Optimisation-based wavefront sensorless adaptive optics for microscopy

Jacopo Antonello

Typeset by the author using LaTeX and BibTeX.
Date: Wed Oct 29 13:26:43 2014 +0100.
Commit: 3a1162e687295bde6575dd2317fb21e7e8944d16.

OPTIMISATION-BASED WAVEFRONT SENSORLESS ADAPTIVE OPTICS FOR MICROSCOPY

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op

maandag 10 november 2014 om 10:00 uur

door

Jacopo ANTONELLO

Master of Science in Engineering of Computer Systems,
Politecnico di Milano, Italië
geboren te Camposampiero, Italië

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. ir. M. Verhaegen

Samenstelling promotiecommisie:

Rector Magnificus,	voorzitter
Prof. dr. ir. M. Verhaegen,	Technische Universiteit Delft, promotor
Prof. dr. M. J. Booth,	University of Oxford
Prof. dr. N. J. Doelman,	Universiteit Leiden
Prof. dr. C. U. Keller,	Universiteit Leiden
Dr. S. F. Pereira,	Technische Universiteit Delft
Prof. dr. ir. B. De Schutter,	Technische Universiteit Delft
Prof. dr. G. Vdovin,	Technische Universiteit Delft
Prof. dr. ir. J. Hellendoorn,	Technische Universiteit Delft, reservelid

Enabling new technology

This research is supported under project number 10433 by the Dutch Technology Foundation (STW), which is part of the Netherlands Organization for Scientific Research (NWO).

Author's e-mail: jack@antonello.org
Author's website: antonello.org

ISBN: 978-94-6203-673-4

Printed by CPI – Koninklijke Wöhrmann B.V., Zutphen, The Netherlands.

Acknowledgements

I thank my promotor, Michel, for the opportunity to obtain a PhD. I am grateful for his support and for the freedom in my research. I am very grateful to Rufus and 宋宏 (Song Hong) for helping me a lot at the beginning, especially to get started with the experimental work. A very special thanks to my friend and colleague Tim, whose help and contribution were really invaluable to finish my PhD. I thank Christoph and Hans Gerritsen for their support. Ik wil graag STW en Nederland bedanken om buitenlandse onderzoekers en talent te uitnodigen en ondersteunen. I also thank Prof. M. Lovera for suggesting me the opportunity to work in Delft.

I would like to thank Ivo Houtzager, Stefan, Gabriel, Emilie, Paweł, Gijs, Paolo, Justin, Marco, Ilya, Federico and Andrea. Ik wil mijn nederlandse collega's Ivo Grondman, Pieter, Edwin en Mernout bedanken voor onze gesprekken in het Nederlands, ook al was ik niet altijd even spraakzaam. I thank 유한응 (HanWoong Yoo), for being the only other person around in the laboratory during many long weekends and evenings. I thank Aleksandar, Coen, Ruxandra, Raluca, Jeroen, Hans Verstraete, Elisabeth, Patricio, Yu Hu, Yashar, Mohammad, Noortje, Alessandro Scotti, Arturo, Alessandro Abate, Mathieu, Amol, Visa and Tope. I am very thankful for the great time that I had at Lindobeach and to everybody that I met there.

I thank Kees and Will for their support in the laboratory. I would like to thank Olaf, Kitty, Esther, Marieke, Heleen and Saskia for their help. I am very grateful for the friendly atmosphere at the Molecular Biophysics Group in Utrecht. I thank them for being sensible and for granting me access to their laboratory and the pulsed laser. I must say that such flexibility is unheard of at 3mE, where a lot of effort is spent on revoking access and locking doors.

Finally, I am thankful for the most important support of all, which came from my family and Αγγελική, and was especially invaluable in these last two very difficult years. I thank my parents Maria and Matteo, my brother Niccolò and my sister Elettra. I thank Αγγελική for her encouragement and understanding.

's-Gravenhage, oktober 2014
Jacopo Antonello

Contents

Chapter 1

Introduction

This chapter provides a brief introduction to microscopy and adaptive optics, with references to the relevant literature. The concept of resolution in microscopy is introduced by considering the simple case of a self-luminous point object. We then consider scanning microscopy, with references to confocal and two-photon microscopy. We discuss the detrimental effects that aberrations have on image-forming systems by mentioning the geometrical wavefront, the phase aberration function, and Zernike polynomials. Aberrations can be minimised using adaptive optics, which we introduce in its original application for astronomy. The problem of wavefront sensing is considered by examining the classical Shack–Hartmann wavefront sensor. Finally, we discuss adaptive optics in microscopy by considering specimen-induced aberrations, direct wavefront sensing, and wavefront sensorless adaptive optics.

1.1 Microscopy

1.1.1 Introduction

A microscope, from the ancient Greek μικρός "small" and σκοπεῖν "see", is an instrument used to examine objects that are too small for the naked eye [1]. Some evidence [2] suggests that the first microscope was invented in the Netherlands in 1595 by Jansen Sacharias, a Dutch spectacle-maker originally from Den Haag. Sacharias's microscope allowed to view an object magnified up to nine times with respect to its original size [2, 3]. The name "microscope" was coined later in 1625 by Giovanni Faber [1], who used it to denote the microscope developed by Galileo Galilei. Robert Hooke popularised microscopy as an investigation tool by publishing in 1665 his book Micrographia, which contained drawings of his observations made with a microscope probably built by Christopher Cock. A decade later, thanks to the discoveries of Antoni van Leeuwenhoek, microscopy had already become an established tool for scientific research, and has been developed further ever since.

Nowadays, fluorescence microscopy [4] is widely used in life sciences. This technique allows to selectively analyse a structure of interest inside a specimen, by labelling it with

a fluorescent dye, or by relying on the intrinsic fluorescence of the structure. The functional properties of the structure can be studied by recording 3D views with a scanning microscope. Laser light is used to excite the fluorophores of the dye, which subsequently re-emit light with a longer wavelength. The fluorescence emission is conveniently separated from the excitation light and collected onto a detector to generate high-contrast images.

1.1.2 The resolving power

A fundamental property in microscopy is the resolving power [5], which we consider in this section. In Fig 1.1, the reference frame centred in O denotes the object space, where the specimen is positioned. The reference frame centred in O' denotes the image space, where the image the of the specimen is formed. A point source is located in O and emits a monochromatic radiation with wavelength λ. The microscope maps the field distribution in O to a corresponding field distribution in O'. Due to the diffraction of light [5, 6] at the circular pupil of the microscope objective, the intensity of the field in O' is proportional to [7, 8, 9]

$$I_w(v) = \left| \frac{2J_1(v)}{v} \right|^2, \tag{1.1}$$

where $J_1(\cdot)$ is the first-order Bessel function of the first kind [10], and v is the normalised lateral coordinate, which is defined as [7]

$$v = \frac{2\pi}{\lambda} r n \sin(\alpha). \tag{1.2}$$

The normalised lateral coordinate v is related to r, the radial coordinate in the $x'y'$ plane, i.e., $r = (x'^2 + y'^2)^{1/2}$. n is the index of refraction [5] in the object space, and α is the acute angle that the marginal ray [5] makes with the optical axis OO'. The product $n \sin(\alpha)$ is the object-side numerical aperture, which is denoted as NA.

Throughout this introduction, we always assume that the paraxial approximation [7, 5] is valid, and only consider the Fraunhofer approximation [5, 11] for the diffraction of light. More refined models, which are suitable for modern microscope objectives, have been developed in the literature, see for example [12, 13] and the references therein.

For a given microscope objective, the manufacturer provides the equivalent paraxial quantities for the magnification M and the numerical aperture NA. These quantities are related to the focal length f and to the radius of the pupil a of the microscope objective by [12]

$$f = F/M, \qquad a = \frac{F \cdot \mathrm{NA}}{M}, \tag{1.3}$$

where F is the focal length of the tube lens of the microscope, which is a constant fixed to 200 mm for Leika and Nikon [12].

A plot of $I_w(v)$ is found in Fig. 1.2, where it can be seen that the intensity of the field due to a point source in O has a maximum value in O' but is also non-zero in the rest of the $x'y'$ plane. The Rayleigh criterion [5] is employed as a reference to quantify the spread of $I_w(v)$ within the $x'y'$ plane. This criterion considers the first minimum of $I_w(v)$, which

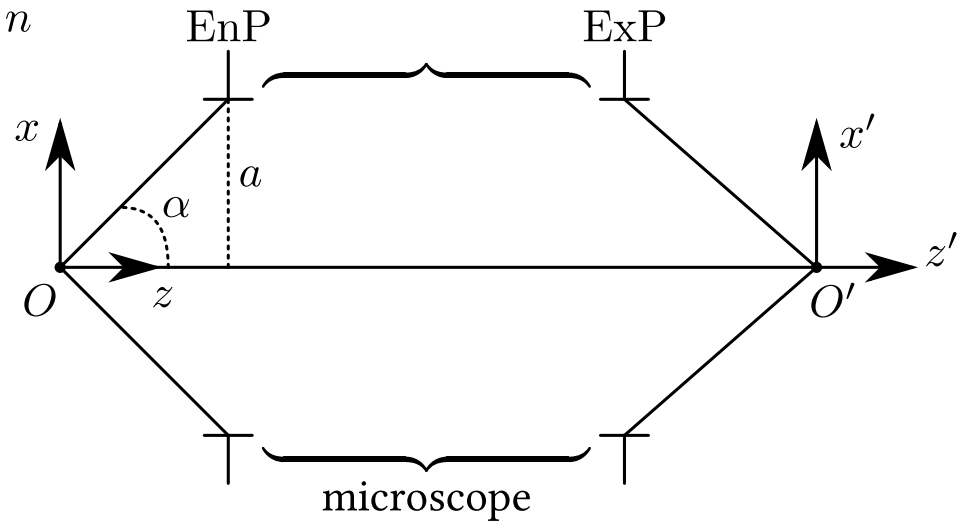

Figure 1.1: Simplified illustration of a microscope. The reference frame centred in O is the object space, where the index of refraction [5] is n. The reference frame centred in O' is the image space, where the image is formed. The marginal rays [5] depart from O, touch the edges of the entrance pupil (EnP), emerge from the edges of the exit pupil (ExP), and meet again in O'. The circular entrance pupil has radius a. The angle between the optical axis OO' and the marginal ray is α.

occurs for $v \approx 1.220\pi$. Using this criterion, the lateral resolution is defined as

$$r_l \approx 1.220\lambda/(2 \cdot \mathrm{NA}). \tag{1.4}$$

If the distance between two points in the object space is less than r_l, then the two points are said to be unresolved. The lateral resolving power of the microscope is defined as [5] $1/r_l$.

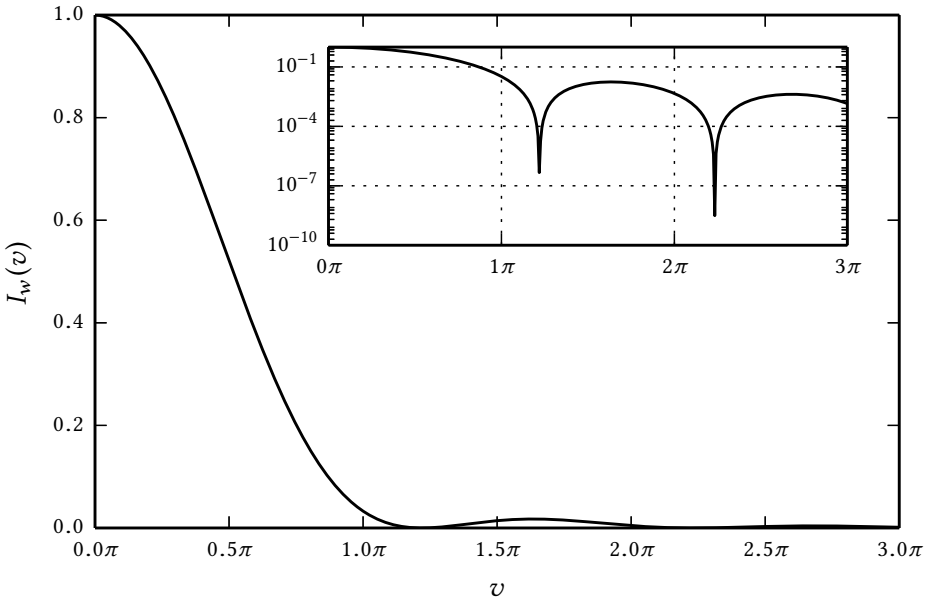

Figure 1.2: Profile of the intensity $I_w(v)$ of the field due to a point source in O. The profile is radially symmetric in the $x'y'$ plane, and is reported here as a function of the normalised coordinate v. The main lobe centred in O, called the Airy disk, is surrounded by attenuated concentric rings. The full Airy disk, obtained by rotation along the z' axis is reported later in Fig. 1.6. Note that the local minima in the inset plot are actually zeros.

To examine the Rayleigh criterion in more detail, consider two incoherent point sources located at O and P in the xy plane. The intensity of the total field due to the two sources in the $x'y'$ plane is shown in Fig. 1.3 for four different locations of P along the radial coordinate r. The separate contributions due to O and P can be distinguished more clearly as the distance $d(O,P)$ is larger than r_l. On the contrary, for $d(O,P) < r_l$, it appears that a single point source is present in the xy plane.

Similarly, one can also consider the intensity of the field along the z' axis, and define the axial resolution as [4]

$$r_a \approx 2\lambda n/\mathrm{NA}^2, \tag{1.5}$$

where n is the index of refraction [5] in the object space. The axial resolving power is then defined as $1/r_a$.

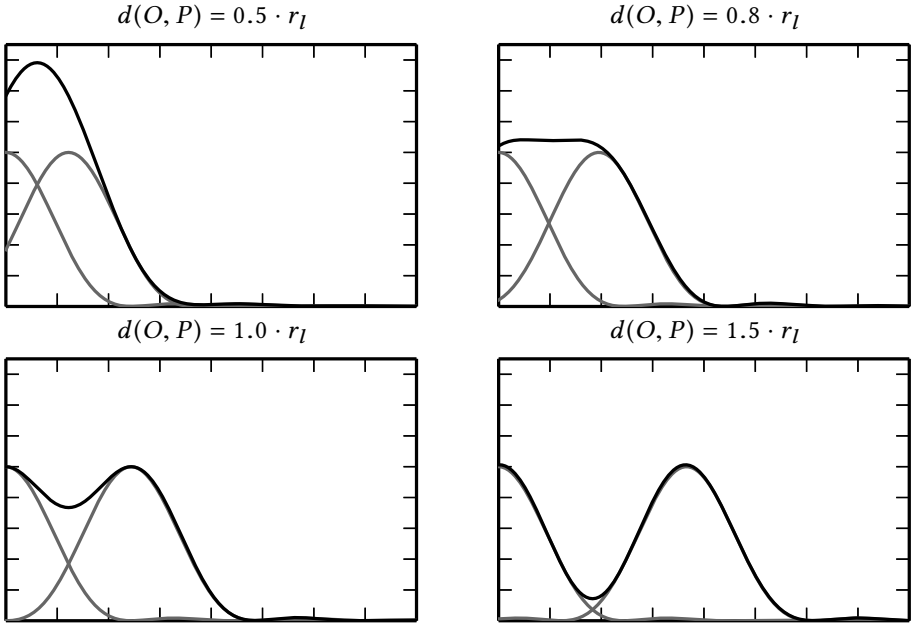

Figure 1.3: Illustration about the Rayleigh criterion. The figure shows the intensity of the field due to two incoherent point sources in O and P within the xy plane in Fig. 1.1. It is easier to tell the two sources apart as $d(O,P)$ becomes larger than r_l.

The conclusion from this section is that the resolving power in conventional microscopy is limited by the wavelength (λ) and by the numerical aperture (NA) according to Eq. (1.4) and Eq. (1.5). To improve the resolving power, i.e., to achieve smaller values for r_l and r_a, one possibility is to consider a shorter λ. Nevertheless, λ is subject to constraints determined, for example, by the type of fluorophores used and by the availability of suitable microscope objectives. A second possibility is to choose a microscope objective with a higher NA. When an oil-immersion objective [4] is used, the numerical aperture can be as high as 1.4 [14]. In the next subsections we introduce two different imaging techniques that allow to improve the resolving power with respect to the conventional microscope described in this section. It should also be noted that measurement noise, which was not considered in this section, is also a limiting factor to the resolving power [15].

1.1.3 Scanning microscopy

Scanning microscopy is a sequential image acquisition technique, whereby the image is built point by point executing a raster-type scan [7, 4]. This technique provides higher resolving powers with respect to conventional microscopy [7], and 3D views of biological specimens. In the following subsections two types of scanning microscopes are briefly introduced.

1.1.4 Confocal microscopy

Confocal microscopy (CM) was invented and patented by Minsky in 1957 [16]. In combination with fluorescent labels, CM provides 3D views of biological specimens [17, 18].

An illustration of a confocal microscope is found in Fig. 1.4. In a modern confocal microscope, laser light (laser) is focused by the objective towards a point O within the specimen (S), creating a double cone illumination profile along the optical axis. The fluorescence emission from the double cone is collected by the objective and separated from the excitation light using a dichroic beam splitter (DB). A pinhole (P), which is located in the image space before the detector, lets through the fluorescence emitted from O, but blocks the fluorescence emitted from the out-of-focus planes $\mathcal{I}$ and $\mathcal{J}$. An image can be composed by moving the specimen (S) within the xy plane.

We now consider the lateral resolution of a confocal microscope when a point object is located in O, which is equivalent to assume that the concentration of fluorophores is nonzero only in O. Neglecting for simplicity the finite size of the pinhole [19, 20] P and the difference in wavelength between the excitation light (λ_{ill}) and the fluorescence emission (λ_{fl}), it can be shown that [7, 21] the intensity of the field in O' is proportional to

$$I_c(v) = \left| \frac{2J_1(v)}{v} \right|^4 . \tag{1.6}$$

In Fig. 1.5, it can be seen that $I_c(v)$ has a sharper main lobe and reduced side lobes with respect to $I_w(v)$, which results in an improved lateral resolution [7, 21]. Similarly it can be shown that the axial resolution [22, 7, 21] is also improved. A comparison between $I_w(v)$ and $I_c(v)$ is also found in Fig. 1.6, where the case of an extended object is considered.

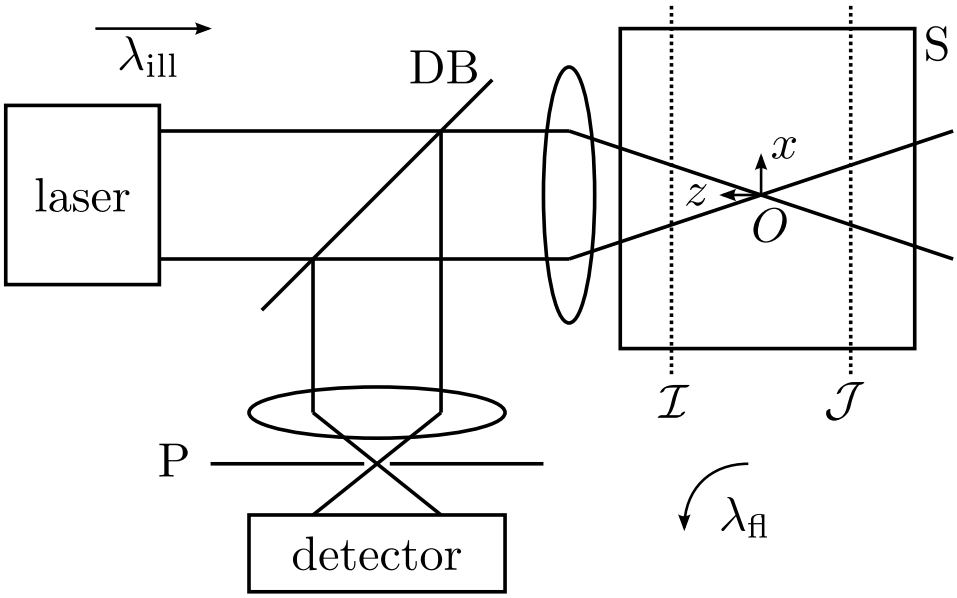

Figure 1.4: Illustration of a confocal microscope. Laser light propagates from left to right (λ_{ill}), passes through a dichroic beam splitter (DB) and is focused inside the specimen, which is depicted as a box (S). A double-cone illumination profile is created inside the specimen. The fluorescence emission from each point in the double cone propagates right to left (λ_{fl}), is collected by the microscope objective, and is reflected by DB towards a detector covered by a pinhole (P). Because the pinhole aperture and O are confocal, only the fluorescence emission from O reaches the detector, whereas the remaining emission is blocked.

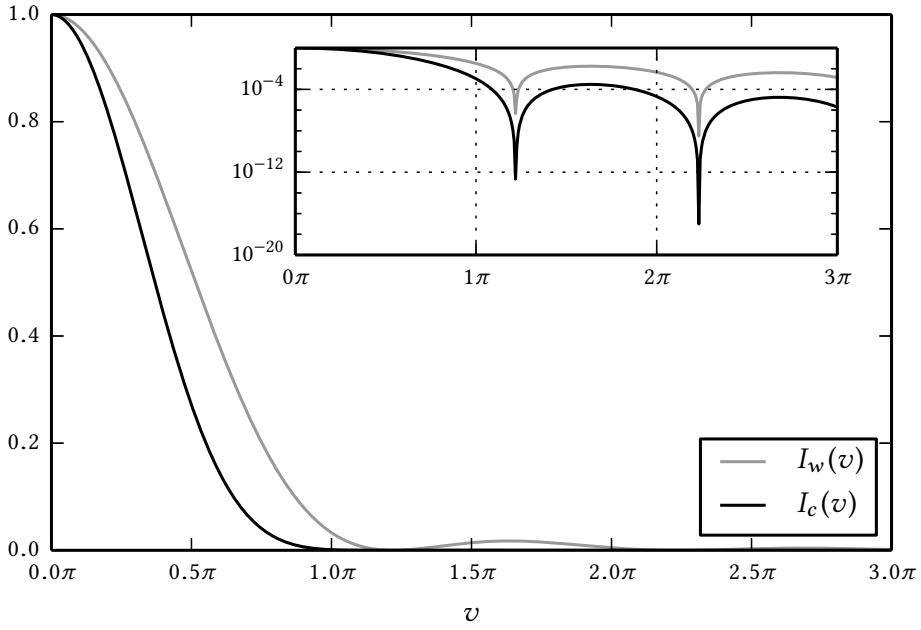

Figure 1.5: Comparison of the intensity profiles of a conventional ($I_w(\upsilon)$) and a confocal microscope ($I_c(\upsilon)$). $I_c(\upsilon)$ exhibits a thinner main lobe and reduced outer rings. As a consequence, a higher resolving power is expected when using confocal microscopy.

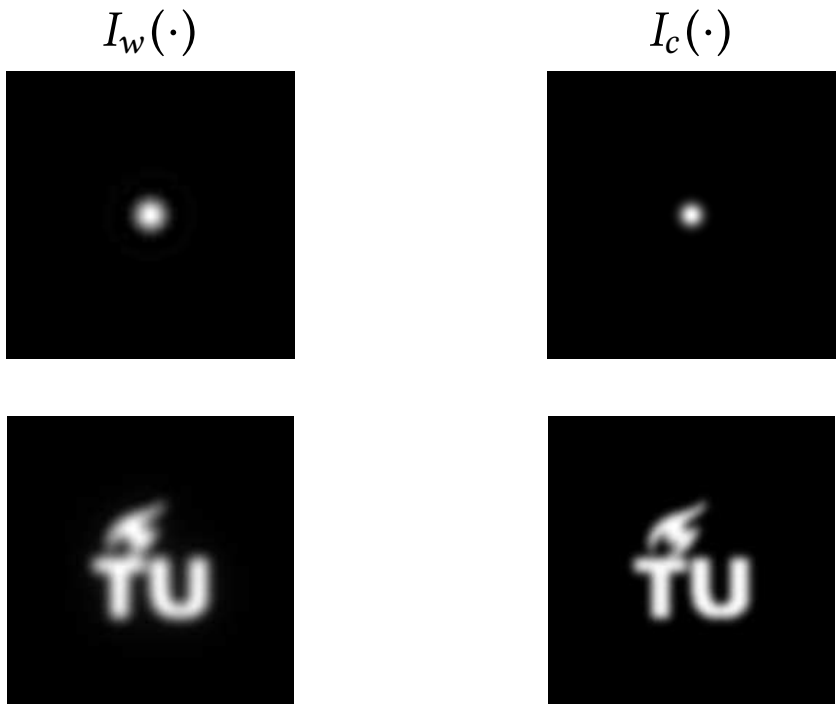

Figure 1.6: The top row shows the intensity of the field due to a point source in O for a conventional ($I_w(\cdot)$) and for a confocal ($I_c(\cdot)$) microscope. The disk in the top left image is called Airy disk. In the bottom left-hand corner, the intensity of the field in the image space is shown when a self-luminous extended object is examined with a conventional microscope. In the bottom right-hand corner, the same object is assumed to be labelled with a fluorescent dye and imaged using a confocal microscope.

1.1.5 Two-photon excitation microscopy

Two-photon excitation microscopy (2PEM) was developed by Denk *et al.* in 1990 [23], and uses the nonlinear light-matter interaction phenomenon of two-photon absorption, predicted by Göppert-Mayer [24] in 1931, to induce the fluorescence emission [25]. Contrary to CM, where the emission originates from the whole double cone illumination profile as depicted in Fig.1.4, in 2PEM most of the emission originates from a small focal volume centred in O, where the intensity of the field is sufficiently high [25]. In 2PEM one uses near-infrared excitation light, which can penetrate deeper inside the specimen [25, 26]. An improved resolving power is also obtained without including the confocal pinhole in front of the detector [23, 26]. Nevertheless, an expensive pulsed laser source must be used to achieve the instantaneous peak intensity necessary for the two-photon absorption [25], and a broader focal spot is generated due to the longer illumination wavelength.

It can be shown [27] that the intensity of the field in O', due to a point object in O, is proportional to

$$I_2(v) = \left| \frac{2J_1(v/2)}{v/2} \right|^4, \tag{1.7}$$

where the $1/2$ factor accounts for using an illumination wavelength of 2λ to generate the fluorescence emission with wavelength λ. A plot of $I_2(v)$ is reported in Fig. 1.7. From this simple analysis, it appears that the resolution achievable with 2PEM is worse with respect to conventional and confocal microscopy. In spite of this, CM and 2PEM achieve comparable resolutions in practice [28, 29]. This can be concluded by a more refined analysis, where the effect of the finite pinhole [30, 31] and the Stokes shift [28] are considered for CM. In addition, the excitation wavelength used for 2PEM is often shorter [28] than 2λ, where λ is the wavelength that one would use for CM. The resolving power can further be improved in 2PEM by including a confocal pinhole [32, 33].

1.2 Aberrations

1.2.1 Introduction

A microscope that achieves one of the resolving powers outlined in the previous sections is said to be diffraction-limited, i.e., its resolving power is only constrained by the phenomenon of the diffraction of light [5, 6]. Unfortunately, this ideal case is never attained in practice, and the actual resolving power is limited, instead, by the presence of aberrations [8, 11, 34]. Aberrations can be caused by imperfections in the optical components, such as manufacturing defects in the profiles of lenses, or by incorrect alignment within the optical system, for example when the axis of a lens does not coincide with the optical axis of the rest of the system. More importantly, aberrations are caused when light passes through a medium in which the index of refraction n is not constant, but varies as a function of time or space. Two notable examples of such media are the turbulent atmosphere of the earth [35] and biological specimens [36, 37, 38].

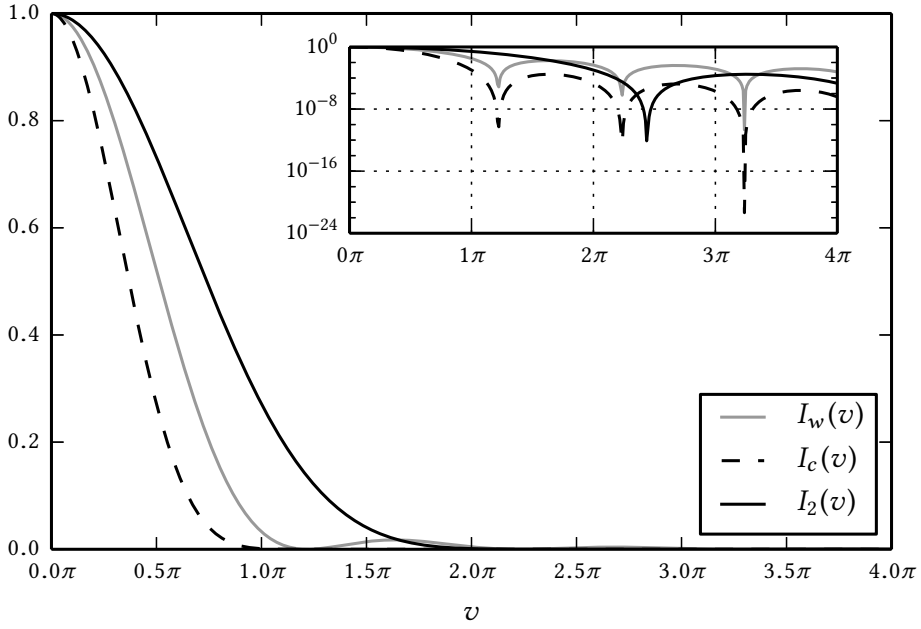

Figure 1.7: Comparison of the intensity profiles of a conventional ($I_w(\upsilon)$), confocal ($I_c(\upsilon)$), and two-photon microscopy ($I_2(\upsilon)$) where the same emission wavelength is assumed.

1.2.2 The geometrical wavefront

We use some intuitive geometrical optics arguments to try to convey how spatial variations of n within an optical system lead to aberrations. With reference to Fig. 1.8, consider a point source in O, and assume that the propagation of light from O up to the exit pupil (ExP) is described using geometrical optics [8, 5]. Also assume that the index of refraction is equal to one in both the object and the image space. The ray that departs from O, follows the optical axis and finishes at the centre of the exit pupil P_0, has travelled a certain optical path length, i.e,

$$[OP_0] = \int_O^{P_0} n\, ds, \tag{1.8}$$

where the line integral is taken along the curvilinear coordinate s, which follows the path of the ray through the optical system. Similarly, we can take all the other rays that depart from O, pass through the optical system and travel the same amount of optical path length $[OP_0]$. The surface that passes through the end-points of all these rays is called a wavefront. For a diffraction-limited optical system, the wavefront coincides with the Gaussian reference sphere [5, 8] V_g, which is a spherical surface with centre of curvature in O'. We can repeat the same procedure when the optical system is affected by aberrations, and define a different surface V. If we consider the ray depicted as a dashed line in Fig. 1.8, the wave aberration is given by the optical path length difference $[\bar{P}_1 P_1] = [OP_1] - [O\bar{P}_1]$. In the example in Fig. 1.8, the path length difference is caused by a patch depicted in grey

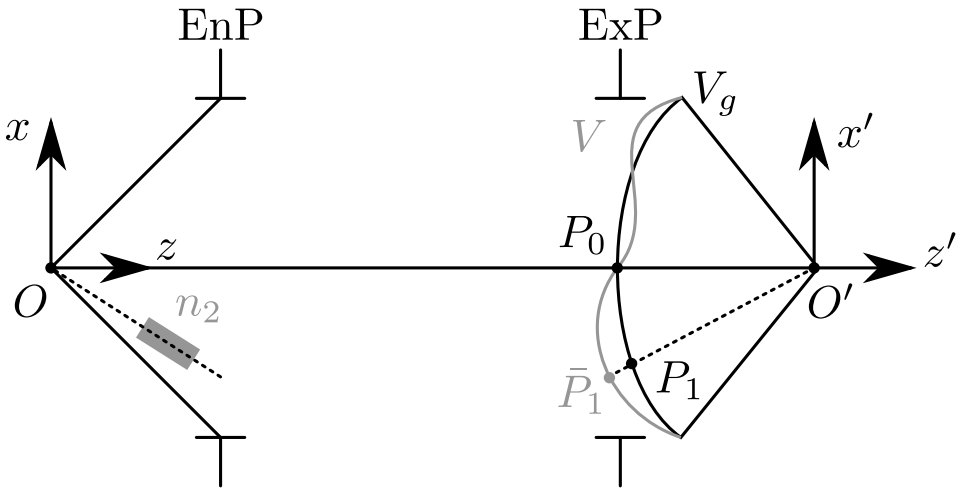

Figure 1.8: Geometrical optics description of the wavefront. Light rays depart from O, pierce the entrance pupil (EnP) and travel through the optical system exiting at the exit pupil (ExP) and converging to O'. One light ray depicted as a dashed line travels through a region where the index of refraction is $n_2 \neq 1$. The corresponding wave aberration is given by $[\bar{P}_1 P_1]$.

in the object space, where the index of refraction is $n_2 \neq 1$.

1.2.3 The phase aberration function

We now consider the scalar diffraction theory [8, 11] and conveniently assume that the Fraunhofer approximation is valid [5, 11]. The effects of the aberrations can be modelled using the generalised pupil function (GPF) [11], i.e.,

$$P(\rho,\theta) = A(\rho,\theta)\exp(i\Phi(\rho,\theta)). \tag{1.9}$$

The GPF is a complex-valued function defined over the normalised pupil of the optical system, which we assume to be circular. The real-valued function $A(\rho,\theta)$ accounts for the amplitude aberrations, e.g., due to amplitude apodisation [11]. Instead, the real-valued function $\Phi(\rho_0,\theta_0)$ accounts for the phase aberrations. Considering a point (ρ_0,θ_0) in the pupil disk, we have that that the phase aberration $\Phi(\rho_0,\theta_0)$ is equal to $(2\pi/\lambda)\cdot\text{OPD}$, where OPD is the optical path difference discussed in the previous subsection.

Assuming that the exit pupil of the optical system is the unit disk and that a point source is located in O, the intensity of the field in the $x'y'$ plane is proportional to [5, 11]

$$I(v,\phi) = \frac{1}{\pi}\left|\int_0^1\int_0^{2\pi} P(\rho,\theta)\exp(iv\rho\cos(\theta-\phi))\,\rho\,d\rho\,d\theta\right|^2. \tag{1.10}$$

Note that for $A(\rho,\theta) = 1$ and $\Phi(\rho,\theta) = 0$, one can compute the integral analytically [7] and recover $I(v,\phi) = I_w(v)$ from Eq. (1.1).

1.2.4 Zernike polynomials

It is useful to analyse the phase aberrations by decomposing $\Phi(\rho,\theta)$ into a series of Zernike polynomials [39, 5, 40], which are a complete set of orthogonal polynomials defined over the unit disk. Orthogonal polynomials have also been derived for other pupil geometries in [41, 42, 43]. For a circular pupil, we have

$$\Phi(\rho,\theta) = \sum_{n,m}\alpha_n^m\mathcal{Z}_n^m(\rho,\theta), \tag{1.11}$$

where indices $n\in\mathbb{N}_0$ and $m\in\mathbb{Z}$ denote respectively the radial order and the azimuthal frequency of the Zernike polynomial $\mathcal{Z}_n^m$, and are such that $n-|m|\geq 0$ and even [1] The coefficients of the Zernike polynomials are denoted by $\alpha_n^m\in\mathbb{R}$. Each Zernike polynomial $\mathcal{Z}_n^m$ is given by the product of a radial polynomial $R_n^{|m|}(\rho)$ and a trigonometric function $\Theta_n^m(\theta)$,

$$\mathcal{Z}_n^m(\rho,\theta) = c_n^m R_n^{|m|}(\rho)\Theta_n^m(\theta). \tag{1.12}$$

The coefficients c_n^m and the functions $R_n^m(\rho)$ and $\Theta_n^m(\theta)$ are defined as follows,

$$c_n^m = \begin{cases}\sqrt{n+1} & m=0\\ \sqrt{2(n+1)} & m\neq 0\end{cases}, \qquad \Theta_n^m(\theta) = \begin{cases}\cos(m\theta) & m\geq 0\\ -\sin(m\theta) & m<0\end{cases}, \tag{1.13}$$

[1] We use the symbol n to denote both the radial order of a Zernike polynomial and the index of refraction, as is commonly done in the literature [5]. No confusion should arise since the distinction is clear from the context.

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\,(\frac{n+m}{2}-s)!\,(\frac{n-m}{2}-s)!}\rho^{n-2s}. \tag{1.14}$$

Here, we have ordered and normalised the Zernike polynomials according to Noll [44]. A table of the first 37 Zernike polynomials and a list of properties are available in [40].

Zernike polynomials represent classical aberrations [5] that are combined to minimise the aberration variance over the pupil [45, 42], and are widely employed in optical design and adaptive optics. The first three Zernike polynomials, the piston ($\mathcal{Z}_0^0$), x-tilt ($\mathcal{Z}_1^1$), and y-tilt ($\mathcal{Z}_1^{-1}$), are reported in Fig. 1.9. The column on the left shows a plot of $\Phi = 0.8\mathcal{Z}_n^m$, the central column shows the intensity of the field in the image plane due to a point source, and the column on the right reports the intensity of the field in the image plane due to a self-luminous object, assuming an incoherent shift-invariant imaging system [11]. As can be seen, these three aberrations affect neither the image quality nor the resolving power. In fact, the piston aberration is not detectable from the intensity in the image plane, and the x-tilt and y-tilt correspond to shifts in the image plane. Therefore, the images are diffraction-limited and the Rayleigh criterion is applicable. Finally, the image generated by a point source can be computed analytically, and is equivalent to $I_w(v)$.

In Fig. 1.10, the defocus ($\mathcal{Z}_2^0$), and the primary astigmatisms ($\mathcal{Z}_2^{-2}$ and $\mathcal{Z}_2^2$) are shown. For the defocus aberration, the same arguments outlined in the first paragraphs could be applied if the image plane were moved to the defocused position along the z' axis [2]. Nevertheless, in the current position of the image plane ($z' = 0$), the image is not diffraction-limited, as the image of the point source shows a larger main lobe than the one expected from the Rayleigh criterion. The astigmatism aberrations lead to a decreased resolving power. Note that in this case it is difficult to assess the resolving power, since the intensity profile of the field due to a point source is not radially symmetric. This analysis holds also for the primary coma aberrations ($\mathcal{Z}_3^{-1}$ and $\mathcal{Z}_3^1$), and a trefoil aberration ($\mathcal{Z}_3^{-3}$) reported in Fig. 1.10.

It is useful to associate a scalar indicator to a given phase aberration function $\Phi(\rho, \theta)$, so that one can compare the severity of two different phase aberration functions by comparing the two indicators. Using the following functionals [40, 46],

$$\mathrm{E}_k[\Phi] = \frac{1}{\pi}\int_0^1 \int_0^{2\pi} \Phi(\rho,\theta)^k\, \rho\, d\rho\, d\theta, \tag{1.15}$$

for $k = 1$ and $k = 2$, one can define [40] the variance [3] and the rms of $\Phi(\rho, \theta)$ as

$$\mathrm{var}(\Phi) = \mathrm{E}_2[\Phi] - (\mathrm{E}_1[\Phi])^2, \qquad \mathrm{rms}(\Phi) = (\mathrm{E}_2[\Phi])^{1/2}. \tag{1.16}$$

Exploiting the orthogonality properties of the Zernike polynomials and the normalisation factors c_n^m in Eq. (1.13), one has that [40] $\mathrm{E}_1[\Phi] = \alpha_0^0$ and $\mathrm{E}_2[\Phi] = \sum_{n,m}(\alpha_n^m)^2$. Therefore

[2] It should be remarked that if the hypothesis of shift-invariance is not valid or other effects such as vignetting are not negligible [11], then $\mathcal{Z}_1^1$, $\mathcal{Z}_1^{-1}$ and $\mathcal{Z}_2^0$ may indeed affect the resolving power.

[3] Note that even though the name "variance" is commonly used, the function $\Phi(\rho, \theta)$ is deterministic, and the functionals $\mathrm{E}_k[\Phi]$ compute the definite integrals of $\Phi(\rho, \theta)^k$ in the unit disk. No probability density function is considered here.

