

OPTIMIZATION-BASED ADAPTIVE OPTICS FOR OPTICAL COHERENCE TOMOGRAPHY

OPTIMIZATION-BASED ADAPTIVE OPTICS FOR OPTICAL COHERENCE TOMOGRAPHY

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 woensdag 15 maart 2017 om 10:00 uur

door

Hans Roeland Geert Wim VERSTRAETE

Master of Science in Control Systems, Imperial College London, Verenigd Koninkrijk
geboren te Brasschaat, België.

Dit proefschrift is goedgekeurd door de promotor:

Prof. dr. ir. M. Verhaegen

Copromotor: Dr. J. Kalkman

Copromotor: Dr. -Ing. S. Wahls

Samenstelling promotiecommissie:

Rector Magnificus,	voorzitter
Prof. dr. ir. M. Verhaegen,	Technische Universiteit Delft, promotor
Dr. J. Kalkman,	Technische Universiteit Delft, copromotor
Dr. -Ing. S. Wahls,	Technische Universiteit Delft, copromotor

Onafhankelijke commissieleden:

Prof. dr. M. J. Booth,	Oxford University
Prof. dr. A. G. J. M. van Leeuwen,	Universiteit van Amsterdam
Prof. dr. N. J. Doelman,	Universiteit Leiden
Prof. dr. H. P. Urbach,	Technische Universiteit Delft
Prof. dr. ir. B. De Schutter,	Technische Universiteit Delft
Prof. dr. G. V. Vdovine,	Technische Universiteit Delft, reservelid



Rijksdienst voor Ondernemend
Nederland

Keywords: Costly and noisy optimization, adaptive optics, optical coherence tomography, machine learning

Front & Back: Artificial representation of a human iris.

Copyright © 2017 by H. R. G. W. Verstraete

ISBN 978-94-92516-40-4

An electronic version of this dissertation is available at
<http://repository.tudelft.nl/>.

*There is only one real failure possible;
and that is, not to be true to the best one knows.*

Frederic William Farrar

ACKNOWLEDGMENTS

This dissertation summarizes the work done throughout my four years in the DCSC department of the Delft University of Technology. I have had the opportunity to work with many great people and I have enjoyed every day of it. Many people were involved directly and indirectly and have contributed to this dissertation. I am very grateful for all their contributions, big and small.

In the first place, I would like to thank my promotor Prof. Michel Verhaegen for an unforgettable four years. Michel has given me the freedom to work on interesting topics with great people in the most extraordinary locations. I would like to thank my co-promotor Jeroen Kalkman for introducing me to the world of optics. Jeroen gladly shared his knowledge of optics and optical coherence tomography in our shared office. Thank you Jeroen for going with me through the struggle that building an adaptive optics optical coherence tomography setup is. I also thank my second co-promotor Sander Wahls. Without the insights and feedback of these three people this thesis would have looked entirely different. Together we did some great things. Thank you.

I would also like to thank the secretary Kitty, Marieke, Heleen and Kiran. They were always there to help with countless matters. Additionally, I would like to thank Ditske, for keeping track of my working hours and payments of orders. I am grateful to Kees Slinkman and Will van Geest for helping me out with technical problems. Thank you for your support and translation stage for my lemon and thank you for reminding me that a good set of hands and some creativity can solve a lot of problems. I thank our department head prof. Hans Hellendoorn for the occasional small talk and assistance with the graduate school.

I would like to thank Laurens Blik for solving various problems together with me. In the first year, Jacopo Antonello, Han Woong Yoo and me worked in the same optics lab. I am grateful to them for spending hours in the darkness with me. I had a great time working with all my colleagues in DCSC.

As a side-note, I absolutely enjoyed the Italian and English cooking of Paolo and Dean. I had a lot fun skiing with Subramanya and Sachin. I enjoyed the evenings out in Padova with Gleb, Oleg and Hai, the American breakfasts in Virginia with Elisabeth or sunny day at the beach with Baptiste. I really enjoyed the evenings of food, drinks and fun spent with Renshi, Bart, Hildo, Ruxandra, Edwin, Kim and Pieter. We started a great tradition of regularly hosting these evenings and I hope this tradition does not end. I would also like to thank everyone in the department for the occasional small talk, the various insights and the numerous games at the football table. This list of people includes but is not limited to Robert, Bart, Tamas, Ton, Jan-Willem, Simone, Jens, Manuel, Raf, Vishal, Cornelis, Nico, Yiming, Yihui, Chengpu, Tope, Farid, Sjoerd, Zhe, Tim, Reinier, Laura, Amir, Anqi, Yasin, Yu, Ana, Le, Shuai, Sebastiaan, Esmaeil, Vahab, Mohammad, Zhou, Ilya, Noortje, Yashar, Jia, Patricio, Alexander...

I had a great time in the Center for Optical Research and Education (CORE) in Ut-

sunomiya University, Japan. I thank Barry Cense for having given me the opportunity to work in his lab and the great times spent with his group under the burning sun. I also thank the group in Vancouver of the Simon Fraser University with Prof. Marinko Sarunic, Yifan Jian, Morgan Heisler, Daniel Wahl, Myeong Jin Ju and Bengt Haunerland. Thank you for all the great times in Vancouver and the skiing trip. Great thanks goes to Stefano Bonora, who allowed me to use his state-of-the-art lens in the experiments. I also thank the PhD committee of my defense for their insightful comments and feedback.

Last but not least I thank my family, my parents, my brother and my sister for their support over these four years. I thank Ivana for her love and for being patient whenever I was too distracted with work. I am grateful to always have her by my side in good and in rough times.

Hans Roeland Geert Wim Verstraete
Delft, August 2016

CONTENTS

1	Introduction	1
1.1	Optical coherence tomography	2
1.1.1	Time domain optical coherence tomography	3
1.1.2	Fourier domain optical coherence tomography	5
1.1.3	Resolution and depth of field	8
1.2	Adaptive Optics	9
1.2.1	Optical wavefront aberrations	9
1.2.2	Active and adaptive components	13
1.2.3	Adaptive optics in optical coherence tomography	18
1.3	Wavefront sensorless aberration correction	19
1.3.1	Derivative-free, noisy and costly function optimization	19
1.4	Goal of this thesis	22
1.5	Outline thesis	23
2	Towards model-based adaptive optics optical coherence tomography	27
2.1	Introduction	28
2.2	OCT model	29
2.2.1	OCT mirror model	29
2.2.2	Scattering medium OCT model	31
2.2.3	Single step defocus correction	31
2.3	Materials and methods	32
2.4	Results	33
2.5	Discussion	39
2.6	Conclusion	40
3	Model-based sensor-less wavefront aberration correction in optical coherence tomography	41
3.1	Introduction	42
3.2	Materials and Methods	43
3.3	Experiments	45
3.4	Discussion	50
3.5	Conclusion	50
4	Online Optimization with Costly and Noisy Measurements using Random Fourier Expansions	51
4.1	Introduction	52
4.2	Random Fourier Expansions	53
4.2.1	Ideal RFE Weights	54
4.2.2	Convergence of the Least Squares Solution	57

4.3	Online Optimization Algorithm	59
4.3.1	Recursive Least Squares Approach for the Weights	59
4.3.2	DONE Algorithm.	60
4.4	Choice of Hyper-parameters	61
4.4.1	Probability Distribution of Frequencies	62
4.4.2	Upper Bound on the Regularization Parameter	64
4.5	Numerical Examples	65
4.5.1	Analytic Benchmark Problem: Camelback Function	65
4.5.2	Optical Coherence Tomography	66
4.5.3	Tuning of an Optical Beam-forming Network	68
4.5.4	Robot Arm Movement	72
4.6	Conclusions.	74
4.7	Proof of convergence of the least squares solution	75
4.8	Minimum-variance properties	79
5	Wavefront sensorless adaptive optics OCT with the DONE algorithm for <i>in vivo</i> human retinal imaging	85
5.1	Introduction	86
5.2	Materials and methods	88
5.2.1	Hysteresis correction of the multi-actuator adaptive lens	89
5.2.2	The DONE algorithm	90
5.2.3	Human imaging and aberration correction	92
5.3	Results	93
5.3.1	Imaging human photoreceptors	95
5.4	Discussion	96
5.5	Conclusion	100
6	Applying DONE to S-PAM	101
6.1	Introduction	102
6.2	S-PAM.	102
6.3	Methods	103
6.4	Results and discussion	105
6.5	Conclusion	105
7	Conclusions	107
7.1	Conclusions.	108
7.1.1	Modeling aberrations in optical coherence tomography	108
7.1.2	The DONE algorithm	108
7.1.3	Sensorless aberration correction with DONE.	109
7.2	Recommendations for future work	110
	Bibliography	113
	Summary/Samenvatting	133
	List of publications	137
	Curriculum Vitæ	139

1

INTRODUCTION

This chapter provides an introduction to the relevant topics in this thesis. First, the concept of optical coherence tomography (OCT) is introduced. OCT is a non-invasive optical imaging technique capable of making 3D images. Second, optical wavefront aberrations and adaptive optics are described. Understanding the influence of aberrations on OCT images is essential for finding methods to efficiently correct these aberrations without using a wavefront sensor, which is commonly used in adaptive optics. Third, wavefront sensorless aberration correction is explained. We combine these three concepts into wavefront sensorless adaptive optics in optical coherence tomography.

Parts of this chapter have been published in [1].

1.1. OPTICAL COHERENCE TOMOGRAPHY

Optical coherence tomography (OCT) is a high-resolution imaging modality with a depth resolution of 2-10 μm providing cross-sectional images of tissue, in which the contrast is based on differences in light back-scattering [2]. Today, OCT is a standard imaging tool in ophthalmologic clinics used for the detection of retinal malformations and for monitoring the effect of treatment and/or disease progression. In the medical sector, it is also used for imaging the inside of (coronary) blood vessels and as a tool for optical biopsy [3]. It even has found its way from biomedicine to other application areas, such as non-destructive testing of materials: multi-layered foils [4], pharmaceutical tablets [5], ceramics [6], electronics [7], polymer-based materials [8], artwork conservation [9] and turbid media [10]. New research proposes that smartphones could do OCT imaging [11]. We continue by describing some basic principles of OCT imaging. Cross-sectional images, B-scans, are acquired by performing a series of axial depth scans, A-scans, of optical backscattering at different transverse locations. The axial scans pro-

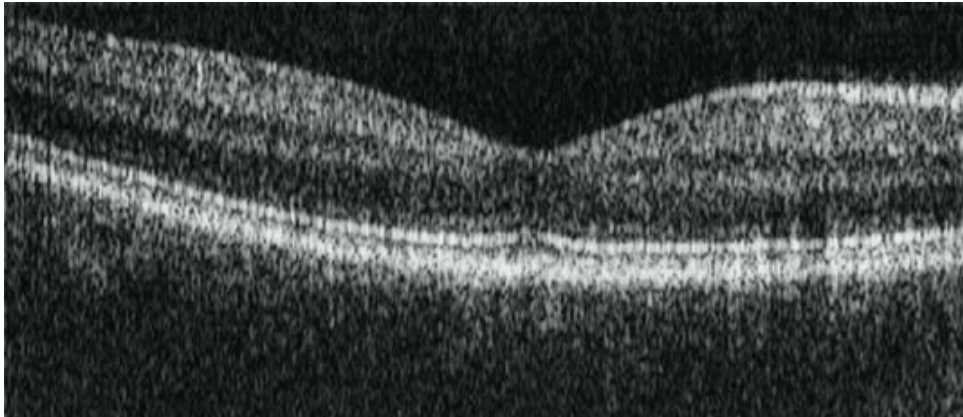


Figure 1.1: Cross-sectional OCT image, showing various layers of a human retina. [12].

vide information of the intensity of backscattered light versus the optical path-length in the sample, which is used to reconstruct a 3D image of the sample. An example of a B-scan of the human retina is shown in Fig. 1.1. A volumetric scan or C-scan is obtained by taking A-scans in two different directions or, in other words, by combining B-scans of different planes. One of the advantages of OCT is that it has a good balance between the penetration depth and axial resolution [13], as shown in Fig. 1.2. OCT has several other advantages over other imaging modalities such as non-invasive measurements, relatively simple setup, and fast scanning.

Initially, OCT systems were operating in the time domain (TD), utilizing low temporal coherence interferometry with a scanning mirror [2, 14, 15]. Temporal coherence is a measure of the average correlation between a wave and a delayed copy of itself. In other words, it characterizes how well a wave can interfere with itself for different delays. A light source with a broad frequency range has low temporal coherence, which plays a crucial role in the axial resolution of OCT. Time domain OCT was superseded by Fourier domain (FD) OCT in which the A-scan is determined from the interference spectrum

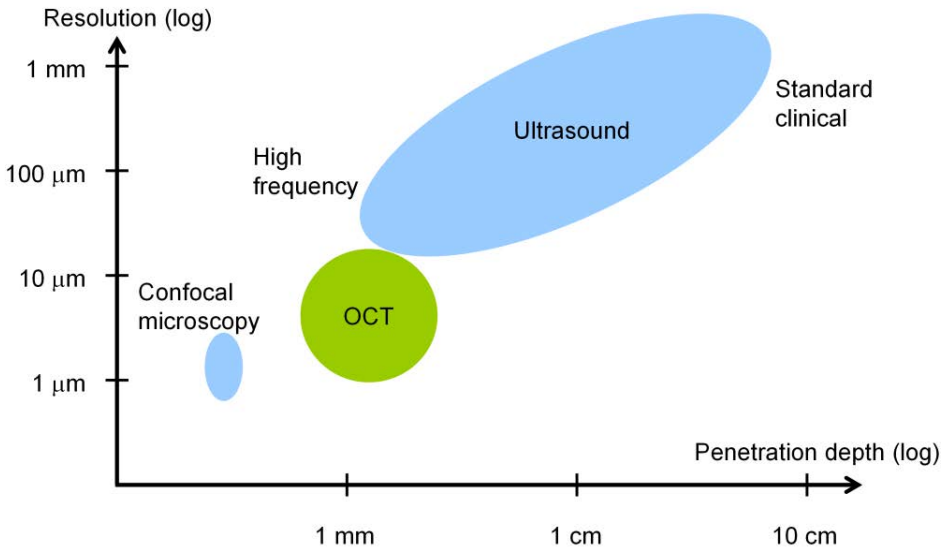


Figure 1.2: Schematic overview of the axial resolution and imaging depth of OCT compared to confocal microscopy and ultrasound [13].

and no longer requires a moving reference mirror in the reference arm to reconstruct the depth information in axial scans [16]. Instead it uses the Fourier transform to go from the wavenumber or spectral domain to the spatial domain, revealing the depth-resolved optical structure of a sample. We will discuss this further in Section 1.1.2. Other functional OCT imaging techniques were developed, such as polarization sensitive OCT and Doppler OCT used for birefringence and flow detection [17], respectively. These techniques are not treated in this thesis and will not be discussed in detail.

1.1.1. TIME DOMAIN OPTICAL COHERENCE TOMOGRAPHY

In 1988, scanning interferometry was first used to measure the eye length [18]. In 1990, a first two-dimensional *in vivo* depiction of a human eye fundus along a horizontal meridian based on white light interferometric depth scans was presented [19]. After further development, the concept of TD-OCT was introduced in 1991 [2, 20]. A typical TD-OCT setup is shown in Fig. 1.3. TD-OCT is based on measuring the interference between backscattered light from a sample and a reference arm using a Michelson interferometer and a single photodetector. To determine the depth at which the reflection in the sample takes place the interference is based on low temporal coherence light. The light from the source is divided by a beamsplitter and directed to the reference arm and the sample arm. The returning light of the sample and reference arm are then aligned in the same optical path and the interference of the two waves is measured by a single photodiode detector. By moving the mirror in the reference arm at constant velocity, the reflectivity profile is modulated at the Doppler frequency of the moving mirror. By demodulating the interference signal, the axial reflectivity profile of the sample can be determined from

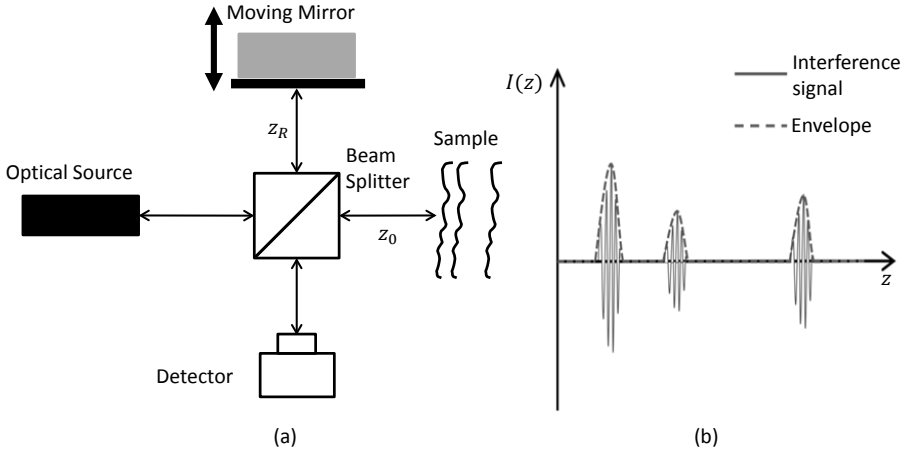


Figure 1.3: (a) Schematic of a time domain OCT system. (b) Schematic of an A-scan of the sample. The envelope represents the axial reflectivity profile of the sample.

the envelope of the interference signal. An example of a possible interference signal and its envelope, which represents the reflectivity profile of a sample, is given in Fig. 1.3.

The signal from the sample is a combination of light backscattered from different depths z in the sample. The reflection amplitude of the sample at depth z is denoted by $a(z)$. The light from the sample is combined with light returning from the reference arm. The intensity measured by the detector I in terms of the wavenumber k is given by

$$I(k) = |E_R(k) + E_S(k)|^2 \quad (1.1)$$

$$= |E_R(k)|^2 + |E_S(k)|^2 + 2 \operatorname{Re} [E_R(k)E_S(k)] \quad (1.2)$$

Here, E_R and E_S are the reference and sample electrical fields at the detector, respectively. It should be noted that losses from the beamsplitter for instance are taken into account by E_R and E_S and that $I(k)$ is implicitly dependent on z_R , the additional distance that is added by the moving mirror. Where appropriate we write $I(k, z_R)$ instead of $I(k)$ to emphasize this dependence. The intensity $I(k, z_R)$ for a sample with reflectivity $a(z)$ and refractive index $n(z)$ is

$$I(k, z_R) = S(k) \left| a_R e^{2ikz_R} + \int_0^\infty a(z) e^{2ik|n(z)z|} dz \right|^2. \quad (1.3)$$

The source's spectral intensity, which expresses the intensity of the source in terms of the wavenumber, is given by $S(k)$, a_R is the total reflection amplitude coefficient of the reference arm and i is the imaginary number $\sqrt{-1}$.

In the following example, we simplify the expression in Eq. (1.3). By convention, the sample has the reflectivity profile $a(z) = 0$ when $z < z_0$, where z_0 is the distance from the beamsplitter to the start of the sample as shown in Fig. 1.3. We ignore the dispersion and assume the refractive index of the sample $n(z)$ to be a constant. The refractive index in the reference and sample arm before the sample is assumed to be one, which is similar

to the refractive index of air. Without further loss of generality, we assume and $a(z) < a_R$ and that $a_R = 1$, which leads to

$$I(k, z_R) = S(k) \left| e^{2ikz_R} + \int_0^\infty a(z) e^{2ik[nz]} dz \right|^2 \quad (1.4)$$

$$= S(k) \left[1 + 2 \operatorname{Re} \left(\int_0^\infty a(z) e^{2ik(z_R - nz)} dz \right) + \int_0^\infty \int_0^\infty a(z) a^*(\hat{z}) e^{2ikn[z - \hat{z}]} dz d\hat{z} \right]. \quad (1.5)$$

Here, $a^*(z)$ is the complex conjugate of $a(z)$. In practice, the third term including $a(z)a^*(\hat{z})$ is a lot smaller because it is not amplified by the reference arm, hence it is ignored. To further simplify this example, we assume that $a(z)$ is real and the optical source has a Gaussian spectral intensity

$$S(k) = \frac{1}{\Delta k \sqrt{\pi}} e^{-\left[\frac{k-k_0}{\Delta k}\right]^2}, \quad (1.6)$$

where $2\Delta k$ is the spectral bandwidth at $\frac{1}{e}$ of its maximum and k_0 represents the center wavenumber of the light source spectrum. Since the detector detects the light at all wavenumbers, we integrate over the wavenumbers as follows [17, Eq. 2.12]

$$I(z_R) = \int S(k) \left[1 + 2 \operatorname{Re} \left(\int_0^\infty a(z) e^{2ik(z_R - nz)} dz \right) \right] dk \quad (1.7)$$

$$= S_0 \left[1 + 2 \int_0^\infty a(z) e^{-(z_R - nz)^2 \Delta k^2} \cos[2k_0(z_R - nz)] dz \right]. \quad (1.8)$$

Here $S_0 = \int_{-\infty}^\infty S(k) dk$ is the spectral integrated intensity from the optical source. For example, if the sample consists of 2 very thin layers such that $a(z) = a_1 \delta(z_{S1}) + a_2 \delta(z_{S2})$ and $n = 1$, then

$$I(z_R) = S_0 \left[1 + 2 \sum_{m=1}^2 a_m e^{-(z_R - z_{Sm})^2 \Delta k^2} \cos[2k_0(z_R - z_{Sm})] \right]. \quad (1.9)$$

The signal $I(z_R)$ in Eq. (1.9) is shown in Fig. 1.4. From Eq. (1.9) it becomes clear that the axial resolution of the OCT system is dependent on the spectral bandwidth $2\Delta k$ of the source. If the spectral bandwidth increases, the interference envelope becomes smaller and the axial resolution improves. The main disadvantage of TD-OCT is that it requires the mechanical movement of a reference mirror. Further research led to the development of Fourier domain OCT that no longer requires a moving reference mirror.

1.1.2. FOURIER DOMAIN OPTICAL COHERENCE TOMOGRAPHY

In Fourier domain (FD) OCT the detector input $I(k)$ Eq. (1.5) is captured, which no longer requires moving the reference mirror. The measured data is then processed using an inverse Fourier transform to reconstruct the sample reflectivity profile. In 1995, spectral domain interferometry was first used to measure intraocular distances [21]. Soon after, spectral interferometry was used in optical coherence tomography [16, 22, 23]. There are two ways to implement FD-OCT.

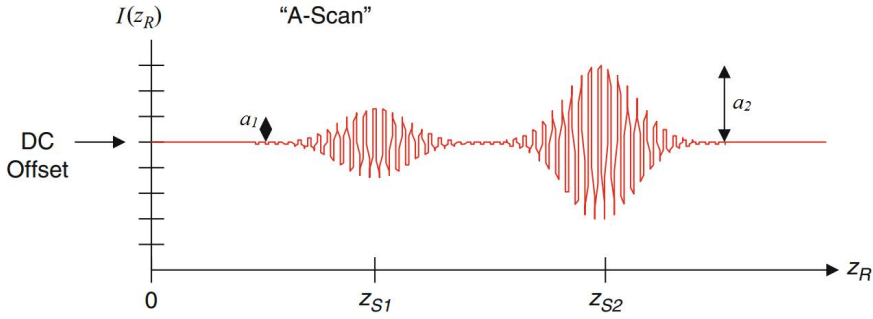


Figure 1.4: TD-OCT A-scan for two thin layers at z_{S1} and z_{S2} [17].

The first way to do FD-OCT is spectral domain (SD) OCT. In SD-OCT the reference arm mirror is stationary and a spectrometer is used to measure $I(k)$ on an array of detectors in space instead of detecting all the light with a single detector. Figure 1.5(a) shows a simplified layout of a SD-OCT setup.

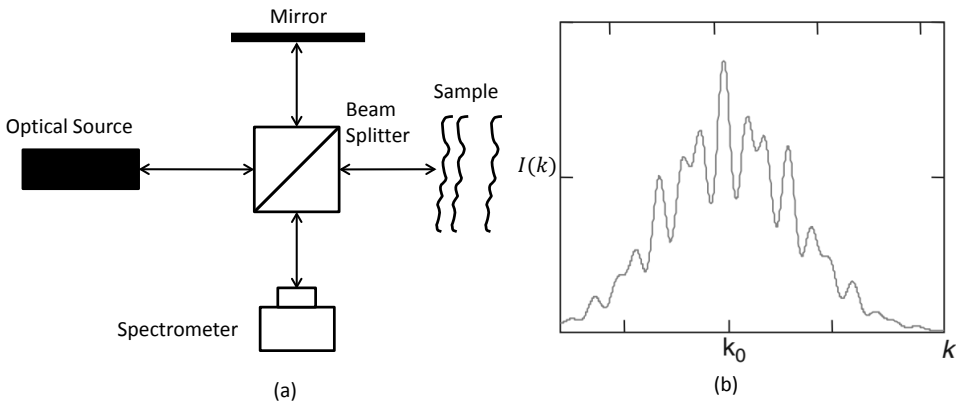


Figure 1.5: (a) Spectral domain optical coherence tomography layout. (b) Possible interference spectrum of the sample on the spectrometer [17].

The second method for FD-OCT is swept-source (SS) OCT. In SS-OCT the reference mirror is also stationary and a tunable laser is used to measure $I(k)$ in time on a single photodiode detector. A tunable laser is a narrowband optical source that can sweep through its wavelengths or wavenumbers. During the sweep through the wavenumbers synchronous measurements are taken with the detector to measure $I(k)$. For both SD-OCT and SS-OCT the A-line is obtained through inverse Fourier transformation of $I(k)$.

To explain the principles of FD-OCT we start from Eq. (1.5). The added distance z_R from the moving mirror is set to zero, because in FD-OCT no moving mirror is required to obtain $I(k)$. To simplify this example, we assume that reflectivity profile $a(z)$ is real.

This leads to

$$I(k) = S(k) \left[1 + 2 \int_0^\infty a(z) \cos(2knz) dz + \int_0^\infty \int_0^\infty a(z) a(\hat{z}) e^{2ikn[z-\hat{z}]} dz d\hat{z} \right]. \quad (1.10)$$

In this equation three terms can be seen. The first term is a constant offset. The second term is the term that encodes the depth information of the object. The depth z is embedded in a sum of cosines with angular frequencies $2kn$. The third term is an auto-correlation term that represent the mutual interference of all the waves from the sample. We continue rewriting Eq. (1.10) using the definition $\hat{a}(z) = a(z) + a(-z)$, which gives

$$I(k) = S(k) \left[1 + \int_{-\infty}^\infty \hat{a}(z) \cos(2knz) dz + \frac{1}{4} \int_{-\infty}^\infty \int_{-\infty}^\infty \hat{a}(z) \hat{a}(\hat{z}) e^{-2ikn[z-\hat{z}]} dz d\hat{z} \right] \quad (1.11)$$

$$= S(k) \left[1 + \int_{-\infty}^\infty \hat{a}(z) e^{-2iknz} dz + \frac{1}{4} \int_{-\infty}^\infty \text{AC}[\hat{a}(z)] e^{-2iknz} dz \right], \quad (1.12)$$

with $\text{AC}[\hat{a}(z)] = \int_{-\infty}^\infty \hat{a}(z) \hat{a}(\hat{z}) e^{2ikn\hat{z}} d\hat{z}$. We define the Fourier transform as $\mathfrak{F}_{\tilde{z}}[f(\tilde{z})] = \int_{-\infty}^\infty f(\tilde{z}) e^{-ik\tilde{z}} d\tilde{z}$ and substitute $\tilde{z} = 2nz$. Following [22, Eq. 4 to 6], this results in

$$I(k) = S(k) \left[1 + \frac{1}{2n} \mathfrak{F}_{\tilde{z}} \left(\hat{a} \left(\frac{\tilde{z}}{2n} \right) \right) + \frac{1}{8n} \mathfrak{F}_{\tilde{z}} \left(\text{AC} \left[\hat{a} \left(\frac{\tilde{z}}{2n} \right) \right] \right) \right], \quad (1.13)$$

where AC essentially represents an autocorrelation. Taking the inverse Fourier transform \mathfrak{F}_k^{-1} gives us

$$\mathfrak{F}_k^{-1}[I(k)] = \mathfrak{F}_k^{-1}[S(k)] * \left[\delta(z) + \frac{1}{2n} \hat{a}(z) + \frac{1}{8n} \text{AC}[\hat{a}(z)] \right], \quad (1.14)$$

with $*$ indicating a convolution. The convolution with the delta peak results in an inverse Fourier transform of the spectrum of the source located at $z = 0$. The third autocorrelation term is not amplified by the reference arm signal, therefore, it will often be a lot weaker and is considered negligible. Additionally, this term is located around $z = 0$ and can easily be ignored by moving the sample away from $z = 0$. The second term contains the reflectivity profile of the sample. Figure 1.6 shows an exemplary-scan of a sample reconstructed with the FD-OCT methods described in this section.

