient estimator $\hat{\boldsymbol{\alpha}}$ as follows:

$$\hat{\boldsymbol{\alpha}} = (\mathbf{C}_1^{\top}\mathbf{C}_1)^{-1}\mathbf{C}_1^{\top}\mathbf{g}. \tag{5-15}$$

The information in this section together with Chapter 3 on B-splines provide the framework to solve the problem at hand.

5-2-3 Subaperture Local Problem

In this phase retrieval method, we propose that both the pupil and detection planes of the wavefront sensor are divided in N (where N is the number of subapertures) equal and adjacent square regions. Each cell in the pupil plane contains one subaperture. Each cell in the detection plane corresponds to the area that "sees" the local wavefront sampled by a subaperture. We will use the data from the subimage on the detection plane to estimate the phase only in its corresponding subaperture.

Provided that the diffraction effects on the detection plane are small, one can guarantee that the propagated beams from one subaperture have a minimal effect in the subimage corresponding to another subaperture. Therefore, the optimization procedure that can be defined when minimizing J_{lin} in Eq. (5-13) can be applied locally and almost independently to estimate the phase distribution in each of the subapertures.

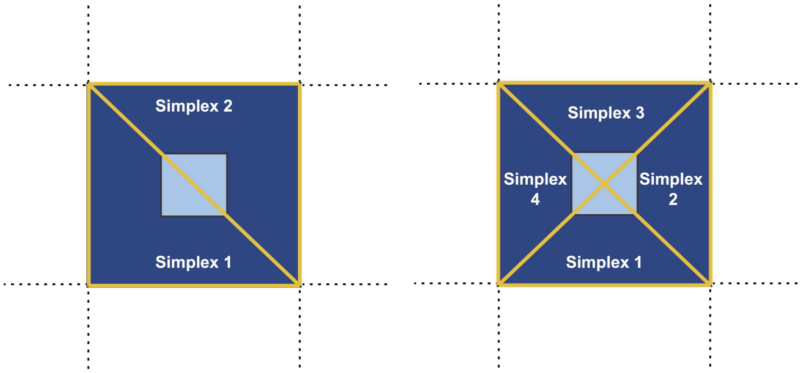

Figure 5-1: Close-up on one of the subapertures. In the left (right) figure, a Type I (II) triangulation [5, Section 2.3] was defined using 2 (4) simplices per subaperture.

The local areas that have been defined in both planes are square, whereas the B-splines are defined in a simplex with a triangular shape. Therefore, the phase distribution was parametrized using T triangles that divide the subaperture. An example with 2 (Type I triangulation) and 4 (Type II triangulation) triangles can be seen in Figure 5-1.

In order to ensure continuity in the same subaperture one may enforce the continuity constraints (Eq. (3-15)) inherent to the spline framework. The word may is used as these constraints are not nearly as important as the ones that connect the different subapertures which will be presented in the next section. The formulation of the local problem is presented in Eq. (5-16) for a certain subaperture n:

$$\begin{aligned} &\text{minimize} \, ||\mathbf{g}_n - \mathbf{C}_{1,n}\boldsymbol{\alpha}_n|| \\ &\text{subject to} \quad \mathbf{H}^{\text{local}}\boldsymbol{\alpha}_n = 0, \end{aligned} \tag{5-16}$$

where

$$\mathbf{g}_n = \begin{bmatrix} \mathbf{g}_n^1 \\ \vdots \\ \mathbf{g}_n^T \end{bmatrix}, \quad \mathbf{C}_{1,n} = \begin{bmatrix} \mathbf{C}_{1,n}^1, & \cdots & , \mathbf{C}_{1,n}^T \end{bmatrix}, \quad \boldsymbol{\alpha}^n = \begin{bmatrix} \boldsymbol{\alpha}_n^1 \\ \vdots \\ \boldsymbol{\alpha}_n^T \end{bmatrix}.$$

The vector $\mathbf{g}_n$ is defined in $\mathbb{R}^{M_n \times 1}$, where M_n denotes the number of pixels in the detector surface that correspond to the aperture n. Defining $\hat{d} = \frac{(2+d)!}{2d!}$ as the number of basis functions coefficients per simplex, we have that $\boldsymbol{\alpha}_n$ is defined in $\mathbb{R}^{T\hat{d} \times 1}$ and matrix $\mathbf{C}_{1,n}$ in $\mathbb{R}^{M_n \times T\hat{d}}$.

Each of the submatrices $\mathbf{C}^t_{1,n} \in \mathbb{R}^{M_n \times \hat{d}}$ is the Jacobian associated with the coefficients $\boldsymbol{\alpha}^j_n$ from the simplex t. The vector $\mathbf{g}^t_n \in \mathbb{R}^{M_t \times 1}$ has information from the M_t pixels belonging to simplex t.

Global Optimization

The natural extension to Eq. (5-16) is to perform this optimization for all the subapertures simultaneously. For the optimization to be successful it is not enough to purely replicate the procedure presented in Eq. (5-13). Additional boundary constraints must be inserted in addition to those that provide continuity between the two neighbouring polynomials in each of the subapertures.

Let us provide an example. Take, for instance, the upper simplex in the left subaperture in Figure 5-1. Assuming that the adjacent subapertures have the same simplex layout, continuity constraints should be imposed to connect the aforementioned simplex with the lower simplices of the subapertures which are located upwards and to the right.

These constraints should be included in the global smoothness matrix $\mathbf{H}$ so as to generate a global problem for N subapertures. In the end, the global smoothness constraint matrix $\mathbf{H}$ will contain the constraints in $\mathbf{H}^{\text{local}}$ from (5-16) and the new constraints that regulate continuity between neighbouring subapertures.

The global optimization problem is then defined as follows:

$$\begin{aligned} &\text{minimize}\, ||\mathbf{g} - \mathbf{C}_1\boldsymbol{\alpha}||_2^2 \\ &\text{subject to} \quad \mathbf{H}\boldsymbol{\alpha} = 0 \end{aligned} \tag{5-17}$$

where,

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_N \end{bmatrix}, \quad \mathbf{C}_1 = \begin{bmatrix} \mathbf{C}_{1,1} & & \\ & \ddots & \\ & & \mathbf{C}_{1,N} \end{bmatrix}, \boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_N \end{bmatrix}.$$

The global B-coefficient vector $\boldsymbol{\alpha}$, the linearisation offset $\mathbf{c}_0$ and the intensity measurement vector $\mathbf{i}_{\text{meas}}$ are simply created by stacking their local elements in the appropriate order. Due to the local linear relation between one partition in the imaging plane and the phase sampled by a subaperture, $\mathbf{C}_1 \in \mathbb{R}^{NM \times NK}$ is block diagonal, where M is the number of intensity measurements in one partition and K is the number of B-coefficients that describe the phase in a subaperture. The blocks of $\mathbf{C}_1$ are the local Jacobians $\mathbf{C}_{1,n} \in \mathbb{R}^{M \times T_s\hat{d}}$, where $\hat{d} = \frac{(2+d)!}{2d!}$ is the number of B-coefficients to be estimated per spline and T_s the number of simplices per subaperture. The global smoothness matrix $\mathbf{H} \in \mathbb{R}^{QE \times T_sN\hat{d}}$, where Q represents the continuity conditions per simplex edge and E represents the number of edges between the simplices, contains the equality constraints to guarantee order r continuity between the simplices.

The solution for this problem can be found by solving the following matrix inversion problem [20] that satisfies the Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{bmatrix} \mathbf{C}_1^\top \mathbf{C}_1 & \mathbf{H}^\top \\ \mathbf{H} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^\top \mathbf{g} \\ 0 \end{bmatrix}, \tag{5-18}$$

where the vector $\mathbf{y}$ represents the dual variables associated with the constraints specified by matrix $\mathbf{H}$. This vector does not need to be computed as it has no influence on the phase coefficients $\boldsymbol{\alpha}$.

One way to solve this problem is by first eliminating the equality constraints by working with a transformed and reduced set of coefficients $\boldsymbol{\alpha}$ that satisfies $\mathbf{H}\boldsymbol{\alpha} = 0$. This is accomplished by using the nullspace of the constraint matrix[3], $\text{Null}(\mathbf{H}) \in \mathbb{R}^{NT_s\hat{d} \times N_f}$, where N_f represents the degrees of freedom that the coefficients still have if the equation $\mathbf{H}\boldsymbol{\alpha} = 0$ is respected. The transformation that retrieves the full coefficient vector from its reduced version $\boldsymbol{\alpha}^{\text{red}}$ is given by

$$\boldsymbol{\alpha} = \text{Null}(\mathbf{H})\boldsymbol{\alpha}^{\text{red}},$$

which leads to the following linear least-squares *unconstrained* problem:

$$\text{minimize} \left|\left| \mathbf{g} - \mathbf{C}_1^{\text{red}} \boldsymbol{\alpha}^{\text{red}} \right|\right|_2^2, \tag{5-19}$$

where $\mathbf{C}_1^{\text{red}} = \mathbf{C}_1 \text{Null}(\mathbf{H})$. The optimum solution for this problem is given by

$$\boldsymbol{\alpha}^\star = \text{Null}(\mathbf{H}) \big((\mathbf{C}_1^{\text{red}})^\top \mathbf{C}_1^{\text{red}} \big)^{-1} (\mathbf{C}_1^{\text{red}})^\top \mathbf{g}. \tag{5-20}$$

Given that it is not guaranteed that $(\mathbf{C}_1^{\text{red}})^\top \mathbf{C}_1^{\text{red}}$ can be inverted, the solution should be found using QR (see Section C-2-1) or SVD decomposition. The statistical properties of this estimator can be consulted in Appendix B.

5-3 Closed-loop Convergence Analysis

In [13], Smith *et al.* proposed a very simple, but effective, method to correct phase aberrations in real-time also using an intensity-based algorithm. The iterative estimation at each time step $t + 1$ is given by

$$\boldsymbol{\alpha}(t+1) = \boldsymbol{\alpha}(t) - \hat{\boldsymbol{\alpha}}(t), \tag{5-21}$$

with $\boldsymbol{\alpha}(t + 1)$ and $\boldsymbol{\alpha}(t)$ representing the residual phase at time step $t + 1$ and t, respectively, and $\hat{\boldsymbol{\alpha}}(t)$ denoting the estimate of the true coefficients $\boldsymbol{\alpha}(t)$ at time k.

Smith *et al.* also present a bound on the relative residual error for a wavefront reconstruction algorithm that is also based on intensity measurements. If this bound has a value smaller than 1, the convergence of the algorithm is guaranteed. The method and its proof are supported by three main implicit assumptions.

Assumption 1. *The deformable mirror is perfect, i.e., the estimated phase can be perfectly represented by the mirror.*

Assumption 2. *The intensities are not affected by any measurement noise in the sensor.*

[3] A more detailed explanation of the nullspace projection method is presented in the context of distributed wavefront reconstruction, in Section 8-3, where more significant transformations have to occur.

Assumption 3. *No delay is present in the loop. The estimate at time step t affects the residual wavefront also at time step t.*

This convergence proof can be very easily extrapolated to prove firstly the convergence of the central and then of the distributed methods. For the reduced centralized problem in (5-19) the bound on the relative error follows the following expression:

$$\frac{||\hat{\boldsymbol{\alpha}}^{\mathrm{red}}(t) - \boldsymbol{\alpha}^{\mathrm{red}}(t)||}{||\boldsymbol{\alpha}^{\mathrm{red}}(t)||} \lesssim \frac{||\Delta\mathbf{i}(\boldsymbol{\alpha}(t))||}{||\mathbf{g}(t)||}\left\{\frac{2\mathrm{cond}(\mathbf{C}_1^{\mathrm{red}})}{\cos(\theta)} + \tan(\theta)\mathrm{cond}(\mathbf{C}_1^{\mathrm{red}})^2\right\}, \tag{5-22}$$

where the quantity $\Delta\mathbf{i}(\boldsymbol{\alpha}(t))$ represents the higher order terms which were neglected when the linearisation was performed. The matrix $\mathbf{C}_1^{\mathrm{red}}$ and the vector $\mathbf{g}(t)$ represent the leftmost matrix and the rightmost vector in (5-18). The angle θ represents the acute angle between the vectors $\mathbf{C}_1^{\mathrm{red}}\hat{\boldsymbol{\alpha}}(t)$ and $\mathbf{g}(t)$. The operator $\mathrm{cond}(\cdot)$ computes the condition number of the matrix given as an argument (see Section 1-5).

In order to better characterize this bound, the term $\Delta\mathbf{i}(\boldsymbol{\alpha}(t))$ that represents the error made when linearising the model must also be bounded. If the function that we are approximating (in this case, the expression that relates the intensity distribution with the phase aberration in (5-5)) is twice differenciable on its domain, we can analytically compute the Lagrange remainder K_{lag} [76, § 7.7] which leads to the following definition of the modelling errors:

$$\Delta\mathbf{i}(\boldsymbol{\alpha}(t)) = K_{\mathrm{Lag}}||\boldsymbol{\alpha}^{\mathrm{red}}(t)||^2, \tag{5-23}$$

where K_{Lag} is the Lagrange remainder. Hence, Eq. (5-22) can be rewritten as

$$\frac{||\hat{\boldsymbol{\alpha}}^{\mathrm{red}}(t) - \boldsymbol{\alpha}^{\mathrm{red}}(t)||}{||\boldsymbol{\alpha}^{\mathrm{red}}(t)||} \lesssim \frac{K_{\mathrm{Lag}}||\boldsymbol{\alpha}^{\mathrm{red}}(t)||^2}{||\mathbf{g}(t)||}\left\{\frac{2\mathrm{cond}(\mathbf{C}_1^{\mathrm{red}})}{\cos(\theta)} + \tan(\theta)\mathrm{cond}(\mathbf{C}_1^{\mathrm{red}})^2\right\}. \tag{5-24}$$

For very small aberrations (that is, taking the limit when $||\boldsymbol{\alpha}(t)|| \to 0$) and for a non-singular Jacobian matrix $\mathbf{C}_1^{\mathrm{red}}$, it is clear that the expression in (5-24) is less than unity, given that the upper bound tends to zero.

The assumptions stated in the beginning of the section can also be relaxed as follows. Assumption 1 can be relaxed in order to cope with the lack of precision that a non-perfect mirror introduces by adding two terms: the first term is also a Lagrange remainder, which bounds the error made by choosing a linear model for the DM; the second term is an error that is caused by the inherent limitations of the mirror in terms of its approximation power, *i.e*, even if the mirror is modelled perfectly it will be unable to describe the reconstructed phase with infinite precision. These errors should be added to the model error $\Delta\mathbf{i}(\boldsymbol{\alpha}(t))$.

Regarding Assumption 2, if we consider the noise to be modelled by a Gaussian distribution, guarantees of monotonic convergence according to this criterion are impossible, given the small but existing possibility of having very large noise perturbations. However, a truncated Gaussian probability distribution can be used, and the worst-case scenario can be added to the model error and thus included in the bound calculation.

For relaxing Assumption 3, a controller is introduced in the loop that minimizes the effect of the delay, such that we can approximately consider that (5-21) is satisfied, that is, the incoming wavefront at time instant t is subtracted the estimation calculated also at time instant t.

5-4 Numerical Results

The nominal parameters of the wavefront sensing setup simulated in this section are similar to the ones presented in [12]. A Hartmann wavefront sensor was simulated where each hole has a side length of 200 µm and is separated from the adjacent hole by a distance of 562.5 µm and the wavelength is $\lambda = 638$ nm. The pupil plane configuration and the intensity distribution for an unaberrated wavefront are depicted in Figure 5-2.

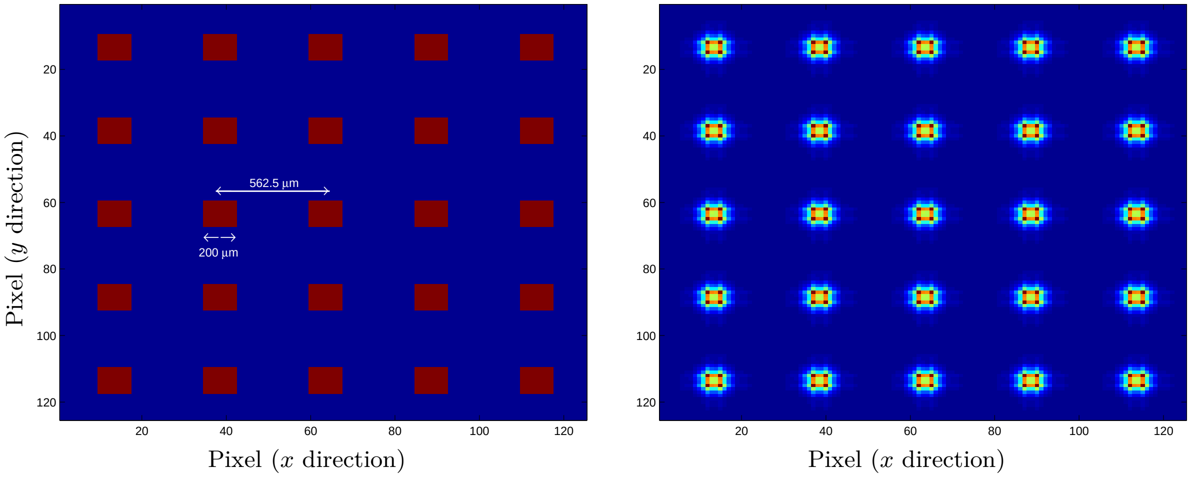

Figure 5-2: (Left) Mask used to sample the wavefront. The red squares represent the subapertures and the blue background is assumed to be completely opaque. (Right) Intensity distribution (aberration-free) captured in the detection plane given the sampling performed by the mask on the left part of the figure.

The sensor was considered to suffer predominantly the effects of read-out noise on the CCD camera. This noise was modelled as white and Gaussian with zero-mean and standard deviation σ_{ccd}. The noise is additive and affects each of the normalized (between 0 and 1) intensity distribution values.

The aberrations introduced will be modelled by Zernike polynomials or Kolmogorov turbulence models [78]. The strength of those aberrations in terms of their RMS value will be measured by the coefficient γ. This coefficient corresponds to the 2-norm of the Zernike coefficients vector and is also used to depict the RMS value of Kolmogorov turbulence aberrations. The values of γ used are between $10^{-6}\lambda$ [rad] and $10^{2}\lambda$ [rad].

In the following sections, an analysis regarding the influence of simulation parameters and setup configuration will be made. Afterwards, comparative results with the modal reconstruction method are presented in Section 5-4-6 in an open-loop setup which involves no deformable mirror but only the wavefront sensing and reconstruction. Finally, in Section 5-4-7, the closed-loop results using a perfect deformable mirror are presented followed by some final remarks.

5-4-1 Triangulation Selection

The type of triangulation used can influence significantly the quality of the estimation. In Table 5-1, it can be seen that Type II triangulations are in general preferable to Type I triangulations, as using them to parametrize the wavefront yields smaller reconstruction errors. The superior reliability and robustness of Type II regarding Type I triangulations had been previously asserted in Chapter 3 based on the results achieved in [5, § 3] concerning the propagation factor.

Table 5-1: Comparison between the RMS values of the reconstruction errors for Type I and Type II triangulations using different data related parameters (number of subapertures and strength of the aberration). The RMS errors presented were averaged over 10 different realizations of a Kolmogorov turbulence screen. The normalized intensity values were corrupted with zero-mean, white, Gaussian noise with a standard deviation $\sigma = 4 \times 10^{-4}$. The B-splines used had degree $d = 2$ and continuity $r = 1$.

Grid	$[2 \times 2]$		$[5 \times 5]$		$[10 \times 10]$	
Tri.	Type I	Type II	Type I	Type II	Type I	Type II
λ	3.96×10^{-1}	6.99×10^{-1}	1.99×10^{-1}	1.71×10^{-1}	4.51×10^{-1}	1.78×10^{-1}
0.1λ	2.78×10^{-2}	5.10×10^{-2}	1.27×10^{-2}	1.12×10^{-2}	1.24×10^{-2}	8.94×10^{-3}
0.01λ	8.05×10^{-3}	8.16×10^{-3}	3.35×10^{-3}	1.33×10^{-3}	1.43×10^{-3}	1.18×10^{-3}

However, there is an exception. For small grids (less than three subapertures per side) the Type I triangulation provides better estimates than Type II.

Given that our main focus will be to work with grids larger than $[10 \times 10]$, we will use Type II triangulations in the remaining sections of the Chapter, except when clearly stated otherwise.

5-4-2 Propagation distance Analysis

An assumption was implicitly made in Section 5-2-3, where it is said that the local linearisation is only possible if the diffraction pattern is almost completely contained in the image plane partition corresponding to the subaperture. If the diffraction pattern from one subaperture is such that it interferes with the pattern created by a neighbouring subaperture the assumption no longer holds and the method starts to fail.

In Figure 5-3, several intensity profiles are compared with one another. It is clear that, when the imaging plane is at a bigger distance, the interference between the partitions in the imaging plane is more prominent leading to higher intensity values around sample 26, where the border between the two partitions lies.

The results in Figure 5-3 show that for a similar setup as the one used in [12], this method is only applicable for much smaller propagation distances ($z < 10$ mm) as they allow for a less spread out (and thus, less interfering) diffraction pattern.

Notice that the row of pixels that goes trough the center of the diffraction pattern does not necessarily correspond to the row where the maximum occurs.

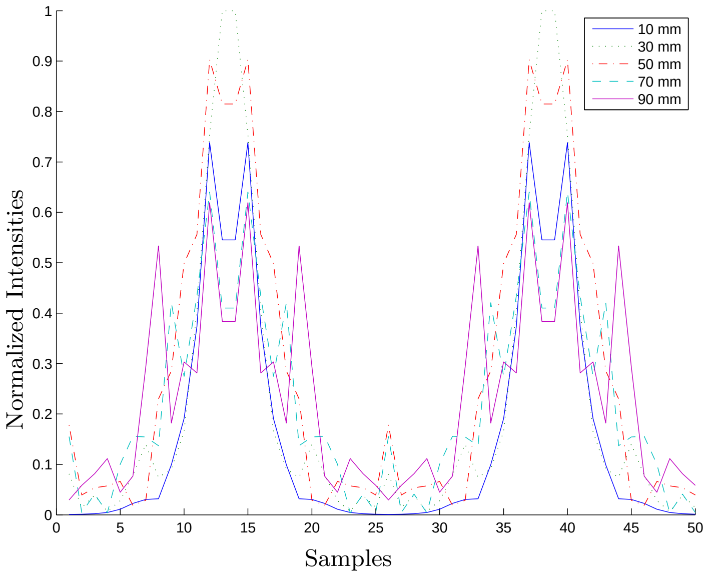

Figure 5-3: Depiction of the row of pixels that goes through the geometrical center of the diffraction pattern of the normalized intensity distribution. This intensity profile was computed for propagation distances z of 10, 30, 50, 70 and 90 mm.

5-4-3 Influence of the Number of Subapertures

The effect of having a different grid size subject to the same aberrations was tested in this section. We chose four different grid sizes and three different aberration sizes. All the aberrations are a defocus 5-th order Zernike polynomial. The results are presented in Table 5-2.

Table 5-2: Comparison between the RMS values of the reconstruction errors for a different number of subapertures using different aberrations. The RMS errors presented were averaged over 10 different realizations of a defocus modelled by a 5-th order Zernike polynomial (according to Noll's notation). The normalized intensity values were corrupted with zero-mean, white, Gaussian noise with a standard deviation $\sigma_{\text{ccd}} = 4 \times 10^{-4}$. The B-splines used had degree $d = 2$ and continuity $r = 1$. The number of pixels corresponding to each subaperture is given by $M_s = 25$.

Grid	$[2 \times 2]$	$[5 \times 5]$	$[10 \times 10]$
λ	1.11×10^{-1}	2.90×10^{-2}	5.55×10^{-3}
0.1λ	7.67×10^{-4}	6.59×10^{-4}	4.09×10^{-3}
0.01λ	4.64×10^{-4}	3.84×10^{-4}	4.06×10^{-4}

When increasing the number of subapertures, we sample more information from the wavefront and are able to reconstruct it more effectively. Such trend is clearly visible on the data from Table 5-2 where for all aberration strengths the reconstruction error decreases as the grid size increases.

Unfortunately, the fact that the centralized method does not allow for grids bigger than $[12 \times 12]$, due to memory constraints, restrains our ability to draw more general conclusions, specifically for large-scale systems.

5-4-4 Influence of the Number of Pixels

In the following paragraphs, a tentative analysis of the results presented in Figure 5-4 will be presented regarding the influence of the number of pixels on the quality of the linearisation and hence, how their number influences the reconstruction error.

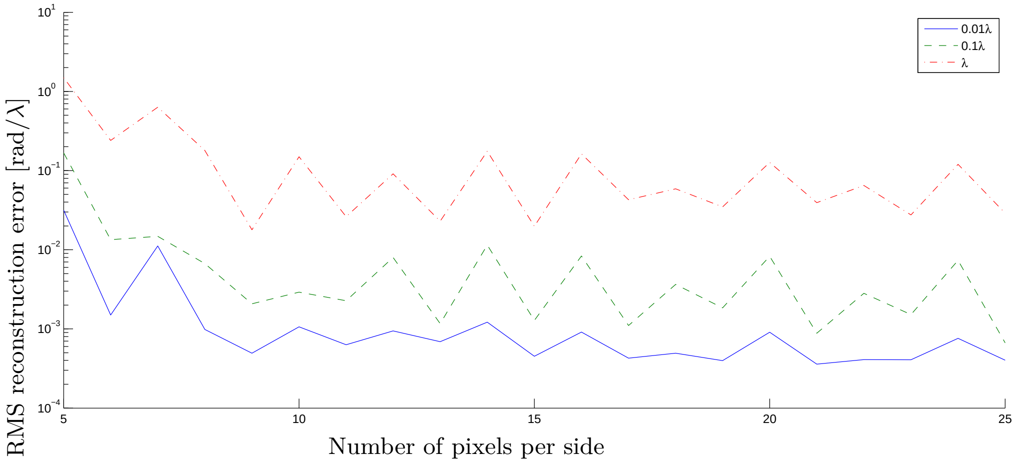

Figure 5-4: Comparison between the RMS values of the reconstruction errors for a different number of pixels per subaperture using aberrations with different magnitudes. The RMS errors presented were averaged over 10 different realizations of a defocus modelled by a 5-th order Zernike polynomial (according to Noll's notation). The normalized intensity values were corrupted with zero-mean, white, Gaussian noise with a standard deviation $\sigma_{\text{ccd}} = 4 \times 10^{-4}$. The grid used had $[5 \times 5]$ subapertures. The B-splines used had degree $d = 2$ and continuity $r = 1$.