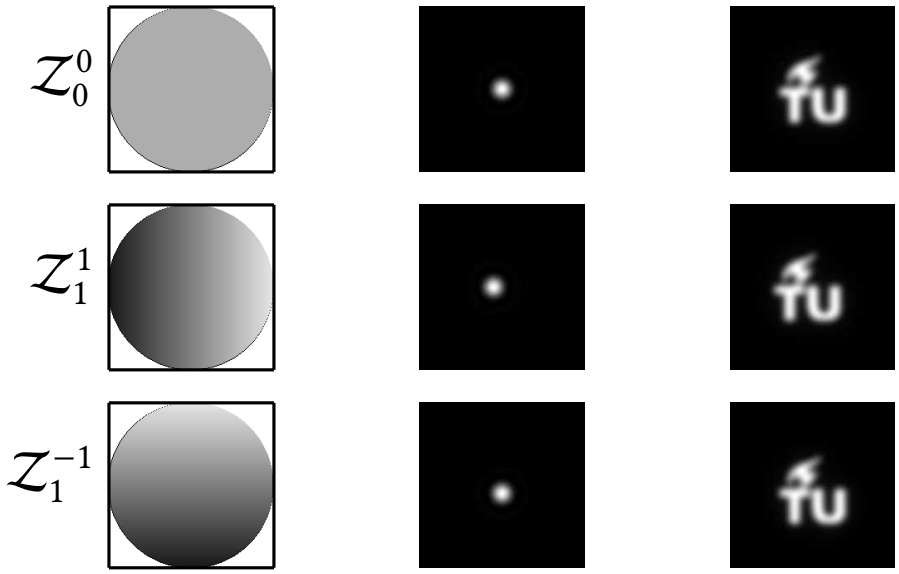

Figure 1.9: Examples of Zernike aberrations. In each row, we have $\Phi(\rho,\theta) = 0.8\mathcal{Z}_n^m(\rho,\theta)$. A plot of $\Phi(\rho,\theta)$ is reported in the left column. The intensity of the field in the image plane is reported in the central and right columns, respectively when a point object and a self-luminous extended object are present in the object space. The piston ($\mathcal{Z}_0^0$), x-tilt ($\mathcal{Z}_1^1$), and y-tilt ($\mathcal{Z}_1^{-1}$) polynomials are shown in each row.

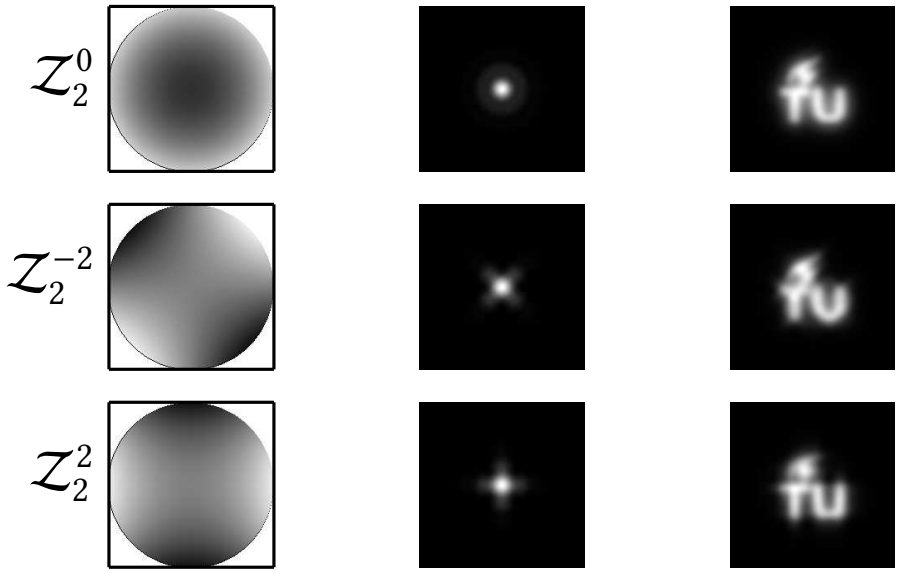

Figure 1.10: See the caption of Fig. 1.9. The defocus ($\mathcal{Z}_2^0$), and the primary astigmatisms ($\mathcal{Z}_2^{-2}$ and $\mathcal{Z}_2^2$) polynomials are shown in each row.

one can evaluate the functionals above using the following simple formulas,

$$\mathrm{var}(\Phi) = \sum_{n\neq 0, m\neq 0} (\alpha_n^m)^2, \qquad \mathrm{rms}(\Phi) = \left(\sum_{n,m} (\alpha_n^m)^2\right)^{1/2}. \tag{1.17}$$

These results are motivated by the fact that, except for the piston $\mathcal{Z}_0^0$, Zernike polynomials have unit variance and zero mean value ($\mathrm{E}_1[\cdot]$) over the unit disk. In fact, the piston mode is commonly neglected in adaptive optics literature, since it does not affect the image as seen in the first row in Fig. 1.9. When a finite set of N_α Zernike polynomials is considered, one can collect the corresponding Zernike coefficients into a vector $\boldsymbol{\alpha} \in \mathbb{R}^{N_\alpha}$, and neglect the piston by arbitrarily setting $\alpha_0^0 = 0$. In this case, we have that $\mathrm{rms}(\Phi) = (\mathrm{var}(\Phi))^{1/2}$, and one can easily compute the rms by evaluating the 2-norm of $\boldsymbol{\alpha}$, i.e., $\mathrm{rms}(\Phi) = \|\boldsymbol{\alpha}\|$.

1.3 Adaptive optics

1.3.1 Introduction

Adaptive optics (AO) is concerned with minimising the aberrations in an optical system, and was initially conceived by Babcock [47] in 1953 to counteract "seeing", the detri-

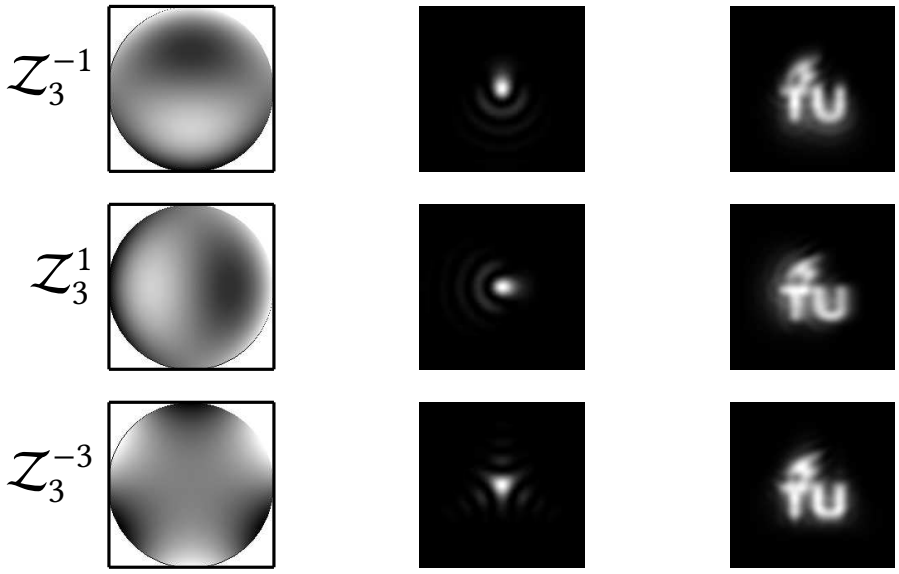

Figure 1.11: See the caption of Fig. 1.9. The primary coma ($\mathcal{Z}_3^{-1}$ and $\mathcal{Z}_3^{1}$), and a trefoil aberration ($\mathcal{Z}_3^{-3}$) polynomials are shown in each row.

mental effect on astronomical observations caused by the turbulent atmosphere of the earth. An adaptive optics system was also independently described by Linnik in 1957 [48]. When considering only phase aberrations [35], the objective of AO is to render the phase aberration function $\Phi(\rho, \theta)$ identically zero over the pupil [4], which implies that all the Zernike coefficients α_n^m are also zero.

An illustration of an adaptive optics system is found in Fig. 1.12. Light emitted from a distant celestial object propagates through space, where n is uniformly equal to one, and reaches the atmosphere of the earth. At this point no phase aberration is present. As light propagates through the atmosphere and reaches the aperture of a telescope, it passes through a turbulent medium where n varies randomly as function of time and space. As a consequence, a phase aberration Φ_{ab} is found in the entrance pupil of the telescope. The aperture of the telescope is reimaged onto a deformable mirror [35, 49, 50] (DM) with N_a actuators, which introduces a controllable phase aberration Φ_{dm}, such that the residual phase aberration after reflection by the DM becomes $\Phi_r = \Phi_{ab} + \Phi_{dm}$. An image of the celestial object is finally formed by focusing the light onto a detector (CCD). A beam splitter (BS) directs part of the light onto a wavefront sensor (SH), which provides an estimate of Φ_r, for example in the form of a set of N_α Zernike coefficients collected into a vector $\hat{\boldsymbol{\alpha}} \in \mathbb{R}^{N_\alpha}$. A controller ($C$) receives $\hat{\boldsymbol{\alpha}}$ as input, and computes a vector $\mathbf{u} \in \mathbb{R}^{N_a}$ that contains the control signals of the N_a actuators of the DM. The objective

[4] The phase aberrations are also completely suppressed when $\Phi(\rho, \theta) = \alpha_0^0 \cdot \mathcal{Z}_0^0$ for a non-zero α_0^0. As seen earlier, we can neglect the piston coefficient and assume $\alpha_0^0 = 0$.

of the controller is to minimise $\|\hat{\boldsymbol{\alpha}}\|$.

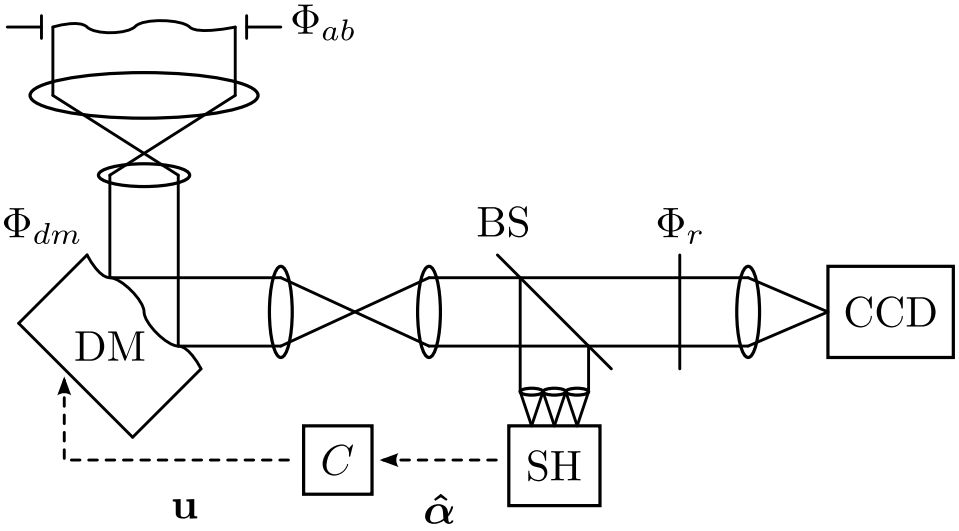

Figure 1.12: Example of an adaptive optics system. An aberrated wavefront is incident on the aperture of a telescope, which is reimaged onto a DM. After reflection on the DM, light passes through beam splitter BS and is focused onto a CCD detector. The residual phase aberration is $\Phi_r = \Phi_{ab} + \Phi_{dm}$, where Φ_{ab} and Φ_{dm} are respectively the initial wavefront aberration and the aberration introduced by the DM. Part of the light is directed by BS onto a wavefront sensor SH, which estimates a finite set of Zernike coefficients. The controller C uses $\hat{\boldsymbol{\alpha}}$ to drive the actuators of the DM $\mathbf{u}$ and cancel the aberration.

The problem of controlling an AO system has been extensively studied for astronomy applications [35]. For example, optimal control algorithms [51, 52, 53, 54] and adaptive algorithms [55, 56] have been investigated. A comprehensive review of control strategies for AO is found in [57, 58]. More recently, effort has been spent in devising control algorithms for large-scale AO systems [59, 60, 61, 62, 63, 64], where the wavefront correcting element is expected to have up to 40000 degrees of freedom [64].

Adaptive optics has also found numerous applications in other fields. For example, AO has been recently considered to counteract thermally-induced aberrations in EUV lithography [65, 66, 67], and to suppress spherical aberration in laser machining [68]. Extensive use of AO is now common in fields such as ophthalmology [69, 70, 34, 71], optical coherence tomography [72, 73, 74, 75, 76], and microscopy [36, 37, 38].

1.3.2 Shack–Hartmann wavefront sensing

The wavefront sensor has a pivotal role in AO, as it allows the controller to compute the necessary aberration correction. An example of a wavefront sensor is the Shack–Hartmann wavefront sensor [77, 78, 79] (SHWFS), which has been extensively studied

and experimentally validated in astronomy applications [35]. It consists of an aperture with small lenses, called the lenslet array, that focus light onto a CCD detector, see Fig. 1.13. Each lens defines a subaperture that samples a different part of the wavefront in the pupil of the optical system. In this section we describe in some detail the

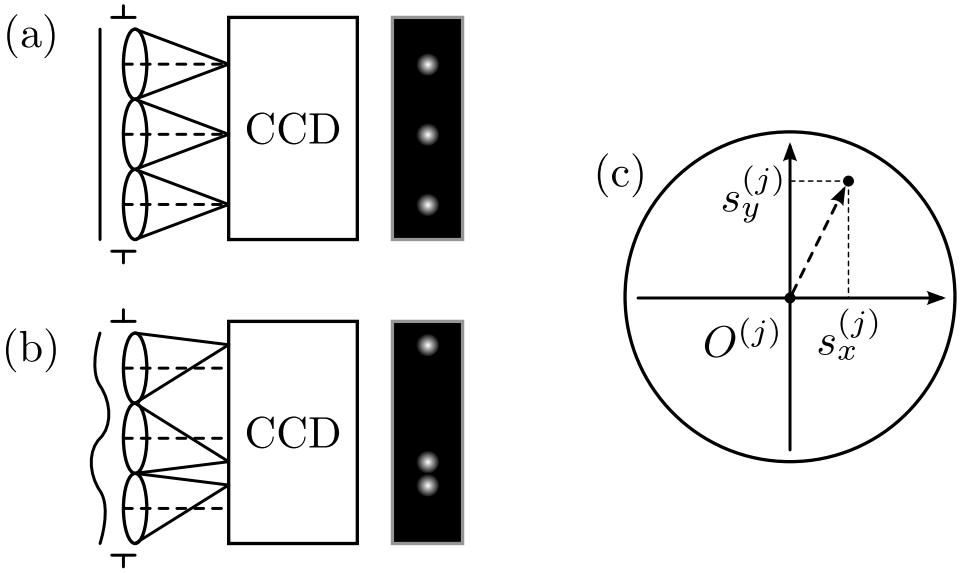

Figure 1.13: Illustration of the Shack–Hartmann wavefront sensing principle. (a) A flat wavefront is incident on the lenslet array of the SHWFS. The CCD detector records a focal spot for each subaperture. (b) An aberrated wavefront is incident on the lenslet array. The focal spots are displaced from their reference position. (c) Plot of the displacement vector $(s_x^{(j)}, s_y^{(j)})$ referenced to the centre $O^{(j)}$ of the subaperture (j).

implementation of a modal-based wavefront reconstruction method. We provide the formulas to compute the definite integrals of the Zernike polynomial gradients over each subaperture. These integrals can be easily computed since the domain of integration is normal in θ, and the Zernike polynomials are separable in polar coordinates. To the best knowledge of the author, these formulas have not been reported in the literature.

We assume that the SHWFS has a circular aperture so that the phase aberration is given by $\Phi(\rho,\theta) = \sum_{n,m} \alpha_n^m \mathcal{Z}_n^m(\rho,\theta)$. To calibrate the SHWFS, light from a point source is collimated generating a plane wave with negligible phase aberration ($\Phi(\rho,\theta) \approx 0$), which is then directed towards the SHWFS. As a consequence, a focal spot appears in each region of the CCD that corresponds to a subaperture, see Fig. 1.13(a). For each subaperture (j), the centre of the focal spot $(x_c^{(j)}, y_c^{(j)})$ is found by computing the centroids

$$x_c^{(j)} = \frac{\sum_i x_i p_i}{\sum_i p_i}, \qquad y_c^{(j)} = \frac{\sum_i y_i p_i}{\sum_i p_i}, \tag{1.18}$$

where p_i is the intensity measured by the pixel in location (x_i, y_i), in a global reference frame within the CCD surface, which is centred to aperture of the SHWFS.

During normal operation, an aberrated wavefront is incident on the SHWFS, as illustrated in Fig. 1.13(b), causing a displacement of the focal spots. As depicted in Fig. 1.13(c), the displacements $s_x^{(j)}$ and $s_y^{(j)}$ are found by computing the centroids with respect to the reference position $O^{(j)}$ located at $(x_c^{(j)}, y_c^{(j)})$ in each subaperture, i.e.,

$$s_x^{(j)} = \frac{\sum_i (x_i - x_c^{(j)}) p_i}{\sum_i p_i}, \qquad s_y^{(j)} = \frac{\sum_i (y_i - y_c^{(j)}) p_i}{\sum_i p_i}. \tag{1.19}$$

If one replaces the summation symbols with continuous integrals in Eq. (1.19), it can be shown [80, 81] that

$$\begin{aligned} s_x^{(j)} &\approx \frac{\lambda f}{2\pi A_{\text{sa}}} \int_{A_{\text{sa}}} \frac{\partial \Phi(\rho,\theta)}{\partial x} \, \rho \, d\rho \, d\theta, \\ s_y^{(j)} &\approx \frac{\lambda f}{2\pi A_{\text{sa}}} \int_{A_{\text{sa}}} \frac{\partial \Phi(\rho,\theta)}{\partial y} \, \rho \, d\rho \, d\theta, \end{aligned} \tag{1.20}$$

where f is the focal length of each lenslet, A_{sa} is the area of each subaperture, and the integrals are restricted to A_{sa}.

As suggested by Dai [82], one can obtain the derivatives of each Zernike polynomial with respect to x and y in polar coordinates,

$$\begin{aligned} \frac{\partial \mathcal{Z}_n^m(\rho,\theta)}{\partial x} &= \frac{\partial R_n^m(\rho)}{\partial \rho} \Theta_n^m(\theta) \cos(\theta) - \frac{R_n^m(\rho)}{\rho} \frac{\partial \Theta_n^m(\theta)}{\partial \theta} \sin(\theta), \\ \frac{\partial \mathcal{Z}_n^m(\rho,\theta)}{\partial y} &= \frac{\partial R_n^m(\rho)}{\partial \rho} \Theta_n^m(\theta) \sin(\theta) + \frac{R_n^m(\rho)}{\rho} \frac{\partial \Theta_n^m(\theta)}{\partial \theta} \cos(\theta). \end{aligned} \tag{1.21}$$

We consider a fixed number N_α of Zernike polynomials, i.e., a truncation of Eq. (1.11). By stacking the displacements $s_x^{(j)}$ and $s_y^{(j)}$ into a vector $\mathbf{s} \in \mathbb{R}^{2N_{\text{sa}}}$, where N_{sa} is the number of subapertures present in the SHWFS, one finds the following linear relationship,

$$\mathbf{s} \approx E\boldsymbol{\alpha}, \tag{1.22}$$

where each element in matrix $E \in \mathbb{R}^{2N_{\text{sa}} \times N_\alpha}$ is given by

$$\begin{aligned} e_{(j),x}^{n,m} = \frac{\lambda f}{2\pi A_{\text{sa}}} \int_{\theta_a^{(j)}}^{\theta_b^{(j)}} \Bigg(& \int_{\rho_a^{(j)}}^{\rho_b^{(j)}} \frac{\partial R_n^m(\rho)}{\partial \rho} \, \rho \, d\rho \cdot \Theta_n^m(\theta) \cos(\theta) - \\ & \int_{\rho_a^{(j)}}^{\rho_b^{(j)}} R_n^m(\rho) \, d\rho \cdot \frac{\partial \Theta_n^m(\theta)}{\partial \theta} \sin(\theta) \Bigg) \, d\theta, \\ e_{(j),y}^{n,m} = \frac{\lambda f}{2\pi A_{\text{sa}}} \int_{\theta_a^{(j)}}^{\theta_b^{(j)}} \Bigg(& \int_{\rho_a^{(j)}}^{\rho_b^{(j)}} \frac{\partial R_n^m(\rho)}{\partial \rho} \, \rho \, d\rho \cdot \Theta_n^m(\theta) \sin(\theta) + \\ & \int_{\rho_a^{(j)}}^{\rho_b^{(j)}} R_n^m(\rho) \, d\rho \cdot \frac{\partial \Theta_n^m(\theta)}{\partial \theta} \cos(\theta) \Bigg) \, d\theta. \end{aligned} \tag{1.23}$$

The subscripts $\cdot_{(j),x}$ and $\cdot_{(j),y}$ denote the two rows of E corresponding to aperture (j), and the superscript $\cdot^{n,m}$ denotes the column of E that corresponds to $\mathcal{Z}_n^m$.

As illustrated in Fig. 1.14, the boundaries of the integration intervals are

$$\begin{aligned}
\theta_a^{(j)} &= \theta_0 - \arctan(\rho_{\mathrm{sa}}/\rho_0), \\
\theta_b^{(j)} &= \theta_0 + \arctan(\rho_{\mathrm{sa}}/\rho_0), \\
\rho_a^{(j)} &= \rho_0 \cos(\theta - \theta_0) - \sqrt{\rho_0^2(\cos(\theta - \theta_0)^2 - 1) + \rho_{\mathrm{sa}}}, \\
\rho_b^{(j)} &= \rho_0 \cos(\theta - \theta_0) + \sqrt{\rho_0^2(\cos(\theta - \theta_0)^2 - 1) + \rho_{\mathrm{sa}}},
\end{aligned} \tag{1.24}$$

where (ρ_0, θ_0) are the polar coordinates of $O^{(j)}$ and ρ_{sa} is the radius of the subapertures. In this case we have assumed circular subapertures. Nevertheless, the integrals on the right-hand side of Eq. (1.20) can also be easily computed without approximation in case of square subapertures, by using the Zernike polynomials expressed in cartesian coordinates. Lenslet arrays are usually fabricated using square or hexagonal lenses [83], to minimise the space between each pair of lenses and to collect more light. For hexagonal lenses, we use the circular approximation of A_{sa} as outlined in this section.

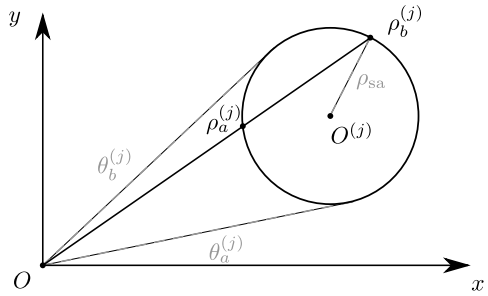

Figure 1.14: Illustration of the boundaries of the integration intervals reported in Eq. (1.24). O is the centre of the global reference frame over the CCD. $O^{(j)}$ is the centre of subaperture (j).

Using the reference image (see Fig. 1.13(a)), matrix E can be precomputed at calibration time, by numerically evaluating the integrals in θ. During operation of the SHWFS, an aberrated image is recorded with the CCD (Fig. 1.13(b)), the displacements in Eq. (1.19) are computed, and finally an estimate of $\boldsymbol{\alpha}$ is obtained by solving Eq. (1.22) in a least-squares sense, under the assumption that $N_\alpha < 2N_{\mathrm{sa}}$. One can consider the condition number [84] of E to select the number of Zernike polynomials N_α to estimate. The condition number depends [82, 85, 86] on N_{sa} and on the arrangement of the subapertures, which are fixed parameters once the lenslet array is manufactured.

Numerous wavefront reconstruction [87] methods have been proposed in the literature. The so called zonal-based methods were initially developed by Hudgin and Fried [88, 89, 90, 91, 92]. These methods establish a rectangular grid where each node represents an unknown value of the phase aberration function $\Phi[i,j] \in \mathbb{R}$, and the displacements $s_x^{(j)}$ and $s_y^{(j)}$ are linear combinations of neighbouring nodes, e.g., $s_x^{(j)} = \Phi[i+1,j] - \Phi[i,j]$. Zonal methods do not provide a Zernike analysis of $\Phi(\rho,\theta)$, which can be obtained by fitting [93] the Zernike polynomials in a second step.

Hudgin developed a geometry that uses a single displacement, either $s_x^{(j)}$ or $s_y^{(j)}$, from each subaperture, and formulates the wavefront reconstruction problem into a least-squares problem. Fried [90] proposed a different grid arrangement whereby both displacements are used. This geometry turns out to be insensitive to the "waffle" mode [35]. Southwell [94] compared the zonal-based methods of Hudgin and Fried with a modal-based method proposed by Cubalchini [95], and concluded that the modal-based method appeared be superior in terms of ease of implementation and noise propagation properties.

A modal-based estimation method that uses the discrete Fourier transform was studied by Freischlad [96], who showed that it is equivalent to a filtering operation [96]. More recently, in [59], Poyneer discussed the computational advantages of using a method based on the fast Fourier transform (FFT), which is suitable for large-scale adaptive optics systems. In her paper, a discussion is found about the issue of the missing boundary displacements (see Fig. 3 and Fig. 4 in [59]), which arises when a method based on a rectangular grid is used with a non-rectangular arrangement of apertures. Padding the missing displacements with zeros leads to a large reconstruction error [59]. More recently, a wavefront reconstruction method based on splines has also been proposed [97]. This method accommodates non-rectangular arrangements of subapertures. The issue of aliasing with higher-order modes for modal-based methods was first discussed by Herrmann in [98] and later by Dai [82, 85, 86].

Finally, we remark that the method outlined in this section is also independent of the arrangement of the subapertures and does not suffer from the issue of the missing boundary displacements. Because the integrals in Eq. (1.20) have been computed numerically, the displacements $s_x^{(j)}$ and $s_y^{(j)}$ are not defined as the difference between two nodes in a grid, as is the case for zonal methods and for the modal-based method described in [95]. A MATLAB toolbox that implements the method described in this subsection is freely available [99].

1.4 Adaptive optics in microscopy

Aberrations in microscopy arise from the fact that specimens are heterogeneous media. To illustrate this point, three different cases are considered in Fig. 1.15. In Fig. 1.15(a), a microscope objective converts a flat wave into a spherical wave, which converges to the focal point. In this case, the medium under the objective is homogeneous and has a constant index of refraction equal to n_1. A diffraction-limited focal spot is created in the focal point, i.e., the intensity of the field is proportional to $I_w(v)$ in Eq. (1.1). In Fig. 1.15(b), instead, the medium is heterogeneous and has a non-constant index of refraction. As a consequence, an aberrated focal spot is created. As shown in Fig. 1.15(c), adaptive optics [100, 37, 38] can be used to introduce an aberration in the pupil of the objective, so that some of the specimen aberration is cancelled.

The effects of aberrations in confocal microscopy were initially studied in [101, 102]. Experiments showed, for both confocal [103, 104] and multi-photon microscopy [105], a substantial degradation of the fluorescence emission and of the resolving power when focusing through media with refractive index mismatches. This is a common situation in microscopy where the indices of refraction of the immersion liquid, cover glass, and

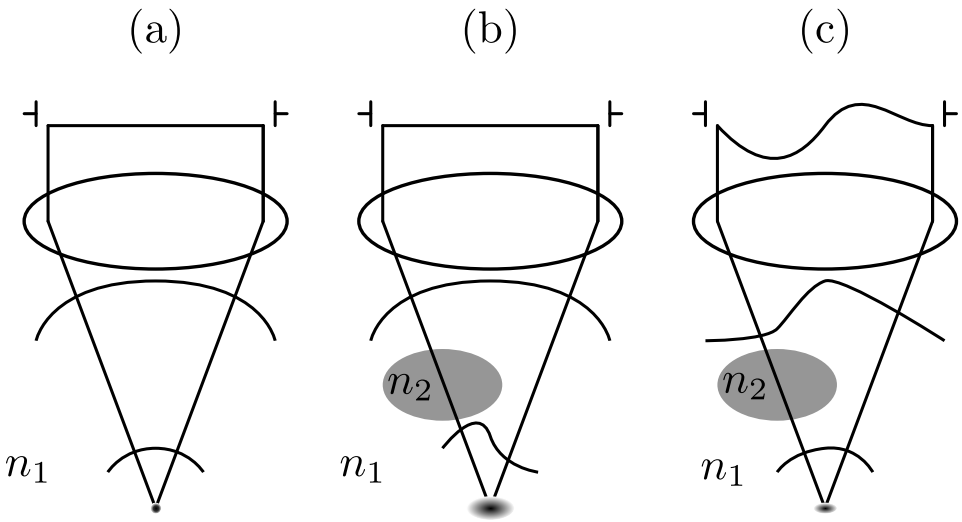

Figure 1.15: Illustration of a microscope objective that focuses light. (a) the medium under the objective is homogeneous with index of refraction n_1, and a diffraction-limited focal spot is formed. (b) the medium under the objective is heterogeneous ($n_1 \neq n_2$), and an aberrated focal spot is formed. (c) AO is used to minimise the aberration at the focus by introducing an aberration in the pupil of the objective.

specimen usually cannot match, resulting in a depth-dependent spherical aberration. Efforts were made to model the image formation in the presence of stratified media for confocal [31, 104, 106, 107] and multi-photon [32] microscopy. Hell [31] notes that several other factors, not considered in the image formation models, can also contribute to loss of fluorescence intensity and resolution, such as the diffusion of fluorophores inside the specimen, losses due to scattering and absorption, and fluorescence saturation [108]. Nevertheless, aberrations are expected to be the predominant factor, as shown in experimental verification [31].

Incorporating an additional lens to compensate the spherical aberration was considered in [109]. Instead, the benefits of the more general approach of AO were considered in [30], where it was concluded that correcting up to the third-order ($\mathcal{Z}_8^0$) spherical aberration essentially recovers diffraction-limited imaging. Early demonstrations of aberration correction are found in [110, 111, 112, 113] for two-photon microscopy and in [114] for confocal microscopy.

1.4.1 Specimen-induced aberrations in microscopy

The aberrations induced by a number of typical biological specimens were measured using phase step interferometry in [115, 116]. A Zernike analysis of the aberrations showed that high-order Zernike polynomials have only a limited contribution [115, 116, 117] with respect to the overall aberration and therefore a significant improvement is expected when applying AO to correct low-order Zernike aberrations. Two other observations

were that, as expected, spherical aberration was dominant, and the aberrations varied over the field of view. This insight was later further supported in [118], where correcting low-order Zernike aberrations was also found to significantly reduce the overall aberration, even though aberration correction may not be feasible in some parts of the specimen where the distortions are too large. In [118], the authors also studied the correlation between the spatial variations of the aberrations and the structure of the specimen. For example, in skin tissue, the topology of the outermost layers determines the predominant part of the aberration. On the contrary, in mouse hippocampus, the aberrations were mostly determined by in-depth heterogeneity. The combined effect of the specimen structure and of the aberrations has also been studied in [119]. Finally, in [120, 121], using multi-conjugate adaptive optics [122] has been investigated to counteract the spatial variations of the aberrations over the field of view.

From the discussion presented in the previous paragraph, it should be noted that aberrations in microscopy are fundamentally different from aberrations in astronomy. The air flow in the atmosphere corresponds to fully developed turbulence [123], which is mathematically modelled using statistical theory [124, 123, 125]. One has that the physical quantities that affect the index of refraction, such as temperature, exhibit random fluctuations [124]. In scanning microscopy, one can assume that the time necessary to acquire an image is much smaller than the time scale in which biological processes evolve (see for example [126]), since otherwise a distorted image would be obtained. Therefore the spatial variations of the index of refraction are more relevant, and these are deterministically given by the structure of the specimen and by the path followed during the scanning process. As a consequence, modelling the aberrations with statistical theory is more difficult in microscopy [38]. For example, if the scanning acquisition is repeated multiple times, one would not expect significant changes in the aberration maps obtained for the skin specimens in [118], unless the topology of the strata is also changing or photodamage occurred. On the contrary, in astronomy, one would not expect to obtain the same sequence of aberration measurements if an observation period is repeated multiple times, due to the turbulence of the atmosphere. In astronomy, under the same optimal observing conditions, two different seeing periods can be expected to provide similar statistics for the aberrations [127]. In microscopy, unless the same region of a given specimen is acquired multiple times, similar statistics for the aberrations are not guaranteed, as can be seen by qualitatively examining the aberrations maps in [118, 38].

1.4.2 Direct wavefront sensing

In astronomy applications, the wavefront sensor is positioned after the aberrating medium and before the imaging lens, as shown in Fig. 1.12. Such a configuration is not possible in microscopy applications, where the aberrating medium is positioned just after the microscope objective, as shown in Fig. 1.15. As a consequence, measuring the aberrations directly with a wavefront sensor is more involved [128, 129] in microscopy, and specific solutions must be developed.

One solution that uses the back-scattered illumination light was investigated in [130, 131, 132]. The authors used coherence-gating to select only the light originating from the focal region, and a phase stepping interferometry algorithm to retrieve the complex amplitude. In [133], coherence-gating was combined with a confocal pinhole to

reduce ghost reflections and speckles. In this case, the reference beam was tilted and Fourier analysis was employed to recover the complex amplitude from a single fringe pattern. In [130, 131, 132, 133], once the complex amplitude was obtained, virtual Shack-Hartmann wavefront sensing was used to recover the phase aberrations. A disadvantage of this solution is given by the complexity of implementing the interferometric setup. Furthermore, this solution is weakly sensitive to odd-symmetry Zernike aberrations [5], such as coma, due to the double-pass effect [135].

Another solution was investigated in [136], where Shack–Hartmann wavefront sensing was applied directly to the back-scattered light. In this case, the light originating from the out-of-focal regions was rejected using a confocal pinhole. In [137, 138], the influence of the size of the confocal pinhole was studied. In [137], the sensitivity and cross-talk of the measured Zernike aberrations were analysed. This solution also suffers from weak sensitivity to odd-symmetry Zernike aberrations.

A different approach was followed in [139, 140, 141, 142, 143], where the fluorescence emission from point objects inside the specimen is used to perform Shack–Hartmann wavefront sensing. The objects can be endogenous fluorescent microspheres that must be inserted into the specimen [139, 143], fluorescent proteins that label appropriate structures inside the specimen [141] or autofluorescence from specific structures [142].

1.4.3 Wavefront sensorless adaptive optics

Aberration correction in microscopy can also be achieved using wavefront sensorless adaptive optics, where the aberrations are determined indirectly, by analysing the amount of fluorescence emission. This approach only requires the addition of a DM to an existing microscope and avoids the complexity of implementing a wavefront sensor. In practice, an image quality metric is established, and a series of trial aberrations are sequentially applied with the DM until the metric is maximised. The drawback is that the number of necessary trial aberrations can be large [36, 144, 145], consequently leading to increased bleaching and phototoxicity.

Due to its experimental simplicity, sensorless adaptive optics was employed early in microscopy related applications. In [111], a parabolic mirror was used to focus pulsed laser light into a sample. A genetic algorithm was used to maximise the second-harmonic signal emitted from the focal region. This led to the correction of the aberrations caused when the beam is scanned off-axis. In [112], correction of the spherical aberration was demonstrated by applying a genetic algorithm to maximise the emitted fluorescence. A comparison of the performance of general optimisation algorithms used to maximise the fluorescence emission is found in [146, 147, 148], concerning both confocal and two-photon microscopy. Other general optimisation algorithms that have been applied in

[5] A Zernike polynomial $\mathcal{Z}_n^m$ is said to be even when $m > 0$ and odd when $m < 0$ [134, 40]. This denotation refers to Noll's [44] single-index ordering, whereby the polynomials are ordered using a single index j such that, for $m \neq 0$, an even j and an odd j correspond respectively to $\Theta_n^m(\theta) = \cos(m\theta)$ and $\Theta_n^m(\theta) = -\sin(m\theta)$. This denotation should not be confused with the even- and odd-symmetry about the origin in the pupil plane, which instead is determined by whether m and n are even or odd respectively. Direct wavefront measurement using the back-scattered illumination light is weakly sensitive to odd-symmetry aberrations, i.e., Zernike aberrations where m and n are odd.

sensorless adaptive optics include hill-climbing algorithms [113, 146], imaged-based algorithms [149], stochastic parallel gradient descent methods [150] and the Nelder–Mead algorithm [151, 152, 153, 154]. The solutions listed in this paragraph can be denoted as model-free, since they employ off-the-shelf optimisation algorithms that have no prior knowledge about the image quality metric. When a new aberration must be corrected, these algorithms start the optimisation from scratch, as if the image quality metric were a completely general function, and no use is made of the information gained from the previous executions of the algorithms.

For small aberrations, a model of the image quality metric can be exploited to accelerate the correction procedure [155, 156, 157, 158]. In [155], a quadratic polynomial was used to model the image quality metric. Aberration correction is achieved using a closed-form expression that requires $N_\alpha+1$ trial aberrations to correct N_α Zernike aberrations. A model-based solution was also devised for incoherent optical systems in [159]. Model-based wavefront sensorless algorithms have been applied to correct aberrations in a number of different microscopy techniques that include structured illumination microscopy [160], two-photon microscopy [161, 144, 118], second-harmonic microscopy [162, 163], third-harmonic microscopy [126], and STED microscopy [164, 165]. In general, the minimum number of trial aberrations required to apply the correction is linear in N_α, and the accuracy of the correction depends on the number of trial aberrations [144].

It should be mentioned that algorithms that are not based on optimisation have also been developed for two-photon microscopy. In [166, 167], the pupil is divided into segments. Illuminating each segment at a time generates a set of shifted images. By analysing the shifts of the images, the global wavefront can be reconstructed. In [168], a heuristic for rejecting the background fluorescence was considered.

1.5 Contributions & outline of this thesis

This thesis comprises five chapters. The current chapter provides some introductory notions about microscopy, adaptive optics and wavefront sensorless adaptive optics. The contributions of the thesis are collected into three chapters. Each chapter corresponds to a separate scientific publication, uses a self-contained notation, and can be read independently of the other chapters.

- Chapter 2 considers a wavefront sensorless adaptive optics system that is implemented with an optical breadboard. The signal recorded using a photodiode covered by a pinhole is selected as the image quality metric. The metric is modelled with a quadratic polynomial. Quadratic polynomials have been widely employed to model image quality metrics in different optical systems, e.g, in [161, 126, 169, 144, 155, 170, 159, 157, 160, 128]. A general procedure to compute the parameters of the quadratic polynomial directly from input–output measurements is developed. This procedure is implemented and shown to outperform another procedure, previously described in the literature [169].

  A new closed-form expression to estimate the aberration is also developed. Provided that the quadratic polynomial is a valid model of the metric, this expression requires a minimum of $N_\alpha + 1$ trial aberrations to estimate N_α Zernike aberrations.

Aberration correction experiments are performed using the optical breadboard, and a comparison is made between the proposed expression to estimate the aberration and two other aberration correction algorithms.

Reference: J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller, "Semidefinite programming for model-based sensorless adaptive optics," J. Opt. Soc. Am. A **29**, 2428–2438 (2012).

- Chapter 3 describes the results of applying wavefront sensorless adaptive optics to a second-harmonic microscope. A set of basis functions used for controlling the deformable mirror is obtained via the singular value decomposition. This set of basis functions can be made approximately orthogonal to the x-tilt, y-tilt and defocus Zernike aberrations. This is of interest in scanning microscopy, as applying these Zernike aberrations with the DM introduces distortions in the acquired images. This is also relevant for astronomy applications, where the x-tilt and y-tilt correction is usually applied with a separate mirror.

  A collagen fibre specimen is used in the aberration correction experiments. The mean image intensity [162] is selected as the image quality metric, which is again modelled with a quadratic polynomial. The parameters of the polynomial are computed using the procedure developed in Chapter 2, which is verified using a biologically relevant specimen for the first time.

  A new algorithm that computes the least-squares estimate of the aberration by solving a non-convex [171] optimisation problem is considered. With the assumption that the quadratic polynomial is a valid model of the image quality metric, the algorithm requires a minimum of $N_\alpha + 1$ trial aberrations. Aberration correction experiments are performed using the second-harmonic microscope.