In FD-OCT, the maximum depth of the measurements z_{max} is limited by the number of pixels in the detector array. The maximum depth range z_{max} for reconstruction in FD-OCT is

$$z_{max} = \frac{\pi}{2\delta_s k}, \quad (1.15)$$

where $\delta_s k$ is the sampling interval in the k domain. In practice, the depth range of OCT is limited by the scattering and absorption of tissue to about 2 mm deep because at greater depths the proportion of light returning in the optical path is too small to be detected.

In comparison with TD-OCT, FD-OCT can take scans faster because it no longer requires a moving reference mirror. Therefore, in applications where measurement speed is of the essence, FD-OCT is generally preferred. Additionally, FD-OCT has an improved sensitivity over TD-OCT [24]. However, when imaging samples, not only the sensitivity is important, but also the resolution will determine which features you will be able to distinguish.

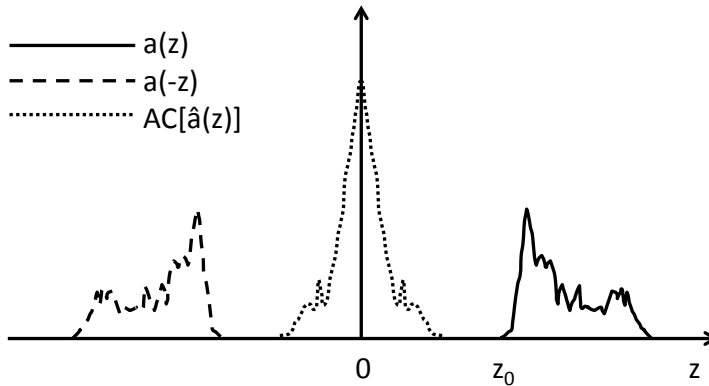


Figure 1.6: FD-OCT A-scan of an arbitrary sample.

1.1.3. RESOLUTION AND DEPTH OF FIELD

In OCT, the axial resolution and lateral resolution are independent. As demonstrated in Eq. (1.9) for TD-OCT and in Eq. (1.14) for FD-OCT, the axial resolution is given by the width of the axial point spread function, which is the inverse Fourier transform of the source spectrum. For a Gaussian source spectrum, the width of the axial PSF is inversely proportional to the bandwidth of the light source and the axial resolution is [25]

$$\Delta z = \frac{2 \ln(2) \lambda_0^2}{\pi \Delta \lambda} \quad (1.16)$$

where Δz is the full-width-at-half-maximum (FWHM) of the axial point spread function, $\Delta \lambda$ is the FWHM of the spectral intensity $S(k)$ and λ_0 is the center wavelength of the optical source. From Eq. (1.9) it can be seen that a broad bandwidth optical source is necessary to achieve a high axial resolution [26].

The lateral resolution in OCT is decoupled from the axial resolution and determined by the focal spot size. In OCT, the lateral intensity distribution of the beam is often Gaussian [27]. This Gaussian beam is relayed onto the entrance pupil of the final sample arm lens and, if no optical wavefront aberrations are present, a diffraction limited focal spot is generated at the sample. The diffraction limited spot size is proportional to the wavelength and inversely proportional to the numerical aperture or pupil size of the focused beam. The e^{-2} Gaussian beam waist w_0 for a spot the focus plane is approximated by,

$$w_0 = \frac{\lambda f}{\pi w}, \quad (1.17)$$

where w is the e^{-2} beam waist in the pupil plane, λ is the wavelength and f is the focal length [28]. Higher lateral resolutions are obtained by using a larger numerical aperture, i.e., using a larger d or smaller f . Unfortunately, a higher numerical aperture also results in a decreased depth of field. If the beam is strongly focused, objects in the focus plane are sharp, however, out of focus objects are strongly blurred. This is characterized by the Rayleigh length of the optical system. The Rayleigh length is the distance along the

propagation direction of a beam from the waist to the place where the area of the cross section of the beam is doubled. A Gaussian beam focal spot has a Rayleigh length of [28]

$$z_{Rayleigh} = \frac{\pi w_0^2}{\lambda}. \quad (1.18)$$

To summarize, in OCT the axial resolution is independent of the lateral resolution and can be improved by getting a broad bandwidth optical source. However, improving the lateral resolution requires a trade-off between diffraction limited spot-size and the depth of field. Moreover, using a larger pupil results in large wavefront aberrations that deteriorate the lateral resolution. Adaptive optics can correct these wavefront aberrations.

1.2. ADAPTIVE OPTICS

Optical wavefronts can be reshaped using adaptive optics (AO). In most cases, AO is used for modifying the phase of an optical beam using deformable mirrors, lenses or other adaptive phase shaping devices. But, AO is more than just phase-only correction. Many techniques use intensity and phase correction for the reshaping of an optical wavefront and AO is also applicable to incoherent imaging techniques. However, in this thesis we will focus on phase modification and the removal of phase aberrations to improve the image quality. More and more optical systems that suffer from wavefront aberrations are using AO to improve their signal to noise ratio, image quality or optical beam quality [29]. An example of phase aberrations are the defocus and astigmatism aberrations introduced by the human eye, which result in poor vision and can be corrected for by wearing contact lenses or spectacles. To correct wavefront aberrations that result from a large entrance pupil, the wavefront can be corrected before entering the pupil of the eye. For a spectral-domain OCT system this can be done using the setup as shown in Fig. 1.7. The spectral domain OCT system in Fig. 1.7 is combined with an active optical element, such as a deformable mirror or lens to change the phase of the wavefront entering the eye, such that a perfect spherical wavefront is incident on the retina. A wavefront sensor, in this case a Shack-Hartmann wavefront sensor, is used to measure the wavefront returning from the eye.

1.2.1. OPTICAL WAVEFRONT ABERRATIONS

In many optics applications, including OCT, higher resolutions and higher signal to noise ratios are desired [30]. In practice, imaging systems rarely reach the diffraction limit [31, 32], because optical wavefront aberrations lead to a larger focal spot, which deteriorates the achievable resolution of the imaging system [33, 34]. When free-space light is coupled into a fiber, the coupling efficiency of the light into the fiber is affected by the optical wavefront aberrations as well. In this case, aberrations do not only lead to a loss in resolution, but also to a decreased signal to noise ratio. Aberrations can be caused by imperfections of optical lenses and mirrors or misalignments in the optical path. They are also induced by inhomogeneous media whose refractive index n varies in space or time. Examples are turbulent atmospheres in astronomy [35] and imaging of biological specimens [36,37]. When *in vivo* imaging the human retina, the largest optical wavefront aberrations are introduced by imperfections of the eye lens and the cornea [38–40].

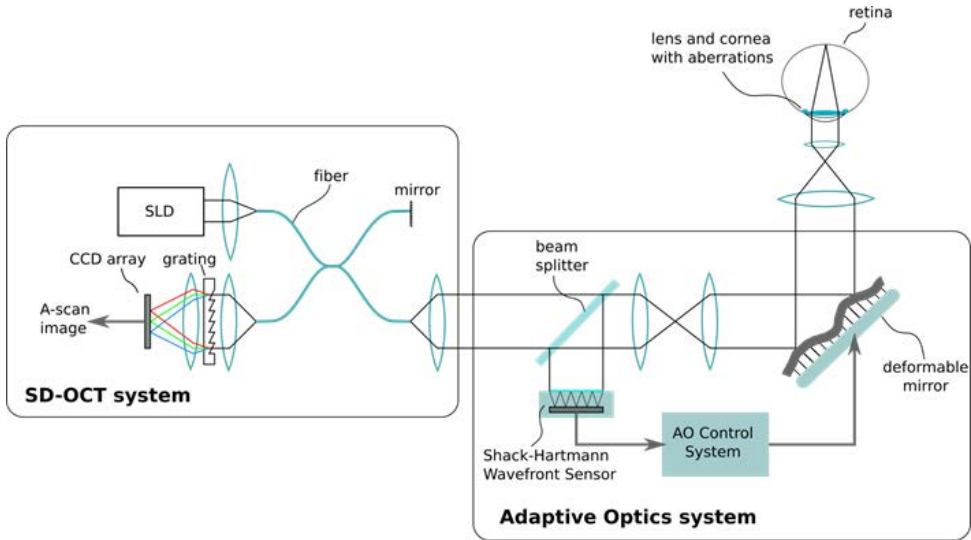


Figure 1.7: An adaptive optics spectral domain optical coherence tomography setup.

Optical aberrations are often modeled by the generalized pupil function (GPF) [33],

$$P(x, y) = A(x, y)e^{i\Phi(x, y)}. \quad (1.19)$$

Here, $\Phi(x, y)$ is called the phase aberration function and $A(x, y)$ is the amplitude distribution. The complex GPF is defined over the pupil of the optical system. The amplitude distribution in the pupil plays an important role in the size of the focal point and can account for amplitude apodization [41], where apodization refers to cutting off the edges of the optical beam. The phase of the wavefront in the pupil $\Phi(x, y)$ also has a great influence on the resolution and can be decomposed into Zernike polynomials, which are a set of orthogonal polynomials defined over the unit disk [42, 43]. A visualization of some Zernike polynomials is shown in Fig. 1.8. The polynomials are named after the Dutch physicist Frits Zernike, who won the Nobel prize in physics in 1953 [44]. An advantage of Zernike polynomials is not only that they can be obtained in closed form, but also that the first few modes represent the classical aberrations familiar to opticians and represent the largest aberrations in the human eye [45, 46]. To simplify the notations of the Zernike polynomials, a polar coordinate system (ρ, ϕ) is introduced in which $x = \rho \cos(\phi)$ and $y = \rho \sin(\phi)$. The phase Φ is decomposed in Zernike polynomials as follows,

$$\Phi(\rho, \phi) = kW(\rho, \phi) = k \sum_{n, m} \alpha_{n, m} Z_{n, m}(\rho, \phi). \quad (1.20)$$

The indices $n \in \mathbb{N}_0$ and $m \in \mathbb{Z}$, represent the radial order and the azimuthal frequency of the Zernike polynomials and are chosen such that $n - |m|$ is even and greater than or equal to zero.¹ The Zernike coefficients $\alpha_{n, m} \in \mathbb{R}$ are a measure of how much each

¹Both the radial order of the Zernike polynomials and the index of refraction are denoted by n as done in literature, the context and usage should avoid any ambiguity.

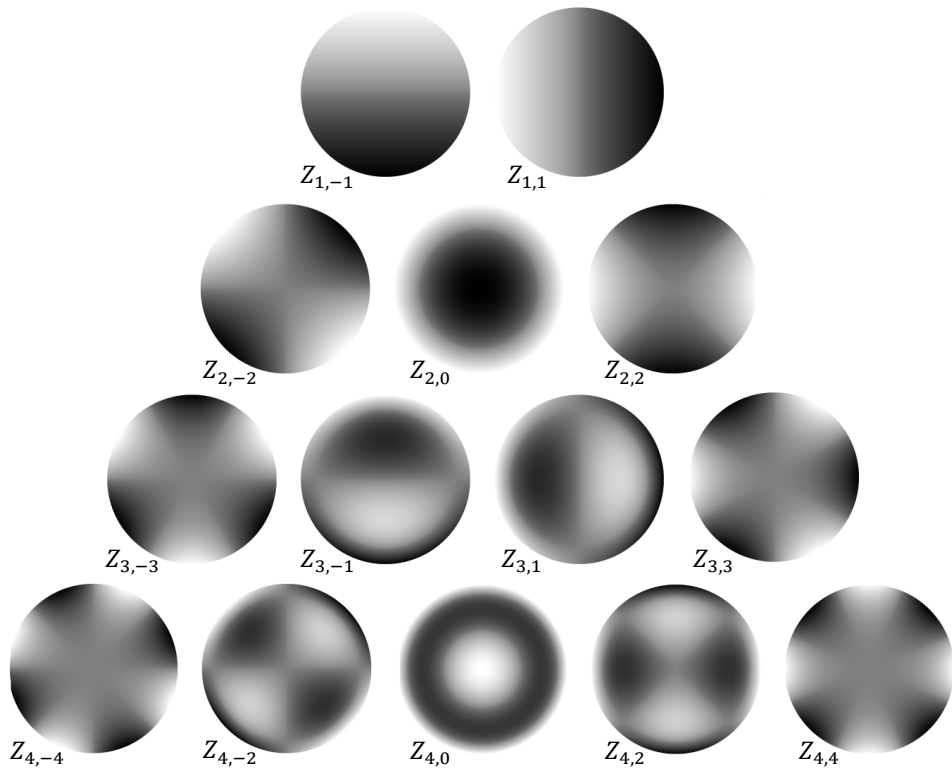


Figure 1.8: Visualization of the first to fourth order Zernike polynomials.

aberration contributes to the phase. The Zernike polynomials are defined as:

$$Z_{n,m}(\rho, \phi) = \begin{cases} c_{n,m} R_{n,|m|}(\rho) \cos(m\phi) & m \geq 0 \\ -c_{n,m} R_{n,|m|}(\rho) \sin(m\phi) & m < 0 \end{cases}, \quad (1.21)$$

where $c_{n,m}$ is a normalization constant and $R_{n,|m|}(\rho)$ is the radial polynomial. The normalization constant is defined as

$$c_{n,m} = \begin{cases} \sqrt{n+1} & m = 0 \\ \sqrt{2(n+1)} & m \neq 0 \end{cases}, \quad (1.22)$$

and the radial polynomial $R_{n,|m|}(\rho)$ is

$$R_{n,|m|}(\rho) = \sum_{j=0}^{(n-|m|)/2} \frac{(-1)^j (n-j)!}{j! \binom{n+|m|}{2-j}! \binom{n-|m|}{2-j}!} \rho^{n-2j}. \quad (1.23)$$

The Zernike polynomials are normalized to have unit variance over the unit disk [46]. Throughout the thesis we also use the Zernike polynomials in Cartesian coordinates $Z_{n,m}(x, y)$. An overview of Zernike polynomials in Cartesian coordinates is given in [47].

A decomposition of the phase in Zernike polynomials makes it easy to express the aberration variance in terms of the classical aberrations and clearly shows how much each Zernike mode contributes to the root mean square (RMS) of the phase aberration function [48, 49]. The RMS of the phase aberration $\Phi(\rho, \phi)$ is calculated as follows [50],

$$\text{RMS}[\Phi(\rho, \phi)] = \sqrt{\frac{1}{\pi} \int_0^1 \int_0^{2\pi} \Phi(\rho, \phi)^2 \rho d\rho d\phi}. \quad (1.24)$$

Using the orthogonality of Zernike polynomials and the normalization factor $c_{n,m}$ this can be simplified to

$$\text{RMS}[\Phi(\rho, \phi)] = k \sqrt{\sum_{\substack{n,m \\ n \neq 0}} \alpha_{n,m}^2}. \quad (1.25)$$

Zernike polynomials, except for the piston $Z_{0,0}$, have zero mean and a variance of one over the unit disk. The piston $Z_{0,0}$ is an extra delay and often neglected as it does not deteriorate the focal spot. This means that the variance of $\Phi(\rho, \phi)$ is

$$\text{var}([\Phi(\rho, \phi)]) = k^2 \sum_{\substack{n,m \\ n \neq 0}} \alpha_{n,m}^2. \quad (1.26)$$

In OCT, the piston will result in an axial shift of the sample in an A-scan. This axial shift can be countered by adjusting the length of the reference arm. The tip and tilt aberrations, $Z_{1,-1}$ and $Z_{1,1}$, change the angle of the beam. In OCT, this principle is taken advantage of for laterally scanning over the sample with galvanometric scanning mirrors that introduce tip and tilt. The tip and tilt are also often ignored as they do not directly lead to a deterioration of the focal spot. However, they can introduce other aberrations by steering the beam of the optical axis. The second order Zernike polynomials, $n = 2$,

namely the defocus, the vertical astigmatism and the oblique astigmatism, are the largest aberrations in a healthy human eye [51]. When imaging the human retina, correcting the second radial order modes typically results in the largest improvement [40,52]. However, the influence of aberrations on the OCT signal is not well defined. In Chapter 2 we investigate the influence of phase aberrations on the OCT system. By doing this, we find a relation between the optical aberrations and the OCT signal. This is exploited to identify and correct undesired aberrations without using a wavefront sensor as shown in Fig. 1.7. Undesired aberrations in the optical beam path are corrected for by an active or adaptive component in the optical path.

1.2.2. ACTIVE AND ADAPTIVE COMPONENTS

Adaptive optics (AO) is used in optical setups to remove undesired wavefront aberrations and to shape optical wavefronts, in other words, you try to minimize the undesired phase aberrations and add desired aberrations to $\Phi(x, y)$ with AO. If $\Phi(x, y)$ is decomposed in Zernike polynomials, this also implies minimizing certain Zernike coefficients $\alpha_{n,m}$. In this section we will describe the principles of deformable mirrors and active or deformable lenses that are commonly used to reshape the phase of the wavefront $\Phi(x, y)$.

DEFORMABLE MIRROR

Deformable mirrors (DM) are among the most commonly used wavefront correctors. DMs are mirrors with a deformable surface whose shape can be adjusted actively for the

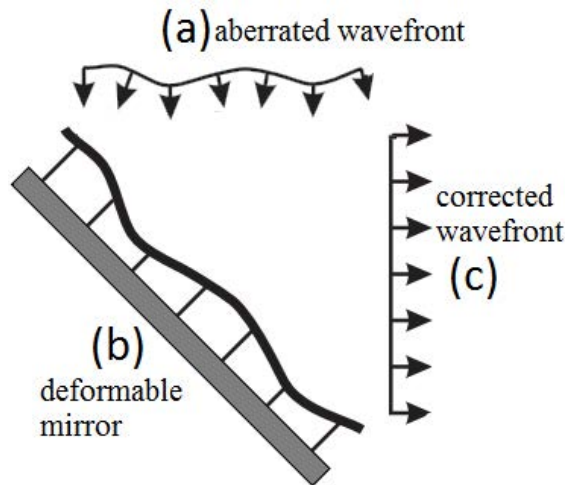


Figure 1.9: (a) Aberrated wavefront propagating towards a deformable mirror. (b) Deformable mirror with continuous surface. (c) Corrected wavefront.

correction of optical aberrations [53,54]. Several different technologies exist for creating deformable mirrors. Examples are segmented mirrors [55], magnetic actuated DMs [56], microelectromechanical systems (MEMS) based DMs [57], membrane DMs [58], bi-morph DMs [59] and ferrofluid based DMs [60]. Every technology limits the achievable DM properties and has its specific advantages and disadvantages [61]. An example of

the operation of a DM is shown in Fig. 1.9. Figure 1.9 explains the principle of a DM by analyzing the phase differences of the light. Phase modulation devices, which includes DMs, are based on changing the optical path difference (OPD), which can be written as $n\Delta z$, where n is the refractive index and Δz is the length of the path travelled. The OPD is related to the phase as $\Phi = 2\pi\text{OPD}/\lambda$, which can be decomposed in Zernike polynomials. Deformable mirrors change the length of the path Δz by shaping their reflective surface as seen in Fig. 1.9. As such they can correct for the path length differences and the phase aberrations of the incoming wavefront. Alternatively, they can shape the wavefront in a pre-determined shape to compensate for aberrations further down in the system. The shape of a DM's surface is controlled by multiple actuators at the backside of the mirror [62]. DMs can be used in combination with optical wavefront sensors and control systems [63, 64]. Often, they can be modelled and approximated by linear systems, such that linear control theory applies [65–67]. DMs have several advantages. They are reflective and therefore do not introduce chromatic aberrations. This is important for broadband applications such as OCT, because the imaging performance would suffer from these chromatic aberrations. Also, their coatings can be highly reflective so that there is almost no power loss in the optical system and, therefore, are usable for high-power optics. A disadvantage is that DMs are expensive and require folded optical paths to conjugate the various pupil planes in the system.

Several properties characterize each DM [68]. A key property is how accurately a target phase can be reproduced. For example, a segmented mirror has a discontinuous surface and in contrast to the continuous surface DM in Fig. 1.9, the surface of a segmented DM can result in discontinuous phases. Being a mechanical element, every actuator has a region where it has influence on the DM surface shape, which can be described by the influence function. The influence function is the characteristic phase change corresponding to the action of a single actuator. If the influence functions of all actuators in a deformable mirror are linear with respect to the input and the displacement, then a single influence matrix can describe all the corresponding characteristic phase changes, also see Chapter 6. The influence functions of the majority of DMs are quite different from Zernike polynomials, however, in most cases Zernike polynomials can be formed approximately by the DM as a combination of its influence functions. The reproduction quality of Zernike polynomials is mainly determined by the number of actuators and the shape of the influence functions. Moreover, different actuators can influence each other through coupling. Actuator coupling occurs when the movement of one actuator displaces its neighbors [69] and depends on the pitch between the actuators, the materials used and the stroke. The actuator stroke is the maximum possible actuator displacement and limits the amplitude of the introduced wavefront. The inter-actuator stroke and the pitch limit the gradients and the amplitude of higher-order aberrations. For DMs typically, the actuator stroke is between 1 to 50 micrometers.

In practice, for each application different demands are placed on the performance of the deformable mirror. For high-speed applications and the correction of dynamic aberrations the response and settling time of the actuators are important [70]. For applications in which large wavefront corrections are required, e.g. in ophthalmology, the position and number of actuators, their accuracy and (inter-actuator) stroke determine the maximum amplitude and number of Zernike modes the deformable mirror can cor-

rect [68, 71, 72]. Even though all the available technologies still have their limits, it is possible for a specific application to formulate general requirements for a wavefront corrector. For example, in astronomy a DM should have a large number of degrees of freedom, i.e. many actuators, fast response time and accuracy. The shape and the amplitude of its influence functions should be able to match the statistics of the wavefronts to be corrected [29]. Finally, the response time should be fast enough to cover the whole temporal range of the targeted aberrations. The cost of a DM is strongly related to the requirements of the application and are often considered too expensive in many applications. Other alternatives, such as adaptive or deformable lenses, for correcting aberrations in the phase of the wavefront exist.

ACTIVE AND ADAPTIVE LENSES

In contrast with DMs, adaptive and deformable lenses are used in transmission rather than in a reflective manner. With an adaptive lens you can increase or decrease the distance the waves travel through a medium with a different refractive index. In this way, it can change the OPD and introduce variations in the phase delays of the optical wavefront. An example of a deformable lens changing its focal length is shown in Fig. 1.10.

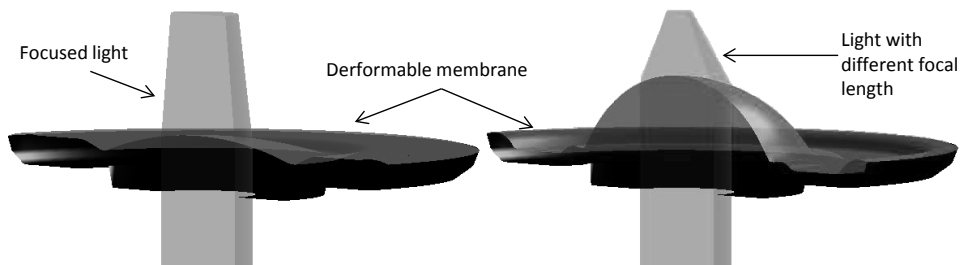


Figure 1.10: Example of a deformable lens. The lens surface changes shape, therefore the focus of the lens changes from far away (left) to close by (right) [73].

When using active lenses, different wavefront aberrations can be introduced by translating or rotating fixed lenses with different shapes with respect to each other in the optical beam [74–78]. In [79] an optical lens setup is described to correct astigmatism by rotating two cylindrical lenses. It is shown that the astigmatism can be removed using the two cylindrical lenses and that defocus can be removed using a lens on a translation stage. In [80] a setup with two low cost cylindrical lenses is used for astigmatism correction and a description of how the angles of the rotational lenses can be related to the size of the astigmatism aberrations is given. This setup has been tested in a fundus imaging device, which makes *en face* retinal images, but has no active defocus correction. In [1], an optical lens setup to correct both large defocus and astigmatism was presented. For defocus correction a deformable defocus lens (Optotune EL-10-30) is used and cylindrical lenses are placed on rotation stages behind the offset lens to introduce the astigmatism, shown in Fig. 1.11. In Fig. 1.12 we demonstrate the potential of this setup by showing the images of a resolution target with different aberrations.

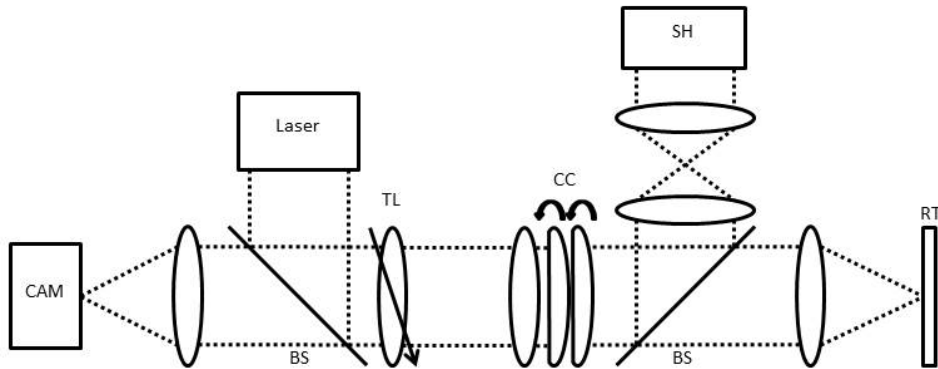


Figure 1.11: Full lens setup for defocus and astigmatism correction. CAM is the camera, BS are beamsplitters, TL is the deformable lens, CC are the rotational cylindrical lenses, SH is the Shack-Hartmann wavefront sensor, RT is the resolution target.

The images in Fig. 1.12 show that the lens setup can accurately change the defocus and astigmatism aberrations in the optical path and that it is capable of correcting most of the remaining defocus and astigmatism aberrations in the system. The lens setup can correct large defocus and astigmatism aberrations ranging from at least -6 to 4 diopters for the defocus and from -1.5 up to 1.5 diopters for both astigmatism with a root mean square (RMS) error of 0.17 and 0.075 diopters for the defocus and the astigmatism, respectively. However, this setup has the disadvantage that it is bulky and can only introduce defocus and astigmatism aberrations. Furthermore, the beam size on the resolution target changes slightly when different aberrations are introduced by the setup. Other more compact adaptive lens technologies exist that can introduce aberrations with a single lens [81, 82]. There are two principal approaches to make adaptive lenses.