The most clear trend is that the reconstruction error tends to decrease for an increasing resolution. Notice that the difference between the results for the three different aberration strengths from the $[5 \times 5]$ grid to the $[25 \times 25]$ one is of almost two orders of magnitude.

However, the error does not decrease monotonically with the increasing resolution as it can be seen from the erratic profile of the curves, specially for $\gamma = \lambda$ [rad] (red line) and $\gamma = 0.1\lambda$ [rad] (green line). One possible reason for this phenomenon can be attributed to quantization errors that stem from the discretization of the intensity distribution. Some particular discretizations may sample the signal in such way that important information is lost and leads to a poor reconstruction. This erratic behaviour can not be due to the presence of noise given that the results were averaged over 10 different distributions and the behaviour is consistent for all 3 aberrations (*i.e.*, there is a consistency in the peaks and valleys of the curves).

5-4-5 Noise Influence

The influence of the noise in the reconstruction can be seen clearly in Figure 5-5.

It is clear that for a high noise power ($\sigma_{\text{ccd}} \geq 10^{-4}$), the reconstruction yields progressively worse results as the noise is increased. For noise powers smaller than 10^{-5} the reconstruction performance does not improve further, which means that for the aberration specified, the

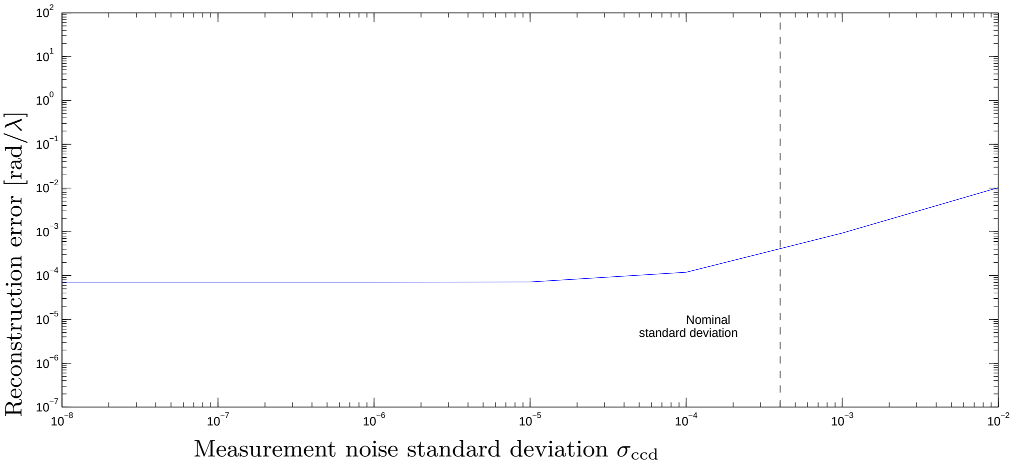

Figure 5-5: The grid used had $[10 \times 10]$ subapertures. The aberration was a 4-th order Zernike aberration $\alpha_4 = 0.1\lambda$. The B-splines used had degree $d = 2$ and continuity $r = 1$.

lower threshold was reached, *i.e.*, for lower noise powers the performance of the algorithm is equivalent to the noiseless case.

5-4-6 Comparison against Modal Reconstruction

In this section, the performance of our method is compared against a modal wavefront reconstruction algorithm presented in [47]. The principle of operation of this modal reconstruction method is explained in Section 2-3-2. The centroid algorithm used to compute the slope measurements is presented in [30].

Regarding the configuration of the setup, the propagation distance (*i.e.*, the distance between the aperture plane and the detection plane) is set at 10 mm. A Hartmann grid of 10 by 10 subapertures is used. The number of pixels per subaperture side is 25. The standard deviation of the noise is given by $\sigma_{\text{ccd}} = 4 \times 10^{-4}$.

The wavefronts were modelled using a 4th order Zernike mode (according to Noll's notation [77]) with varying strength and using Kolmogorov atmospheric turbulence statistics [78]. The polynomial functions of the B-splines were chosen to have varying degrees. A Type II triangulation partitions the subapertures, as in Figure 5-1, on the right.

All the aforementioned quantities are the nominal values of the parameters, and will be the ones used in the rest of this section except when explicitly stated otherwise.

The following sections will focus on the results for the noise-free and the noisy cases.

Noiseless Intensity Measurements

Firstly, let us compare the novel method with the modal reconstruction in the noiseless case, where it is assumed that no read-out noise corrupts the intensity measurements (*i.e.*, $\sigma_{\text{ccd}} = 0$). All the other parameters have their nominal values.

(d, r)	$(1, 0)$	$(2, 1)$	$(2, 2)$	$(3, 1)$	$(3, 2)$
Number of coefficients	1200	2400	2400	4000	4000
Degrees of freedom	221	143	6	583	66

Table 5-3: Degrees of freedom for different splines parameters.

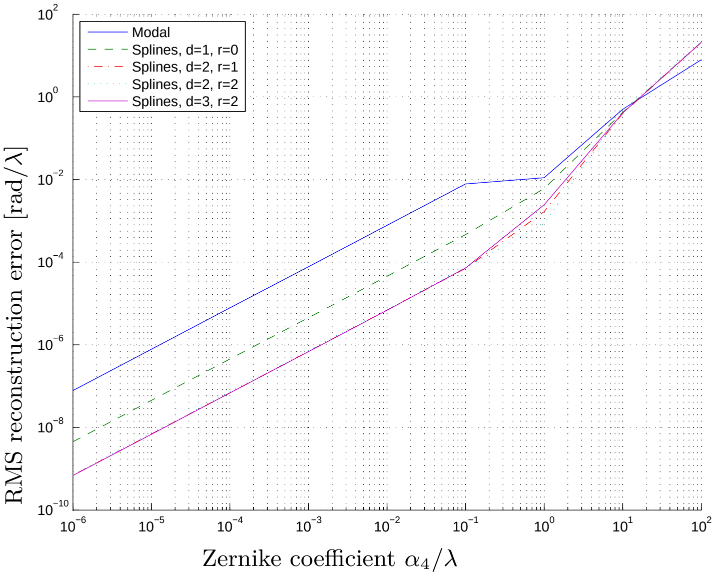

Figure 5-6: Comparison of the RMS value of the wavefront reconstruction error between the new spline based WFR method and the modal reconstruction method (in open-loop) for the noiseless case and an astigmatic Zernike aberration.

In Figure 5-6 it can be seen that for aberrations with an RMS value smaller than 10λ [rad] the splines method outperforms modal reconstruction. When the aberrations have an RMS value smaller than 0.1λ [rad] the improvement can be up to two orders of magnitude depending on the splines parameter chosen.

Comparing the results for the different splines parameters, we observe that the (d, r) pair which produces better results is $(2, 1)$ and $(2, 2)$, rather than $(1, 0)$. For this experiment, the correct aberration can be estimated using just 6 degrees of freedom (given that the Zernike astigmatism is described by a second-order polynomial) which explains the good results for the pair $(2, 2)$. Although the number of degrees of freedom increases from 6 to 143 for the pair $(2, 1)$, there is no noise corrupting the measurements. Therefore, the extra smoothness imposed by $(2, 2)$ that should filter out the noise has virtually no effect which leads to very similar results for both cases. In the case where $(d, r) = (1, 0)$, the approximating power of the polynomials is simply not enough to reconstruct the aberration which leads to the poorer results (although still one order of magnitude better than modal reconstruction).

The Zernike aberration studied is a relatively low frequency mode. We can test the robustness of the method for a more complex aberration with higher frequency components by, *e.g.*, using a phase screen characterized statistically by the model of Kolmogorov atmospheric turbulence [78].

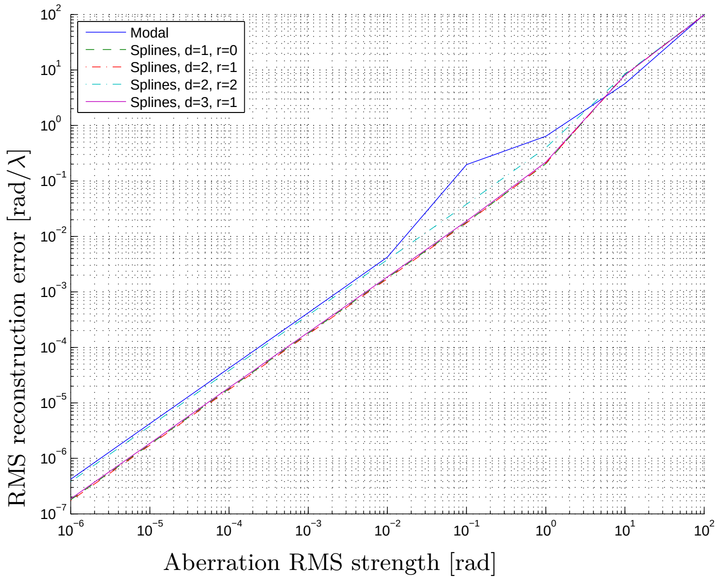

Figure 5-7: Comparison of the RMS value of the wavefront reconstruction error between the new spline based WFR method and the modal reconstruction method (in open-loop) for the noiseless case and an aberration generated using Kolmogorov atmospheric turbulence.

For the Kolmogorov case portrayed in Figure 5-7, the improvements the new method yields are less significant than the ones shown before, for low order aberrations. None of the different configurations of (d, r) can achieve more than 1 order of magnitude of difference regarding modal reconstruction.

It is interesting to notice that increasing the degree of the polynomial does not, by default, decrease the reconstruction error. That can be seen when comparing $(2, 1)$ with $(3, 1)$ and $(3, 2)$, where little or no improvements can be seen by increasing the polynomial degree.

An extra fact that must be pointed out concerns the parameters $(d, r) = (2, 2)$. This pair gave good results for the Zernike astigmatism case, however, for more complex aberrations, having only 6 degrees of freedom drastically reduces the performance. In Figure 5-7 it can be seen that the performance is very similar to modal reconstruction.

Noisy Intensity Measurements

In this section, an analysis will be made regarding the performance of the algorithm when the intensity measurements are corrupted by noise. Firstly, the results for a Zernike astigmatic aberration are shown, followed by the reconstruction of a Kolmogorov atmospheric turbulence phase screen.

The results presented in Figure 5-9 refer to an astigmatic aberration and show that our method yields an RMS error approximately 1 order of magnitude lower in relation to the modal reconstruction method in the presence of noise for aberrations smaller than λ for $d = 2$ and $r = 1$ or $r = 2$. For $d = 1$ and $r = 0$ the improvement of 1 order of magnitude is only observable for aberrations smaller than 0.1λ.

For aberrations larger than 10λ, the diffraction pattern corresponding to a subaperture will affect the intensity pattern originating from other subapertures. In that case, our locality

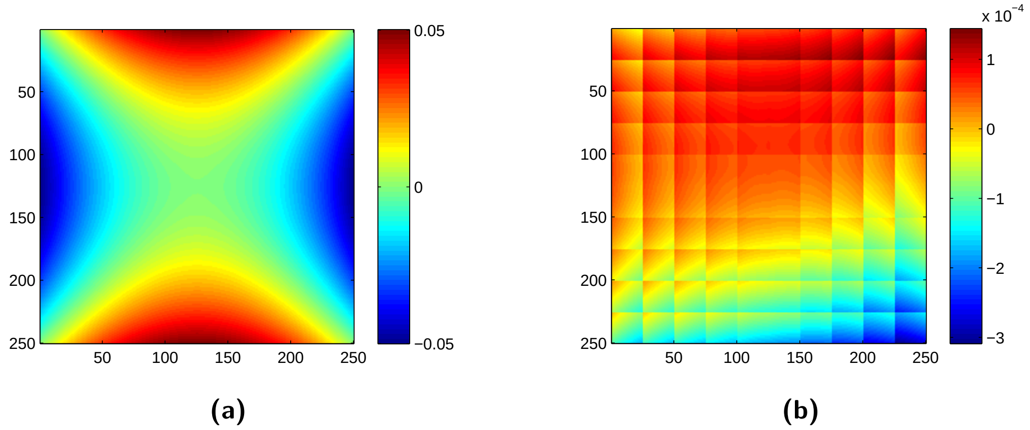

Figure 5-8: (a) Reconstructed phase distribution for $\alpha_4 = 0.1\lambda$ and (b) its error (both normalized to the wavelength λ) when the intensity distribution was subject to noise ($\sigma = 4 \times 10^{-4}$)

assumption presented in Section 5-2-3 is not verified and the method performs poorer than the modal reconstruction, specially for $\alpha_4 \geq 100\lambda$.

Regardless of the magnitude of the aberration and of the parameters that characterize the splines functions, the variance of the 100 RMS errors obtained for aberrations with different strengths, was 2 orders of magnitude higher in the modal reconstruction than in the splines reconstruction as per Figure 5-9. An empirical analysis was carried out using the characterization of the two estimators presented in Appendix B. However, the theoretical predictions did not match the practical results.

Concerning the choice of d and r, as we have imposed a smooth aberration, increasing r generally leads to better results as we are imposing a smoother phase reconstruction. Moreover, it was also verified experimentally that choosing d and r such that $d - r = 1$ yields the best results, otherwise, we would encounter overfitting problems due to the great amount of degrees of freedom.

Further analysis of Figure 5-9, reveals that, due to the influence of the noisy intensity measurements, the RMS error values reach a lower threshold when the aberrations imposed are smaller than 0.01λ. This threshold is approximately equal to $1.5 \times 10^{-3}\lambda$ [rad] for modal reconstruction and $2.5 \times 10^{-4}\lambda$ [rad] for spline-based reconstruction.

Other experiments were done using different wavefronts aberrated by other Zernike polynomials and combinations thereof and the results were similar to the ones presented above.

The splines algorithm was also benchmarked against the modal algorithm for a phase screen generated using the Kolmogorov atmospheric turbulence model. The average and the variance of the reconstruction error are presented in Figure 5-10.

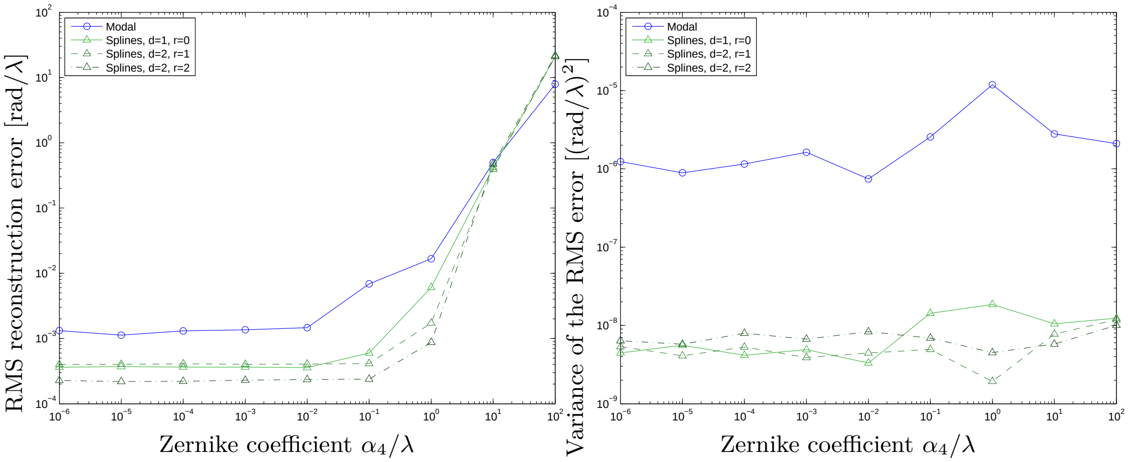

Figure 5-9: Comparison of the RMS value of the wavefront reconstruction error and its sample variance between the new spline based WFR method and the modal reconstruction method (in open-loop). The results were averaged over 100 different reconstructions where the intensities were subject to different noise realizations.

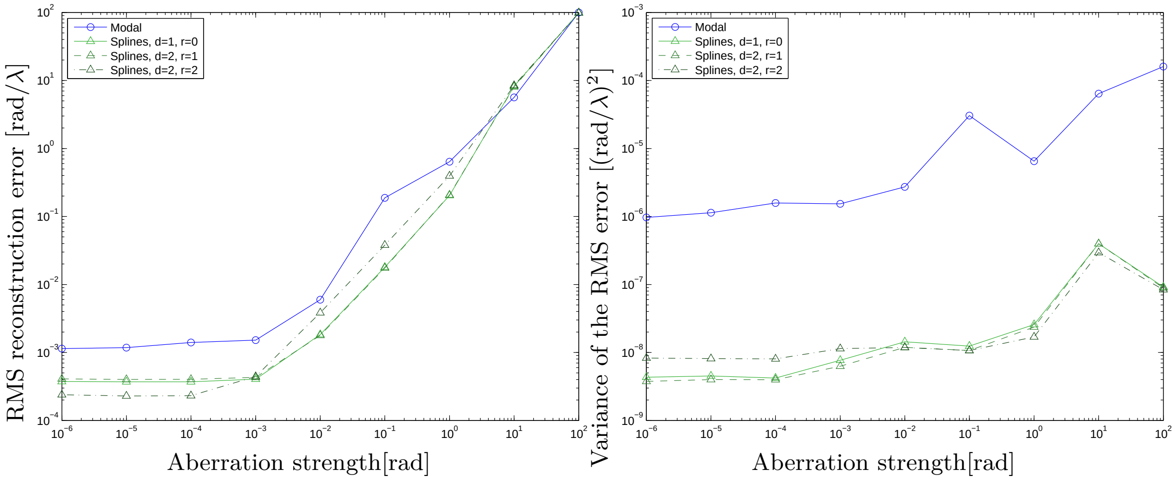

Figure 5-10: Comparison of the mean RMS value of the wavefront reconstruction error and its sample variance between the new spline based WFR method and the modal reconstruction method (in open-loop) for an incoming wavefront that satisfies the Kolmogorov turbulence model with a Fried parameter of 0.2 [m] and an RMS value of 0.1λ [rad]. The results were averaged over 100 different reconstructions where the intensities were subject to different noise realizations.

For a phase screen created using the Kolmogorov turbulence statistical properties, the results are not as good in terms of mean RMS error as they were for the Zernike. The high frequency components of the Kolmogorov wavefront can not be as accurately depicted as the previously used smooth Zernike polynomials.

One aspect that is worth extra emphasis is the different behaviour between the pair (d, r) for $(2, 1)$ and $(2, 2)$. In the case of the wavefront aberrated by a Zernike polynomial in Figure 5-9 the pair $(2, 2)$ performed always better than $(2, 1)$ which happens only because the Zernike aberration can be described perfectly by a 2nd order polynomial. Thus, the strict smoothness conditions imposed by $r = 2$ are very suitable for the reconstruction the aberration modelled by the Zernike polynomial.

In Figure 5-10, it can be seen that the pair $(2, 2)$ performs worst than $(2, 1)$ for aberrations with an RMS value larger than $10^{-3}\lambda$ [rad] and performs better for aberrations smaller than that value. The worsening in performance is due to the fact that the pair $(2, 2)$ does not enable enough degrees of freedom to accurately depict the high frequency components of the Kolmogorv phase screen. The pair $(2, 1)$, given its lower continuity constraint, allows for more degrees of freedom which yield a better performance.

When dealing with very small aberrations, which are very sensitive to noise perturbations in the intensity measurements, having a higher continuity r actually helps to filter out that noise and avoid overfitting. This translates into a lower mean RMS reconstruction error for very small aberrations.

In terms of sample variance, it can be seen that the spline-based method has a very low variance, two orders of magnitude less than the modal reconstruction method, which shows that the modal reconstruction is much more sensitive to the noise realization.

5-4-7 Closed-loop Comparison

So far, the results presented only concerned the open-loop reconstruction where neither a feedback loop nor a deformable mirror is included. In this section, the methods used previously for wavefront reconstruction were integrated in a classical AO feedback loop (Figure 2-1) in order to analyse the empirical convergence properties and sensitivity to noise of the wavefront reconstruction error. The convergence proof in the limit case for very small aberrations is presented in Section 5-3.

The closed-loop setup includes the HWS and the reconstruction algorithm with the same characteristics as specified in Section 5-4-6. It also includes the deformable mirror, which in this case is assumed to be perfect, meaning that it will take the exact shape of the reconstructed wavefront. Furthermore, a delay was added in the loop simulating the time consumed by the computations and communications in a real-time implementation. To counteract the effect of the delay a PI controller was also integrated and tuned in order to minimize the effect of the delay.

Small Aberrations Let us first start by analysing the results when the incoming wavefront has a small aberration. Using a wavefront characterized by an astigmatism aberration ($\alpha_4 = 0.1\lambda$) we obtain the results presented in Figure 5-11.

Applying our novel approach to a small aberration leads to the the convergence of the reconstruction error to a lower RMS error value. Moreover, the reconstruction error is also less sensitive to noise than the closed loop with modal reconstruction. Computing the sample variance over the samples after the 10-th time step, where we have reached steady-state, we obtain, for the modal case, 1.0×10^{-6} $[(\text{rad}/\lambda)^2]$ and for the splines 7.6×10^{-9} $[(\text{rad}/\lambda)^2]$, which is in accordance with the difference of 2 orders of magnitude registered in open-loop. A direct consequence of this experimental result is that, in the presence of noise, the centroid-based modal reconstruction method can not achieve the threshold for small aberrations in open-loop in Figure 5-9 while our intensity-based method can.

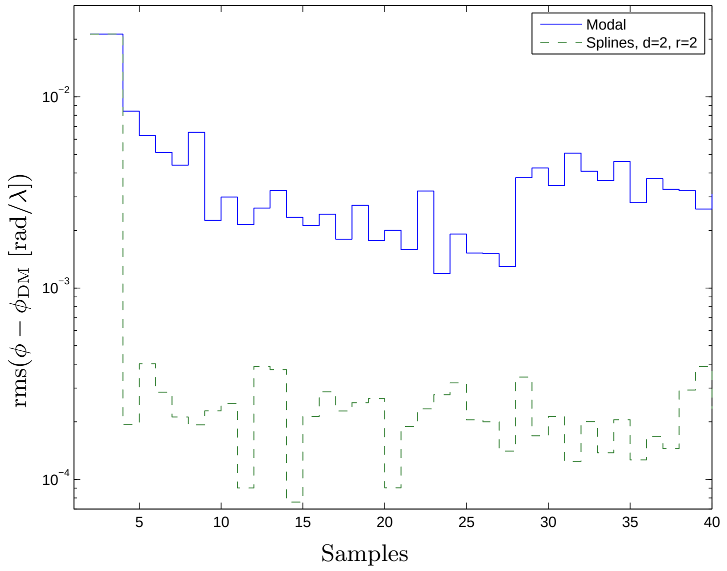

Figure 5-11: RMS reconstruction error evolution of the modal and the spline-based reconstruction method, embedded in a classical AO control loop.

Large Aberrations For larger aberrations, our approach does not reach the resolution previously attained in Figure 5-11 as can be seen in Figure 5-12.

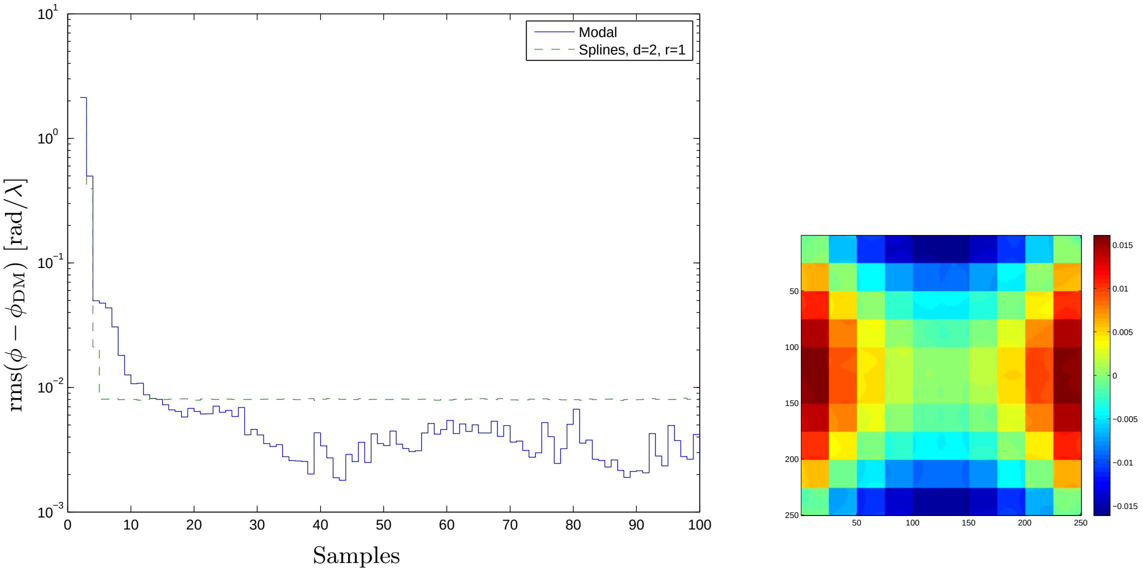

Figure 5-12: (Left) RMS reconstruction error evolution of the modal and the spline-based reconstruction method, embedded in a classical AO control loop for a static astigmatism aberration with a Zernike coefficient of 10λ. (Right) Wavefront reconstruction error after 100 iterations, using the spline-based method.

The simulation was run with a delay of one sample and the PI controller had a proportional gain of 1 unit and an integral gain of -0.99. The mirror was assumed to be perfect.

In Figure 5-12, it can be seen that the spline-based reconstruction method reaches an RMS error of 8×10^{-3} [rad] which is one order of magnitude higher than the steady-state reconstruction error obtained earlier (see Figure 5-11). This is due to the fact that whenever a reconstruction is made using our method, the continuity constraints are never completely satisfied. This leads to the accumulation of the piston modes associated with the individual subapertures[4], producing the result on the right of Figure 5-12. This type of phase screen cannot be estimated by any type of reconstruction method given that the local piston present in each subaperture can not be estimated. A possibility to counteract this error will be presented later.

Also in Figure 5-12, an oscillatory behaviour in the modal reconstruction method is apparent. This behaviour is due to the delay introduced in the loop, since disappears once the delay is removed. Several controller parameters were chosen but none yielded better results.

In conclusion, our new method is much more robust regarding noise and the controller parameters, but it does not achieve the performance attained for smaller aberrations.

Alternating Modal and Splines Reconstruction One way to circumvent the problems verified in the simulations for larger aberrations is to join the advantages of the modal (or any other method that reconstructs the wavefront using slope approximations of the phase) and the splines reconstruction method.