  Reference: J. Antonello, T. van Werkhoven, M. Verhaegen, H. H. Truong, C. U. Keller, and H. C. Gerritsen, "Optimization-based wavefront sensorless adaptive optics for multiphoton microscopy," J. Opt. Soc. Am. A **31**, 1337–1347 (2014).

- Chapter 4 investigates using a phase retrieval [172] algorithm to correct the aberrations in a wavefront sensorless adaptive optics system. Using the extended Nijboer–Zernike theory [173, 174], the phase retrieval problem is formulated into a matrix rank minimisation problem [175, 176, 177]. A solution of the phase retrieval problem is obtained using PhaseLift [178, 179], a convex relaxation [180, 181, 182] of the rank minimisation problem.

  The wavefront sensorless adaptive optics system is implemented using an optical breadboard and aberration correction experiments are performed. The performance of the aberration correction is assessed using a Shack–Hartmann wavefront sensor.

  Although this phase retrieval algorithm, as presented in this chapter, cannot be directly applied to correct aberrations in scanning microscopy, it is useful to characterise [183, 184] the deformable mirror.

  Reference: J. Antonello and M. Verhaegen, "Modal-based phase retrieval for adaptive optics," (in preparation).

The conclusions are drawn in Chapter 5. The author implemented the setups used in Chapter 2 and Chapter 4. The second-harmonic microscope used in Chapter 3 was implemented by Dr. T. van Werkhoven.

Chapter 2

Semidefinite programming for model-based sensorless adaptive optics

Wavefront sensorless adaptive optics methodologies are widely considered in scanning fluorescence microscopy where direct wavefront sensing is challenging. In these methodologies, aberration correction is performed by sequentially changing the settings of the adaptive element until a predetermined image quality metric is optimised. An efficient aberration correction can be achieved by modelling the image quality metric with a quadratic polynomial. We propose a new method to compute the parameters of the polynomial from experimental data. This method guarantees that the quadratic form in the polynomial is semidefinite, resulting in a more robust computation of the parameters with respect to existing methods. In addition, we propose an algorithm to perform aberration correction requiring a minimum of $N+1$ measurements, where N is the number of considered aberration modes. This algorithm is based on a closed-form expression for the exact optimisation of the quadratic polynomial. Our arguments are corroborated by experimental validation in a laboratory environment.

Reference: J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller, "Semidefinite programming for model-based sensorless adaptive optics," J. Opt. Soc. Am. A **29**, 2428–2438 (2012).

2.1 Introduction

Adaptive optics is concerned with the active suppression of disturbances in optical systems. The sources of the disturbances can be different, according to the application in question. Notable examples are atmospheric turbulence for astronomy and heterogeneity in the index of refraction within specimens for microscopy. As a consequence, phase aberrations develop in the pupil of the objective lens, severely affecting the quality of

the image [185]. The principle of adaptive optics is that by measuring such phase variations with a sensor, they can be cancelled by appropriately driving an active wavefront correction element. In astronomy this practice is well established with the use of a Shack-Hartmann wavefront sensor and a deformable mirror [185].

Nonetheless, there are instances where the deployment of a wavefront sensor is challenging. This is the case for scanning fluorescence microscopy [36], due to difficulties in the rejection of out-of-focus light and in the lack of reference point sources within specimens [139, 136, 130, 186, 187, 140, 141].

Alternatively, sensorless adaptive optics schemes have been considered, where the fluorescence emission is used as a feedback signal for the suppression of the aberrations. One approach involves the rejection of out-of-focus background [168]. More commonly, instead, aberration correction is achieved by sequentially modulating the adaptive element until a selected image quality metric is optimised. The assumption is that the global extremum of the metric is attained when the aberrations have been maximally suppressed. Examples of such metrics are, among others, sharpness measures for images [153] and the amount of fluorescence emission.

In the literature, a number of proposed solutions make use of model-free optimisations. These include hill-climbing algorithms [113, 146], genetic algorithms [111, 148, 112, 147, 146], image-based algorithms [149, 166], conjugate gradient methods [188], stochastic parallel gradient descent methods [150], and the Nelder–Mead simplex algorithm [151, 152, 153]. Such general methodologies require a large number of measurements of the metric [36, 144, 145] and may not converge to the global optimum [152, 155]. Reducing the number of necessary measurements is a critical factor for the overall image acquisition time [113, 146] and for inhibiting side effects, such as phototoxicity and photobleaching [36].

It has been shown [155] that physical modelling of the image quality metric allows for direct and deterministic optimisation methods, requiring a reduced number of measurements with respect to model-free solutions. Initially, model-based methodologies were proposed for optical systems where the object is a point source. In [170, 155], a quadratic polynomial was employed to model a Strehl-based metric. For small aberrations, it was shown that the proposed model-based approach outperforms model-free algorithms. This result was extended to encompass larger aberrations in [128], by using a metric based on the Lukosz-Zernike functions and a nonlinear detector. In [157] a generalisation was provided to handle arbitrary functions other than the Lukosz-Zernike functions. The case of incoherent imaging was analysed in [159]. Here first principles derivations motivated employing a quadratic polynomial in order to model a metric based on the low spatial frequency content of the recorded images. Similarly, in [160], theoretical derivations supported using a quadratic polynomial to model an image quality metric that is appropriate for structured illumination microscopy. Experimental validation of model-based approaches was also provided for two-photon microscopy [161] and for multiharmonic microscopy [126].

One challenge of model-based approaches is found in the need to compute the parameters of the quadratic polynomial for a given real optical system. Initially, this task was performed using first principles, i.e., by computing the theoretical value of each parameter [170, 155, 128, 157, 159]. In this way, however, imperfections in the real optical

system are not accounted for [169]. Also, experimentally computing the parameters is more suited, for example, in the case of coherent microscopies such as third-harmonic generation [169]. To address these shortcomings, experimental methods for the computation of the parameters were developed [160, 169]. Such methods, nevertheless, fail to guarantee that the quadratic form in the polynomial used to model the image quality metrics is semidefinite. This latter property always follows from the theoretical analysis of the image formation processes [161, 126, 169, 144, 170, 159, 160, 128]. In this paper, we present and validate a new method that guarantees that the semidefiniteness property is satisfied. We compare our procedure with the previously proposed methods [160, 169] and show that a more accurate fitting of the experimental data is achieved. We remark that an inaccurate computation of the parameters of the polynomial adversely affects the performance in the correction of the aberrations as shown elsewhere [169, 144].

Once the parameters of the polynomial are known, the correction of an arbitrary aberration is performed by solving an optimisation problem that exploits the knowledge about the quadratic polynomial. For the imaging system considered in [170, 155, 128, 157], an approximate solution of the optimisation was proposed in [170, 155, 128], using $N + 1$ measurements. In [157] an exact solution was provided, using $N + 1$ measurements. For the remaining imaging systems [161, 126, 169, 144, 159, 160], an exact solution of the optimisation was provided using a minimum of $2N + 1$ measurements. In this paper, we derive an exact solution of the optimisation requiring a minimum of $N+1$ measurements. Because our formulas are derived for a quadratic polynomial in its most general form, all the model-based approaches mentioned so far are encompassed as special cases.

This paper is organised as follows. Section 2.2 provides a first principles derivation showing that a quadratic polynomial can model the image quality metric used in our experimental validation. Section 2.3 considers the experimental computation of the parameters of a quadratic polynomial used to model an image quality metric. Section 2.4 focuses on the algorithms used for aberration correction. Section 2.5 provides a description of the optical system used in the experimental validation. Experimental results are reported in Section 2.6. Finally, conclusions are found in Section 2.7.

2.2 Quadratic modelling of a wavefront sensorless adaptive optics system

2.2.1 Problem formulation

Consider the problem of correcting a static aberration in a wavefront sensorless adaptive optics system. Such a problem can be formulated as follows

$$\max_{\mathbf{u}(k)} \quad \tilde{y}(k) \tag{2.1}$$

where $\tilde{y}(k) \in \mathbb{R}$ is the value of a metric quantifying the image quality, $k \in \mathbb{Z}$ is the discrete time index, and $\mathbf{u}(k) \in \mathbb{R}^N$ is the control signal applied to an active element with N degrees of freedom. An instance of this problem is found when imaging a single focal spot in a fluorescence scanning microscope [4, 23], where static specimen-induced

aberrations are to be suppressed. In this case, the value of metric $\tilde{y}(k)$ depends on the amount of fluorescence emission originating from the focal spot. A phase deformation can be applied to the illumination light in the pupil of the objective lens, for instance by employing a deformable mirror that is controlled via vector $\mathbf{u}(k)$. When the deformation induced by the deformable mirror maximally suppresses the specimen-induced aberration, a solution of Eq. (2.1) is found.

In general we have that $\tilde{y}(k) = f(\mathbf{u}(k))$, where $f(\cdot)$ is a function with a global maximum and possibly multiple local extrema. For this reason, a general nonlinear optimisation algorithm can be employed in order to solve Eq. (2.1) as discussed for the model-free methodologies in the introduction. Instead, model-based methodologies exploit the fact that within a suitable neighbourhood of the global maximum, $f(\cdot)$ can be approximated by a quadratic polynomial. Here, metric $\tilde{y}(k) = f(\mathbf{u}(k))$ can be modelled with an approximate metric $y(k) = q(\mathbf{u}(k))$, where $q(\cdot)$ is a quadratic polynomial. The knowledge about $q(\cdot)$ allows us to efficiently solve Eq. (2.1). In the next section we provide a derivation for $q(\cdot)$ based on first principles for the optical system that was used in our experimental validation. This serves as an example in order to highlight the advantage of experimentally determining $q(\cdot)$ as proposed in this paper.

2.2.2 Modelling of a wavefront sensorless adaptive optics imaging system

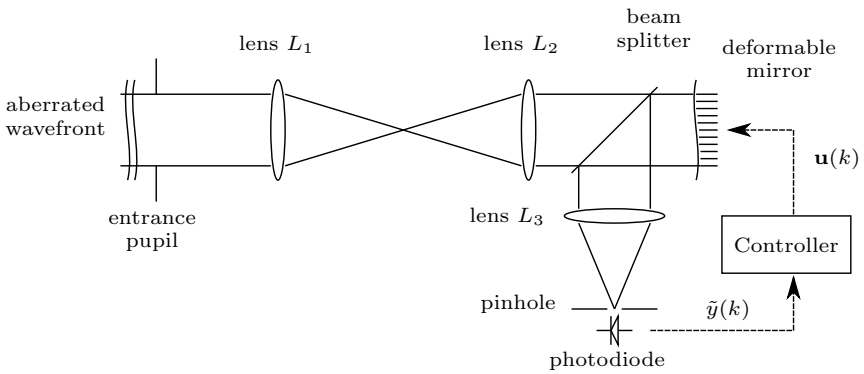

Figure 2.1: Schema representing a sensorless adaptive optics system. An unknown aberration applied at the entrance pupil of the system must be corrected by a deformable mirror that is conjugated to the entrance pupil. The measurement $\tilde{y}(k)$ made with a photodiode covered by a pinhole is an indicator of the residual aberration in the wavefront. The controller changes control signal $\mathbf{u}(k)$ in order to maximise $\tilde{y}(k)$.

Consider the optical configuration in Fig. 2.1. A disturbance in the entrance pupil of L_1 induces an unknown time-invariant aberration to the wavefront. The entrance pupil is reimaged by lenses L_1 and L_2 onto the membrane of the deformable mirror. An image is formed by lens L_3 onto a photodiode, which is covered by a pinhole aperture. Let $\tilde{y}$ denote the integral over the pinhole aperture of the intensity distribution in the focal plane of

L_3. Such quantity will be hereafter referred to with the general term "intensity," as is commonly done in the literature [170, 155, 156]. The intensity is taken to be proportional to the voltage recorded at the output of the photodiode. As outlined in [170, 155, 156], it is assumed that $\tilde{y}$ is a valid image quality metric, i.e., maximising $\tilde{y}$ minimises the aberration of the wavefront. In Section 2.6.3 such a conjecture is experimentally verified by measuring the residual aberration with a Shack-Hartmann wavefront sensor.

The intensity $\tilde{y}(k)$ can be modelled [11] by

$$\tilde{y}(k) = \left(\frac{A}{\lambda f}\right)^2 \int_{\Sigma_2} \left| \int_{\Sigma_1} \exp\left[j\frac{2\pi}{\lambda} W(\xi,\eta,k) - j\frac{2\pi}{\lambda f}(\xi\alpha + \eta\beta) \right] d\xi d\eta \right|^2 d\alpha d\beta + w(k), \quad (2.2)$$

where k is the time index, A and λ are, respectively, the amplitude and wavelength of the monochromatic wave, f is the focal distance of L_3, Σ_2 is the pinhole aperture, Σ_1 is the pupil, $j = \sqrt{-1}$, $W(\xi,\eta,k)$ is the wavefront aberration and $w(k)$ is the measurement noise.

If the effects due to the finite size of the pinhole are neglected (see [170, 155]),

$$\tilde{y}(k) \approx \left(\frac{A}{\lambda f}\right)^2 \left| \int_{\Sigma_1} \exp\left[j\frac{2\pi}{\lambda} W(\xi,\eta,k) \right] d\xi d\eta \right|^2 + w(k). \quad (2.3)$$

Let $\tilde{\Phi}(\xi,\eta,k) = \frac{2\pi}{\lambda} W(\xi,\eta,k)$ and $S = \int_{\Sigma_1} d\xi d\eta$. As done in [5, 46] for k fixed, the exponential in Eq. (2.3) is expanded into a Taylor series and the terms of order higher than 2 are neglected. The approximation becomes

$$\tilde{y}(k) \approx \left(\frac{AS}{\lambda f}\right)^2 \left[1 - \left(\frac{1}{S} \int_{\Sigma_1} \tilde{\Phi}(\xi,\eta,k)^2 d\xi d\eta - \left(\frac{1}{S} \int_{\Sigma_1} \tilde{\Phi}(\xi,\eta,k) d\xi d\eta \right)^2 \right) \right] + w(k). \quad (2.4)$$

We assume that $\tilde{\Phi}$ can be expanded into the following series

$$\tilde{\Phi}(\xi,\eta,k) = \sum_{i=1}^{\infty} \Psi_i(\xi,\eta) v_i(k). \quad (2.5)$$

Possible choices for $\Psi_i(\cdot,\cdot)$ include, among others, Zernike polynomials [44], mirror modes [189], or simply the influence functions of the actuators of a deformable mirror [151]. Let $\mathcal{F}(\xi,\eta) = \begin{bmatrix} \Psi_1(\xi,\eta) & \dots & \Psi_N(\xi,\eta) \end{bmatrix}^T$ and $\mathbf{v}(k) = \begin{bmatrix} v_1(k) & \dots & v_N(k) \end{bmatrix}^T$, then the N-th order truncation of Eq. (2.5) is denoted as

$$\tilde{\Phi}(\xi,\eta,k) \approx \mathcal{F}(\xi,\eta)^T \mathbf{v}(k). \quad (2.6)$$

Substituting Eq. (2.6) into Eq. (2.4) and performing the integrations leads to the following quadratic approximation,

$$\tilde{y}(k) \approx \tilde{c}_0 - \mathbf{v}(k)^T \tilde{Q} \mathbf{v}(k) + w(k), \quad (2.7)$$

where $\tilde{c}_0 = (\frac{AS}{\lambda f})^2$ and

$$\tilde{Q} = \tilde{c}_0 \left(\frac{1}{S} \int_{\Sigma_1} \mathcal{F}(\xi,\eta)\mathcal{F}(\xi,\eta)^T d\xi d\eta - (\frac{1}{S} \int_{\Sigma_1} \mathcal{F}(\xi,\eta) d\xi d\eta)(\frac{1}{S} \int_{\Sigma_1} \mathcal{F}(\xi,\eta)^T d\xi d\eta) \right). \tag{2.8}$$

Note that the elements of $\mathcal{F}$ need not be orthogonal over the pupil. If such elements are chosen among N Zernike polynomials, the Strehl ratio can be recognised and $\tilde{Q}$ is diagonal. It can be seen that $\tilde{Q}$ must be positive semidefinite, i.e., $\tilde{Q} \succeq 0$. Vector $\mathbf{v}(k)$ accounts for both the contribution due to the unknown aberration $\mathbf{x}$ in the entrance pupil of L_1 and the aberration $\mathbf{u}(k)$ induced by the deformable mirror. Letting $\mathbf{v}(k) = \mathbf{x} - \mathbf{u}(k)$, Eq. (2.7) is rewritten into

$$\tilde{y}(k) \approx \tilde{c}_0 - (\mathbf{x} - \mathbf{u}(k))^T \tilde{Q} (\mathbf{x} - \mathbf{u}(k)) + w(k). \tag{2.9}$$

Quadratic polynomials, such as the right-hand side of Eq. (2.9) without the noise term $w(k)$, have been used in order to model image quality metrics in a variety of different imaging techniques [161, 126, 169, 144, 155, 170, 159, 157, 160, 128]. This suggests that Eq. (2.9) can be employed as a generalised approximate metric. In each of these publications, a thorough analysis of the image formation process led to a quadratic polynomial where matrix $\tilde{Q}$ was positive semidefinite [respectively negative semidefinite if optimisation Eq. (2.1) is formulated as a minimisation problem as in [128] for example]. The semidefiniteness property of $\tilde{Q}$ stems from the fact that $\tilde{y}$ exhibits a global extremum. Nevertheless, when applying the experimental computations proposed in [160, 169], such a property can be violated. This shortcoming is addressed in the procedure proposed in Section 2.3.2.

2.3 Identification of the parameters for quadratic approximate metrics

Once metric $\tilde{y}$ has been selected, depending on the imaging system (see [161, 126, 169, 144, 170, 159, 160, 128]), parameters $\tilde{c}_0$ and $\tilde{Q}$ must be computed before aberration correction can be applied. This operation is required once only. As in [160, 169], the unknown aberration $\mathbf{x}$ is assumed to be zero throughout this calibration procedure.

One possibility is to compute $\tilde{c}_0$ and $\tilde{Q}$ from their first principles definitions, e.g., Eq. (2.8) for $\tilde{Q}$. Such an approach can be cumbersome. First, accurately measuring a number of quantities in a real system is both prone to errors and inconvenient. Also additional equipment may be necessary. In Eq. (2.8), amplitude A and the surface of the pupil S must be measured. Second, a different numerical integration must be solved for each different choice of the basis functions $\Psi_i(\cdot,\cdot)$ in Eq. (2.5). In addition, a first principles computation of the parameters does not account for defects in the real optical system such as misalignment, imperfect illumination profile, non-circular pupils, etc. (see the discussion in [169]). A final drawback is related to the fact that the overall modelling error is neither explicitly defined nor minimised. As remarked earlier, an inaccurate choice

for $\tilde{Q}$ leads to a decreased performance in the correction of aberrations [169, 144]. The alternative followed in this paper, instead, is to select the values of $\tilde{c}_0$ and $\tilde{Q}$ by minimising the fitting error, e.g., $\sum_k (\tilde{c}_0 - \mathbf{u}(k)^T \tilde{Q} \mathbf{u}(k) - \tilde{y}(k))^2$, over some set of real input–output measurements. For these reasons an experimental computation of $\tilde{c}_0$ and $\tilde{Q}$ is desirable.

2.3.1 Débarre's experimental identification procedure

We briefly report the latest, most accurate experimental procedure to compute $\tilde{Q}$, which was proposed in [169], in order to compare it with our proposed method. The idea is to compute $\tilde{Q}$ by independently estimating each of its elements $\tilde{q}_{i,j}$. First, the diagonal elements are recovered. Afterwards, the off-diagonal elements are computed by estimating each of the $N(N-1)/2$ submatrices of $\tilde{Q}$ of dimension 2×2.

As an example, consider Eq. (2.9) when $N = 3$,

$$y(u_1, u_2, u_3) = \tilde{c}_0 - \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}^T \begin{bmatrix} \tilde{q}_{1,1} & \tilde{q}_{2,1} & \tilde{q}_{3,1} \\ \tilde{q}_{2,1} & \tilde{q}_{2,2} & \tilde{q}_{3,2} \\ \tilde{q}_{3,1} & \tilde{q}_{3,2} & \tilde{q}_{3,3} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}. \tag{2.10}$$

By keeping u_2 and u_3 fixed to zero, an input–output data set is collected. Subsequently, element $\tilde{q}_{1,1}$ and $\tilde{c}_0$ can be estimated by fitting the resulting parabola $y(u_1) = \tilde{c}_0 - \tilde{q}_{1,1} u_1^2$. Repeating this step allows us to recover $\tilde{c}_0$ and all the diagonal elements of $\tilde{Q}$. This demands $p_1 N$ input–output data points altogether, where $p_1 \geq 2$.

The estimation of $\tilde{q}_{2,1}$ can be achieved by taking input–output data sets where u_3 is fixed to zero and u_2 is fixed to a constant $\bar{u}$. A parabola in u_1 results

$$y(u_1, \bar{u}) = \tilde{c}_1 - \begin{bmatrix} u_1 \\ \bar{u} \end{bmatrix}^T \begin{bmatrix} \tilde{q}_{1,1} & \tilde{q}_{2,1} \\ \tilde{q}_{2,1} & \tilde{q}_{2,2} \end{bmatrix} \begin{bmatrix} u_1 \\ \bar{u} \end{bmatrix}. \tag{2.11}$$

The extremum of Eq. (2.11) is reached when $u_1 = -(\tilde{q}_{1,2}/\tilde{q}_{1,1})\bar{u}$. Fitting this latter linear relation allows us to compute $\tilde{q}_{1,2}$. This demands $p_3 \geq 3$ input–output data points for $p_2 \geq 1$ different fixed values of $\bar{u}$. This step needs to be repeated $N(N-1)/2$ times. Altogether, $\tilde{c}_0$ and $\tilde{Q}$ can be estimated using $p_1 N + p_2 p_3 N(N-1)/2$ input–output data points.

One shortcoming of such a methodology is that the total amount of necessary measurements can be large (see Section 2.6.3). Measurements in each input–output data set are only exploited for estimating a subset of the parameters instead of all the parameters at once. Most importantly, this procedure cannot ensure that the resulting matrix $\tilde{Q}$ be semidefinite, as predicted by the theoretical derivations [161, 126, 169, 144, 170, 159, 160, 128]. Indeed, noise in the measurements of $\tilde{y}$ and numerical errors can lead to computing an indefinite $\tilde{Q}$. A more robust estimation of $\tilde{Q}$ (see Section 2.6.1) is achieved including the semidefinite constraint in the estimation. In this way indefinite matrices are excluded *a priori*.

2.3.2 Data driven identification procedure

In [190], we first proposed using semidefinite programming [191] for estimating $\tilde{Q}$ and $\tilde{c}_0$. This allows us to recast the computation of the parameters into a single mathematical optimisation. The constraint $\tilde{Q} \succeq 0$ is also satisfied.

Metric Eq. (2.9) can be slightly generalised by including a linear term. This allows to relax the assumption that no aberration is present during the experimental computation of the parameters. Hence Eq. (2.9) is redefined as

$$y(k) = c_0 + \mathbf{c}_1^T(\mathbf{x} - \mathbf{u}(k)) - (\mathbf{x} - \mathbf{u}(k))^T Q(\mathbf{x} - \mathbf{u}(k)) + w(k) \tag{2.12}$$

where $c_0 \in \mathbb{R}$, $\mathbf{c}_1 \in \mathbb{R}^N$, $Q \in \mathbb{R}^{N\times N}$, $Q \succeq 0$. Term $w(k)$ represents the uncertainty in approximating $\tilde{y}$ with Eq. (2.12), and as such it cannot be measured by definition. Quantities c_0, $\mathbf{c}_1$ and Q are the new set of parameters that must be estimated.

Again, as in Section 2.3.1, we temporarily assume that $\mathbf{x} = \mathbf{0}$. A collection of input vectors $\{\mathbf{u}(k) \in \mathbb{R}^N \mid k = 1,\dots,D\}$ is applied as the input to the deformable mirror and the corresponding measurement of $\tilde{y}$ is recorded. This results into the identification data set $\{(y(k),\mathbf{u}(k)) \mid k = 1,\dots,D\}$. Such a collection of input vectors can be selected arbitrarily, in contrast with the methods proposed in [160, 169], where specific pupil functions must be generated. In addition all the data points are used at once to estimate all the parameters.

Minimal fitting error could be attained by solving a linear least-squares problem [192]. However, such an approach does not guarantee that the constraint $Q \succeq 0$ be satisfied. For this purpose, the following constrained optimisation problem is defined

$$\begin{aligned} &\min_{c_0,\mathbf{c}_1,Q} \sum_{k=1}^{D} |y(k) - (c_0 - \mathbf{c}_1^T\mathbf{u}(k) - \mathbf{u}(k)^T Q\mathbf{u}(k))|^2 \\ &\text{s.t.} \quad Q \succeq 0 \end{aligned} \tag{2.13}$$

where $c_0 \in \mathbb{R}$, $\mathbf{c}_1 \in \mathbb{R}^N$, and $Q \in \mathbb{R}^{N\times N}$ are decision variables. Optimisation Eq. (2.13) belongs to the realm of semidefinite programming [191]. A convenient tool for formulating Eq. (2.13) is the modelling suite YALMIP [193]. The widely used numerical solver SeDuMi [194] is employed to solve Eq. (2.13).

In [190], we included a regularisation condition on Eq. (2.13). This leads to a matrix Q that is strictly positive definite. Such a constraint, however, should be removed as some modes naturally correspond to a null space in Q. Examples of these modes are the piston mode if Zernike polynomials are used, or high-frequency aberrations that deform the Airy disk while keeping the encircled energy stationary. A better approach is to include regularisation only when inverting Q, for instance truncating negligible singular values.

The computational complexity of SeDuMi is a function of the number of scalar decision variables and of the number of rows in the total linear matrix inequality constraint [195]. This latter is related to D in Eq. (2.13). A QR factorisation can be used in order to compress the second-order cone constraint implicit in the cost function of Eq. (2.13). For this

purpose, it is convenient to rewrite the least-squares cost function in Eq. (2.13) as

$$\min_{\mathbf{z}} \|\mathbf{b} - A\mathbf{z}\|_2^2, \tag{2.14}$$

where $\mathbf{b} = \begin{bmatrix} y(1) & \dots & y(D) \end{bmatrix}^T$,

$$A = \begin{bmatrix} 1 & -\mathbf{u}(1)^T & -\mathbf{u}(1)^T \otimes \mathbf{u}(1)^T \\ \vdots & \vdots & \vdots \\ 1 & -\mathbf{u}(D)^T & -\mathbf{u}(D)^T \otimes \mathbf{u}(D)^T \end{bmatrix} \in \mathbb{R}^{D \times (1+N+N^2)}, \tag{2.15}$$

and $\mathbf{z} = \begin{bmatrix} c_0 & \mathbf{c}_1^T & \mathrm{vec}(Q)^T \end{bmatrix}^T$. Here, function $\mathrm{vec}(\cdot)$ denotes the vectorisation transformation and $\otimes$ the Kronecker product [192]. Assuming the rank r of A is $1+N+N(N+1)/2$, let $A = \bar{Q} \begin{bmatrix} \bar{R}^T & 0^T \end{bmatrix}^T$ be the QR factorisation of A, where $\bar{Q} \in \mathbb{R}^{D \times D}$ is an orthogonal matrix and $\bar{R} \in \mathbb{R}^{r \times r}$ is an upper triangular matrix. Left multiplication by $\bar{Q}^T$ inside the norm in Eq. (2.14) leads to the following equivalent formulation for Eq. (2.13)

$$\begin{aligned} &\min_{\mathbf{z}} \|\mathbf{b}_1 - \bar{R}\mathbf{z}\|_2^2 \\ &\text{s.t.} \quad Q \succeq 0 \end{aligned} \tag{2.16}$$

where $\mathbf{b}_1 = \begin{bmatrix} I_{r \times r} & 0 \end{bmatrix} \bar{Q}^T \mathbf{b}$. Solving Eq. (2.16) is preferable to Eq. (2.13) as that leads to a smaller semidefinite programme (see Section 2.6.3).

2.4 Aberration correction for quadratic approximate metrics

With the parameters c_0, $\mathbf{c}_1$ and Q known, we now discuss correcting for the unknown aberration $\mathbf{x}$. In the scanning microscope example, this situation corresponds to introducing the specimen and recording the fluorescence emitted from a given focal volume.

2.4.1 Independent parabolic optimisation algorithm

We briefly outline the correction method proposed in [159] and used in [161, 126, 170, 159, 160, 128]. For simplicity, we take both $\mathbf{c}_1$ and $w(\cdot)$ to be zero. Since $Q = Q^T$, there exists an orthogonal matrix V such that $Q = V \Delta V^T$, where Δ is a diagonal matrix. Let $\mathbf{z} = V^T \mathbf{x}$ and $\mathbf{p} = V^T \mathbf{u}$ and let z_i, p_i, $\lambda_{i,j}$ denote respectively the elements of $\mathbf{z}$, $\mathbf{p}$ and Δ. Eq. (2.12) can be rewritten as

$$y(\mathbf{p}) = c_0 - \sum_{i=1}^{N} \lambda_{i,i} (z_i - p_i)^2 \tag{2.17}$$

The diagonalisation of Q is referred to as "linear crosstalk removal" in [169] and has the purpose of reformulating the N dimensional optimisation of Eq. (2.12) into N independent one-dimensional parabola optimisations. If Q is semidefinite, a global optimum of

Eq. (2.17) is found by composing the result of the one-dimensional optimisations. Consider the i-th parabola optimisation. As $\mathbf{p}$ is the independent variable, we can set $p_j = 0$ for $i \neq j$ so that optimising Eq. (2.17) results in

$$\max_{p_i} \quad \alpha_1 p_i^2 + \alpha_2 p_i + \alpha_3, \tag{2.18}$$

where the coefficients $\alpha_1, \alpha_2, \alpha_3$ are unknown as they depend on $\mathbf{z}$. If three measurements are taken: y_1 for $p_i = -b$, y_2 for $p_i = 0$, and y_3 for $p_i = b$, where $b \in \mathbb{R}_+$ is a bias, a Vandermonde system can be solved giving

$$\begin{cases} \alpha_1 = (y_1 - 2y_2 + y_3)/(2b^2) \\ \alpha_2 = (y_1 - y_3)/(2b) \\ \alpha_3 = y_2 \end{cases} . \tag{2.19}$$

Consequently the extremum of the parabola is found by setting $p_i = -\alpha_2/(2\alpha_1)$ which evaluates to

$$-b(y_1 - y_3)/(2y_1 - 4y_2 + 2y_3) \tag{2.20}$$

(Eq. (33) in [159]). If the measurement for $p_i = 0$ is shared among all the modes, this requires a minimum of $2N + 1$ measurements.

We note that in this way not all information derived from $Q = V\Delta V^T$ has been exploited. In fact, α_1 is known *a priori* as the opposite of the i-th eigenvalue of Q, namely, $-\lambda_{i,i}$. Henceforth, only two coefficients α_2 and α_3 are unknown for each mode. Sharing one measurement among all the modes, one sees that $N + 1$ measurements are sufficient to exactly optimise Eq. (2.17).

We illustrate this fact by examining Eq. (2.17) for $N = 1$. Taking 0 and $\bar{p} \neq 0$ for the independent variable p, we have

$$\begin{cases} y(0) &= c_0 - \lambda z^2 \\ y(\bar{p}) &= c_0 - \lambda z^2 + 2\lambda \bar{p} z - \lambda \bar{p}^2 \end{cases} . \tag{2.21}$$

Considering the difference $\delta = y(\bar{p}) - y(0)$ between two measurements of $\tilde{y}$, we have $p^* = (\delta + \lambda \bar{p}^2)/(2\lambda \bar{p})$. A generalisation to N dimensions is reported in the next section. Note that this result is achieved via a closed-form expression, i.e., without resorting to an approximate solution as was proposed in [155, 170, 128].

2.4.2 Linear least-squares optimisation

In this section we provide formulas for the exact optimisation of Eq. (2.12) in a minimum of $N + 1$ measurements of $\tilde{y}$. In our proposed solution, there is no need to diagonalise Q and the optimisation is solved simultaneously for all the modes in a linear least-squares sense.

The system is excited with M input vectors $\{\mathbf{u}(k) \in \mathbb{R}^N \mid k = 1, \ldots, M\}$, where $M \in \mathbb{N}_+$, $M \geq 2$ and the rank of $\begin{bmatrix} \mathbf{u}(1) & \ldots & \mathbf{u}(M) \end{bmatrix}$ is $\min(N, M)$. The corresponding output values are collected $\{y(k) \in \mathbb{R} \mid k = 1, \ldots, M\}$.

Define $\delta y(k,l) = y(k) - y(l)$. Then, we obtain

$$\begin{aligned}\delta y(k,l) = &- \mathbf{c}_1^T(\mathbf{u}(k) - \mathbf{u}(l)) + 2(\mathbf{u}(k) - \mathbf{u}(l))^T Q\mathbf{x} \\ &+ \mathbf{u}(l)^T Q\mathbf{u}(l) - \mathbf{u}(k)^T Q\mathbf{u}(k) + w(k) - w(l).\end{aligned} \tag{2.22}$$

It can be seen that Eq. (2.22) is linear in the unknown $\mathbf{x}$. By stacking a number of such expressions together, a linear set of equations in $\mathbf{x}$ is established.

From $\{y(k) \in \mathbb{R} \mid k = 1,\dots,M\}$, $M-1$ values of $\delta(\cdot,\cdot)$ can be computed, namely, $\delta(M,1),\dots,\delta(M,M-1)$. The linear set of equations in $\mathbf{x}$ is arranged as

$$F_M\mathbf{x} + L_M\mathbf{e}_M = \mathbf{d}_M, \tag{2.23}$$

where

$$F_M = \begin{bmatrix} 2(\mathbf{u}(M) - \mathbf{u}(1))^T Q \\ \vdots \\ 2(\mathbf{u}(M) - \mathbf{u}(M-1))^T Q \end{bmatrix} \in \mathbb{R}^{(M-1)\times N}, \tag{2.24}$$

$L_M = \begin{bmatrix} -I_{(M-1)\times(M-1)} & \mathbf{1}_{M-1} \end{bmatrix}$, $\mathbf{1}_{M-1}$ denotes a vector of ones, $\mathbf{e}_M^T = \begin{bmatrix} w(1) & \dots & w(M) \end{bmatrix} \in \mathbb{R}^M$ and

$$d_M = \begin{bmatrix} \delta y(M,1) + \mathbf{c}_1^T(\mathbf{u}(M) - \mathbf{u}(1)) - \mathbf{u}(1)^T Q\mathbf{u}(1) + \mathbf{u}(M)^T Q\mathbf{u}(M) \\ \vdots \\ \delta y(M,M-1) + \mathbf{c}_1^T(\mathbf{u}(M) - \mathbf{u}(M-1)) \\ -\mathbf{u}(M-1)^T Q\mathbf{u}(M-1) + \mathbf{u}(M)^T Q\mathbf{u}(M) \end{bmatrix} \in \mathbb{R}^{M-1}. \tag{2.25}$$

The solution $\hat{\mathbf{x}}_M$ of the weighted least-squares problem

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{e}_M^T\mathbf{e}_M \\ \text{s.t.} \quad & F_M\mathbf{x} + L_M\mathbf{e}_M = \mathbf{d}_M, \end{aligned} \tag{2.26}$$

satisfies the normal equation

$$(F_M^T W_M F_M)\hat{\mathbf{x}}_M = F_M^T W_M \mathbf{d}_M, \tag{2.27}$$

where $W_M = (L_M L_M^T)^{-1}$.

We consider the case where the set of equations Eq. (2.23) is not underdetermined, i.e., $M \geq N+1$. Between time $k=1$ and $k=N+1$ inclusive, the system is excited with an $N+1$ input sequence. In Section 2.6, the column vectors of $b\begin{bmatrix}\mathbf{0} & V\end{bmatrix}$ were used for such an input sequence, where b is a bias and V a base of eigenvectors of Q as in Section 2.4.1. At time $k = N+2$, output $y(N+1)$ has been acquired and $\hat{\mathbf{x}}_{N+1}$ is computed. Hence, an estimate of the input $\mathbf{u}(N+2)$ maximising the intensity $y(N+2)$ is

$$\mathbf{u}(N+2) = \hat{\mathbf{x}}_{N+1} - \tfrac{1}{2}Q^{-1}\mathbf{c}_1. \tag{2.28}$$

Similarly, in the following time instants, a refined estimation of $\mathbf{x}$ is obtained by solving the overdetermined system, where $M > N+1$. The control law for sample time $k > N+1$

is

$$\mathbf{u}(k) = \hat{\mathbf{x}}_{k-1} - \tfrac{1}{2}Q^{-1}\mathbf{c}_1. \tag{2.29}$$

Note that if the additive noise $w(k) \in \mathbb{R}$ is assumed to be white noise with zero mean and covariance $E[w(k)w(j)] = \delta(k-j)$, where $\delta(\cdot)$ is the unit pulse, then $\hat{\mathbf{x}}_{N+1}$ is the minimum variance unbiased linear estimator of $\mathbf{x}$ (see [192] for further details). The choice of fixing index k in Eq. (2.22) to M is arbitrary, and different arrangements are possible as long as at least $M-1$ non-zero values of $\delta y(\cdot,\cdot)$ can be computed.

Equation (2.12) models the image quality metric for small aberrations, e.g., when the approximations made in Eq. (2.4) are valid (see Section 2.6.2). For larger aberrations, the error in modelling $\tilde{y}$ with a quadratic polynomial is not negligible and therefore applying Eq. (2.29) fails to maximally suppress the aberration. Nevertheless, experimental evidence suggests that applying multiple iterations of Eq. (2.29) enables to gradually suppress large amounts of aberration. This experimental observation is also reported in [144]. One feasible approach is to implement a window-based aberration correction. Let i denote the i-th iteration. Each iteration consists of a data acquisition part followed by a correction part. In the first part, input–output data is collected by exciting the system with input vectors taken from a neighbourhood centred at the estimate of $\mathbf{x}$ in the previous iteration. In the second part, aberration correction takes place by solving Eq. (2.26) over the input–output data acquired in the first part. This process is illustrated in Fig. 2.2.

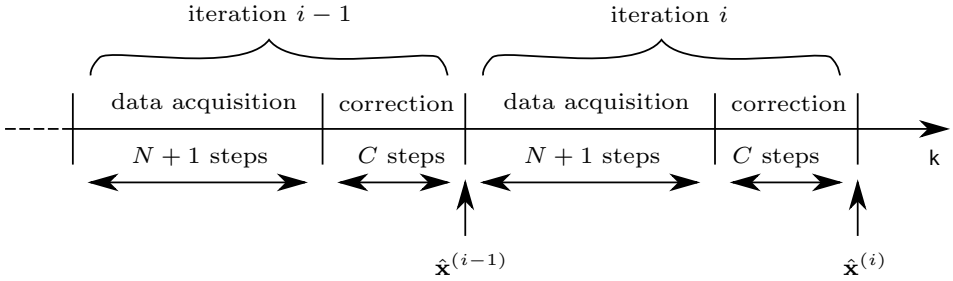

Figure 2.2: Timeline of the iterative aberration correction algorithm. Each iteration consists of a data acquisition part where $N+1$ data points are acquired and a correction part where correction is performed using Eq. (2.29) for C time instants. The vectors in the i-th data acquisition are taken from a neighbourhood of the estimate of $\mathbf{x}$ at iteration $i-1$, i.e., $\hat{\mathbf{x}}^{(i-1)}$.