The first principal approach is to actively change the shape of a lens. The first example of this approach is the deformable defocus lens (Optotune EL-10-30) used in the setup mentioned before in Fig. 1.11. This lens changes shape based on the amount of liquid in the lens and the stiffness of the membrane. The liquid is concealed in a container with at least one side being an elastic polymer or membrane. The membrane deforms under changes of pressure in the container, hence changing the shape of the lens. The change of the pressure of the liquid can be controlled mechanically or electrically [73]. The advantages of this type of deformable lens are a large tuning range and a low polarization dependence. Another way to change the shape of the deformable lens is based on electrowetting [83–85]. These lenses consist of two liquids with a similar density but a different refractive index, such as water and oil. A voltage is applied to an insulated metal substrate and the fluids, one conductive and the other insulating, changes shape according to the applied voltage. This technology has several advantages, such as relatively fast response times, a compact design, inexpensive fabrication and low power consumption. However, for both of these technologies, gravity induces a coma optical aberration when the lens is in upright position (optical axis horizontal) and the density of the two liquids does not match, which is usually the case. Varioptics, a manufacturer of electrowetting

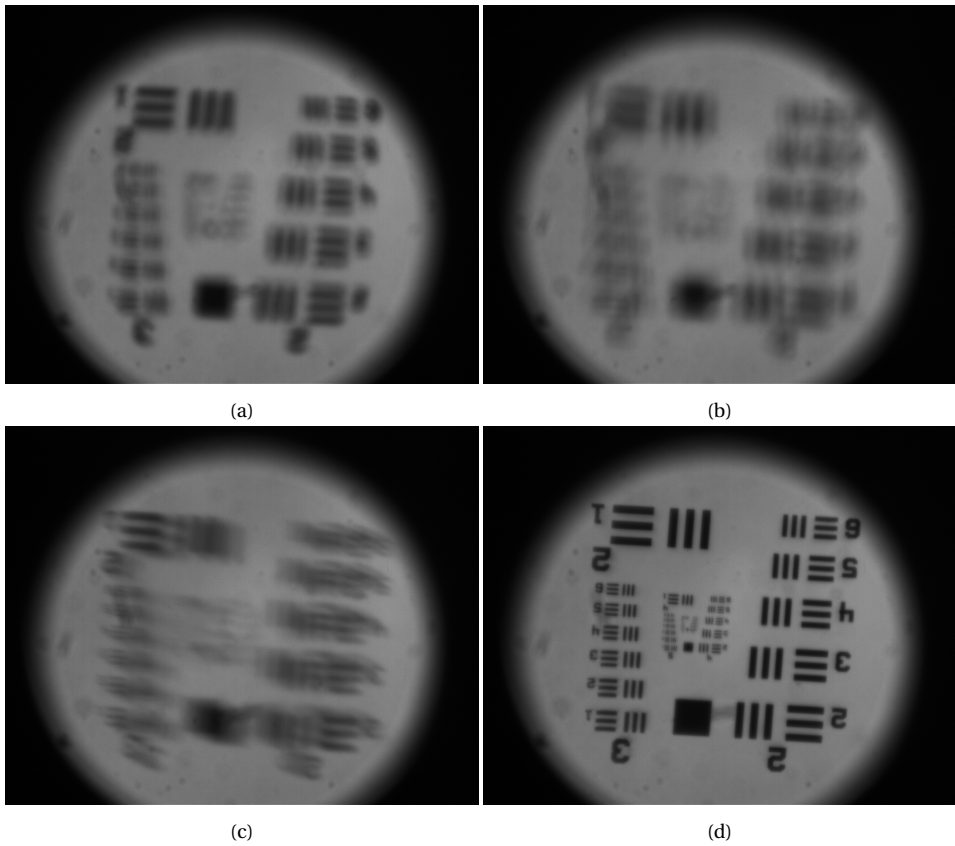


Figure 1.12: (a) Image with 0.39 diopters defocus. (b) Image with vertical and oblique astigmatisms of 0.33 diopters and 0.44 diopters. (c) Image with 0.61 diopters defocus and astigmatisms of 0.27 diopters and 0.19 diopters. (d) Image with aberrations removed.

lenses, recently announced a deformable lens with a clear aperture of 8 mm and a defocus range of -12 to +12 diopters [86]. Moreover, this lens can also correct astigmatisms up to -6 diopters, which compared to other deformable lenses is very large. Recently, another multi-actuator lens that changes the shape was built in Italy. This is the only available deformable lens with 18 piezoelectric actuators and it can correct up to the fourth order of Zernike aberrations [87, 88]. However, the maximum amplitude of the aberrations is limited.

The second principal approach to adaptive lenses is based on local changes in refractive index, which can be induced by electro-optic or acousto-optic effects. A popular technology for this approach are liquid crystal lenses [89]. They have the advantage of small power dissipation, low voltages and ease of miniaturization. However, liquid crystals are sensitive to the polarization of the light and slow in response due to reordering of molecules with respect to other technologies.

Throughout the thesis, we examine different alternatives for an affordable adaptive component in an AO-OCT system that can swiftly correct human eye aberrations. In Fig. 1.7, the AO-OCT setup uses a deformable mirror. In Chapter 2 and Chapter 3 we use a deformable mirror and in Chapter 5, we explore the use of a deformable lens for AO-OCT.

1.2.3. ADAPTIVE OPTICS IN OPTICAL COHERENCE TOMOGRAPHY

The development of new broad spectrum light sources [90–93] has improved the axial resolution for retinal imaging below $3\ \mu\text{m}$ in OCT systems in 2001. However, the lateral resolution was still hampered by optical wavefront aberrations in the eye and did not reach theoretical diffraction-limited performance. As mentioned earlier, the lateral resolution of OCT improves by using a larger pupil. However, when the pupil size is increased for *in vivo* imaging of the retina, large ocular aberrations are introduced. In practice, most conventional OCT systems used for retinal imaging are limited to lateral resolutions of approximately 15 to 20 μm .

The concept of an AO-OCT system is shown in Fig. 1.7. A high lateral resolution can be obtained by using adaptive optics (AO) to correct the optical wavefront aberrations on large pupils ($>2\ \text{mm}$). A wavefront sensor and a deformable mirror are used in combination to correct the undesired aberrations. Initially, a flat wavefront is launched into the eye. A small part of the aberrated light returning from the human retina is directed towards a wavefront sensor. The wavefront sensor estimates the aberrations that the light experienced from its propagation to and from the retina. The shape of the deformable mirror is set to correct the undesired aberrations. In practice this means that the wavefront is deformed in such a way that after traversing the cornea and the lens a perfect spherical wavefront is incident on the retina.

AO has been used in human retinal imaging to correct ocular wavefront aberrations. In 2003, a report was made of improved lateral resolution for TD-OCT using *en face* OCT and AO [94]. Soon after in 2004, a high-resolution TD-OCT system was demonstrated with an improved transverse resolution of 5 to 10 μm [95]. The first implementation of AO correction in FD-OCT imaging has been reported in 2005 with a lateral resolution of 3 μm [96]. The combination of AO with the high axial resolution of OCT results in three-dimensional ultra-high resolution imaging. Such systems have been reported [96–99]

and demonstrated lateral and axial resolution up to $3\ \mu\text{m}$ and 2 to $3\ \mu\text{m}$, respectively. As a result, AO-OCT has made it possible to image the 3D architecture of individual rods and cones *in vivo* in the human eye [99, 100]. These results demonstrate that the combination of AO and OCT has the potential to provide researchers and clinicians with near cellular-resolution information on retinal morphology [101]. To reduce the cost and size of an AO-OCT setup, wavefront sensorless aberration correction has been pursued as will be discussed in the next section.

1.3. WAVEFRONT SENSORLESS ABERRATION CORRECTION

In Sec. 1.2, an example of an AO-OCT setup with a Shack-Hartmann wavefront sensor was given in Fig. 1.7. Imaging systems with SH wavefront sensors can suffer from several disadvantages, such as non-common path wavefront errors and undesired reflections on the wavefront sensor if they are not taken into account and removed. However, the biggest disadvantage is that wavefront sensors lead to an increased cost and size of an AO-OCT setup. These drawbacks can be avoided if wavefront sensor-less adaptive optics (WFSL-AO) methods are used that no longer require a wavefront sensor. WFSL methods are based on the optimization of quality metrics of the acquired image in order to minimize the aberrations and create a sharper image. Examples of metrics are image sharpness or signal strength. In WFSL-AO-OCT, aberration correction methods rely on the measurement of a metric based on the OCT image (e.g. the strength of the OCT signal) rather than on the wavefront measurement using of the wavefront sensor. WFSL-AO optimization procedures are often confused with post-processing methods that attempt to remove aberrations from already acquired images [102, 103]. Post-processing algorithms do not improve the imaging resolution or signal to noise ratio of the system by actively correcting aberrations in the optical beam path, but rather remove them after the image has already been taken and create image artefacts instead. If aberrations are too large or the signal to noise ratio is too low, these methods often do not succeed in creating a better image. In short, WFSL-AO algorithms aim to optimize the image quality by actively changing the wavefront of an optical imaging system without using a wavefront sensor.

1.3.1. DERIVATIVE-FREE, NOISY AND COSTLY FUNCTION OPTIMIZATION

The search for the wavefront aberrations that maximize the image quality metric is an optimization process. Optimization can be defined as finding the values of variables with the most cost effective or highest achievable performance of the objective function f under the given constraints. In practice, the lack of full information, the presence of noise and the lack of time restricts the optimization process in finding the optimum [104]. In the case of WFSL-AO the variables used in the optimization are often related to the inputs of the adaptive components, such as the deformable mirror or the deformable lens. For example, the variables that are used to maximize the image quality can be the Zernike coefficients of the aberrations which are displayed by the adaptive component. In this case, the image quality metric is an objective function that depends on certain characteristics of the system.

The goal of the optimization is to find values of the variables in vector \mathbf{x} that optimize the objective function or metric f . The objective function or metric f is evaluated by tak-

ing a measurement or an image. The variables are often constrained to a certain domain or restricted to have certain properties [105]. The minimization problem is defined as

$$\min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}), \quad (1.27)$$

where \mathcal{X} is the domain to which \mathbf{x} is restricted. A plethora of algorithms exist that tackle variations of this problem. Different gradient descent or gradient based algorithms exist that use the Jacobian $\Delta_{\mathbf{x}}f$ in the optimization process [106]. Other examples are quasi-Newton methods [107, 108] and conjugate directions methods [109, 110]. The step size per iteration in these optimization algorithms is often limited or determined by trust region methods [111] or line-search methods [104, 112]. Trust region methods are methods that limit the domain of the next guess for the minimizing argument to a region where the current model is considered accurate enough, while line search methods attempt to find a sufficiently decreasing step-size for a given descent direction. In practice, the derivatives of the cost or objective function $f(\mathbf{x})$ are often not available or difficult to calculate due to the presence of noise. If the derivatives are not explicitly available, they can be approximated by finite differences. However, the calculation of finite differences for the derivatives requires additional function evaluations and is in general not accurate in the presence of noise [113].

An alternative method is derivative-free optimization [114]. Some of the first and most simple derivative-free algorithms are random search algorithms, coordinate descent algorithms [115] and the Nelder-Mead or simplex method [116]. Michael Powell wrote one of his first breakthrough derivative-free optimization methods in 1964, which is referred to as Powell's method [117]. Later he developed the NEWUOA method in which he updates a quadratic model in a particular way to reduce computational complexity and to improve the convergence speed [118]. Each iteration, the algorithm establishes a quadratic model function and then finds the minimizing argument of the model within a trust region. Some of his other derivative-free optimization algorithms are based on similar principles, but allow the addition of (linear) constraints [119, 120]. These algorithms are amongst the top performing derivative-free local optimization methods in terms of number of measurements (convergence rate) and computational complexity [121]. However, they are not very robust with respect to high levels of noise [122].

Other more frequently used derivative-free optimization algorithms include pattern search methods [123, 124], simulated annealing [125], genetic or evolution based algorithms [126, 127], particle swarm optimization [128], ant colony optimization [129], or combinations such as hybrid genetic and swarm optimization [130]. These methods are often not very robust to noisy outliers in case the function $f(\mathbf{x})$ is noisy and the number of measurements are limited [131, 132]. In many practical applications, such as retinal OCT, it often happens that function evaluations of $f(\mathbf{x})$ are noisy and costly to obtain. For example, a costly function can be expensive in terms of actual cost or time. Hence, during the optimization the number of measurements needs to be limited but still needs to get close to the optimum of the objective function.

In some methods function evaluations are used to fit a surrogate model. In turn, the model is used to find an approximate optimum of f by finding the minimum of the surrogate model. Some of the previously mentioned methods are already based surro-

gate functions, those that do not can be used in combination with surrogate functions in order to improve robustness or limit the number of function evaluations [133, 134]. In Fig. 1.13 an example is given of derivative-free optimization using a non-linear surro-

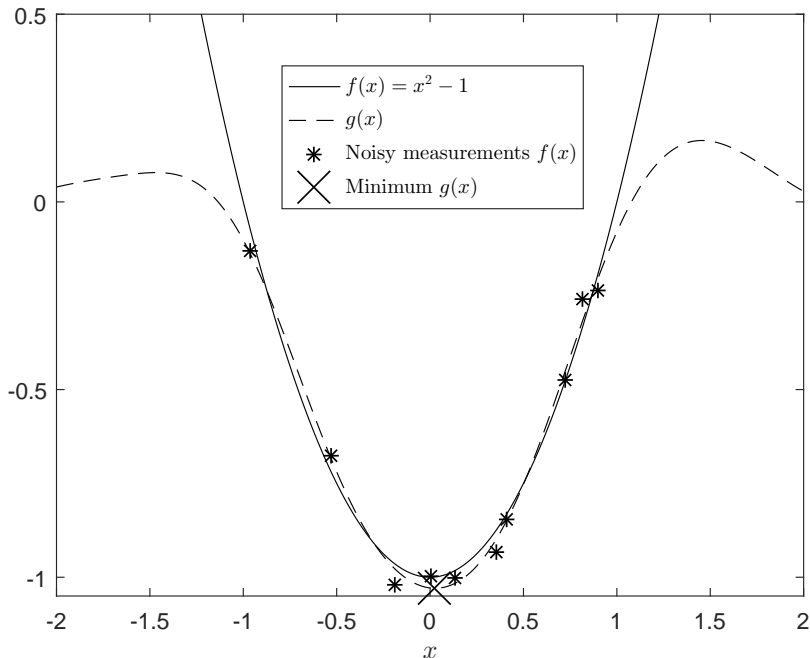


Figure 1.13: Example of optimization with a non-linear surrogate function. The unknown function $f(x)$ is approximated by $g(x)$ after ten noisy measurements. The minimum of $g(x)$ is then found and approximates the minimum of $f(x)$.

gate function. The surrogate function $g(x)$ is fitted through the measurements such that it approximates the unknown function $f(x)$. With only ten measurements, the optimum of $g(x)$ is already close to the optimum of $f(x)$. An example of a method that attempts to reduce the number of measurements is Bayesian optimization [135]. Bayesian optimization is becoming more and more popular for derivative-free optimization, because it aims to do global optimization over the space \mathcal{X} by using a surrogate model and it is robust towards zero-mean noise [136–139].

Many well-known derivative-free optimization methods have been used for WFS-AO [140]. Examples are hill-climbing algorithms [141], stochastic parallel-gradient-descent algorithms [142], the Nelder-Mead method [143], coordinate search methods [144], simulated annealing [145] and genetic algorithms [146]. When the number of measurements for the evaluation of the metric is limited, many of these algorithms falsely identify noisy outliers as an approximation of the optimum. Some WFS-AO methods rely on the repetition of an optimization procedure, such as pupil segmentation where the same optimization procedure is repeated to adjust the phase of each segment

in order to maximize the constructive interference of light in the focal spot [147].

It was shown that using model-based algorithms can improve the convergence rate of WFSL-AO methods [148–151]. In [148, 151], a quadratic polynomial was used to model the region around the optimum of the metric. The quadratic model was based on prior knowledge or prior measurements of the metric. This means that if there are large aberrations where the quadratic approximation is no longer valid, these methods will no longer work accurately. Others have focused on approximating certain metrics by linear models or using specific linear metrics [150, 152]. Their methods are often limited to specific metrics or applications. In [149, 153] neural networks are used to make an approximating model of the metric to aid in correcting aberrations. One advantage of neural networks is that they can approximate arbitrary continuous functions [154]. This gives them the advantage that the information of more measurements with large aberrations can be kept in the model. A drawback of these algorithms is that they are often computationally expensive and thus take more time to perform an optimization.

WFSL-AO-OCT was first attempted in 2009 with a simulated annealing algorithm [145] and in 2013 with a coordinate search algorithm [144]. In this thesis, we

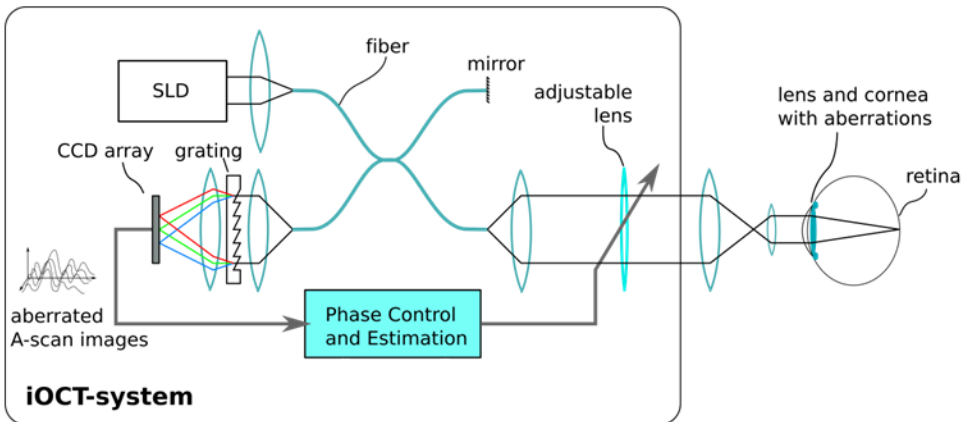


Figure 1.14: A wavefront sensorless adaptive optics optical coherence tomography setup.

combine OCT, sensorless AO and optimization to obtain a WFSL-AO algorithm that can correct the largest phase aberrations present in the human eye, is sufficiently fast to image human patients, and is robust with respect to noise. Furthermore, we demonstrate that the aberration correction can be performed with a deformable lens. These two improvements reduce the complexity, cost and size of AO-OCT systems. An example of the WFSL-AO-OCT system we will demonstrate is given in Fig. 1.14.

1.4. GOAL OF THIS THESIS

In ophthalmic use, the lateral resolution of OCT, which is given by the spot size of the sample arm light at the retina ($\sim 20 \mu\text{m}$), is hampered by aberrations present in the lens and cornea of the eye. Consequently, the detection of one of the most important parts of the retina, the cones, is not possible with standard commercial OCT systems. Therefore,

AO is used to compensate for the aberrations in the eye lens and cornea and combined with OCT systems to facilitate high-resolution imaging of the retina. AO has demonstrated in the last decade to expand the OCT imaging resolution in three different ways: an increased lateral resolution complementary to the high axial resolution of OCT, reduced speckle size and increasing the sensitivity to weakly reflecting biological objects. Despite these major advantages, the wide-scale clinical integration of AO in OCT technology is hampered by several reasons.

The first reason is the cost and complexity of this technology. Classical AO consists of the use of expensive wavefront sensors and deformable mirrors that can increase the cost of a commercial spectral-domain OCT (SD-OCT) up to forty percent. Moreover, the existing AO technology is difficult to integrate in OCT systems, because of the large footprint.

The second reason is that part of the photons reflected by the object (retina) are used by the wavefront sensor in AO systems, thus lowering the signal to noise ratio (SNR) of the OCT system. The loss in photons could also require a longer integration time for the camera and this would reduce the imaging speed. The temporal dynamics of the optical aberrations show a correlation with the cardiopulmonary system (1 to 2 Hz), which complicates the use of AO systems if the imaging speed is too low [155].

The third reason is that most AO correction in OCT is done for aberrations measured by the wavefront sensor. The light coming from the biological specimen and going to the OCT camera has a slightly different optical path than the wavefront sensor due to small misalignments. This results in non-common path errors in the aberration estimation and will result in poor AO performance.

The key objective of this thesis is to simplify the hardware of an AO-OCT system and to overcome the drawbacks of AO-OCT. This will be done in three stages that will transform the full AO-OCT setup shown in Fig. 1.7 into the more compact WFSL-AO-OCT setup shown in Fig. 1.14. First, the influence of the aberrations on the eye and the resulting OCT signal is determined in order to understand what wavefront sensorless (WFSL) AO approaches are applicable. Second, we exclude the use of the wavefront sensor and correct the wavefront aberrations directly with aberrations obtained from the OCT signal itself. In this step the aim is to develop a fast and robust algorithm capable of correcting the large aberrations that are present in the human eye. Third, we will replace the expensive deformable mirror with a cheap phase-diversity device, for example a deformable lens that can induce different aberrations in the optical path and correct wavefront aberrations in the OCT image. In combination these steps yield ultra-high resolution OCT based on novel optimization methods that corrects the wavefront aberrations without the wavefront sensor. Meeting these goals allows the realization of ultra-high resolution OCT to become more economically attractive on a wide clinical scale.

1.5. OUTLINE THESIS

In this section, we will discuss the outline of the other chapters in this thesis.

Chapter 2

In this chapter, we derive a model for optical wavefront aberrations in optical coherence

tomography. More precisely, the transfer function for optical wavefront aberrations in single-mode fiber based optical coherence tomography is determined. The loss in measured OCT signal due to optical wavefront aberrations is quantified using Fresnel propagation and the calculation of overlap integrals. A distinction is made between a model for a mirror and a scattering medium. The model predictions are validated with measurements on a mirror and a scattering medium obtained with an adaptive optics optical coherence tomography setup. Furthermore, a one-step defocus correction, based on a single A-scan measurement, is derived from the model and verified. Finally, the pseudo-convex structure of the optical coherence tomography transfer function is validated by showing convergence of the wavefront optimization with a hill climbing algorithm. The implications of this model for wavefront sensorless aberration correction are discussed.

The chapter is based on the following publications:

H. R. G. W. Verstraete, B. Cense, R. Bilderbeek, M. Verhaegen, and J. Kalkman, "Towards model-based adaptive optics optical coherence tomography," *Opt. Express*, vol. 22, no. 26, pp. 32 406–32 418, Dec 2014.

H. R. G. W. Verstraete, M. Verhaegen, and J. Kalkman, "Modeling the effect of wave-front aberrations in fiber-based scanning optical microscopy," in *Imaging and Applied Optics*. Optical Society of America, 2013, p. JTU4A.13.

Chapter 3

In this chapter, we describe the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm, a data-based optimization algorithm that is robust towards noisy measurements. Several sensor-less wavefront aberration correction methods that correct wavefront aberrations by maximizing the OCT signal are tested on an OCT setup. A conventional coordinate search method is compared to two model-based optimization methods. The first model-based method takes advantage of the well-known optimization algorithm (NEWUOA) and utilizes a quadratic model. The second model-based method (DONE) is developed by us and utilizes a random multi-dimensional Fourier basis expansion. The model-based algorithms achieve lower wavefront errors with up to ten times fewer measurements. Furthermore, the newly proposed DONE method outperforms the NEWUOA method significantly. The DONE algorithm is tested on OCT images and shows a significantly improved image quality.

The chapter is based on the following publications:

H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Model-based sensor-less wavefront aberration correction in optical coherence tomography," *Opt. Lett.*, vol. 40, no. 24, pp. 5722–5725, Dec 2015.

H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Numerical evaluation

of advanced optimization algorithms for wavefront aberration correction in OCT,” in *Imaging and Applied Optics 2015*. Optical Society of America, 2015, p. AOM3E3.

Chapter 4

This chapter analyzes the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm, an online optimization algorithm that iteratively minimizes an unknown function based on costly and noisy measurements. The algorithm maintains a surrogate of the unknown function in the form of a random Fourier expansion (RFE). The surrogate is updated whenever a new measurement is available and is subsequently used to determine the next measurement point. The algorithm is comparable to Bayesian optimization algorithms, but its computational complexity per iteration does not depend on the number of measurements. We derive several theoretical results that provide insight on how the hyper-parameters of the algorithm should be chosen. The algorithm is compared to a Bayesian optimization algorithm for a benchmark problem and three applications, namely, optical coherence tomography, optical beam-forming network tuning and robot arm control. It is found that the DONE algorithm is significantly faster than Bayesian optimization in the discussed problems, while achieving a similar or better performance.

The chapter is based on the following publications:

H. R. G. W. Verstraete, L. Bliet, M. Verhaegen, and S. Wahls, “Online optimization with costly and noisy measurements using random Fourier expansions,” *IEEE Trans. Neural Netw. Learn. Syst.* [Accepted], 2016.

Chapter 5

In this chapter, we apply the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm to *in vivo* AO-OCT imaging. We optimize the optical coherence tomography signal and image quality during *in vivo* imaging of the human retina using wavefront sensorless adaptive optics. The optical wavefront aberrations are corrected using a novel multi-actuator deformable lens which is linearized for the hysteresis in the piezoelectric actuators.

The DONE algorithm succeeds in drastically improving the optical coherence tomography signal while achieving a computational time of 1 ms per iteration, making it applicable for many high-speed applications.

The chapter is based on the following publications:

H. R. G. W. Verstraete, M. Heisler, M. J. Ju, D. Wahl, L. Bliet, J. Kalkman, S. Bonora, M. Sarunic, Y. Jian, and M. Verhaegen, “*In vivo* wavefront sensorless adaptive lens OCT with the DONE algorithm,” [Submitted], 2016.

1

Chapter 6

The Data-based Online Nonlinear Extremum-seeker (DONE) algorithm is not only capable of optimizing the signal and quality of images in optical coherence tomography, but can be applied to a multitude of optimization problems. In this chapter, we apply the DONE algorithm to a Smart Programmable Array Microscope (S-PAM) to correct optical wavefront aberrations. In this example, the DONE algorithm was modified to work with a sliding-window principle. Older measurements are forgotten so that the algorithm adapts to slowly changing aberrations.

The chapter is based on the following publications:

P. Pozzi, D. Wilding, O. Soloviev, H. R. G. W. Verstraete, L. Bliet, G. Vdovin, and M. Verhaegen, "Real time wavefront sensorless aberration correction in digital micromirror based confocal microscopy," *Opt. Express [Accepted]*, 2016.

Finally, in the last chapter the conclusions drawn throughout this thesis are repeated and recommendations for future work are given.

2

TOWARDS MODEL-BASED ADAPTIVE OPTICS OPTICAL COHERENCE TOMOGRAPHY

The transfer function for optical wavefront aberrations in single-mode fiber based optical coherence tomography is determined. The loss in measured OCT signal due to optical wavefront aberrations is quantified using Fresnel propagation and the calculation of overlap integrals. A distinction is made between a model for a mirror and a scattering medium model. The model predictions are validated with measurements on a mirror and a scattering medium obtained with an adaptive optics optical coherence tomography setup. Furthermore, a one-step defocus correction, based on a single A-scan measurement, is derived from the model and verified. Finally, the pseudo-convex structure of the optical coherence tomography transfer function is validated with the convergence of a hill climbing algorithm. The implications of this model for wavefront sensorless aberration correction are discussed.

Parts of this chapter have been published in [156].

©2014 Optical Society of America. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this chapter are prohibited.

<https://www.osapublishing.org/oe/abstract.cfm?uri=oe-22-26-32406>

2.1. INTRODUCTION

Optical coherence tomography (OCT) is a technique for non-invasive, *in vivo* imaging of tissue [2, 21]. Its main application is found in ophthalmology, where it is used for the 3D imaging of the cornea and retina [24, 92, 93]. The axial resolution of OCT is obtained through low coherence interferometry and is inversely proportional to the source bandwidth. Using ultra-broadband sources, axial resolutions below $1\ \mu\text{m}$ have been recorded [157]. The lateral resolution of OCT is determined by conventional optical lens focusing. Hence, the lateral resolution improves by using a larger pupil. In OCT imaging of the retina the lateral resolution is hampered by the small pupil size ($< 2\ \text{mm}$). Moreover, when the pupil size is increased, large ocular aberrations are introduced. It has been demonstrated that high lateral resolutions can be obtained by using adaptive optics (AO) to correct the optical wavefront aberrations on large pupils ($> 2\ \text{mm}$). Combining the high axial resolution of OCT with the high lateral resolution of AO results in ultra-high resolution AO-OCT imaging in three dimensions. Such systems have been reported in [96–99] and demonstrated lateral and axial resolution up to $3\ \mu\text{m}$ and $2\text{--}3\ \mu\text{m}$, respectively. As a result, AO-OCT has made it possible to image the 3D architecture of individual rods and cones *in vivo* in the human eye [99, 100].

Current AO-OCT setups usually rely on wavefront sensors such as the Shack-Hartmann wavefront sensor. In general, wavefront sensors have several drawbacks. First, a loss of signal to noise ratio in the OCT signal occurs because light from the object is directed away from the sensor (if no alternative imaging device is used as a wavefront sensor). Second, specular reflection of optical components or from the eye prevent the wavefront sensor from giving an accurate estimation of the aberrations. Third, the magnitude and accuracy of the measured wavefront aberrations are limited by the wavefront sensor design. Due to the large aberrations present in the human eye, a trade-off is usually made between the maximum aberration that can be measured and the accuracy of the wavefront estimation. Fourth, since wavefront sensors are located outside the imaging path, non common-path errors with the imaged signal occur. In other words different aberrations are estimated by the SH wavefront sensor than are present in the optical path for the OCT image. Finally, the cost of the wavefront sensor increases the overall cost of any adaptive optics system. This increased cost hampers commercial and medical use of AO-OCT devices. Several authors have successfully applied wavefront sensorless (WFS) AO algorithms to OCT [144, 145, 158] to mitigate these disadvantages. WFS approaches are based on phenomenological image quality metrics that are optimized to achieve aberration free images. If analytical expressions for the transfer function of the OCT signal for different aberrations are known, more efficient metrics and algorithms can be found and the region of convergence can be determined. For example, in [148, 151] it is shown that the use of models can improve the convergence speed of WFS algorithms. In [103] a general model is used, based on a synthetic aperture and Zernike modal decomposition, for a software based image optimization. The optimization is based on a sharpness and intensity metric.