As the slope-based methods can deal with much larger aberrations than the intensity-based method, the first iterations of the adaptive optics loop would have the wavefront reconstruction being made by a slope-based method. When the steady-state was reached, our intensity-based method would be triggered and a better performance could be achieved. One problem to this approach is that, given that we are unable to compute the RMS reconstruction error as the incoming wavefront is unknown, we would have to make some assumptions regarding the moment when the intensity-based method would be triggered. If no assumptions can be made, an off-line image processing technique could be used to assess the resolution of the image captured by the science camera. If this resolution was above a certain threshold, the intensity-based method would be activated. All of these triggering rules are highly heuristic and would depend greatly on the application.

An example of this switching method is depicted in Figure 5-13.

[4] Notice that this happens for the particular case of a perfect mirror that can take non-smooth shapes. In a smooth mirror such an effect would not be present.

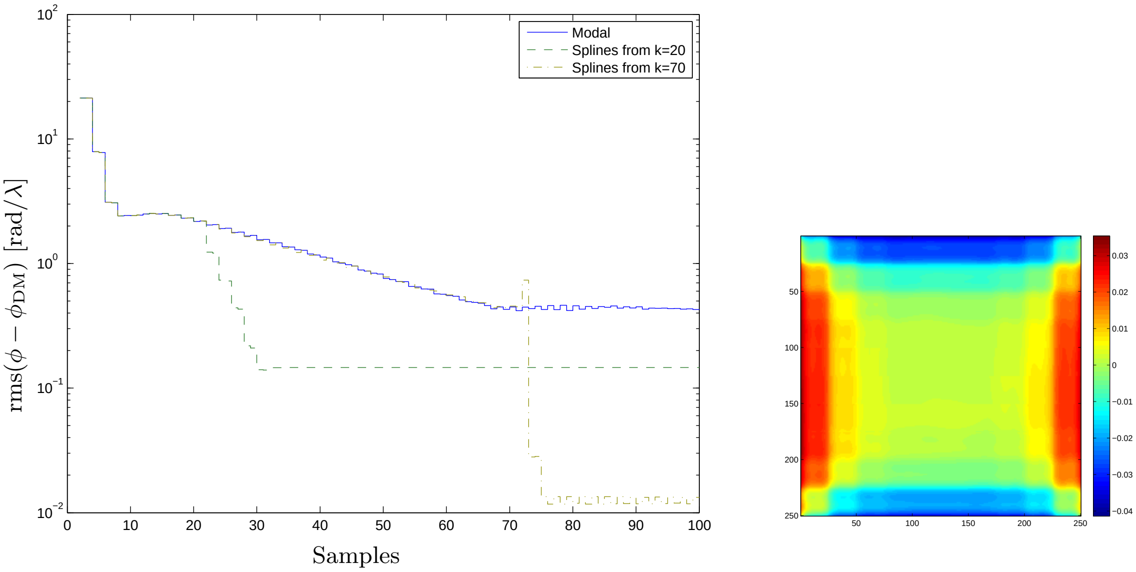

Figure 5-13: RMS reconstruction error evolution of the modal and the spline-based reconstruction method, embedded in a classical AO control loop.

The intensity-based method is triggered in two different moments: at time instant $k = 20$ and after reaching steady-state, at $k = 70$. In both cases the improvements are of one order of magnitude. It is also clear that when the intensity method is triggered after the modal reconstruction reaches a lower value, the final reconstructed wavefront also has a lower error.

5-5 Final Remarks

With the intensity-based method which makes use of a parametrization of the wavefront via B-splines, a successful alternative for the classical slope measurement phase retrieval techniques is presented.

The first results presented show the performance of the method for different setup configurations and simulation parameters. It was concluded that a triangulation of Type II (see Figure 3-1) yields slightly better reconstruction results than a Type I triangulation. For a [10 × 10] subaperture grid and an aberration with an RMS value of λ [rad], a Type II triangulation yielded 2.5 times smaller.

Regarding the propagation distance between the pupil- and the detector plane, it was shown that the diffraction pattern produced by a subaperture does not interfere with the neighbouring diffraction patterns significantly if the propagation distance is smaller than 10 mm (given the nominal setup configuration parameters provided in the beginning of the chapter).

Increasing the number of the subapertures lead to a better reconstruction performance, in general. However, due to the computational complexity of the method it was not possible to run the simulations for very big grids such that more general conclusions can be drawn regarding the influence of the number of subapertures. Besides the number of subapertures, we also analysed the influence of the number of pixels in the detector screen and it was seen that reducing the number of pixels per subaperture to values lower than 10 significantly decreased the performance in terms of the RMS reconstruction error, up to two orders of magnitude.

In open-loop, we can distinguish two different results obtained for low (Zernike defocus and astigmatism modes) and high spatial frequency aberrations (phase screen with Kolmogorov statistics). When reconstructing low spatial frequency aberrations, our wavefront reconstruction method provides an improvement of approximately one order of magnitude in terms of RMS error compared to modal reconstruction method using Zernike polynomials for aberrations with an RMS value smaller than λ [rad] when the intensity measurements were subject to noise. If the readings were hypothetically noiseless the improvement on the RMS reconstruction error would be of two orders of magnitude. The variance of the RMS error was experimentally verified to be two orders of magnitude lower for our intensity method.

Regarding Kolmogorov phase screens, an improvement of one order of magnitude in the reconstruction error was verified in the noiseless case. In the presence of noise, the method proved to outperform modal reconstruction only for aberrations with an RMS value smaller than 0.1λ [rad]. The sample variance of the reconstruction error was also verified to be two orders of magnitude lower for our intensity method, which is a strong indication of its robustness to noisy measurements.

Concerning the splines degree d and continuity degree r, the most important results are the reconstructions without the presence of noise. In this case, it was seen that increasing the degree of the approximating polynomials more than 2 does not necessarily lead to better performance. This is due to the limitations of the linearised model itself, and it not being able to capture complex enough information such that it can be described by a higher order polynomial. The best results were obtained for $(d, r) = (2, 1)$ and $(1, 0)$.

From an AO control perspective, this method can be easily implemented in real-time feedback setups given its linearity. Besides that, integrating this novel method in a classical AO feedback loop yields, once steady-state is reached, a reconstruction error one order of magnitude smaller than the modal method, provided that the strength of the aberration is smaller than λ [rad] and that the deformable mirror has sufficient resolution to accurately depict the reconstructed phase.

In closed-loop and for larger aberrations ($\geq 10\lambda$), and under the assumption that the mirror is perfect, the intensity based method is not able to achieve such low residual wavefronts (in terms of RMS value) as for smaller aberrations. This is due to the incremental creation of waffle modes due to the smoothness constraints not being perfectly matched. To minimize this effect, we propose that another method (*e.g.*, modal reconstruction) is used for reducing the aberration error until a certain threshold, and then the intensity based method is triggered to fine tune the results and reach an even lower reconstruction error.

Given the structure of the (Shack-)Hartmann sensor the problem has the potential to be solved in a distributed way, which will be presented in Chapter 7 and 8.

Future Work In closed-loop, the new algorithm was never tested with dynamic aberrations. A good follow-up research would be to test its performance when the phase screen is changing with time. A theoretical analysis could also be attempted where the rate of change of the phase screen would be bounded and the performance of the algorithm would be quantified based on that bound. In other words, how fast can the phase-screen change such that the reconstruction algorithm is able to keep the reconstruction error below a certain tolerance.

This method can be also applied to phase reconstruction using intensity measurements from a Shack-Hartmann wavefront sensor. The work made in [13] proposed a phase-retrieval algorithm using a single lenslet.

Furthermore, in [12], the subapertures were rotated 25 degrees such that the overlap between the intensity patterns from different subapertures was reduced. In this thesis, that effect was not studied as it created a mismatch between the diffraction patterns and the square and symmetric mesh that defined the triangulations. Further research can be done in this topic to try to find a suitable triangulation such that the interference between intensity patterns is minimized.

Another interesting path to try would be to have a gain scheduling approach that adapts the linearization point on-line such that the model is better suited to the aberration that we are trying to estimate. Let us provide an example. Consider that we are trying to estimate a defocus aberration but that we start with a linearization around a flat wavefront. When steady-state is assumed to be reached in closed-loop the model is changed to one whose linearization was made around phase coefficients that better represent the defocus aberration.

Chapter 6

Compressive Sampling Using Jacobian Analysis

This chapter is based on the work by Brunner et al. in [79].

To guarantee real time applicability of the spline based method presented in the previous chapter with intensity measurements, we propose two approaches to reduce the computational complexity. The authors in [53] have shown that the local nature of the B-spline framework allows highly distributed computation of the LS estimate. The approach followed in [53] which deals with slope-based reconstruction methods will be extended to the centralized method presented in Chapter 5 by formulating a novel distributable reconstruction method in Chapter 8. In addition to the distributed implementation, a second strategy based on compressive sampling [80] to reduce the computational complexity can be implemented. In this chapter, we focus on the latter improvement. We present simulation results which show that the number of pixels used for estimation of the wavefront can be reduced dramatically without significant loss of accuracy and robustness of the reconstruction.

In Section 6-1, the compressive sampling technique is explained. The results are then presented in Section 6-2 for both open- (see Section 6-2-1) and closed-loop (see Section 6-2-2). The chapter ends with some final remarks in Section 6-3.

6-1 Acceleration Study through Compressive Sampling

Motivated by the work of [81] and others, who applied principles of the theory of compressive sensing and sampling to solve the phase retrieval problem, the idea is to only use $M' := \frac{cr}{100}M$ for $0 < cr \leq 100$, of the total number of M given intensity measurements stored in $\mathbf{i}_{\text{meas}}$ to estimate the spline coefficients. This is implemented by introducing a selection matrix $\mathbf{S} \in \mathbb{R}^{M' \times M}$ in the cost function of the global reconstruction problem in (5-17), which yields a new cost function

$$J_{\text{lin}}^{\text{CS}} = ||\mathbf{S}(\mathbf{i}_{\text{meas}} - \mathbf{c}_0) - \mathbf{S}\mathbf{C}_1\boldsymbol{\alpha}||_2^2 \,. \tag{6-1}$$

Parameter cr denotes the undersampling of the intensity images and is further referred to as compression ratio. The LS estimate of the B-coefficient vector is then computed as

$$\boldsymbol{\alpha}^{\star} = \mathbf{N_H}\big((\mathbf{C}_1^{\text{CS}})^{\top}\mathbf{C}_1^{\text{CS}}\big)^{-1}(\mathbf{C}_1^{\text{CS}})^{\top}(\mathbf{i}_{\text{meas}} - \mathbf{c}_0)\,, \tag{6-2}$$

where $\mathbf{C}_1^{\text{CS}} := \mathbf{S}\mathbf{C}_1\mathbf{N}_H \in \mathbb{R}^{M' \times k_H}$ with k_H denoting the nullspace dimension of constraint matrix $\mathbf{H}$.

Choosing at each time instance, the $M_s' = crM_s$ largest intensities in each subaperture of the sensor would guarantee maximal signal to noise ratio. However, this approach is not feasible for real time computations as it would turn the selection matrix $\mathbf{S}$ to become time variant. The consequence of this is that the reconstruction matrix $\mathbf{R} = \mathbf{N_H}\big((\mathbf{C}_1^{\text{CS}})^{\top}\mathbf{C}_1^{\text{CS}}\big)^{-1}(\mathbf{C}^{\text{CS}}{}_1)^{\top}$ can no longer be precomputed, thereby increasing the real-time computational complexity (unacceptably for large scale problems such as for extreme large telescopes). Two time-invariant procedures to construct a precomputed selection matrix $\mathbf{S}$ are considered here. The first is inspired by the random sampling often performed in compressive sampling [80]. This results in determining $\mathbf{S}$ such that at each time instance an identical (though a priori determined random) selection of M_s' pixels is made in each subaperture. The selection can differ for different subapertures. The second option is to choose the M_s' intensity measurements which are most “favoured” by the linear model in the following manner. Firstly, for each subaperture n we determine a vector $\mathbf{k}_n \in \mathbb{R}^{M_s}$ as

$$\mathbf{k}_n(m) = \sum_{j=1}^{K} |\mathbf{C}_{1,n}(m, j)|, \tag{6-3}$$

for $m = 1, ..., M_s$. The entries of this vector reflect the averaged (in terms of the $\ell1$-norm) sensitivity of the intensity measurements to the spline coefficients corresponding to a local aperture. The selection matrix $\mathbf{S}_n \in \mathbb{R}^{M_s' \times M_s}$ is in this second option constructed such that the M_s' pixels $(\mathbf{i}_{\text{meas}})_n(m)$ with the highest values for $\mathbf{k}_n(m)$ are selected. The global block diagonal selection matrix is given by

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 & & \\ & \ddots & \\ & & \mathbf{S}_N \end{bmatrix} \in \mathbb{R}^{M' \times M}\,. \tag{6-4}$$

For a regular triangulation with B-spline polynomials of degree d in all simplices, the local matrices $\mathbf{S}_n$ are identical for all $n = 1, ..., N$. The motivation for the second selection option is that the “most senstive” pixels will also be those that have the best signal to noise ratio. This heuristic argument will be further illustrated in the experimental section 6-2 where a significant advantage of the Jacobian based over the randomly computed selection matrix is shown.

6-2 Numerical Simulations

For the simulations, the same setup was used as in Section 5-4-6. An astigmatism, the 4th Zernike mode in Noll’s notation, is used to model the incoming wavefront with small aberrations of $\alpha_4 = 0.1\lambda$ where α_4 denotes the 4th Zernike coefficient. It was shown by [82] that the spline based method for intensity measurements provides stable results for aberrations smaller than λ. The wavefront is estimated on the aperture plane of the Hartmann sensor, where each subaperture is covered by 4 simplices. The phase screen is then approximated with a B-form polynomial of degree $d = 2$, subject to continuity constraints of order $r = 2$. To evaluate the performance of the wavefront reconstruction the RMS values of the residual wavefront, the difference between the simulated and the estimated phase screen (both normalised to the wavelength λ), have been computed for several noise realisations to obtain an averaged result.

Selection Matrix and Computational Gain

In this section, the advantage of the Jacobian based selection matrix $\mathbf{S}$ over its randomly computed counterpart is shown. It becomes especially significant for the compression ratios of interest $cr < 20\%$. Furthermore, a short complexity analysis of the real time computations which have to be performed in the presented reconstruction method shows the acceleration achieved with the compressive sampling approach.

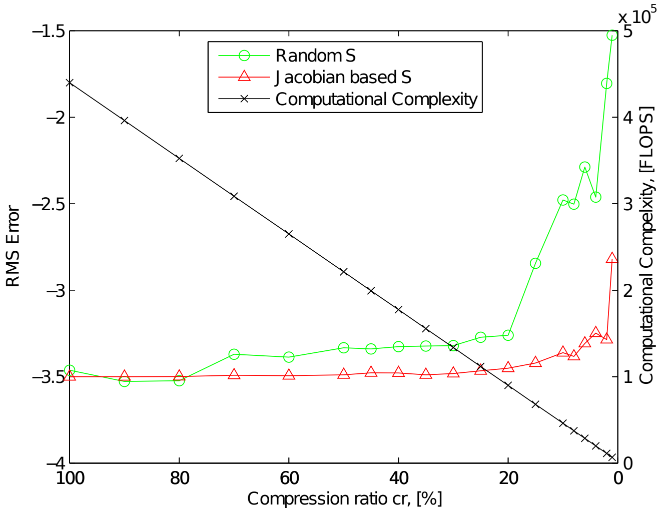

Figure 6-1: RMS errors between the reconstructed and the orginal 4th order Zernike wavefront for decreasing percentage $cr\%$ of intensity measurements used. Triangles: Randomly computed selection matrix. Circles: Jacobian based selection matrix. Crosses: Real time computational complexity depending on $cr\%$.

In Figure 6-1, the RMS values of the normalised residual wavefront are plotted in logarithmic scale for wavefront estimates computed with $cr\%$ of the noisy CCD image pixels simulated for the described Hartmann set up and an astigmatic incoming wavefront. One can see that the accuracy of the estimation performed for randomly selected pixels decreases for less than 20% of compression ratio, where as the Jacobian based selection matrix provides stable results up to 10% and shows a steep increase in the RMS error only at a compression ratio of 1%. Even

though these results do not replace a full analysis of the Jacobian based selection, they give sufficiently strong indication to use this approach for the compressive sampling.
Next to the evolution of the RMS error, the computational complexity of the real time computations, which have to be performed in order to obtain the values of the estimated wavefront at N' point stacked in coordinate vector $\mathbf{x} \in \mathbb{R}^{N' \times 2}$, is shown in Figure 6-1 for decreasing compression ratio cr. This real time computation consists of the following 3 steps:

$$\phi^*(\mathbf{x}) = \big(\mathbf{B}(\mathbf{x})\mathbf{N_H}\big)\big(\mathbf{C_1^{\mathrm{CS}}}\big)^+ \mathbf{S}\big(\mathbf{i}_{\mathrm{meas}}^{\mathrm{glob}} - \mathbf{c_0}^{\mathrm{glob}}\big)\,, \tag{6-5}$$

where the product of the spline evaluation matrix $\mathbf{B}(\mathbf{x})$ and the nullspace projector $\mathbf{N_H}$ as well as the pseudo inverse of the modified Jacobian $\mathbf{C_1^{\mathrm{CS}}}$ (introduced in (6-2)) are precomputed. The selection of M' intensities with sparse matrix $\mathbf{S}$ and the subtraction of the respective linearisation offsets can be scaled at M' FLOPs (Floating Point Operations). The computational complexity is then given by

$$\mathcal{C} = (N'k_H + k_H M' + M')\mathrm{FLOPs}$$

where the compressed total number of intensity measurements is $M' = crM_sN$, with $M_s = 625$ pixels per subaperture and $N = 100$ subapertures in the considered case. Note that for a real case scenario of an extremely large telescope N scales at $\mathcal{O}(10^4)$. The real time computation is applied in the described way as dimension k_H of the nullspace of constraint matrix $\mathbf{H}$ is much smaller then K and M'. k_H is a function of the total number of simplices and internal edges in the triangulation as well as degree d and continuity order r of the spline model [52] and equals 6 for the chosen setup and model. For p evaluation points per subaperture the computational complexity is obtained with

$$\mathcal{C} = (p\,k_H N + (k_H + 1)M_s N cr)\mathrm{FLOPs}$$

as linearly decaying function of compression ratio cr, which is plotted for $p = 4$ evaluation points per subaperture in Figure 6-1 .

6-2-1 Open-loop Comparison

In the following section, we present simulation results for open loop reconstruction which show that the spline based method for intensity measurements with compressive sensing of ratio $cr = 10\%$ suffers only minor to negligible losses in peformance to variations in aberration strength and to different noise standard deviations, compared to the original method using the full CCD output. To allow further comparison to a standard wavefront reconstruction matrix, a modal wavefront reconstruction method using slope measurements was used for the same set up. For this very common method, the center of gravity of the intensity distribution is computed with the centroid algorithm for each subaperture of the Hartmann sensor. It approximates the averaged slopes of the wavefront seen by the respective subaperture. From these slope measurements the wavefront which is parametrized using Zernike polynomials can be estimated by solving the least squares problem for optimal weighting coefficients.

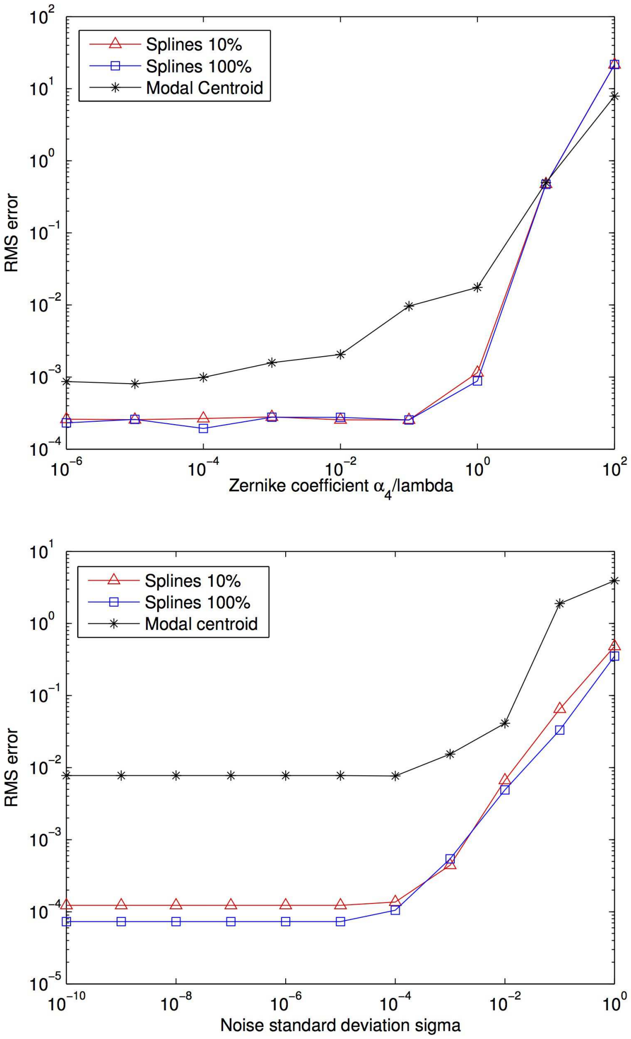

Figure 6-2: RMS errors for different reconstruction methods: Spline based method for 10% of the intensity measurements (Triangles), spline based method using all given intensity measurements (Squares), classical modal reconstruction method for slope measurements (Stars). Top: Fixed noise standard deviation $\sigma_{\text{ccd}} = 4{\times}10^{-4}$, increasing aberration strength simulated by augmenting the Zernike coefficient α_4. Bottom: Fixed astigmatic aberration for $\alpha_4 = 0.1\lambda$, noise with varying standard deviation σ_{ccd}.

Figure 6-2 shows the RMS values of the absolute error maps between the estimated wavefront and the original incoming wavefront simulated by an astigmatism.

In the first plot, a fixed noise standard deviation was assumed while the open loop reconstructions were performed for varying aberration strength simulated by varying the Zernike weight α_4. One can see that the spline and intensity based method give almost the same accuracy for reconstruction from 100% or 10% of the measurements. For aberrations smaller than λ, the assumption of locally independent imaging holds and aberrations of higher polynomial orders can be retrieved from the intensity measurements. The centroid based method processes only information about the local slopes of the wavefront which yields a less accurate approximation. The RMS errors of both methods reach a threshold for very small aberrations due to the influence of the measurement noise. For aberrations larger than 10λ, the diffraction pattern corresponding to one subaperture affects the intensity pattern of the neighbouring such that the assumption for independent imaging is not valid anymore. In this case, slope measurements give better information about the shape of the wavefront.

The second plot shows the behaviour of the RMS error for the same simulations with a fixed $\alpha_4 = 0.1\lambda$ aberration where different noise levels on the intensity measurements where simulated. Again, the spline method shows the same behaviour using 10% of the intensity measurements as for reconstruction from all the pixels. Only a minor loss in accuracy was observed for the reduced version. Since the used spline polynomials of degree 2 cannot approximate higher modes in the wavefront, the performance reaches a limit due to fitting errors.

6-2-2 Closed-loop Comparison

In this section, the discussed wavefront reconstruction methods were integrated in a classical AO feedback loop. It is shown that the compressive sampling preserves the convergence properties and the sensitivity of the wavefront reconstruction error to noise of the spline and intensity based method.

The closed-loop setup includes a simulator of the Hartmann sensor specified at the beginning of Section 6-2 which computes the intensity measurements and models the read out noise. The control loop configuration is the same as presented in Section 5-4-7.

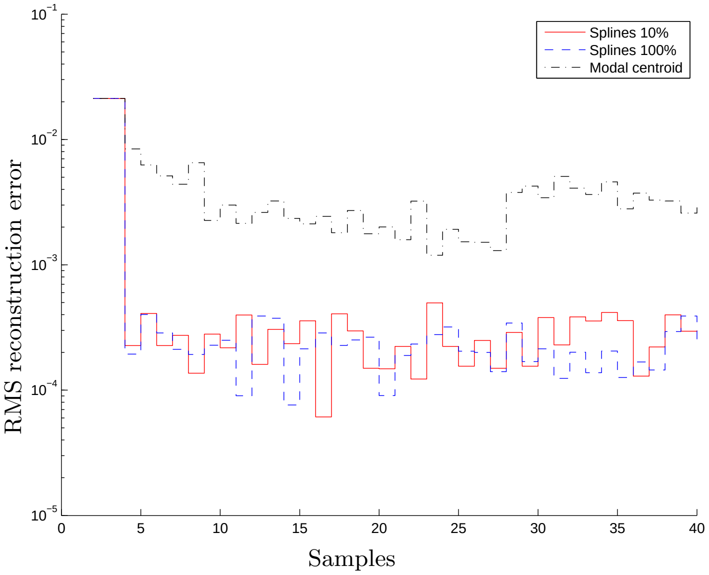

Figure 6-3: RMS value of the residual wavefront for a classical closed loop AO setup for reconstruction from intensities with the spline based method (Solid: 10% of pixels; Dashed: all pixels) and from slopes with the classical modal method (Dot-dashed).

Figure 6-3 shows the results obtained for an 0.1λ astigmatism aberration with measurement noise of standard deviation 10^{-4}. The splines based method reaches the same convergence and noise sensitivity levels for compressed number of intensity measurements as for the full number of pixels. The RMS error values of the reconstruction error with the modal method using slope measurements emphasizes that the new intensity based spline approach converges to a lower error level and is less sensitive to noise than the classical method. The variance of the RMS reconstruction error once steady-state was reached was 1.04×10^{-6}[rad^2] for the modal reconstruction. For the splines with no compression, the variance achieved was 7.64×10^{-9} [rad^2] and with a compression ratio of 10% the variance did not increase substantially, attaining a value of 1.22×10^{-8} [rad^2].

6-3 Final Remarks

In this section, we introduced a procedure to accelerate the real time computation part of a wavefront reconstruction method for intensity measurements of a wavefront sensor without compromising the performance of the reconstruction. The compressive sampling reconstruction method significantly reduces the number of intensity measurements used for wavefront reconstruction to only a few percent of the full image information. Firstly, simulations of open-loop and closed-loop AO systems gave very promising results. This indicates that performance and robustness as well as the convergence gain that was established with the new intensity based reconstruction method [82] could be preserved with a significantly reduced computational complexity. The novel compressive sensing method is highly suitable for a distributed implementation, since the selection process is independently performed for each intensity pattern in the sensor's subapertures which stands in contrast to the globally applied random sampling in many compressive sensing methods. Due to the local nature of the presented compressive sensing reconstruction method for intensity measurements, it can be integrated with our recent work of [53] which has shown that the locally defined B-spline framework allows highly distributed computation of the LS wavefront estimate.