2.5 Experimental setup

Experimental validation was performed with the system depicted in Fig. 2.3. As was done in [128, 159, 144], a deformable mirror was simultaneously used as the source of the aberration and as the correcting element. Aberrations were generated by adding an offset $\mathbf{x}$ to the control signal of the deformable mirror. Assuming $\mathbf{x}$ to be unknown, an aberration correction algorithm was subsequently applied. The performance of the aberration correction was quantitatively assessed by measuring the residual aberration with a Shack-Hartmann wavefront sensor. Obviously, the measurements from the photodiode were the

only information that was supplied to the aberration correction algorithms. With reference to Fig. 2.3, light from a He–Ne laser source (632.8nm wavelength) is spatially filtered and collimated using lens L_1 (11mm), pinhole P_1 (30 μm), lens L_2 (500mm), and I_1 in order to fill two thirds (10mm) of the membrane of the deformable mirror *DM* (MMDM37, OKOTech, The Netherlands). The pupil is demagnified and reimaged by L_3 (200mm) and L_4 (100mm) onto the hexagonal microlens array *MLA* (127 microlenses, 18 mm focal distance, 300 μm pitch, OKOTech, The Netherlands). The image of the microlens array is recorded with the camera C_1 (svs340, 648×492 pixels, 7.4 μm pixel size, SYSVISTEK, Germany). The beam splitter BS_2 divides light between the Shack-Hartmann wavefront sensor and the pinhole-photodiode sensor. Lens L_5 (200mm) focuses the beam onto the photodiode (TSL250R-LF, TAOS, Korea), which is covered by the 50 μm pinhole P_2. Voltage to the electrodes of the deformable mirror is supplied by a high-voltage amplifier (OKOTech, The Netherlands) with 40 channels. An external power supply provides 150V to the high-voltage amplifier. The system is operated using a desktop PC running Linux. The high voltage amplifier is controlled with a 16-bit analogue output card (PD2-AO-96/16A, United Electronic Industries, United States). Voltage from the photodiode is acquired with a 16-bit analogue input card (PCI-6220, National Instruments, United States). A framegrabber card (Leonardo CL Full, Arvoo, The Netherlands) is used to acquire images from camera C_1. Customised software written in C and MATLAB (Version R2011a, The MathWorks, United States) is used to perform the experiments.

A modal wavefront reconstruction method was implemented [82] using the first 15 Zernike polynomials defined and enumerated as in [44]. We estimated with a least-squares fit a linear relationship between the square root of the voltage applied to the electrodes of the deformable mirror and the Zernike coefficients [196]. As done in [160, 144], 11 Zernike coefficients (Z_5 to Z_{15}, see [44]) were controlled, so that $N = 11$ in the previous formulas. As suggested by OKOtech, bidirectional operation of the deformable mirror was achieved by slightly misaligning L_3 to compensate for the defocus that is introduced when the mirror is biased [196].

2.6 Experimental results

Experimental validation has been performed using the system described in the previous section.

2.6.1 Comparison of the identification procedures for the approximate metric

A comparison was made between Débarre's experimental computation method (see Section 2.3.1) and our proposed procedure (see Section 2.3.2). First the system was initialised by flattening the deformable mirror with the Shack-Hartmann wavefront sensor. Subsequently, the Nelder–Mead simplex method [152] was briefly applied to correct for relative misalignment between the Shack-Hartmann wavefront sensor and the pinhole-photodiode detector. In this way, any residual aberration was removed from the system so that both $\mathbf{x}$ and $\mathbf{c}_1$ are zero in Eq. (2.12). From this initial condition, Débarre's method and Eq. (2.16) were repeatedly applied to compute c_0 and Q.

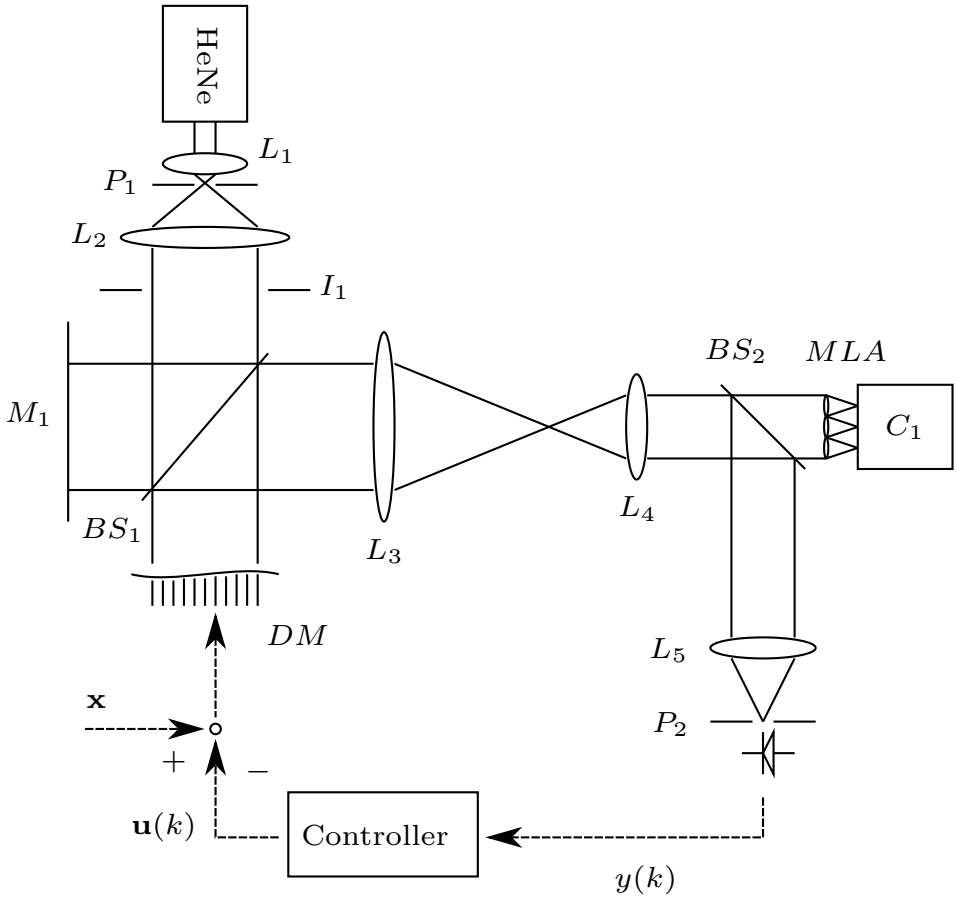

Figure 2.3: The spherical wavefront is generated by spatially filtering a laser beam with lens L_1 and pinhole P_1. The beam is collimated by lens L_2 and clipped by iris I_1 to fill 10mm of the aperture of the deformable mirror *DM*. The membrane of *DM* is reimaged by lenses L_3 and L_4 onto a microlens array *MLA*. C_1 and *MLA* implement a Shack-Hartmann wavefront sensor. Lens L_5 focuses the beam onto a photodiode that is covered by a pinhole P_2. Flat mirror M_1 is used to calibrate the Shack-Hartmann wavefront sensor. An aberration is introduced as an unknown offset $\mathbf{x}$ to the control signal of *DM*. An aberration correction experiment consists of suppressing $\mathbf{x}$ when only the measurements of the photodiode are available. Afterwards, a measurement of the residual aberration is obtained with the wavefront sensor to assess the performance of the correction.

The result of each identification experiment was assessed by evaluating the variance accounted for (VAF). This is defined as $\mathrm{VAF}(\mathbf{y},\hat{\mathbf{y}}) = \max\left(1 - \frac{\mathrm{var}(\mathbf{y}-\hat{\mathbf{y}})}{\mathrm{var}(\mathbf{y})}, 0\right) \times 100\%$, where the elements of vector $\mathbf{y}$ are taken from real output measurements of the system, whereas the elements of vector $\hat{\mathbf{y}}$ are the corresponding output predictions computed using the right-hand side of Eq. (2.12). If the VAF is 100% for one identification experiment, then that indicates that metric $\tilde{y}$ has been perfectly modelled. Forty different identification experiments were executed. In each, an input–output data set was collected and both Débarre's procedure and Eq. (2.16) were applied in order to compute c_0 and Q. The VAF was computed using this identification data. Afterwards, a second input–output data set was collected in order to perform a cross-validation. In this second set, the input was randomly selected. The VAF was then computed using this cross-validation data.

In Fig. 2.4(a), the mean value, maximum, minimum and standard deviation of the VAFs computed with the identification data sets are reported. These show that using Eq. (2.16) guarantees a higher mean value for the VAF with respect to Débarre's procedure. The maximum values for the VAFs show comparable performance between the two procedures. Instead the minimum values and standard deviations for the VAF show that a robust performance is achieved by including the semidefinite constraint in the identification procedure. In fact an indefinite matrix Q was recovered using Débarre's procedure in 32 out of the 40 trials. In Fig. 2.4(b), the mean value, maximum, minimum, and standard deviation of the VAFs computed with the cross-validation data sets are reported. These numbers also support that our identification procedure produces an accurate result. Figure 2.4(c) shows the mean value of Q for both identification procedures. As expected, both matrices have large elements in the diagonal. Nonetheless, some cross-talk elements are also present due to the finite size of the detection pinhole (see [197]). Whereas the diagonal elements are quite similar for both identification procedures, differences are found in the off-diagonal elements. The consequences of employing an inaccurate matrix Q in the correction of aberrations have already been discussed elsewhere [144, 169].

2.6.2 Empirical analysis of the quadratic approximation

Optimisation Eq. (2.16) can also be used to empirically study the region of validity within which metric $\tilde{y}$ can be approximated with Eq. (2.12). For this purpose, a large input–output data set with 50,000 tuples was recorded where the maximum input aberration was 2 rad rms. Afterwards, optimisation Eq. (2.16) was solved by choosing 10 different subsets of the identification data. In each subset, the maximum rms for the input aberrations, denoted by $\rho_{\max}$, was increased and a quarter of the tuples were reserved for validation. The resulting identification and validation VAFs are reported in Fig. 2.5. It can be seen that between 0.4 and 0.6 rad rms, the effect of the modelling error becomes appreciable and $\tilde{y}$ begins to deviate from its value as predicted by a quadratic polynomial.

Experimentally, we found that employing an iterative aberration correction scheme (e.g., the one depicted in Fig 2.2) provided a better performance instead of solving Eq. (2.26) for $M \gg N+1$. This empirical observation is in agreement with Fig. 2.5. Assuming that after each iteration the residual aberration is decreased, then from Fig. 2.5 one can see that the VAF computed with the input–output tuples within each iteration increases. Henceforth, the input–output tuples of the previous iterations should not be reused to solve Eq. (2.26) as they are fitted less accurately by Eq. (2.12).

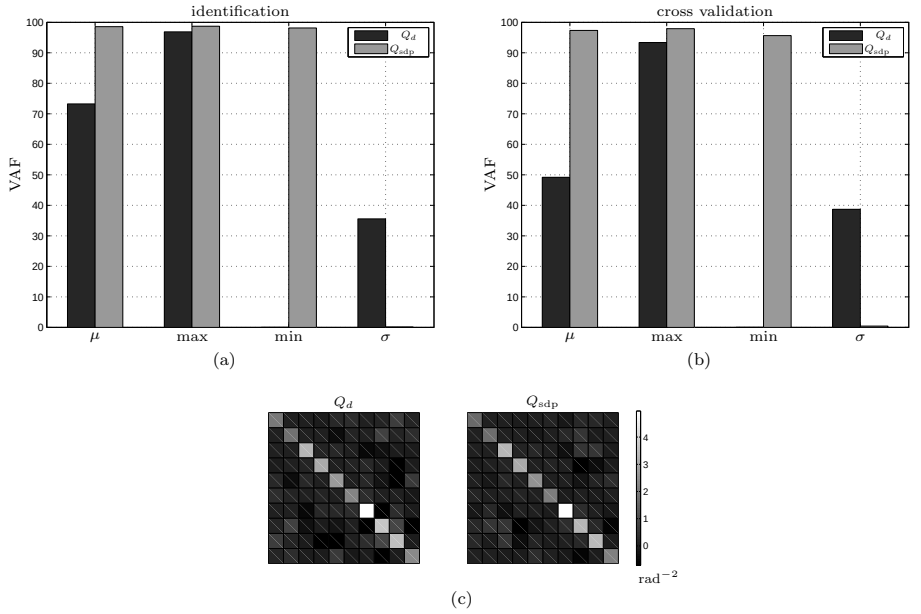

Figure 2.4: Comparison of the experimental computation of matrix Q with Débarre's method (see Section 2.3.1) and our proposed procedure (see Section 2.3.2). First a data set of 15246 input–output tuples is acquired ($p_1 = 21$, $p_2 = 13$ and $p_3 = 21$) and Q is computed with Débarre's method, resulting in Q_d. The same input–output data set is used to compute Q with Eq. (2.16), resulting in Q_{sdp}. The VAFs for Q_d and Q_{sdp} are computed over the identification data set. Subsequently, a new input–output data set with 15,000 tuples is acquired for cross-validation. In this second set the input aberrations are chosen randomly. The VAFs for Q_d and Q_{sdp} are computed using this latter validation set. Such steps are repeated 40 times. (a) Mean value μ, maximum max, minimum min, and standard deviation σ of the identification VAFs for Q_d and Q_{sdp}. (b) Mean value μ, maximum max, minimum min, and standard deviation σ of the cross-validation VAFs for Q_d and Q_{sdp}. (c) Mean value of Q_d and Q_{sdp} over the 40 realisations. The colour map is scaled to the maximum and minimum of the elements of Q_{sdp} in order to preserve contrast between the two matrices. Matrix Q_d resulted indefinite 32 times out of the 40 trials.

An additional observation regards the selection of the bias b in the initial $N+1$ excitation sequence (see Section 2.4). Using a small bias (e.g. < 0.5 rad rms) is preferable as this ensures that the data points collected during the $N+1$ data acquisition (see Fig 2.2) have a similar VAF.

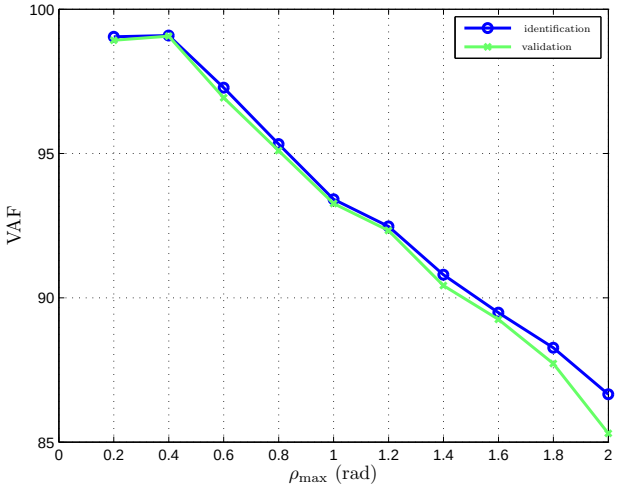

Figure 2.5: Optimisation Eq. (2.16) was solved for 10 different input–output data sets where the maximum rms of the input aberration ($\rho_{\max}$) is linearly increasing up to 2 rad rms. The VAF is reported for both identification (3750 data points) and validation (1250 data points). Between 0.4 and 0.6 rad rms, the difference between $\tilde{y}$ and Eq. (2.12) becomes noticeable.

2.6.3 Aberration correction using the quadratic approximate metric

We report a number of aberration correction experiments where we compare our proposed method with the model-based 3N and 5N algorithms (see [144] and Section 2.4.1) and the model-free Nelder–Mead simplex method [152]. For the model-based algorithms, i.e., the 3N, 5N, and our proposed solution, the same parameters c_0, $\mathbf{c}_1$, and Q were used. Such parameters were identified by collecting an input–output data set with 6000 tuples. The input aberrations were randomly chosen with a maximum rms of 0.5, as motivated in Section 2.6.2. The data set was split into two sets, reserving 4500 tuples for identification and 1500 tuples for validation. Identification was completed in a couple of seconds by solving Eq. (2.16) using YALMIP [193] and SeDuMi [194]. Note that SeDuMi took less time to solve Eq. (2.16) instead of Eq. (2.13) (the ratio between the two computation times was 0.17). A VAF of 98.27% and 98.16% was found respectively for identification and validation.

Figures 2.6(a)–2.6(d) show the results of the correction of random aberrations with magnitudes of 0.3, 0.4, 0.6, and 0.8 rad rms. Such magnitudes are of interest in microscopy applications, where moderate amounts of aberrations are corrected at intermediate depth

levels as one focuses deep within a sample (see Section 5 of [144]). For each figure, 50 random aberrations were generated, where the Zernike coefficients of each aberration were randomly selected so that the resulting rms phase profile had a determined value. In the upper plot, the initial aberration and the final residual aberrations are plotted. The circles, vertical bars, and horizontal bars denote respectively the mean value, standard deviation and minimum and maximum for the correction of the 50 aberrations. The horizontal dashed-dotted magenta line denotes the rms corresponding to a Strehl ratio of 0.9. The lower plot reports the mean value of the intensity against sample time for the 50 correction experiments. The vertical axis is normalised to the maximum intensity recorded when no aberration is applied.

Figure 2.6(a) reports a summary of the correction of 50 random aberrations of 0.3 rad rms with one iteration of our proposed method (LS1) [see Eq. (2.29)], the 3N (3N) algorithm (see [144] and Section 2.4.1) and the model-free Nelder–Mead simplex method [152] (Simplex). The residual aberration is comparable for the three algorithms. Nevertheless, a value of 0.95 for the normalised intensity is already achieved at sample time 13 for (LS1), whereas the other two algorithms reach about 0.94 at sample time 34. A bias of 0.5 rad is used for the 3N algorithm, as suggested in [144] for small aberrations. Instead, a smaller bias of 0.02 rad was used for our proposed solution, as motivated in Section 2.6.2. Such a small bias can be advantageous in scanning image acquisition processes, as it leads to less excitation of the dynamics of the deformable mirror and to a smoother variation of the image as aberration correction is being applied.

Figure 2.6(b) shows a summary of the correction of 50 random aberrations of 0.4 rad rms with one iteration of our proposed method (LS1), the 3N algorithm (3N) and the simplex method (Simplex). Also in this case the two model-based approaches outperform the model-free one. Lower mean values of the residual aberrations are reached for both (LS1) and (3N). In addition, the average normalised intensity is also higher than in (Simplex). This time, at sample time 13, (LS1) reaches a normalised mean intensity of 0.91, which is exceeded by (3N) only at sample time 32. (Simplex) instead stops at 0.9 at sample time 34. Note that both in Fig. 2.6(a) and in Fig. 2.6(b) the mean intensity of (LS1) does not improve significantly after the first correction is applied at sample time $N + 2$, as was discussed in Section 2.6.2.

Fig. 2.6(c) reports a summary of the correction of 50 random aberrations of 0.6 rms with two iterations of our proposed method (LS2) (see Fig. 2.2), the 3N algorithm (3N) and the simplex method (Simplex). In this case, the 3N algorithm is underperforming. A similar behaviour was also reported in [144] (see Fig.4 therein). Nevertheless, a robust performance is shown by (LS2), as can be seen by the reduced standard deviation of the residual aberration and the mean value of the normalised intensity.

In Fig. 2.6(d), 50 random aberrations with a magnitude of 0.8 rms were corrected. Here, four iterations of our proposed method (LS4), the 5N (5N) algorithm (see [144]) and the simplex method [152] (Simplex) are applied. The performance of (5N) is worse than what is reported in [144] (see, for instance, Fig.5 in [144]). Such a discrepancy can be explained by three facts. First, in [144], both the imaging system and the selected metric were different from our case. Second, a maximum bias of 0.5 rad was used, instead of the suggested maximum of 2 rad as in [144]. We found that the deformable mirror could not reproduce such a large deformation without saturating the actuators or producing an inaccurate phase profile, especially for coma, spherical aberration, and second-order

astigmatism. Finally, in [144], the square root of a Lorentzian curve was used instead of Eq. (2.12) (see also the discussion in the next section). Nevertheless, on average, (LS4) reaches a Strehl ratio higher than 0.9 (see the horizontal dashed-dotted line in the upper plot) and outperforms (Simplex).

2.6.4 Aberration correction using non-quadratic approximate metrics

As seen in Section 2.6.2, approximating $\tilde{y}$ with a quadratic polynomial is accurate within a restricted neighbourhood of the global maximum of $\tilde{y}$. Nonetheless, empirically, it was found that a broader range of aberrations could be encompassed when using Gaussian or Lorentzian functions to approximate $\tilde{y}$ (see [160, 126] and [144, 159, 161] respectively). Neglecting issues related to numerical computations and the measurement noise, the quadratic polynomial is recovered by assuming that $t(\tilde{y}(k)) \approx c_0 + \mathbf{c}_1^T(\mathbf{x} - \mathbf{u}(k)) - (\mathbf{x} - \mathbf{u}(k))^T Q(\mathbf{x} - \mathbf{u}(k))$, where $t(\cdot)$ is the logarithm when Gaussian functions are used. The use of such an output transformations to recover the quadratic polynomial follows what has been done in [160, 126, 144, 159, 161].

In Fig. 2.7 we report a summary of the corrections of 50 random aberrations with a magnitude of 1.0 rad rms. Here a Gaussian function was used to model $\tilde{y}$ and a new identification was performed by solving Eq. (2.16). A comparison is made between three iterations of our proposed solution (LS3E) and the simplex method (Simplex). It can be seen that (LS3E) outperforms (Simplex), since a lower mean value is achieved for the residual aberration and the mean intensity is consistently higher than the one of (Simplex) after sample time 11. A study of the amount of aberration that can be corrected with model-based approaches is found in [144] and is not worth repeating here. Similar results should be expected if our aberration correction algorithm is employed.

2.7 Conclusions

In this paper, a new experimental procedure to compute the parameters of quadratic approximate metrics in wavefront sensorless adaptive optics has been presented. Such metrics are applicable to a broad spectrum of different imaging techniques [161, 126, 169, 144, 155, 170, 159, 157, 160, 128]. Our proposed procedure has been shown to produce a more robust computation of the parameters with respect to existing procedures [160, 169]. Arbitrary input–output data can be used without the need to generate specific pupil functions as in [160, 169]. An additional benefit is found in the possibility to empirically study the region of applicability of the quadratic approximate metric.

A second contribution is found in the algorithm used for aberration correction. Formulas have been provided where aberration correction is achieved by exactly optimising the quadratic approximate metric using a closed-form expression in a minimum of $N + 1$ measurements. Since these expressions are given for a quadratic polynomial in its most general form, they are widely applicable [161, 126, 169, 144, 155, 170, 159, 157, 160, 128] and they represent an improvement to previously employed algorithms that required a minimum of $2N + 1$ measurements [161, 126, 169, 144, 159, 160]. Finally, our arguments have been corroborated by experimental validation in a laboratory environment.

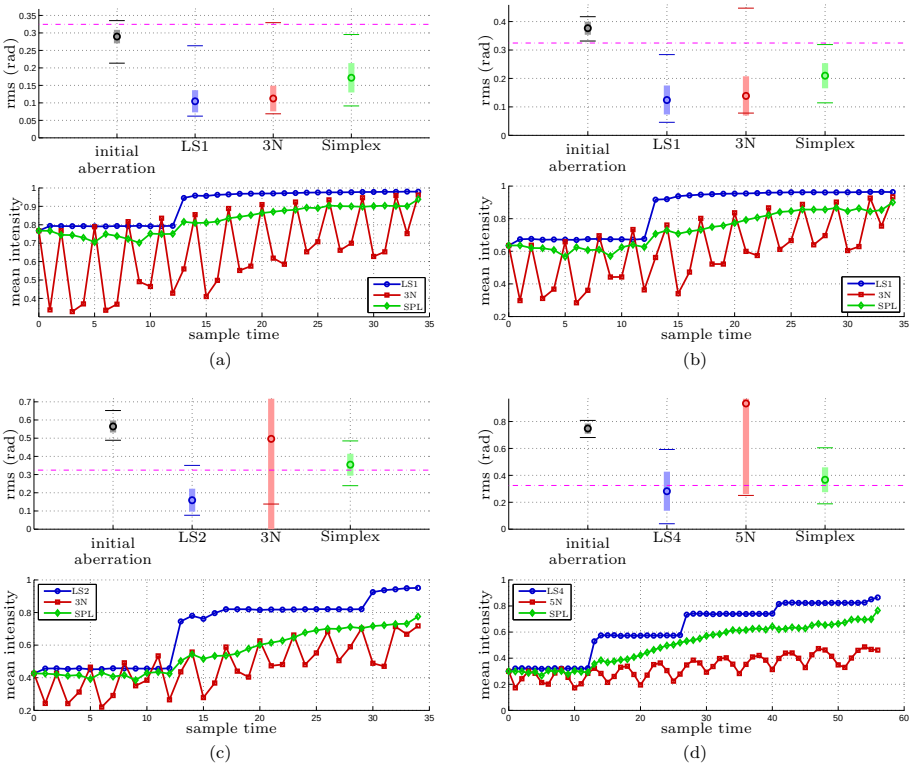

Figure 2.6: Each figure reports a summary of the correction of a set of 50 random aberrations. In the upper plot, the mean value, standard deviation, minimum, and maximum of the residual aberrations after the correction are reported in radians. These are denoted, respectively, by a circle, a thick vertical bar, and thin horizontal lines. The same indicators are also reported for the random initial aberrations before correction. (LS1), (LS2), and (LS4) denote respectively 1, 2, and 4 iterations of Eq. (2.29) as depicted in Fig. 2.2. (3N) and (5N) are described in Section 2.4.1 and [144]. (Simplex) and (SPL) denote the Nelder–Mead simplex method [152]. The horizontal dashed-dotted magenta line denotes a Strehl ratio of 0.9. The lower plots report the mean value of the normalised intensity against sample time for the 50 aberration correction experiments.

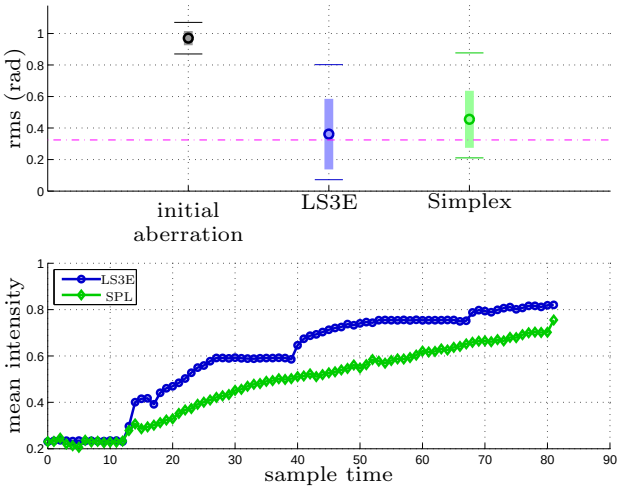

Figure 2.7: Summary of 50 random aberration correction experiments. The same conventions as in Fig.2.6 are employed to report the results. In this case, a Gaussian function instead of Eq. (2.12) was used to model $\tilde{y}$, as outlined in Section 2.6.4. (LS3E) denotes three iterations of Eq. (2.29) as depicted in Fig. 2.2.

Acknowledgements

This research is supported by the Technology Foundation STW under project number 10433, which is part of the Smart Optics Systems perspective programme directed by Prof. M. Verhaegen. The 37 channels micromachined membrane deformable mirror used in the experimental verification in Section 2.6 was provided by Dr. Niek Doelman (TNO). His support is gratefully acknowledged. We would like to thank Mr. Will van Geest and Mr. Kees Slinkman from TU Delft for their contribution in the implementation of the experimental setup.

Chapter 3

Optimisation-based wavefront sensorless adaptive optics for multiphoton microscopy

Optical aberrations have detrimental effects in multiphoton microscopy. These effects can be curtailed by implementing model-based wavefront sensorless adaptive optics, which only requires the addition of a wavefront shaping device, such as a deformable mirror to an existing microscope. The aberration correction is achieved by maximising a suitable image quality metric. We implement a model-based aberration correction algorithm in a second-harmonic microscope. The tip, tilt and defocus aberrations are removed from the basis functions used for the control of the deformable mirror, as these aberrations induce distortions in the acquired images. We compute the parameters of a quadratic polynomial that is used to model the image quality metric directly from experimental input–output measurements. Finally, we apply the aberration correction by maximising the image quality metric using the least-squares estimate of the unknown aberration.

Reference: J. Antonello, T. van Werkhoven, M. Verhaegen, H. H. Truong, C. U. Keller, and H. C. Gerritsen, "Optimization-based wavefront sensorless adaptive optics for multiphoton microscopy," J. Opt. Soc. Am. A **31**, 1337–1347 (2014).

3.1 Introduction

Multiphoton microscopy techniques such as two-photon fluorescence microscopy [23] and second-harmonic microscopy [198] are commonly employed to image biological specimens. Exploiting the image sectioning properties of these processes, one can create high-resolution 3D reconstructions that are invaluable for biomedical research. One limiting factor is the presence of specimen-induced aberrations. Because the index of refraction is not homogeneous within the specimen, aberrations affect both the resolution and the maximum depth of penetration [36]. Using adaptive optics [35], these detrimental

effects can be minimised by reducing the phase aberrations. A phase aberration can be introduced in the excitation beam by means of a deformable mirror, for example. Chosen correctly, such a phase aberration can suppress some amount of the specimen-induced aberrations.

Direct measurement of the specimen-induced aberrations is challenging. One solution involves measuring the aberrations of the excitation light that is back-scattered from the specimen [137, 133, 187, 131, 136, 199]. In this case, the difficulty arises in excluding the light that is reflected from the out-of-focus layers of the specimen [137, 133]. Additionally, these measurements are weakly sensitive to odd aberrations [137], due to the double pass effect [135]. In another solution, instead, the emission from a point source inside the specimen is used to perform Shack–Hartmann wavefront sensing [139, 140, 138, 142, 143]. Here the difficulty stems from the lack of such reference point sources within the specimen and from the limited number of photons available in the emission signal.

An alternative, indirect approach involves deducing the specimen-induced aberrations solely by examining the emission signal. This approach only requires the addition of a deformable mirror into the excitation path of an existing microscope. A solution that is based on the segmentation of the pupil has been proposed [166, 167]. Other solutions are based on the optimisation of an image quality metric, which attains its global maximum when the residual aberration is maximally suppressed. In practice, different trial aberrations are sequentially applied with the deformable mirror until the image quality metric attains its maximum.

General optimisation algorithms can be used to maximise the image quality metric [151, 111, 150, 113, 153, 146]. However, because these algorithms have no prior knowledge about the metric, a large number of trial aberrations must be evaluated before the metric is maximised [36, 155, 158]. Reducing the number of trial aberrations is critical in achieving short image acquisition times and in limiting side effects, such as photobleaching and phototoxicity. For small aberrations, the response of the image quality metric can be approximated using a simple model, such as a quadratic polynomial [197, 200, 155, 159, 128, 160, 161, 162, 144, 118, 158]. Model-based aberration correction algorithms [197, 200, 155, 159, 128, 160, 161, 162, 156, 157, 201, 144, 118, 158] exploit the knowledge about the model of the metric to quicken the aberration correction, thus curtailing the aforementioned side effects.

In this paper, we investigate applying a model-based wavefront sensorless aberration correction algorithm to a second-harmonic microscope.

The paper is organised as follows. In Section 3.2 we discuss the definition of the basis functions for the control of the deformable mirror. In Section 3.3 we outline our proposed algorithm for the aberration correction. In Section 3.4 we report the experimental results. The conclusions are drawn in Section 3.5.

3.2 Definition of the basis functions for the control of the deformable mirror

The basis functions should satisfy two different requirements. In scanning microscopy, the aberration correction should not introduce x-tilt, y-tilt and defocus Zernike aberra-

tions [44]. These aberrations do not affect the image quality. Instead, they affect the position of the focal point within the specimen and they induce translations or distortions in the acquired images [160, 161, 202]. For this reason, a first requirement is that the basis functions be orthogonal to the x-tilt, y-tilt and defocus aberrations.

A second requirement is that the basis functions express the capabilities of the deformable mirror in an accurate and concise form, by taking into account the mechanical limitations of the mirror and the misalignment in the optical system as much as possible. This requirement is not satisfied when using Zernike polynomials as the basis functions since a deformable mirror with N_a actuations cannot accurately induce a set of N_a Zernike polynomials.

We now discuss a simple procedure to define a new set of basis functions that satisfy the two requirements discussed so far. This procedure is based on the Singular Value Decomposition (SVD) of a matrix H that approximately describes a linear relationship between the control signals of the deformable mirror and a set of Zernike coefficients. For completeness, we first report how H can be computed from input–output measurements.

3.2.1 Computation of matrix H from input–output measurements

Let N_a be the number of actuators of the deformable mirror. Assuming that the deformable mirror is a linear device, the phase aberration $\Phi(\boldsymbol{\xi})$ is given by the superposition of the influence functions [203, 50] $\psi_i(\boldsymbol{\xi})$ of each actuator, where $\boldsymbol{\xi}$ is the spatial coordinate in the pupil and u_i is the control signal of the i-th actuator

$$\Phi(\boldsymbol{\xi}) = \sum_{i=1}^{N_a} u_i \psi_i(\boldsymbol{\xi}). \tag{3.1}$$

For a suitable number N_z of Zernike polynomials $\mathcal{Z}_j(\boldsymbol{\xi})$, the phase aberration is approximated by

$$\Phi(\boldsymbol{\xi}) \approx \sum_{j=2}^{1+N_z} z_j \mathcal{Z}_j(\boldsymbol{\xi}), \tag{3.2}$$

where z_j is the j-th Zernike coefficient. We neglect the piston mode $\mathcal{Z}_1$ since this does not affect the image and assume that $\Phi(\boldsymbol{\xi})$ and $\psi_i(\boldsymbol{\xi})$ have zero mean value over the pupil.

The coefficients u_i and z_j are collected respectively into vectors $\mathbf{u} \in \mathbb{R}^{N_a}$ and $\mathbf{z} \in \mathbb{R}^{N_z}$. By considering a grid defined in the pupil, N_c samples of $\Phi(\boldsymbol{\xi})$ are collected into a vector $\boldsymbol{\phi} \in \mathbb{R}^{N_c}$. Similarly, we evaluate $\psi_i(\boldsymbol{\xi})$ and $\mathcal{Z}_j(\boldsymbol{\xi})$ over the grid and define two matrices $\Psi \in \mathbb{R}^{N_c \times N_a}$ and $Z \in \mathbb{R}^{N_c \times N_z}$. Using Eqs. (3.1) and (3.2), we find $\boldsymbol{\phi} = \Psi \mathbf{u}$ and $\boldsymbol{\phi} \approx Z\mathbf{z}$.

We would like to recover a matrix H that maps an actuation vector $\mathbf{u}$ into the corresponding vector of Zernike coefficients $\mathbf{z}$, i.e., $\mathbf{z} \approx H\mathbf{u}$. H can be computed using input–output measurements, so that the misalignment in the optical system is accounted for. Using a Shack–Hartmann wavefront sensor or interferometric methods [189, 204, 50, 183] one can collect a set of measurements of the phase $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_D$ corresponding to different settings of the deformable mirror $\mathbf{u}_1, \ldots, \mathbf{u}_D$.

We compute H by minimising the following criterion,

$$\min_{H} \sum_{i=1}^{D} \|\boldsymbol{\phi}_i - ZH\mathbf{u}_i\|^2. \tag{3.3}$$

Setting the derivative with respect to H to zero leads to the following normal equation

$$Z^T ZH(\sum_{i=1}^{D} \mathbf{u}_i \mathbf{u}_i^T) = Z^T(\sum_{i=1}^{D} \boldsymbol{\phi}_i \mathbf{u}_i^T), \tag{3.4}$$

which can be solved by multiplying from the left and from the right by the inverse matrices of $Z^T Z$ and $\sum_{i=1}^{D} \mathbf{u}_i \mathbf{u}_i^T$. For a properly defined grid, the inverse of $Z^T Z$ exists, since Z is full column rank due to the orthogonality property of the Zernike polynomials. Additionally, vectors $\mathbf{u}_i$ can be selected so that $\sum_{i=1}^{D} \mathbf{u}_i \mathbf{u}_i^T$ is full rank.

In our system we have $N_a = 17$ and $N_c = 75912$. We performed $D = 4N_a$ measurements of the phase $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_{4N_a}$. In each measurement, a single actuator is poked while the other actuators are at rest. We empirically chose $D = 4N_a$, other choices are possible provided $D \geq N_a$. The choice of N_z is more critical. With a poor choice of N_z, the accuracy requirement discussed at the beginning of Section 3.2 may not be fulfilled and the approximation $\mathbf{z} \approx H\mathbf{u}$ may be too rough. We chose $N_z = 35$ by evaluating the error in approximating the phase measurements $\boldsymbol{\phi}_i$ using an increasing number of Zernike polynomials.

3.2.2 SVD-based removal of the x-tilt, y-tilt and defocus aberrations

From the previous section, we conclude that, in our system, the $N_a = 17$ influence functions approximately span a subspace of the space spanned by the first $N_z = 35$ Zernike polynomials. Because $\text{rank}(H) < N_z$, there exist non-zero vectors $\mathbf{z}$ that do not belong to the range of H and the Zernike polynomials should not be used as the basis functions for the control of the deformable mirror.

We can split vector $\mathbf{z}$ and matrix H so that $\mathbf{z} \approx H\mathbf{u}$ is partitioned as

$$\begin{bmatrix} \mathbf{z}_l \\ \mathbf{z}_h \end{bmatrix} \approx \begin{bmatrix} H_l \\ H_h \end{bmatrix} \mathbf{u}, \tag{3.5}$$

where the x-tilt, y-tilt and defocus coefficients are collected into $\mathbf{z}_l = [z_2, z_3, z_4]^T$. The SVD of H_l is

$$H_l = U_l \begin{bmatrix} \Sigma_l & \mathbf{0} \end{bmatrix} \begin{bmatrix} V_{l1}^T \\ V_{l2}^T \end{bmatrix}, \tag{3.6}$$

where Σ_l has dimensions 3×3, V_{l1} has dimensions $N_a \times 3$ and V_{l2} has dimensions $N_a \times N_p$ with $N_p = N_a - 3$. The required constraint that $\mathbf{z}_l \approx \mathbf{0}$ is enforced if we choose $\mathbf{u}$ such that $H_l\mathbf{u} = 0$. This is achieved by parametrising $\mathbf{u}$ using the columns of V_{l2}, i.e., letting

$\mathbf{u} = V_{l2}\mathbf{p}$, where $\mathbf{p} \in \mathbb{R}^{N_p}$. Therefore, the phase aberration is

$$\Phi(\xi) = \sum_{i=1}^{N_a} u_i \psi_i(\xi), \quad \text{s.t.} \quad \mathbf{u} = V_{l2}\mathbf{p} \tag{3.7}$$

or equivalently

$$\Phi(\xi) = \sum_{i=1}^{N_p} p_i \omega_i(\xi), \tag{3.8}$$

where $\omega_i(\xi)$ are the new basis functions. These functions are defined by

$$\omega_i(\xi) = \sum_{j=1}^{N_a} (V_{l2})_{ji} \psi_j(\xi),$$

where $(V_{l2})_{ji}$ denotes the element of V_{l2} at position (j, i). For a given vector $\mathbf{p}$, we can compute the control signals of the actuators with $\mathbf{u} = V_{l2}\mathbf{p}$. Similarly, for a given $\mathbf{p}$, the Zernike analysis of the induced wavefront aberration is given by $\mathbf{z} \approx HV_{l2}\mathbf{p}$.

In our experiments, we also applied regularisation [205] by truncating the SVD of HV_{l2} to $U_1\Sigma_1 V_1^T$. Using no more than 80% of the sum of the singular values, Σ_1 was a 7×7 matrix and the deformable mirror was controlled with a vector $\mathbf{r} \in \mathbb{R}^N$, where $N = 7$. For a given $\mathbf{r}$, the control signals of the actuators of the deformable mirror are computed using $\mathbf{u} = V_{l2}V_1\mathbf{r}$. The Zernike analysis of the induced wavefront aberration is computed using $\mathbf{z}_h \approx H_h V_{l2} V_1 \mathbf{r}$ and the rms of the phase profile is given by computing the 2-norm, i.e., $\|\mathbf{z}_h\|$. This is equivalent to applying another parametrisation to Eq. (3.8). We remark that in this way, no pseudo-inverse is ever computed or used to control the deformable mirror, differently from what is done in [203], for example.