Until now, only single aberrations have been modeled for OCT. The tip/tilt aberrations are modeled in the context of galvanometric mirrors, that are used for scanning across the sample. The defocus aberration is studied in the context of the axial transfer function for single-mode fiber (SMF) based OCT systems. Studies [159–162] showed that

the amount of backscattered signals, which is due to optical properties of the tissue, is also influenced by the optical components of the OCT setup. A simplified axial transfer function is proposed in [162] based on the overlap of the Gaussian beam profile arising from the SMF. The proposed axial transfer function assumes the propagation of a perfect Gaussian beam profile without lateral optical wavefront aberrations other than a fixed defocus introduced by the lens. This transfer function is used to correct for OCT system properties in the backscattered measured signal to determine the attenuation coefficient of tissue from OCT scans. In general, OCT transfer function models distinguish between a specular reflector and a backscattered sample. In the axial defocus model in [162] the interaction of a Gaussian beam with the scattering medium is modeled by removing the defocus curvature of the Gaussian beam at the sample.

The goal of this paper is to model and validate the effect of arbitrary optical wavefront aberrations on the OCT signal for mirror and scattering media reflectors. This model is used to correct for defocus and predict the convergence of a WFS algorithm.

2.2. OCT MODEL

OCT is based on a combination of confocal and coherent gating [22]. Confocal gating is implemented by single mode fibers that act as pinholes for outgoing and returned light. Coherence gating is based on low coherence, i.e. large optical bandwidth interference, to detect only a thin section of the sample. The confocal and coherent detection of the OCT system all take place in the OCT sample arm, which can be modeled as in Fig. 2.1. A first lens (CL), conjugated to the sample lens, is used to collimate the beam. A second

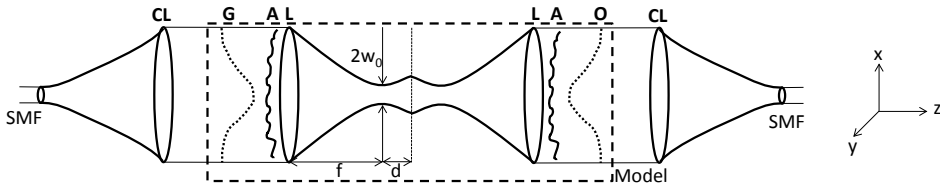


Figure 2.1: Geometry model for SMF-OCT. A collimator lens (CL) collimates the beam to a planar wavefront with Gaussian distribution (G). The sample lens (L) with aberrations (A) focuses the light on the sample at distance $f + d$. Light is reflected/scattered by the sample, shown here in transmission. The overlap integral (O) with the Gaussian mode is calculated to quantify the coupling efficiency into the SMF.

lens (L), the sample lens, which contains aberrations (A), is used to focus the light on the sample located a distance d from the actual focus f of the sample lens. Light reflected from a thin slice of the sample goes through the same path back to the SMF, which is shown in this model as transmission. The sample can be a mirror or a thin slice of tissue, which we model here separately.

2.2.1. OCT MIRROR MODEL

After collimation by the collimator lens the optical field emitted from the single mode fiber has a Gaussian distribution with a flat wavefront. This planar Gaussian wavefront (G) is described by

$$G(x, y) = C \exp\left(-\frac{1}{w^2}(x^2 + y^2)\right). \quad (2.1)$$

Here, C is the amplitude, and w is the beam waist of the collimated fundamental Gaussian fiber mode. Next, the wavefront traverses optical aberrations. The aberrations (A) in the pupil plane are denoted as

$$A(x, y, \alpha) = \exp(ikW(x, y)). \quad (2.2)$$

Here, k is the wave number $\frac{2\pi}{\lambda}$, i is the imaginary number $\sqrt{-1}$. Normalized Zernike polynomials are used over a circular pupil [163] to represent $W(x, y) = \alpha^T Z(x, y)$, the vector α contains the corresponding Zernike coefficients, $Z(x, y)$ is a vector of the corresponding Zernike polynomials.

The aberrated optical wavefront is focused by a lens. The paraxial approximation of the focusing sample lens (L) with a circular pupil is given by

$$L(x, y) = \exp(-ikf) \exp\left(\frac{-ik}{2f}(x^2 + y^2)\right) \text{circ}\left(\frac{\sqrt{x^2 + y^2}}{r_{pupil}}\right). \quad (2.3)$$

For the lens the Fresnel approximation is used with focal length f . Following Goodman [33], the circular pupil function $\text{circ}(\sqrt{x^2 + y^2}/r_{pupil})$ is a circle with radius r_{pupil} . Values inside the circle are set to 1, outside the circle to 0, and on the border to 0.5. The pupil function is determined by the smallest pupil in the optical setup, in our case the deformable mirror, which limits the beam radius to $r_{pupil} = 3.61$ mm.

Fresnel diffraction (F) [33] approximates the propagation of an arbitrary wavefront $U(x_1, y_1)$ to the plane (x_2, y_2) over an OPL z as,

$$U_z(x_2, y_2) = F(U(x_1, y_1), z) = \frac{\exp(ikz)}{i\lambda z} \iint U(x_1, y_1) \exp\left(\frac{ik}{2z}[(x_2 - x_1)^2 + (y_2 - y_1)^2]\right) dx_1 dy_1. \quad (2.4)$$

In the mirror model the Gaussian distributed wavefront (G) passes through aberrations (A) and the sample lens (L) and propagates (Fresnel diffraction) over a distance $f + d$ to reflect on the mirror and then propagates backwards over the distance $f + d$. We describe the latter by a Fresnel propagation over $2(f + d)$. Hence, the wavefront right before the sample lens (L) is

$$U_1(x_3, y_3, d, \alpha) = F[G(x_1, y_1)L(x_1, y_1)A(x_1, y_1, \alpha), 2(f + d)]. \quad (2.5)$$

Again, wavefront U_1 traverses the sample lens (L) with aberrations (A). Finally, an overlap integral (O) of the field with the Gaussian mode of the collimated beam is taken. Since the SMF only accepts a Gaussian wavefront identical to the collimated Gaussian mode, we take the overlap integral of the reflected light with this Gaussian wavefront to determine the total field coupled back into the fiber from the sample. The overlap integral [164] for an arbitrary wavefront $U(x, y)$ with the Gaussian mode $G(x, y)$ is defined as

$$O(U(x, y)) = \frac{\iint_{-\infty}^{\infty} U(x, y)G^*(x, y) dx dy}{\iint_{-\infty}^{\infty} |G(x, y)|^2 dx dy}. \quad (2.6)$$

The total intensity coupled back into the SMF is then denoted by

$$h_1(d, \alpha) = |O[U_1(x_3, y_3, d, \alpha)L(x_3, y_3)A(x_3, y_3, \alpha)]|^2. \quad (2.7)$$

2.2.2. SCATTERING MEDIUM OCT MODEL

The scattering medium transfer function is very similar to the mirror transfer function. The only difference is the interaction with the scattering medium. The Gaussian beam (G) traverses the sample lens (L) with aberrations (A) and interacts with the sample after propagating an OPL $f + d$. In the scattering medium we assume that the lateral phase relation of the wavefront is lost and a new plane wave is formed with a perfect planar wavefront. This is modeled by taking the modulus of the field, $|U(x, y)|$, at the position of the sample, which reduces the original phase of the wavefront laterally to a constant phase. The wavefront U_2 represents the wavefront after having interacted with the sample.

$$U_2(x_2, y_2, d, \alpha) = |F[G(x_1, y_1)L(x_1, y_1)A(x_1, y_1, \alpha), f + d]| \quad (2.8)$$

The wavefront U_2 propagates an OPL of $f + d$ to the aberrated sample lens, which is represented by the wavefront U_3 ,

$$U_3(x_3, y_3, d, \alpha) = F[U_2(x_2, y_2, d, \alpha), f + d]. \quad (2.9)$$

Wavefront U_3 again traverses the sample lens (L) with aberrations (A) and the intensity coupled into the fiber is determined with the overlap integral (O). The total intensity coupled into the SMF fiber is

$$h_2(d, \alpha) = |O[U_3(x_3, y_3, d, \alpha)L(x, y)A(x, y, \alpha)]|^2. \quad (2.10)$$

The functions $h_1(d, \alpha)$ and $h_2(d, \alpha)$ are called OCT transfer functions for the mirror and scattering medium, respectively. These equations represent the intensity that is coupled back into the fiber. It is important to note that the OCT signal, defined as the magnitude of the Fourier transform, scales with the sample arm field and thus with $\sqrt{h_1}$ and $\sqrt{h_2}$. The results for the transfer functions are numerically calculated in MATLAB using the numerical Fresnel propagation code in [165].

2.2.3. SINGLE STEP DEFOCUS CORRECTION

If the AO-OCT setup and sample lens are well-known and calibrated, then a single step correction for the defocus can be implemented. The change in focal length d is related to the Zernike defocus coefficient α_4 and the original focal length f as follows,

$$-\frac{1}{2(f+d)}(x^2 + y^2) = -\frac{1}{2f}(x^2 + y^2) + 2\sqrt{3}\alpha_4 \frac{(x^2 + y^2)}{r_{pupil}^2}. \quad (2.11)$$

In Eq. (2.11) the left hand side is the quadratic phase factor for the lens focused at a distance $f + d$ and the second term on the right hand side is the Zernike defocus. As a result, the defocus Zernike coefficient is derived as

$$\alpha_4 = \frac{r_{pupil}^2}{4\sqrt{3}} \left(\frac{1}{f} - \frac{1}{f+d} \right). \quad (2.12)$$

In the equations above r_{pupil} is the radius of the pupil, defined by the smallest pupil in the setup (the deformable mirror), on which the Zernikes are defined. The index proposed by Thibos et al. [166] is used to order the Zernike coefficients α (4 corresponds to the defocus). Equation (2.12) allows to directly calculate the defocus coefficient α_4 from the distance d the sample is out of focus.

2.3. MATERIALS AND METHODS

AO-OCT measurements are performed on an adaptive optics OCT system based on single mode fibers and a Michelson interferometer, shown in Fig. 2.2. The light source is a fiber coupled super luminescent diode (Superlum Broadlighter D-840-HP-I) with a bandwidth of 100 nm and a center wavelength of 840 nm. The fiber from the source is a single mode fiber with a $4\ \mu\text{m}$ core diameter coupled to a fiber coupler (Gould Fiber Optics Corning HI-780) with a splitting ratio of 50/50. Light is detected with a spectrometer (Wasatch Photonics Cobra UHR) with 4096 pixels that cover a wavelength range from 650 nm to 950 nm. The integration time and line-time are set to $23\ \mu\text{s}$ and $25\ \mu\text{s}$, respectively.

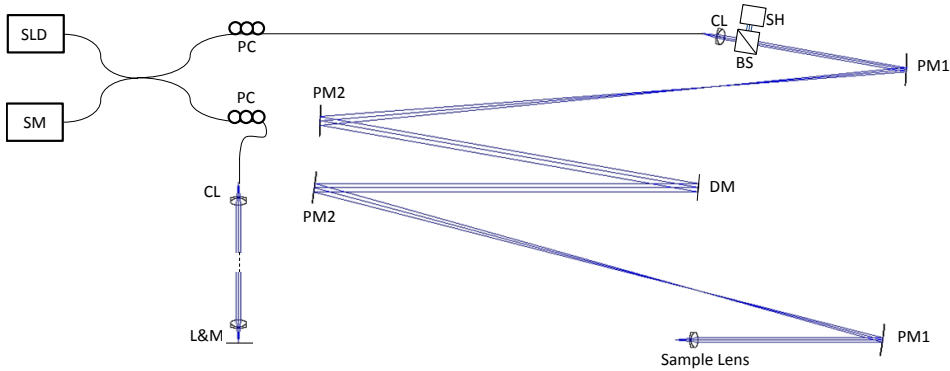


Figure 2.2: Schematic overview of the AO-OCT Setup. SMF fibers are indicated by a single line. The free space light is indicated by three lines. SLD = super luminescent diode. SM = spectrometer. PC = polarization controllers. CL = fiber collimator lens. L&M are the lens and mirror for the reference arm. SH = Shack-Hartmann wavefront sensor. BS = beam-splitter. PM1 and PM2 are parabolic mirrors. DM = deformable mirror.

The two arms from the fiber-based interferometer both have 3-paddle polarization controllers (Thorlabs FPC560) and light exiting the sample and reference arm fibers is collimated by a fiber collimator lens (Thorlabs AC254-030-B). The reference arm consists of a folded, collimated beam covering a distance of 3.8 m. In both sample and reference arm a collimated Gaussian beam with a $3.4\ \text{mm}$ beam waist (e^{-2}) starts from the fiber collimator lens. The sample arm consists of pairs of parabolic mirrors that are used to conjugate the planes of the fiber collimator, deformable mirror (DM) and sample lens. The parabolic mirrors PM1 (Edmund Optics $50.8 \times 304.8\ \text{mm}$ PFL 15°) and PM2 (Edmund Optics $50.8 \times 635.0\ \text{mm}$ PFL 15°) in Fig. 2.2 are off-axis parabolic gold mirrors that (de)magnify the beam 2.08 times. The deformable mirror (Imagine Eyes Mirao52) has 52 actuators and a stroke of $50\ \mu\text{m}$. At the end of the reference and sample arm the sample lens with a focal length of $45\ \text{mm}$ (Thorlabs AC254-045-B) focuses the light on a mirror (Thorlabs KM100-E03) and the sample, respectively. An OKotech Shack-Hartmann (SH) wavefront sensor, 1-inch optical format, with a lenslet array pitch of $300\ \mu\text{m}$ and focal length $18.6\ \text{mm}$, is placed perpendicular to the optical beam path. A pellicle beam splitter (Thorlabs BP108) directs 8% of the light towards the SH and 92% to the fiber. The SH pupil is conjugated to the pupil plane. The SH mask to CCD distance is $16.56\ \text{mm}$.

From the SH images the modal wavefront is reconstructed using the method described in [167]. The diameter of the DM is the smallest pupil and acts as a field stop

that determines the maximum pupil size. The DM diameter is 15 mm and after conjugation through the parabolic mirrors the pupil radius is $r_{pupil} = 3.61$ mm. This pupil radius, r_{pupil} , is used for the wavefront reconstruction in the SH sensor as well as in the circular pupil function for the model.

The acquired spectra are processed into OCT A-scans in the following way. First, the spectrum of the reference arm is subtracted. Second, since the pixels of the spectrometer correspond to a non-equidistant k axis, the data is linearly interpolated in k based on a calibration table provided by the manufacturer. After interpolation the fast Fourier transform (FFT) is taken over the spectrum. The magnitude of the FFT is used as the OCT-signal.

The measured axial resolution is $3.9 \mu\text{m}$ (FWHM), which is in good agreement with the coherence length of the Superlum source spectrum, which is $3.5 \mu\text{m}$ (FWHM). The sensitivity is determined to be -94 dB by measuring the signal to noise ratio between peak signal and the RMS noise.

Odd and even Zernike aberrations are set on the DM, such that perfect Zernike aberrations are displayed in the conjugate plane of the sample lens. Normalized Zernike aberrations with index 1 to 9 are applied to the DM (tip, tilt, defocus, 2 astigmatisms, 2 comas and 2 trefoils). The tip and tilt are left out for the scattering medium, because these aberrations correspond to lateral scanning over the sample, which is performed by a translation stage. The aberration coefficients are varied between $-0.45 \mu\text{m}$ to $0.45 \mu\text{m}$.

For the mirror measurements 512 A-scans are taken at the same lateral position. All of the OCT signal in the A-scans is then averaged over time and all depth to a scalar and normalized (divided by the maximum value in a set of measurements). This scalar represents the OCT signal for the mirror measurements.

For the scattering medium measurement, 512 A-scans are taken of a sample consisting of 4 layers of Scotch tape attached on a translation stage. These 512 A-scans are averaged over time and out of the averaged A-scan four depth points at zero aberrations are taken. Several different aberrations are applied to the DM and the same four depth points are scaled by the same normalization factor. These four points represent the OCT signal for the scattering medium measurements.

2.4. RESULTS

Figure 2.3 shows the OCT measurements on a mirror for varying aberrations applied on the DM. The measurements are plotted together with numerical results obtained for the mirror model. The mirror model is in good agreement with the measurements, especially for small aberrations. The model predicts almost no loss in OCT signal for the odd aberrations (index 1-2, 6-9), because these aberrations are canceled due to the double pass through the optical system. The higher order odd aberrations (comas and trefoils) show a loss of OCT signal stronger than the results predicted by the model. We attribute this to slight misalignments, which cause an imperfect cancellation of the odd aberrations leading to the introduction of other aberrations in the wavefront.

Figure 2.4 shows the measured Zernike coefficients on the SH sensor, for modes with index 1 to 20, obtained from the SH wavefront sensor. For even aberrations applied on the DM (index 3 to 5) the measured wavefront aberrations on the SH are about twice as large as those that have been put on the DM, i.e. the slope of the measurements on the

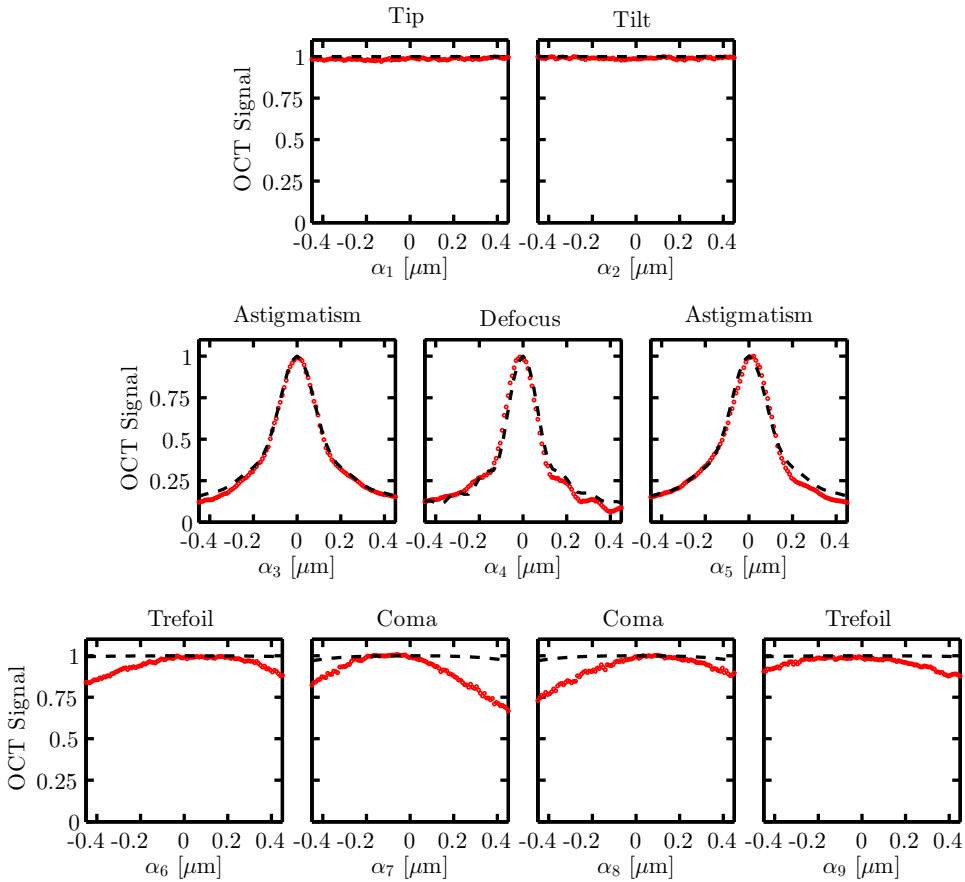


Figure 2.3: OCT signal measurements on a mirror versus applied aberration for index 1-9. Measurements (red markers) and numerical results (dashed black line) of the transfer function $\sqrt{h_1(0, \alpha)}$ for the OCT signal are shown. The standard deviation of the measurements is smaller than the marker size.

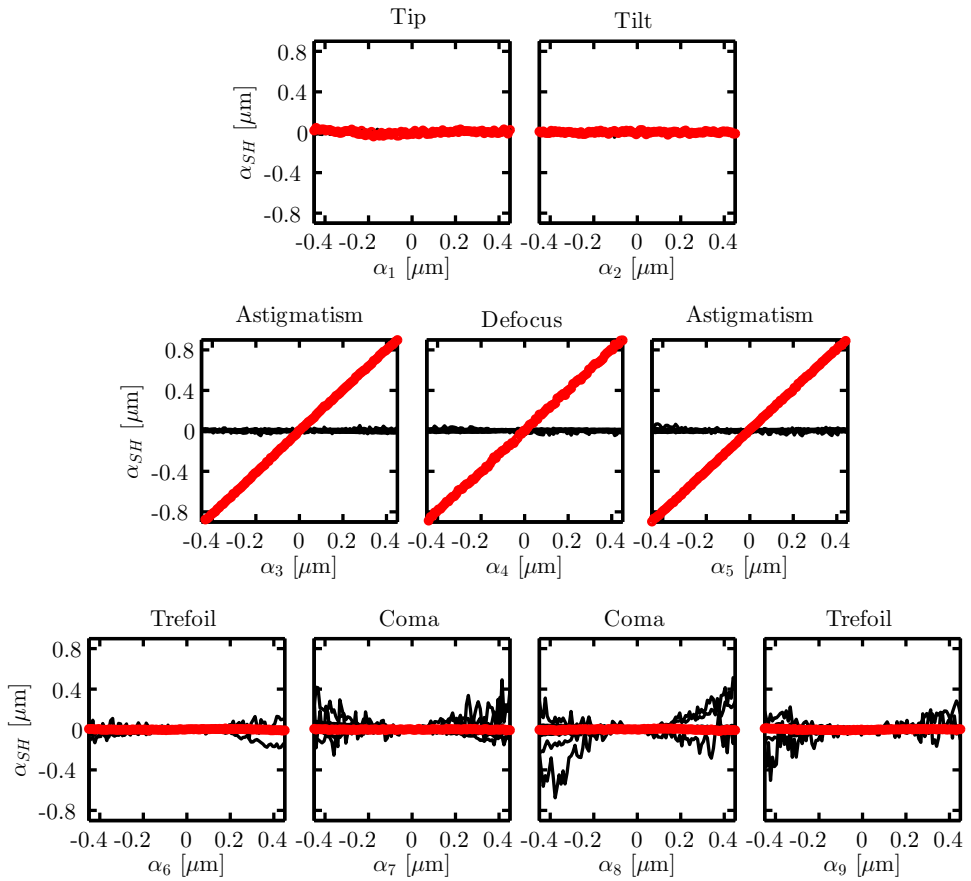


Figure 2.4: Shack-Hartmann wavefront measurements on a mirror versus applied wavefront aberration, for index 1-9. Measured wavefront aberrations are index 1-20. The measured aberration that is identical to the applied aberration is in red, all other aberrations are indicated in black.

SH sensor for even aberrations in Fig. 2.4 is about 2. The Zernike coefficients of the SH sensor show that for odd aberrations (index 1 to 2 and 6 to 9) the wavefront aberration cancels in the SH measurement. However, other aberrations appear (mostly tip, tilt, and defocus) due to imperfect cancellation of the odd aberrations.

Figure 2.5 shows OCT measurements on the scattering medium versus the applied aberration for index 3 to 9. The OCT measurements have a greater standard deviation than the mirror measurements due to movement of the sample and the much lower signal strength. The measurements are compared to the scattering medium model and show good agreement to the scattering medium model. It can be observed that the obtained functions for the scattering medium for astigmatisms and defocus are broader than the corresponding functions for the mirror measurements. The FWHM for the defocus curve of the mirror model is approximately $0.2 \mu\text{m}$, while the FWHM for the defocus curve of the scattering medium model is approximately $0.4 \mu\text{m}$. Moreover, the transfer functions have a global maximum at zero aberrations for the given aberrations from -0.45 to $0.45 \mu\text{m}$. In comparison to the mirror model a clear maximum is visible for both odd and even applied aberrations.

The corresponding Zernike coefficients, index 1 to 20, measured on the SH wavefront sensor are shown in Fig. 2.6. The SH sensor measurement gives aberrations similar to the aberration applied to the DM, even though the optical path passes the DM twice. This means that light is not specularly reflected, but diffuse and that the phase of the wavefront is lost after interaction with the scattering medium.

Based on the validity of our wavefront model we apply the model for one step defocus correction, as shown in Fig. 2.7. For this experiment the sample is moved a fixed distance out of focus (0.39 mm in Fig. 2.7). The optical path length moved out of focus, d , is determined by the displacement of the target depth of the sample on the OCT A-scan measurement and used to calculate α_4 based on Eq. (2.12). Other parameters used in this equation are the pupil radius $r_{pupil} = 3.61 \text{ mm}$, and the focal length of the sample lens $f = 45 \text{ mm}$. The theoretical prediction of Eq. (2.12) for all displacements is shown in Fig. 2.7(c) as the black line. The red dots correspond to the Zernike defocus coefficient, α_4 , that maximized the OCT signal at that depth in the A-scan. Figure 2.7 shows the sample before Fig. 2.7(a) and after Fig. 2.7(b) defocus correction, both are displayed with the same intensity scale. Clearly, the sample in focus has a higher OCT signal.

Finally, we demonstrate that the model gives us insight into sensorless wavefront correction. We test a sequential optimization algorithm similar to [144] for aberrations with RMS error smaller than $0.4 \mu\text{m}$. The applied aberrations are limited to the Zernike modes with index 3 to 9 (similar to Fig. 2.5). The results for this test are shown in Fig. 2.8. Figure 2.8(a) shows a single optimization for a random aberration applied to the DM. It shows the magnitude of the merit function (OCT signal) while the algorithm scans through Zernike modes with index 3 to 9 for the first time and adds the argument corresponding to the maximum for each mode to the DM to correct this aberration. In Fig. 2.8(b) the value of the merit function is shown after the algorithm has converged for 100 different starting aberrations, all with RMS error smaller than $0.4 \mu\text{m}$. The test shows that the sequential optimization algorithm converges to the same magnitude for 100 different initial aberrations all with an RMS wavefront error smaller than $0.4 \mu\text{m}$. On average it takes 20 seconds for the algorithm to converge.

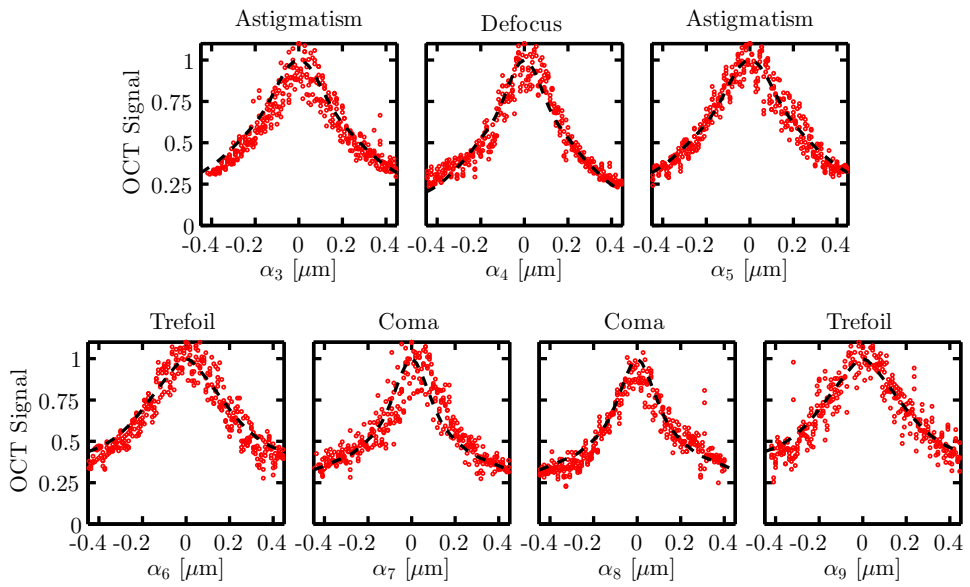


Figure 2.5: OCT signal measurements on a scattering medium versus applied aberration for index 3-9. Measurements (red markers) and numerical results (dashed black line) of the transfer function $\sqrt{h_2(0, \bar{\alpha})}$ for the OCT signal are shown.