Chapter 7

Distributed and Unstructured Wavefront Reconstruction

Alongside the compressive sampling technique presented in the previous chapter, the computational complexity of the wavefront reconstruction method in Chapter 5 can be alleviated by performing the estimation in a distributed way. The inherent local nature of the wavefront reconstruction problem using a (Shack-)Hartmann sensor, together with the locality of the spline phase parametrization allows for the formulation of the wavefront reconstruction problem in a distributed way.

In simplex B-spline approximation [83], linear least-squares regression problems naturally arise because of inherent domain decomposition of the function domain into a triangulation. This leads to a regression matrix with a block-diagonal structure, where the number of block-diagonal elements N equals the number of partitions needed to divide the function domain into a triangulation. In recent times, simplex B-spline based regression has found numerous applications in various domains, such as modelling of aerodynamic data [84], wavefront reconstruction in adaptive optics systems [53], numerical solutions to Navier-Stokes equations [85], and numerical solutions to time-dependent Hamilton-Jacobi-Bellman equations [86]. To improve scalability of B-spline regression in the various application domains, finding fast, efficient, and distributed methods to solve the regression problems will be essential.

The method which is presented in the following sections does not take advantage of the sparse communication structure which is present in the splines smoothness matrix (hence, the title of the chapter referring to an unstructured method). This implies that the update of the splines coefficients in one particular subaperture can depend on the coefficients of all other subapertures and not only on the neighbours.

The chapter is organized as follows. The problem presented in Chapter 5 is reformulated to fit the ADMM framework in Section 7-1. The ADMM is applied to the specific problem of wavefront reconstruction in Section 7-2. The converge in open- and closed- loop is analysed in Section 7-3. In Section 7-4 the results are presented and the influence of the stepsize and the number of iterations is analysed. The chapter is concluded with some final remarks in Section 7-5.

7-1 Separable and Unstructured Problem Formulation

The wavefront reconstruction method presented in [82] and presented again in Section 5-2-3 is formulated as a linear least-squares problem subject to linear constraints by

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & ||\mathbf{i}_{\text{meas}} - (\mathbf{c}_0 + \mathbf{C}_1\boldsymbol{\alpha})||_2^2\,, && (7\text{-}1)\\ \text{subject to} \quad & \mathbf{H}\boldsymbol{\alpha} = 0. && (7\text{-}2)\end{aligned}$$

Due to the block diagonal structure of $\mathbf{C}_1$ the objective function can be split in N different functions, each pertaining to one subaperture, which gives

$$||\mathbf{i}_{\text{meas}} - (\mathbf{c}_0 + \mathbf{C}_1\boldsymbol{\alpha})||_2^2 = ||(\mathbf{i}_{\text{meas}})_1 - (\mathbf{c}_{0,1} + \mathbf{C}_{1,1}\boldsymbol{\alpha}_1)||_2^2 + \cdots + ||(\mathbf{i}_{\text{meas}})_N - (\mathbf{c}_{0,N} + \mathbf{C}_{1,N}\boldsymbol{\alpha}_N)||_2^2\,. \tag{7-3}$$

The constraints can also be separated according to the subapertures by selecting the columns associated with each set of variables, such that

$$\mathbf{H}\boldsymbol{\alpha} = \begin{bmatrix}\mathbf{H}_1 & \dots & \mathbf{H}_N\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_1\\ \vdots \\ \boldsymbol{\alpha}_N\end{bmatrix} = 0, \tag{7-4}$$

where, for each subaperture i, the local variables $\boldsymbol{\alpha}_i$ are defined in $\mathbb{R}^{T_s\hat{d}\times 1}$, and the constraint matrix $\mathbf{H}_i$ is defined in $\mathbb{R}^{QE\times T_s\hat{d}}$, following the notation already defined in Section 3-4. Hence the full problem separated in the coefficients that belong to each of the subapertures can be written as follows:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \;\; & ||\mathbf{g} + \mathbf{C}_1\boldsymbol{\alpha}||_2^2 = \sum_{i=1}^{N} ||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 \\ \text{subject to} & \begin{bmatrix}\mathbf{H}^1 & \dots & \mathbf{H}^N\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}^1\\ \vdots \\ \boldsymbol{\alpha}^N\end{bmatrix} = 0, \end{aligned} \tag{7-5}$$

where the quantity $\mathbf{g} = \mathbf{i}_{\text{meas}} - \mathbf{c}_0$ was previously defined in Section 5-2-2, and $\mathbf{g}_i \in \mathbb{R}^{M_s\times 1}$ contains data related to subaperture i.

Thus, the objective is to find a method that enables us to solve this problem in a distributed way. The next section illustrates the use of ADMM to solve the problem.

7-2 Unstructured ADMM Application to Wavefront Reconstruction

The problem described in Eq. (7-5) can be seen as a generalization of the canonical exchange problem presented in Section 4-4-1. The generalization is made such that we deal with optimization variables that are vectors and not scalars and that the constraints can have a more general format, rather than just enforcing equality between unweighted local variables.

Let us first define the augmented Lagrangian $\mathcal{L}_\rho(\boldsymbol{\alpha}, \mathbf{y})$ (see Section 4-4) as follows:

$$\mathcal{L}_\rho(\boldsymbol{\alpha}, \mathbf{y}) = \sum_{i=1}^{N} ||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 + \mathbf{y}^\top \sum_{i=1}^{N} \mathbf{H}_i\boldsymbol{\alpha}_i + \frac{\rho}{2}||\sum_{i=1}^{N} \mathbf{H}_i\boldsymbol{\alpha}_i||_2^2, \tag{7-6}$$

where the dual variable $\mathbf{y} \in \mathbb{R}^{QE \times T_s \hat{d}}$ has the same size as the number of constraints in the smoothness matrix $\mathbf{H}$.

Using ADMM, the augmented Lagrangian is not minimized exactly (the minimization would be exact if we had a centralized update using the Method of Multipliers in Section 4-3-2), but approximately through parallel minimization steps on each variable $\boldsymbol{\alpha}_i$ as follows:

$$\begin{aligned}
\boldsymbol{\alpha}_i(k+1) \quad &= \arg\min_{\boldsymbol{\alpha}_i} \mathcal{L}_{\rho,i} && \text{(7-7)}\\
&= \arg\min_{\boldsymbol{\alpha}_i} \left(||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 + \mathbf{y}(k)^\top \mathbf{H}_i\boldsymbol{\alpha}_i + \frac{\rho}{2}||\mathbf{H}_i\boldsymbol{\alpha}_i + \sum_{j=1, j\neq i}^{N} \mathbf{H}_j\boldsymbol{\alpha}_j|| \right), && \text{(7-8)}\\
\mathbf{y}(k+1) \quad &= \mathbf{y}(k) + \rho\frac{\partial \mathcal{L}_\rho(\boldsymbol{\alpha}, \mathbf{y})}{\partial \mathbf{y}}. && \text{(7-9)}
\end{aligned}$$

Solving these minimization problems yields the following closed-form update equations:

$$\begin{aligned}
\boldsymbol{\alpha}_i(k+1) \quad &= (2\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i} + \rho\mathbf{H}_i^\top \mathbf{H}_i)^{-1} \Big(2\mathbf{C}_{1,i}^\top \mathbf{g}_i - \mathbf{H}_i^\top \mathbf{y}(k) + \rho\mathbf{H}_i^\top \sum_{j=1, j\neq i}^{N} \mathbf{H}_j\boldsymbol{\alpha}_j(k) \Big) && \text{(7-10)}\\
\mathbf{y}(k+1) \quad &= \mathbf{y}(k) + \rho\mathbf{H}\boldsymbol{\alpha}(k+1). && \text{(7-11)}
\end{aligned}$$

The derivations needed to obtain the update formulas of $\boldsymbol{\alpha}_i(k+1)$ presented in the previous equations can be found in Appendix C-1.

The $\boldsymbol{\alpha}_i$ updates are performed completely in parallel for each iteration[1]. After one iteration the results are joined and a central operation to update the dual variable $\mathbf{y}$ is performed. This update tries to maximize the dual variable by following the gradient of the augmented Lagrangian with respect to the dual variable weighted by $\rho > 0$. This gradient is simply given by $\mathbf{H}\boldsymbol{\alpha}(k+1)$.

A problem may arise in computing $\boldsymbol{\alpha}_i(k+1)$, due to the possible singularity of the matrix $(2\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i} + \rho\mathbf{H}_i^\top \mathbf{H}_i)$ that needs to be inverted. One heuristic way to cope with this singularity is to tune the splines parameters so that we don't have modes that can not be retrieved and thus make the matrix non-invertible.

[1] This parallel update (Jacobi type of update where the calculation of the local primal variables at time $k+1$ depends on those variables at time k) is not the standard method of update in the current papers regarding ADMM. Most of the most recent papers that concern proofs regarding convergence and convergence rates, *e.g.*, [87, 88, 89], deal with a Gauss-Seidel update where, for instance, the calculation of $\boldsymbol{\alpha}_2(k+1)$ depends on $\boldsymbol{\alpha}_1(k+1)$ and the calculation of $\boldsymbol{\alpha}_3(k+1)$ depends on $\boldsymbol{\alpha}_2(k+1)$ and $\boldsymbol{\alpha}_1(k+1)$. The issue of convergence when performing a Jacobi update has been addressed in [87, § 4.2] for the case where the equality constraint matrix $\mathbf{H}$ has full rank.

7-3 Convergence Analysis

Open-loop Convergence For a proof of convergence under very mild assumptions, we refer the interested reader to Appendix C-1-2, where the outline of the proof are found based on the work of [88, 87].

Closed-loop Convergence In the chapter containing the centralized method, the closed loop convergence is based on a bound over the relative error (5-24) and following certain assumptions regarding the delay, the measurement noise, and the mirror.

To prove the convergence, the same reasoning can be used to prove that in the limit, *i.e.*, for very small aberrations, convergence is guaranteed. Let $\hat{\boldsymbol{\alpha}}$ represent the centralized estimate, let $\boldsymbol{\alpha}^{\text{red}}$ denote the reduced set of coefficients after the projection into the nullspace of the smoothness matrix $\mathbf{H}$, and let $\hat{\hat{\boldsymbol{\alpha}}}$ represent the distributed estimate. Furthermore, let us define a constant $\gamma > 1$ that bounds the distributed algorithm error with respect to the centralized error as follows:

$$\frac{||\hat{\hat{\boldsymbol{\alpha}}}(t) - \boldsymbol{\alpha}(t)||}{||\boldsymbol{\alpha}(t)||} \leq \gamma \frac{||\hat{\boldsymbol{\alpha}}^{\text{red}}(t) - \boldsymbol{\alpha}^{\text{red}}(t)||}{||\boldsymbol{\alpha}^{\text{red}}(t)||}. \tag{7-12}$$

Thus, using the bound in (5-24) to upper bound the error in the centralized method yields

$$\frac{||\hat{\hat{\boldsymbol{\alpha}}}(t) - \boldsymbol{\alpha}(t)||}{||\boldsymbol{\alpha}(t)||} \lesssim \gamma \frac{K_{\text{Lag}}||\boldsymbol{\alpha}^{\text{red}}(t)||^2}{||\mathbf{g}(t)||} \left\{ \frac{2\text{cond}(\mathbf{C}_1^{\text{red}})}{\cos(\theta)} + \tan(\theta)\text{cond}(\mathbf{C}_1^{\text{red}})^2 \right\}. \tag{7-13}$$

For very small aberrations (that is, taking the limit when $||\boldsymbol{\alpha}(t)||_2 \rightarrow 0$), it is clear that the expression in (7-13) is less than unity, given that the upper bound tends to zero. Hence, the proof is only valid in the limit case.

7-4 Numerical Results

In this section, the results for the wavefront reconstruction in open-loop are presented. The simulations were made using the open-loop specifications of the simulation setup previously described in Chapter 5.

There are mainly two new parameters that we need to analyse for this distributed method: the stepsize ρ and the number of iterations. Firstly, we will analyse the influence of the stepsize parameter ρ on the wavefront reconstruction in Section 7-4-1. Then, the effect of the number of iterations will be presented in Section 7-4-2. Finally, a comparison will be made between the unstructured ADMM method and the centralized method from Chapter 5 in Section 7-4-3.

7-4-1 Influence of the Stepsize ρ

Studies on the impact and optimality of the stepsize parameter have been made in [72], where the optimal stepsize ρ was found for a specifical problem structure. In this section, the analysis that will be undertaken is purely empirical, however.

After performing several experiments it was concluded that the stepsize ρ has a very small range of convergence for the unstructured ADMM approach. In Figure 7-1, two final reconstruction results, after 50 iterations, are shown. On the left-hand side, the reconstruction result is presented for $\rho = 10^{-3}$, which is approximately the highest step-size achievable using a $[10 \times 10]$ subaperture grid. On the right-hand side of Figure 7-1, we chose $\rho = 2 \times 10^{-3}$ and the method starts to diverge.

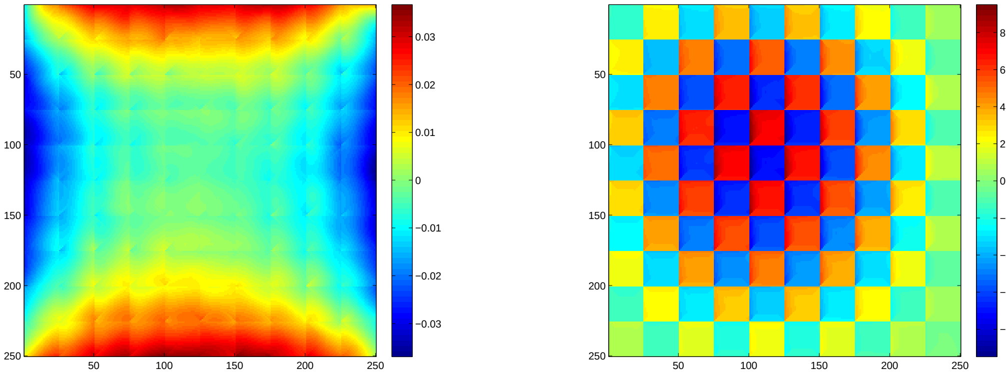

Figure 7-1: Reconstructed wavefronts for (left) $\rho = 10^{-3}$ and (right) $\rho = 2 \times 10^{-3}$ for $d = 2$, $r = 1$. The aberration is given by a Zernike astigmatism aberration with magnitude 0.1λ.

The fact that choosing $\rho > 0$ is not enough to ensure convergence is due to the fact that we chose to split the problem into N different parts, instead of just into two. In [71], it is proven that ADMM converges for any $\rho > 0$, but only for the 2-splitting method. Some authors [87, 88, 89] have successfully shown convergence for N-splitting methods, provided that the stepsize is small enough and that some constraints on the convexity of the problem and rank of the matrices involved are met (which does not happen in this problem).

7-4-2 Influence of the Number of Iterations

Besides choosing the parameter ρ that guarantees the fastest convergence, the number of iterations that the method takes to converge to an acceptable tolerance are a very relevant indicator regarding the real-time implementability of the algorithm. In Figure 7-2, the evolution of the reconstruction is shown along 500 iterations of the ADMM method.

The experiment consisted of the estimation of a Zernike astigmatism aberration with $\alpha_4 = 0.1\lambda$ in a $[10 \times 10]$ subaperture grid. The reconstruction error clearly converges to the centralized solution, although very slowly. In the example in Figure 7-2 the final reconstruction error normalized to the wavelength after 500 iterations is approximately 7.5×10^{-4} whereas the centralized result yields an error of approximately 3.5×10^{-4}.

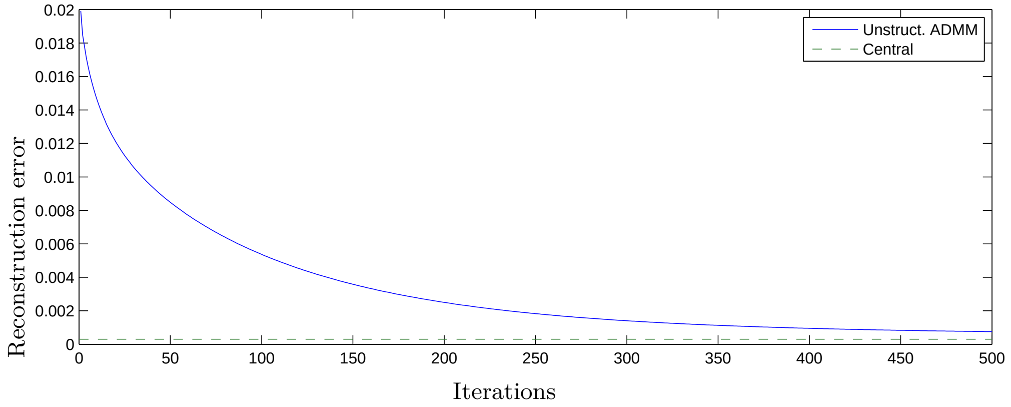

Figure 7-2: Evolution of the RMS value of the reconstruction error as a function of the number of iterations of the unstructured ADMM method.

The relative error between the distributed and the centralized reconstruction methods should be below 10% (preferably 1%) after a few dozens of iterations. In the experiment in Figure 7-2 it is clear that even with a very large number of iterations the accuracy will hardly reach the 10% margin.

7-4-3 Comparison with the Centralized Method

The incoming wavefront is fully described by a 4-th order Zernike polynomial (according to Noll's notation). The ADMM method was run for 100 iterations and the stepsize was set to $\rho = 1 \times 10^{-3}$.

As could be foreseen from the results presented in Section 7-4-2 and in Figure 7-2, this distributed method converges very slowly to the central solution, thus needing a very large number of iterations ($\geq$ 10000) to converge to a value within a 10% margin of the centralized solution.

7-5 Final Remarks

The previously presented method was developed via a straightforward ADMM approach without taking into consideration the underlying structure of the equality constraints (smoothness) matrix $\mathbf{H}$. This approach allows solving the centralized problem presented in Chapter 5 in a distributed way, thus having only to store in memory small matrix and not having to invert the big matrix in (5-18). With this method we are able only to invert very small matrices in (7-10).

The fact that only small matrices are inverted means that there can be an on-line adaptation of the model without extensive off-line pre-computations.

The method was proven to work and to converge for a sufficiently small step-size ρ. However, three pitfalls must be taken into consideration:

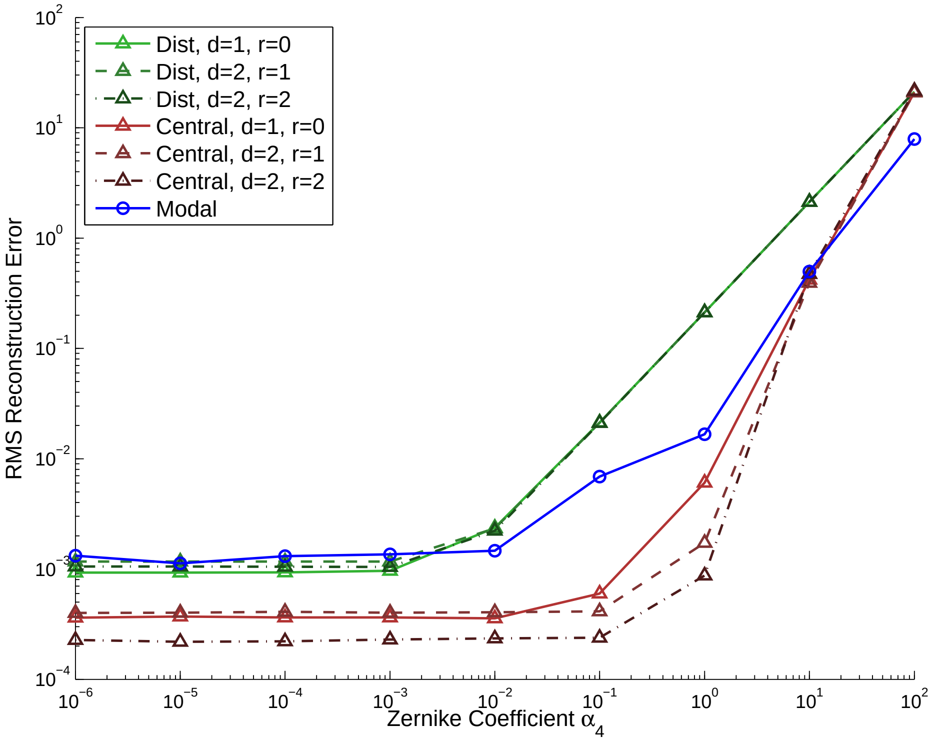

Figure 7-3: Comparison between the centralized splines method (red lines), the modal reconstruction method (full blue line) and the unstructured distributed method (green lines). The algorithm was ran for 100 iterations and ρ was chosen to be 1×10^{-3}.

- The method does not converge for all $\rho > 0$ as the standard ADMM method in [71]. This is due to the fact that the method is based on an N-splitting and not a 2-splitting procedure, which restrains the convergence conditions.

- The convergence rate is very slow. One of the reasons that may have lead to such a poor result is the fact that ADMM is based on a Gauss-Seidel type of update, where the coefficients $\boldsymbol{\alpha}_i(k+1)$ should be calculated based on $\boldsymbol{\alpha}_j(k+1)$, $\forall j < i$ and $\boldsymbol{\alpha}_j(k)$, $\forall j > i$. As the update, in this case, was made in parallel (Jacobi update), each $\boldsymbol{\alpha}_i(k+1)$ could only depend on coefficients from the previous iteration, k, rendering the approximate minimization even poorer than with a Gauss-Seidel approach.

- All the updates require knowledge of the local primal variables in every subaperture which makes this problem rely heavily on communication between the different local nodes if an implementation in an FPGA/GPU platform is to be made.

In the following chapter, a new method which relies only on communication between neighbouring subapertures and takes advantage of the structure of the smoothness matrix will be presented. This method will solve all the pitfalls that one must take into consideration when working with the unstructured ADMM approach.

Chapter 8

Distributed and Structured Wavefront Reconstruction

In this capter a structured and distributed approach to solving the centralized wavefront reconstruction problem in Chapter 5 is presented. This reconstruction method is based on the inherent structure of the splines partioning of the domain and on the fact that the continuity constraints between simplices only affect the neighbouring simplices. Thus, to estimate a function in a particular simplex t under some continuity constraints, the only information needed is contained in the simplices that share an edge with t. The method presented in this chapter takes that structure into consideration.

This chapter is organized as follows. The problem is formulated in Section 8-1 under a completely general framework. ADMM is then applied in Section 8-2 to solve the general structured problem. Then, in Section 8-3, ADMM is applied to solve the wavefront reconstruction problem. Finally the results are presented in Sec. 8-6.

8-1 Structured Problem Formulation

After finding a suitable method for distributed wavefront reconstruction, we realized that that method could be cast in a more general framework and thus have applicability in other areas rather than just in Adaptive Optics. To better understand the terminology and nomenclature used in this section, the reader is advised to read Chapter 4 on distributed optimization, and the references therein. Besides that, very simple graph theory knowledge will be used in this section. The interested reader can refer to [90, § 1.3] for more information on graph theory for distributed algorithms.

Let $\mathcal{G}$ be an undirected and strongly connected graph with N vertices/nodes and an arbitrary number of edges connecting those nodes. The nodes i and the edges (i,j) are contained in the sets $\mathcal{V} = \{1, ..., N\}$ and $\mathcal{E} = \{(i,j) : i \neq j, i,j \in \mathcal{V}\}$, respectively. The problem that we are dealing with tries to minimize, in each node i, a cost function f_i with respect to local variables

$\mathbf{x}_i$. This minimization problem is subject to equality constraints of the form $\mathbf{H}_i\mathbf{x}_i = \mathbf{b}_i$ which only affect the local variables. Therefore, each node has to solve an equality constrained minimization problem. However, besides the local node optimizations it must be guaranteed that the equality constraints of the form

$$\mathbf{H}_{ij}\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{bmatrix} = \mathbf{b}_{ij} \tag{8-1}$$

between all neighbouring pairs of nodes $(i, j) \in \mathcal{E}$ also hold.

All of the aforementioned conditions can be condensed in the following problem formulation:

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i)\\ \text{subject to} \quad & \mathbf{H}_i\mathbf{x}_i = \mathbf{b}_i,\ i = 1, \dots, N\\ & \mathbf{H}_{ij}\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{bmatrix} = \mathbf{b}_{ij},\ \forall(i,j) \in \mathcal{E},\end{aligned} \tag{8-2}$$

where $f_i\ :\ \mathbb{R}^L \to \mathbb{R}, \forall i = 1, ..., N$ are assumed to be convex functions and $\mathbf{x}^\top = \left(\mathbf{x}_1^\top \cdots \mathbf{x}_N^\top\right) \in \mathbb{R}^{LN}$ with $\mathbf{x}_n \in \mathbb{R}^L$. The matrix $\mathbf{H}_i$ belongs to $\mathbb{R}^{C_i \times L}$ and $\mathbf{H}_{ij}$ to $\mathbb{R}^{C_{ij} \times 2L}$, where C_i, $C_{ij} \in \mathbb{R}$ denote the number of constraints associated with the according linear equality constraint. Let us also define the set $\mathcal{N}(i)$ which represents the neighbouring nodes of node i. This set has cardinality M_i.