3.3 Least-squares estimation of the unknown aberration

In this section we discuss the aberration correction algorithm. In [161, 144, 118], the authors show that, for small aberrations, the image quality metric can be modelled using a quadratic polynomial. We denote a measurement of the image quality metric at time instant k with $\tilde{y}_k$, so that

$$\tilde{y}_k = c_0 - (\mathbf{x} + \mathbf{r}_k)^T Q(\mathbf{x} + \mathbf{r}_k) + \epsilon_k, \tag{3.9}$$

where c_0 and Q are the parameters of the quadratic polynomial. Matrix Q is a positive semidefinite matrix, i.e., $Q \succeq 0$ [158]. Vector $\mathbf{x}$ represents the unknown aberration whereas $\mathbf{r}_k$ accounts for the aberration induced by the deformable mirror. The term ϵ_k is a placeholder that collects both the uncertainty in modelling the image quality metric and the measurement noise, and as such it cannot be measured by definition. By including this term, a measurement $\tilde{y}_k$ can be set equal to the right-hand side of Eq. (3.9). Excluding ϵ_k, the right-hand side of Eq. (3.9) is referred to as the approximate image quality metric in [158]. The parameters c_0 and Q can be computed using the input–output

measurements recorded in a calibration experiment described in Section 3.4.3 or using the methods described in [160, 169].

3.3.1 Definition of the least-squares problem

The aberration correction is achieved by maximising the image quality metric, i.e., by letting $\mathbf{r}_k = -\mathbf{x}$ in Eq. (3.9). For this reason, we must first estimate the unknown vector $\mathbf{x}$. This can be done by applying $m \geq N + 1$ trial aberrations $\mathbf{r}_1, \ldots, \mathbf{r}_m$ with the deformable mirror and by taking the corresponding measurements $\tilde{y}_1, \ldots, \tilde{y}_m$.

Collect $\epsilon_1, \ldots, \epsilon_m$ into a vector $\boldsymbol{\epsilon}$ and $\tilde{y}_1, \ldots, \tilde{y}_m$ into a vector $\tilde{\mathbf{y}}$. By stacking m instances of Eq. (3.9), we can define a vector-valued function $\mathbf{g}$,

$$\mathbf{g}(\mathbf{x}) = \begin{bmatrix} c_0 - (\mathbf{x} + \mathbf{r}_1)^T Q(\mathbf{x} + \mathbf{r}_1) \\ \vdots \\ c_0 - (\mathbf{x} + \mathbf{r}_m)^T Q(\mathbf{x} + \mathbf{r}_m) \end{bmatrix}, \tag{3.10}$$

such that

$$\tilde{\mathbf{y}} = \mathbf{g}(\mathbf{x}) + \boldsymbol{\epsilon}. \tag{3.11}$$

The least-squares estimate of $\mathbf{x}$ is obtained by minimising $\|\boldsymbol{\epsilon}\|^2$, i.e., by solving

$$\min_{\mathbf{x}} f(\mathbf{x}), \tag{3.12}$$

where

$$f(\mathbf{x}) = \|\tilde{\mathbf{y}} - \mathbf{g}(\mathbf{x})\|^2. \tag{3.13}$$

We note that if vector $\boldsymbol{\epsilon}$ follows a multivariate normal distribution with zero mean and covariance proportional to the identity matrix, solving Eq. (3.12) provides the maximum likelihood [206] estimate of $\mathbf{x}$.

3.3.2 Analysis of the least-squares problem

Finding the global minimum of $f(\mathbf{x})$ appears to be non-trivial as $f(\mathbf{x})$ may be non-convex. This is illustrated with a two-dimensional example in Fig. 3.1. Here, the contour plot of Eq. (3.13) is shown, when $m > N + 1$ measurements of $\tilde{y}$ are taken. The measurement noise is zero, i.e., $\epsilon = \mathbf{0}$. Nevertheless, $f(\mathbf{x})$ is not convex and exhibits two critical points. In addition to the least-squares solution $\mathbf{x}^{\text{ls}}$ of Eq. (3.12), which is the global minimum and for which $f(\mathbf{x}^{\text{ls}}) = 0$, a local minimum $\mathbf{x}^{\text{loc}}$ is present. In case one uses $\mathbf{r}_k = -\mathbf{x}^{\text{loc}}$ to perform the aberration correction, then the residual aberration is not zero and the image quality metric is not maximised.

Because the convexity property is not satisfied in general, it is unclear how $\mathbf{x}^{\text{ls}}$ can be computed. For example, a gradient based method applied to solve Eq. (3.12) may fail to compute $\mathbf{x}^{\text{ls}}$. Alternatively, more sophisticated algorithms may be unsuitable to meet the requirements of a real time implementation. Nevertheless, the global solution of Eq. (3.12) can be computed efficiently even when $f(\mathbf{x})$ is not convex, as is outlined in the following section.

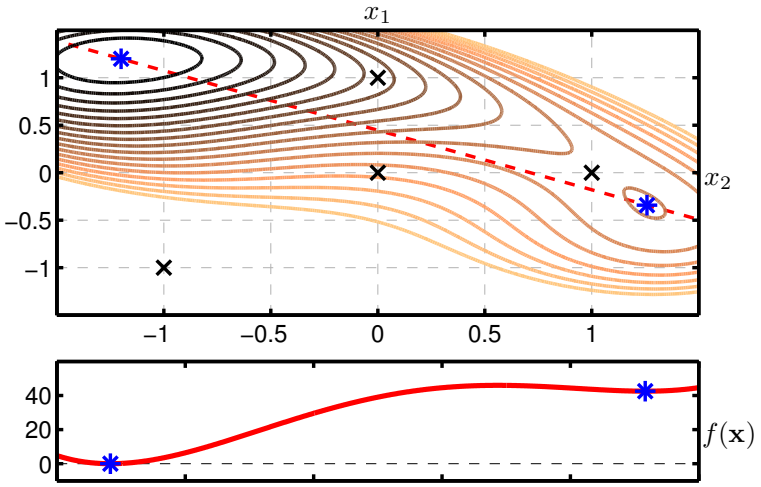

Figure 3.1: Contour plot of Eq. (3.13). In this example, $f(\mathbf{x})$ is not convex and exhibits a local minimum. The parameters are $c_0 = 100$ and $Q = [1.25, 0.433; 0.433, 1.25]$. Four measurements of $\tilde{y}$, taken at $\mathbf{r}_1 = [0,0]^T$, $\mathbf{r}_2 = [1,0]^T$, $\mathbf{r}_3 = [0,1]^T$ and $\mathbf{r}_4 = [0,-1]^T$ are marked with $\times$ symbols. The global minimum $\mathbf{x}^{\text{ls}} = [-1.2, 1.2]^T$ and the local minimum $\mathbf{x}^{\text{loc}} \approx [1.2582, -0.3421]^T$ are indicated with $*$ symbols. Isolines with an elevation greater than 70 have been removed for clarity. A cross section along the dashed line is reported in the plot in the bottom.

3.3.3 Efficient computation of $\mathbf{x}^{\mathbf{ls}}$

In [207], an efficient algorithm to find the global solution of a possibly non-convex optimisation similar to Eq. (3.12) was developed, in the context of localisation problems. In this section we show how the solution proposed in [207] can be applied to our problem. For the remaining part of the paper we assume that Q is strictly positive definite. This assumption is reasonable since if there are aberrations that do not affect the image quality metric then these cannot be corrected and they should be neglected during the aberration correction [158].

Introducing an additional scalar variable α, we can reformulate Eq. (3.12) into the following equivalent constrained optimisation

$$\begin{aligned} \min_{\mathbf{x},\alpha} \sum_{k=1}^{m} \left(-\alpha - 2\mathbf{r}_k^T Q\mathbf{x} + c_0 - \mathbf{r}_k^T Q\mathbf{r}_k - \tilde{y}_k\right)^2 \\ \text{s.t.} \quad \alpha = \mathbf{x}^T Q\mathbf{x}. \end{aligned} \tag{3.14}$$

Problem (3.14) is written concisely in matrix form as

$$\min_{\mathbf{w}} \|A\mathbf{w} - \mathbf{b}\|^2 \quad \text{s.t.} \quad \mathbf{w}^T D\mathbf{w} + 2\mathbf{f}^T\mathbf{w} = 0, \tag{3.15}$$

where

$$\begin{aligned} \mathbf{w}^T = \begin{bmatrix}\mathbf{x}^T & \alpha\end{bmatrix}, R = \begin{bmatrix}\mathbf{r}_1 & \dots & \mathbf{r}_m\end{bmatrix}, \\ A = \begin{bmatrix}-2R^T Q & -\mathbf{1}\end{bmatrix}, \mathbf{b} = \begin{bmatrix}\mathbf{r}_1^T Q\mathbf{r}_1 + \tilde{y}_1 - c_0 \\ \vdots \\ \mathbf{r}_m^T Q\mathbf{r}_m + \tilde{y}_m - c_0\end{bmatrix}, \\ D = \begin{bmatrix}Q & \mathbf{0} \\ \mathbf{0} & 0\end{bmatrix}, \mathbf{f}^T = \begin{bmatrix}\mathbf{0} & -1/2\end{bmatrix}^T \end{aligned} \tag{3.16}$$

and $\mathbf{1}$ and $\mathbf{0}$ denote vectors of appropriate dimensions where all components are respectively ones and zeros. The authors in [207] note that Eq. (3.15) is a Generalised Trust Region Subproblem [208]. Such problems, although non-convex in general, have necessary and sufficient optimality conditions [208]. In particular, from [207, 208], we know that $\mathbf{w}$ is a global minimiser of Eq. (3.15) if and only if there exists a Lagrange multiplier ν such that

$$\begin{aligned} (A^T A + \nu D)\mathbf{w} = A^T\mathbf{b} - \nu\mathbf{f} \\ \mathbf{w}^T D\mathbf{w} + 2\mathbf{f}^T\mathbf{w} = 0 \\ A^T A + \nu D \succeq \mathbf{0}. \end{aligned} \tag{3.17}$$

We assume that matrix A is full column rank, which in turn implies that $m \geq N + 1$. This assumption on A is by no means restrictive. Because $Q \succ 0$, it can be factored $Q = V\Delta V^T$, where Δ is diagonal and full rank. Choose $R = [V, \mathbf{0}]$, where $\mathbf{0}$ is a vector of zeros, then A is full column rank. We further assume that the optimal Lagrange multiplier ν^* is such that $A^T A + \nu^* D$ is strictly positive definite. The authors in [207] point out that this more restrictive assumption could be removed with a more refined analysis. However the case where ν^* is such that $A^T A + \nu^* D$ is not strictly positive definite is unlikely to occur both

in theory and in practice [207].

Under these assumptions, one can compute

$$\mathbf{w}(\nu) = (A^T A + \nu D)^{-1}(A^T \mathbf{b} - \nu \mathbf{f}) \tag{3.18}$$

for a fixed value of ν. By replacing $\mathbf{w}$ in the second equation in Eq. (3.17) with the right-hand side of Eq. (3.18), one finds a univariate rational polynomial equation in ν:

$$\mathbf{w}(\nu)^T D\mathbf{w}(\nu) + 2\mathbf{f}^T \mathbf{w}(\nu) = 0. \tag{3.19}$$

The optimal Lagrange multiplier ν^* can be found examining the solutions of Eq. (3.19). From the assumption $A^T A + \nu D \succ 0$, it can be derived [208, 207] that ν must be in the interval $(I_l, +\infty)$, where $I_l = -1/\lambda_{\max}(\Delta_A^{-1/2} V_A^T D V_A \Delta_A^{-1/2})$, and we used the factorisation $A^T A = V_A \Delta_A V_A^T$. In addition, it is known [208, 207] that Eq. (3.19) is strictly decreasing in ν within the considered interval. Therefore the desired root ν^* of Eq. (3.19) can be found efficiently, for example via a bisection algorithm [207]. Once ν^* is found, the estimate of $\mathbf{x}$ is extracted from the first N components of $\mathbf{w}(\nu^*)$.

The aberration correction algorithm is therefore applied in the following manner. First the data collection step takes place, whereby the $m \geq N + 1$ trial aberrations are applied and the corresponding measurements $\tilde{y}_1, \ldots, \tilde{y}_m$ are taken. Then, ν^* is computed by finding the root of Eq. (3.19) within $(I_l, +\infty)$. The estimate $\mathbf{x}^{\text{ls}}$ of the aberration is found in the first N components of $\mathbf{w}(\nu^*)$. The second step involves applying the aberration correction with the deformable mirror, by letting $\mathbf{r} = -\mathbf{x}^{\text{ls}}$. These steps can be repeated in the following time instants by including more than m measurements to achieve a refined correction. We note that in [158], the least-squares estimate of $\mathbf{x}$ was not computed since the quadratic constraint in Eq. (3.14) was neglected to obtain a linear least-squares problem.

3.4 Experimental results

We implemented the model-based wavefront sensorless algorithm and report the experimental results in this section. Following [162], we employ the mean image intensity as a metric to correct aberrations in our second-harmonic microscope. Our experiments show successful aberration correction using this metric (see Section 3.4.7).

Our first purpose is to validate a previously proposed method [158] to compute the parameters c_0 and Q of Eq. (3.9) using input–output measurements. This validation has not been previously done in a realistic setting, since in [158] no microscope and no specimen were involved. Additionally, we intend to validate the aberration correction algorithm described in Subsection 3.3.3. We report our results in the following sections.

3.4.1 Description of the experimental setup

A schematic of the experimental setup is shown in Fig. 3.2. The source is a Coherent Chameleon Ultra II Ti:Sa 140 fs pulsed, near-infrared laser, with a beam diameter of 1.2 mm. This beam is expanded to a 14 mm wide beam by lenses L1 and L2.

The beam is stopped down to 9.5 mm (AP) before it is reflected under an angle of about 10° by the deformable mirror (DM; Okotech, 17-channel micromachined deformable mirror with tip-tilt stage). The deformable mirror is reimaged one-to-one onto the objective back-aperture by lenses L3 and L4 (focal lengths, 300 mm). Because the deformable mirror can only introduce negative deflection, we bias the mirror so that we can apply positive and negative deflections to correct the wavefront (see [50]). In addition, the relationship between the control signal u_i of each actuator of the deformable mirror and the voltage applied to the corresponding electrode is quadratic, so that a linear displacement of the membrane is expected [50]. Due to this bias, the collimated beam coming from L2 is converging after being reflected by the deformable mirror. We corrected this by using lenses L4 and L3, so that a collimated beam is fed into the objective.

The sample is mounted on an xyz-piezo stage (XYZ; PI, Nanocube P-611.3S). The second-harmonic signal from the specimen is collected by the objective and split off by a 705 nm cut-off dichroic beam splitter (DBS; Semrock, FF705-Di01-25x36). This light is focused onto a 600 μm multi-mode fibre which is connected to a photomultiplier tube (PMT; Hamamatsu, GaAsP photocathode H7422P-40). The objective used (OBJ) is a 40×/0.9 NA Nikon air objective with spherical correction collar. We manually adjusted the collar to correct for the spherical aberration due to the cover glass and the specimen at the selected depth.

For characterisation of the deformable mirror, we interfere a tilted reference beam with a sample beam deflected off the deformable mirror to create fringes that encode the wavefront deformation. To allow this, a 50/50 beam splitter (BS1; Thorlabs, BS016) splits off part of the light into a reference arm beam, which is relayed onto the camera (CCD; AVT, Guppy Pro F-033b) by mirrors M3, M4, and M5. The sample arm beam is deflected by the deformable mirror once before flip mirror FM1 directs the light into the calibration arm. Lenses L6 and L7 reimage the deformable mirror onto a camera. For this calibration we use the alignment laser, which is a continuous wave. We used the method described in [133] to decode the wavefront from the fringe patterns.

The piezo stage is controlled with a data acquisition board (National Instruments, PCI-e 6259) on a Windows computer running LabView. The deformable mirror is controlled through a PCI DAC card on a Linux computer running MATLAB and custom C code.

In the aberration correction experiments we imaged collagen fibre extracted from rat tail washed four times in distilled water. Following fixation in 4% paraformaldehyde, the fibre was washed in phosphate buffer saline and then embedded in 3% agarose (Sigma-Aldrich chemie GmbH) in a 35 mm glass bottom dish (MatTek Corporation). We used 900 nm excitation light to generate the second-harmonic signal.

3.4.2 Preparation of the experiments

We first imaged a 20 μm×20 μm region, approximately 33 μm deep into the collagen fibre. The region is labelled with A in Fig. 3.3. The influence of the size of the region used for the aberration correction has been studied elsewhere [118]. For a certain setting of the deformable mirror $\mathbf{r}$, the region is scanned using the xyz stage. The corresponding value of the image quality metric $\tilde{y}$ is measured as the mean image intensity [162, 161, 144, 118], i.e., the mean pixel value recorded over the region. The pixel dwell time is 0.5 ms and the

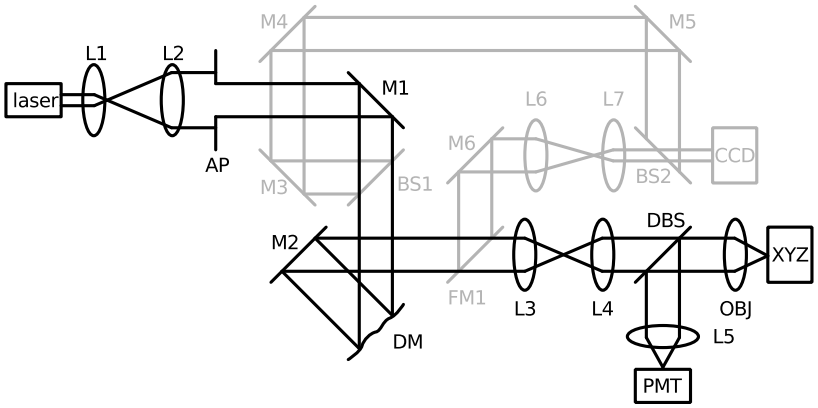

Figure 3.2: Illustration of the optical setup. The components in black are used throughout the aberration correction experiments. The components in grey are used only for the initial characterisation of the deformable mirror (DM). A pulsed laser beam is expanded with lenses L1 and L2, clipped by aperture AP and reflected by flat mirror M1 onto the DM. The DM is in an image of the back aperture of the microscope objective (OBJ), using lenses L3 and L4. The DM is illuminated under an angle of about 10° using the flat mirrors M1 and M2. The microscope objective (OBJ) focuses the light onto the specimen, which is supported by an xyz stage (XYZ). The second-harmonic signal emitted from the focal point inside the specimen is collected with the objective and separated from the illumination beam using a dichroic beam splitter (DBS). The emitted signal is focused by lens L5 onto a photomultiplier tube (PMT). For characterising the DM, the surface of the DM is reimaged onto a CCD camera (CCD) using the flip mirror FM1, flat mirror M6 and lenses L6 and L7. A reference arm is created using beam splitter BS1, flat mirrors M3, M4, M5 and beam splitter BS2. A coherence-gated fringe analysis method described elsewhere [133] is applied to the fringe pattern generated onto CCD.

sampling is 24 pixels × 24 pixels. With these settings, the xyz stage does not reach the full 20 μm distance in the x scanning direction, which is the fast axis. This was not an issue since such a coarse sampling was only used to perform the aberration correction in a short time [144]. The final images taken after the aberration correction were recorded with a higher sampling. The image deformation due to both the non-linearity and non-uniform speed of the xyz stage were removed from the final images, using interpolation and the signals recorded with the position sensors of the xyz stage.

First, the static aberrations in the system due to misalignment and imperfections in the optical components were corrected. We used the non-zero initial aberration that was found during the calibration of the deformable mirror in Section 3.2 (about 0.79 rms rad at 900 nm, mostly astigmatism). At this point the Nelder–Mead algorithm [152] was executed four times to find a value $\mathbf{r}$ that maximises $\tilde{y}$. Unfortunately, this led to the saturation of two actuators, indicating that the stroke of the deformable mirror may be insufficient to completely suppress the aberration in this region. We selected a slightly sub-optimal vector $\mathbf{r}$ from the vectors generated by the Nelder–Mead algorithm. For the selected vector, the maximum normalised voltage of the actuators was 0.72, i.e., $\|\mathbf{u}\|_\infty \leq 0.72$, $\tilde{y}$ improved by 3% and a total aberration of about 0.18 rms rad was suppressed. This state was used as the new initial condition for the rest of the experiments, i.e., $\mathbf{r} = \mathbf{0}$ is mapped to this setting of the deformable mirror. In the following sections, all the units in rad are referenced to the 900 nm excitation laser light.

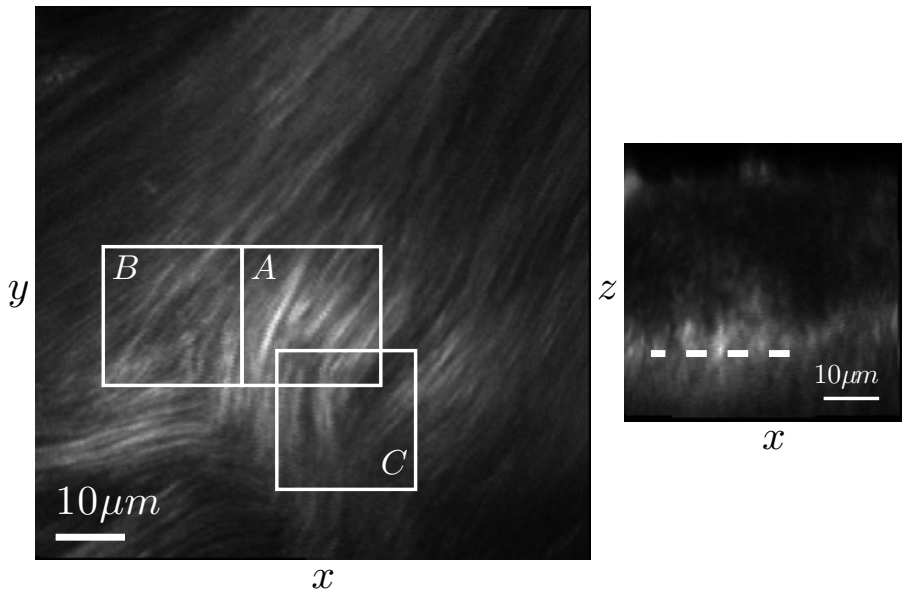

Figure 3.3: Cross sections of rat tail collagen fibre used in our experiments. The smaller image on the right-hand side is an xz cross section [50 μm × 50 μm, 128 pixels × 128 pixels]. The dashed line denotes an xy cross section [80 μm × 80 μm, 256 pixels × 256 pixels] approximately 33 μm deep, which is shown on the left-hand side. Three different 20 μm × 20 μm regions are marked.

3.4.3 Computation of the parameters of the quadratic polynomial using input–output measurements

We executed the computation of the parameters of the quadratic polynomial used for modelling the image quality metric multiple times. Each time, the sequence of input vectors consisted of two sub-sequences. The first sub-sequence contained random vectors $\mathbf{r}_1, \ldots, \mathbf{r}_{250}$ and was used for the validation and the cross-validation. The second sub-sequence contained 70 fixed vectors (each vector having a single non-zero component). The second sub-sequence was included because the 250 random vectors may be insufficient to uniformly sample the N-dimensional space of the inputs. In [158], 3750 random vectors were used, but this was impractical here, due to the time necessary to move the xyz stage. The maximum rms rad of the input aberrations did not exceed 0.81. This value was empirically tuned by examining the goodness of fit as a function of the maximum rms [158].

For each input vector in the sequence, we measured the corresponding output of $\tilde{y}$. The resulting input–output data, i.e., collections $\tilde{y}_1, \ldots, \tilde{y}_{320}$ and $\mathbf{r}_1, \ldots, \mathbf{r}_{320}$ were used to formulate the following optimisation problem [158]

$$\begin{aligned} &\min_{\tilde{\mathbf{x}}} \|\tilde{A}\tilde{\mathbf{x}} - \tilde{b}\| \quad \text{s.t.} \\ &\tilde{A} = \begin{bmatrix} 1 & \mathbf{r}_1^T & -\mathbf{r}_1^T \otimes \mathbf{r}_1^T \\ \vdots & \vdots & \vdots \\ 1 & \mathbf{r}_{320}^T & -\mathbf{r}_{320}^T \otimes \mathbf{r}_{320}^T \end{bmatrix}, \\ &\tilde{b} = \begin{bmatrix} \tilde{y}_1 & \ldots & \tilde{y}_{320} \end{bmatrix}^T, \\ &\tilde{\mathbf{x}} = \begin{bmatrix} c_0 & \mathbf{c}_1^T & \mathrm{vec}(Q)^T \end{bmatrix}^T, \\ &\quad Q \succeq 0, \end{aligned} \tag{3.20}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorisation operation and $\otimes$ the Kronecker product. This programme was solved using cvxopt [209] (see [210] for further details).

3.4.4 Validation and cross-validation of the computed parameters

The results of applying Eq. (3.20) in region A marked in Fig. 3.3 are shown in Fig. 3.4. The computation of the parameters was repeated six times. Each time, a new input–output data set was acquired, $\mathcal{D}_1, \ldots, \mathcal{D}_6$. For each input–output data set, optimisation (3.20) was solved generating six sets of parameters, each set comprising c_0, $\mathbf{c}_1$ and Q. The sets are denoted as $\mathcal{M}_1, \ldots, \mathcal{M}_6$. We quantified the goodness of fit for all combinations of $\mathcal{D}$ and $\mathcal{M}$ by means of the $\mathcal{R}^2$ indicator. Using the random input sub-sequence of $\mathcal{D}_i$ and $\mathcal{M}_j$, we computed the predicted output $\hat{\mathbf{o}} \in \mathbb{R}^{250}$ of the image quality metric. The input–output data points obtained from the deterministic input sub-sequence of 70 vectors were discarded and were not included in the computation of the $\mathcal{R}^2$ indicator, which is obtained using the following equations:

$$\begin{aligned} &\mathcal{R}^2 = 1 - S_r/S_t, \qquad S_r = \|\mathbf{o} - \hat{\mathbf{o}}\|^2, \\ &S_t = \|\mathbf{o} - \bar{o}\mathbf{1}\|^2, \quad \bar{o} = (1/250)\mathbf{1}^T\mathbf{o}, \end{aligned} \tag{3.21}$$

where $\mathbf{o}$ is the measured output of $\mathcal{D}_i$. An $\mathcal{R}^2 = 1$ implies a perfect fit of the experimental data.

Fig. 3.4 reports $\mathcal{R}^2$ indicators that are close to one, implying a good fit of the experimental data. A good fit is also found for the combinations that are off the main diagonal. Here the parameters c_0, $\mathbf{c}_1$ and Q allow to accurately predict cross-validation output data. Similar results were found for the other two regions marked in Fig. 3.3.

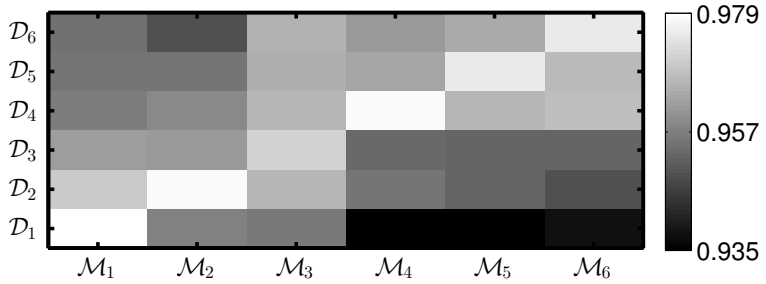

Figure 3.4: Validations and cross-validations of the computation of c_0, $\mathbf{c}_1$ and Q using Eq. (3.20). The computation has been performed six times in region A in Fig. 3.3. $\mathcal{D}_i$ denotes the input–output data taken during the i-th time. $\mathcal{M}_i$ denotes the set of parameters [c_0, $\mathbf{c}_1$ and Q in Eq. (3.20)] computed from $\mathcal{D}_i$. For each combination $\mathcal{M}_i$ and $\mathcal{D}_j$, the i-th random input sub-sequence and $\mathcal{M}_j$ are used to compute the predicted output $\hat{\mathbf{o}} \in \mathbb{R}^{250}$. Each rectangle reports the goodness of fit [$\mathcal{R}^2$, see Eq. (3.21)] computed comparing $\hat{\mathbf{o}}$ with the corresponding measured output $\mathbf{o} \in \mathbb{R}^{250}$ of $\mathcal{D}_j$. A value of one for the goodness of fit indicates that the model fits the data without error. High values of the goodness of fit are reported in all combinations showing that Eq. (3.20) is a robust method to compute the parameters.

3.4.5 Correction of the residual aberration

In this section we apply the aberration correction algorithm described in Section 3.3.3. First, we attempt to further reduce the residual aberration in region A, which is marked in Fig. 3.3. Some aberration may not have been completely suppressed by the Nelder–Mead algorithm, which was applied to region A in Subsection 3.4.2. We therefore expect no improvement or a small improvement in region A. Second, we apply the aberration correction to regions B and C, where the Nelder–Mead algorithm was not applied. Here we expect some improvement, provided that the aberrations found in regions B and C are different from the aberration found in region A.

We take the parameters c_0, $\mathbf{c}_1$ and Q that were computed using $\mathcal{D}_6$ in Subsection 3.4.4. In order to apply the algorithm, the following modified parameters need to be used, i.e., $c_0' = c_0 + (1/4)\mathbf{c}_1^T Q^{-1} \mathbf{c}_1$, $\mathbf{c}_1' = \mathbf{0}$ and $Q' = Q$. This modification is necessary since, for simplicity, in Section 3.3 we neglected the linear term $\mathbf{c}_1$. This term corresponds to the aberration that is present when computing the parameters of the quadratic polynomial, see [158] for further details. The aberration correction experiment is applied in the three

regions using the corresponding parameters for each region. A summary of the results is given in Fig. 3.5.

In Fig. 3.5(a), the normalised measurements of the image quality metric are reported for region A (curve with $\circ$ markers), region B (curve with $\square$ markers) and region C (curve with $\star$ markers). The measurements are normalised using the corresponding maximum recorded measurement $\tilde{y}_{\max}$ in each region. The initial value of $\tilde{y}$ is reported at sample time $k = 0$. This measurement is not supplied to the aberration correction algorithm. The data collection step is performed between time $k = 1$ and $k = 8$ inclusive, where $N+1 = 8$ trial aberrations are applied. From time $k = 9$ onwards, the aberration correction step is applied.

As expected, a marginal improvement is found in region A (curve with $\circ$ markers), where an aberration of about 0.38 rms rad is corrected. The rms of each aberration is estimated using $\|H_h V_{l2} V_1 \mathbf{x}^{\text{ls}}\|$, adjusted for the 900 nm excitation light. Also in region B (curve with $\square$ markers), a small aberration of about 0.37 rms rad is corrected. In region C (curve with $\star$ markers), an estimated 1.27 rms rad aberration is corrected, leading to an improvement of 20% of the image quality metric. Nevertheless, two actuators of the deformable mirror are saturated.

Two 256 pixels $\times$ 256 pixels images of region C are reported in Fig. 3.5(b) and Fig. 3.5(c). These images are recorded before [$k = 0$, Fig. 3.5(b)] and after [$k = 24$, Fig. 3.5(c)] the aberration correction. The cross sections marked in the images are reported in a single graph in Fig. 3.5(d). The image taken at time $k = 24$ is 18% brighter and shows finer detail in the bottom and right parts. Here some structure of the fibre was not visible at time $k = 0$. The improvement is less clear when examining the left and top parts of the region. One possible reason for the variability of the improvement is that the aberration is not spatially invariant over the considered region. We also note that the applied correction was not optimal, due to the saturation of two actuators of the deformable mirror. We conclude by observing that this improvement after the aberration correction is compatible with what was achieved by running four iterations of the Nelder–Mead algorithm in Section 3.4.2.

3.4.6 Validation of the aberration correction algorithm

To assess whether the aberration correction algorithm is effectively removing aberration we performed a different kind of experiment. First we introduce a known amount of aberration using the deformable mirror. We then apply the aberration correction algorithm to suppress this aberration. The algorithm is not supplied with any information about the known aberration. Finally we evaluate the residual aberration by comparing the estimate of the aberration provided by the algorithm with the known aberration. This experiment is commonly employed in the literature to assess the effectiveness of the aberration correction [159, 144, 118, 158].

Fig. 3.6 reports a summary of the correction of 20 random aberrations introduced with the deformable mirror in region A. The upper plot in Fig. 3.6 shows some statistical indicators of the normalised measurements of the image quality metric. The measurements have been normalised using the maximum measurement of the metric $\tilde{y}_{\max}$ that is recorded throughout the 20 experiments. The median, 25th and 75th percentiles are computed in

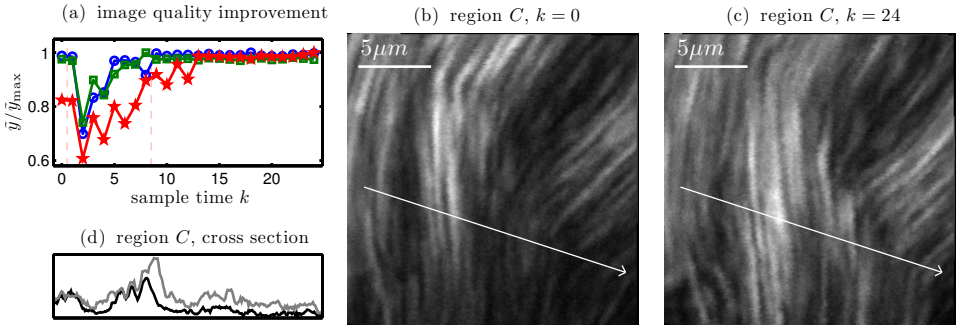

Figure 3.5: Summary of three aberration correction experiments. (a) evolution of the normalised image quality metric. The experiments were performed in region A (curve with $\circ$ markers), B (curve with $\square$ markers) and C (curve with $\star$ markers), which are marked in Fig. 3.3. For each region, the corresponding parameters computed by solving Eq. (3.20) were used. $\tilde{y}_{\max}$ is the maximum measurement of $\tilde{y}$ in each region. The estimated rms rad of each aberration is 0.38 for region A, 0.37 for region B and 1.27 for region C. (b) 256 pixels $\times$ 256 pixels image of region C at sample time $k = 0$. (c) 256 pixels $\times$ 256 pixels image of region C at sample time $k = 24$. (d) cross sections taken along the arrows marked in (b) and (c), black for (b) and grey for (c).

each time instant, see the caption of Fig. 3.6 for a detailed legend. The same analysis has been made for the residual aberration and is reported in the lower plot in Fig. 3.6. The rms of the residual aberration is computed as the rms of the difference between the known aberration introduced by the deformable mirror and the respective estimate provided by the algorithm. From this figure, we conclude that the image quality metric is consistently maximised, as the median is close to 1 after the aberration correction is applied from sample time $k = 9$ onwards. This is consistent with the reduction in the residual aberration reported in the lower plot.

The same experiments were performed in region B and C. In both cases we used the corresponding modified parameters, computed using $\mathcal{D}_6$ in Subsection 3.4.4. The results are reported in Figs. 3.7 and 3.8. Whereas the results for region C are similar to the results obtained in region A, the results in region B do not show a good performance, since the medians of the residual aberration are comparable with the initial aberration before the correction.

Out of the 20 trials in region B, we report respectively the ones that resulted in the maximum and in the minimum improvement of $\tilde{y}$ in Fig. 3.9. In Fig. 3.9(a), some fine structure of the fibre is more visible after the correction, which is compatible with a successful aberration correction. On the other hand, in Fig. 3.9(b), the aberration correction failed, as both the image after the correction is visually worse and the intensity is slightly decreased.

The experiments resulting in the maximum and in the minimum improvement of $\tilde{y}$ in region C are also reported in Fig 3.10. In Fig 3.10(a), a successful aberration correction is shown, with a clear maximisation of $\tilde{y}$ and a noticeable improvement in the contrast of the image after the correction. In Fig 3.10(b), the improvement is more marginal.

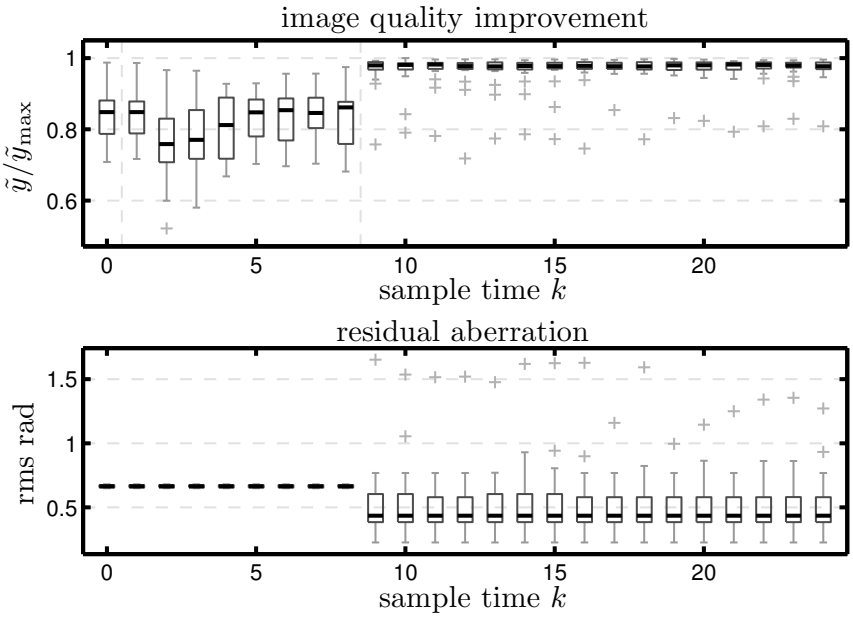

Figure 3.6: Summary of the correction of 20 random aberrations induced by the deformable mirror in region A, which is marked in Fig. 3.3. The upper plot reports the normalised measurements of the image quality metric. The measurements are normalised using the maximum measurement $\tilde{y}_{\max}$ that is recorded throughout the 20 experiments. At time $k = 0$ the initial value of $\tilde{y}$ is reported, this data point is not supplied to the aberration correction algorithm. Between time $k = 1$ and $k = 8$, the data collection step is executed. From time $k = 9$ onwards, the aberration correction step is applied. A statistical analysis is made at each time instant using the function boxplot from MATLAB. The tops and bottoms of the rectangles denote the 25th and 75th percentiles, the horizontal lines in the middle of the rectangles denote the medians, the whiskers extend to the furthest measurements not considered as outliers. The + symbols denote single outliers. The same statistical analysis is performed for the residual aberration and the results are shown in the lower plot.

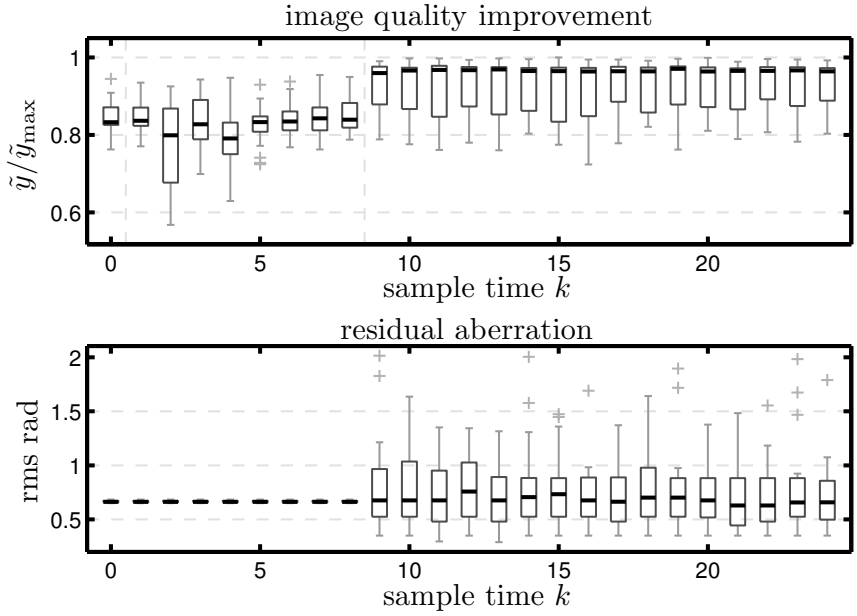

Figure 3.7: Summary of the correction of 20 random aberrations induced by the deformable mirror in region *B*. See the caption of Fig. 3.6 for a legend of the plots.