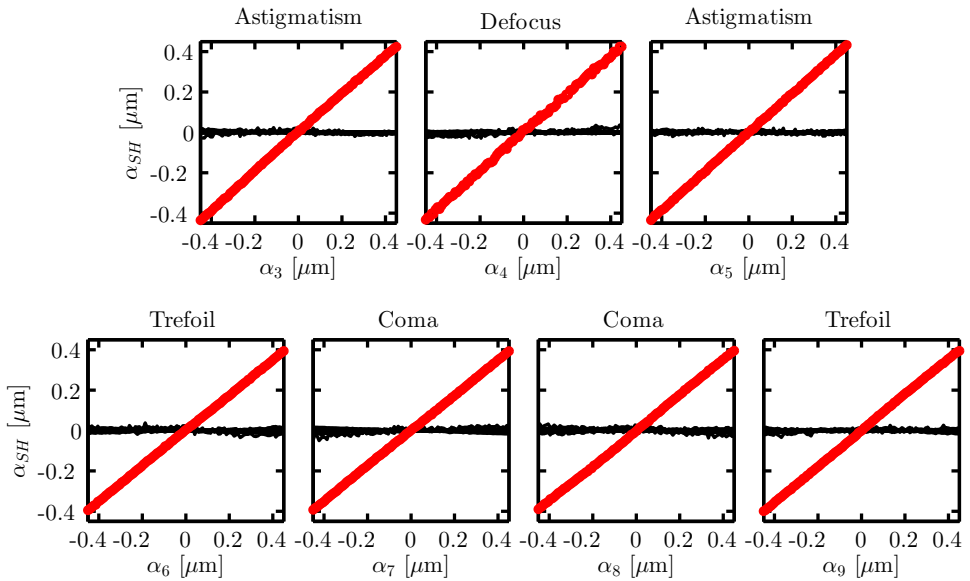


Figure 2.6: Shack-Hartmann wavefront measurements on a scattering medium versus applied wavefront aberration, for index 3-9. Measured wavefront aberrations are index 1-20. The measured aberration that is identical to the applied aberration is in red, all other aberrations are indicated in black.

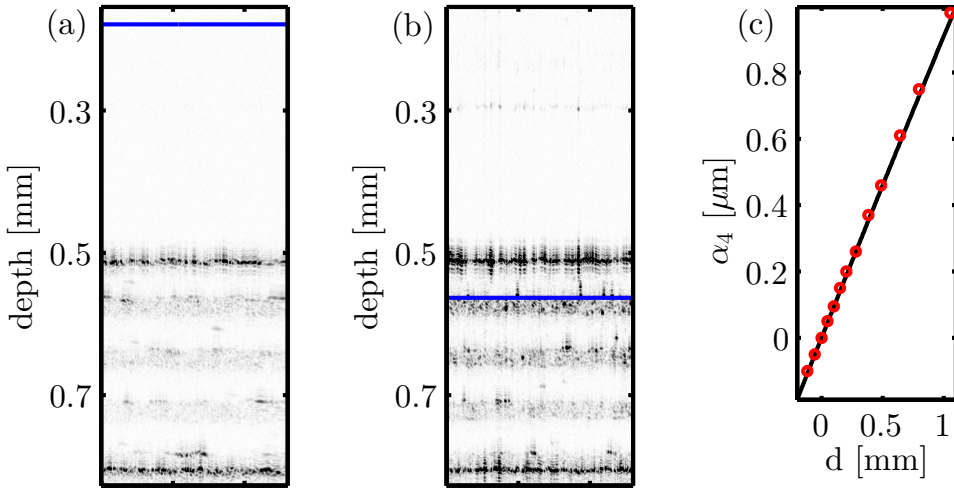


Figure 2.7: OCT B-scans of the tissue phantom. (a) Out of focus. (b) After single shot focus correction. Blue lines indicate the location of the focal plane. (c) Results of the one shot defocus correction. Zernike coefficients are determined from Eq. (2.7) (solid black line) and from measurements (red circles).

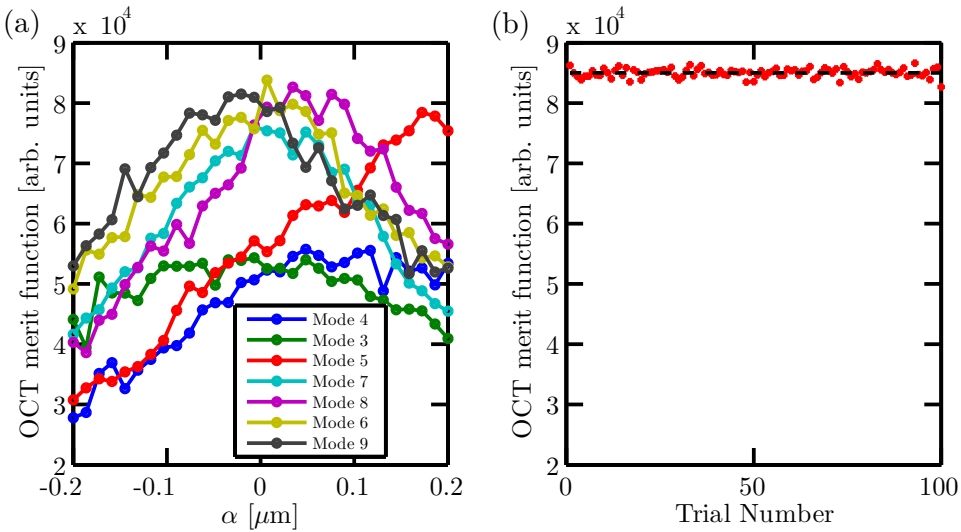


Figure 2.8: (a) Merit function (value of OCT signal) during a single step in the sequential optimization process. Only the first search iteration of every Zernike mode is shown for index 3-9. Further iterations resulted in an increased value of the merit function. (b) Optimized merit function after complete optimization process for 100 random aberrations with RMS wavefront error $<0.4 \mu\text{m}$. For all 100 aberrations the maximum OCT signal is found.

2.5. DISCUSSION

The transfer functions for the OCT signal for mirror and scattering medium in Eq. (2.7) and Eq. (2.10) are calculated based only on experimental parameters. Both transfer functions match the experimental results well, as shown in Fig. 2.3 and 2.5.

The Shack-Hartmann measurements show a large difference between the wavefront returning from the mirror and the scattering medium. The phase aberrations of the reflected wavefront are doubled for even aberrations and are canceled for odd aberrations for measurements on the mirror. The specular reflection of the mirror causes the odd aberrations to cancel and the even aberrations to double. Hence, these aberrations are difficult to detect from a decrease in the OCT signal of a specular reflector. Even though the OCT signal for the mirror measurements does not decrease a lot for the odd aberrations (index 1-2 and 6-9) in Fig. 2.3, our calculations show that the lateral resolution (quantified by the FWHM of the focal spot on the sample) decreases for an increasing magnitude of the odd aberrations [168]. Hence, image quality metrics should not only be based on intensity but also on sharpness when dealing with specular reflecting objects, which might also occur in OCT. This has been proposed on an ad-hoc basis [169, 170], but clearly follows from our analysis.

Our model and measurements show that the lateral phase relations are lost when the wavefront interacts with the scattering sample as the Shack-Hartmann measures the same size of aberration as applied to the DM. The axial defocus model of van Leeuwen et al. [162] for a scattering medium is also based on removal of defocus wavefront curvature of the Gaussian beam after it interacted with the scattering medium. This supports the use of our phase reset operator in the scattering medium model, which is based on the same principle. Others have reported partial loss of wavefront phase upon reflection from a scattering medium [171]. This may be attributed to the presence of both specular and diffuse reflection components in their measurement.

To further validate the proposed OCT transfer functions, they are compared with the well established Lorentzian axial transfer functions from [162]. The numerical results of the OCT signal transfer functions $h_1(d, 0)$ and $h_2(d, 0)$ for an axial displacement of the sample, d , for the mirror and scattering medium are similar to the two different Lorentzian functions proposed in [162] when the pupil is large enough compared to the Gaussian waist of the beam. For large pupils, $h_1(d, 0)$ and $h_2(d, 0)$ correspond to the Lorentzian functions, similar to the analytical expression in [162, 168]. The developed transfer functions can be used to correct OCT A-scans for system properties including all aberrations. Hence, potentially improving the estimation of tissue attenuation coefficients from OCT data in the presence of other aberrations besides defocus.

These models can be related to another imaging technique, scanning laser ophthalmoscopy (SLO), however, this technique is not based on interferometry. The image is formed by collecting all the backscattered light from different depths in the sample and this requires further research.

Theoretical estimation of the defocus Zernike coefficient to correct for a defocus caused by an axial displacement is in excellent agreement with the Zernike coefficient obtained from an optimization maximizing the OCT signal at the given depth. The one step defocus correction can be applied to scanning OCT systems to quickly set the optimal defocus or to easily change the focal point to another depth in the sample. Only well

calibrated and well known optical systems allow for this one step defocus correction, in other cases it might be considered as a good first estimate. Note that the defocus is the primary aberration in the human eye [166] and that its corrections leads to the strongest image quality improvement of all aberrations.

2

The scattering medium OCT signal transfer functions for the single aberrations all exhibit pseudo-convex properties, i.e. they have a global maximum for zero aberrations. A validation for the pseudo-convex properties is given by the convergence results shown in Fig. 2.8, which show that the algorithm always converged towards the same metric value. The pseudo-convexity shows why hill climbing algorithms such as proposed in [144, 145] eventually converge towards a maximum signal. A clearer understanding of how aberrations influence the OCT signal can help in selecting metrics for wavefront sensorless aberration correction, which can lead to faster signal convergence. The OCT transfer functions, as developed here, also allow testing and simulation of new optimization algorithms prior to *in vivo* application.

2.6. CONCLUSION

Using Fresnel propagation, overlap integrals and Gaussian distributions two OCT transfer functions modeling the effect of lateral aberrations for a mirror and a scattering medium on the OCT signal are derived. Measurements on a mirror and a scattering medium with an AO-OCT system closely resemble the proposed transfer functions. A one step defocus correction method has theoretically been derived and successfully applied. The pseudo-convex nature of the transfer function is validated by the 100% convergence success of a hill-climbing algorithm.

3

MODEL-BASED SENSOR-LESS WAVEFRONT ABERRATION CORRECTION IN OPTICAL COHERENCE TOMOGRAPHY

Several sensor-less wavefront aberration correction methods that correct nonlinear wavefront aberrations by maximizing the OCT signal are tested on an optical coherence tomography (OCT) setup. A conventional coordinate search method is compared to two model-based optimization methods. The first model-based method takes advantage of the well-known optimization algorithm (NEWUOA) and utilizes a quadratic model. The second model-based method (DONE) is new and utilizes a random multi-dimensional Fourier basis expansion. The model-based algorithms achieve lower wavefront errors with up to ten times fewer measurements. Furthermore, the newly proposed DONE method outperforms the NEWUOA method significantly. The DONE algorithm is tested on OCT images and shows a significantly improved image quality.

Parts of this chapter have been published in [172].

©2014 Optical Society of America. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this chapter are prohibited.

<https://www.osapublishing.org/ol/abstract.cfm?uri=ol-40-24-5722>

3.1. INTRODUCTION

Non-invasive 3D imaging of the retina is one of the main applications of optical coherence tomography (OCT) [2]. The lateral resolution of the OCT system can be improved by increasing the pupil size. In general, this leads to increased optical wavefront aberrations that limit the resolution. Adaptive optics (AO) has been successfully used to correct these optical wavefront aberrations on large pupils (>2 mm), leading to an improved image quality. Ultra-high lateral and axial resolutions up to $3\ \mu\text{m}$ and $2 - 3\ \mu\text{m}$, respectively, have been obtained by using adaptive optics in OCT [96,97]. In general, the optical aberrations in AO-OCT setups are determined by wavefront sensors such as the Shack-Hartmann (SH) wavefront sensor. Imaging systems with SH wavefront sensors suffer from several disadvantages, such as an increased cost and non-common path wavefront errors. These drawbacks can be avoided if wavefront sensor-less adaptive optics (WFSL-AO) methods are used. WFSL methods optimize image quality metrics (e.g. the strength of the OCT signal) in order to minimize the aberrations and create a sharper image.

The ideal WFSL-AO algorithm for OCT is robust with respect to noise and converges to its optimum in a small number of measurements. Here, noise includes shot noise, speckle, and variation of sample structure in the B-scans with different lateral positions. Finite difference approximations for the explicit calculation of derivatives, such as a forward difference in the Newton-Raphson method, require extra measurements to determine the individual partial derivatives and are not robust with respect to noise. Therefore, derivative-free optimization algorithms are preferred. It has been shown that various derivative-free optimization algorithms can successfully improve the quality and signal-to-noise ratio of OCT images [144, 145, 156, 158]. We demonstrated that the coordinate search (CS) algorithm [144] reaches the maximum OCT signal if the aberrations are not too large [156]. Additionally, simulated annealing [145] and the stochastic parallel gradient descent (SPGD) algorithm [173] have been successfully used for WFSL-AO in OCT and scanning laser ophthalmoscopy, respectively. However, the final obtained root mean square (RMS) wavefront error in these algorithms is very susceptible to noise, because past measurements are not exploited. It was shown that the use of models can improve the convergence rate of WFSL algorithms [148, 151]. In this case, the model, which is fit to the measurements, is used to estimate the derivatives. The final result is much less susceptible to noise, because past measurement information is used in the fit of the model. In [151] a quadratic model was fit to a set of prior measurements of an image metric, so that the model could be used to correct the aberrations. However, the aberration correction is limited to the relatively small region where the quadratic model is a good approximation of the metric.

Recently, we developed and validated a transfer function for optical wavefront aberrations in OCT [156]. This model predicts the loss of the OCT signal caused by optical wavefront aberrations. Based on the shape of the transfer functions and simulations on this model we propose two advanced model-based optimization algorithms for WFSL-AO in OCT, the NEWUOA and the DONE algorithm, and compare these two algorithms with the CS algorithm. The NEWUOA algorithm [174], based on an adaptive quadratic model, is chosen because it is one of the most well known and best performing derivative-free algorithms in optimization. Our DONE algorithm, *Data-based Online Nonlinear Extremum-seeker*, is based on a random Fourier basis. It is called an online

method, because the aberrations of each measurement is chosen based on the outcome of previous measurements. We demonstrate experimentally that both algorithms outperform the CS algorithm.

3.2. MATERIALS AND METHODS

The experimental results are obtained using the AO-OCT setup described in [156], except the spectrometer is a Cobra VHR (Wasatch Photonics). The three investigated wavefront correction algorithms maximize a metric based on the magnitude of the OCT signal. The OCT signal is obtained by sampling the spectrometer signal equidistantly in the k domain using interpolation, taking the absolute value of the Fourier transform of the spectrometer signal with the reference arm intensity subtracted. The time and space averaged signal from a selected depth segment of the complete OCT signal is used as the image quality metric. The metric is a real valued function, $f(\alpha)$, of the optical wavefront aberrations represented in the vector α . The aberration coefficients α_i in α are the coefficients for the normalized Zernike polynomials in μm based on our model [156].

The first algorithm we consider is the coordinate search algorithm from [144]. The CS algorithm successively scans in S steps through a predefined domain of each aberration α_i with a step size s . It retains the value of α_i of the scan that maximized the metric before going to the next aberration α_{i+1} . A performance trade-off exists between the number of measurements and the size of the scanning domain. The step size s has to be small enough to ensure a small final error and large enough to have fast convergence to the maximum. The coordinate search optimization method is simple and converges slowly.

The second algorithm is the NEWUOA optimization algorithm [174]. The NEWUOA algorithm is a computationally efficient derivative-free optimization method based on a quadratic model. The m variables of the initial multi-dimensional quadratic model are fit to the first m measurements of the metric based on the OCT signal. When a new measurement is taken to update the quadratic model, an old measurement is thrown away. An optimization routine minimizes the quadratic model of $-f(\alpha)$ within a bounded region, in which the quadratic model is considered accurate, leading to a new estimate of the metric's maximum. This process is iterated until some stopping criterion is fulfilled. The NEWUOA algorithm has three parameters. The first parameter, m , is the number of variables the quadratic model is based on, the default value is $(2d + 1)$. Here, d is the number of aberration coefficients that are optimized. The second parameter is an initial step size ρ_α . This parameter should not be chosen too large such that the fundamental features of the function are not skipped and also not too small such that larger variations in $f(\alpha)$ are probed. The third parameter ρ_Ω determines the final step size, which should be small. It should be smaller than the RMS wavefront error corresponding to $\frac{\lambda}{14}$, the Maréchal criterion for the diffraction limit.

The third algorithm, which we call *Data-based Online Nonlinear Extremum-seeker* (DONE), fits a multi-dimensional random Fourier basis to the measurements [175]. With every new measurement taken, the function *FourierRegression*, shown below in pseudo-code with MATLAB like notation, computes a new model $\hat{f}(\alpha)$ of the metric function $f(\alpha)$. After the model is obtained, a well known optimization routine, `fmincon` (MATLAB R2012b) (see [176, 177] for the theoretical background on `fmincon`), is run

on the model. Since we are looking for a maximum of $f(\alpha)$, we minimize $-f(\alpha)$. The initial vector of the optimization algorithm is α_{init} , which is bounded element-wise by the lower bound lb and upper bound ub . These bounds are chosen such that the aberration, which is to be corrected, is never out of the bounded region. Before the bounding, a random perturbation is added to avoid the algorithm from getting stuck in an insignificant local minimum. The coefficients of the next measurement α_{i+1} are determined by adding a small normally distributed perturbation with standard deviation σ_1 to the last found minimizer α_{min} and enforcing the bounds as before. This second perturbation is added to keep the algorithm from concentrating on a too narrow part of the search space.

3

Algorithm 1 DONE Algorithm

```

1: procedure DONE( $\alpha_0, N, lb, ub$ )
2:    $d = \text{length}(\alpha_0)$ 
3:   for  $i = 0$  to  $N - 1$  do
4:      $f_i = f(\alpha_i) + \text{measurement noise}$ 
5:      $\hat{f}(\alpha) = \text{FourierRegression}([\alpha_0 \dots \alpha_i], [f_0; \dots; f_i], d)$ 
6:      $\alpha_{init} = \max(\min(\alpha_i + \sigma_1 \text{randn}(d, 1), ub), lb)$ 
7:      $\alpha_{min} = \text{fmincon}(-\hat{f}(\alpha), \alpha_{init}, lb, ub)$ 
8:      $\alpha_{i+1} = \max(\min(\alpha_{min} + \sigma_1 \text{randn}(d, 1), ub), lb)$ 
9:   return  $\alpha_{min}$ 

```

Algorithm 2 Fourier Regression

```

1: procedure FOURIERREGRESSION( $A, F, d$ )
2:    $\omega = \sqrt{2\sigma} \text{randn}(D, d)$ 
3:    $b = 2\pi \text{rand}(D, 1)$ 
4:    $Z = \sqrt{\frac{2}{D}} \cos(\omega A + [b \dots b])$ 
5:    $w = (ZZ^T + \lambda \text{length}(F) \text{eye}(D))^{-1} (ZF)$ 
6:    $\hat{f}(\alpha) = \sqrt{\frac{2}{D}} w^T \cos(\omega \alpha + b)$ 
7:   return  $\hat{f}(\alpha)$ 

```

▷ $\hat{f}(\alpha)$ is a function

In the function *FourierRegression*, Line 5 solves the least squares problem $\|Z^T w - F\|_2^2 + \lambda \text{length}(F) \|w\|_2^2$. The constant D specifies the number of random Fourier basis functions. More basis functions will lead to a better representation of the original function, but also to an increased computational load. The parameter λ is a regularization parameter used to avoid over-fitting of the model to the measurements. The parameter σ sets the standard deviation of the frequencies of the cosine bases. It should be chosen such that the higher frequencies of the unknown function $f(\alpha)$ are still captured, however the bandwidth should not be too high such that noise in higher frequencies is also filtered. The MATLAB commands `randn` and `rand` create normally and uniformly distributed matrices, respectively. The command `eye` creates an identity matrix.

Table 3.1: Parameter values for the three wavefront correction algorithms (wavefront aberrations are defined in μm)

CS		NEWUOA		DONE	
S	50	ρ_α	1	D	1000
s	0.01	ρ_Ω	$1\text{e-}8$	σ_1	$0.1\sqrt{\frac{3}{d}}$
		m	$2d+1$	σ	1
				λ	0.001

Table 3.1 shows the proposed parameter settings for the three algorithms. The parameters of the CS and NEWUOA algorithms are chosen such that the diffraction limit can be reached, the maximum value of the added aberrations is covered, and fundamental features of the metric function $f(\alpha)$ are not missed. In order to find suitable values for the parameters of the DONE algorithm, it was simulated using the OCT model [156] for random aberrations between $-0.45 \mu\text{m}$ and $+0.45 \mu\text{m}$.

3.3. EXPERIMENTS

We perform two experiments with three and seven aberrations applied to the DM shown in Fig. 3.1 and Fig. 3.2, respectively. Each trial we add a random combination of these wavefront aberrations with a total maximum RMS wavefront aberration of $0.45 \mu\text{m}$ to the deformable mirror. Subsequently, the metric calculated from one B-scan of a Scotch tape sample is optimized using the three algorithms. The Scotch tape sample is shown in Fig. 3.3. In each iteration only one B-scan with 512 A-scans is taken. The remaining RMS error after the optimization is measured with the Shack-Hartmann wavefront sensor. In the first experiment, shown in Fig. 3.1, a combination of Zernike modes three to five, i.e. defocus and two astigmatisms, is applied to the DM and corrected. In Fig. 3.1(a) the maximum achieved values of the OCT signal with respect to the number of iterations are averaged over 100 trials. In Fig. 3.1(b) the final RMS wavefront errors of the 100 trials are shown in a box plot together with the black line indicating the Maréchal criterion for the diffraction limit based on the OCT center wavelength of 850 nm. On each box, the central red line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually with red markers. The coordinate search algorithm converges slowly and does not have a good final RMS wavefront error after it was stopped at 1050 measurements. However, it manages to get a strong OCT signal. Before elaborating on this observation, we continue with the description of the results. The NEWUOA algorithm converges faster than CS. It obtains smaller final RMS wavefront errors than CS, even though the OCT signals are weaker. The DONE algorithm outperforms both other algorithms in final RMS error with ten times less measurements than the CS algorithm. Fig. 3.2 shows a similar experiment, however in this case a combination of Zernike modes three to nine are applied and corrected by the DM. The results for three and seven aberrations have similar characteristics. However, the performance of

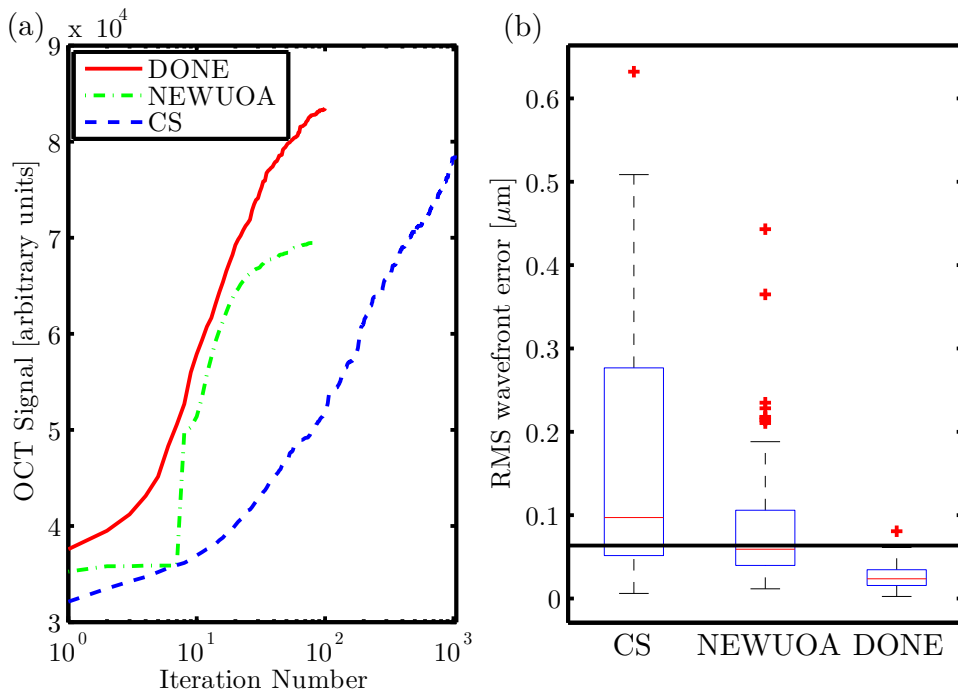


Figure 3.1: AO-OCT wavefront correction for Zernike modes three to five. (a) OCT signal averaged over 100 trials versus iteration number. (b) Box plots of the 100 final RMS wavefront errors measured with the SH wavefront sensor. The black line indicates the Maréchal criterion for the diffraction limit.

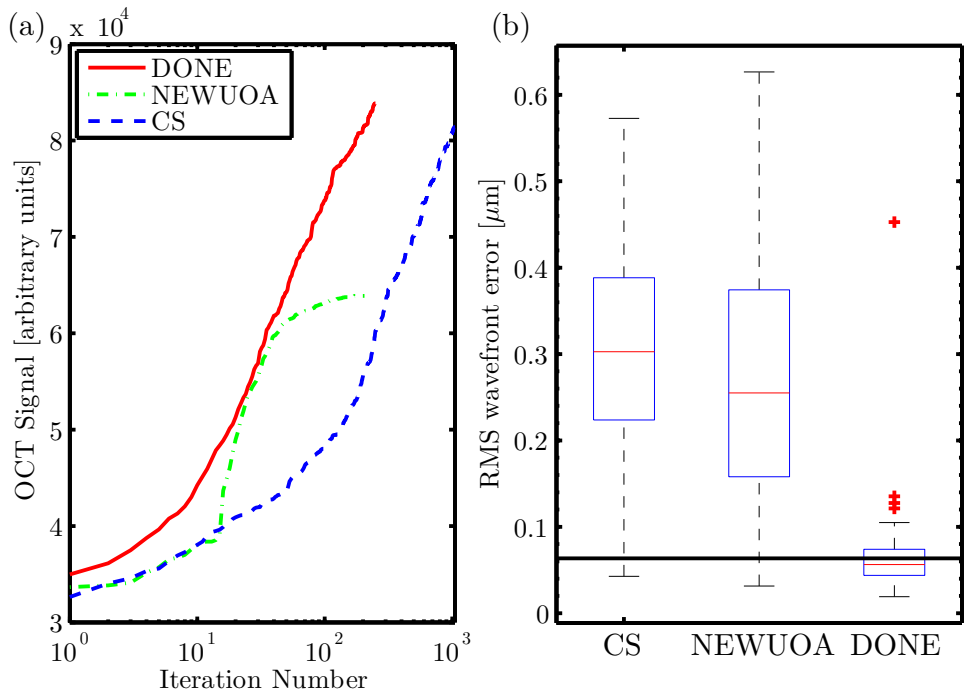


Figure 3.2: Similar to Fig. 3.1 for wavefront aberrations consisting of Zernike modes three to nine.

the DONE algorithm in contrast with the other algorithms in terms of the final RMS error for seven aberrations has become even better.

Since the NEWUOA algorithm has its own stopping criterion, the longest runs of NEWUOA took 87 and 205 measurements for three and seven aberrations, respectively. The DONE algorithm was set to stop after $N = 100$ and $N = 250$ measurements for three and seven aberrations, respectively. For seven aberrations the CS and the NEWUOA algorithm often don't reach the diffraction limit, which is clear from the box plots of the final RMS wavefront errors. The DONE algorithm clearly outperforms both the CS and the NEWUOA algorithm in terms of final RMS error. In three dimensions DONE converges 99% of the starting aberrations converges below the Maréchal criterion, compared with 65% for seven dimensions.