In order to be able to solve this problem in a distributed way, we can introduce one coupling variable $\mathbf{z}_{ij} \in \mathbb{R}^{C_{ij}}$ per each link $(i, j) \in \mathcal{E}$. The coupling variable was introduced by performing the following algebraic manipulation. Let us first split matrix $\mathbf{H}_{ij}$ as

$$\mathbf{H}_{ij} = \begin{bmatrix}\mathbf{G}_{ij} & \mathbf{F}_{ij}\end{bmatrix},$$

with matrices $\mathbf{G}_{ij}, \mathbf{F}_{ij} \in \mathbb{R}^{C_{ij} \times L}$. Then we can insert a coupling variable $\mathbf{z}_{ij}$ per each link in $\mathcal{E}$. Thus, the new constraints are rewritten as follows:

$$\mathbf{H}_{ij}\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_j\end{bmatrix} - \mathbf{b}_{ij} = \mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 + \mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 = 0$$

$$\begin{cases}\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 = \mathbf{z}_{ij}\\ \mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 = -\mathbf{z}_{ij}\end{cases}$$

Having introduced the coupling, an equivalent problem to (8-2) is formulated as

$$\underset{\mathbf{x}}{\text{minimize}} \qquad f(\mathbf{x}) = \sum_{i=1}^{N} f_i(\mathbf{x}_i) \tag{8-3}$$

$$\text{subject to} \qquad \mathbf{H}_i\mathbf{x}_i = \mathbf{b}_i,\ i = 1, \dots, N \tag{8-4}$$

$$\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 = \mathbf{z}_{ij} \tag{8-5}$$

$$\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 = -\mathbf{z}_{ij},\ \forall(i,j) \in \mathcal{E}. \tag{8-6}$$

Using this new problem formulation with coupling constraints, the ADMM method can be applied to find the optimizer $\mathbf{x}^\star$ in an iterative and distributed way. Following the procedure presented in Section 4-4, we are now able to write the augmented Lagrangian for the new formulation as

$$\begin{aligned}\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\nu}, \mathbf{w}, \mathbf{y}) &= \sum_{i=1}^{N} \mathcal{L}_{\rho,i} \\ &= \sum_{i=1}^{N} \Big(f_i(\mathbf{x}_i) + \boldsymbol{\nu}_i^\top (\mathbf{H}_i \mathbf{x}_i - \mathbf{b}_i) \Big) \\ &+ \sum_{(i,j)\in\mathcal{E}} \Big(\rho \mathbf{w}_{ij}^\top (\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}) + \rho/2 ||\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}||_2^2 \Big) \\ &+ \sum_{(i,j)\in\mathcal{E}} \Big(\rho \mathbf{y}_{ij}^\top (\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij}) + \rho/2 ||\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij}||_2^2 \Big), \end{aligned} \tag{8-7}$$

where we introduced three dual column vectors $\boldsymbol{\nu}_i \in \mathbb{R}^{C_i}$, $\mathbf{w}_{ij} \in \mathbb{R}^{C_{ij}}$ and $\mathbf{y}_{ij} \in \mathbb{R}^{C_{ij}}$ that are associated with the constraints in (8-4), (8-5) and (8-6), respectively.

8-2 ADMM Application to the Structured Problem

In order to find the optimum $p^\star$ of the problem in (8-3) and its optimizer $\mathbf{x}^\star$, the update equations for the primal variables $\mathbf{x}$, for the coupling variables $\mathbf{z}$, and for the dual variables $\mathbf{w}$ and $\mathbf{y}$ must be obtained. For that, first we need to minimize the augmented Lagrangian with respect to the primal and coupling variables. Then, the augmented Lagrangian has to be maximized with respect to the dual variables. This process which is done in an alternate and iterative manner.

During the rest of the chapter we will use $k \in \mathbb{Z}$ to denote the iteration of the ADMM algorithm (it is important not to mistake this counter with t which denotes the time steps in closed-loop).

The derivations to obtain the following update equations were based on the work done in [91].

Primal variable update

The first minimization of $\mathcal{L}_{\rho,i}$ is performed with respect to the local primal variables $\mathbf{x}_i$ in each of the nodes. If we aggregate the terms of the Lagrangian (8-7) in $\mathbf{x}_i$ and disregard the remaining ones (as the minimization is done with respect to $\mathbf{x}_i$), we can write the update equation for the local primal variable as

$$\mathbf{x}_i(k+1) = \arg\min \Big\{ f_i(\mathbf{x}_i) + \boldsymbol{\nu}_i^\top(\mathbf{H}_i\mathbf{x}_i - \mathbf{b}_i), \tag{8-8}$$

$$+ \sum_{\substack{i \text{ fixed}\\(i,j)\in\mathcal{E}}} \Big(\rho\mathbf{w}_{ij}^\top(k)(\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}(k)) + \rho/2||\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}(k)||_2^2\Big), \tag{8-9}$$

$$+ \sum_{\substack{i \text{ fixed}\\(l,i)\in\mathcal{E}}} \Big(\rho\mathbf{y}_{li}^\top(k)(\mathbf{F}_{li}\mathbf{x}_i - \mathbf{b}_{li}/2 + \mathbf{z}_{li}(k)) + \rho/2||\mathbf{F}_{li}\mathbf{x}_i - \mathbf{b}_{li}/2 + \mathbf{z}_{li}(k)||_2^2\Big)\Big\}, \tag{8-10}$$

where $k \in \mathbb{Z}$ denotes the current iteration.

A general and explicit update law can not be provided given the fact that the functions f_i have not been specified.

Coupling variable update

By aggregating the terms in $\mathbf{z}_{ij}$ in the augmented Lagrangian and disregarding the remaining ones, we derive the update law for the coupling variables $\mathbf{z}_{ij}$ which is given by

$$\begin{aligned}\mathbf{z}_{ij}(k+1) = \arg\min_{\mathbf{z}_{ij}} \Big\{ & \rho\mathbf{w}_{ij}^\top(\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}) + \rho/2\big|\big|\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}\big|\big|_2^2 \\ & + \rho\mathbf{y}_{ij}^\top(\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij}) + \rho/2\big|\big|\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij}\big|\big|_2^2\Big\}.\end{aligned}$$

After completing the squares we obtain

$$\begin{aligned}\mathbf{z}_{ij}(k+1) = \arg\min_{\mathbf{z}_{ij}} \Big\{ & \rho/2\big|\big|\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij} + \mathbf{w}_{ij}\big|\big|_2^2 - \rho/2||\mathbf{w}_{ij}||_2^2 \\ & + \rho/2\big|\big|\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij} + \mathbf{y}_{ij}\big|\big|_2^2 - \rho/2||\mathbf{y}_{ij}||_2^2\Big\}.\end{aligned}$$

Setting the gradient with respect to $\mathbf{z}_{ij}$ of the quantity inside braces to zero yields

$$-\rho\big(\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij} + \mathbf{w}_{ij}\big) + \rho\big(\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij} + \mathbf{y}_{ij}\big) = 0,$$

which can be rewritten as

$$\mathbf{z}_{ij}(k+1) = \tag{8-11}$$

$$frac12(\mathbf{G}_{ij}\mathbf{x}_i(k+1) - \mathbf{F}_{ij}\mathbf{x}_j(k+1) + \mathbf{w}_{ij}(k) - \mathbf{y}_{ij}(k)) \tag{8-12}$$

where the iteration update is now explicit.

Dual variables update

Now that the primal variable and the coupling variable have their update rules defined we can focus on the dual variables. The dual variables update rule is given by

$$\mathbf{w}_{ij}(k+1) = \arg\max_{\mathbf{w}_{ij}} \Big\{\rho/2||\mathbf{G}_{ij}\mathbf{x}_i - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij} + \mathbf{w}_{ij}||_2^2 - \rho/2||\mathbf{w}_{ij}||_2^2\Big\},$$

$$\mathbf{y}_{ij}(k+1) = \arg\max_{\mathbf{y}_{ij}} \Big\{\rho/2||\mathbf{F}_{ij}\mathbf{x}_j - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij} + \mathbf{y}_{ij}||_2^2 - \rho/2||\mathbf{y}_{ij}||_2^2\Big\}.$$

Following the direction of the gradient yields

$$\mathbf{w}_{ij}(k+1) = \mathbf{w}_{ij}(k) + \mathbf{G}_{ij}\mathbf{x}_i(k+1) - \mathbf{b}_{ij}/2 - \mathbf{z}_{ij}(k+1) \tag{8-13}$$

$$\mathbf{y}_{ij}(k+1) = \mathbf{y}_{ij}(k) + \mathbf{F}_{ij}\mathbf{x}_j(k+1) - \mathbf{b}_{ij}/2 + \mathbf{z}_{ij}(k+1). \tag{8-14}$$

Simplification of the coupling variables update

Substituting (8-13) and (8-14) in (8-12) shows that $\mathbf{w}_{ij}(k+1) - \mathbf{y}_{ij}(k+1) = 0$ which simplifies (8-12) to

$$\mathbf{z}_{ij}(k+1) = \frac{1}{2}(\mathbf{G}_{ij}\mathbf{x}_i(k+1) - \mathbf{F}_{ij}\mathbf{x}_j(k+1)) \tag{8-15}$$

8-3 ADMM Application for Wavefront Reconstruction

In the preceding section, ADMM was applied to a general problem formulation between nodes connected via equality constraints that force a certain relation between their local variables. Before presenting the distributed and iterative wavefront reconstruction method, let us particularize the optimization problem to the estimation problem that we want to solve.

The main goal is to distribute the centralized phase retrieval method presented in Chapter 5. For that purpose, consider that the underlying graph that supports this alogrithm is a regular, symmetric, square mesh with N nodes in the set $\mathcal{V}$ and $N(N-1)/2$ edges in set $\mathcal{E}$. The characteristics of this graph are related to the structure of the Hartmann sensor which is presented in Figure 5-2 and to the parametrization of the wavefront using B-Splines in Figure 5-1. Each node i will represent subaperture i and the edges represent its connections with the neighbouring subapertures.

The functions f_i can now be assigned the following expression:

$$f_i(\boldsymbol{\alpha}) = ||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2, \tag{8-16}$$

where the notation was kept consistent with that of Chapter 5.

In the centralized problem, the equality constraints are defined based on the full smoothness matrix $\mathbf{H}$ (see Eq. (**??**)). This matrix contains two types of constraints: internal constraints that regulate continuity inside each node and neighbouring constraints that regulate continuity between neighbouring nodes.

The internal equality constraints are applied locally in each node and correspond to the continuity constraints between the splines coefficients of the simplices that parametrize the wavefront in a certain subaperture. In order to make the link between the centralized and the distributed method clearer, when introducing the centralized method, we started by presenting this type of constraints in Section 5-2-3.

The notation used will be

$$\mathbf{H}_i\boldsymbol{\alpha}_i = 0, \tag{8-17}$$

with $\mathbf{H}_i$ denoting the local smoothness matrix and where $\boldsymbol{\alpha}_i \in \mathbb{R}^{T_s\hat{d}\times 1}$ represents the local splines coefficients in node i.

The equality constraints that impose continuity between subapertures i and j are denoted as

$$\mathbf{H}_{ij}\boldsymbol{\alpha}_{ij} = \begin{bmatrix}\mathbf{G}_{ij} & \mathbf{F}_{ij}\end{bmatrix}\begin{bmatrix}\boldsymbol{\alpha}_i \\ \boldsymbol{\alpha}_j\end{bmatrix} = 0. \tag{8-18}$$

The complete problem is then defined as

$$\begin{aligned}
\underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & f(\boldsymbol{\alpha}) = \sum_{i=1}^{N} f_i(\boldsymbol{\alpha}_i) = \sum_{i=1}^{N} ||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 \\
\text{subject to} \quad & \mathbf{H}_i\boldsymbol{\alpha}_i = 0,\ i = 1, \ldots, N \\
& \mathbf{G}_{ij}\boldsymbol{\alpha}_i = \mathbf{z}_{ij} \\
& \mathbf{F}_{ij}\boldsymbol{\alpha}_j = -\mathbf{z}_{ij},\ \forall (i,j) \in \mathcal{E},
\end{aligned} \tag{8-19}$$

and the augmented Lagrangian is given by

$$\begin{aligned}
\mathcal{L}_\rho(\boldsymbol{\alpha}, \mathbf{z}, \boldsymbol{\nu}, \mathbf{w}, \mathbf{y}) = & \sum_{i=1}^{N}\Big(||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 + \boldsymbol{\nu}_i^\top(\mathbf{H}_i\boldsymbol{\alpha}_i)\Big) \\
& + \sum_{(i,j)\in\mathcal{E}}\Big(\rho\mathbf{w}_{ij}^\top(\mathbf{G}_{ij}\boldsymbol{\alpha}_i - \mathbf{z}_{ij}) + \rho/2||\mathbf{G}_{ij}\boldsymbol{\alpha}_i - \mathbf{z}_{ij}||_2^2\Big) \\
& + \sum_{(i,j)\in\mathcal{E}}\Big(\rho\mathbf{y}_{ij}^\top(\mathbf{F}_{ij}\boldsymbol{\alpha}_j + \mathbf{z}_{ij}) + \rho/2||\mathbf{F}_{ij}\boldsymbol{\alpha}_j + \mathbf{z}_{ij}||_2^2\Big).
\end{aligned} \tag{8-20}$$

Primal variable update

Now that the cost function is perfectly defined, an explicit update law for the general expression in (8-10).

Firstly, let us take the derivative of the augmented Lagrangian with respect to a particular $\boldsymbol{\alpha}_i$ and set it to zero as follows:

$$\begin{aligned}
&\underbrace{\Big(2\mathbf{C}_{1,i}^\top\mathbf{C}_{1,i} + \rho\sum_{\substack{i\text{ fixed}\\(i,j)\in\mathcal{E}}}\mathbf{G}_{ij}^\top\mathbf{G}_{ij} + \rho\sum_{\substack{i\text{ fixed}\\(l,i)\in\mathcal{E}}}\mathbf{F}_{li}^\top\mathbf{F}_{li}\Big)}_{\mathbf{X}_i}\boldsymbol{\alpha}_i(k+1) + \mathbf{A}_i^\top\boldsymbol{\nu}_i(k+1) \\
&\quad \underbrace{= 2\mathbf{C}_{1,i}^\top\mathbf{g}_i - \rho\sum_{\substack{i\text{ fixed}\\(i,j)\in\mathcal{E}\\(l,i)\in\mathcal{E}}}\Big\{(\mathbf{G}_{ij}^\top(\mathbf{w}_{ij}(k) - \mathbf{z}_{ij}(k))) - (\mathbf{F}_{li}^\top(\mathbf{y}_{li}(k) + \mathbf{z}_{li}(k)))\Big\}}_{\mathbf{d}_i}
\end{aligned}$$

which can be more compactly described by

$$\mathbf{X}_i\boldsymbol{\alpha}_i + \mathbf{A}_i^\top\boldsymbol{\nu}_i = \mathbf{d}_i. \tag{8-21}$$

If we join the resulting equation (8-21) with the local constraints defined by $\mathbf{H}_i\boldsymbol{\alpha}_i = 0$ in (8-19) we will have a system of equations that corresponds to the Karush-Kuhn-Tucker

(KKT) conditions. Using this compact notation we can write the system as

$$\begin{bmatrix} \mathbf{X}_i & \mathbf{H}_i^\top \\ \mathbf{H}_i & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_i(k+1) \\ \boldsymbol{\nu}_i(k+1) \end{bmatrix} = \begin{bmatrix} \mathbf{d}_i \\ 0 \end{bmatrix}, \tag{8-22}$$

$$\bar{\mathbf{X}}_i \bar{\boldsymbol{\alpha}}_i(k+1) = \bar{\mathbf{d}}_i. \tag{8-23}$$

The solution to this problem is then straightforward to obtain provided that the matrix on the left-hand side of 8-22 can be inverted. Notice that the update iterations do not depend on the previous values $\boldsymbol{\alpha}_i(k)$ and $\boldsymbol{\nu}_i(k)$. Therefore, the primal and dual local variables need not be stored. Moreover, the dual local variable does not even need to be computed.

In order to analyse the invertibility of the left-hand side matrix $\bar{\mathbf{X}}_i$ in (8-23) we should analyse the ranks of $\mathbf{X}_i$ and $\mathbf{H}_i$. Although the rank of the matrices $\mathbf{X}_i$ and $\mathbf{H}_i$ does not have a direct relationship with the rank of $\bar{\mathbf{X}}_i$, the rank of the former matrices provides good indications towards a possible rank deficiency of $\bar{\mathbf{X}}_i$.

Starting by the analysis of $\mathbf{X}_i$, the most influential term of $\mathbf{X}_i$, concerning its rank, is $\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i}$. When using, *e.g.*, a Hartmann sensor, this symmetric matrix may become singular when there are too few data points (very few pixels per subaperture) with respect to the amount of coefficients that have to be estimated. Regarding $\mathbf{H}_i$, the matrix has usually full row rank and, if that is not the case, the linearly dependent rows can be removed as they are redundant constraints.

In order to solve the rank deficiency problems of $\mathbf{X}_i$ and $\mathbf{H}_i$, two alternatives are proposed:

QR decomposition In order to have a robust performance over all possible scenarios, we can use QR factorization [92] with pivoting. This factorization separates the free variables (which can take an arbitrary value and gave rise to the singularity) from the determined variables. The full derivation is presented in Appendix C-2-1.

This method reduces the dimensionality of the vector $\bar{\boldsymbol{\alpha}}_i$ in (8-23) to the size given by the rank of $\bar{\mathbf{X}}_i$. In practice, the majority of the cases where this method is useful occur when there are many degrees of freedom and very little data, which will yield poor results even if the centralized method is used.

Nullspace projection Given that the local coefficients should respect $\mathbf{H}_i\boldsymbol{\alpha}_i = 0$, we perform the following linear transformation

$$\boldsymbol{\alpha}_i = \mathbf{N}_{\mathbf{H}_i}\boldsymbol{\alpha}_i^{\text{red}}, \quad \text{where } \mathbf{N}_{\mathbf{H}_i} = \text{null}(\mathbf{H}_i).$$

The nullspace of $\mathbf{H}_i$ spans all the possible combinations of variables $\boldsymbol{\alpha}_i$ that verify the constraints. In other words, if a certain coefficient vector $\boldsymbol{\alpha}_i$ is in the column space of $\mathbf{H}_i$ then the internal constraints $\mathbf{H}_i\boldsymbol{\alpha}_i = 0$ are immediately satisfied. This transformation into the constrained space allows working with a smaller primal variable vector, $\boldsymbol{\alpha}_i^{\text{red}}$. In terms of the update formula, the matrices $\mathbf{C}_{1,i}$, $\mathbf{G}_{ij}$ and $\mathbf{H}_{ij}$ must be replaced by their “reduced” versions in problem (8-19), as follows:

$$\mathbf{C}_{1,i}^{\text{red}} = \mathbf{C}_{1,i}\mathbf{N}_{\mathbf{H}_i} \quad \mathbf{G}_{ij}^{\text{red}} = \mathbf{G}_{ij}\mathbf{N}_{\mathbf{H}_i} \quad \mathbf{F}_{ij}^{\text{red}} = \mathbf{F}_{ij}\mathbf{N}_{\mathbf{H}_i}. \tag{8-24}$$

Working in this reduced spaced eliminates the need for internal constraints. Hence, the dual variable is not needed and the primal variables are obtained as

$$\boldsymbol{\alpha}_i(k+1) = \mathbf{N}_{\mathbf{H}_i} \underbrace{(\mathbf{X}_i^{\text{red}})^{+}\mathbf{d}_i^{\text{red}}}_{\boldsymbol{\alpha}_i^{\text{red}}}, \tag{8-25}$$

where $\mathbf{X}_i^{\text{red}}$ and $\mathbf{d}_i^{\text{red}}$ are the transformed matrices $\mathbf{X}_i$ and $\mathbf{d}_i$ after performing the substitutions in (8-24) given by:

$$\mathbf{X}_i^{\text{red}} = \left(2(\mathbf{C}_{1,i}^{\text{red}})^\top \mathbf{C}_{1,i}^{\text{red}} + \rho \sum_{\substack{i \text{ fixed} \\ (i,j)\in\mathcal{E}}} (\mathbf{G}_{ij}^{\text{red}})^\top \mathbf{G}_{ij}^{\text{red}} + \rho \sum_{\substack{i \text{ fixed} \\ (l,i)\in\mathcal{E}}} (\mathbf{F}_{li}^{\text{red}})^\top \mathbf{F}_{li}^{\text{red}}\right)$$

$$\mathbf{d}_i = 2(\mathbf{C}_{1,i}^{\text{red}})^\top \mathbf{g}_i - \rho \sum_{\substack{i \text{ fixed} \\ (i,j)\in\mathcal{E} \\ (l,i)\in\mathcal{E}}} \left\{ ((\mathbf{G}_{ij}^{\text{red}})^\top (\mathbf{w}_{ij}(k) - \mathbf{z}_{ij}(k))) - ((\mathbf{F}_{li}^{\text{red}})^\top (\mathbf{y}_{li}(k) + \mathbf{z}_{li}(k))) \right\}$$

Notice that, when we want to change back to the original basis and retrieve the full primal variable vector, we have to multiply by the nullspace $\mathbf{N}_{\mathbf{H}_i}$ again.

There might be cases where, for instance, $\mathbf{X}_i^{\text{red}}$ is still rank deficient. In these cases, the user is advised to either supply more (non-redundant) data, decrease the degree of the spline polynomials, or use a factorization to separate the coefficients that can be estimated from those that are free.

The conclusion is that, to solve the KKT conditions, performing a QR factorization and solving the resulting reduced system yields always a solution, whatever the number of degrees of freedom and amount of data available. However, in realistic scenarios the increased robustness brought by the QR factorization is not relevant. For that reason, nullspace projection is preferable given that it reduces the dimensionality of the problem much more than QR factorization by exploring the characteristics of the B-splines framework. There is also a possibility of joining both methods and firstly projecting the coefficients on the nullspace of the local smoothness matrix and then solve the resulting system using QR factorization.

Coupling and dual variable update

The update laws for these variables are similar to the ones obtained for the general problem in (8-15), (8-13), and (8-14). The only change that ought to be made is in replacing the local variable vector $\mathbf{x}$ by the coefficient vector $\boldsymbol{\alpha}$ and eliminating the constant vector $\mathbf{b}_{ij}$ as follows:

$$\mathbf{z}_{ij}(k+1) = \frac{\mathbf{G}_{ij}\boldsymbol{\alpha}_i(k+1) - \mathbf{F}_{ij}\boldsymbol{\alpha}_j(k+1)}{2} \tag{8-26}$$

$$\mathbf{w}_{ij}(k+1) = \mathbf{w}_{ij}(k) + \mathbf{G}_{ij}\boldsymbol{\alpha}_i(k+1) - \mathbf{z}_{ij}(k+1) \tag{8-27}$$

$$\mathbf{y}_{ij}(k+1) = \mathbf{y}_{ij}(k) + \mathbf{F}_{ij}\boldsymbol{\alpha}_j(k+1) + \mathbf{z}_{ij}(k+1), \tag{8-28}$$

where the update must be performed for all edges $(i,j) \in \mathcal{E}$.

8-4 Scalability and Computational Complexity

In this section, we will analyse the computational complexity for both the centralized and the structured and distributed method. The analysis focus on the computations that need to be performed for real-time implementation of the algorithm. The time consumed by the precomputation of matrix inverses and nullspaces is not taken into account into this analysis.

The big-O notation which uses the operator $\mathcal{O}(\cdot)$ and the convention for the definition of floating-point operations (FLOPs) is defined in Section 1-5.

Centralized method

Using the nullspace projection method to solve the KKT conditions yields the equation in (5-20) which will be rewritten below:

$$\boldsymbol{\alpha}^{\star} = \mathbf{N_H}(\mathbf{C}_1^{\text{red}})^{+}\mathbf{g}, \tag{8-29}$$

where $\mathbf{N_H} = \text{Null}(\mathbf{H}) \in \mathbb{R}^{N_\alpha \times N_f}$, $\left(\mathbf{C}_1^{\text{red}}\right)^{+} \in \mathbb{R}^{N_f \times M}$, and $\mathbf{g} \in \mathbb{R}^{M\times 1}$. The constant N_f represents the number of degrees of freedom of the coefficients $\boldsymbol{\alpha}$ allowed by the smoothness matrix $\mathbf{H}$. The constant N_α represents the total number of coefficients and is given by $LN = T_s\hat{d}N$, where T_s is the number of simplex per subaperture and $\hat{d}$ denotes the number of coefficients per simplex. The constant M was defined before and denotes the number of data points. Dividing M by the number of subapertures N, we obtain M_s, the number of points per subaperture.

If the model is assumed to be time invariant, then the matrix $\left(\mathbf{C}_1^{\text{red}}\right)^{+}$ is constant for each computation performed in closed-loop. The nullspace of $\mathbf{H}$ is also time invariant given that it is calculated based on the splines parameters and on the triangulation defined, which remain constant.

This implies that the matrix multiplication $\mathbf{N_H}(\mathbf{C}_1^{\text{red}})^{+} \in \mathbb{R}^{N_\alpha \times M}$ can be stored, such that at each iteration only a matrix vector multiplication with vector $\mathbf{g}$ must be performed. The overall cost of the computations is given by the multiplication of the number of data points and the number of coefficients as follows:

$$\mathcal{O}(M_s L N^2) \text{ FLOPs}, \tag{8-30}$$

where we used the fact that $M = M_s N$, and $N_\alpha = LN$ to explicit the scaling with N^2.

Structured and distributed method

This analysis must be performed in three distinct steps, given the sequential update of the primal, coupling, and dual variables. The primal variable update equation, using nullspace projection to solve the KKT conditions is presented in (8-25) and rewritten below:

$$\boldsymbol{\alpha}_i(k+1) = \mathbf{N}_{\mathbf{H}_i}(\mathbf{X}_i^{\text{red}})^{+}\mathbf{d}_i^{\text{red}}, \tag{8-31}$$

where $\mathbf{N}_{\mathbf{H}_i} \in \mathbb{R}^{L \times N_{f,i}}$, $(\mathbf{X}_i^{\text{red}})^+ \in \mathbb{R}^{N_{f,i} \times M_s}$, and $\mathbf{d}_i^{\text{red}} \in \mathbb{R}^{M_s}$. The constant $N_{f,i}$ represents the number of degrees of freedom that $\boldsymbol{\alpha}_i$ is allowed to have given the constraints imposed by matrix $\mathbf{H}_i$. The quantity $M_s = M/N$ denotes the number of data points in one node.

Provided that the model is time-invariant, the real-time computational complexity to compute the primal variables for each node is given by

$$\mathcal{O}(M_s L) \text{ FLOPs.} \tag{8-32}$$

The total number of operations for all N nodes should be obtained by multiplying the previous complexity by N. However, given that the computations for each node can be performed completely in parallel, the computational complexity remains at $\mathcal{O}(M_s L)$ FLOPs, if the implementation is done in a GPU/FPGA board.

Let us now perform the same analysis for the coupling variable update. The coupling variable update presented in (8-26) has a simple computational complexity given by:

$$\mathcal{O}(C_{ij} L) \text{ FLOPs,} \tag{8-33}$$

where C_{ij} represents the number of constraints in the matrix $\mathbf{H}_{ij}$.

Notice that this complexity represents a single coupling variable update. If the implementation allows $\frac{(N-1)N}{2}$ parallel computations corresponding to each of the coupling variables, the complexity remains $\mathcal{O}(C_{ij} L)$.

The dual variable updates in (8-14) and (8-13) consist of small vector sums, so its complexity will be disregarded in this analysis.