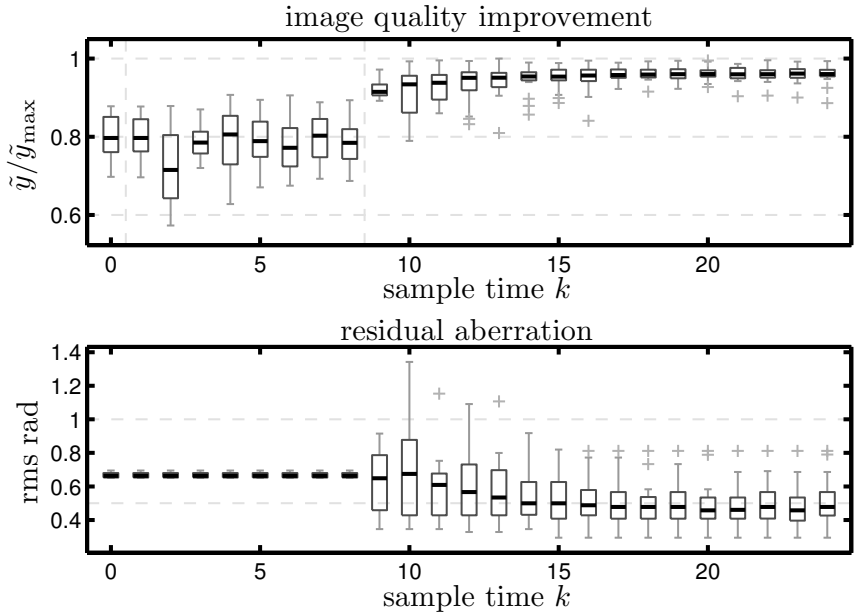

Figure 3.8: Summary of the correction of 20 random aberrations induced by the deformable mirror in region *C*. See the caption of Fig. 3.6 for a legend of the plots.

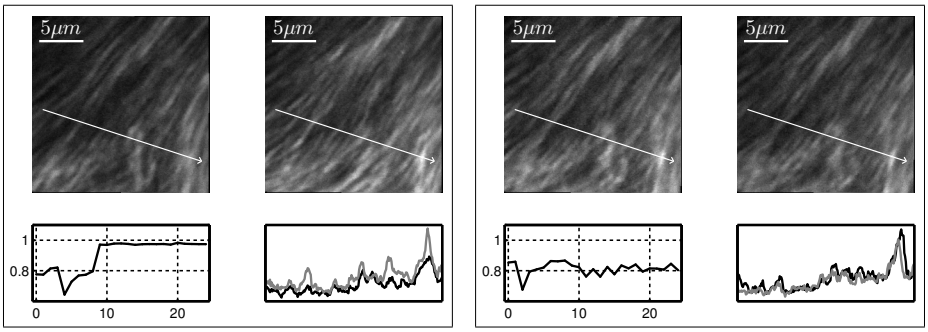

Figure 3.9: Two aberration correction experiments from the set of experiments reported in Fig. 3.7. These two experiments resulted, respectively, in the maximum (a) and the minimum (b) improvement of $\tilde{y}$. Both in (a) and (b), a 256 pixels × 256 pixels image is taken before [on the left, $k = 0$] and after [on the right, $k = 24$] the aberration correction. The graphs in the bottom of (a) and (b) show respectively the evolution of the normalised metric [on the left] and the cross sections indicated by the arrows in the images [on the right]. In the cross section graphs, the dark and the light lines correspond respectively to $k = 0$ and $k = 24$.

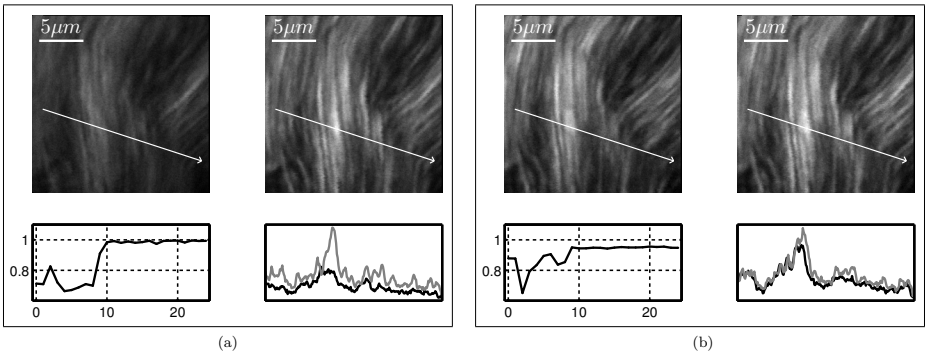

Figure 3.10: Two aberration correction experiments from the set of experiments reported in Fig. 3.8. These two experiments resulted, respectively, in the maximum (a) and the minimum (b) improvement of $\tilde{y}$. Refer to the caption of Fig. 3.9 for a detailed legend.

3.4.7 Analysis of the experimental results

We computed the correlation among different quantities to concisely assess the results of the 60 aberration correction experiments reported in Figs. 3.6, 3.7 and 3.8. Considering the last time instant $k = 24$, we set up a saturation indicator variable $s_1, \ldots, s_{60}$ that is 1 if saturation of some actuators of the deformable mirror occurred and -1 otherwise.

We computed a correlation of -0.2477 between the normalised measurement of the image quality metric and the saturation indicator variable. These two quantities are slightly inversely correlated, meaning that saturation of some actuators negatively affected the final value of the image quality metric. The correlation between the final rms of the residual aberration and the saturation indicator variable was 0.3457. This positive correlation shows that a larger amount of residual aberration was found when the stroke of the deformable mirror was exhausted. Finally we computed a correlation of -0.7388 between the normalised measurement of the image quality metric and the rms of the residual aberration, which confirms that some aberration is removed by maximising the image quality metric. We conclude that saturation of the deformable mirror was an issue that hampered the results in our experimental validation.

3.4.8 Variations of the parameters over the field of view

We report spatial variations in the parameters c_0, $\mathbf{c}_1$ and Q. The parameters differed when computed respectively in region A, B and C in Fig. 3.3. Due to the variations, we were not able to apply the aberration correction algorithm using a single set of parameters, e.g., by correcting aberrations in region B and C using the parameters computed from region A. Variations in the parameters represent a challenge for model-based aberration correction algorithms, since the parameters are computed once only using a calibration experiment [197, 200, 155, 159, 128, 160, 161, 162, 156, 157, 144, 118, 158].

Parameter $\mathbf{c}_1$ depends on the non-zero aberration that is present when collecting the input–output measurements used in Eq. (3.20). This parameter can be removed by applying the aberration correction algorithm as done in Subsection 3.4.5. Instead, c_0 is dependent on the maximum value of the image quality metric, which differed in the three regions. We found variations in Q. For example, the largest eigenvalue of Q varied by about 30% in region B and by about 22% in region C with respect to its value in region A. The eigenvectors of Q were also affected. For instance, the eigenvector corresponding to the second largest eigenvalue of Q was rotated by about 7° in region B and by about 21° degrees in region C with respect to its orientation in region A.

From Section 3.4.5 and Section 3.4.6 we conclude that the minimum number of measurements necessary to apply the aberration correction when c_0 and Q are known is $N + 1$, as also found in [155, 128, 158]. If the parameters c_0 and Q vary during the acquisition of different regions of the specimen, then additional measurements are necessary to update the parameters before the aberration correction can be applied. This is consistent with [159, 160, 161, 162, 144, 118], where algorithms that use a minimum of $2N + 1$ measurements were employed. By approximating the solution of Eq. (3.12), these algorithms use the additional N measurements to estimate all the eigenvalues of Q each time the aberration correction is applied (see Section 4 in [158]). Nevertheless, variations in the orientations of the eigenvectors, such as the ones detected during our experiments, are not

accounted for and affect the accuracy of the aberration correction (see Section 3 in [144]). For these reasons, detecting variations and updating the parameters of the model as different regions of the specimen are acquired is an open research challenge.

3.5 Conclusions

In this paper we present a procedure to define a new set of basis functions for the control of the deformable mirror. The new basis functions can be made approximately orthogonal to a set of Zernike polynomials. This is necessary for applying aberration correction in scanning microscopy applications, where the deformable mirror must not induce the x-tilt, y-tilt and defocus aberrations.

The second contribution concerns the algorithm used for the aberration correction. We consider computing the least-squares estimate of the unknown aberration. Although this problem is non-convex in general, the least-squares estimate can be computed efficiently by exploiting results already applied in the solution of localisation problems [207]. Once the estimate is computed, the aberration correction is applied by maximising the image quality metric.

We implement the aberration correction algorithm in a second-harmonic microscope. First, we are able to compute the parameters of the quadratic polynomial used to model the image quality metric directly from input–output measurements, using a previously proposed method [158]. Second, we validate the aberration correction algorithm discussed in this paper. We also report the measurement of variations in the parameters of the quadratic polynomial over the field of view.

Acknowledgements

This research is supported by the Technology Foundation STW under project number 10433, which is part of the Smart Optics Systems perspective programme directed by Prof. M. Verhaegen.

Chapter 4

Modal-based phase retrieval for adaptive optics

We consider a phase retrieval (PR) algorithm to correct phase aberrations in an optical system. Our solution uses three measurements of the point-spread function (PSF) to estimate the aberration. We apply a different defocus aberration with a deformable mirror (DM) in each of the three measurements of the PSF. Once the aberration is estimated using the PR algorithm, we minimise it with the DM. Using the extended Nijboer–Zernike theory, the PR problem is formulated into a matrix rank minimisation problem. A solution to the rank minimisation problem is obtained with PhaseLift, a signal recovery method based on convex optimisation. We demonstrate the feasibility of this algorithm by performing aberration correction experiments with an optical breadboard.

Reference: J. Antonello and M. Verhaegen, "Modal-based phase retrieval for adaptive optics," (in preparation).

4.1 Introduction

The phase retrieval (PR) problem consists in finding the unknown phase of a complex-valued function from a set of measurements of the magnitude of its Fourier transform [172]. This inverse problem has a number of applications in crystallography [211], astronomy [212], optical imaging [213], microscopy [214, 215, 216, 217], single-molecule localisation [218] and adaptive optics [219], which we consider in this paper. In this latter application, one is concerned with minimising the phase aberration in the pupil of an optical system using an adaptive element, such as a deformable mirror (DM). To achieve this, an estimate of the aberration must be obtained. One can estimate the aberration by applying a PR algorithm to a set of measurements of the point-spread function (PSF) of the optical system. This is an attractive way to estimate the aberration for some optical systems [220, 221, 65], due to the experimental simplicity in recording the measurements of the PSF. Further, non-common path errors are avoided and no additional optical components must be included, as is required for implementing interferometric methods or Shack–Hartmann wavefront sensing.

A number of different algorithms have been developed to solve PR problems, for a recent, comprehensive review we refer to [222]. The most widely used algorithms are based on alternating projections (AP), such as the Gerchberg–Saxton algorithm [223] and its variants developed by Fienup [224]. AP algorithms formulate the PR problem into a set theoretic framework, where one seeks the unknown function that lies in the intersection of two or more sets. A set can express the constraint about the magnitude of the Fourier transform of the function, or some additional *a priori* information, such as knowledge about the support of the function. Successful applications of these algorithms, which are relevant for our case of aberration estimation, have been reported, e.g., in [214, 184, 218]. Here, the authors collect measurements of the PSF at different defocus planes, to improve the convergence of the AP algorithm. This is similar to the practice in the more difficult problem of phase diversity (PD) [219, 225, 226], which we do not consider in this paper, and where both the aberration of the optical system and the object that is imaged are jointly estimated. In general, AP algorithms are not guaranteed to converge to the unique solution [227, 228, 229, 230], provided a unique solution to the PR problem exists up to trivial ambiguities [231, 222, 232].

More recently, new algorithms based on convex optimisation [171] have been proposed to address the PR problem. The principle is that a solution of the PR problem is obtained if one is able to solve a related matrix rank minimisation problem [233, 178, 179, 234, 235], which is known to be an NP-hard problem [175, 176, 177]. For this reason, one does not attempt to solve the rank minimisation problem itself, but instead considers a convex relaxation [180]. In this paper we employ the convex relaxation proposed in [178, 179], which the authors call PhaseLift. Using semidefinite programming [191], we can efficiently compute the solution of this convex relaxation and, as a result, we obtain an approximate solution to the PR problem. As pointed out in [178, 179], one advantage of PhaseLift is that it can be made robust to measurement noise. On the contrary, measurement noise is a serious issue that can lead to inconsistency in AP algorithms, whereby the intersection of the sets that express the constraints becomes empty, see [229] and the references therein.

An important issue encountered in PR algorithms is the computational burden. For example, when considering AP algorithms, one does not work with the unknown function itself, which is a continuous function. Instead, the discrete signal obtained by sampling the unknown function over a finite grid is considered, and the 2D fast Fourier transform (FFT) is used to approximate the Fourier transform in each iteration of the AP algorithm. This approach can lead to an overall significant computational burden [236]. In this paper, instead of working with a sampling grid, we employ a modal decomposition of the unknown function to reduce the computational burden. By applying the extended Nijboer–Zernike theory (ENZ) [173, 174], we approximate the unknown function using a low number of complex-valued Zernike polynomials [44], only 21 in our case. Instead of evaluating the unknown function over a sampling grid and computing the FFT, we use the ENZ formulas to compute the complex PSF via a matrix-vector product, which is computationally cheaper, since we considered only 21 Zernike polynomials.

To corroborate our arguments, we perform aberration correction experiments using an optical breadboard. We introduce an unknown aberration in the optical system with a DM. After recording three measurements of the PSF, where in each measurement the DM applies a different amount of additional defocus aberration, we estimate the unknown

aberration with the PR algorithm. Using the estimate, we apply the aberration correction with the DM, and we measure the residual aberration with a Shack–Hartmann wavefront sensor.

The paper is organised as follows. In Section 4.2, the PR problem is formulated using the ENZ theory. Section 4.3 discusses how to solve the PR problem using PhaseLift. In Section 4.4, we report the experimental results. The conclusions are drawn in Section 4.5.

4.2 Formulation of the phase retrieval problem using the extended Nijboer–Zernike theory

For completeness, we briefly recall some necessary results of the ENZ theory, which is discussed in detail elsewhere [173, 174, 237, 238]. The generalised pupil function (GPF) of an aberrated optical system is a complex-valued function which is defined as [11]

$$P(\rho,\theta) = A(\rho,\theta)\exp(i\Phi(\rho,\theta)), \tag{4.1}$$

where ρ and θ are the normalised polar coordinates in the exit pupil plane of the optical system, $A(\rho,\theta)$ is the amplitude apodisation function, and $\Phi(\rho,\theta)$ is the phase aberration function. Both $A(\rho,\theta)$ and $\Phi(\rho,\theta)$ are real-valued. We can approximate the GPF using a truncated series of complex-valued Zernike polynomials [44],

$$\hat{P}(\rho,\theta) = \sum_{n,m} \beta_n^m \mathcal{N}_n^m(\rho,\theta), \tag{4.2}$$

so that $P(\rho,\theta) \approx \hat{P}(\rho,\theta)$. In Eq. (4.2), n and m are respectively the radial order and the azimuthal frequency of the complex-valued Zernike polynomial $\mathcal{N}_n^m(\rho,\theta)$, which is defined in Eq. (4.21) within Appendix 4.A. The coefficient β_n^m of $\mathcal{N}_n^m(\rho,\theta)$ is a complex number. Eq. (4.2) is a generalisation for complex-valued functions of the real-valued Zernike series, which is commonly used to analyse $\Phi(\rho,\theta)$ in adaptive optics literature [44]. When a maximum radial order n_M is considered, the summation in Eq. (4.2) extends over $N_\beta = (n_M+1)(n_M+2)/2$ addends, and the coefficients can be collected into a vector $\boldsymbol{\beta} \in \mathbb{C}^{N_\beta}$.

The normalised complex PSF corresponding to $P(\rho,\theta)$ is given by [173, 174]

$$\begin{aligned} U(r,\phi,f) =& \frac{1}{\pi}\int_0^1\int_0^{2\pi} \exp(if\rho^2)P(\rho,\theta) \\ &\times \exp(i2\pi r\rho\cos(\theta-\phi))\,\rho\,d\rho\,d\theta, \end{aligned} \tag{4.3}$$

where r and ϕ are coordinates in the image plane that are normalised by the diffraction unit λ/NA, NA is the image-side numerical aperture of the optical system, and f is the defocus parameter [173, 174]. The authors in [173, 174], show that by replacing $P(\rho,\theta)$

with $\hat{P}(\rho,\theta)$, one has the following approximation for $U(r,\phi,f)$,

$$\hat{U}(r,\phi,f,\boldsymbol{\beta}) = 2\sum_{n,m}\beta_n^m\sqrt{n+1}i^m V_n^m(r,f)\exp(im\phi), \tag{4.4}$$

where the complex terms $V_n^m(r,f)$ are defined in Eq. (2.47) in [237]. These terms can be computed using semi-analytic formulas (Eq. (2.48) in [237]) for the required accuracy [173]. Note that we employ a different normalisation for $\mathcal{N}_n^m(\rho,\theta)$ than what is adopted in [173, 174, 237, 238], which is useful to evaluate the error in approximating $P(\rho,\theta)$ with $\hat{P}(\rho,\theta)$ (see Section 4.4.4 and Appendix 4.C).

Using the ENZ theory to compute the complex PSF can be computationally advantageous with respect to the conventional method of propagating the field using the FFT. While a 2D FFT has a complexity of $O(N^2\log(N))$ for a square $N\times N$ sampling grid, Eq. (4.4) can be expressed as a matrix-vector product (see also Section 4.3), where the matrix can be precomputed once only, and the vector is $\boldsymbol{\beta}\in\mathbb{C}^{N_\beta}$. The complexity of evaluating the matrix-vector product is $2N^2(2N_\beta-1)$, which is $O(N^2)$. Henceforth, provided that the constant N_β is small enough, using Eq. (4.4) is computationally profitable.

We can now formulate the PR problem. One can measure the PSF, using a CCD detector with N_p pixels, at N_f different locations along the optical axis, which correspond to N_f different values of the defocus parameter f. As a result, the collected pixel values can be stored into a vector $\mathbf{I}\in\mathbb{R}^{N_m}$, where $N_m = N_pN_f$. Each element I_k of $\mathbf{I}$ corresponds to the intensity recorded by the pixel at position (r_k,ϕ_k) in the image plane, and at defocus position f_k along the optical axis. In practice, one can obtain $\mathbf{I}$ by moving the CCD detector with a motorised stage. In this paper, we collected $\mathbf{I}$ by applying the necessary defocus with the DM instead of requiring mechanical movement of the CCD detector.

The error between the recorded measurements and the expected PSF is given by a vector $\boldsymbol{\epsilon}\in\mathbb{R}^{N_m}$, where for each element ϵ_k of $\boldsymbol{\epsilon}$,

$$\epsilon_k = I_k - |\hat{U}(r_k,\phi_k,f_k,\boldsymbol{\beta})|^2. \tag{4.5}$$

The solution of the PR problem is found by minimising the norm of the error, i.e.,

$$\min_{\boldsymbol{\beta}\in\mathbb{C}^{N_\beta}} \|\boldsymbol{\epsilon}\|. \tag{4.6}$$

In [220, 221], the authors assume a small phase aberration consisting of cosine terms only, and describe a procedure to linearise $|\hat{U}(r_k,\phi_k,f_k,\boldsymbol{\beta})|^2$, so that the PR problem can be solved via a linear system of equations. This procedure is not adequate for medium to large aberrations [238], and was improved in [221] by considering a predictor-corrector method to overcome the linearisation error. In the following section, instead, we formulate a convex relaxation of Eq. (4.6) using PhaseLift. In our case, a general aberration comprising sine and cosine terms can be handled, as in [238]. Nevertheless, contrary to [238], we do not need to explicitly derive the linearisation of $|\hat{U}(r_k,\phi_k,f_k,\boldsymbol{\beta})|^2$, since the task of computing the solution is left to the convex optimisation solver.

4.3 Solution of the phase retrieval problem using PhaseLift

By defining the complex vectors $\mathbf{a}_k \in \mathbb{C}^{N_\beta}$ for $k = 1, \ldots, N_m$ as

$$\mathbf{a}_k = \begin{bmatrix} \vdots \\ 2\sqrt{n+1} V_n^m(r_k, f_k)^H e^{i(f_k - m\left(\phi_k + \frac{\pi}{2}\right))} \\ \vdots \end{bmatrix}, \tag{4.7}$$

where $\cdot^H$ denotes the conjugate transpose, we can expand Eq. (4.6) into

$$\min_{\boldsymbol{\beta} \in \mathbb{C}^{N_\beta}} \left(\sum_{k=1}^{N_m} \left(I_k - |\langle \boldsymbol{\beta}, \mathbf{a}_k \rangle|^2 \right)^2 \right)^{1/2}, \tag{4.8}$$

where the angle brackets denote the inner product, i.e., $\langle \boldsymbol{\beta}, \mathbf{a}_k \rangle = \mathbf{a}_k^H \boldsymbol{\beta}$. Exploiting the properties of the trace operator $\mathrm{tr}(\cdot)$, one can show [233, 178, 179, 234, 235] that $|\langle \boldsymbol{\beta}, \mathbf{a}_k \rangle|^2 = \mathrm{tr}(\mathbf{a}_k \mathbf{a}_k^H \boldsymbol{\beta} \boldsymbol{\beta}^H)$. Letting $A_k = \mathbf{a}_k \mathbf{a}_k^H$ and $B = \boldsymbol{\beta}\boldsymbol{\beta}^H$, Eq. (4.8) is equivalent to

$$\begin{aligned} \min_{\substack{B \in \mathbb{C}^{N_\beta \times N_\beta} \\ \boldsymbol{\epsilon} \in \mathbb{R}^{N_m}}} \quad & \|\boldsymbol{\epsilon}\| \quad \text{s.t.} \\ & \epsilon_k = I_k - \mathrm{tr}(A_k B) \quad k = 1, \ldots, N_m \\ & B \succeq 0 \\ & \mathrm{rank}(B) = 1. \end{aligned} \tag{4.9}$$

The problem in Eq. (4.9), is known to be NP-hard [178, 179], due to the rank constraint. Following [178, 179], the rank constraint is removed and the cost function is modified from $\|\boldsymbol{\epsilon}\|$ to $\|\boldsymbol{\epsilon}\| + \lambda_r \, \mathrm{rank}(B)$, where λ_r is a regularisation parameter [178]. A convex relaxation of Eq. (4.9) is obtained [178, 179] by replacing $\mathrm{rank}(B)$ with its convex surrogate [177] $\mathrm{tr}(B)$, which results in the following optimisation problem,

$$\begin{aligned} \min_{\substack{B \in \mathbb{C}^{N_\beta \times N_\beta} \\ \boldsymbol{\epsilon} \in \mathbb{R}^{N_m}}} \quad & \|\boldsymbol{\epsilon}\| + \lambda_r \, \mathrm{tr}(B) \quad \text{s.t.} \\ & \epsilon_k = I_k - \mathrm{tr}(A_k B) \quad k = 1, \ldots, N_m \\ & B \succeq 0. \end{aligned} \tag{4.10}$$

The problem in Eq. (4.10) is a complex semidefinite programme, which can be solved using cvxopt [239] as outlined in Appendix 4.D. Once a solution B has been computed, the estimate of $\boldsymbol{\beta}$ is obtained by considering the rank-1 approximation of B that has the largest eigenvalue, as done in [178, 179].

We note that, by employing the ENZ theory, one can select a low value of N_β, which results in low dimensions for the complex semidefinite constraint $B \succeq 0$. In this paper, B is a 21×21 complex matrix. For comparison, consider the case where one uses the discrete signal generated by sampling the GPF over a square grid instead of the modal decompos-

ition of the GPF (see Eq. (4.2)). In this case, a 16×16 complex semidefinite constraint is necessary if the GPF is sampled using a 2×2 grid, which is a rough approximation of the GPF. Choosing a 3×3 grid, would require a 81×81 complex semidefinite constraint. We select $N_\beta = 21$ and numerically estimate the approximation error between Eq. (4.1) and Eq. (4.2) in Section 4.4.4. This choice has the additional benefit that it is practical to use an interior-point method [210] to solve Eq. (4.10), instead of the first-order methods employed in [178, 179]. Finally, we remark that, as outlined in the introduction, this formulation is robust to measurement noise, since the error $\|\boldsymbol{\epsilon}\|$ is minimised, as it appears in the cost function of Eq. (4.10).

4.4 Experimental results

In this section we report the experimental results of applying the PR algorithm to correct static phase aberrations. Although we only consider phase aberrations, we note that the PR algorithm also provides estimates of the amplitude apodisation function.

We implemented an adaptive optics system comprising a DM, a Shack–Hartmann wavefront sensor, and a CCD detector to measure the PSF. A single aberration correction experiment consists in the following three steps. First, a static aberration is applied with the DM and the corresponding PSF is measured at N_f different defocus positions. The PR algorithm has, of course, no knowledge about the static aberration. Second, the PR algorithm is applied to obtain an estimate of the static aberration, which is used for the aberration correction. Third, both the static aberration and the aberration correction are applied simultaneously using the DM. The Shack–Hartmann wavefront sensor is used to measure the residual aberration and to assess the performance of the aberration correction.

4.4.1 Description of the experimental setup

The layout of the experimental setup is depicted in Fig. 4.1. The He–Ne laser source (LASER; wavelength 632.8 nm) is spatially filtered using lens L1 (focal length 11 mm) and pinhole P (diameter 35 μm). The resulting diverging beam is collimated by lens L2 (focal length 300 mm) and stopped down with an iris diaphragm (AP) to a 9 mm wide circular section with uniform amplitude. After passing through beam splitter BS1, the beam is reflected by the DM (DM; Okotech, 17-channel micromachined deformable mirror with tip-tip stage), and then is directed towards the sensors by BS1. The DM is reimaged by lenses L3 (focal length 200 mm) and L4 (focal length 100 mm) onto a Shack–Hartmann wavefront sensor (SH), which consists of a hexagonal microlens array (Okotech, hexagonal array with 127 microlenses, 18 mm focal distance, 300 μm pitch) and a CCD sensor (Sys-Vistek, svs340, 648×492 pixels, 7.4 μm pixel size). A second beam is created with beam splitter BS2, and is focused by lens L5 (focal length 500 mm) onto a CCD detector (CCD; Sys-Vistek, svs340, 648×492 pixels, 7.4 μm pixel size), which measures the PSF.

A desktop PC with a 3 GHz processor (Intel, Xeon X5472) running Linux is used to control the equipment. The voltage for the actuators of the DM is supplied by a high voltage amplifier (Okotech, 20-channel HV unit), which is connected to a 16 bit analogue output

card (United Electronic Industries, PD2-AO-96/16A). The CCD detectors in CCD and in SH are connected to two framegrabber cards (Arvoo, Leonardo CL Full). The hardware is controlled via customised code written in C and in Python.

For the calibration of the Shack–Hartmann wavefront sensor, we temporarily replaced the DM with a flat mirror and recorded a reference image. The control signal of each actuator u_k of the DM is quadratically related to the corresponding voltage applied with the high voltage amplifier, so that a linear displacement of the mirror is expected [50]. We operate the DM from a biased reference position, so that both positive and negative deflections of the wavefront can be induced [50], which causes the beam reflected by BS1 to be converging. To restore a collimated beam, we move lens L3 along the optical axis until the defocus measured with the Shack–Hartmann wavefront sensor is minimised. The coefficients of the Zernike analysis of the wavefront are estimated by applying a modal-based wavefront reconstruction [82].

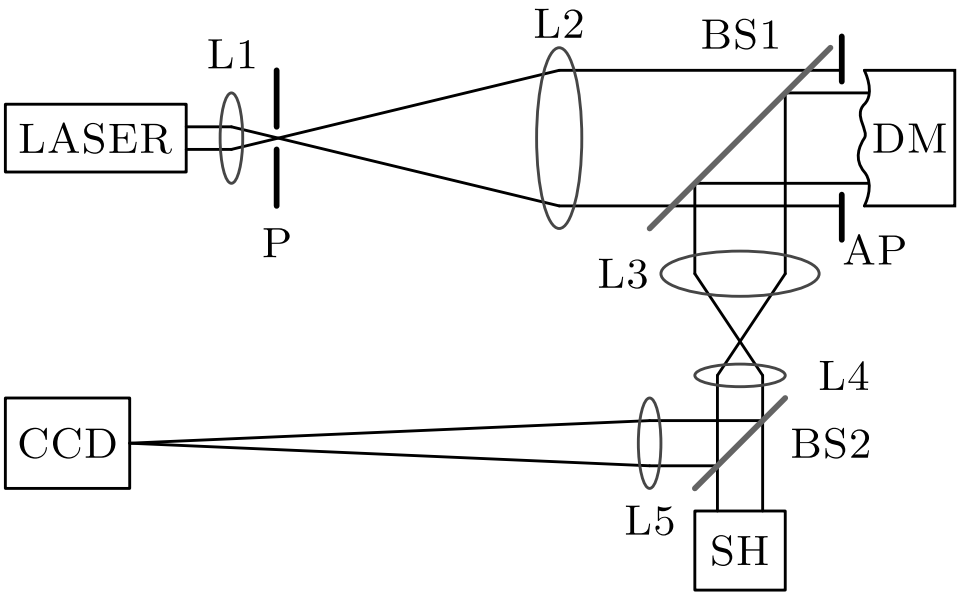

Figure 4.1: Illustration of the optical setup used in the experiments. The laser beam is spatially filtered using lens L1 and pinhole P, and collimated by lens L2. A 9 mm wide circular section of the beam is selected with diaphragm AP to illuminate the DM. The DM is reimaged two-to-one using lenses L3 and L4 onto a Shack–Hartmann wavefront sensor (SH). A second beam is directed by beam splitter BS2 through lens L5, and is focused onto a CCD sensor (CCD).

4.4.2 Preparation of the experiments

We use the Nelder–Mead (NM) algorithm [152] to align the tip-tilt stage of the DM. The NM algorithm is applied to find the orientation of the stage that minimises the 2-norm of

the vector $[\alpha_1^1, \alpha_1^{-1}]^T$, which is estimated using the Shack–Hartmann wavefront sensor. Once aligned, the stage is kept at the same orientation throughout all the experiments, and the two actuators used to operate the stage are neglected.

The control signals of the $N_a = 17$ actuators, i.e., $u_1, \dots, u_{N_a}$, are collected into a vector $\mathbf{u} \in \mathbb{R}^{N_a}$. Each control signal u_k is normalised so that $-1 \leq u_k \leq 1$, where -1 and 1 correspond, respectively, to the minimum and to the maximum voltage that is applicable to the actuator. In our experiments, the Shack–Hartmann wavefront sensor provided estimates of the Zernike coefficients up to the sixth radial order, so that we consider $N_\alpha = 28$ Zernike polynomials to describe the aberration.

Assuming linearity, $\Phi(\rho, \theta)$ is given by the superposition of the influence functions [50] $\psi_k(\rho, \theta)$ of each actuator of the DM,

$$\Phi(\rho, \theta) = \sum_{i=1}^{N_a} u_k \psi_k(\rho, \theta). \tag{4.11}$$

Using the first $N_\alpha = 28$ Zernike polynomials, we have the following approximation,

$$\Phi(\rho, \theta) \approx \sum_{n,m} \alpha_n^m \mathcal{Z}_n^m(\rho, \theta). \tag{4.12}$$

We model the DM using a matrix $H \in \mathbb{R}^{N_\alpha \times N_a}$ that approximately maps a certain actuation vector $\mathbf{u}$ to the corresponding vector of Zernike coefficients $\boldsymbol{\alpha}$, i.e., $\boldsymbol{\alpha} \approx H\mathbf{u}$. Note that in this paper we always assume $\alpha_0^0 = 0$ (see also Appendix 4.C). Although we do not make this explicit in the formulas involving H to maintain a simple notation, only the sub-matrix obtained by neglecting the first row of H is used in our computations. To estimate H, we collected input–output data by applying $D \geq N_a$ random actuation vectors $\mathbf{u}_1, \dots, \mathbf{u}_D$ and by recording the corresponding estimates of the Zernike coefficients $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_D$ obtained with the Shack–Hartmann wavefront sensor. H is subsequently recovered by solving a least-squares problem. Once H is available, we can suppress the initial aberration in the system, which is due to the misalignment of the optical components and to the non-flat initial shape of the DM.

4.4.3 Generation of the random Zernike aberrations

We performed 100 random aberration correction experiments. The random aberrations were generated by drawing 100 random actuation vectors $\mathbf{u}_1, \dots, \mathbf{u}_{100}$ from a set of normal multivariate distributions. A boxplot of the expected Zernike aberrations, i.e., $\boldsymbol{\alpha}_1$, …, $\boldsymbol{\alpha}_{100}$, is reported in Fig. 4.2. The expected Zernike aberrations can be estimated using $\boldsymbol{\alpha} \approx H\mathbf{u}$. From the boxplot, it can be seen that the magnitude of the Zernike coefficients decreases for increasing radial order and azimuthal frequency. Therefore, the DM can correct significant amounts of low-order Zernike aberrations, i.e., aberrations whose Zernike polynomials have low radial order and low azimuthal frequency. On the contrary, the mirror can introduce only a negligible amount of high-order Zernike aberrations, due to the finite number of actuators and to the mechanical limitations of the mirror. Therefore, we regard $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{100}$ as typical aberrations that should be estimated by the PR algorithm, since they express the correction capabilities of our DM.

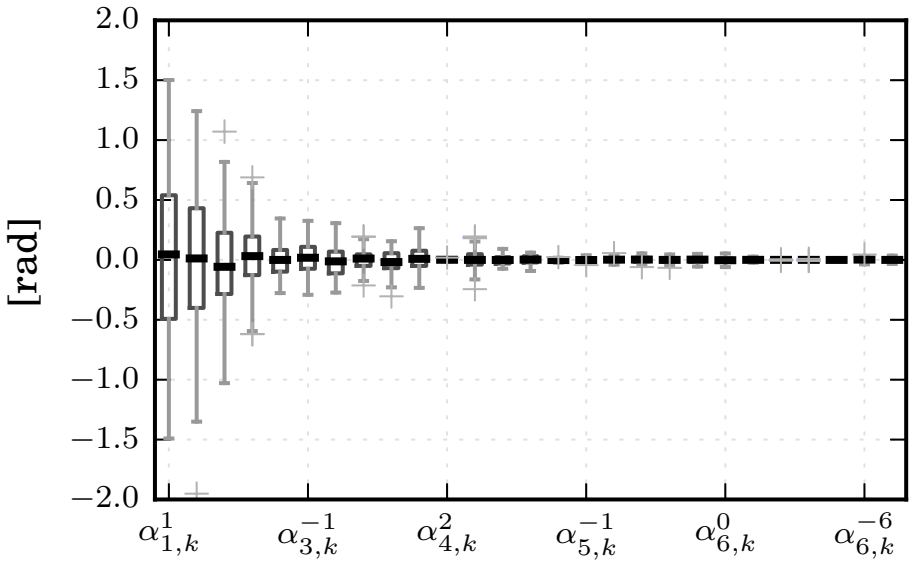

Figure 4.2: Boxplot of the 100 random Zernike aberrations. Each aberration is generated by drawing a random actuation voltage $\mathbf{u}_k$. The corresponding Zernike analysis $\boldsymbol{\alpha}_k$ is estimated using $\boldsymbol{\alpha} \approx H\mathbf{u}$. The Zernike coefficients are ordered and normalised according to Noll [44]. For each position in the x-axis, the elements $\alpha^m_{n,1}, \ldots, \alpha^m_{n,100}$ of vectors $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{100}$ are considered to compute the boxplot indicators. The tops and bottoms of the boxes denote the 25th and 75th percentiles, the horizontal lines in the middle denote the medians, the whiskers extend to the furthest measurements not considered as outliers using a 1.5 interquartile range. This description holds for all the boxplots reported in this paper.

4.4.4 Error in approximating the generalised pupil function

In this subsection, we estimate the error that is expected when the GPF is approximated using 21 complex-valued Zernike polynomials. For each vector $\boldsymbol{\alpha}_k \in \mathbb{R}^{N_\alpha}$, we consider the corresponding GPF with uniform unit amplitude, i.e.,

$$P_k(\rho,\theta) = \exp\left(i \sum_{n,m} \alpha^m_{n,k} \mathcal{Z}^m_n(\rho,\theta)\right), \tag{4.13}$$

and its approximation obtained with a finite set of Zernike polynomials

$$\hat{P}_k(\rho,\theta) = \sum_{n,m} \beta^m_{n,k} \mathcal{N}^m_n(\rho,\theta). \tag{4.14}$$

The approximation error is $E_k(\rho,\theta) = P_k(\rho,\theta) - \hat{P}_k(\rho,\theta)$. Although E_k is a complex-valued function defined over the pupil, we employ the real-valued $\mathrm{rms}(E_k)$, to concisely quantify the error using a scalar number. The definition of $\mathrm{rms}(\cdot)$ is given in Appendix 4.C, where we also discuss that this indicator is a generalisation of the rms indicator that is commonly used for the real-valued phase aberration function. First, we employ $N_\beta = 136$ Zernike polynomials in Eq. (4.14), to obtain accurate estimates of the GPFs. For each k, we estimate the Zernike coefficients $\beta^m_{n,k} \in \mathbb{C}$ by sampling Eq. (4.13) over a finite grid and by taking the approximate inner products, as described in Appendix 4.B. This results in 100 vectors $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_{100}$, where each vector belongs to $\mathbb{C}^{136}$. We estimated that $\mathrm{rms}(E_k) \leq 8.74e-5$ for all k when $N_\beta = 136$ in Eq. (4.14). We neglect this error and assume $E_k \approx 0$ for the rest of this subsection. In the following paragraph, we use the vectors $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_{100}$ to estimate the error in approximating the GPF using $N_\beta = 21$ instead of $N_\beta = 136$ in Eq. (4.14).

Each vector $\boldsymbol{\beta}_k$ can be split into two vectors $\boldsymbol{\beta}^l_k$ and $\boldsymbol{\beta}^h_k$, where the first 21 elements of $\boldsymbol{\beta}_k$ are collected into $\boldsymbol{\beta}^l_k$ and the remaining elements are collected into $\boldsymbol{\beta}^h_k$. By exploiting the orthogonality properties of the Zernike polynomials and the normalisation employed in Appendix 4.A, we have $\mathrm{rms}(E_k) \approx \|\boldsymbol{\beta}^h_k\|$ when only 21 Zernike polynomials are used in Eq. (4.14). The approximation sign is due to the error in computing the approximate inner products and in considering the first 136 Zernike polynomials instead of an infinite number of polynomials. In Fig. 4.3, we report each estimate of $\mathrm{rms}(E_k)$ using $\times$ symbols, against the corresponding rms of the phase aberration, $\mathrm{rms}(\hat{\Phi}_k)$, which is computed as $\|\boldsymbol{\alpha}_k\|$ (see Appendix 4.C). It can be seen that the error in using 21 Zernike polynomials in Eq. (4.14) grows for an increasing rms of the phase aberration.