3

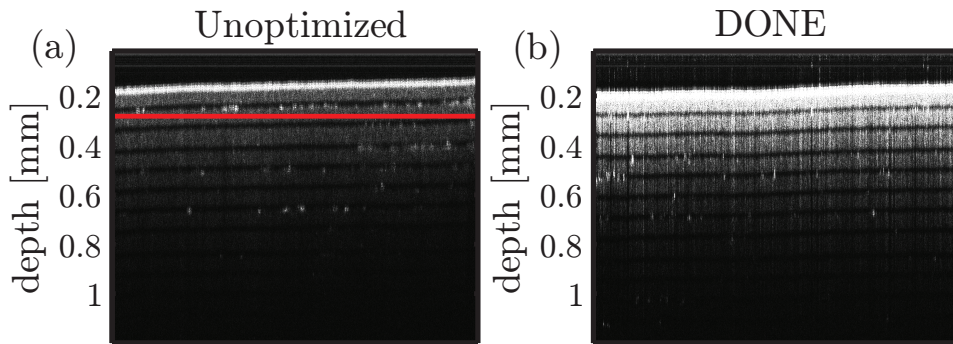


Figure 3.3: OCT B-scans of Scotch tape before aberration correction (a) and after 100 iterations of the DONE algorithm (b). The red line indicates the depth of the OCT signal optimization.

To demonstrate the feasibility of our approach to actual OCT imaging, we image the Scotch tape sample with a defocus of $-0.54 \mu\text{m}$, a vertical astigmatism of $0.02 \mu\text{m}$, and an oblique astigmatism of $0.30 \mu\text{m}$ in Fig. 3.3(a). Figure 3.3(b) shows an OCT image where the RMS wavefront error measured by the SH wavefront sensor has been reduced to $0.03 \mu\text{m}$ after 100 iterations of the DONE algorithm. Using the same linear intensity scale, the corrected OCT image has a much stronger OCT signal. The outcomes of the other algorithms suffered from a weaker OCT signal but had a similar structure. Thus, they are omitted.

In Fig. 3.4(a) we show a B-scan of a lemon slice with an added defocus of $-0.49 \mu\text{m}$, a vertical astigmatism of $0.18 \mu\text{m}$, and an oblique astigmatism of $0.32 \mu\text{m}$. Figure 3.4(b), Fig. 3.4(c), and Fig. 3.4(d) show the lemon slice image after 100 iterations of the CS, NEWUOA, and DONE algorithm, respectively. A zoomed area for each image is added to demonstrate the clear difference in the visible structural features, besides the strong difference in OCT signal strength. For the large area images the same linear intensity scale is used, however, the zoomed areas all have an enhanced contrast scaling to show the features more clearly. The remaining RMS wavefront errors read from the coefficients applied to the DM for Fig. 3.4(b), Fig. 3.4(c), and Fig. 3.4(d) are $0.36 \mu\text{m}$, $0.23 \mu\text{m}$, and $0.06 \mu\text{m}$, respectively.

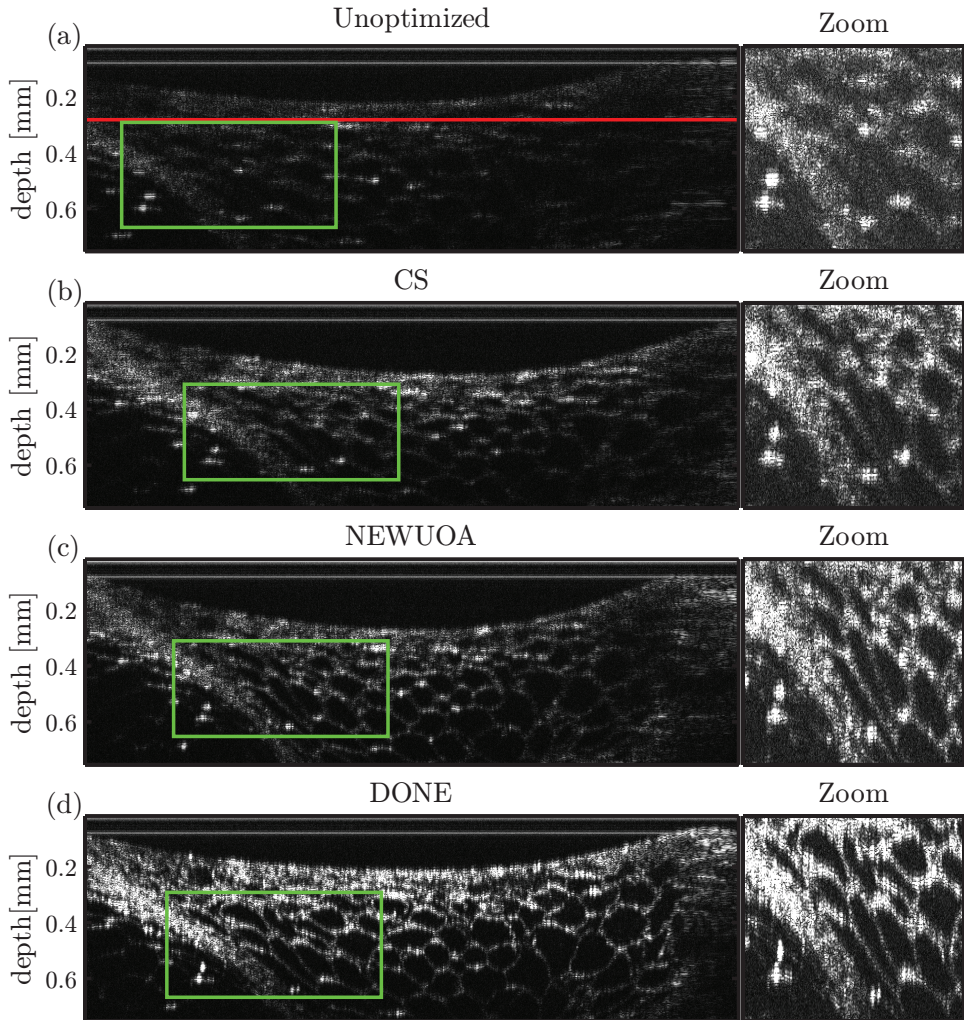


Figure 3.4: OCT B-scans of a lemon slice before aberration correction (a) and after 100 iterations of the CS algorithm (b), of the NEWUOA algorithm (c), and of the DONE algorithm (d). The red line indicates the depth of the OCT signal optimization, the green rectangle indicates the zoomed area. An enhanced contrast scaling is applied to the zoomed areas.

3.4. DISCUSSION

In our experiments, we observed that even though the CS algorithm obtained a similarly strong OCT signal strength in the end, the resulting RMS error is much lower for the DONE algorithm. We attribute this phenomenon to the noise in the OCT signal measurements. The CS algorithm does not take the noise into account and can therefore yield OCT signals that are strong only due to noise. In contrast, the DONE algorithm builds its model based on all known measurements, making it less susceptible to noise as the number of measurements increases. The NEWUOA algorithm can keep track of a fixed number of measurements, but this number is inherently limited by the complexity of the quadratic model it uses. It is therefore less susceptible to noise as well, but, as can be seen in the experimental results, the restriction to a fixed number of measurements results in a considerable increase in the final RMS wavefront error. The superior performance of the DONE algorithm, however, comes at the price of an increased computational complexity. While the computational complexity for CS and NEWUOA is independent on the number of measurements, the computational complexity of a naive implementation of DONE increases with every iteration. The CS and NEWUOA algorithms have a computational time below 1 ms. However, the computational time of the DONE algorithm is in the order of 60 ms during the experiments. This computational time is equal to the acquisition and processing time of a B-scan in our OCT system, therefore it is not a limiting factor. We expect that the computational complexity can be reduced below 1 ms by using a recursive least squares algorithm in the FourierRegression function, optimized compiled code, and parallel processing on a GPU.

3.5. CONCLUSION

In short, the performance of three methods for WFSL-AO in OCT has been investigated experimentally. The DONE algorithm outperforms both NEWUOA and CS in terms of final RMS error and it converges up to ten times faster than the CS algorithm. We implemented the DONE algorithm in OCT imaging and showed a significantly improved image quality. We showed that for large aberrations the DONE algorithm succeeds in correcting a noisy signal with high accuracy.

4

ONLINE OPTIMIZATION WITH COSTLY AND NOISY MEASUREMENTS USING RANDOM FOURIER EXPANSIONS

This paper analyzes DONE, an online optimization algorithm that iteratively minimizes an unknown function based on costly and noisy measurements. The algorithm maintains a surrogate of the unknown function in the form of a random Fourier expansion (RFE). The surrogate is updated whenever a new measurement is available, and then used to determine the next measurement point. The algorithm is comparable to Bayesian optimization algorithms, but its computational complexity per iteration does not depend on the number of measurements. We derive several theoretical results that provide insight on how the hyper-parameters of the algorithm should be chosen. The algorithm is compared to a Bayesian optimization algorithm for an analytic benchmark problem and three applications, namely, optical coherence tomography, optical beam-forming network tuning, and robot arm control. It is found that the DONE algorithm is significantly faster than Bayesian optimization in the discussed problems, while achieving a similar or better performance.

Parts of this chapter have been published in [178].

©2016 IEEE. One print or electronic copy may be made for personal use only. Systematic reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modifications of the content of this chapter are prohibited.

4.1. INTRODUCTION

Many optimization algorithms use the derivative of an objective function, but often this information is not available in practice. Regularly, a closed form expression for the objective function is not available and function evaluations are costly. Examples are objective functions that rely on the outcome of a simulation or an experiment. Approximating derivatives with finite differences is costly in high-dimensional problems, especially if the objective function is costly to evaluate. More efficient algorithms for derivative-free optimization (DFO) problems exist. Typically, in DFO algorithms a model is used that can be optimized without making use of the derivative of the underlying function [114, 179]. Some examples of commonly used DFO algorithms are the simplex method [116], NEWUOA [118], BOBYQA [119], and DIRECT [180]. Additionally, measurements of a practical problem are usually corrupted by noise. Several techniques have been developed to cope with a higher noise level and make better use of the expensive objective functions evaluations. Filtering and pattern search optimization algorithms such as implicit filtering [181] and SID-PSM [182] can handle local minima resulting from high frequency components. Bayesian optimization, also known as sequential Kriging optimization, deals with heteroscedastic noise and perturbations very well. One of the first and best known Bayesian optimization algorithms is EGO [135]. Bayesian optimization relies on a surrogate model that represents a probability distribution of the unknown function under noise, for example Gaussian processes or Student's-t processes [139, 183–185]. In these processes different kernels and kernel learning methods are used for the covariance function [186, 187]. The surrogate model is used to decide where the next measurement should be taken. New measurements are used to update the surrogate model. Bayesian optimization has been successfully used in various applications, including active user modeling and reinforcement learning [136], robotics [138], hyper-parameter tuning [184], and optics [188].

Recently, the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm was proposed in [172]. It is similar to Bayesian optimization, but simpler and faster. The DONE algorithm uses random Fourier expansions [175] (RFEs) as a surrogate model. The nature of the DONE algorithm makes the understanding of the hyper-parameters easier. In RFE models certain parameters are chosen randomly. In this paper, we derive a close-to-optimal probability distribution for some of these parameters. We also derive an upper bound for the regularization parameter used in the training of the RFE model.

The advantages of the DONE algorithm are illustrated in an analytic benchmark problem and three applications. We numerically compare DONE to BayesOpt [139], a Bayesian optimization library that was shown to outperform many other similar libraries in [139]. The first application is optical coherence tomography (OCT), a 3D imaging method based on interference often used to image the human retina [144, 145, 172]. The second application we consider is the tuning of an optical beam-forming network (OBFN). OBFNs are used in wireless communication systems to steer phased array antennas in the desired direction by making use of positive interference of synchronized signals [189–194]. The third application is a robot arm of which the tip has to be directed to a desired position [195].

This paper is organized as follows. Section 4.2 gives a short overview and provides new theoretical insights on random Fourier expansions, the surrogate model on which

the DONE algorithm is based. We have noticed a gap in the literature, where approximation guarantees are given for ideal, but unknown RFE weights, while in practice RFE weights are computed via linear least squares. We investigate several properties of the ideal weights and combine these results with existing knowledge of RFEs to obtain approximation guarantees for least-square weights. Section 4.3 explains the DONE algorithm. Theoretically optimal as well as more practical ways to choose the hyperparameters of this algorithm are given in Section 4.4. In Section 4.5 the DONE algorithm and BayesOpt are compared for a benchmark problem and for the three aforementioned applications. We conclude the paper in Section 4.6.

4.2. RANDOM FOURIER EXPANSIONS

In this section, we will describe the surrogate model that we will use for optimization. There is a plethora of black-box modeling techniques to approximate a function from measurements available in the literature, with neural networks, kernel methods, and of course classic linear models probably being the most popular [137, 196, 197]. In this paper, we use random Fourier expansions (RFEs) [175] to model the unknown function because they offer a unique mix of computational efficiency, theoretical guarantees and ease of use that make them ideal for online processing. While general neural networks are more expressive than random Fourier features, they are difficult to use and come without theoretical guarantees. Standard kernel methods suffer from high computational complexity because the number of kernels equals the number of measurements. RFEs have been originally introduced to reduce the computational burden that comes with kernel methods, as will be explained next [175, 198, 199].

Assume that we are provided N scalar measurements y_i taken at measurement points $\mathbf{x}_i \in \mathbb{R}^d$ as well as a kernel $k(\mathbf{x}_i, \mathbf{x}_j)$ that, in a certain sense, measures the closeness of two measurement points. To train the kernel expansion

$$g_{KM}(\mathbf{x}) = \sum_{i=1}^N a_i k(\mathbf{x}, \mathbf{x}_i), \quad (4.1)$$

a linear system involving the kernel matrix $[k(\mathbf{x}_i, \mathbf{x}_j)]_{i,j}$ has to be solved for the coefficients a_i . The computational costs of training and evaluating (4.1) grow cubically and linearly in the number of datapoints N , respectively. This can be prohibitive for large values of N . We now explain how RFEs can be used to reduce the complexity [175]. Assuming the kernel k is shift-invariant and has Fourier transform p , it can be normalized such that p is a probability distribution [175]. That is, we have

$$k(\mathbf{x}_i - \mathbf{x}_j) = \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) e^{-i\boldsymbol{\omega}^T (\mathbf{x}_i - \mathbf{x}_j)} d\boldsymbol{\omega}. \quad (4.2)$$

We will use several trigonometric properties and the fact that k is real to continue the derivation. This gives

$$\begin{aligned} k(\mathbf{x}_i - \mathbf{x}_j) &= \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) \cos(\boldsymbol{\omega}^T (\mathbf{x}_i - \mathbf{x}_j)) d\boldsymbol{\omega} \\ &= \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) \cos(\boldsymbol{\omega}^T (\mathbf{x}_i - \mathbf{x}_j)) + p(\boldsymbol{\omega}) \int_0^{2\pi} \cos(\boldsymbol{\omega}^T (\mathbf{x}_i + \mathbf{x}_j) + 2b) db d\boldsymbol{\omega} \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{2\pi} \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) \int_0^{2\pi} \cos(\boldsymbol{\omega}^T(\mathbf{x}_i - \mathbf{x}_j)) \\
 &\quad + \cos(\boldsymbol{\omega}^T(\mathbf{x}_i + \mathbf{x}_j) + 2b) db d\boldsymbol{\omega} \\
 &= \frac{1}{2\pi} \int_{\mathbb{R}^d} p(\boldsymbol{\omega}) \int_0^{2\pi} 2 \cos(\boldsymbol{\omega}^T \mathbf{x}_i + b) \\
 &\quad \cdot \cos(\boldsymbol{\omega}^T \mathbf{x}_j + b) db d\boldsymbol{\omega} \\
 &= \mathbb{E}[2 \cos(\boldsymbol{\Omega}^T \mathbf{x}_i + B) \cos(\boldsymbol{\Omega}^T \mathbf{x}_j + B)] \\
 &\approx \frac{2}{D} \sum_{k=1}^D \cos(\boldsymbol{\omega}_k^T \mathbf{x}_i + b_k) \cos(\boldsymbol{\omega}_k^T \mathbf{x}_j + b_k), \tag{4.3}
 \end{aligned}$$

4

if $\boldsymbol{\omega}_k$ are independent samples of the random variable $\boldsymbol{\Omega}$ with probability distribution function (p.d.f.) p , and $b_k \in [0, 2\pi]$ are independent samples of the random variable B with a uniform distribution. For $c_k = \sum_{i=1}^N \frac{2}{D} a_i \cos(\boldsymbol{\omega}_k^T \mathbf{x}_i + b_k)$ we thus have:

$$g_{KM}(\mathbf{x}) \approx \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k). \tag{4.4}$$

Note that the number of coefficients D is now independent of the number of measurements N . This is especially advantageous in online applications where the number of measurements N keeps increasing. We use the following definition of a random Fourier expansion.

Definition 1. A *Random Fourier Expansion (RFE)* is a function of the form $g: \mathbb{R}^d \rightarrow \mathbb{R}$,

$$g(\mathbf{x}) = \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k), \tag{4.5}$$

with $D \in \mathbb{N}$, the b_k being realizations of independent and identically distributed (i.i.d.) uniformly distributed random variables B_k on $[0, 2\pi]$, and with the $\boldsymbol{\omega}_k \in \mathbb{R}^d$ being realizations of i.i.d. random vectors $\boldsymbol{\Omega}_k$ with an arbitrary continuous p.d.f. $p_{\boldsymbol{\Omega}}$. The B_k and the $\boldsymbol{\Omega}_k$ are assumed to be mutually independent.

We finally remark that there are other approaches to reduce the complexity of kernel methods and make them suitable for online processing, which are mainly based on sparsity [200–203]. However, these are much more difficult to tune than using RFEs [199]. It is also possible to use other basis functions instead of the cosine, but the cosine was among the top performers in an exhaustive comparison with similar models [204]. Moreover, the parameters of the cosines have intuitive interpretations in terms of the Fourier transform.

4.2.1. IDEAL RFE WEIGHTS

In this section, we deal with the problem of fitting a RFE to a given function f . We derive ideal but in practice unknown weights c . We start with the case of infinitely many samples and basis functions (see also [205, 206]), which corresponds to turning the corresponding sums into integrals.

Theorem 1. Let $f \in L^2(\mathbb{R}^d)$ be a real-valued function and let

$$\bar{c}(\boldsymbol{\omega}, b) = \begin{cases} \frac{1}{\pi} |\hat{f}(\boldsymbol{\omega})| \cos(\angle \hat{f}(\boldsymbol{\omega}) - b), & b \in [0, 2\pi], \\ 0, & \text{otherwise.} \end{cases} \quad (4.6)$$

Then, for all $\mathbf{x} \in \mathbb{R}^d$,

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega}. \quad (4.7)$$

Here, $|\hat{f}|$ and $\angle \hat{f}$ denote the magnitude and phase of the Fourier transform $\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i\boldsymbol{\omega}^T \mathbf{x}} d\mathbf{x}$. The sets L^2 and L^∞ denote the space of square integrable functions and the space of all essentially bounded functions, respectively.

Proof. For $b \in [0, 2\pi]$, we have

$$\begin{aligned} \bar{c}(\boldsymbol{\omega}, b) &= \frac{1}{\pi} |\hat{f}(\boldsymbol{\omega})| \cos(\angle \hat{f}(\boldsymbol{\omega}) - b) \\ &= \frac{1}{\pi} \operatorname{Re} \left\{ \hat{f}(\boldsymbol{\omega}) e^{-ib} \right\}. \end{aligned} \quad (4.8)$$

Using that $f(\mathbf{x})$ is real, we find that

$$\begin{aligned} f(\mathbf{x}) &= \operatorname{Re} \left\{ \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \mathbf{x}} d\boldsymbol{\omega} \right\} \\ &= \operatorname{Re} \left\{ \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \left(\hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \mathbf{x}} \frac{1}{2\pi} \int_0^{2\pi} 1 db + \right. \right. \\ &\quad \left. \left. \hat{f}(\boldsymbol{\omega}) e^{-i\boldsymbol{\omega}^T \mathbf{x}} \underbrace{\int_0^{2\pi} e^{-2ib} db}_{=0} \right) d\boldsymbol{\omega} \right\} \\ &= \operatorname{Re} \left\{ \frac{1}{\pi} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \hat{f}(\boldsymbol{\omega}) e^{-ib} \right. \\ &\quad \left. \frac{1}{2} \left[e^{i(\boldsymbol{\omega}^T \mathbf{x} + b)} + e^{-i(\boldsymbol{\omega}^T \mathbf{x} + b)} \right] db d\boldsymbol{\omega} \right\} \\ &= \operatorname{Re} \left\{ \frac{1}{\pi} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \hat{f}(\boldsymbol{\omega}) e^{-ib} \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega} \right\} \\ &\stackrel{(4.8)}{=} \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega}. \end{aligned} \quad (4.9)$$

□

For $b \in [0, 2\pi]$, we have another useful expression for the ideal weights that is used later on in this section, namely

$$\bar{c}(\boldsymbol{\omega}, b) = \frac{1}{\pi} \operatorname{Re} \left\{ \hat{f}(\boldsymbol{\omega}) e^{-ib} \right\}$$

$$\begin{aligned}
 &= \frac{1}{\pi} \operatorname{Re} \left\{ \int_{\mathbb{R}^d} f(\mathbf{x}) e^{-i(\boldsymbol{\omega}^T \mathbf{x} + b)} d\mathbf{x} \right\} \\
 &= \frac{1}{\pi} \int_{\mathbb{R}^d} f(\mathbf{x}) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) d\mathbf{x}. \tag{4.10}
 \end{aligned}$$

The function \bar{c} in Theorem 1 is not unique. However, of all functions c that satisfy (4.7), the given \bar{c} is the one with minimum norm.

Theorem 2. *Let \bar{c} be as in Theorem 1. If $\bar{c} : \mathbb{R}^d \times [0, 2\pi] \rightarrow \mathbb{R}$ satisfies*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega} \quad \text{a.e.} \tag{4.11}$$

then $\|\bar{c}\|_{L^2}^2 \geq \|\bar{c}\|_{L^2}^2 = \frac{(2\pi)^d}{\pi} \|f\|_{L^2}^2$, with equality if and only if $\bar{c} = \bar{c}$ in the L^2 sense.

Proof. First, using Parseval's theorem and $\int_0^{2\pi} \cos(a - b)^2 db = \pi$ for any real constant a , note that

$$\begin{aligned}
 \|\bar{c}\|_{L^2}^2 &= \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b)^2 db d\boldsymbol{\omega} \\
 &\stackrel{(4.6)}{=} \int_{\mathbb{R}^d} \int_0^{2\pi} \frac{1}{\pi^2} |\hat{f}(\boldsymbol{\omega})|^2 \cos(\angle \hat{f}(\boldsymbol{\omega}) - b)^2 db d\boldsymbol{\omega} \\
 &= \int_{\mathbb{R}^d} \frac{1}{\pi^2} |\hat{f}(\boldsymbol{\omega})|^2 \int_0^{2\pi} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b)^2 db d\boldsymbol{\omega} \\
 &= \int_{\mathbb{R}^d} \frac{1}{\pi} |\hat{f}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\
 &= \frac{(2\pi)^d}{\pi} \int_{\mathbb{R}^d} f(\mathbf{x})^2 d\mathbf{x} = \frac{(2\pi)^d}{\pi} \|f\|_{L^2}^2. \tag{4.12}
 \end{aligned}$$

Assume that $\bar{c}(\boldsymbol{\omega}, b) = \bar{c}(\boldsymbol{\omega}, b) + q(\boldsymbol{\omega}, b)$. Then we get

$$\begin{aligned}
 &\int_{\mathbb{R}^d} f(\mathbf{x})^2 d\mathbf{x} \\
 &\stackrel{(4.11)}{=} \int_{\mathbb{R}^d} f(\mathbf{x}) \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega} d\mathbf{x} \\
 &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \int_{\mathbb{R}^d} f(\mathbf{x}) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) d\mathbf{x} db d\boldsymbol{\omega} \\
 &\stackrel{(4.10)}{=} \frac{\pi}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \bar{c}(\boldsymbol{\omega}, b) db d\boldsymbol{\omega} \\
 &= \frac{\pi}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b)^2 + \bar{c}(\boldsymbol{\omega}, b) q(\boldsymbol{\omega}, b) db d\boldsymbol{\omega} \\
 &\stackrel{(4.12)}{=} \int_{\mathbb{R}^d} f(\mathbf{x})^2 d\mathbf{x} + \frac{\pi}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) q(\boldsymbol{\omega}, b) db d\boldsymbol{\omega}. \tag{4.13}
 \end{aligned}$$

Following the above equality we can conclude that $\int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) q(\boldsymbol{\omega}, b) db d\boldsymbol{\omega} = 0$. The following now holds:

$$\|\bar{c}\|_{L^2}^2 = \|\bar{c} + q\|_{L^2}^2$$

$$\begin{aligned}
&= \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b)^2 + 2\bar{c}(\boldsymbol{\omega}, b)q(\boldsymbol{\omega}, b) + q(\boldsymbol{\omega}, b)^2 db d\boldsymbol{\omega} \\
&= \|\bar{c}\|_{L^2}^2 + \|q\|_{L^2}^2 \geq \|\bar{c}\|_{L^2}^2.
\end{aligned} \tag{4.14}$$

Furthermore, equality holds if and only if $\|q\|_{L^2} = 0$. That is, the minimum norm solution is unique in L^2 . \square

These results will be used to derive ideal weights for a RFE with a finite number of basis functions as in Definition 1 by sampling the weights in (4.6). We prove unbiasedness in the following theorem, while variance properties are analyzed in Appendix 4.8.

Theorem 3. *For any continuous p.d.f. $p_{\boldsymbol{\Omega}}$ with $p_{\boldsymbol{\Omega}}(\boldsymbol{\omega}) > 0$ if $|\hat{f}(\boldsymbol{\omega})| > 0$, the choice*

$$C_k = \frac{2}{D(2\pi)^d} \frac{|\hat{f}(\boldsymbol{\Omega}_k)|}{p_{\boldsymbol{\Omega}}(\boldsymbol{\Omega}_k)} \cos(\angle \hat{f}(\boldsymbol{\Omega}_k) - B_k) \tag{4.15}$$

makes the (stochastic) RFE $G(\mathbf{x}) = \sum_{k=1}^D C_k \cos(\boldsymbol{\Omega}_k^T \mathbf{x} + B_k)$ an unbiased estimator, i.e., $f(\mathbf{x}) = \mathbb{E}[G(\mathbf{x})]$ for any $\mathbf{x} \in \mathbb{R}^d$.

Proof. Using Theorem 1, we have

$$\begin{aligned}
f(\mathbf{x}) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \int_0^{2\pi} \bar{c}(\boldsymbol{\omega}, b) \cos(\boldsymbol{\omega}^T \mathbf{x} + b) db d\boldsymbol{\omega} \\
&= \mathbb{E}_{\boldsymbol{\Omega}_1, B_1} \left[\frac{1}{(2\pi)^d p_B(B_1) p_{\boldsymbol{\Omega}}(\boldsymbol{\Omega}_1)} \bar{c}(\boldsymbol{\Omega}_1, B_1) \cos(\boldsymbol{\Omega}_1^T \mathbf{x} + B_1) \right] \\
&= \mathbb{E}_{\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_D, B_1, \dots, B_D} \left[\sum_{k=1}^D \frac{2\pi \bar{c}(\boldsymbol{\Omega}_k, B_k)}{D(2\pi)^d p_{\boldsymbol{\Omega}}(\boldsymbol{\Omega}_k)} \cos(\boldsymbol{\Omega}_k^T \mathbf{x} + B_k) \right] \\
&\stackrel{(4.6)}{=} \mathbb{E} \left[\sum_{k=1}^D \frac{2}{D(2\pi)^d} \frac{|\hat{f}(\boldsymbol{\Omega}_k)|}{p_{\boldsymbol{\Omega}}(\boldsymbol{\Omega}_k)} \cos(\angle \hat{f}(\boldsymbol{\Omega}_k) - B_k) \right. \\
&\quad \left. \cos(\boldsymbol{\Omega}_k^T \mathbf{x} + B_k) \right] = \mathbb{E}[G(\mathbf{x})].
\end{aligned} \tag{4.16}$$

\square

These ideal weights enjoy many other nice properties such as infinity norm convergence [207]. In practice, however, a least squares approach is used for a finite D . This is investigated in the next subsection.