In the end, the complexity of the full algorithm in a single core machine is given by

$$\mathcal{O}\Big(\beta(N)\Big(M_s L N + \frac{C_{ij} L (N-1) N}{2}\Big)\Big) \text{ FLOPs.} \tag{8-34}$$

If the implementation is done on a machine that allows for multi-core implementation, and can have $\max\left\{N, \frac{(N-1)N}{2}\right\}$ threads running in parallel, the complexity boils down to

$$\mathcal{O}\Big(\beta(N)(M_s L + C_{ij} L)\Big) \text{ FLOPs,} \tag{8-35}$$

where $\beta(N)$ is an integer which represents the required number of iterations to achieve the performance required.

Example To exemplify the improvement obtained by using a distributed algorithm, we considered a setup whose subapertures are sampled using $M_s = 625$ pixels (25 pixels per side) and the splines parameters are $d = 2$ and $r = 1$. The number of iterations for the distributed method to converge to a reasonable accuracy is assumed to grow linearly with the number of nodes and is given by the expression $\beta(N) = 0.5N$. The results are presented in Figure 8-1.

A substantial increase in speed can be verified, specially for larger N. Note, however, that these complexity calculations assume that there is no communication overhead. If we could quantify that communication delay between nodes, then the curve of the distributed method would have to be shifted up by a constant quantity.

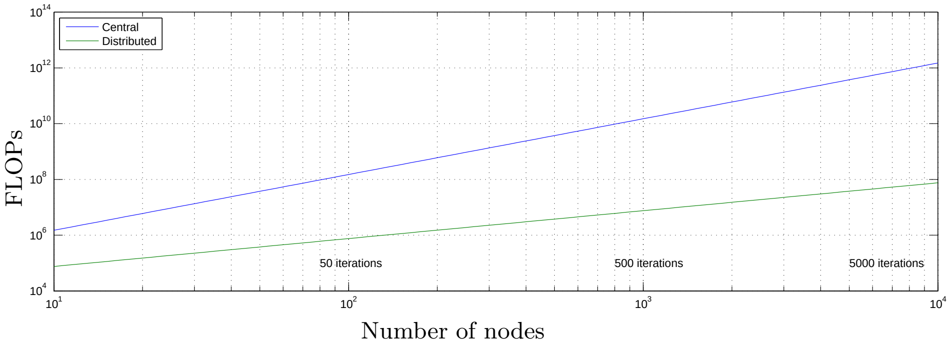

Figure 8-1: Complexity in terms of FLOPs with respect to the number of subapertures/nodes in the algorithms.

8-5 Convergence Analysis

Open-loop Given that one particular node only needs to communicate with its neighbours via the coupling variables $\mathbf{z}$ at its edges, this method turns into a simple 2-splitting method. First, all the primal variables $\boldsymbol{\alpha}$ are updated, completely in parallel. Then, all the coupling variables $\mathbf{z}$ are updated, also in parallel. As proved in [71], the ADMM 2-splitting approach converges for all $\rho > 0$.

Closed-loop The proof follows exactly the same reasoning presented in Section 7-3.

8-6 Numerical Results

In this section, the results for the wavefront reconstruction in open- and closed-loop using a structured ADMM application are presented. The simulations were made using the open-loop specifications of the simulation setup previously described in Chapter 5.

The outline of this section will closely resemble the sequence of topics presented in Section 7-4 of the previous chapter. Firstly, we will analyse the influence of the stepsize parameter ρ on the wavefront reconstruction in Section 8-6-1. Then, the effect of the number of iterations will be presented in Section 8-6-2, followed by a comparison between the unstructured ADMM method and the centralized method from Chapter 5 in Section 8-6-4. Finally, we will end the chapter with some final remarks in Section 8-7 and possible future developments in Section 8-8.

8-6-1 Influence of the Stepsize ρ

The most important parameter in this simulation is the one that regulates the stepsize ρ in the ADMM method. The results presented in Figure 8-2 show the reconstruction error averaged over 100 different noise distributions for 10 different stepsize ρ values. It can be seen that ρ can influence the results significantly up to one order of magnitude.

The experiment was conducted using the nominal setup configuration parameters specified in Section 5-4. The splines parameters were set to $(d, r) = (2, 1)$

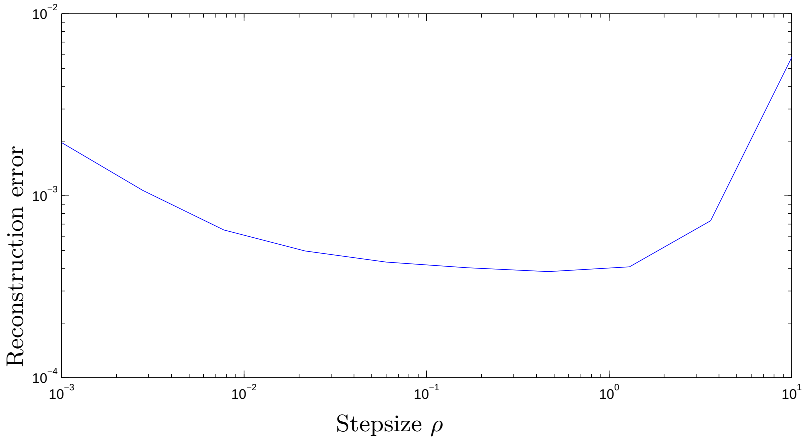

Figure 8-2: Evolution of the RMS value of the reconstruction error as a function of the stepsize ρ of the structured ADMM method.

For the remaining part of the chapter the default value of ρ is going to be set at 0.7 which corresponds to the minimum in Figure 8-2 except when explicitly stated otherwise.

8-6-2 Influence of the Number of Iterations

In Figure 8-3, the evolution of the reconstruction is shown along 500 iterations of the structured ADMM method.

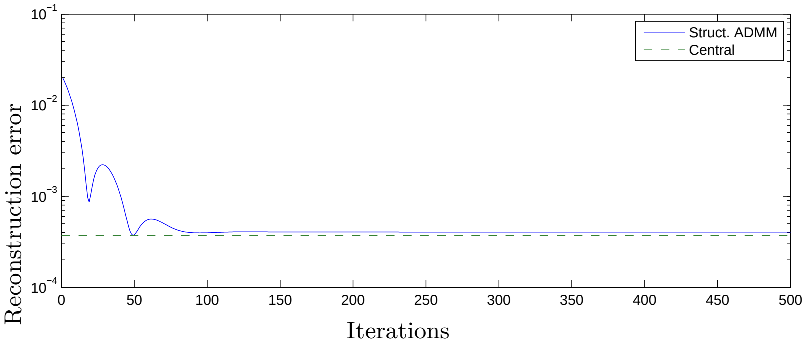

Figure 8-3: Evolution of the RMS value of the reconstruction error as a function of the number of iterations of the structured ADMM method.

The experiment consisted of the estimation of a Zernike astigmatism aberration with $\alpha_4 = 0.1\lambda$ in a $[10 \times 10]$ subaperture grid. The reconstruction error clearly converges to the centralized solution, although not in a monotonically decreasing way.

In fact, the convergence results for ADMM [71, § 3.2.1] show that for the 2-splitting method used in this chapter, there is no guarantee that the convergence towards the optimum is monotonic.

The only convergence results that are guaranteed in the general case are the ones postulated in Section 4-4. Therefore, it is also not guaranteed, in the general case, that the primal variables $\boldsymbol{\alpha}$ and the coupling variables $\mathbf{z}$ converge to optimal values. Under some assumptions, *e.g.*, strong convexity on the cost function [88], we can guarantee that we converge to a primal-dual optimum pair.

In comparison with the results obtained for the unstructured ADMM method in Chapter 7 it can be observed that the convergence rate is much higher and that the method converges to a value much closer to the centralized solution. This can be attributed mainly to the fact that in the unstructured approach, due to the fact that we solve an N-splitting problem, the algorithm converges only for a limited interval of stepsizes ρ, which is not the case in the 2-splitting structured method. Besides that, the unstructured approach using the N-splitting method performs a Jacobi type of update, which involves that all the primal variable updates are made in parallel. This slows down the convergence of the unstructured method even more, as in the 2-splitting method, the Gauss-Seidel update is used, where the primal variables $\boldsymbol{\alpha}$ are used to compute the coupling $\mathbf{z}$ (opposite to the Jacobi update, where $\boldsymbol{\alpha}$ and $\mathbf{z}$ are computed in parallel).

8-6-3 Scalability Analysis

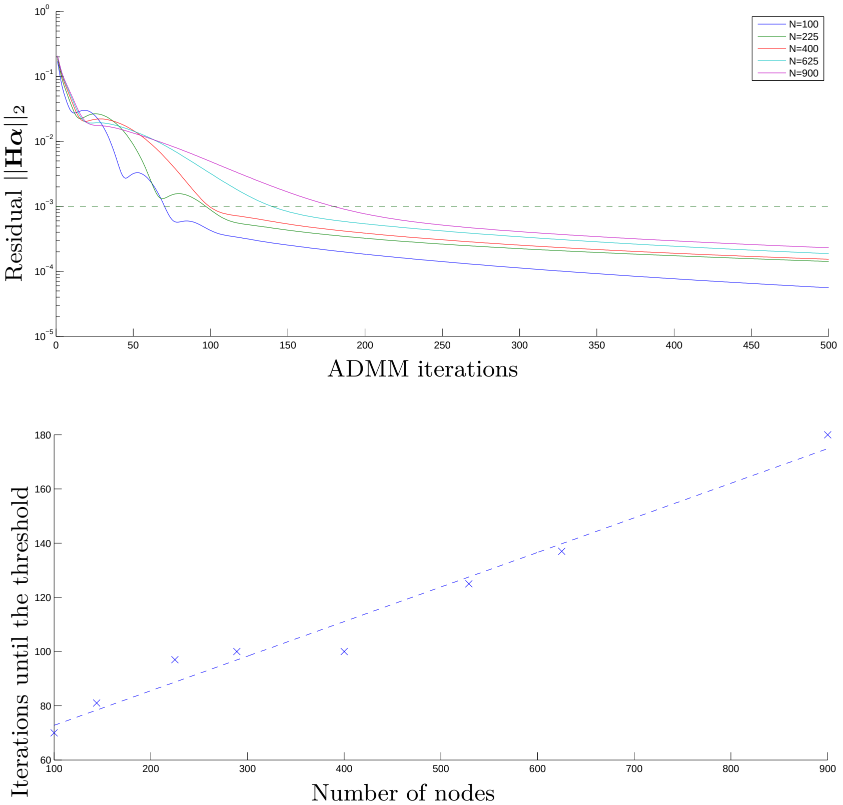

Figure 8-4: Residual error $||\mathbf{H}\boldsymbol{\alpha}||_2$ for problems with different sizes

Due to the fact that it is not possible to get a centralized solution due to the memory usage and computational complexity of the method, the threshold that will be used to ensure that the distributed solution is close enough to the optimum value is the residual $||\mathbf{H}\boldsymbol{\alpha}||_2$. If this value falls below a certain threshold, we consider the solution yielded to be acceptable.

Before analysing the results it is important to understand the dimensions of the problem at hand. Let us consider a grid with $N = 900$. Applying a Type II triangulation to this problem will result in 3600 simplices and 21600 coefficients to be estimated. The total amount of data points used in this estimation is $M = 562500$. These are the dimensions associated with the biggest problem analysed in Figure 8-4.

The results for different grid sizes are presented in Figure 8-4. In the upper plot, the convergence of the residual is depicted along with the threshold defined to ensure a certain level of feasibility. In the plot below, we plotted the number of iterations necessary to reach the threshold. Using linear regression it was determined that the number of iterations $\beta(N)$ necessary to reach the threshold is given by the following expression:

$$\beta(N) = 60.02 + 0.1276N. \tag{8-36}$$

This linear fit validates the theoretical analysis made in Section 8-4, where β was assumed to evolve linearly with N.

8-6-4 Comparison with Centralized Method

In order to compare the results obtained via the centralized and the structured ADMM application, we simulated wavefront reconstruction for an astigmatism aberration with different strengths using the nominal setup parameters and changing the splines parameters d and r. The ADMM method was run for 50 iterations using a stepsize of $\rho = 0.7$.

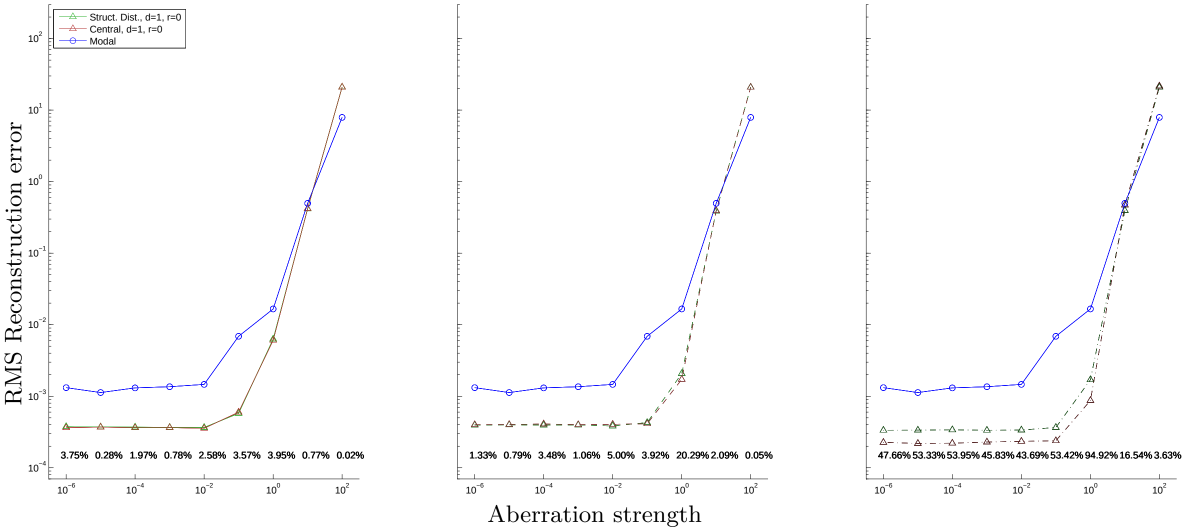

Figure 8-5: Comparison between the RMS values of the reconstruction error for modal reconstruction method, the centralized method, and the structured distributed method. The simulations were made for (left) $d = 1$ and $r = 0$, (center) $d = 2$ and $r = 1$, (right) $d = 2$ and $r = 2$. The reconstruction errors were averaged over 100 different noise distributions affecting the intensity measurements. Near the horizontal axis, the relative error between the centralized and distributed methods with respect to the centralized method is computed. The distributed method was simulated using $\rho = 0.7$ and 50 iterations.

Comparing the unstructured (see Figure 7-3) and the structured approach, it can be seen that using the same 50 iterations as in Section 7-4-3, the structured method led to a much lower reconstruction error.

The influence of the splines parameters d and r is also interesting to analyse, specially for the case where $d = 2$ and $r = 2$. Notice that, for $(d, r) = (1, 0)$ and for $(d, r) = (2, 1)$ the relative error is almost always below 5%, however, the error increases dramatically when $(d, r) = (2, 2)$. That is due to the fact that the aberration can be completely described by a second order polynomial and thus, the centralized method with tight continuity constraints is perfectly suitable for this specific case. The reason why the distributed method can not reach the centralized one in terms of accuracy can be explained with the fact that the satisfaction of the constraints has a much higher degree of slackness than the centralized method, where the feasibility is guaranteed and not approximated.

8-6-5 Closed-loop

The reconstruction errors for a static aberration in a closed-loop setting are presented in Figure 8-6. The same setup as in the open-loop comparison was used.

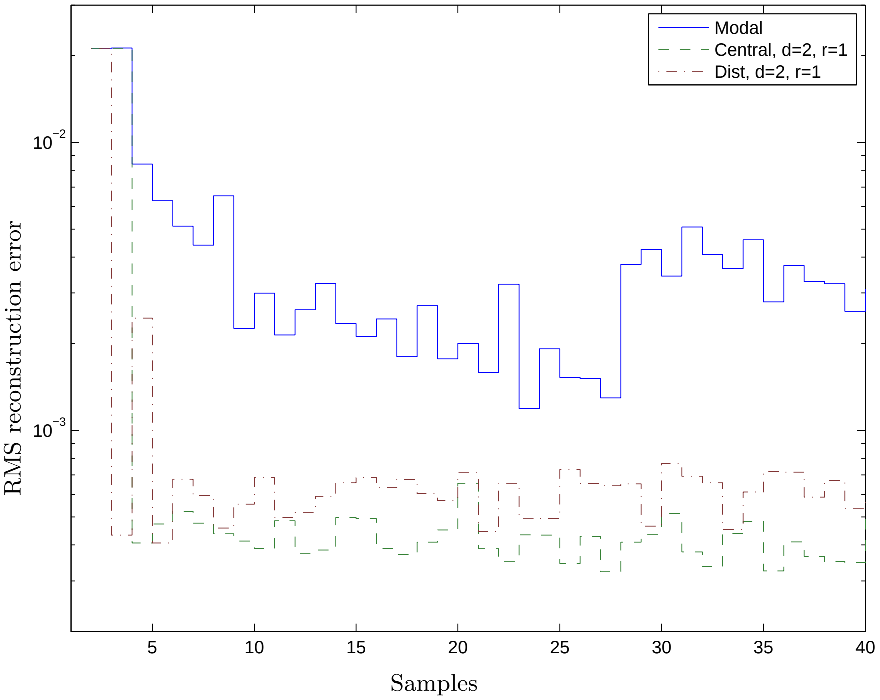

Figure 8-6: Comparison between the RMS values of the reconstruction error for modal reconstruction method, the centralized method, and the structured distributed method in closed loop for $d = 2$ and $r = 1$. The distributed method was simulated using $\rho = 0.7$ and 50 iterations. The aberration introduced was an astigmastism with the Zernike coefficient $\alpha_4 = 0.1\lambda$.

This distributed reconstruction method intrinsically relaxes the continuity constraints, whereas the centralized method abides by the continuity constraints much more. If we calculate the norm of the feasibility residual $||\mathbf{H}\boldsymbol{\alpha}||_2$ we get for the centralized method results between approximately 10^{-16} and 10^{-14}. In the distributed method that residual is between approximately 10^{-3} and 10^{-2}, thus, significantly higher.

There is however a pitfall associated with the fact that we do not get a low feasibility residual when the reconstruction is performed using a perfect mirror. Due to the fact that the mirror can take non-smooth shapes, the errors in completely satisfying the continuity constraints are accumulated until a irretrievable waffle-mode is created. This problem is completely associated with the presence of a perfect mirror and would not exist if a smooth mirror was used.

8-7 Final Remarks

The structured ADMM implementation was able to eliminate the pitfalls the unstructured implementation suffered from, namely:

- Due to the fact that the structured approach is based on a 2-splitting (canonical separation in $\mathbf{x}$ and $\mathbf{z}$) and not in an N-splitting the method converges for any stepsize $\rho > 0$. Finding the optimum stepsize is still a heuristic procedure that depends on the setup configuration and on the type of aberrations involved.

- The convergence rate is much faster in the structured method. For most of the splines configuration parameters we can reach a relative error of 5% with respect to the centralized method within 50 iterations. We used a Gauss-Seidel approach where the coupling variables $\mathbf{z}$ are only updated after all the primal variables $\mathbf{x}$ have been calculated which leads to a better *approximate* joint minimization than the Jacobi update used in the unstructured update equations. It is noteworthy that the convergence is not guaranteed to be monotonic for this method.

- All the updates of the primal variables $\mathbf{x}_i$ rely only on information contained in the neighbouring nodes of node i. The updates of the coupling variable and dual variables rely only on knowing the primal variable values on the edges. This leads to a fully distributed approach where no global update and global broadcast operations are required.

These improvements led to a much faster convergence rate and thus made us achieve a very low reconstruction error close to the centralized optimum.

8-8 Future Work

This section presents three major improvement topics, namely, speeding up the computations, using a gain scheduling approach to improve convergence, and develop a distributed termination criterion.

8-8-1 Further speed-up

This method can be further sped up if the updates of the coupling variables $\mathbf{z}$ are made in parallel to the update of the primal variables $\mathbf{x}$. That would change the update equation (8-26) into:

$$\mathbf{z}_{ij}(k+1) = \frac{\mathbf{G}_{ij}\boldsymbol{\alpha}_i(k) - \mathbf{F}_{ij}\boldsymbol{\alpha}_j(k)}{2}. \tag{8-37}$$

Having this type of update would most likely jeopardize the convergence speed. However, for a specific setup it might be more profitable in terms of time to have this update made in parallel.

Changing the penalty parameter ρ at certain time steps can also improve the rate of convergence, although it is difficult to present theoretical results to justify those heuristic changes. However, most of these heuristics involve computing the residuals on a centralized fashion (see [71, § 3.4.1], and the references therein).

8-8-2 Gain Scheduling Approach

Regarding future developments of this method, it would be interesting to analyse the reconstruction results for a hierarchical constraint application. Such a method would start by applying C^0 continuity constraints, then C^1, and so on. The intuition behind such a proposal is that the biggest errors are due to errors in abiding by C^0 constraints. If the method was made such that we could force continuity very fast and then proceed to smoothen the first-derivative continuity, we might be able to achieve faster convergence.

If we implemented such an approach in this ADMM scheme, we would introduce extra constraints at pre-determined iteration steps turning the problem into a hybrid optimization. For instance, from iteration 1 to 20, only C^0 constraints would be active. From iteration 21 to 30, both C^0 and C^1 continuity would be active. And so on for increasing degrees of continuity.

However, this approach has two major pitfalls:

- Due to the hybrid nature of the algorithm, even if a certain stepsize guarantees convergence for a certain set of active constraints, it may not for another. A solid convergence proof under this conditions may be very cumbersome.
- For each different set of active constraints, different matrices need to be inverted. Thus, if the method involves a great number of switches, many different matrices must be pre-computed to guarantee real-time applicability.

An additional remark ought to be made. Due to the linear least-squares nature of the problem, the primal variables update does not depend on the previous estimate of the primal variables. Thus, when we change the method used and introduce additional constraints, the dual variable will play the role of initial condition.

8-8-3 Distributed Rermination Criteria

An important question that needs to be addressed is the termination of the algorithm. Until this moment the main aspect that determined the accuracy of the algorithm was given by a fixed number of iterations. However, it would be interesting for applications that rely on ADMM to come up with a distributed termination criterion that does not depend on collecting all the variables from all the different nodes as per Section 4-4.

The termination procedure presented in [71] requires a global gathering of the primal variables and the coupling variables. Having such a termination criterion is perfectly adequate when the connectivity constraints that we are dealing with are virtual, such as parallel implementations in a GPU framework. However, if the application of the algorithm concerns a multi-agent environment with no central computing unit, having a distributed way of finding when to stop the algorithm is of the utmost importance. If we use a fixed number of iterations based on some heuristics or experimental data we are always in danger of not reaching the performance level we wanted, or of optimizing beyond the necessary requirements.

In [71], the algorithm stops when, at a certain iteration k, the following bounds on the primal and dual residuals are respected:

$$||\mathbf{r}(k)||_2 \leq \epsilon^{\text{pri}}, \quad ||\mathbf{s}(k)||_2 \leq \epsilon^{\text{dual}}, \tag{8-38}$$

where the residuals are given by

$$\mathbf{r}(k) = \mathbf{A}_{\text{res}}\mathbf{x}(k) + \mathbf{B}_{\text{res}}\mathbf{z}(k), \tag{8-39}$$

$$\mathbf{s}(k) = -\rho\mathbf{A}_{\text{res}}^{\top}\mathbf{B}_{\text{res}}\big(\mathbf{z}(k) - \mathbf{z}(k-1)\big). \tag{8-40}$$

For now let us assume that the termination thresholds ϵ^{pri} and ϵ^{dual} are constants chosen *a priori*.

Problem formulation. The primal residual vector $\mathbf{r}$ in the k-th iteration can be decomposed in N sub vectors each corresponding to a sub-problem associated with a time step in the horizon (the boundary sub-problem at time $t = 0$ can be discarded). The primal residual vector can be written as

$$\mathbf{r}^{\top}(k) = \begin{bmatrix} (\mathbf{r}_1^k)^{\top} & (\mathbf{r}_2^k)^{\top} & \cdots & (\mathbf{r}_N^k)^{\top} \end{bmatrix}, \tag{8-41}$$

where the vector $(\mathbf{r}_i^k)^{\top}$ is known to node i.

We are now faced with the problem of the undecomposability of the 2-norm in (8-38), that is, this norm can not be calculated in a distributed manner. However, taking the square of the norm allows us to write the following inequality:

$$||\mathbf{r}(k)||_2^2 = \mathbf{e}_{\text{pri}} \leq (\epsilon^{\text{pri}})^2. \tag{8-42}$$

Let us now focus on the dual residual. We can find an upper bound for the 2-norm of the residual by making use of an induced matrix norm which leads to the following inequalities:

$$||\mathbf{s}(k)||_2 \leq \rho\big|\big|\mathbf{A}_{\text{res}}^{\top}\mathbf{B}_{\text{res}}\big|\big|_2\big|\big|\mathbf{z}(k) - \mathbf{z}(k-1)\big|\big|_2 \leq \epsilon^{\text{dual}} \tag{8-43}$$

Considering $\mathbf{e}_{\text{dual}}(k) = \big|\big|\mathbf{z}(k) - \mathbf{z}(k-1)\big|\big|_2^2$ and $\xi^{\text{dual}} = \epsilon^{\text{dual}}/\big|\big|\mathbf{A}_{\text{res}}^{\top}\mathbf{B}_{\text{res}}\big|\big|_2$, we can rewrite the last inequality as

$$\mathbf{e}_{\text{dual}}(k) \leq (\xi^{\text{dual}})^2, \tag{8-44}$$

where the vector $\mathbf{e}_{\text{dual}}(k)$ can be decomposed as

$$\mathbf{e}_{\text{dual}}(k) = \left[||\mathbf{z}_1(k) - \mathbf{z}_1(k-1)||_2^2, \; \cdots \; , ||\mathbf{z}_{N(N-1)/2}(k) - \mathbf{z}_{N(N-1)/2}(k-1)||_2^2 \right], \tag{8-45}$$

allowing each pair of nodes (link) to compute its correspondent error $||\mathbf{z}_j(k) - \mathbf{z}_j(k-1)||_2^2$ with only local information.

Now that we have redesigned the stopping criteria for the primal and dual residual, we only have to solve a consensus problem with an arbitrary number of iterations M after each iteration of the main algorithm.