For comparison, we can evaluate the error when one uses the linear approximation of the GPF,

$$\hat{P}^l_k(\rho,\theta) = 1 + i \sum_{n,m} \alpha^m_{n,k} \mathcal{Z}^m_n(\rho,\theta), \tag{4.15}$$

which is widely employed in the literature, e.g. in [220, 221, 65]. To evaluate the error $E^l_k(\rho,\theta) = P_k(\rho,\theta) - \hat{P}^l_k(\rho,\theta)$ using the vectors $\boldsymbol{\beta}_1,\dots,\boldsymbol{\beta}_{100}$, we must compute the coefficients $\gamma^m_{n,k} \in \mathbb{C}$ of the complex Zernike polynomials $\mathcal{N}^m_n(\rho,\theta)$ that express the right-hand

side of Eq. (4.15). We use the following equations,

$$\begin{cases} \gamma_0^0 = 1 & n = 0, m = 0 \\ \gamma_n^m = i\alpha_n^m & n \neq 0, m = 0 \\ \gamma_n^m = (\alpha_n^{-m} + i\alpha_n^m)/\sqrt{2} & m > 0 \\ \gamma_n^m = (-\alpha_n^m + i\alpha_n^{-m})/\sqrt{2} & m < 0 \end{cases}, \tag{4.16}$$

where the index k is removed for clarity, and we have used the assumptions of unit amplitude and that $\alpha_{0,k}^0 = 0$. The rms of the error $E_k^l(\rho,\theta)$ is approximately given by $\|\boldsymbol{\beta}_k - \boldsymbol{\gamma}_k\|$, where the first N_α elements of vectors $\boldsymbol{\gamma}_k \in \mathbb{C}^{136}$ are given by the corresponding coefficients $\gamma_{n,k}^m$. The approximation sign here is intended in the same sense as in the previous paragraph, whereas the relations between $\alpha_{n,k}^m$ and $\gamma_{n,k}^m$ given in Eq. (4.16) are exact. We report $\mathrm{rms}(E_k^l)$ using + symbols in Fig. 4.3. Note that even though in the linear approximation $N_\alpha = 28$ Zernike polynomials are considered, the approximation error is still larger than that of using only 21 Zernike polynomials in Eq. (4.14). The error in the linear approximation stems from the fact that the coefficients $\boldsymbol{\alpha}_k$ and $\boldsymbol{\gamma}_k$ are non-linearly related [174] by the complex exponential function. In addition, Eq. (4.15) introduces a larger amplitude error with respect to Eq. (4.14), see Section 2.1 in [63].

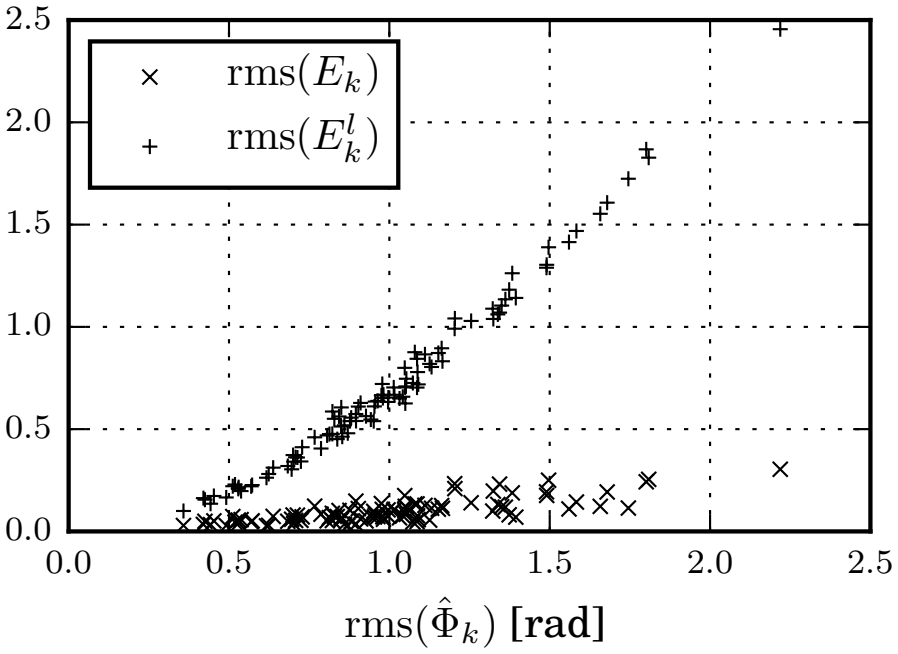

Figure 4.3: Expected error in approximating $P_k(\rho,\theta)$ with $\hat{P}_k(\rho,\theta)$ (× symbols, $\mathrm{rms}(E_k)$) and $\hat{P}_k^l(\rho,\theta)$ (+ symbols, $\mathrm{rms}(E_k^l)$). $\mathrm{rms}(E_k)$ is estimated as $\|\boldsymbol{\beta}_k^h\|$ and $E_k^l(\rho,\theta)$ is estimated as $\|\boldsymbol{\beta}_k - \boldsymbol{\gamma}_k\|$. Both errors increase as the rms of the phase aberration $\mathrm{rms}(\hat{\Phi}_k)$ increases.

4.4.5 Aberration correction experiments

We first apply the PR algorithm to correct for the non-common path error between the arm with the Shack–Hartmann wavefront sensor and the arm with the CCD, see Fig. 4.1. All the following measurements of the residual aberrations performed with the Shack–Hartmann wavefront sensor are referenced to this initial aberration correction, which we denote with (R). We perform 100 aberration correction experiments and record three measurements of the PSF in each experiment, respectively at defocus position $f_1 = 0.0$, $f_2 = -1.0$, and $f_3 = 1.7$, so that $N_f = 3$. A measurement consists of a 35 pixels $\times$ 35 pixels image.

Each experiment comprises the following steps. First the DM applies the aberration $\boldsymbol{\alpha}_k$, and the initial rms of the aberration is measured with the Shack–Hartmann wavefront sensor. At this point, the three measurements of the PSF at defocus position f_1, f_2, and f_3 are recorded, by simultaneously applying with the DM the aberration $\boldsymbol{\alpha}_k$ and the corresponding Zernike defocus $\alpha_2^0 = f_j/(2\sqrt{3})$ for $j = 1,\dots,3$. The PR algorithm is subsequently applied, by solving Eq. (4.10) using the three PSF measurements and letting $\lambda_r = 1$. As a result, an estimate $\hat{\boldsymbol{\beta}}_k \in \mathbb{C}^{21}$ is obtained. To correct the aberration, we must obtain the estimate of $\boldsymbol{\alpha}_k$ from $\hat{\boldsymbol{\beta}}_k$. We replace $\hat{\boldsymbol{\beta}}_k$ into Eq. (4.2), and evaluate the phase $\arg(\hat{P}(\rho,\theta))$ over a grid, by applying a simple phase unwrapping algorithm [240]. The estimate of $\boldsymbol{\alpha}_k$ is obtained by evaluating the the inner products as discussed in Appendix 4.B. Finally, we execute the aberration correction, by simultaneously applying with the DM both the phase aberration $\boldsymbol{\alpha}_k$ and the aberration correction $-\hat{\boldsymbol{\alpha}}_k$. We conclude the single experiment by measuring the final rms of the residual aberration with the Shack–Hartmann wavefront sensor, and the corrected PSF with the CCD. When solving the PR problem, we neglect the pixel measurements that are below 6% of the maximum value that can be measured with the CCD. In doing this, we avoid most of the contribution of the background noise of the CCD and we decrease the computational cost of the PR algorithm.

The results of the experiments are reported using boxplots in Fig. 4.4 and Fig. 4.5, where the 100 experiments are divided into two sets of 50 experiments according to the magnitude of the initial aberration. The boxplot indicators are computed for the initial rms $\|\boldsymbol{\alpha}_k\|$ (initial) on the left, and the final rms $\|\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k\|$ (final) on the right, both measured using the Shack–Hartmann wavefront sensor. It can be seen that the algorithm consistently removes some aberration, which confirms that a successful phase retrieval is achieved. The computational time is dominated by the time necessary to solve Eq. (4.10), whereas the time spent in the phase unwrapping operation and in the numerical evaluation of the inner products is negligible. Throughout the 100 experiments, the median time spent to solve Eq. (4.10) is about 30 seconds, and the maximum time is about 45 seconds.

We also report the PSFs recorded for three single experiments from the 100 trials in Fig. 4.6, Fig. 4.7, and Fig. 4.8. The PSF recorded after correcting the non-common path error at the beginning of this section is reported on the left (R) for reference. At time instant $t = 0$, the random aberration $\boldsymbol{\alpha}_k$ is introduced, and the PSF and the initial rms are recorded for the defocus position $f_1 = 0$. The same measurements are repeated at time instants $t = 1$ and $t = 2$, which correspond to the defocus positions $f_2 = -1.0$ and $f_3 = 1.7$ respectively. The PSF after the aberration correction and the final rms are reported for

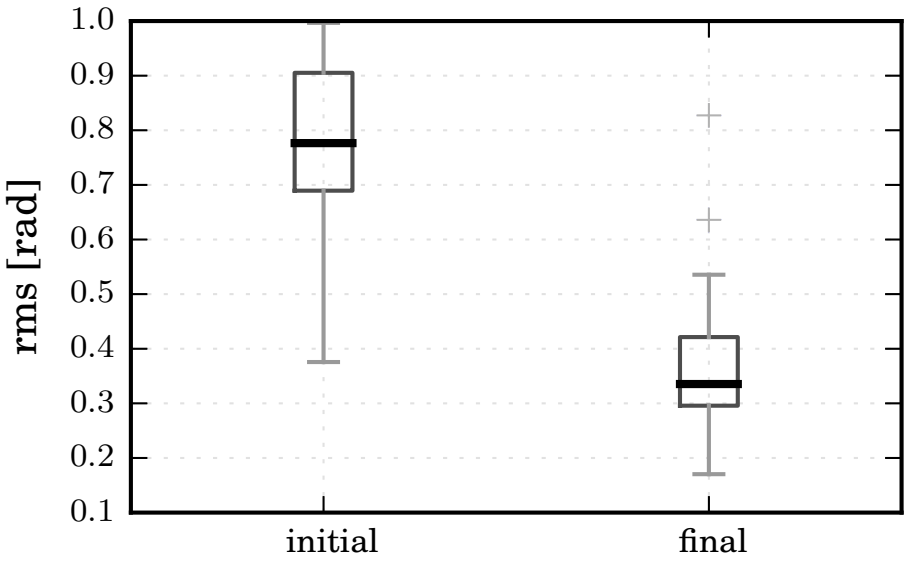

Figure 4.4: Boxplot of 50 aberration correction experiments. The boxplot indicators are computed for the initial rms $\|\boldsymbol{\alpha}_k\|$ (initial) on the left, and the final rms $\|\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}_k\|$ (final) on the right, both measured with the Shack–Hartmann wavefront sensor. A lower final rms denotes a successful aberration correction.

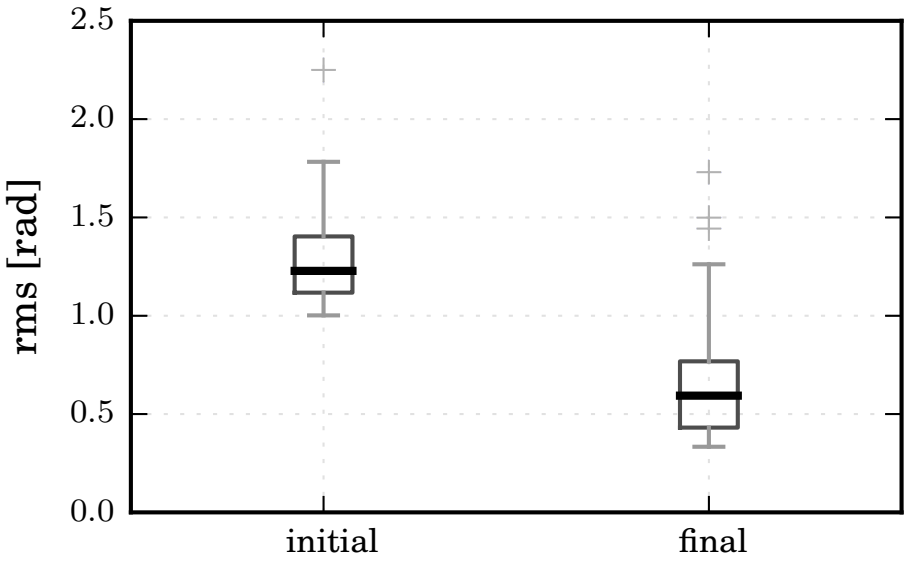

Figure 4.5: Boxplot of the remaining 50 aberration correction experiments. See the caption in Fig. 4.4.

time instant $t = 3$ on the right. Fig. 4.6 shows the maximum intensity jump from the 100 experiments, i.e., the experiment in which we recorded the maximum difference between the brightest pixel at $t = 3$ and the brightest pixel at $t = 0$. Whereas Fig. 4.7 and Fig. 4.8, report respectively the experiment with the minimum and the median intensity jump. In Fig. 4.6 and Fig. 4.8, the PR algorithm is successful, and an improved PSF is observed after the aberration correction. In Fig. 4.7, instead, the PR algorithm fails to significantly improve the PSF.

Figure 4.6: Single aberration correction experiment. Each image reports the PSF measured with the CCD detector. The rms of the residual aberration is measured with the Shack–Hartmann wavefront sensor and is referenced to the correction of the non-common path error (see the beginning of Section 4.4.5), which is reported in (R). At time instant $t = 0$, the aberration is introduced and the PSF and the initial rms are recorded for the defocus position $f_1 = 0$. Two more measurements of the PSF are collected at time instants $t = 1$ (defocus position $f_2 = -1.0$) and $t = 2$ (defocus position $f_2 = 1.7$). The PSF after the aberration correction and the final rms are reported for time instant $t = 3$. Only the measurements in $t = 0$, $t = 1$ and $t = 2$ are supplied to the PR algorithm. This experiment refers to the maximum intensity jump from the 100 experiments, i.e., the experiment with the maximum difference between the brightest pixel at $t = 3$ and the brightest pixel at $t = 0$. A noticeable improvement of the PSF is observed. All five images use the same colour scale.

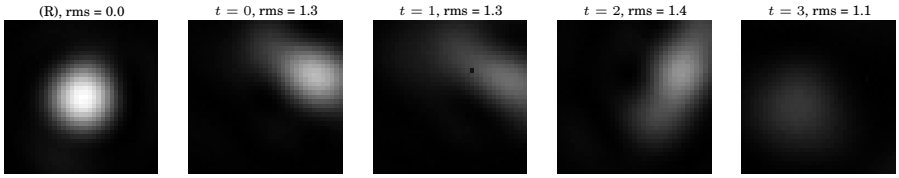

Figure 4.7: For the legend, see the caption in Fig. 4.6. This experiment refers to the minimum intensity jump from the 100 experiments. In this case, the PR algorithm fails to significantly improve the PSF.

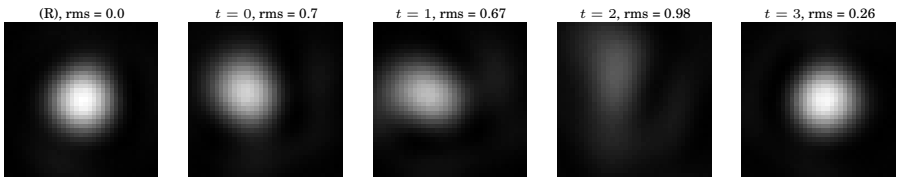

Figure 4.8: For the legend, see the caption in Fig. 4.6. This experiment refers to the median intensity jump among the 100 experiments. The PR is successful, and an improved PSF is observed.

4.5 Conclusions

In this paper we consider the problem of correcting phase aberrations in an optical system using a phase retrieval (PR) algorithm. The proposed solution uses three measurements of the point-spread function (PSF) of the optical system, where each measurement is taken at a different position along the optical axis. The three measurements can be obtained by applying an appropriate Zernike defocus aberration with the deformable mirror (DM). We formulate the PR problem using the extended Nijboer–Zernike theory [173, 174], and compute a solution using PhaseLift, a signal recovery method based on convex optimisation that is robust to additive measurement noise. Finally, we demonstrate the feasibility of this approach by performing aberration correction experiments with an optical breadboard, which includes a DM to induce and correct phase aberrations, a CCD detector to measure the PSF, and a Shack–Hartmann wavefront sensor to assess the the aberration correction.

Acknowledgements

The authors would like to acknowledge the useful discussions about this work with Prof. J. Braat and Dr. S. van Haver from TU Delft. We would like to thank Mr. W. van Geest and Mr. K. Slinkman from TU Delft for their contribution in the implementation of the experimental setup.

4.A Zernike polynomials for the analysis of real- and complex-valued functions

The truncated Zernike analysis of the real-valued phase aberration function $\Phi(\rho, \theta)$, which is defined over the unit disk, is given by

$$\hat{\Phi}(\rho, \theta) = \sum_{n,m} \alpha_n^m \mathcal{Z}_n^m(\rho, \theta), \tag{4.17}$$

where indices $n \in \mathbb{N}_0$ and $m \in \mathbb{Z}$ denote respectively the radial order and the azimuthal frequency of the Zernike polynomial $\mathcal{Z}_n^m$, and are such that $n - |m| \geq 0$ and even.

Terms $\alpha_n^m \in \mathbb{R}$ are the coefficients of the real-valued Zernike polynomials $\mathcal{Z}_n^m$, which are defined [5, 44, 40] by

$$\mathcal{Z}_n^m(\rho,\theta) = c_n^m R_n^{|m|}(\rho)\Theta_n^m(\theta), \tag{4.18}$$

where

$$\begin{aligned} c_n^m &= \begin{cases} \sqrt{n+1} & m = 0 \\ \sqrt{2(n+1)} & m \neq 0 \end{cases} \\ \Theta_n^m(\theta) &= \begin{cases} \cos(m\theta) & m \geq 0 \\ -\sin(m\theta) & m < 0 \end{cases}. \end{aligned} \tag{4.19}$$

The definition of $R_n^m(\rho)$ can be found in [5, 44, 40]. According to Noll [44], the coefficients α_n^m can be ordered using a single index, which is a function of n and m. By considering a maximum radial order of n_M and Noll's single index, we can collect the coefficients α_n^m into the elements of a vector $\boldsymbol{\alpha} \in \mathbb{R}^{N_\alpha}$, where $N_\alpha = (n_M + 1)(n_M + 2)/2$ [40].

Similarly, we can consider the truncated Zernike analysis of the complex-valued GPF $P(\rho,\theta)$, which is given by

$$\hat{P}(\rho,\theta) = \sum_{n,m} \beta_n^m \mathcal{N}_n^m(\rho,\theta), \tag{4.20}$$

where terms $\beta_n^m \in \mathbb{C}$ are the coefficients of the complex-valued Zernike polynomials $\mathcal{N}_n^m$, which are defined in [237] as

$$\mathcal{N}_n^m(\rho,\theta) = \sqrt{n+1} R_n^{|m|}(\rho)\exp(im\theta). \tag{4.21}$$

The coefficients $\beta_n^m \in \mathbb{C}$ can also be ordered using Noll's single index, and collected into a vector $\boldsymbol{\beta} \in \mathbb{C}^{N_\beta}$, where $N_\beta = (n_M + 1)(n_M + 2)/2$ and n_M is the maximum radial order considered. Note that in (4.21), we used a different normalisation than [237]. Clearly, the complex-valued Zernike polynomials may also be used to analyse a real-valued function $\Phi(\rho,\theta)$.

4.B Numerical evaluation of the inner products

Each coefficient $\alpha_n^m \in \mathbb{R}$ in Eq. 4.17 can be estimated by evaluating the inner products [44, 40],

$$\langle \Phi, \mathcal{Z}_n^m \rangle = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \Phi(\rho,\theta)\mathcal{Z}_n^m(\rho,\theta)\, \rho\, d\rho\, d\theta. \tag{4.22}$$

As suggested in [241], we evaluate $\Phi(\rho,\theta)$ using a separable cosine sampling in ρ. Define a grid $\{\theta_l = 2\pi l/L,\ l = 0,\ldots,L-1\} \times \{\rho_k = \cos((K-k-1/2)\pi/(2K),\ k = 0,\ldots,K-1\}$. Assuming $\Phi(\rho,\theta)$ is approximately constant in each neighbourhood of (ρ_k,θ_l), we can

approximate the right-hand side of Eq. (4.22) with

$$\frac{c_n^m}{\pi} \sum_{l,k} \Phi(\rho_k, \theta_l) \int_{I_{\rho_k}} R_n^m(\rho) \rho \, d\rho \int_{I_{\theta_l}} \Theta_n^m(\theta) \, d\theta, \tag{4.23}$$

where $I_{\rho_k} = [\cos((K-k)\pi/(2K)), \cos((K-k-1.0)\pi/(2K))]$ and $I_{\theta_l} = [(2l-1)\pi/L, (2l+1)\pi/L]$. One can collect the values $\Phi(\rho_k, \theta_l)$ into a vector and use Eq. (4.23) to define a matrix, so that the estimates of $\boldsymbol{\alpha}$ are computed via a matrix-vector product. In our experiments, we chose $L = 200$ and $K = 200$.

Similarly, the coefficient $\beta_n^m \in \mathbb{C}$ in Eq. 4.20 can be estimated by evaluating the inner products,

$$\langle P, \mathcal{N}_n^m \rangle = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} P(\rho, \theta) \mathcal{N}_n^m(\rho, \theta)^H \, \rho \, d\rho \, d\theta, \tag{4.24}$$

where $\cdot^H$ denotes complex conjugation. Using the same sampling defined for Eq. 4.23, we have the following approximation for the right-hand side of Eq. (4.24)

$$\frac{\sqrt{n+1}}{\pi} \sum_{l,k} P(\rho_k, \theta_l) \int_{I_{\rho_k}} R_n^m(\rho) \rho \, d\rho \int_{I_{\theta_l}} \exp(-im\theta) \, d\theta, \tag{4.25}$$

which allows to estimate $\boldsymbol{\beta}$ via a matrix-vector product.

4.C Root-mean-square value of the phase aberration and the generalised pupil functions

Using the following functional [40, 46],

$$\mathrm{E}_2[\hat{\Phi}] = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \hat{\Phi}(\rho, \theta)^2 \, \rho \, d\rho \, d\theta, \tag{4.26}$$

one can define the rms of a real-valued function $\hat{\Phi}(\rho, \theta)$ as [40]

$$\mathrm{rms}(\hat{\Phi}) = (\mathrm{E}_2[\hat{\Phi}])^{1/2}. \tag{4.27}$$

Due to the normalisation coefficient c_n^m in Eq. (4.18), we have that [40] $\mathrm{rms}(\hat{\Phi}) = \|\boldsymbol{\alpha}\|$. Note that in this paper we always assume $\alpha_0^0 = 0$ when considering the rms of the phase aberration function, since the piston coefficient does not affect the image quality.

The definition of the rms indicator can be extended to the complex field to evaluate the error $E(\rho, \theta)$ between two GPFs, when these are expressed using Eq. (4.20) as done in Section 4.4.4. We therefore use the following functional,

$$\mathrm{E}_2[E] = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} |E(\rho, \theta)|^2 \, \rho \, d\rho \, d\theta, \tag{4.28}$$

which is real-valued. Using the definitions of rms in Eq. (4.27), and the normalisation in Eq. (4.21), we have that $\mathrm{rms}(E) = \|\boldsymbol{\beta}\|$.

4.D Solving the convex relaxation with cvxopt

Following [242], we can reformulate the complex semidefinite programme in Eq. (4.10) into a real programme, which can be solved using cvxopt [239]. Let $s_0 \in \mathbb{R}$, $S_1 \in \mathbb{R}^{N_\beta \times N_\beta}$ such that $S_1 = S_1^T$, $S_2 \in \mathbb{R}^{N_\beta \times N_\beta}$ such that $S_2 = -S_2^T$, and $\hat{\mathbf{u}} \in \mathbb{R}^{N_m}$ be the variables of the following real semidefinite programme,

$$\begin{aligned} \min\ & s_0 + \lambda_r \operatorname{tr}(S_1) \quad \text{s.t.} \\ & s_0 \geq \|\mathbf{I} - \hat{\mathbf{u}}\| \\ & \begin{bmatrix} S_1 & -S_2 \\ S_2 & S_1 \end{bmatrix} \geq 0, \\ & \operatorname{tr}(D_k S_1) - \operatorname{tr}(F_k S_2) = \hat{u}_k \end{aligned} \tag{4.29}$$

In Eq. (4.29), the last equality must hold for $k = 1, \ldots, N_m$, and $\hat{u}_k$ is an element of $\hat{\mathbf{u}}$. Further, we have $D_k = \mathbf{d}_k \mathbf{d}_k^T + \mathbf{f}_k \mathbf{f}_k^T$ and $F_k = \mathbf{f}_k \mathbf{d}_k^T - \mathbf{d}_k \mathbf{f}_k^T$, where $\mathbf{d}_k$ and $\mathbf{f}_k$ are respectively the real and imaginary parts of the complex vector $\mathbf{a}_k$, i.e., $\mathbf{a}_k = \mathbf{d}_k + i\mathbf{f}_k$. By solving Eq. (4.29), and letting $B = S_1 + iS_2$ one finds a solution to Eq. (4.10). In our case, we used the default settings for cvxopt [239]. A more efficient interior-point method could be implemented by exploiting techniques to handle large second-order cone constraints [210], as one can expect $N_m \gg N_\beta$ for this application.

Chapter 5

Conclusions

5.1 Summary of this research

In this thesis we considered the problem of correcting aberrations in a wavefront sensorless adaptive optics system by maximising an image quality metric, which we model using a quadratic polynomial. In Chapter 2, we developed a procedure to compute the parameters of the polynomial directly from input–output measurements. The procedure was experimentally validated using a second-harmonic microscope and a biologically relevant specimen in Chapter 3.

A closed-form expression to estimate the unknown aberration was derived in Chapter 2. This expression requires a minimum of $N_\alpha + 1$ measurements to estimate N_α Zernike coefficients of the unknown aberration. Each measurement is obtained by applying a trial aberration with the deformable mirror (DM), and by recording the corresponding value of the image quality metric. In Chapter 3, we solve a non-convex optimisation problem to compute the least-squares estimate of the unknown aberration. This algorithm is the preferred method to estimate the aberration, since the expression reported in Chapter 2 does not provide the least-squares estimate.

In Chapter 4, we used a phase retrieval algorithm to correct the aberrations in a wavefront sensorless adaptive optics system. Using the extended Nijboer–Zernike theory [173, 174] (ENZ), the phase retrieval problem is formulated into a matrix rank minimisation problem [175, 176, 177]. A solution of the phase retrieval problem is obtained using PhaseLift [178, 179], a convex relaxation [180, 181, 182] of the rank minimisation problem.

5.2 Limitations encountered

A number of limitations that we encountered in applying wavefront sensorless adaptive optics in a multiphoton microscope have been reported at the end of Section 1.6 in [129]. In particular, we recommend using a DM that is capable of introducing several radians of the first 15 Zernike aberrations. We found that our 17-actuators DM was able to introduce a limited amount of spherical aberration (less than 1 radian) and secondary astigmatism at 900 nm. One should also consider that some of the available stroke of the DM is depleted to correct the static aberrations in the optical system, and therefore is not available for

correcting specimen-induced aberrations. The lack of stroke affects the accuracy of the aberration correction, since when one applies a trial aberration that results in saturation of the DM, the effective aberration introduced in the system is not known with precision. One may be forced to use a smaller rms for the trial aberration, to avoid saturation. Nevertheless, in this case it is worth to note that the Hessian of (3.12) in Chapter 3 is more likely not to be positive definite, so that a non-convex aberration estimation problem must be solved (see Fig.3.1). In this case, an accurate estimate of the aberration is still obtained using the algorithm found in Chapter 3.

A delicate matter is determining the minimum number of measurements that are necessary to estimate N_α Zernike coefficients of the aberration. Assuming that the quadratic polynomial is a valid model of the metric, this minimum number is $N_\alpha + 1$, as determined in Chapter 2, and as experimentally verified also in Chapter 3. The assumption about the validity of the model is certainly satisfied in a number of optical systems, e.g., the ones considered in [155, 128, 157, 158]. In Chapter 3, we detected variations of the parameters of the quadratic polynomial when considering different regions of the specimen, and therefore a single model was not valid throughout all the regions. As a consequence, we were not able to successfully apply the aberration correction in all the regions using the parameters computed from a single region. The authors in [161, 144, 118] report excellent aberration correction when using a minimum of $2N_\alpha + 1$ measurements. In their case, the extra N_α measurements are used to recompute a subset of the parameters from scratch, as outlined in Section 2.4.1 in Chapter 2.

As far as Chapter 4 is concerned, the main limitation was found to be the dynamic range of the CCD camera. We chose the maximum exposure time for which no saturation of the pixels in the CCD occurs when the system is almost diffraction limited. For this exposure time, we measure a non-zero noise level of about 0.035 when no light is incident on the CCD, where 1.0 corresponds to saturation. For the selected exposure time we found that aberrations in the order of 1.5 rms rad or more can lead to a point-spread function that becomes comparable with the background noise, thus affecting the aberration correction accuracy.

5.3 Suggestions for future work

As far as Chapter 2 and Chapter 3 are concerned, we did not consider the problem of detecting and tracking the variations of the parameters of the quadratic polynomial, which is used to model the image quality metric. This topic could be addressed in future research, for example using fault detection techniques or adaptive filtering. It is possible that the variations are due to the structure of the specimen. In this case, this problem has also been detected in direct wavefront sensing using back-scattered illumination light, since in [129, 186] the authors report variations in the sensitivity to the measured Zernike coefficients.

As far as Chapter 4 is concerned, modulating the exposure time could be considered to broaden the range of the aberrations that can be corrected. Assuming that the aberrations are not known *a priori*, an algorithm that obtains the optimal exposure time should be developed, and the pixel measurements should be processed to account for the different exposure times. With more research effort, the phase retrieval algorithm could

be extended and applied in scanning microscopy. On the one hand the benefits would include that the ENZ theory has been developed to provide a vectorial representation of the field [243] and to model high numerical aperture systems [244]. On the other hand, one could exploit recently developed algorithms in convex optimisation [171] to solve the phase retrieval problem efficiently. In addition, one could possibly try to obtain guarantees about the exactness of the retrieval, such as the ones derived in [178, 179].

Eventually, research could aim at estimating both the aberrations and the 3D structure that emits the fluorescence, thereby extending the work of [225] to scanning microscopy.

5.4 Notes about the software

All illustrations in this thesis have been created using Inkscape. The graphs in Chapter 1 and Chapter 4 were created using matplotlib [245], whereas the graphs in Chapter 2 and Chapter 3 were produced using MATLAB [246].

The software that the author developed for this thesis was implemented using C [247], MATLAB [246] and Python [248]. Some notable software packages and applications that the author used during his research are cvxopt [239], CVX [249, 250], YALMIP [193], SeDuMi [194], SDPT3 [251, 252], scipy and numpy [253, 254, 255], IPython [256], comedi [257], and most of the standard GNU/Linux software.

Bibliography

[1] Wikipedia, "Microscope — wikipedia, the free encyclopedia," Available at http://en.wikipedia.org/w/index.php?title=Microscope&oldid=620374043 (2014).

[2] M. W. Davidson, "Introduction to optical microscopy, digital imaging, and photomicrography," Available at http://micro.magnet.fsu.edu/primer/museum/ (2009).

[3] B. Shmaefsky, *Biotechnology 101* (Greenwood Publishing Group, 2006).

[4] J. Pawley, *Handbook of Biological Confocal Microscopy* (Springer, 2006).

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

[6] R. K. Luneburg, *Mathematical Theory of Optics* (University of California Press, 1964).

[7] T. Wilson and C. J. R. Sheppard, *Theory and Practice of Scanning Optical Microscopy* (Academic Press, 1984).

[8] V. N. Mahajan, *Optical Imaging and Aberrations, Part II. Wave Diffraction Optics* (SPIE Press, 2004), 2nd ed.

[9] H. H. Barrett and K. J. Myers, *Foundations of Image Science* (Wiley-Interscience, 2004).

[10] G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge university press, 1966), 2nd ed.

[11] J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company Publishers, 2005), 3rd ed.

[12] P. Török and F.-J. Kao, eds., *Optical Imaging and Microscopy: Techniques and Advanced Systems* (Springer, 2007), 2nd ed.

[13] M. R. Foreman and P. Török, "Computational methods in vectorial imaging," J. Mod. Opt. **58**, 339–364 (2011).

[14] S. Hell and E. H. K. Stelzer, "Properties of a 4Pi confocal fluorescence microscope," J. Opt. Soc. Am. A **9**, 2159–2166 (1992).

[15] A. J. den Dekker and A. van den Bos, "Resolution: a survey," J. Opt. Soc. Am. A **14**, 547–557 (1997).

[16] M. Minsky, "Memoir on inventing the confocal scanning microscope," Scanning **10**, 128–138 (1988).

[17] J. G. White, W. B. Amos, and M. Fordham, "An evaluation of confocal versus conventional imaging of biological structures by fluorescence light microscopy." J. Cell Biol. **105**, 41–48 (1987).

[18] D. M. Shotton, "Confocal scanning optical microscopy and its applications for biological specimens," J. Cell Sci. **94**, 175–206 (1989).

[19] M. Gu and C. J. R. Sheppard, "Confocal fluorescent microscopy with a finite-sized circular detector," J. Opt. Soc. Am. A **9**, 151–153 (1992).

[20] M. Gu and C. J. R. Sheppard, "Effects of a finite-sized pinhole on 3D image formation in confocal two-photon fluorescence microscopy," J. Mod. Opt. **40**, 2009–2024 (1993).

[21] T. Wilson, "Resolution and optical sectioning in the confocal microscope," J. Microsc. **244**, 113–121 (2011).

[22] C. J. R. Sheppard, "Axial resolution of confocal fluorescence microscopy," J. Microsc. **154**, 237–241 (1989).

[23] W. Denk, J. Strickler, and W. Webb, "Two-photon laser scanning fluorescence microscopy," Science **248**, 73–76 (1990).

[24] M. Göppert-Mayer, "Über Elementarakte mit zwei Quantensprüngen," Ann. Phys. **401**, 273–294 (1931).

[25] F. Helmchen and W. Denk, "Deep tissue two-photon microscopy," Nat. Methods **2**, 932–940 (2005).

[26] V. E. Centonze and J. G. White, "Multiphoton excitation provides optical sections from deeper within scattering specimens than confocal imaging," Biophys. J. **75**, 2015 – 2024 (1998).

[27] C. J. R. Sheppard and M. Gu, "Image formation in two-photon fluorescence microscopy," Optik **86**, 104–106 (1990).

[28] G. Cox and C. J. R. Sheppard, "Practical limits of resolution in confocal and nonlinear microscopy," Microsc. Res. Techniq. **63**, 18–22 (2004).

[29] M. J. Booth and T. Wilson, "Refractive-index-mismatch induced aberrations in single-photon and two-photon microscopy and the use of aberration correction," J. Biomed. Opt. **6**, 266–272 (2001).

[30] M. J. Booth, M. A. A. Neil, and T. Wilson, "Aberration correction for confocal imaging in refractive-index-mismatched media," J. Microsc. **192**, 90–98 (1998).

[31] S. Hell, G. Reiner, C. Cremer, and E. H. K. Stelzer, "Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index," J. Microsc. **169**, 391–405 (1993).

[32] H. Jacobsen, P. Hänninen, E. Soini, and S. W. Hell, "Refractive-index-induced aberrations in two-photon confocal fluorescence microscopy," J. Microsc. **176**, 226–230 (1994).

[33] R. Gauderon, P. B. Lukins, and C. J. R. Sheppard, "Effect of a confocal pinhole in two-photon microscopy," Microsc. Res. Techniq. **47**, 210–214 (1999).

[34] G.-m. Dai, *Wavefront Optics for Vision Correction* (SPIE Press, 2008).

[35] R. Tyson, *Principles of Adaptive Optics* (CRC Press, 2010), 3rd ed.

[36] M. J. Booth, "Adaptive optics in microscopy," Phil. Trans. R. Soc. A **365**, 2829–2843 (2007).

[37] J. A. Kubby, ed., *Adaptive Optics for Biological Imaging* (CRC press, 2013).

[38] M. J. Booth, "Adaptive optical microscopy: the ongoing quest for a perfect image," Light Sci. Appl. **3**, e165 (2014).

[39] F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasenkontrastmethode," Physica **1**, 689–704 (1934).

[40] V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," Appl. Opt. **33**, 8121–8124 (1994).

[41] V. N. Mahajan and G.-m. Dai, "Orthonormal polynomials for hexagonal pupils," Opt. Lett. **31**, 2462–2464 (2006).

[42] V. N. Mahajan and G.-m. Dai, "Orthonormal polynomials in wavefront analysis: analytical solution," J. Opt. Soc. Am. A **24**, 2994–3016 (2007).

[43] G.-m. Dai and V. N. Mahajan, "Zernike annular polynomials and atmospheric turbulence," J. Opt. Soc. Am. A **24**, 139–155 (2007).

[44] R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66**, 207–211 (1976).

[45] V. N. Mahajan, "Zernike polynomials and aberration balancing," Proc. SPIE **5173**, 1–17 (2003).

[46] T. S. Ross, "Limitations and applicability of the Maréchal approximation," Appl. Opt. **48**, 1812–1818 (2009).

[47] H. W. Babcock, "The possibility of compensating astronomical seeing," Publ. Astron. Soc. Pac. **65**, 229–236 (1953).

[48] V. P. Linnik, "On the possibility of reducing the influence of atmospheric seeing on the image quality of stars," Opt. Spectrosc. **3**, 401–402 (1957).

[49] G. Vdovin and P. M. Sarro, "Flexible mirror micromachined in silicon," Appl. Opt. **34**, 2968–2972 (1995).

[50] G. Vdovin, O. Soloviev, M. Loktev, and V. Patlan, *OKO Guide to Adaptive Optics* (Flexible Optical B.V., 2013), 4th ed.

[51] B. L. Roux, C. Kulcsár, L. M. Mugnier, T. Fusco, H.-F. Raynaud, and J.-M. Conan, "Optimal control law for classical and multiconjugate adaptive optics," J. Opt. Soc. Am. A **21**, 1261–1276 (2004).

[52] C. Kulcsár, H.-F. Raynaud, C. Petit, J.-M. Conan, and P. V. de Lesegno, "Optimal control, observers and integrators in adaptive optics," Opt. Express **14**, 7464–7476 (2006).

[53] N. Doelman, R. Fraanje, I. Houtzager, and M. Verhaegen, "Adaptive and real-time optimal control for adaptive optics systems," Eur. J. Control **15**, 480–488 (2009).

[54] K. Hinnen, M. Verhaegen, and N. Doelman, "Exploiting the spatiotemporal correlation in adaptive optics using data-driven $\mathcal{H}_2$-optimal control," J. Opt. Soc. Am. A **24**, 1714–1725 (2007).

[55] B. L. Ellerbroek and T. A. Rhoadarmer, "Real-time adaptive optimization of wavefront reconstruction algorithms for closed-loop adaptive optical systems," Proc. SPIE **3353**, 1174–1185 (1998).

[56] J. S. Gibson, C.-C. Chang, and B. L. Ellerbroek, "Adaptive optics: wave-front correction by use of adaptive filtering and control," Appl. Opt. **39**, 2525–2538 (2000).

[57] K. J. G. Hinnen, "Data-driven optimal control for adaptive optics," Ph.D. thesis, Delft University of Technology (2007). [URL].

[58] H. Song, "Model-based control in adaptive optics systems," Ph.D. thesis, Delft University of Technology (2011). [URL].

[59] 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 **19**, 2100–2111 (2002).

[60] R. Fraanje, J. Rice, M. Verhaegen, and N. Doelman, "Fast reconstruction and prediction of frozen flow turbulence based on structured Kalman filtering," J. Opt. Soc. Am. A **27**, A235–A245 (2010).

[61] R. Fraanje, P. Massioni, and M. Verhaegen, "A decomposition approach to distributed control of dynamic deformable mirrors," Int. J. Optomechatronics **4**, 269–284 (2010).

[62] P. Massioni, C. Kulcsár, H.-F. Raynaud, and J.-M. Conan, "Fast computation of an optimal controller for large-scale adaptive optics," J. Opt. Soc. Am. A **28**, 2298–2309 (2011).

[63] C. U. Keller, V. Korkiakoski, N. Doelman, R. Fraanje, R. Andrei, and M. Verhaegen, "Extremely fast focal-plane wavefront sensing for extreme adaptive optics," Proc. SPIE **8447**, 844721 (2012).

[64] V. Korkiakoski, C. U. Keller, N. Doelman, M. Kenworthy, G. Otten, and M. Verhaegen, "Fast & Furious focal-plane wavefront sensing," Appl. Opt. **53**, 4565–4579 (2014).

[65] A. Polo, "Adaptive optics for EUV lithography," Ph.D. thesis, Delft University of Technology (2014). [URL].