4.2.2. CONVERGENCE OF THE LEAST SQUARES SOLUTION

The ideal weights \bar{c} depend on the Fourier transform of the unknown function f that we wish to approximate. Of course, this knowledge is not available in practice. We therefore assume a finite number of measurement points $\mathbf{x}_1, \dots, \mathbf{x}_N$ that have been drawn independently from a p.d.f. $p_{\mathbf{X}}$ that is defined on a compact set $\mathcal{X} \subseteq \mathbb{R}^d$, and corresponding measurements y_1, \dots, y_N , with $y_n = f(\mathbf{x}_n) + \eta_n$, where η_1, \dots, η_N have been drawn independently from a zero-mean normal distribution with finite variance σ_H^2 . The input and

noise terms are assumed independent of each other. We determine the weights c_k by minimizing the squared error

$$\begin{aligned} J_N(\mathbf{c}) &= \sum_{n=1}^N \left(y_n - \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x}_n + b_k) \right)^2 + \lambda \sum_{k=1}^D c_k^2 \\ &= \|\mathbf{y}_N - \mathbf{A}_N \mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_2^2. \end{aligned} \quad (4.17)$$

Here,

$$\begin{aligned} \mathbf{y}_N &= [y_1 \cdots y_N]^T, \\ \mathbf{A}_N &= \begin{bmatrix} \cos(\boldsymbol{\omega}_1^T \mathbf{x}_1 + b_1) & \cdots & \cos(\boldsymbol{\omega}_D^T \mathbf{x}_1 + b_D) \\ \vdots & \ddots & \vdots \\ \cos(\boldsymbol{\omega}_1^T \mathbf{x}_N + b_1) & \cdots & \cos(\boldsymbol{\omega}_D^T \mathbf{x}_N + b_D) \end{bmatrix}, \end{aligned} \quad (4.18)$$

and λ is a regularization parameter added to deal with noise, over-fitting and ill-conditioning.

Since the parameters $\boldsymbol{\omega}_k, b_k$ are drawn from continuous probability distributions, only the weights c_k need to be determined, making the problem a linear least squares problem. The unique minimizer of J_N is

$$\mathbf{c}_N = (\mathbf{A}_N^T \mathbf{A}_N + \lambda \mathbf{I}_{D \times D})^{-1} \mathbf{A}_N^T \mathbf{y}_N. \quad (4.19)$$

The following theorem shows that RFEs whose coefficient vector have been obtained through a least squares fit as in (4.19) can approximate the function f arbitrarily well. Similar results were given in [205–208], but we emphasize that these convergence results did concern RFEs employing the ideal coefficient vector given earlier in Theorem 3 that is unknown in practice. Our theorem, in contrast, concerns the practically relevant case where the coefficient vector has been obtained through a least-squares fit to the data.

Theorem 4. *The difference between the function f and the RFE trained with linear least squares can become arbitrarily small if enough measurements and basis functions are used. More precisely, suppose that $f \in L^2 \cap L^\infty$ and that $\sup_{\boldsymbol{\omega} \in \mathbb{R}^D, b \in [0, 2\pi]} \left| \frac{\bar{c}(\boldsymbol{\omega}, b)}{p_{\boldsymbol{\Omega}}(\boldsymbol{\omega}) p_B(b)} \right| < \infty$. Then, for every $\epsilon > 0$ and $\delta > 0$, there exist constants N_0 and D_0 such that*

$$\int_{\mathcal{X}} \left(f(\mathbf{x}) - \sum_{k=1}^D C_{Nk} \cos(\boldsymbol{\Omega}_k^T \mathbf{x} + B_k) \right)^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \epsilon \quad (4.20)$$

for all $N \geq N_0, D \geq D_0, 0 < \lambda \leq N\Lambda$ with probability at least $1 - \delta$. Here, C_{Nk} is the k -th element of the random vector corresponding to the weight vector given in (4.19), and $\Lambda \geq 0$ is the solution to

$$\begin{aligned} &\left\| (\mathbf{A}_N^T \mathbf{A}_N + N\Lambda \mathbf{I}_{D \times D})^{-1} \mathbf{A}_N^T \mathbf{y}_N \right\|_2^2 = \\ &\sum_{k=1}^D \left(\frac{\bar{c}(\boldsymbol{\omega}_k, b_k)}{(2\pi)^d D p_{\boldsymbol{\Omega}}(\boldsymbol{\omega}_k) p_B(b_k)} \right)^2. \end{aligned} \quad (4.21)$$

The proof of this theorem is given in Appendix 4.7. In Section 4.4.2 we show how to obtain Λ in practice.

4.3. ONLINE OPTIMIZATION ALGORITHM

In this section, we will investigate the DONE algorithm, which locates a minimum of an unknown function f based on noisy evaluations of this function. Each evaluation, or *measurement*, is used to update a RFE model of the unknown function, based on which the next measurement point is determined. Updating this model has a constant computation time of order $O(D^2)$ per iteration, with D being the number of basis functions. We emphasize that this is in stark contrast to Bayesian optimization algorithms, where the computational cost of adding a new measurement increases with the total number of measurements so far. We also remark that the DONE algorithm operates *online* because the model is updated after each measurement. The advantage over offline methods, in which first all measurements are taken and only then processed, is that the number of required measurements is usually lower as measurement points are chosen adaptively.

4

4.3.1. RECURSIVE LEAST SQUARES APPROACH FOR THE WEIGHTS

In the online scenario, a new measurement y_n taken at the point \mathbf{x}_n becomes available at each iteration $n = 1, 2, \dots$. These are used to update the RFE. Let $\mathbf{a}_n = [\cos(\boldsymbol{\omega}_1^T \mathbf{x}_n + b_1) \cdots \cos(\boldsymbol{\omega}_D^T \mathbf{x}_n + b_D)]$, then we aim to find the vector of RFE weights by minimizing the regularized mean square error

$$J_n(\mathbf{c}) = \sum_{i=1}^n (y_i - \mathbf{a}_i \mathbf{c})^2 + \lambda \|\mathbf{c}\|_2^2. \quad (4.22)$$

Let \mathbf{c}_n be the minimum of J_n ,

$$\mathbf{c}_n = \underset{\mathbf{c}}{\operatorname{argmin}} J_n(\mathbf{c}). \quad (4.23)$$

Assuming we have found \mathbf{c}_n , we would like to use this information to find \mathbf{c}_{n+1} without solving (4.23) again. The recursive least squares algorithm is a computationally efficient method that determines \mathbf{c}_{n+1} from \mathbf{c}_n as follows [209, Sec. 21]:

$$\gamma_n = 1 / (1 + \mathbf{a}_n \mathbf{P}_{n-1} \mathbf{a}_n^T), \quad (4.24)$$

$$\mathbf{g}_n = \gamma_n \mathbf{P}_{n-1} \mathbf{a}_n^T, \quad (4.25)$$

$$\mathbf{c}_n = \mathbf{c}_{n-1} + \mathbf{g}_n (y_n - \mathbf{a}_n \mathbf{c}_{n-1}), \quad (4.26)$$

$$\mathbf{P}_n = \mathbf{P}_{n-1} - \mathbf{g}_n \mathbf{g}_n^T / \gamma_n, \quad (4.27)$$

with initialization $\mathbf{c}_0 = 0$, $\mathbf{P}_0 = \lambda^{-1} \mathbf{I}_{D \times D}$.

We implemented a square-root version of the above algorithm, also known as the inverse QR algorithm [209, Sec. 21], which is known to be especially numerically reliable. Instead of performing the update rules (4.24)-(4.27) explicitly, we find a rotation matrix Θ_n that lower triangularizes the upper triangular matrix in Eq. (5.3) below and generates a post-array with positive diagonal entries:

$$\begin{bmatrix} 1 & \mathbf{a}_n \mathbf{P}_{n-1}^{1/2} \\ \mathbf{0} & \mathbf{P}_{n-1}^{1/2} \end{bmatrix} \Theta_n = \begin{bmatrix} \gamma_n^{-1/2} & \mathbf{0} \\ \mathbf{g}_n \gamma_n^{-1/2} & \mathbf{P}_n^{1/2} \end{bmatrix}. \quad (4.28)$$

The rotation matrix Θ_n can be found by performing a QR decomposition of the transpose of the matrix on the left hand side of (5.3), or by the procedure explained in [209, Sec. 21]. The computational complexity of this update is $O(D^2)$ per iteration.

4.3.2. DONE ALGORITHM

We now explain the different steps of the DONE algorithm. The DONE algorithm is used to iteratively find a minimum of a function $f \in L^2$ on a compact set $\mathcal{X} \subseteq \mathbb{R}^d$ by updating a RFE $g(\mathbf{x}) = \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k)$ at each new measurement, and using this RFE as a surrogate of f for optimization. It is assumed that the function f is unknown and only measurements perturbed by noise can be obtained: $y_n = f(\mathbf{x}_n) + \eta_n$. The algorithm consists of four steps that are repeated for each new measurement: **1)** take a new measurement, **2)** update the RFE, **3)** find a minimum of the RFE, **4)** choose a new measurement point. We now explain each step in more detail.

Initialization

Before running the algorithm, an initial starting point $\mathbf{x}_1 \in \mathcal{X}$ and the number of basis functions D have to be chosen. The parameters $\boldsymbol{\omega}_k$ and b_k of the RFE expansion are drawn from continuous probability distributions as defined in Definition 1. The p.d.f. $p_{\boldsymbol{\Omega}}$ and the regularization parameter λ have to be chosen a priori as well. Practical ways for choosing the hyper-parameters will be discussed later in Sect. 4.4. These hyper-parameters stay fixed over the whole duration of the algorithm. Let $\mathbf{P}_0^{1/2} = \lambda^{-1/2} \mathbf{I}_{D \times D}$, and $n = 1$.

Step 1: New measurement

Unlike in Section 4.2.2, it is assumed that measurements are taken in a recursive fashion. At the start of iteration n , a new measurement $y_n = f(\mathbf{x}_n) + \eta_n$ is taken at the point \mathbf{x}_n .

Step 2: Update the RFE

As explained in Section 4.3.1, we update the RFE model $g(\mathbf{x}) = \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k)$ based on the new measurement from Step 1 by using the inverse QR algorithm given in (4.24)-(4.27). Only the weights c_k are updated. The parameters $\boldsymbol{\omega}_k$ and b_k stay fixed through-out the whole algorithm.

Step 3: Optimization on the RFE

After updating the RFE, an iterative optimization algorithm is used to find a (possibly local) minimum $\hat{\mathbf{x}}_n$ of the RFE. All derivatives of the RFE can easily be calculated. Using an analytic expression of the Jacobian will increase the performance of the optimization method used in this step, while not requiring extra measurements of f as in the finite difference method. For functions that are costly to evaluate, this is a big advantage. The method used in the proposed algorithm is an L-BFGS method [104,210]. Other optimization methods can also be used. The initial guess for the optimization is the projection of the current measurement point plus a random perturbation:

$$\mathbf{x}_{init} = P_{\mathcal{X}}(\mathbf{x}_n + \zeta_n), \quad (4.29)$$

where $P_{\mathcal{X}}$ is the projection onto \mathcal{X} . The random perturbation prevents the optimization algorithm from starting exactly in the point where the model was trained. Increasing its value will increase the exploration capabilities of the DONE algorithm but might slow down convergence. In the proposed algorithm, ζ_n is chosen to be white Gaussian noise.

Step 4: Choose a new measurement point

The minimum found in the previous step is used to update the RFE again. A perturbation is added to the current minimum to avoid the algorithm getting trapped unnecessarily in insignificant local minima or saddle points [211]:

$$\mathbf{x}_{n+1} = P_{\mathcal{X}}(\hat{\mathbf{x}}_n + \xi_n). \quad (4.30)$$

The random perturbations can be seen as an exploration strategy and are again chosen to be white Gaussian noise. Increasing their variance σ_{ξ} increases the exploration capabilities of the DONE algorithm but might slow down convergence. In practice, we typically use the same distribution for ξ and ζ . Finally, the algorithm increases n and returns to Step 1.

The full algorithm is shown below in Algorithm 3 for the case $\mathcal{X} = [lb, ub]^d$.

4

Algorithm 3 DONE Algorithm

```

1: procedure DONE( $f, \mathbf{x}_1, N, lb, ub, D, \lambda, \sigma_{\zeta}, \sigma_{\xi}$ )
2:   Draw  $\omega_1 \dots \omega_D$  from  $p_{\Omega}$  independently.
3:   Draw  $b_1 \dots b_D$  from Uniform( $0, 2\pi$ ) independently.
4:    $\mathbf{P}_0^{1/2} = \lambda^{-1/2} \mathbf{I}_{D \times D}$ 
5:    $\mathbf{c}_0 = [0 \dots 0]^T$ 
6:    $\hat{\mathbf{x}}_0 = \mathbf{x}_1$ 
7:   for  $n = 1, 2, 3, \dots, N$  do
8:      $\mathbf{a}_n = [\cos(\omega_1^T \mathbf{x}_n + b_1) \dots \cos(\omega_D^T \mathbf{x}_n + b_D)]$ 
9:      $y_n = f(\mathbf{x}_n) + \eta_n$ 
10:     $g(\mathbf{x}) = \text{updateRFE}(\mathbf{c}_{n-1}, \mathbf{P}_{n-1}^{1/2}, \mathbf{a}_n, y_n)$ 
11:    Draw  $\zeta_n$  from  $\mathcal{N}(0, \sigma_{\zeta}^2 \mathbf{I}_{d \times d})$ .
12:     $\mathbf{x}_{init} = \max(\min(\mathbf{x}_n + \zeta_n, ub), lb)$ 
13:     $[\hat{\mathbf{x}}_n, \hat{g}_n] = \text{L-BFGS}(g(\mathbf{x}), \mathbf{x}_{init}, lb, ub)$ 
14:    Draw  $\xi_n$  from  $\mathcal{N}(0, \sigma_{\xi}^2 \mathbf{I}_{d \times d})$ .
15:     $\mathbf{x}_{n+1} = \max(\min(\hat{\mathbf{x}}_n + \xi_n, ub), lb)$ 
16:  return  $\hat{\mathbf{x}}_n$ 

```

Algorithm 4 updateRFE

```

1: procedure UPDATERFE( $\mathbf{c}_{n-1}, \mathbf{P}_{n-1}^{1/2}, \mathbf{a}_n, y_n$ )
2:   Retrieve  $\mathbf{g}_n \gamma_n^{-1/2}$ ,  $\gamma_n^{-1/2}$  and  $\mathbf{P}_n^{1/2}$  from (5.3)
3:    $\mathbf{c}_n = \mathbf{c}_{n-1} + \mathbf{g}_n(y_n - \mathbf{a}_n \mathbf{c}_{n-1})$ 
4:    $g(\mathbf{x}) = [\cos(\omega_1^T \mathbf{x} + b_1) \dots \cos(\omega_D^T \mathbf{x} + b_D)] \mathbf{c}_n$ 
5:  return  $g(\mathbf{x})$ 

```

4.4. CHOICE OF HYPER-PARAMETERS

In this section, we will analyze the influence of the hyper-parameters of the DONE algorithm and, based on these results, provide practical ways of choosing them. The perfor-

mance of DONE depends on the following hyper-parameters:

- number of basis functions D ,
- p.d.f. p_{Ω} ,
- regularization parameter λ ,
- exploration parameters σ_{ζ} and σ_{ξ} .

The influence of D is straight-forward: increasing D will lead to a better performance (a better RFE fit) of the DONE algorithm at the cost of more computation time. Hence, D should be chosen high enough to get a good approximation, but not too high to avoid unnecessarily high computation times. It should be noted that D does not need to be very precise. Over-fitting should not be a concern for this parameter since we make use of regularization. The exploration parameters determine the trade-off between exploration and exploitation, similar to the use of the acquisition function in Bayesian optimization [136, 187]. The parameter σ_{ζ} influences the exploration of the RFE surrogate in Step 3 of the DONE algorithm, while σ_{ξ} determines exploration of the original function. Assuming both to be close to each other, σ_{ζ} and σ_{ξ} are usually chosen to be equal. If information about local optima of the RFE surrogate or of the original function is available, this could be used to determine good values for these hyper-parameters. Alternatively, similar to Bayesian optimization the expected improvement could be used for that purpose, but this remains for future work. The focus of this section will be on choosing p_{Ω} and λ .

4

4.4.1. PROBABILITY DISTRIBUTION OF FREQUENCIES

Recall the parameters ω_k and b_k from Definition 1, which are obtained by sampling independently from the continuous probability distributions p_{Ω} and $p_B = \text{Uniform}(0, 2\pi)$, respectively. In the following, we will investigate the first and second order moments of the RFE and try to find a distribution p_{Ω} that minimizes the variance of the RFE.

Unfortunately, as shown in Theorem 7 in Appendix 4.8, it turns out that the optimal p.d.f. is

$$p_{\Omega}^*(\omega) = \frac{|\hat{f}(\omega)|\sqrt{\cos(2\angle\hat{f}(\omega) + 2\omega^T \mathbf{x}) + 2}}{\int_{\mathbb{R}^d} |\hat{f}(\tilde{\omega})|\sqrt{\cos(2\angle\hat{f}(\tilde{\omega}) + 2\tilde{\omega}^T \mathbf{x}) + 2d\tilde{\omega}}}. \quad (4.31)$$

This distribution depends on the input \mathbf{x} and both the phase and magnitude of the Fourier transform of f . But if both $|\hat{f}|$ and $\angle\hat{f}$ were known, then the function f itself would be known, and standard optimization algorithms could be used directly. Furthermore, we would like to use a p.d.f. for ω_k that does not depend on the input \mathbf{x} , since the ω_k parameters are chosen independently from the input in the initialization step of the algorithm.

In calibration problems, the objective function f suffers from an unknown offset, $f(\mathbf{x}) = \tilde{f}(\mathbf{x} + \Delta)$. This unknown offset does not change the magnitude in the Fourier domain, but it does change the phase. Since the phase is thus unknown, we choose a

uniform distribution for p_B such that $b_k \in [0, 2\pi]$. However, the magnitude $|\hat{f}|$ can be measured in this case. Section 4.5.2 describes an example of such a problem. We will now derive a way to choose p_Ω for calibration problems.

In order to get a close to optimal p.d.f. for ω_k that is independent of the input \mathbf{x} and of the phase $\angle \hat{f}$ of the Fourier transform of f , we look at a complex generalization of the RFE. In this complex problem, it turns out we can circumvent the disadvantages mentioned above by using a p.d.f. that depends only on $|\hat{f}|$.

Theorem 5. Let $\tilde{G}(\mathbf{x}) = \sum_{k=1}^D \tilde{C}_k e^{i\Omega_k^T \mathbf{x} + B_k}$, with Ω_k being i.i.d. random vectors with a continuous p.d.f. \tilde{p}_Ω over \mathbb{R}^d that satisfies $\tilde{p}_{\Omega_k}(\omega) > 0$ if $|\hat{f}(\omega)| > 0$, and B_k being random variables with uniform distribution from $[0, 2\pi]$. Then $\tilde{G}(\mathbf{x})$ is an unbiased estimator of $f(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$ if

$$\tilde{C}_k = \frac{\hat{f}(\Omega_k) e^{-iB_k}}{D(2\pi)^d \tilde{p}_\Omega(\Omega_k)}. \quad (4.32)$$

For this choice of \tilde{C}_k , the variance of $\tilde{G}(\mathbf{x})$ is minimal if

$$\tilde{p}_\Omega(\omega) = \frac{|\hat{f}(\omega)|}{\int_{\mathbb{R}^d} |\hat{f}(\tilde{\omega})| d\tilde{\omega}}, \quad (4.33)$$

giving a variance of

$$\text{Var}[\tilde{G}(\mathbf{x})] = \frac{1}{D(2\pi)^{2d}} \left(\int_{\mathbb{R}^d} |\hat{f}(\omega)| d\omega \right)^2 - f(\mathbf{x})^2. \quad (4.34)$$

Proof. The unbiasedness follows directly from the Fourier inversion theorem,

$$\begin{aligned} \mathbb{E}[\tilde{G}(\mathbf{x})] &= \sum_{k=1}^D \int_{\mathbb{R}^d} \int_0^{2\pi} \frac{\hat{f}(\omega_k) e^{-ib_k} e^{i\omega_k^T \mathbf{x} + b_k}}{D(2\pi)^d \tilde{p}_\Omega(\omega_k) 2\pi} db_k \tilde{p}_\Omega(\omega_k) d\omega_k \\ &= D \int_{\mathbb{R}^d} \int_0^{2\pi} \frac{\hat{f}(\omega) e^{-ib}}{D(2\pi)^d \tilde{p}_\Omega(\omega)} e^{i\omega^T \mathbf{x} + b} \frac{1}{2\pi} db \tilde{p}_\Omega(\omega) d\omega \\ &= D \int_{\mathbb{R}^d} \frac{\hat{f}(\omega)}{D(2\pi)^d \tilde{p}_\Omega(\omega)} e^{i\omega^T \mathbf{x}} \tilde{p}_\Omega(\omega) \int_0^{2\pi} \frac{1}{2\pi} db d\omega \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{i\omega^T \mathbf{x}} d\omega \\ &= f(\mathbf{x}). \end{aligned} \quad (4.35)$$

The proof of minimum variance is similar to the proof of [212, Thm. 4.3.1]. \square

Note that the coefficients \tilde{C}_k can be complex in this case. Next, we show that the optimal p.d.f. for a complex RFE, \tilde{p}_Ω , is still close-to-optimal (in terms of the second moment) when used in the real RFE from Definition 1.

Theorem 6. Let \tilde{p}_Ω be as in (4.33) and let G with weights C_k be as in Theorem 3. Let P be the set of probability distribution functions for Ω_k that are positive when $|\hat{f}(\omega)| > 0$. Then, we have

$$\mathbb{E}_{\tilde{p}_\Omega, p_B} [G(\mathbf{x})^2] \leq \sqrt{3} \min_{p_\Omega \in P} \mathbb{E}_{p_\Omega, p_B} [G(\mathbf{x})^2]. \quad (4.36)$$

The proof is given in Appendix 4.8. We now discuss how to choose p_Ω in practice.

If no information of $|\hat{f}|$ is available, the standard approach of choosing p_Ω as a zero-mean normal distribution can be used. The variance σ^2 is an important hyper-parameter in this case, and any method of hyper-parameter tuning can be used to find it. However, most hyper-parameter optimization methods are computationally expensive because they require running the whole algorithm multiple times. In the case that $|\hat{f}|$ is not exactly known, but some information about it is available (because it can be estimated or measured for example), this can be circumvented. The variance σ^2 can simply be chosen in such a way that p_Ω most resembles the estimate for $|\hat{f}|$, using standard optimization techniques or by doing this by hand. In this approach, it is not necessary to run the algorithm at all, which is a big advantage compared to most hyper-parameter tuning methods. All of this leads to a rule of thumb for choosing p_Ω as given in Algorithm 5.

4

Algorithm 5 Rule of thumb for choosing p_ω

- 1: **if** $|\hat{f}|$ is known exactly **then**
 - 2: Set $p_\Omega = |\hat{f}| / \int |\hat{f}(\omega)| d\omega$.
 - 3: **else**
 - 4: Measure or estimate $|\hat{f}|$.
 - 5: Determine σ^2 for which the pdf of $\mathcal{N}(0, \sigma^2 \mathbf{I}_{d \times d})$ is close in shape to $|\hat{f}| / \int |\hat{f}(\omega)| d\omega$.
 - 6: Set $p_\Omega = \mathcal{N}(0, \sigma^2 \mathbf{I}_{d \times d})$.
-

4.4.2. UPPER BOUND ON THE REGULARIZATION PARAMETER

The regularization parameter λ in the performance criterion (4.17) is used to prevent under- or over-fitting of the RFE under noisy conditions or when dealing with few measurements. Theorem 4 guarantees the convergence of the least squares solution only if the regularization parameter satisfies $\lambda \leq N\Lambda$, where N is the total number of samples and Λ is defined in (4.21). Here we will provide a method to estimate Λ .

During the proof of Theorem 4, it was shown that the upper bound Λ corresponds to the λ that satisfies

$$\begin{aligned} & \left\| \left(\mathbf{A}_N^T \mathbf{A}_N + N\lambda \mathbf{I}_{D \times D} \right)^{-1} \mathbf{A}_N^T \mathbf{y}_N \right\|_2^2 \\ &= \sum_{k=1}^D \left(\frac{\bar{c}(\omega_k, b_k)}{(2\pi)^d D p_\Omega(\omega_k) p_B(b_k)} \right)^2 = M^2. \end{aligned} \quad (4.37)$$

The left-hand side in this equation is easily evaluated for different values of λ . Thus, in order to estimate Λ , all we need is an approximation of the unknown right hand M^2 .

Like in Section 4.4.1, it is assumed that no information about $\angle \hat{f}$ is available, but that $|\hat{f}|$ can be measured or estimated. Under the assumptions that D is large and that p_{Ω} is a good approximation of $\tilde{p}_{\Omega} = |\hat{f}(\boldsymbol{\omega})| / \int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| d\boldsymbol{\omega}$ as in Algorithm 5, we obtain the following approximation of M :

$$\begin{aligned}
M &= \frac{2}{(2\pi)^d} \sqrt{\frac{1}{D^2} \sum_{k=1}^D \left(\frac{|\hat{f}(\boldsymbol{\omega}_k)|}{p_{\Omega}(\boldsymbol{\omega}_k)} \cos(\angle \hat{f}(\boldsymbol{\omega}_k) - b_k) \right)^2} \\
&\approx \frac{2}{(2\pi)^d} \sqrt{\frac{1}{D} \mathbb{E} \left[\left(\frac{|\hat{f}(\boldsymbol{\Omega}_1)|}{p_{\Omega}(\boldsymbol{\Omega}_1)} \cos(\angle \hat{f}(\boldsymbol{\Omega}_1) - B_1) \right)^2 \right]} \\
&= \frac{2}{(2\pi)^d} \sqrt{\frac{1}{2\pi D} \int_{\mathbb{R}^d} \int_0^{2\pi} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{p_{\Omega}(\boldsymbol{\omega})} \cos^2(\angle \hat{f}(\boldsymbol{\omega}) - b) db d\boldsymbol{\omega}} \\
&= \frac{\sqrt{2}}{(2\pi)^d \sqrt{D}} \sqrt{\int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{p_{\Omega}(\boldsymbol{\omega})} d\boldsymbol{\omega}} \\
&\approx \frac{\sqrt{2}}{(2\pi)^d \sqrt{D}} \sqrt{\int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\tilde{p}_{\Omega}(\boldsymbol{\omega})} d\boldsymbol{\omega}} \\
&= \frac{\sqrt{2}}{(2\pi)^d \sqrt{D}} \int |\hat{f}(\boldsymbol{\omega})| d\boldsymbol{\omega} = M_a. \tag{4.38}
\end{aligned}$$

The squared cosine was removed as in Eq. (4.12). Using the exact value or an estimate of $\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| d\boldsymbol{\omega}$ as in Algorithm 5 to determine M_a , we calculate the left-hand in (4.37) for multiple values of Λ and take the value for which it is closest to M_a^2 . The procedure is summarized in Algorithm 6.

Algorithm 6 Rule of thumb for finding an estimate of Λ

- 1: Run Algorithm 5 to get $\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| d\boldsymbol{\omega}$.
 - 2: Take N measurements to get \mathbf{A}_N and \mathbf{y}_N .
 - 3: Determine Λ for which the left-hand side of (4.37) is close to $M_a^2 = \frac{2}{(2\pi)^{2d} D} \left(\int |\hat{f}(\boldsymbol{\omega})| d\boldsymbol{\omega} \right)^2$.
-

4.5. NUMERICAL EXAMPLES

In this section, we compare the DONE algorithm to the Bayesian optimization library BayesOpt [139] in several numerical examples.