Let $\mathcal{G}$ be a time invariant graph (its structure does not change with time) with vertices contained in the set $\mathcal{V}$ and edges in the set $\mathcal{E}$. The problem of finding the primal and dual residual (respectively) can then be stated as follows

$$\mathbf{e}_{\text{pri}}(t+1) = \mathbf{W}_1 \mathbf{e}_{\text{pri}}(k+1), \tag{8-46}$$
$$\mathbf{e}_{\text{dual}}(t+1) = \mathbf{W}_2 \mathbf{e}_{\text{dual}}(k+1), \tag{8-47}$$

where t denotes the iteration counter for the consensus algorithm, the matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ are doubly stochastic (square matrix of non-negative values of which the rows and columns sum to 1) and the graph obeys some mild connectivity assumptions. The values of the elements of $\mathbf{e}_{\text{pri}}$ and $\mathbf{e}_{\text{dual}}$ will eventually converge to the mean value. Multiplying this mean value by the number of subvectors in the residual vectors will give us the sum.

After a number of iterations M, we may be able to guarantee that we are within some tolerance of the true mean residual value. A way to stop the algorithm would be to have each node check if they comply with the pre-determined stopping criterion given by $(\epsilon^{\text{pri}})^2$ and $(\epsilon^{\text{dual}})^2$. If a node reaches this situation, it could broadcast a stopping signal to all other nodes and halt the algorithm.

Related references. According to the recent survey from Olshevsky and Tsitsiklis [93][1]the consensus method which seems to provide the fastest convergence rate was presented by Xiao and Boyd [94]. In [94], the authors propose an optimization method to design the matrix $\mathbf{W}$ such that the convergence rate is maximized.

[1] This survey was first published in CDC'06 and then slightly updated for SIAM in 2011.

Chapter 9

Conclusions

For more detailed conclusions and further research that needs to be done concerning each of the methods, we refer the reader to the section containing the final remarks at the end of each chapter.

The research conducted during this thesis led to the development of a novel wavefront reconstruction method that encompasses both pupil-plane wavefront sensing and detector-plane wavefront reconstruction, by making use of the structure of the (Shack-)Hartmann sensor and the locality of the phase parametrization by B-splines.

The fundamental idea of the novel method was to replace the slope approximations and use directly the relationship between the phase and the intensity measurements, thus not performing any approximation and using all the raw data. The other idea upon which the method is based is the fact that the phase can be reconstructed locally from the local intensity measurements resulting from the sampling of the wavefront. The third aspect of this method relies on the parametrization of the wavefront using B-splines which provides a solid framework to smoothly connect the local reconstructions. The use of B-splines also supports the parallelization of the estimation method due to the fact that the splines basis functions are only defined locally and not globally.

This method is applicable under the assumptions that the distance that separates the pupil- and the focal-plane is small (≤ 10 mm), the resolution of each pixel of the CCD is continuous (*i.e.*, the quantization effects were not taken into account), and that the noise affecting the normalized intensities is small ($\sigma_{\text{ccd}} \approx 10^{-4}$).

If those specifications are met, the centralized method presented in Chapter 5 provides better results in terms of wavefront reconstruction errors if compared to a modal reconstruction method (that uses wavefront slope approximations and parametrizes the wavefront using a global Zernike mode), for small aberrations (with an RMS value $\leq \lambda$, where λ represents the wavelength). The variance of the reconstruction error using the novel estimator is also significantly improved, up to 2 orders of magnitude for any aberration strength.

Regarding the compressive sampling method presented in Chapter 6, we were able to empirically show that we can reduce the amount of data by 90% without any loss of accuracy, both in open- and closed-loop. This method was based on an analysis of the Jacobian of the intensity measurements with respect to the phase which yields the pixels which are most relevant and contain more useful information.

Two distributed approaches were developed and tested to speed-up the estimation of the wavefront aberration compared to the centralized method. The first approach, presented in Chapter 7 is called the unstructured method and is a simple application of the Alternating Direction Method of Multipliers (ADMM) that does not fully explore the properties of the smoothness matrix inherent to the B-splines framework. This approach converges for small enough stepsizes but at a very slow rate.

The second approach is presented in Chapter 8 and is designated the structured method. The name stems from the fact that contrary to the unstructured method, the structured one takes into account the properties of the smoothness matrix which allows for a complete parallel update of all the variables involved requiring that only the neighbouring nodes exchange information. The results showed that within dozens of iterations the distributed solution reaches the optimizer of the centralized method within a 5% margin in terms of the RMS value of the wavefront reconstruction error.

A challenging future research possibility would be to extend the B-splines parametrization to control the deformable mirror. One of the most interesting results that could come out of this research would be to provide a unified distributed estimation and control method operating always in the reduced space of splines coefficients.

Provided that an experimental setup that matches the assumptions postulated in the first paragraphs of this conclusion, the method can be implemented and tested in a real-time application. The distributed method can even be implemented in a GPU/FPGA environment which would increase the scalability and the speed of the computations immensely.

Appendix A

Model of the Shack-Hartmann Wavefront Sensor

A-1 Intensity Distribution Model

The deductions in this section were based mainly on Chapters 3 to 5 of [26] and the work done in [95]. Some notation (e.g., the one used to represent the magnetic and electric fields) does not comply with the nomenclature in this thesis. However, we believe that given the context this ambiguity should not confuse the reader.

When using model-based wavefront reconstruction algorithms, the model chosen must be as accurate as possible. To this end, one must fully understand how to derive implementable algorithms (such as the ones in [75] and [96]) from plain Maxwell's equations so that all approximations and assumptions are taken into consideration.

The first step is, then, to write Maxwell's equations in the absence of free charge and free current:

$$\nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t}, \tag{A-1}$$

$$\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t}, \tag{A-2}$$

$$\nabla \cdot \epsilon \mathbf{E} = 0, \tag{A-3}$$

$$\nabla \cdot \mu \mathbf{H} = 0, \tag{A-4}$$

where $\mathbf{E}$ represents the electric field, $\mathbf{H}$ the magnetizing field, ϵ the permittivity of the medium and μ the permeability of the medium.

Before proceeding with the derivations some properties of the medium must be explicited.

- The medium is *linear* if it does not change its properties due to the intensity of the light.

- The medium is *isotropic* when its properties do not depend on the direction of the fields.
- *Homogeneity* guarantees that the permittivity ϵ of the medium is constant.
- The medium is *non-dispersive* if it does not depend on the wavelength region of the wave.
- The medium is *non-magnetic* when the permeability μ equals the permeability in vacuum μ_0.

If all of these properties are verified (or assumed to be verified) one can decouple each of the components of the fields **E** and **H** from one another and, thus, write the following expression

$$\nabla^2 u(\mathbf{x},t) - \frac{n^2}{c^2}\frac{\partial u(\mathbf{x},t)}{\partial t^2} = 0, \tag{A-5}$$

where the scalar term u can be replaced by any of the components of **E** or **H**. The vector $\mathbf{x} \in \mathbb{R}^3$ represents the position. The refractive index of the medium is given by n and the velocity of light propagation by c.

For a monochromatic wave we can write

$$u(\mathbf{x},t) = A(\mathbf{x})\cos(2\pi\nu t + \phi(\mathbf{x})), \tag{A-6}$$

which can be also written as

$$u(\mathbf{x},t) = \mathrm{Re}\{U(\mathbf{x})\exp(-j2\pi\nu t)\}, \tag{A-7}$$

with

$$U(\mathbf{x}) = A(\mathbf{x})\exp(-i\phi(\mathbf{x})), \tag{A-8}$$

Equations (A-6) and (A-7) show that the time dependence can be dropped as it can be immediately retrieved from $U(\mathbf{x})$. Substituting Eq. (A-7) in (A-5) yields the Helmoltz equation in U given by

$$(\nabla^2 + k)U(\mathbf{x}) = 0, \tag{A-9}$$

where $k = 2\pi/\lambda$ is the wavenumber.

The application of Green's theorem [26, Section 3.2.2] under the Rayleigh-Sommerfeld formulation of diffraction [26, Section 3.5] enables us to solve the Helmholtz equation in Eq. (A-9) in terms of the quantity U. The solution obtained is called the Huygens-Fresnel principle. For our purposes, we will consider that the propagation is done in the direction of the coordinate z. Furthermore, the electro-magnetic field in the diffracting aperture is defined in a plane (x, y) and the field after propagation in a parallel plane with its points parametrized by the coordinates (u, v) (similar notation as Eq. (2-5); Figure 2-3 may help visualize the planes to which we refer). Given these parametrizations and imposed propagation conditions, we can write, without loss of generality, the propagated complex field according to the Huygens-Fresnel principle in rectangular coordinates as

$$U(u,v) = \frac{z}{i\lambda}\iint U(x,y)\frac{e^{ikr}}{r^2}\mathrm{d}x\mathrm{d}y, \tag{A-10}$$

where $r = \sqrt{z^2 + (u-x)^2 + (v-y)^2}$. The only assumption made is that the propagation distance z must be much bigger than λ so that we can consider far-field propagation.

In order to calculate this field in an efficient manner, we can rewrite (A-10) into a convolution integral [75, § 4.4] as

$$U(u,v) = \iint U(x,y)h(u-x, v-y)\mathrm{d}x\mathrm{d}y, \tag{A-11}$$

where the general form of the Rayleigh-Sommerfeld impulse response is

$$h(\xi,\eta) = \frac{z}{i\lambda}\frac{e^{ikr}}{r^2} \tag{A-12}$$

with r being defined as

$$r = \sqrt{z^2 + \xi^2 + \eta^2}. \tag{A-13}$$

Expressing the convolution in the spatial frequency domain allows rewriting the expression (A-11) using Fourier transforms as

$$U(u,v) = \mathscr{F}^{-1}[\mathscr{F}[U(x,y)]H(f_x,f_y)], \tag{A-14}$$

where the Rayleigh-Sommerfeld transfer function is defined as

$$H(f_x,f_y) := \mathscr{F}[h(x,y)] = \exp\left(ikz\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right) \tag{A-15}$$

Another approximation can be made to simplify Eq. (A-10) such that it can be evaluated analytically. The approximation is called the Fresnel approximation and it consists in truncating the term r^2. The resulting expression is given by

$$U(u,v) = \frac{e^{ikz}}{i\lambda z}e^{\frac{ik}{2z}(u^2+v^2)}\iint U(x,y)e^{\frac{ik}{2z}(x^2+y^2)}e^{\frac{-2\pi i}{\lambda z}(ux+vy)}\mathrm{d}x\mathrm{d}y. \tag{A-16}$$

When a lens is introduced, the phase of the light is changed due to the introduction of a medium with a different refractive index. The complex field immediately before and after the transmission through the lens can be described by

$$U_{\text{after lens}}(x,y) = e^{-\frac{ik}{2f}(x^2+y^2)}U_{\text{before lens}}(x,y). \tag{A-17}$$

The exponential term provides the phase shift introduced by a lens with a certain focal length f and with its optical axis aligned with the direction z. Note, however, that this expression describes accurately the lens only for (x,y) pairs close to the optical axis of the lens.

If we introduce the resulting complex field from (A-17) in (A-16), we obtain

$$U(u,v) = \frac{e^{ikz}}{i\lambda z}e^{\frac{ik}{2z}(u^2+v^2)}\iint U_{\text{before lens}}(x,y)e^{\frac{-2\pi i}{\lambda z}(ux+vy)}\mathrm{d}x\mathrm{d}y. \tag{A-18}$$

Notice that the first exponential term is a constant phase shift and can thus be dropped. To obtain the intensity distribution $I(u,v)$ that appears in Eq. (2-5) we simply take the square of the magnitude of $U(u,v)$, hence

$$I(u,v) = |U(u,v)|^2. \tag{A-19}$$

A-2 Slope Model

The derivation presented is based on [2].

Let us assume that the wavefront is flat. The distance between the lenslet (or aperture) array is given by z. If the wavefront is aberrated, the displacement of the centroid regarding its ideal position is given by Δx and Δy.

The wavefront can then be approximated as only a tilt and its slope may be computed as follows, provided that α_x and α_y are the degrees of the tilt:

$$\tan(\alpha_x) = \frac{\Delta x}{z}. \tag{A-20}$$

Using the small-angle approximation we have that $\tan(\theta) \approx \theta$, $\theta \approx 0$. Thus,

$$\alpha_x \approx \frac{\Delta x}{z}. \tag{A-21}$$

The same procedure can be replicated for the tilt in the y direction.

By multiplying by the distance z on both sides of Eq. (A-21) and considering the presence of noise we arrive at Eq.(2-4)

Appendix B

Statistical Properties of the Novel Method

Given that the new estimation method is now completely defined, we can proceed to analyse its statistical properties, by comparing the splines phase estimator $\hat{\boldsymbol{\phi}}_{\text{spl}}$ with the modal reconstruction estimator $\hat{\boldsymbol{\phi}}_{\text{mod}}$ using Zernike polynomials as presented in Section 2-3-2. Let us define the reconstruction problems using Zernike polynomials (see Eq. (2-11)) and splines (see Eq. (5-17)) in such a way as to put an extra emphasis on the noise minimization:

$$\begin{aligned} \underset{\boldsymbol{\beta}}{\text{minimize}} \quad & \boldsymbol{\epsilon}_{\text{mod}}^{\top}\boldsymbol{\epsilon}_{\text{mod}} \\ \text{subject to} \quad & \mathbf{s} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}_{\text{mod}} \end{aligned}, \tag{B-1}$$

$$\tag{B-2}$$

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\epsilon}_{\text{spl}}^{\top}\boldsymbol{\epsilon}_{\text{spl}} \\ \text{subject to} \quad & \mathbf{i}_{\text{meas}} = \mathbf{c}_0 + \mathbf{C}_1\boldsymbol{\alpha} + \boldsymbol{\epsilon}_{\text{spl}}, \\ & \mathbf{H}\boldsymbol{\alpha} = 0 \end{aligned} \tag{B-3}$$

with the variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denoting the local splines coefficients and the Zernike coefficients, respectively. The vectors $\mathbf{s}$ and $\mathbf{i}_{\text{meas}}$ represent the measured intensity values and the slope measurements, respectively.

The noise vectors $\boldsymbol{\epsilon}_{\text{spl}}$ and $\boldsymbol{\epsilon}_{\text{mod}}$ are characterized by the following normal distributions:

$$\boldsymbol{\epsilon}_{\text{spl}} \sim \mathcal{N}(0, \sigma_{\text{ccd}}^2\mathbf{I}), \quad \boldsymbol{\epsilon}_{\text{mod}} \sim \mathcal{N}(0, \sigma_s^2\mathbf{I}). \tag{B-4}$$

The noise $\boldsymbol{\epsilon}_{\text{spl}}$ affects directly the intensity values and can be retrieved from the specifications provided by a CCD camera. This noise affecting the pixels is white and Gaussian with zero-mean and covariance matrix $\sigma_{\text{ccd}}^2\mathbf{I}$. The other noise vector $\boldsymbol{\epsilon}_{\text{mod}}$ corrupts the slope measurements $\mathbf{s}$. Due to having uncorrelated noise corrupting the pixels (*i.e.*, the noise in one pixel is independent from the noise in another pixel), the noise in the slopes is also uncorrelated from one subaperture to the other. Hence, the covariance matrix being an identity matrix multiplied by a scalar.

Before proceeding to characterize the phase estimators $\hat{\boldsymbol{\phi}}_{\text{spl}}$ and $\hat{\boldsymbol{\phi}}_{\text{mod}}$, let us provide values for the standard deviations σ_s and σ_{ccd}.

Intensity noise

Concerning the noise that affects the intensity measurements, we considered only measurement noise, and disregarded all other sources of noise. According to [12], a typical value for the standard deviation of the measurement noise on a commercial CCD camera is given by $\sigma_{\text{ccd}} = 4 \times 10^{-4}$.

Slopes noise

In [97], the authors present a relation between the variance of the noise that affects the intensity measurements and the variance of the slope noise which is repeated below:

$$\sigma_s^2 = \frac{1}{z}\frac{\sigma_{\text{ccd}}^2}{V^2}L^2\left(\frac{L^2-1}{12} + \mathbf{s}_{\text{ideal}}\right). \tag{B-5}$$

The constant σ_{ccd} represents the standard deviation of the noise that affects the normalized intensity measurements, considered to be 4×10^{-4}. The constant L is the number of pixels both in the y and the x direction. The value of $\mathbf{s}_{\text{ideal}}$ corresponds to the position of the ideal slope (without any phase aberrations) because we are dealing with static aberrations (at the moment). The value of V corresponds to the sum of the normalized intensity values that are used to compute the center of mass.

Notice the factor $1/z$ that appears due to the transformation from center of the spot deviations to slopes, according to the simplifications presented in Appendix A-2.

B-1 Modal Phase Estimator

The phase coefficients that are the optimal solution of the problem in (B-1) are obtained via a least-squares approach. Given that the covariance matrix of the noise is an identity matrix multiplied by a constant the minimum-variance and unbiased least-squares estimator can be calculated as follows:

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^{+}\mathbf{s}, \tag{B-6}$$

where $\mathbf{A}^{+} = (\mathbf{A}^{\top}\mathbf{A})^{-1}\mathbf{A}^{\top}$ represents the pseudo-inverse of matrix $\mathbf{A}$.

The estimator $\hat{\boldsymbol{\beta}}$ is unbiased, that is, $\langle\hat{\boldsymbol{\beta}}\rangle = \boldsymbol{\beta}$, and its covariance matrix is given by $\Sigma_{\hat{\boldsymbol{\beta}}} = (\mathbf{A}^{\top}\mathbf{A})^{-1}\sigma_s^2$ [44].

The phase can be retrieved by multiplying each of the modal coefficients by its corresponding mode. This is a linear operation that can be denoted as

$$\hat{\boldsymbol{\phi}}_{\text{mod}} = \mathbf{Z}\hat{\boldsymbol{\beta}}, \tag{B-7}$$

where the matrix $\mathbf{Z}$ is generated by placing in each of its columns the corresponding vectorized Zernike mode and where covariance matrix of $\hat{\boldsymbol{\phi}}_{\text{modal}}$ is

$$\Sigma_{\hat{\boldsymbol{\phi}}_{\text{mod}}} = \mathbf{Z}(\mathbf{A}^\top \mathbf{A})^{-1}\mathbf{Z}^\top \sigma_s^2. \tag{B-8}$$

B-2 Splines phase Estimator

The minimization problem in (5-19) can be cast into an unconstrained linear least-squares problem if we apply the following change of variables:

$$\boldsymbol{\alpha} = \mathbf{N}_H \boldsymbol{\alpha}_{\text{red}}, \tag{B-9}$$

which was already presented in Section 5-2-3, and where $\mathbf{N}_H = \text{Null}(\mathbf{H})$.

Using this change of variables, we can rewrite the problem in (B-3) as an unconstrained linear least-squares problem:

$$\begin{aligned} &\underset{\mathbf{c}}{\text{minimize}} && \boldsymbol{\epsilon}_{\text{spl}}^\top \boldsymbol{\epsilon}_{\text{spl}} \\ &\text{subject to} && \mathbf{i}_{\text{meas}} = \mathbf{c}_0 + \mathbf{C}_1 \mathbf{N}_H \mathbf{c}_{\text{red}} + \boldsymbol{\epsilon}_{\text{spl}}. \end{aligned} \tag{B-10}$$

Defining $\mathbf{C}_{1r}$ as $\mathbf{C}_1\mathbf{N}_H$ we can deduce the following least-squares estimator:

$$\hat{\boldsymbol{\alpha}} = \mathbf{N}_H \hat{\boldsymbol{\alpha}}_{\text{red}} = \mathbf{N}_H \mathbf{C}_{1r}^{+}(\mathbf{i}_{\text{meas}} - \mathbf{c}_0). \tag{B-11}$$

The estimator $\hat{\boldsymbol{\alpha}}$ is unbiased and has covariance matrix given by $\Sigma_{\hat{\boldsymbol{\alpha}}} = \mathbf{N}_H(\mathbf{C}_{1r}^\top \mathbf{C}_{1r})\mathbf{N}_H^\top$.

The phase can then be reconstructed using a linear operation such as

$$\hat{\boldsymbol{\phi}}_{\text{spl}} = \mathbf{B}\hat{\boldsymbol{\alpha}}, \tag{B-12}$$

where matrix $\mathbf{B}$ denotes the vectorized local splines basis functions which will be weighted by the coefficients in $\hat{\boldsymbol{\alpha}}$.

This estimator is unbiased and has a covariance matrix given by

$$\Sigma_{\hat{\boldsymbol{\phi}}_{\text{spl}}} = \mathbf{B}\mathbf{N}_H(\mathbf{C}_{1r}^\top \mathbf{C}_{1r})^{-1}\mathbf{N}_H^\top \mathbf{B}^\top \sigma_{\text{ccd}}^2 \tag{B-13}$$

Appendix C

Auxiliar Derivations to support ADMM

In the following sections, some auxiliar derivations that led to the final expressions provided in the corpus of this thesis are presented.

C-1 Unstructured ADMM Derivations

C-1-1 Update laws

This section provides a derivation to the minimization of the local augmented Lagrangian $\mathcal{L}_{i,\rho}$ in Eq. (7-8). To perform the minimization one can set the Jacobian with respect to a certain $\boldsymbol{\alpha}_i$ to zero:

$$\nabla_{\boldsymbol{\alpha}_i}\mathcal{L}_{i,\rho} = \nabla_{\boldsymbol{\alpha}_i}\left(||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2 + \mathbf{y}^\top(k)\mathbf{H}_i\boldsymbol{\alpha}_i + (\rho/2)||\mathbf{H}_i\boldsymbol{\alpha}_i + \sum_{j=1,j\neq i}^{N}\mathbf{H}_j\boldsymbol{\alpha}_j(k)||_2^2\right) = 0 \tag{C-1}$$

One can split the previous equation into three independent terms which can be analysed individually.

$$\nabla_{\boldsymbol{\alpha}_i}\left(||\mathbf{g}_i - \mathbf{C}_{1,i}\boldsymbol{\alpha}_i||_2^2\right) = 2\mathbf{C}_{1,i}^\top\mathbf{C}_{1,i}\boldsymbol{\alpha}_i - 2\mathbf{C}_{1,i}^\top\mathbf{g}_i \tag{C-2}$$

$$\nabla_{\boldsymbol{\alpha}_i}\left(\mathbf{y}^\top(k)\mathbf{H}_i\boldsymbol{\alpha}_i\right) = \mathbf{H}_i^\top\mathbf{y}(k) \tag{C-3}$$

$$\nabla_{\boldsymbol{\alpha}_i}\left(\frac{\rho}{2}||\mathbf{H}_i\boldsymbol{\alpha}_i + \sum_{j=1,j\neq i}^{N}\mathbf{H}_j\boldsymbol{\alpha}_j(k)||_2^2\right) = \rho\left(\mathbf{H}_i^\top\mathbf{H}_i\boldsymbol{\alpha}_i + \mathbf{H}_i^\top\sum_{j=1,j\neq i}^{N}\mathbf{H}_j\boldsymbol{\alpha}_j(k)\right) \tag{C-4}$$

The minimum value for $\mathcal{L}_{\rho,i}$ is achieved with $\boldsymbol{\alpha}_i(k+1)$ which can then be written as

$$\boldsymbol{\alpha}_i(k+1) = (2\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i} + \rho \mathbf{H}_i^\top \mathbf{H}_i)^{-1}(2\mathbf{C}_{1,i}^\top \mathbf{g}_i - \mathbf{H}_i^\top \mathbf{y}(k) + \rho \mathbf{H}_i^\top \sum_{j=1, j\neq i}^{N} \mathbf{H}_j \boldsymbol{\alpha}_j(k)). \tag{C-5}$$

C-1-2 Convergence Proof for N-splitting with Quadratic Cost Function

A foolproof approach that would not depend on any heuristic tuning would be to work under the following assumption:

Assumption 4. *The matrix $\mathbf{C}_1^\top \mathbf{C}_1$ is positive definite on the kernel of $\mathbf{H}$, i.e., $\forall \mathbf{v} \in Null(\mathbf{H})$, $\mathbf{v}^\top \mathbf{C}_1^\top \mathbf{C}_1 \mathbf{v} > 0$.*

Assumption 4 allows $\mathbf{C}_1$ to be rank deficient. Consequently, a cost function defined as $J(\boldsymbol{\alpha}) = ||\mathbf{C}_1\boldsymbol{\alpha} - \mathbf{g}||_2^2$ is strictly convex[1]on the nullspace of $\mathbf{H}$, although it may not be strictly convex over $\mathbb{R}^{\hat{d}T_s N}$.

Assumption 4 can also lead to the following proposition.

Proposition 1. *If the matrix $\mathbf{C}_1^\top \mathbf{C}_1$ is positive definite on the kernel of $\mathbf{H}$ (Assumption 4), then the matrices $\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i}$ are positive definite on the kernels of $\mathbf{H}_i$, respectively.*

Proposition 1 shows that under Assumption 4 we can guarantee that the matrix $(2\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i} + \rho \mathbf{H}_i^\top \mathbf{H}_i)$ is invertible through the following lemma.

Lemma 1. *Consider the centralized and augmented cost function given by*

$$J_{aug}(\boldsymbol{\alpha}) = ||\mathbf{C}_1\boldsymbol{\alpha} - \mathbf{g}||_2^2 + \rho||\mathbf{H}\boldsymbol{\alpha}||_2^2$$

and suppose Assumption 4 is satisfied, then $J_{aug}(\boldsymbol{\alpha})$ is strongly convex[2]on its full domain and the matrix $\mathbf{C}_1^\top \mathbf{C}_1 + \rho \mathbf{H}^\top \mathbf{H}$ is strictly positive definite (and thus, invertible).

This lemma is immediately applicable to the problem in (7-5) and we can thus say that the augmented Lagrangian in (7-6) is strongly convex for each $\boldsymbol{\alpha}_i$ and that $\mathbf{C}_{1,i}^\top \mathbf{C}_{1,i} + \rho \mathbf{H}_i^\top \mathbf{H}_i$ is strictly positive definite. Note that the solution for the equivalent problem with the augmented cost function in Lemma 1 is exactly the same as with the unaugmented cost function given that if we are at a feasible optimum $\boldsymbol{\alpha}^\star$, the quantity $\mathbf{H}\boldsymbol{\alpha}^\star$ is zero, and the extra term $\rho||\mathbf{H}\boldsymbol{\alpha}||_2^2$ disappears.

In order to prove the convergence of the algorithm we need two more lemmas. The first lemma states that the gradient of the augmented Lagrangian function (7-6) at any iteration k can be upper bounded by the change $(\boldsymbol{\alpha}(k+1) - \boldsymbol{\alpha}(k))$ induced in the primal variables.

[1] The difference between a strictly convex and a convex function is that the former only has one global minimum.