[66] A. Haber, A. Polo, I. Maj, S. Pereira, H. Urbach, and M. Verhaegen, "Predictive control of thermally induced wavefront aberrations," Opt. Express **21**, 21530–21541 (2013).

[67] R. Saathof, "Adaptive optics to counteract thermal aberrations," Ph.D. thesis, Delft University of Technology (2013). [URL].

[68] P. S. Salter, M. Baum, I. Alexeev, M. Schmidt, and M. J. Booth, "Exploring the depth range for three-dimensional laser machining with aberration correction," Opt. Express **22**, 17644–17656 (2014).

[69] J. Liang, D. R. Williams, and D. T. Miller, "Supernormal vision and high-resolution retinal imaging through adaptive optics," J. Opt. Soc. Am. A **14**, 2884–2892 (1997).

[70] A. Roorda, F. Romero-Borja, W. J. Donnelly III, H. Queener, T. Hebert, and M. Campbell, "Adaptive optics scanning laser ophthalmoscopy," Opt. Express **10**, 405–412 (2002).

[71] P. Artal, A. Guirao, E. Berrio, and D. R. Williams, "Compensation of corneal aberrations by the internal optics in the human eye," J. Vision **1**, 1–8 (2001).

[72] B. Hermann, E. J. Fernández, A. Unterhuber, H. Sattmann, A. F. Fercher, W. Drexler, P. M. Prieto, and P. Artal, "Adaptive-optics ultrahigh-resolution optical coherencetomography," Opt. Lett. **29**, 2142–2144 (2004).

[73] D. Merino, C. Dainty, A. Bradu, and A. G. Podoleanu, "Adaptive optics enhanced simultaneous *en-face* optical coherence tomography and scanning laser ophthalmoscopy," Opt. Express **14**, 3345–3353 (2006).

[74] R. Zawadzki, S. Jones, S. Olivier, M. Zhao, B. Bower, J. Izatt, S. Choi, S. Laut, and J. Werner, "Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal *in vivo* imaging," Opt. Express **13**, 8532–8546 (2005).

[75] R. J. Zawadzki, S. S. Choi, S. M. Jones, S. S. Oliver, and J. S. Werner, "Adaptive optics-optical coherence tomography: optimizing visualization of microscopic retinal structures in three dimensions," J. Opt. Soc. Am. A **24**, 1373–1383 (2007).

[76] Y. Zhang, B. Cense, J. Rha, R. S. Jonnal, W. Gao, R. J. Zawadzki, J. S. Werner, S. Jones, S. Olivier, and D. T. Miller, "High-speed volumetric imaging of cone photoreceptors with adaptive optics spectral-domain optical coherence tomography," Opt. Express **14**, 4380–4394 (2006).

[77] J. Hartmann, "Bemerkungen über den Bau und die Justirung von Spektrographen," Z. Instrumentenkd. **20**, 17–27 (1900).

[78] R. V. Shack and B. C. Platt, "Abstract. Production and use of a lenticular Hartmann screen," J. Opt. Soc. Am. A **61**, 656–660 (1971).

[79] B. C. Platt and R. Shack, "History and principles of Shack–Hartmann wavefront sensing," J. Refract. Surg. **17**, S573–S577 (2001).

[80] J. Primot, G. Rousset, and J. C. Fontanella, "Deconvolution from wave-front sensing: a new technique for compensating turbulence-degraded images," J. Opt. Soc. Am. A **7**, 1598–1608 (1990).

[81] B. L. Ellerbroek and C. R. Vogel, "Inverse problems in astronomical adaptive optics," Inverse Probl. **25**, 063001 (2009).

[82] G.-m. Dai, "Modified Hartmann–Shack wavefront sensing and iterative wavefront reconstruction," Proc. SPIE **2201**, 562–573 (1994).

[83] G. E. Artzner, "Microlens arrays for shack-hartmann wavefront sensors," Opt. Eng. **31**, 1311–1322 (1992).

[84] G. H. Golub and C. F. Van Loan, *Matrix Computations* (Johns Hopkins University Press, 2013), 4th ed.

[85] G.-m. Dai, "Modal compensation of atmospheric turbulence with the use of Zernike polynomials and Karhunen–Loève functions," J. Opt. Soc. Am. A **12**, 2182–2193 (1995).

[86] G.-m. Dai, "Modal wave-front reconstruction with Zernike polynomials and Karhunen–Loève functions," J. Opt. Soc. Am. A **13**, 1218–1225 (1996).

[87] D. Luke, J. Burke, and R. Lyon, "Optical wavefront reconstruction: Theory and numerical methods," SIAM Rev. **44**, 169–224 (2002).

[88] R. H. Hudgin, "Wave-front reconstruction for compensated imaging," J. Opt. Soc. Am. **67**, 375–378 (1977).

[89] R. H. Hudgin, "Optimal wave-front estimation," J. Opt. Soc. Am. **67**, 378–382 (1977).

[90] D. L. Fried, "Least-square fitting a wave-front distortion estimate to an array of phase-difference measurements," J. Opt. Soc. Am. **67**, 370–375 (1977).

[91] J. Herrmann, "Least-squares wave front errors of minimum norm," J. Opt. Soc. Am. **70**, 28–35 (1980).

[92] E. P. Wallner, "Optimal wave-front correction using slope measurements," J. Opt. Soc. Am. **73**, 1771–1776 (1983).

[93] J. Y. Wang and D. E. Silva, "Wave-front interpretation with Zernike polynomials," Appl. Opt. **19**, 1510–1518 (1980).

[94] W. H. Southwell, "Wave-front estimation from wave-front slope measurements," J. Opt. Soc. Am. **70**, 998–1006 (1980).

[95] R. Cubalchini, "Modal wave-front estimation from phase derivative measurements," J. Opt. Soc. Am. **69**, 972–977 (1979).

[96] 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 **3**, 1852–1861 (1986).

[97] C. C. de Visser and M. Verhaegen, "Wavefront reconstruction in adaptive optics systems using nonlinear multivariate splines," J. Opt. Soc. Am. A **30**, 82–95 (2013).

[98] J. Herrmann, "Cross coupling and aliasing in modal wave-front estimation," J. Opt. Soc. Am. **71**, 989–992 (1981).

[99] J. Antonello, "MSHWFS: A modal Shack–Hartmann wavefront sensor toolbox, version 2dd12134," Available at http://antonello.org (2014).

[100] M. J. Booth, "Adaptive optics in microscopy," in *Optical and Digital Image Processing*, G. Cristòbal, P. Schelkens, and H. Thienpont, eds. (Wiley-VCH Verlag GmbH & Co. KGaA, 2011), chap. 14, pp. 295–322.

[101] C. J. R. Sheppard, "Aberrations in high aperture conventional and confocal imaging systems," Appl. Opt. **27**, 4782–4786 (1988).

[102] T. Wilson and A. R. Carlini, "The effect of aberrations on the axial response of confocal imaging systems," J. Microsc. **154**, 243–256 (1989).

[103] J. P. Rigaut, J. Vassy, P. Herlin, F. Duigou, E. Masson, D. Briane, J. Foucrier, S. Carvajal-Gonzalez, A. M. Downs, and A.-M. Mandard, "Three-dimensional DNA image cytometry by confocal scanning laser microscopy in thick tissue blocks," Cytometry **12**, 511–524 (1991).

[104] K. Carlsson, "The influence of specimen refractive index, detector signal integration, and non-uniform scan speed on the imaging properties in confocal microscopy," J. Microsc. **163**, 167–178 (1991).

[105] C. J. de Grauw, J. M. Vroom, H. T. M. van der Voort, and H. C. Gerritsen, "Imaging properties in two-photon excitation microscopy and effects of refractive-index mismatch in thick specimens," Appl. Opt. **38**, 5995–6003 (1999).

[106] C. J. R. Sheppard and P. Török, "Efficient calculation of electromagnetic diffraction in optical systems using a multipole expansion," J. Mod. Opt. **44**, 803–818 (1997).

[107] P. Török, S. J. Hewlett, and P. Varga, "The role of specimen-induced spherical aberration in confocal microscopy," J. Microsc. **188**, 158–172 (1997).

[108] K. Visscher, G. J. Brakenhoff, and T. D. Visser, "Fluorescence saturation in confocal microscopy," J. Microsc. **175**, 162–165 (1994).

[109] C. J. R. Sheppard and M. Gu, "Aberration compensation in confocal microscopy," Appl. Opt. **30**, 3563–3568 (1991).

[110] M. A. A. Neil, R. Juškaitis, M. J. Booth, T. Wilson, T. Tanaka, and S. Kawata, "Adaptive aberration correction in a two-photon microscope," J. Microsc. **200**, 105–108 (2000).

[111] O. Albert, L. Sherman, G. Mourou, T. B. Norris, and G. Vdovin, "Smart microscope: an adaptive optics learning system for aberration correction in multiphoton confocal microscopy," Opt. Lett. **25**, 52–54 (2000).

[112] L. Sherman, J. Y. Ye, O. Albert, and T. B. Norris, "Adaptive correction of depth-induced aberrations in multiphoton scanning microscopy using a deformable mirror," J. Microsc. **206**, 65–71 (2002).

[113] P. Marsh, D. Burns, and J. Girkin, "Practical implementation of adaptive optics in multiphoton microscopy," Opt. Express **11**, 1123–1130 (2003).

[114] M. J. Booth, M. A. A. Neil, R. Juškaitis, and T. Wilson, "Adaptive aberration correction in a confocal microscope," Proc. Natl. Acad. Sci. USA **99**, 5788–5792 (2002).

[115] M. Schwertner, M. J. Booth, M. A. A. Neil, and T. Wilson, "Measurement of specimen-induced aberrations of biological samples using phase stepping interferometry," J. Microsc. **213**, 11–19 (2004).

[116] M. Schwertner, M. Booth, and T. Wilson, "Characterizing specimen induced aberrations for high NA adaptive optical microscopy," Opt. Express **12**, 6540–6552 (2004).

[117] M. Schwertner, M. J. Booth, and T. Wilson, "Simulation of specimen-induced aberrations for objects with spherical and cylindrical symmetry," J. Microsc. **215**, 271–280 (2004).

[118] J. Zeng, P. Mahou, M.-C. Schanne-Klein, E. Beaurepaire, and D. Débarre, "3D resolved mapping of optical aberrations in thick tissues," Biomed. Opt. Express **3**, 1898–1913 (2012).

[119] R. D. Simmonds, T. Wilson, and M. J. Booth, "Effects of aberrations and specimen structure in conventional, confocal and two-photon fluorescence microscopy," J. Microsc. **245**, 63–71 (2012).

[120] Z. Kam, P. Kner, D. Agard, and J. W. Sedat, "Modelling the application of adaptive optics to wide-field microscope live imaging," J. Microsc. **226**, 33–42 (2007).

[121] R. D. Simmonds and M. J. Booth, "Modelling of multi-conjugate adaptive optics for spatially variant aberrations in microscopy," J. Opt. **15**, 094010 (2013).

[122] B. L. Ellerbroek, "First-order performance evaluation of adaptive-optics systems for atmospheric-turbulence compensation in extended-field-of-view astronomical telescopes," J. Opt. Soc. Am. A **11**, 783–805 (1994).

[123] F. Roddier, "The effects of atmospheric turbulence in optical astronomy," in *Progress in Optics*, vol. 19, E. Wolf, ed. (Elsevier, 1981), chap. 5, pp. 281–376.

[124] V. I. Tatarski, *The Effects of the Turbulent Atmosphere on Wave Propagation* (Israel Program for Scientific Translations, 1971).

[125] D. L. Fried, "Statistics of a geometric representation of wavefront distortion," J. Opt. Soc. Am. **55**, 1427–1431 (1965).

[126] N. Olivier, D. Débarre, and E. Beaurepaire, "Dynamic aberration correction for multiharmonic microscopy," Opt. Lett. **34**, 3145–3147 (2009).

[127] L. Poyneer, M. van Dam, and J.-P. Véran, "Experimental verification of the frozen flow atmospheric turbulence assumption with use of astronomical adaptive optics telemetry," J. Opt. Soc. Am. A **26**, 833–846 (2009).

[128] M. J. Booth, "Wavefront sensorless adaptive optics for large aberrations," Opt. Lett. **32**, 5–7 (2007).

[129] T. van Werkhoven, "Lasers, lenses and light curves: adaptive optics microscopy and peculiar transiting exoplanets," Ph.D. thesis, Leiden University (2014). [URL].

[130] M. Feierabend, M. Rückel, and W. Denk, "Coherence-gated wave-front sensing in strongly scattering samples," Opt. Lett. **29**, 2255–2257 (2004).

[131] M. Rueckel, J. A. Mack-Bucher, and W. Denk, "Adaptive wavefront correction in two-photon microscopy using coherence-gated wavefront sensing," Proc. Natl. Acad. Sci. USA **103**, 17137–17142 (2006).

[132] M. Rueckel and W. Denk, "Properties of coherence-gated wavefront sensing," J. Opt. Soc. Am. A **24**, 3517–3529 (2007).

[133] 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 multiphoton microscopy," Opt. Express **22**, 9715–9733 (2014).

[134] R. J. Noll, "Phase estimates from slope-type wave-front sensors," J. Opt. Soc. Am. **68**, 139–140 (1978).

[135] P. Artal, S. Marcos, R. Navarro, and D. R. Williams, "Odd aberrations and double-pass measurements of retinal image quality," J. Opt. Soc. Am. A **12**, 195–201 (1995).

[136] J. W. Cha, J. Ballesta, and P. T. C. So, "Shack–Hartmann wavefront-sensor-based adaptive optics system for multiphoton microscopy," J. Biomed. Opt. **15**, 046022–046022–10 (2010).

[137] S. A. Rahman and M. J. Booth, "Direct wavefront sensing in adaptive optical microscopy using backscattered light," Appl. Opt. **52**, 5523–5532 (2013).

[138] M. Shaw, K. O'Holleran, and C. Paterson, "Investigation of the confocal wavefront sensor and its application to biological microscopy," Opt. Express **21**, 19353–19362 (2013).

[139] O. Azucena, J. Crest, J. Cao, W. Sullivan, P. Kner, D. Gavel, D. Dillon, S. Olivier, and J. Kubby, "Wavefront aberration measurements and corrections through thick tissue using fluorescent microsphere reference beacons," Opt. Express **18**, 17521–17532 (2010).

[140] R. Aviles-Espinosa, J. Andilla, R. Porcar-Guezenec, O. E. Olarte, M. Nieto, X. Levecq, D. Artigas, and P. Loza-Alvarez, "Measurement and correction of *in vivo* sample aberrations employing a nonlinear guide-star in two-photon excited fluorescence microscopy," Biomed. Opt. Express **2**, 3135–3149 (2011).

[141] X. Tao, B. Fernandez, O. Azucena, M. Fu, D. Garcia, Y. Zuo, D. C. Chen, and J. Kubby, "Adaptive optics confocal microscopy using direct wavefront sensing," Opt. Lett. **36**, 1062–1064 (2011).

[142] X. Tao, A. Norton, M. Kissel, O. Azucena, and J. Kubby, "Adaptive optical two-photon microscopy using autofluorescent guide stars," Opt. Lett. **38**, 5075–5078 (2013).

[143] X. Tao, Z. Dean, C. Chien, O. Azucena, D. Bodington, and J. Kubby, "Shack–Hartmann wavefront sensing using interferometric focusing of light onto guide-stars," Opt. Express **21**, 31282–31292 (2013).

[144] A. Facomprez, E. Beaurepaire, and D. Débarre, "Accuracy of correction in modal sensorless adaptive optics," Opt. Express **20**, 2598–2612 (2012).

[145] A. J. Wright, S. P. Poland, J. M. Girkin, C. W. Freudiger, C. L. Evans, and X. S. Xie, "Adaptive optics for enhanced signal in CARS microscopy," Opt. Express **15**, 18209–18219 (2007).

[146] A. J. Wright, D. Burns, B. A. Patterson, S. P. Poland, G. J. Valentine, and J. M. Girkin, "Exploration of the optimisation algorithms used in the implementation of adaptive optics in confocal and multiphoton microscopy," Microsc. Res. Techniq. **67**, 36–44 (2005).

[147] S. P. Poland, A. J. Wright, and J. M. Girkin, "Evaluation of fitness parameters used in an iterative approach to aberration correction in optical sectioning microscopy," Appl. Opt. **47**, 731–736 (2008).

[148] W. Lubeigt, S. P. Poland, G. J. Valentine, A. J. Wright, J. M. Girkin, and D. Burns, "Search-based active optic systems for aberration correction in time-independent applications," Appl. Opt. **49**, 307–314 (2010).

[149] L. P. Murray, "Smart optics: Wavefront sensor-less adaptive optics - Image correction through sharpness maximisation," Ph.D. thesis, National University of Ireland, Galway. (2006). [URL].

[150] M. A. Vorontsov, "Decoupled stochastic parallel gradient descent optimization for adaptive optics: integrated approach for wave-front sensor information fusion," J. Opt. Soc. Am. A **19**, 356–368 (2002).

[151] G. Vdovin, "Optimization-based operation of micromachined deformable mirrors," Proc. SPIE **3353**, 902–909 (1998).

[152] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder–Mead simplex method in low dimensions," SIAM J. Optimiz. **9**, 112–147 (1998).

[153] L. P. Murray, J. C. Dainty, and E. Daly, "Wavefront correction through image sharpness maximisation," Proc. SPIE **5823**, 40–47 (2005).

[154] C. Bourgenot, C. Saunter, G. Love, and J. Girkin, "Comparison of closed loop and sensorless adaptive optics in widefield optical microscopy," J. Eur. Opt. Soc. **8** (2013).

[155] M. J. Booth, "Wave front sensor-less adaptive optics: a model-based approach using sphere packings," Opt. Express **14**, 1339–1352 (2006).

[156] H. Song, R. Fraanje, G. Schitter, H. Kroese, G. Vdovin, and M. Verhaegen, "Model-based aberration correction in a closed-loop wavefront-sensor-less adaptive optics system," Opt. Express **18**, 24070–24084 (2010).

[157] H. Linhai and C. Rao, "Wavefront sensorless adaptive optics: a general model-based approach," Opt. Express **19**, 371–379 (2011).

[158] J. Antonello, M. Verhaegen, R. Fraanje, T. van Werkhoven, H. C. Gerritsen, and C. U. Keller, "Semidefinite programming for model-based sensorless adaptive optics," J. Opt. Soc. Am. A **29**, 2428–2438 (2012).

[159] D. Débarre, M. J. Booth, and T. Wilson, "Image based adaptive optics through optimisation of low spatial frequencies," Opt. Express **15**, 8176–8190 (2007).

[160] D. Débarre, E. J. Botcherby, M. J. Booth, and T. Wilson, "Adaptive optics for structured illumination microscopy," Opt. Express **16**, 9290–9305 (2008).

[161] D. Débarre, E. J. Botcherby, T. Watanabe, S. Srinivas, M. J. Booth, and T. Wilson, "Image-based adaptive optics for two-photon microscopy," Opt. Lett. **34**, 2495–2497 (2009).

[162] A. Jesacher, A. Thayil, K. Grieve, D. Débarre, T. Watanabe, T. Wilson, S. Srinivas, and M. Booth, "Adaptive harmonic generation microscopy of mammalian embryos," Opt. Lett. **34**, 3154–3156 (2009).

[163] J. Antonello, T. van Werkhoven, M. Verhaegen, H. H. Truong, C. U. Keller, and H. C. Gerritsen, "Optimization-based wavefront sensorless adaptive optics for multiphoton microscopy," J. Opt. Soc. Am. A **31**, 1337–1347 (2014).

[164] T. J. Gould, D. Burke, J. Bewersdorf, and M. J. Booth, "Adaptive optics enables 3D STED microscopy in aberrating specimens," Opt. Express **20**, 20998–21009 (2012).

[165] T. J. Gould, E. B. Kromann, D. Burke, M. J. Booth, and J. Bewersdorf, "Auto-aligning stimulated emission depletion microscope using adaptive optics," Opt. Lett. **38**, 1860–1862 (2013).

[166] N. Ji, D. E. Milkie, and E. Betzig, "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues," Nat. Methods **7**, 141–147 (2009).

[167] D. E. Milkie, E. Betzig, and N. Ji, "Pupil-segmentation-based adaptive optical microscopy with full-pupil illumination," Opt. Lett. **36**, 4206–4208 (2011).

[168] A. Leray and J. Mertz, "Rejection of two-photon fluorescence background in thick tissue by differential aberration imaging," Opt. Express **14**, 10565–10573 (2006).

[169] D. Débarre, A. Facomprez, and E. Beaurepaire, "Assessing correction accuracy in image-based adaptive optics," Proc. SPIE **8253**, 82530F (2012).

[170] M. J. Booth, "Wavefront sensorless adaptive optics, modal wavefront sensing, and sphere packings," Proc. SPIE **5553**, 150–158 (2004).

[171] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, 2004).

[172] H. Stark, ed., *Image Recovery: Theory and Application* (Academic Press, 1987).

[173] A. J. E. M. Janssen, "Extended Nijboer–Zernike approach for the computation of optical point-spread functions," J. Opt. Soc. Am. A **19**, 849–857 (2002).

[174] J. Braat, P. Dirksen, and A. J. E. M. Janssen, "Assessment of an extended Nijboer–Zernike approach for the computation of optical point-spread functions," J. Opt. Soc. Am. A **19**, 858–870 (2002).

[175] M. Mesbahi and G. Papavassilopoulos, "On the rank minimization problem over a positive semidefinite linear matrix inequality," IEEE Trans. Automat. Contr. **42**, 239–243 (1997).

[176] M. Fazel, H. Hindi, and S. Boyd, "A rank minimization heuristic with application to minimum order system approximation," Proc. Amer. Contr. Conf. **6**, 4734–4739 (2001).

[177] M. Fazel, "Matrix rank minimization with applications," Ph.D. thesis, Stanford University (2002). [URL].

[178] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," SIAM J. Imaging Sci. **6**, 199–225 (2013).

[179] E. J. Candès, T. Strohmer, and V. Voroninski, "PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming," Commun. Pur. Appl. Math. **66**, 1241–1274 (2013).

[180] Y. Nesterov, "Semidefinite relaxation and nonconvex quadratic optimization," Optim. Methods Softw. **9**, 141–160 (1998).

[181] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization* (Society for Industrial and Applied Mathematics, 2001).

[182] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming* (Society for Industrial and Applied Mathematics, 2001), 3rd ed.

[183] M. Booth, T. Wilson, H.-B. Sun, T. Ota, and S. Kawata, "Methods for the characterization of deformable membrane mirrors," Appl. Opt. **44**, 5131–5139 (2005).

[184] D. Débarre, T. Vieille, and E. Beaurepaire, "Simple characterisation of a deformable mirror inside a high numerical aperture microscope using phase diversity," J. Microsc. **244**, 136–143 (2011).

[185] J. Hardy, *Adaptive Optics for Astronomical Telescopes* (Oxford University Press, 1998).

[186] S. A. Rahman and M. J. Booth, "Adaptive optics for high-resolution microscopy: wave front sensing using back scattered light," Proc. SPIE **8253**, 82530I (2012).

[187] T. van Werkhoven, H. Truong, J. Antonello, R. Fraanje, H. Gerritsen, M. Verhaegen, and C. Keller, "Coherence-gated wavefront sensing for microscopy using fringe analysis," Proc. SPIE **8253**, 82530E (2012).

[188] J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," J. Opt. Soc. Am. A **20**, 609–620 (2003).

[189] B. Wang and M. J. Booth, "Optimum deformable mirror modes for sensorless adaptive optics," Opt. Commun. **282**, 4467–4474 (2009).

[190] J. Antonello, R. Fraanje, H. Song, M. Verhaegen, H. Gerritsen, C. U. Keller, and T. van Werkhoven, "Data driven identification and aberration correction for model-based sensorless adaptive optics," Proc. SPIE **8436**, 84360S (2012).

[191] L. Vandenberghe and S. Boyd, "Semidefinite programming," SIAM Rev. **38**, 49–95 (1996).

[192] M. Verhaegen and V. Verdult, *Filtering and System Identification: A Least Squares Approach* (Cambridge University Press, 2007).

[193] J. Löfberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," Proc. IEEE Int. Symp. CACSD pp. 284–289 (2004).

[194] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," Optim. Methods Softw. **11**, 625–653 (1999).

[195] Y. Labit, D. Peaucelle, and D. Henrion, "SeDuMi Interface 1.02: A tool for solving LMI problems with SeDuMi," Proc. IEEE Int. Symp. CACSD pp. 272–277 (2002).

[196] M. Loktev, O. Soloviev, and G. Vdovin, *Adaptive Optics Guide* (Flexible Optical B.V., 2008), 3rd ed.

[197] M. A. A. Neil, M. J. Booth, and T. Wilson, "New modal wave-front sensor: a theoretical analysis," J. Opt. Soc. Am. A **17**, 1098–1107 (2000).

[198] P. J. Campagnola, A. Lewis, L. M. Loew, H. A. Clark, and W. A. Mohler, "Second-harmonic imaging microscopy of living cells," J. Biomed. Opt. **6**, 277–286 (2001).

[199] J. M. Bueno, E. J. Gualda, and P. Artal, "Adaptive optics multiphoton microscopy to study *ex vivo* ocular tissues," J. Biomed. Opt. **15**, 066004 (2010).

[200] M. J. Booth, M. A. A. Neil, and T. Wilson, "New modal wave-front sensor: application to adaptive confocal fluorescence microscopy and two-photon excitation fluorescence microscopy," J. Opt. Soc. Am. A **19**, 2112–2120 (2002).

[201] H. W. Yoo, M. Verhaegen, M. van Royen, and G. Schitter, "Automated adjustment of aberration correction in scanning confocal microscopy," Proc. IEEE Int. IMTC pp. 1083–1088 (2012).

[202] A. Thayil and M. J. Booth, "Self calibration of sensorless adaptive optical microscopes," J. Eur. Opt. Soc. **6** (2011).

[203] C. Paterson, I. Munro, and J. Dainty, "A low cost adaptive optics system using a membrane mirror," Opt. Express **6**, 175–185 (2000).

[204] E. Fernandez and P. Artal, "Membrane deformable mirror for adaptive optics: performance limits in visual optics," Opt. Express **11**, 1056–1069 (2003).

[205] P. C. Hansen, *Discrete Inverse Problems* (Society for Industrial and Applied Mathematics, 2010).

[206] D. Torrieri, "Statistical theory of passive location systems," IEEE Trans. Aeros. Electron. Syst. **AES-20**, 183–198 (1984).

[207] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," IEEE Trans. Signal Process. **56**, 1770–1778 (2008).

[208] J. J. Moré, "Generalizations of the trust region problem," Optim. Methods Softw. **2**, 189–209 (1993).

[209] M. Andersen, J. Dahl, and L. Vandenberghe, "CVXOPT: A Python package for convex optimization, version 1.1.6," Available at http://cvxopt.org (2013).

[210] M. Andersen, J. Dahl, Z. Liu, and L. Vandenberghe, "Interior-point methods for large-scale cone programming," in *Optimization for Machine Learning*, S. Sra, S. Nowozin, and S. J. Wright, eds. (MIT Press, 2011), pp. 55–83.

[211] R. P. Millane, "Phase retrieval in crystallography and optics," J. Opt. Soc. Am. A **7**, 394–411 (1990).

[212] J. C. Dainty and J. R. Fienup, "Phase retrieval and image reconstruction for astronomy," in *Image Recovery: Theory and Application*, H. Stark, ed. (Academic Press, 1987), pp. 231–275.

[213] D. J. Lee and A. M. Weiner, "Optical phase imaging using a synthetic aperture phase retrieval technique," Opt. Express **22**, 9380–9394 (2014).

[214] B. M. Hanser, M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, "Phase retrieval for high-numerical-aperture optical systems," Opt. Lett. **28**, 801–803 (2003).

[215] E. B. Kromann, T. J. Gould, M. F. Juette, J. E. Wilhjelm, and J. Bewersdorf, "Quantitative pupil analysis in stimulated emission depletion microscopy using phase retrieval," Opt. Lett. **37**, 1805–1807 (2012).

[216] M. Foreman, C. Giusca, P. Török, and R. Leach, "Phase-retrieved pupil function and coherent transfer function in confocal microscopy," J. Microsc. **251**, 99–107 (2013).

[217] P. Kner, "Phase diversity for three-dimensional imaging," J. Opt. Soc. Am. A **30**, 1980–1987 (2013).

[218] S. Liu, E. B. Kromann, W. D. Krueger, J. Bewersdorf, and K. A. Lidke, "Three dimensional single molecule localization using a phase retrieved pupil function," Opt. Express **21**, 29462–29487 (2013).

[219] R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," Opt. Eng. **21**, 829–832 (1982).

[220] P. Dirksen, J. Braat, A. J. E. M. Janssen, and C. Juffermans, "Aberration retrieval using the extended Nijboer–Zernike approach," J. Micro/Nanolith. MEMS MOEMS **2**, 61–68 (2003).

[221] C. van der Avoort, J. J. M. Braat, P. Dirksen, and A. J. E. M. Janssen, "Aberration retrieval from the intensity point-spread function in the focal region using the extended Nijboer–Zernike approach," J. Mod. Opt. **52**, 1695–1728 (2005).

[222] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging," arXiv:1402.7350 (2014). To appear in IEEE Signal Process. Mag.

[223] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972).

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

[225] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations by using phase diversity," J. Opt. Soc. Am. A **9**, 1072–1085 (1992).

[226] M. G. Löfdahl and G. B. Scharmer, "Wavefront sensing and image restoration from focused and defocused solar images." Astron. Astrophys. Suppl. **107**, 243–264 (1994).

[227] D. Youla and H. Webb, "Image restoration by the method of convex projections: Part 1–theory," IEEE Trans. Med. Imaging **1**, 81–94 (1982).

[228] A. Levi and H. Stark, "Image restoration by the method of generalized projections with application to restoration from magnitude," J. Opt. Soc. Am. A **1**, 932–943 (1984).

[229] H. H. Bauschke, P. L. Combettes, and D. R. Luke, "Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization," J. Opt. Soc. Am. A **19**, 1334–1345 (2002).

[230] V. Elser, "Phase retrieval by iterated projections," J. Opt. Soc. Am. A **20**, 40–55 (2003).

[231] J. Sanz, "Mathematical considerations for the problem of Fourier transform phase retrieval from magnitude," SIAM J. Appl. Math. **45**, 651–664 (1985).

[232] Y. C. Eldar and S. Mendelson, "Phase retrieval: Stability and recovery guarantees," Appl. Comput. Harmon. A. **36**, 473–494 (2014).

[233] A. Chai, M. Moscoso, and G. Papanicolaou, "Array imaging using intensity-only measurements," Inverse Probl. **27**, 015005 (2011).

[234] I. Waldspurger, A. d'Aspremont, and S. Mallat, "Phase recovery, MaxCut and complex semidefinite programming," arXiv:1206.0102 (2012). To appear in Math. Program.

[235] F. Fogel, I. Waldspurger, and A. d'Aspremont, "Phase retrieval for imaging problems," arXiv:1304.7735 (2013).

[236] T. P. Zielinski, "Robust image-based wavefront sensing," Ph.D. thesis, University of Rochester (2011). [URL].

[237] J. J. M. Braat, S. van Haver, A. J. E. M. Janssen, and P. Dirksen, "Assessment of optical systems by means of point-spread functions," in *Progress in Optics*, vol. 51, E. Wolf, ed. (Elsevier, 2008), chap. 6, pp. 349–468.

[238] S. van Haver, "The extended Nijboer–Zernike diffraction theory and its applications," Ph.D. thesis, Delft University of Technology (2010). [URL].

[239] M. Andersen, J. Dahl, and L. Vandenberghe, "CVXOPT: A Python package for convex optimization, version 1.1.7," Available at http://cvxopt.org (2014).

[240] K. Itoh, "Analysis of the phase unwrapping algorithm," Appl. Opt. **21**, 2470–2470 (1982).

[241] O. T. A. Janssen, S. van Haver, A. J. E. M. Janssen, J. J. M. Braat, H. P. Urbach, and S. F. Pereira, "Extended Nijboer–Zernike (ENZ) based mask imaging: Efficient coupling of electromagnetic field solvers and the ENZ imaging algorithm," Proc. SPIE **6924**, 692410 (2008).

[242] M. X. Goemans and D. P. Williamson, "Approximation algorithms for Max-3-Cut and other problems via complex semidefinite programming," J. Comput. Syst. Sci. **68**, 442–470 (2004).

[243] J. J. M. Braat, P. Dirksen, A. J. E. M. Janssen, and A. S. van de Nes, "Extended Nijboer–Zernike representation of the vector field in the focal region of an aberrated high-aperture optical system," J. Opt. Soc. Am. A **20**, 2281–2292 (2003).

[244] S. van Haver, J. Braat, P. Dirksen, and A. Janssen, "High-NA aberration retrieval with the extended Nijboer–Zernike vector diffraction theory," J. Eur. Opt. Soc. **1** (2006).

[245] J. D. Hunter, "Matplotlib: A 2D graphics environment," IEEE Comput. Sci. Eng. **9**, 90–95 (2007).

[246] MATLAB, *version 7.10.0 (R2010a)* (The MathWorks Inc., 2010).

[247] B. W. Kernighan and D. M. Ritchie, *The C Programming Language* (Prentice-Hall, 1988), 2nd ed.

[248] Python, *version 3.4.1* (The Python Software Foundation, 2014). Available at http://www.python.org.

[249] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," Available at http://cvxr.com/cvx (2014).

[250] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, eds. (Springer-Verlag Limited, 2008), Lecture Notes in Control and Information Sciences, pp. 95–110.

[251] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3 – a Matlab software package for semidefinite programming, Version 1.3," Optim. Methods Softw. **11**, 545–581 (1999).

[252] R. H. Tütüncü, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," Math. Program. **95**, 189–217 (2003).

[253] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," Available at http://www.scipy.org/ (2001–).

[254] T. E. Oliphant, "Python for scientific computing," IEEE Comput. Sci. Eng. **9**, 10–20 (2007).

[255] S. van der Walt, S. C. Colbert, and G. Varoquaux, "The NumPy Array: A structure for efficient numerical computation," IEEE Comput. Sci. Eng. **13**, 22–30 (2011).

[256] F. Pérez and B. E. Granger, "IPython: A system for interactive scientific computing," IEEE Comput. Sci. Eng. **9**, 21–29 (2007).

[257] D. Schleef, F. M. Hess, and I. Abbott, "Comedi: Linux control and measurement interface, version 0.7.76," Available at http://www.comedi.org/ (2008).

Summary

Optimisation-based wavefront sensorless adaptive optics for microscopy

Jacopo Antonello

Microscopy is an essential tool for life sciences. Thanks to the development of confocal and multiphoton microscopy, scientists are able to obtain high-resolution 3D views of biological specimens. Nevertheless, spatial variations in the index of refraction within specimens cause aberrations that degrade the quality of the 3D views. One can tackle this issue by implementing adaptive optics (AO) techniques, whereby an active element such as a deformable mirror (DM) is used to suppress the aberrations.

In this thesis we consider the problem of estimating aberrations in microscopy. Well-established methods to measure aberrations, such as Shack–Hartmann wavefront sensing, cannot be easily applied due to the lack of well-defined reference wavefronts within specimens. Instead, one can consider wavefront sensorless AO (WFSless-AO), where aberrations are estimated indirectly using a suitable image quality metric. In practice, a series of trial aberration corrections are applied with the DM until the image quality metric is maximised. One can reduce the number of necessary trial corrections by modelling the image quality metric, so that the overall image acquisition time is minimised, and side effects such as photobleaching and phototoxicity are curtailed.

Quadratic polynomials have been used extensively to model image quality metrics in microscopy. In the first part of this thesis, the problem of computing the parameters of the polynomial directly from input–output measurements is solved using a mathematical optimisation. Once the parameters are known, the aberration estimation problem is formulated into a linear least-squares optimisation, which requires a minimum of $N + 1$ trial corrections to estimate N orthogonal aberration modes, such as Zernike polynomials. Both the computation of the parameters and the aberration estimation are validated experimentally using an optical breadboard.

In the second part of this thesis, we implement a WFSless-AO algorithm in a second-harmonic microscope. To achieve a more refined aberration correction, we compute the least-squares estimate of the aberration by solving a non-convex optimisation problem. Aberration correction experiments are performed using a biologically relevant specimen.

In the last part of this thesis, we consider using a phase retrieval algorithm to correct aberrations. We propose an algorithm that uses three measurements of the point-spread

function of the optical system. The phase retrieval problem is formulated using the extended Nijboer–Zernike theory, and it is solved using PhaseLift, a signal recovery method based on convex optimisation. The feasibility of this approach is demonstrated by performing aberration correction experiments using an optical breadboard.

Samenvatting

Optimalisatiegebaseerde golffrontsensorloze adaptieve optiek voor microscopie

Jacopo Antonello

Microscopie is een belangrijk gereedschap voor wetenschappers, die confocale- en multifotonmicroscopen gebruiken om 3D hogeresolutieafbeeldingen van biologische monsters te maken. Helaas wordt de resolutie van de afbeeldingen door aberraties beperkt. De aberraties worden veroorzaakt omdat de brekingsindex van biologische monsters niet homogeen is. Adaptieve optiek is een oplossing voor dit probleem waarbij een actief element zoals een vervormbare spiegel (VS) gebruikt wordt om de aberraties te corrigeren.

Dit proefschrift gaat over het schatten van aberraties in microscopen. Gewone oplossingen zoals het gebruik van een Shack–Hartmann golffrontsensor zijn niet geschikt voor microscopie omdat een referentiegolffront binnen biologische monsters niet altijd beschikbaar is. Een andere oplossing is golffrontsensorloze adaptieve optiek (GFSloos-AO) te gebruiken waarbij aberraties indirect met een afbeeldingskwaliteitmetriek geschat worden. In de praktijk wordt een aantal pogingen gedaan om de aberratie te corrigeren met de VS totdat de afbeeldingskwaliteitmetriek geoptimaliseerd is. Dankzij een model van de afbeeldingskwaliteitmetriek kan het aantal pogingen verlaagd worden zodat de tijd om een volledige afbeelding te maken geminimaliseerd wordt en negatieve neveneffecten zoals fototoxiciteit en fotobleking worden verminderd.

Kwadratische polynomen werden uitgebreid gebruikt om afbeeldingskwaliteitmetrieken van microscopen te modelleren. In het eerste deel van dit proefschrift wordt het schatten van de polynoomparameters direct uit input–output metingen opgelost met een wiskundige optimalisatie. Nadat de polynoomparameters geschat zijn schatten we de aberratie met een stelsel lineaire vergelijkingen dat minimaal $N + 1$ poging aberratiecorrecties nodig heeft om N orthogonale aberratiefuncties te schatten, zoals Zernikepolynomen. Zowel het schatten van de polynoomparameters en het schatten van de abberratie worden experimenteel geëvalueerd met een optische opstelling.

In het tweede deel van dit proefschrift implementeren we een GFSloos-AO algoritme op een second-harmonic microscoop. Om een betere correctie van de aberratie uit te voeren berekenen we de kleinste-kwadraten schatter van de aberratie die door een niet-convex optimalisatieprobleem geschat kan worden. We voeren een aantal aberratiecorrectie-experimenten uit met een biologisch representatief monster.

In het laatste deel van het proefschrift gebruiken we een fase terugvinden algoritme om de aberratie te corrigeren. We stellen het fase terugvinden probleem op met de uitgebreide Nijboer–Zernike theorie en we lossen het op met PhaseLift, een signaalterugvindenalgoritme gebaseerd op convexe optimisatie. We voeren een aantal aberratiecorrectie-experimenten uit met een optisch opstelling om het algoritme te evalueren.

Curriculum vitae

Jacopo Antonello was born in Camposampiero, Italy, in 1984. After finishing his high school diploma in 2003, he applied to the Polytechnic University of Milan, Italy, where he obtained his BSc and MSc degrees in Engineering of Computer Systems respectively in 2006 and in 2009. His master thesis was entitled "Nonlinear filter design using sum of squares optimisations".

He started to work as a PhD candidate at the Delft Center for Systems and Control in TU Delft, The Netherlands, in September 2009. During his PhD he developed and experimentally validated algorithms for the correction of aberrations in optical systems.