[2] For quadratic functions strong convexity (and strict convexity) imply that the Hessian must be positive definite.

Lemma 2. *Let* $\{\boldsymbol{\alpha}(k), \mathbf{y}\}_{k=0}^{\infty}$ *denote the sequence generated by the ADMM algorithm* (7-8) *and* (7-9)*. There exists a constant* $\eta \in \mathbb{R} > 0$ *such that*

$$||\nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k), \mathbf{y}(k))||_2 \leq \eta \sqrt{N} ||\boldsymbol{\alpha}(k+1) - \boldsymbol{\alpha}(k)||_2.$$

Proof. Check the proof for Lemma 2.5 in [87]. □

The second lemma bounds the error caused by the inexact minimization of the augmented Lagrangian (7-6).

Lemma 3. *Consider the sequence* $\{\boldsymbol{\alpha}(k), \mathbf{y}\}_{k=0}^{\infty}$*. There is a* $\tau \in \mathbb{R}$ *for which the following error bound holds:*

$$||\boldsymbol{\alpha}(k) - \bar{\boldsymbol{\alpha}}(\mathbf{y}(k))||_2 \leq \tau ||\nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k), \mathbf{y}(k))||_2$$

where $\bar{\boldsymbol{\alpha}}(\mathbf{y}(k)) = \arg\min_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}, \mathbf{y}(k))$ *and* $\tau = ||(\mathbf{C}_{1,i}^{\top}\mathbf{C}_{1,i} + \rho \mathbf{H}^{\top}\mathbf{H})^{-1}||_2$

Proof. Let $\mathbf{X}_i = (\mathbf{C}_1^{\top}\mathbf{C}_1 + \rho \mathbf{H}^{\top}\mathbf{H})$. We then have that

$$\begin{aligned}
\nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k), \mathbf{y}(k)) &= \nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k) - \bar{\boldsymbol{\alpha}}(k) + \bar{\boldsymbol{\alpha}}(k), \mathbf{y}(k)) \\
&= \mathbf{X}_i(\boldsymbol{\alpha}(k) - \bar{\boldsymbol{\alpha}}(\mathbf{y}(k))) + \mathbf{X}_i \bar{\boldsymbol{\alpha}}(\mathbf{y}(k)) - \mathbf{C}_1^{\top}\mathbf{g} + \mathbf{H}^{\top}\mathbf{y}(k) \\
&= \mathbf{X}_i(\boldsymbol{\alpha}(k) - \bar{\boldsymbol{\alpha}}(\mathbf{y}(k)))
\end{aligned}$$

or that

$$\boldsymbol{\alpha}(k) - \bar{\boldsymbol{\alpha}}(\mathbf{y}(k)) = (\mathbf{C}_1^{\top}\mathbf{C}_1 + \rho \mathbf{H}^{\top}\mathbf{H})^{-1} \nabla_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k), \mathbf{y}(k)). \tag{C-6}$$

Taking the 2-norm and applying the triangular inequality concludes the proof. □

Let us now define the primal optimality gap as

$$\Delta_p(k) := \mathcal{L}_{\rho}(\boldsymbol{\alpha}(k+1), \mathbf{y}(k)) - \mathcal{L}_{\rho}(\bar{\boldsymbol{\alpha}}(k), \mathbf{y}(k)) \geq 0 \tag{C-7}$$

and the dual optimality gap as

$$\Delta_d(k) := d^{\star} - d(\mathbf{y}(k)) \geq 0, \tag{C-8}$$

where $d := \min_{\boldsymbol{\alpha}} \mathcal{L}_{\rho}(\boldsymbol{\alpha}, \mathbf{y})$ represents the dual function and d^{star} its maximum.

The primal optimality gap represents the error caused by the inexact minimization of the Lagrangian function in (7-8) and (7-9). The dual optimal gap describes how far the dual function is from its optimal value.

The following lemma can then be written in order to prove that the primal and dual optimality gaps go to zero as time progresses.

Lemma 4. *Given that Assumption 4 holds there is a* $\gamma > 0$ *such that the following bound holds*

$$\Big(\Delta_p(k) + \Delta_d(k)\Big) - \Big(\Delta_p(k-1) + \Delta_d(k-1)\Big) \leq \Big(\rho\tau^2\eta^2||\mathbf{H}_2^2|| - \gamma\Big)||\boldsymbol{\alpha}(k+1) - \boldsymbol{\alpha}(k)||_2^2. \tag{C-9}$$

For $\rho > 0$ *sufficiently small it is immediately proven that the sum of the primal and dual optimality gaps is a monotonically decreasing function.*

Proof. The proof is presented for a general cost function in Theorem 3.1 of [87] using the two aforementioned lemmas - Lemma 2 and 3. □

The convergence theorem can now be presented for the unstructured ADMM algorithm in (7-5).

Theorem 1. *Given the results in Lemma 4, and choosing a sufficiently small $\rho > 0$ then the sequence $\{\boldsymbol{\alpha}(k), \mathbf{y}\}_{k=0}^{\infty}$ generated by the unstructured ADMM algorithm converges to primal dual optimal pair $\boldsymbol{\alpha}^{\star}$ and $\mathbf{y}^{\star}$.*

Proof. The convergence is shown by proving that the sequence $\{\boldsymbol{\alpha}(k), \mathbf{y}\}_{k=0}^{\infty}$ converges to the KKT conditions. □

C-2 Structured ADMM Derivations

C-2-1 QR factorization

The QR decomposition with pivoting of matrix $\bar{\mathbf{X}}_i \in \mathbb{R}^{\hat{d}\times\hat{d}}$ yields the following expression:

$$\bar{\mathbf{X}}_i = \mathbf{QRE}^\top,$$

where $\mathbf{Q} \in \mathbb{R}^{m\times m}$ is a full square orthogonal matrix (*i.e.*, $\mathbf{Q}^\top = \mathbf{Q}^{-1}$) and matrix $\mathbf{R} \in \mathbb{R}^{m\times n}$ is upper triangular with the same rank as $\bar{\mathbf{X}}_i$. Moreover, $\mathbf{E} \in \mathbb{R}^{n\times n}$ is a permutation/pivoting matrix. We can then rewrite the linear system of equations as

$$\mathbf{QRE}^\top \bar{\boldsymbol{\alpha}}_i = \bar{\mathbf{d}}_i \tag{C-10}$$

$$\begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\alpha}}_i^{\text{det}} \\ \tilde{\boldsymbol{\alpha}}_i^{\text{free}} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_1^\top \\ \mathbf{Q}_2^\top \end{bmatrix} \bar{\mathbf{d}}_i, \tag{C-11}$$

where $\tilde{\boldsymbol{\alpha}}_i^{\text{det}}$ and $\tilde{\boldsymbol{\alpha}}_i^{\text{free}}$ are given by the following transformation

$$\begin{bmatrix} \tilde{\boldsymbol{\alpha}}_i^{\text{det}} \\ \tilde{\boldsymbol{\alpha}}_i^{\text{free}} \end{bmatrix} = \begin{bmatrix} \mathbf{E}_1 & \mathbf{E}_2 \end{bmatrix}^\top \bar{\boldsymbol{\alpha}}_i$$

and represent the pivoted coefficients that have a determined solution and the ones that are free to take any value.

If we denote the rank of matrix $\bar{\mathbf{X}}_i$ by r, $\mathbf{R}_{11} \in \mathbb{R}^{r\times r}$ also has rank r, and $\mathbf{R}_{12}$ is defined in $\mathbb{R}^{r\times(n-r)}$.

If the system is solvable, then $\mathbf{Q}_2^\top \bar{\mathbf{d}}_i = 0$ and we can solve the system for $\tilde{\boldsymbol{\alpha}}_i = [(\tilde{\boldsymbol{\alpha}}_i^{\text{det}})^\top \, (\tilde{\boldsymbol{\alpha}}_i^{\text{free}})^\top]^\top$, where the first part of the vector corresponds to the non-zero rows of $\mathbf{R}$, as follows:

$$\mathbf{R}_{11}\tilde{\boldsymbol{\alpha}}_i^{\text{det}} + \mathbf{R}_{12}\tilde{\boldsymbol{\alpha}}_i^{\text{free}} = \mathbf{Q}_1^\top \bar{\mathbf{d}}_i. \tag{C-12}$$

This represents an overdetermined system where the vector $\bar{\boldsymbol{\alpha}}_i^{\text{free}}$ is free and can be set immediately to zero. The Equation (C-12) then simplifies to

$$\tilde{\boldsymbol{\alpha}}_i^{\text{det}} = \mathbf{R}_{11}^{-1}\mathbf{Q}_1^\top \bar{\mathbf{d}}_i. \tag{C-13}$$

The last equation solves the system for the pivoted variable $\tilde{\boldsymbol{\alpha}}_i^{\mathrm{det}} = \mathbf{E}_1^\top \bar{\boldsymbol{\alpha}}_i$. We know that $\mathbf{E}_1\mathbf{E}_1^\top$ will yield a diagonal matrix with ones in the places corresponding to $\tilde{\boldsymbol{\alpha}}_i^{\mathrm{det}}$ and zeros elsewhere.

Multiplying (C-13) on the right by $\mathbf{E}_1$ yields then the following expression:

$$\bar{\boldsymbol{\alpha}}_i = \mathbf{E}_1 \mathbf{R}_{11}^{-1} \mathbf{Q}_1^\top \bar{\mathbf{d}}_i. \tag{C-14}$$

$$\tag{C-15}$$

Because we considered that we would set $\tilde{\boldsymbol{\alpha}}_i^{\mathrm{free}} = 0$, we can write the full $\bar{\boldsymbol{\alpha}}_i$ on the right-hand side.

Bibliography

[1] M. Verhaegen, "Lecture notes on control for High Resolution Imaging," May 2012.

[2] Spiricon, ed., *Hartmann Wavefront Analyzer Tutorial.* Spiricon, 2004.

[3] W. H. Southwell, "Wavefront estimation from wavefront slope measurements," *J. Opt. Soc. Am.*, vol. 70, pp. 998–1006, Aug. 1980.

[4] H. Song, *Model-based control in Adaptive Optics Systems.* PhD thesis, Delft University of Technology, Delft Center for Systems and Control, 2011.

[5] C. de Visser, *Global Nonlinear Model Identification with Multivariate Splines.* PhD thesis, Delft University of Technology, 2012.

[6] F. Roddier, *Adaptive Optics in Astronomy.* Cambridge University Press, 1999.

[7] C. A. Roddier and F. J. Roddier, "New optical testing methods developed at the University of Hawaii: results on ground-based telescopes and hubble space telescope," 1992.

[8] M. Cayrel, "E-ELT optomechanics: overview," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 8444 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Sep 2012.

[9] K.-P. Markus, "Overall science goals and top level AO requirements for the E-ELT," in *1st AO4ELT conference - Adaptive Optics for Extremely Large Telescopes* (T. F. Y. Clénet, J.-M. Conan and G. R. (Eds.), eds.), 2010.

[10] V. Bakshi, *EUV Lithography.* SPIE Press, 2009.

[11] P. Mercère, P. Zeitoun, M. Idir, S. L. Pape, D. Douillet, X. Levecq, G. Dovillaire, S. Bucourt, K. A. Goldberg, P. P. Naulleau, and S. Rekawa, "Hartmann wave-front measurement at 13.4 nm with λeuv/120 accuracy," *Opt. Lett.*, vol. 28, pp. 1534–1536, Sep. 2003.

[12] A. Polo, V. Kutchoukov, F. Bociort, S. Pereira, and H. Urbach, "Determination of wavefront structure for a Hartmann Wavefront Sensor using a phase-retrieval method," *Optics Express*, pp. 7822–7832, 2012.

[13] C. S. Smith, R. Marinică, A. J. den Dekker, M. Verhaegen, V. Korkiakoski, C. U. Keller, and N. Doelman, "Iterative linear focal-plane wavefront correction," *J. Opt. Soc. Am. A*, vol. 30, pp. 2002–2011, Oct 2013.

[14] M. J. Booth, D. Débarre, and A. Jesacher, "Adaptive optics for biomedical microscopy," *Opt. Photon. News*, vol. 23, pp. 22–29, Jan. 2012.

[15] T. I. M. van Werkhoven, J. Antonello, H. H. Truong, M. Verhaegen, H. C. Gerritsen, and C. U. Keller, "Snapshot coherence-gated direct wavefront sensing for multi-photon microscopy," *Opt. Express*, vol. 22, pp. 9715–9733, Apr 2014.

[16] A. M. Godara, P. Dubis, A. Roorda, J. L. Duncan, and J. Carroll, "Adaptive optics retinal imaging: Emerging clinical applications," *Optom. Vis. Sci.*, vol. 87, no. 12, pp. 930–941, 2010.

[17] L. D. Thibos and X. Hong, "Clinical applications of the Shack-Hartmann aberrometer," *Optom. Vis. Sci.*, vol. 76, no. 12, pp. 817–825, 1999.

[18] W. Cheney and D. Kincaid, *Numerical Mathematics and Computing.* International Thomson Publishing, 4th ed., 1998.

[19] L. Vandenberghe, "Lecture notes for applied numerical computing," 2011.

[20] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press, Mar. 2004.

[21] H. W. Babcock, "The possibility of compensating astronomical seeing," *Publications of the Astronomical Society of the Pacific*, vol. 65, pp. 229–236, Oct. 1953.

[22] V. P. Linnik, "Active and Adaptive Optics," *Opt. Spektrosk.*, vol. 3, no. 401, 1957.

[23] K. Hinnen, M. Verhaegen, and N. Doelman, "A data driven H2-optimal control approach for adaptive optics," *IEEE Trans. on Control Systems Technology*, vol. 16, no. 3, pp. 381–395, 2008.

[24] G. Rousset, J. C. Fontanella, P. Kern, and F. Rigaut, "First diffraction-limited astronomical images with adaptive optics," *Astronomy and Astrophysics*, vol. 230, pp. L29–L32, Apr. 1990.

[25] J. M. Beckers, "Adaptive optics for astronomy: Principles, performance, and application," *Annual review of astronomy and astrophysics*, vol. 31, pp. 13–62, 1993.

[26] J. Goodman, *Introduction to Fourier Optics, 3rd Ed.* Roberts and Company Publishers, 2005.

[27] B. Platt and R. Shack, "History and principles of shack-hartmann wavefront sensing," *J. of Refractive Surgery*, vol. 17, pp. 573–577, Sep./Oct. 2001.

[28] W. Boyle and G. Smith, “Charge coupled devices,” *Bell Syst. Tech. Journal*, vol. 49, pp. 587–593, 1970.

[29] S. Tisa, F. Zappa, and S. Cova, “Monolithic quad-cells for single-photon timing and tracking,” vol. 6583 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, May 2007.

[30] L. A. Carvalho, “A simple and effective algorithm for detection of arbitrary Hartmann–Shack patterns,” *Journal of Biomedical Informatics*, vol. 37, no. 1, pp. 1 – 9, 2004.

[31] J. Ares and J. Arines, “Influence of thresholding on centroid statistics: Full analytical description,” *Appl. Opt.*, vol. 43, pp. 5796–5805, Nov. 2004.

[32] K. L. Baker and M. M. Moallem, “Iteratively weighted centroiding for Shack-Hartmann wave-front sensors,” *Opt. Express*, vol. 15, pp. 5147–5159, Apr. 2007.

[33] C. Leroux and C. Dainty, “Estimation of centroid positions with a matched-filter algorithm: relevance for aberrometry of the eye,” *Opt. Express*, vol. 18, pp. 1197–1206, Jan. 2010.

[34] S. Thomas, “Optimized centroid computing in a Shack-Hartmann sensor,” *SPIE*, vol. 5490, pp. 1238 – 1246, 2004.

[35] D. L. Fried, “Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements,” *J. Opt. Soc. Am.*, vol. 67, pp. 370–375, Mar. 1977.

[36] R. H. Hudgin, “Wave-front reconstruction for compensated imaging,” *J. Opt. Soc. Am.*, vol. 67, pp. 375–378, Mar. 1977.

[37] C. R. Vogel, “Sparse matrix methods for wavefront reconstruction, revisited,” vol. 5490, pp. 1327–1335, 2004.

[38] E. Thiébaut and M. Tallon, “Fast minimum variance wavefront reconstruction for extremely large telescopes,” *J. Opt. Soc. Am. A*, vol. 27, pp. 1046–1059, May 2010.

[39] M. Rosensteiner, “Cumulative Reconstructor: fast wavefront reconstruction algorithm for extremely large telescopes,” *J. Opt. Soc. Am. A*, vol. 28, pp. 2132–2138, Oct. 2011.

[40] M. Rosensteiner, “Wavefront reconstruction for extremely large telescopes via CuRe with domain decomposition,” *J. Opt. Soc. Am. A*, vol. 29, pp. 2328–2336, Nov. 2012.

[41] K. R. Freischlad and C. L. Koliopoulos, “Modal estimation of a wave front from difference measurements using the discrete Fourier transform,” *J. Opt. Soc. Am. A*, vol. 3, pp. 1852–1861, Nov. 1986.

[42] L. A. Poyneer, D. T. Gavel, and J. M. Brase, “Fast wave-front reconstruction in large adaptive optics systems with use of the Fourier transform,” *J. Opt. Soc. Am. A*, vol. 19, pp. 2100–2111, Oct. 2002.

[43] J. Herrmann, “Least-squares wave front errors of minimum norm,” *J. Opt. Soc. Am.*, vol. 70, pp. 28–35, Jan. 1980.

[44] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach.* New York, NY, USA: Cambridge University Press, 1st ed., 2007.

[45] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[46] R. G. Lane, A. Glindemann, and J. C. Dainty, "Simulation of a Kolmogorov phase screen," *Waves in Random Media*, vol. 2, no. 3, pp. 209–224, 1992.

[47] R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," *J. Opt. Soc. Am.*, vol. 69, pp. 972–977, Jul. 1979.

[48] L. Gilles, C. R. Vogel, and B. L. Ellerbroek, "Multigrid preconditioned conjugate-gradient method for large-scale wave-front reconstruction," *J. Opt. Soc. Am. A*, vol. 19, pp. 1817–1822, Sep. 2002.

[49] Y. Saad, *Iterative Methods for Sparse Linear Systems.* Society for Industrial and Applied Mathematics, 2nd ed., 2003.

[50] C. R. Vogel and Q. Yang, "Multigrid algorithm for least-squares wavefront reconstruction," *Appl. Opt.*, vol. 45, pp. 705–715, Feb. 2006.

[51] P. J. Hampton, P. Agathoklis, and C. Bradley, "A New Wave-Front Reconstruction Method for Adaptive Optics Systems Using Wavelets," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, pp. 781–792, Nov. 2008.

[52] C. de Visser and M. Verhaegen, "Wavefront reconstruction in adaptive optics systems using nonlinear multivariate splines," *Journal of the Optical Society of America A*, vol. 30, no. 1, pp. 8295–8301, 2013.

[53] C. de Visser and M. Verhaegen, "A distributed simplex B-splines based wavefront reconstruction," 2012.

[54] P. Massioni, R. Fraanje, and M. Verhaegen, "Adaptive optics application of distributed control design for decomposable systems," in *Proceedings of the 48th IEEE Conference on Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009.*, pp. 7113–7118, 2009.

[55] P. Massioni, *Decomposition Methods for Distributed Control and Identification.* PhD thesis, Delft University of Technology, 2010.

[56] M. J. Lai and L. L. Schumaker, *Spline Functions on Triangulations.* Cambridge University Press, 2007.

[57] C. de Boor, "What is a multivariate spline? in J. McKenna and R. Temam, editors," *ICIAM '87: Proceedings of the First International Conference on Industrial and Applied Mathematics*, p. 90–101, 1987.

[58] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition," in *Proceedings of the IEEE*, 2007.

[59] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1989.

[60] D. P. Bertsekas and D. P. Bertsekas, *Nonlinear Programming.* Athena Scientific, 2nd ed., Sept. 1999.

[61] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, *Convex Analysis and Optimization.* Athena Scientific, 2003.

[62] S. Boyd, L. Xiao, A. Mutapcic, and J. Mattingley, "Notes on decomposition methods for ee364b, winter 2006-07," Feb. 2007.

[63] G. B. Dantzig and P. Wolfe, "Decomposition principle for linear programs," *Operations Research*, vol. 8, no. 1, pp. 101–111, 1960.

[64] H. Everett, "Generalized lagrange multiplier method for solving problems of optimum allocation of resources," *Operations Research*, vol. 11, no. 3, pp. 399–417, 1963.

[65] M. Hestenes, "Multiplier and gradient methods," *Journal of Optimization Theory and Applications*, vol. 4, no. 5, pp. 303–320, 1969.

[66] M. Powell, *A Method for Nonlinear Constraints in Minimization Problems,.* Academic Press, London, ed. r. fletcher ed., 1969.

[67] R. Glowinski and A. Marocco, "Sur l'approximation, par élements finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires," *Revue française d'automatique*, vol. 9, no. 2, pp. 41–76, 1972.

[68] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers and Mathematics with Applications*, vol. 2, no. 1, pp. 17 – 40, 1976.

[69] R. Glowinski and A. Marrocco, "On the solution of a class of non-linear dirichlet problems by a penalty-duality method and finite element of order one," in *Optimization Techniques* (G. I. Marchuk, ed.), Lecture Notes in Computer Science, pp. 327–333, Springer - Verlag, 1974.

[70] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.).* Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[71] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.

[72] E. Ghadimi, A. Teixeira, I. Shames, and M. Johansson, "On the optimal step-size selection for the alternating direction method of multipliers," in *:*, pp. 139–144, 2012. QC 20130521.

[73] S. L. Wang and L. Z. Liao, "Decomposition method with a variable parameter for a class of monotone variational inequality problems," *J. Optim. Theory Appl.*, vol. 109, pp. 415–429, May 2001.

[74] K. Hinnen, M. Verhaegen, and N. Doelman, "Exploiting the spatio-temporal correlation in adaptive optics using data driven H_2-optimal control," *Journal of the Optical Society of America A*, vol. 24, no. 5, pp. 1714–1725, 2007.

[75] D. Voelz, *Computational Fourier Optics.* SPIE Press, 2011.

[76] T. M. Apostol, *Mathematical Analysis.* Reading, MA: Addison-Wesley, second ed., 1974.

[77] D. Malacara and W. T. Welford, *Optical shop testing.* John Wiley Sons, Inc., 2006.

[78] A. Kolmogorov, “The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds’ Numbers,” *Akademiia Nauk SSSR Doklady*, vol. 30, pp. 301–305, 1941.

[79] E. Brunner, J. Silva, M. Verhaegen, and C. de Visser, “Compressive Sampling in Intensity Based Control for adaptive optics,” in *IFAC WC 2014*, 2014.

[80] E. Candès, “Compressive sampling,” in *European Mathematical Society*, Proceedings of the International Congress of Mathematicians, 2006.

[81] H. Ohlsson, V. Kutchoukov, F. Bociort, S. Pereira, and H. Urbach, “An extension of compressive sensing to the phase retrieval problem,” in *Advances in Neural Information Processing Systems*, 2012.

[82] J. Silva, E. Brunner, M. Verhaegen, A. Polo, and C. de Visser, “Wavefront reconstruction using intensity measurements for real-time adaptive optics,” in *ECC 2014*, 2014.

[83] C. de Visser, Q. Chu, and J. Mulder, “A new approach to linear regression with multivariate splines,” *Automatica*, vol. 45, no. 12, pp. 2903 – 2909, 2009.

[84] C. de Visser, J. Mulder, and Q. Chu, “Global nonlinear aerodynamic model identification with multivariate splines,” *AIAA Atmospheric Flight Mechanics Conference*, vol. 5726, 2009.

[85] G. Awanou and M.-J. Lai, “Trivariate spline approximations of 3D Navier-Stokes equations,” *Math. Comp.*, pp. 585–601, 2004.

[86] N. Govindarajan, C. de Visser, and K. Krishnakumar, “A sparse collocation method for solving time-dependent HJB equations using multivariate B-splines,” *Automatica*, To appear.

[87] M. Hong and Z.-Q. Luo, “On the Linear Convergence of the Alternating Direction Method of Multipliers,” *ArXiv e-prints*, Aug. 2012.

[88] D. Han and X. Yuan, “A note on the alternating direction method of multipliers,” *Journal of Optimization Theory and Applications*, vol. 155, no. 1, pp. 227–238, 2012.

[89] W. Deng and W. Yin, “On the global and linear convergence of the generalized alternating direction method of multipliers,” *Rice University CAAM Technical Report*, no. TR12-14, 2012.

[90] F. Bullo, J. Cortés, and S. Martínez, *Distributed Control of Robotic Networks.* Princeton Series in Applied Mathematics, Princeton University Press, 2009.

[91] G. Stathopoulos, T. Keviczky, and Y. Wang, “A hierarchical time-splitting approach for solving finite-time optimal control problems,” in *Control Conference (ECC), 2013 European*, pp. 3089–3094, July 2013.

[92] D. C. Lay, *Linear Algebra and its Applications.* Addison-Wesley, 1993.

[93] A. Olshevsky and J. N. Tsitsiklis, "Convergence speed in distributed consensus and averaging," in *IN PROC. OF THE 45TH IEEE CDC*, 2006.

[94] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Systems and Control Letters*, vol. 53, pp. 65–78, 2003.

[95] O. Manneberg, "Design and simulation of a high spatial resolution Hartmann-Shack wavefront sensor," Master's thesis, KTH Stockholm, 2005.

[96] J. D. Schmidt, *Numerical Simulation of Optical Wave Propagation With Examples in MATLAB (SPIE Press Monograph Vol. PM199).* SPIE Press, Aug. 2010.

[97] G. Cao and X. Yu, "Accuracy analysis of a hartmann-shack wavefront sensor operated with a faint object," *Optical Engineering*, vol. 33, no. 7, pp. 2331–2335, 1994.

Glossary

List of Acronyms

MVM	Matrix Vector Multiply
GPU	Graphics Processing Unit
FLOPs	Floating-point operations
AO	Adaptive Optics
SH	Shack-Hartmann
HWS	Hartmann Wavefront Sensor
WFR	Wavefront Reconstruction
DM	Deformable Mirror
E-ELT	European Extremely Large Telescope
EUVL	Extreme Ultra-violet Lithography
CCD	Charge-Coupled Device
FrIM	Fractal Iterative Method
CuRe	Cumulative Reconstructor
SABRE	Spline based Aberration Reconstruction
LS	Least-squares
FFT	Fast Fourier Transform
SPD	Semi Positive Definite
ADMM	Alternating Direction Method of Multipliers
KKT	Karush-Kuhn-Tucker conditions