4.5.1. ANALYTIC BENCHMARK PROBLEM: CAMELBACK FUNCTION

The camelback function

$$f(\mathbf{x}) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3} \right) x_1^2 + x_1 x_2 + (-4 + 4x_2^2) x_2^2, \tag{4.39}$$

where $\mathbf{x} = [x_1, x_2] \in [-2, 2] \times [-1, 1]$, is a standard test function with two global minima and two local minima. The locations of the global minima are approximately $(0.0898, -0.7126)$ and $(-0.0898, 0.7126)$ with an approximate function value of -1.0316 . We determined the hyper-parameters for DONE on this test function as follows. First, we computed the Fourier transform of the function. We then fitted a function $h(\boldsymbol{\omega}) = \frac{C}{\sigma\sqrt{2\pi}} e^{-\frac{\boldsymbol{\omega}^2}{2\sigma^2}}$ to the magnitude of the Fourier transform in both directions. This was done by trial and error, giving a value of $\sigma = 10$. To validate, two RFEs were fit to the original function using a normal distribution with standard deviation $\sigma = 10$ (good fit) and $\sigma = 0.1$ (bad fit) for $\boldsymbol{\omega}_k$, using the least squares approach from Section 4.2.2. Here, we used $N = 1000$ measurements sampled uniformly from the input domain, the number of basis functions D was set to 500, and a regularization parameter of $\lambda = 10^{-10}$ was used. The small value for λ still works well in practice because the function f does not contain noise.

Let $g(\mathbf{x})$ denote the value of the trained RFE at point \mathbf{x} . We investigated the root mean squared error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{n=1}^N (f(\mathbf{x}_n) - g(\mathbf{x}_n))^2}, \quad (4.40)$$

for the two stated values of σ . The good fit gave a RMSE of $5.5348 \cdot 10^{-6}$, while the bad fit gave a RMSE of 0.2321, which shows the big impact of this hyper-parameter on the least squares fit.

We also looked at the difference between using the real RFE from Definition 1 and the complex RFE from Theorem 5, for $\sigma = 10$, and for different values of D ($D \in \{10, 20, 40, 80, 160, 320, 640, 1280\}$). Fig. 4.1 shows the mean and standard deviation of the RMSE over 100 runs. We see that the real RFE indeed performs similar to the complex RFE as predicted by Theorem 8 in Appendix 4.8.

Using the hyper-parameters $\sigma = 10$ and $\lambda = 10^{-10}$, we also performed 10 runs of the DONE algorithm and compared it to reproduced results from [139, Table 1] (method ‘‘BayesOpt1’’). The number of basis functions D was set to 500, one of the smallest values with a RMSE of below 10^{-5} according to Fig. 4.1, and the initial guess was chosen randomly. The exploration parameters σ_ζ and σ_ξ were set to 0.01. The resulting distance to the true minimum and the computation time in seconds (with their standard deviations) for 50 and 100 measurements can be found in Table 4.1. As in [139], the computation time for BayesOpt was only shown for 100 samples and the accuracy below 10^{-5} was not shown. It can be seen that the DONE algorithm is several orders of magnitude more accurate and about 5 times faster when compared to BayesOpt for this problem.

4.5.2. OPTICAL COHERENCE TOMOGRAPHY

Optical coherence tomography (OCT) is a low-coherence interferometry imaging technique used for making three-dimensional images of a sample. The quality and resolution of images is reduced by optical wavefront aberrations caused by the medium, e.g., the human cornea when imaging the retina. These aberrations can be removed by using active components such as deformable mirrors in combination with optimization

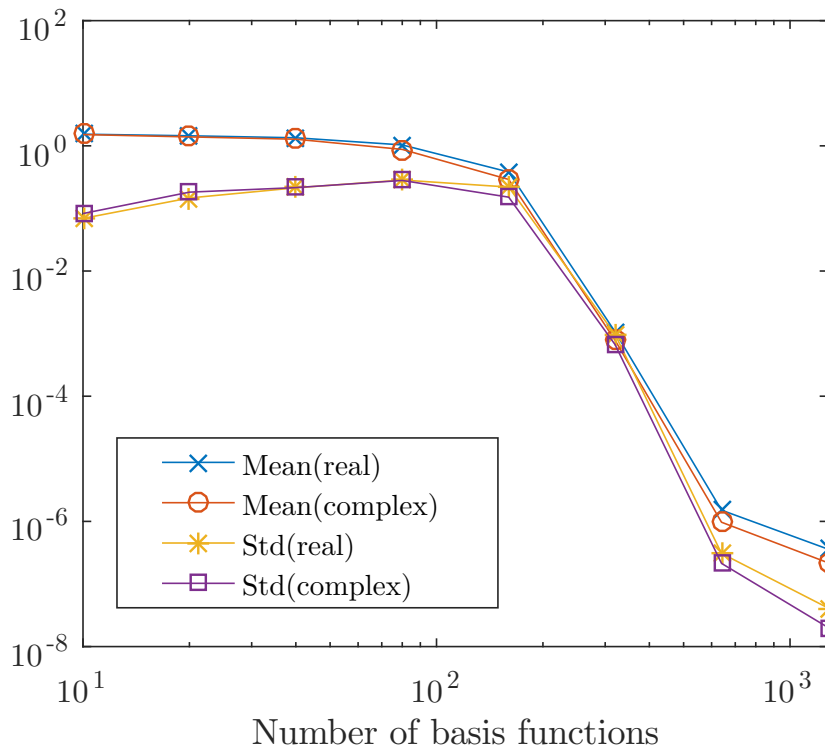


Figure 4.1: Mean and standard deviation of the root mean square error for a real and a complex RFE over 100 runs.

Table 4.1: DONE vs BayesOpt on the Camelback function

	Dist. to min. (50 samp.)	Time (50 samp.)
DONE	$2.1812 \cdot 10^{-9}$ ($8.3882 \cdot 10^{-9}$)	0.0493 (0.0015)
BayesOpt	0.0021 (0.0044)	-
	Dist. to min. (100 samp.)	Time (100 samp.)
DONE	$1.1980 \cdot 10^{-9}$ ($5.2133 \cdot 10^{-9}$)	0.0683 (0.0019)
BayesOpt	$< 1 \cdot 10^{-5}$ ($< 1 \cdot 10^{-5}$)	0.3049 (0.0563)

algorithms [144, 172]. The arguments of the optimization can be the voltages of the deformable mirror or a mapping of these voltages to other coefficients such as the coefficients of Zernike polynomials. The intensity of the image at a certain depth is then maximized to remove as much of the aberrations as possible. In [172] it was shown experimentally that the DONE algorithm greatly outperforms other derivative-free algorithms in final root mean square (RMS) wavefront error and image quality. Here, we numerically compare the DONE algorithm to BayesOpt [139]. The numerical results are obtained by simulating the OCT transfer function as described in [156, 213] and maximizing the OCT signal. The input dimension for this example is three. Three Zernike aberrations are considered, namely the defocus and two astigmatisms. These are generally the largest optical wavefront aberrations in the human eye. The noise of a real OCT signal is approximated by adding Gaussian white noise with a standard deviation of 0.01. The results are shown in Fig. 4.2. For the DONE algorithm the same parameters are used as described in [172], only λ is chosen to be equal to 3. The number of cosines $D = 1000$ is chosen as large as possible such that the computation time still remains around 1 ms. This is sufficiently fast to keep up with modern OCT B-scan acquisition and processing rates. The DONE algorithm is compared to BayesOpt with the default parameters and to BayesOpt with only one instead of 10 prior measurements, the latter is referred to as BayesOpt-1 init. Other values for the parameters of BayesOpt, obtained with trial and error, did not result in a significant performance increase. To use the BayesOpt algorithm, the inputs had to be normalized between 0 and 1. For each input aberration, the region $-0.45 \mu\text{m}$ to $0.45 \mu\text{m}$ was scaled to the region 0 to 1. The results for BayesOpt and DONE are very similar. The mean error of the DONE algorithm is slightly lower than the BayesOpt algorithm. However, the total average computation time for the DONE algorithm was 93 ms, while the total average computation time of Bayesopt was 1019 ms.

4.5.3. TUNING OF AN OPTICAL BEAM-FORMING NETWORK

In wireless communication systems, optical beam-forming networks (OBFNs) can be used to steer the reception or transmission angle of a phased array antenna [189] in the desired direction. In the case of reception, the signals that arrive at the different antenna elements of the phased array are combined in such a way that positive interference of the signals occurs only in a specific direction. A device based on optical ring resonators [190] (ORRs) that can perform this signal processing technique in the optical domain was proposed in [191]. This OBFN can provide accurate control of the reception angle in broadband wireless receivers.

To achieve a maximal signal-to-noise ratio (SNR), the actuators in the OBFN need to be adapted according to the desired group delay of each OBFN path, which can be calculated from the desired reception angle. Each ORR is controlled by two heaters that influence its group delay, however the relation between heater voltage and group delay is nonlinear. Even if the desired group delay is available, controlling the OBFN comes down to solving a nonlinear optimization problem. Furthermore, the physical model of the OBFN can become quite complex if many ORRs are used, and the models are prone to model inaccuracies. Therefore, a black-box approach like in the DONE algorithm could help in the tuning of the OBFN. Preliminary results using RFEs in an offline fashion on this application can be found in [194]. Here, we demonstrate the advantage

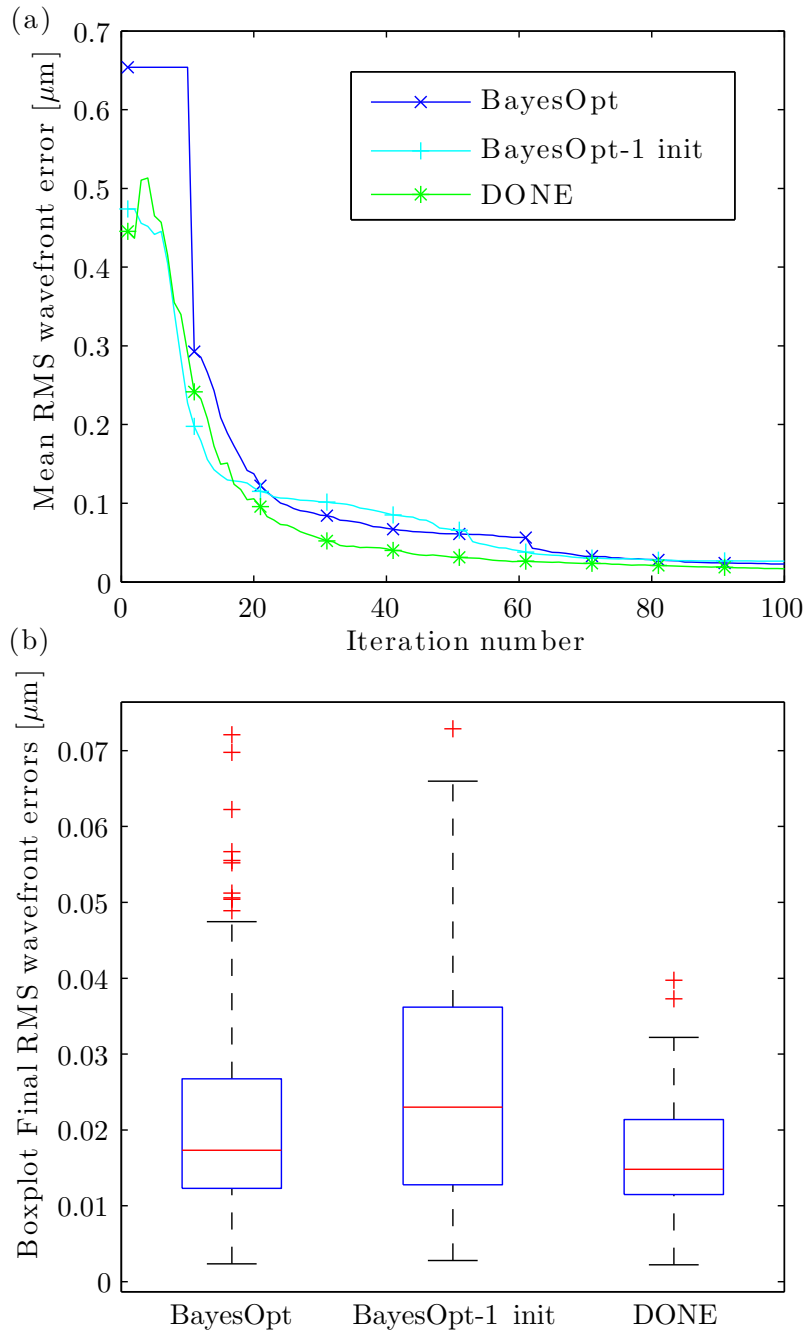


Figure 4.2: (a) The RMS wavefront error of DONE and BayesOpt averaged over 100 simulations versus the number of iterations. (b) A boxplot of 100 final RMS wavefront errors after 100 iterations for DONE and BayesOpt. On each box, the central line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers. Outliers are plotted individually.

of online processing in terms of performance by using DONE instead of the offline algorithm in [194].

An OBFN simulation based on the same physical models as in [194] will be used in this section, with the following differences: 1) the implementation is done in C++; 2) ORR properties are equal for each ORR; 3) heater voltages with offset and crosstalk [193, Appendix B] have been implemented; 4) a small region outside the bandwidth of interest has a desired group delay of 0; 5) an 8×1 OBFN with 12 ORRs is considered; 6) the standard deviation of the measurement noise was set to $7.5 \cdot 10^{-3}$. The input of the simulation is the normalized heater voltage for each ORR, and the output is the corresponding mean square error of the difference between OBFN path group delays and desired delays. The simulation contains 24 heaters (two for each ORR, namely one for the phase shift and one for the coupling constant), making the problem 24-dimensional. Each heater influences the delay properties of the corresponding ORR, and together they influence the OBFN path group delays.

The DONE algorithm was used on this simulation to find the optimal heater voltages. The number of basis functions was $D = 6000$, which was the lowest number that gave an adequate performance. The p.d.f. p_{Ω} was a normal distribution with variance 0.5. The regularization parameter was $\lambda = 0.1$. The exploration parameters were $\sigma_{\zeta} = \sigma_{\xi} = 0.01$. In total, 3000 measurements were taken.

Just like in the previous application, the DONE algorithm was compared to the Bayesian optimization library BayesOpt [139]. The same simulation was used in both algorithms, and BayesOpt also had 3000 function evaluations available. The other parameters for BayesOpt were set to their default values, except for the noise parameter which was set to 0.1 after calculating the influence of the measurement noise on the objective function. Also, in-between hyper-parameter optimization was turned off after noticing it did not influence the results while being very time-consuming.

The results for both algorithms are shown in Fig. 4.3. The found optimum at each iteration is shown for the two algorithms. For DONE, the mean of 10 runs is shown, while for BayesOpt only one run is shown because of the much longer computation time. The dotted line represents an offline approach: it is the average of 10 runs of a similar procedure as in [194], where a RFE with the same hyper-parameters as in DONE was fitted to 3000 random measurements and then optimized. The figure clearly shows the advantage of the online approach: because measurements are only taken in regions where the objective function is low, the RFE model can become very accurate in this region. The figure also shows that DONE outperforms BayesOpt for this application in terms of accuracy. On top of that, the total computation time shows a big improvement: one run of the DONE algorithm took less than 2 minutes, while one run of BayesOpt took 5800 minutes.

The big difference in computation time for the OBFN application can be explained by looking at the total number of measurements N . Even though the input dimension is high compared to the other problems, N is the main parameter that causes BayesOpt to slow down for a large number of measurements. This is because the models used in Bayesian optimization typically depend on the kernel matrix of all samples, which will increase in size each iteration. The runtime for one iteration of the DONE algorithm is, in contrast, independent of the number of previous measurements.

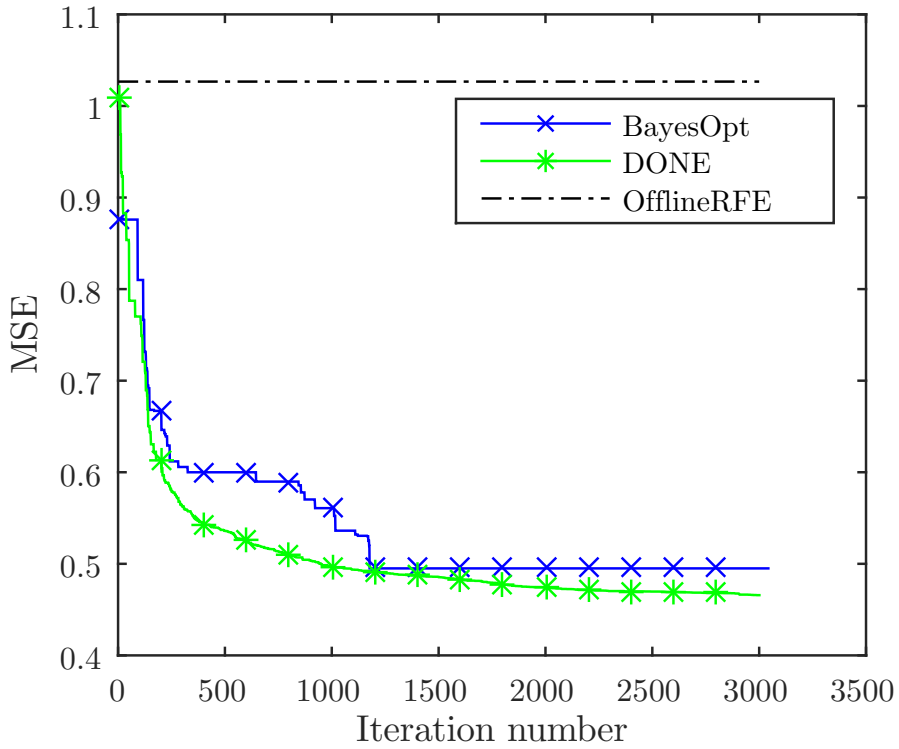


Figure 4.3: The mean square error of DONE and BayesOpt applied to the OBFN application, plotted versus the number of iterations. For DONE, the values are averaged over 10 runs. For BayesOpt only 1 run is shown. The dotted line is the result of fitting a RFE using 3000 random measurements and optimizing that RFE, averaged over 10 runs.

4.5.4. ROBOT ARM MOVEMENT

The previous two examples have illustrated how the DONE algorithm outperforms BayesOpt in terms of speed (both OCT and OFBN) and how its online processing scheme reduces the number of required measurements compared to offline processing (OFBN), respectively. The dimensions in both problems were three and 27, respectively, which is still relatively modest. To illustrate that DONE also works in higher dimensions, we will now consider a toy example from robotics. The following model of a three-link-planar robot, which has been adapted from [195], is considered:

$$a_i(k) = u_i(k) + \sin\left(\pi/180 \sum_{j=1}^i \alpha_j(k-1)\right) \cdot 9.8 \cdot 0.05, \quad (4.41)$$

$$v_i(k) = v_i(k-1) + a_i(k), \quad (4.42)$$

$$\alpha_i(k) = \alpha_i(k-1) + v_i(k), \quad (4.43)$$

$$x(k) = \sum_{j=1}^3 l_j \cos\left(\pi/2 + \pi/180 \sum_{j=1}^i \alpha_j(k)\right), \quad (4.44)$$

$$y(k) = \sum_{j=1}^3 l_j \sin\left(\pi/2 + \pi/180 \sum_{j=1}^i \alpha_j(k)\right). \quad (4.45)$$

Here, $\alpha_i(k)$ represents the angle in degrees of link i at time step k , $v_i(k)$ and $a_i(k)$ are the first and second derivative of the angles, $u_i(k) \in [-1, 1]$ is the control input, $x(k)$ and $y(k)$ denote the position of the tip of the arm, and $l_1 = l_2 = 8.625$ and $l_3 = 6.125$ are the lengths of the links. The variables are initialized as $a_i(0) = v_i(0) = \alpha_i(0) = 0$ for $i = 1, 2, 3$. We use the DONE algorithm to design a sequence of control inputs $u_i(1), \dots, u_i(50)$ such that the distance between the tip of the arm and a fixed target at location $(6.96, 12.66)$ at the 50-th time step is minimized. The input for the DONE algorithm is thus a vector containing $u_i(k)$ for $i = 1, 2, 3$ and $k = 1, \dots, 50$. This makes the problem 150-dimensional. The output is the distance between the tip and the target at the 50-th time step. The initial guess for the algorithm was set to a random control sequence with a uniform distribution over the set $[-1, 1]$ for each robot arm i . We would like to stress that this example has been chosen for its high-dimensional input. We do not consider this approach a serious contender for specialized control methods in robotics.

The hyper-parameters for the DONE algorithm were chosen as follows. The number of basis functions was $D = 3000$, which was the lowest number that gave consistent results. The regularization parameter was $\lambda = 10^{-3}$. The p.d.f. p_{Ω} was set to a normal distribution with variance one. The exploration parameters were set to $\sigma_{\zeta} = \sigma_{\xi} = 5 \cdot 10^{-5}$. The number of measurements N was set to 10000.

No comparison with other algorithms has been made for this application. The computation time of the Bayesian optimization algorithm scales with the number of measurements and would be too long with 10000 measurements, as can be seen in Table 4.2. Algorithms like reinforcement learning use other principles, hence no comparison is given. Our main purpose with this application is to demonstrate the applicability of the DONE algorithm to high-dimensional problems.

Figure 4.4 shows the distance to the target at time step 50 for different iterations of the DONE algorithm, averaged over 10 runs with different initial guesses. The control

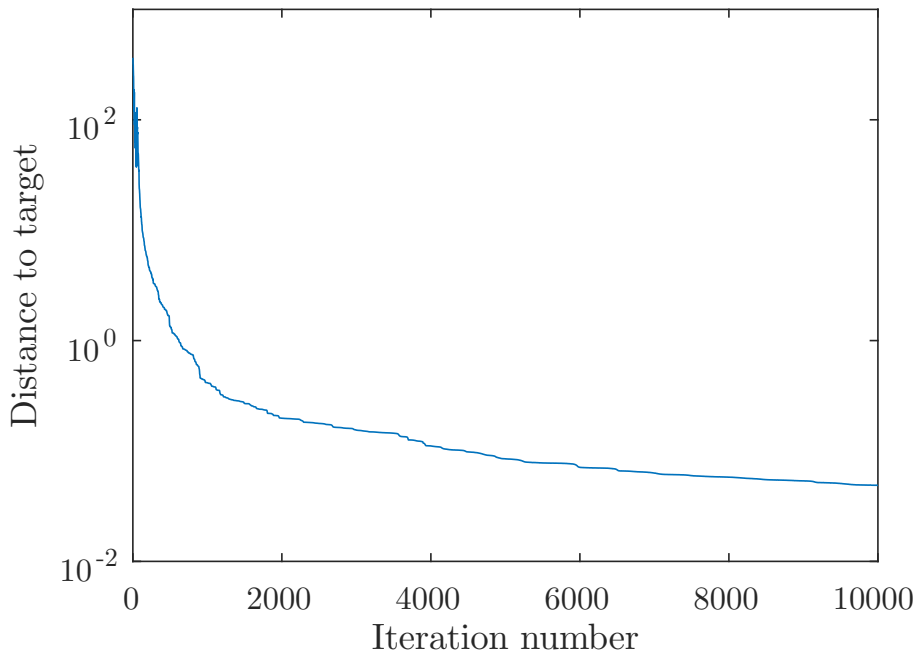


Figure 4.4: The mean distance to target for the robot arm at time step 50, after minimizing this distance with DONE, plotted versus the number of iterations, averaged over 10 runs.

sequences converge to a sequence for which the robot arm goes to the target, i.e., DONE has successfully been applied to a problem with a high input dimension. The number of basis functions required did not increase when compared to the other applications in this paper, although more measurements were required. The computation time for this example and the other examples is shown in Table 4.2.

Table 4.2: Computation Time: DONE vs BayesOpt

Problem	Method	Input dim.	N	D	Time (s)
Camelback	DONE	2	100	50	0.0683
	BayesOpt	2	100	-	0.3049
OCT	DONE	3	100	1000	0.093
	BayesOpt	3	100	-	1.019
OBFN	DONE	24	3000	6000	99.7
	BayesOpt	24	3000	-	$3.48 \cdot 10^5$
Robot arm	DONE	150	10000	3000	99.1

4

4.6. CONCLUSIONS

We have analyzed an online optimization algorithm called DONE that is used to find the minimum of a function using measurements that are costly and corrupted by noise. DONE maintains a surrogate model in the form of a random Fourier expansion (RFE), which is updated whenever a new measurement is available, and minimizes this surrogate with standard derivative-based methods. This allows to measure only in regions of interest, reducing the overall number of measurements required. The DONE algorithm is comparable to Bayesian optimization algorithms, but it has the distinctive advantage that the computational complexity of one iteration does not grow with the number of measurements that have already been taken.

As a theoretical result, we have shown that a RFE that is trained with linear least squares can approximate square integrable functions arbitrarily well, with high probability. An upper bound on the regularization parameter used in this training procedure was given, as well as an optimal and a more practical probability distribution for the parameters that are chosen randomly. We applied the DONE algorithm to an analytic benchmark problem and to three applications: optical coherence tomography, optical beam-forming network tuning, and a robot arm. We compared the algorithm to BayesOpt, a Bayesian optimization library. The DONE algorithm gave accurate results on these applications while being faster than the Bayesian optimization algorithm, due to the fixed computational complexity per iteration.

4.7. PROOF OF CONVERGENCE OF THE LEAST SQUARES SOLUTION

In this section, we show that using the least squares solution in the RFE gives a function that approximates the true unknown function f . To prove this, we make use of the results in [207] and of [214, Thm. 2] and [215, Key Thm.].

Proof of Theorem 4. Let the constant $m > 0$ be given by

$$m = \left\| \left(\frac{1}{N} \mathbf{A}_N^T \mathbf{A}_N + \frac{\lambda}{N} \mathbf{I}_{D \times D} \right)^{-1} \frac{1}{N} \mathbf{A}_N^T \mathbf{y}_N \right\|_2, \quad (4.46)$$

and define the set $C_m = \{\mathbf{c} \in \mathbb{R}^D : \|\mathbf{c}\|_2 \leq m\}$. Note that C_m is a compact set. The least squares weight vector

$$\begin{aligned} \mathbf{c}_N &= (\mathbf{A}_N^T \mathbf{A}_N + \lambda \mathbf{I}_{D \times D})^{-1} \mathbf{A}_N^T \mathbf{y}_N \\ &= \left(\frac{1}{N} \mathbf{A}_N^T \mathbf{A}_N + \frac{\lambda}{N} \mathbf{I}_{D \times D} \right)^{-1} \frac{1}{N} \mathbf{A}_N^T \mathbf{y}_N, \end{aligned} \quad (4.47)$$

is also the solution to the constrained, but unregularized least squares problem (see [216, Sec. 12.1.3])

$$\mathbf{c}_N = \underset{\mathbf{c} \in C_m}{\operatorname{argmin}} \frac{1}{N} \|\mathbf{y}_N - \mathbf{A}_N \mathbf{c}\|_2^2. \quad (4.48)$$

Now, note that a decrease in λ leads to an increase in m . Since $\lambda/N \leq \Lambda$ by assumption and the upper bound Λ in Theorem 4 satisfies

$$\left\| \left(\frac{1}{N} \mathbf{A}_N^T \mathbf{A}_N + \Lambda \mathbf{I}_{D \times D} \right)^{-1} \frac{1}{N} \mathbf{A}_N^T \mathbf{y}_N \right\|_2 = M, \quad (4.49)$$

$$M = \sqrt{\sum_{k=1}^D \left(\frac{\bar{c}(\boldsymbol{\omega}_k, b_k)}{(2\pi)^d D p_{\boldsymbol{\Omega}}(\boldsymbol{\omega}_k) p_B(b_k)} \right)^2}, \quad (4.50)$$

we have that $m \geq M$. We will need this lower bound on m to make use of the results in [207] later on in this proof.

Recall from Section 4.2.2 that the vector \mathbf{y}_N depends on the function evaluations and on measurement noise η that is assumed to be zero-mean and of finite variance σ_H^2 . We first consider the noiseless case, i.e. $y_n = f(\mathbf{x}_n)$. For $\mathbf{x} \in \mathcal{X}$, $\mathbf{c} \in \mathbb{R}^D$, let

$$E(\mathbf{x}, \mathbf{c}) = f(\mathbf{x}) - \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k). \quad (4.51)$$

Using the Cauchy-Schwarz inequality, we have the following bound for all $\mathbf{x} \in \mathcal{X}$, $\mathbf{c} \in C_m$:

$$E(\mathbf{x}, \mathbf{c})^2 = f(\mathbf{x})^2 + \left(\sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k) \right)^2$$

$$\begin{aligned}
 & -2f(\mathbf{x}) \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k) \\
 \leq & f(\mathbf{x})^2 + \left(\sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k) \right)^2 \\
 & + 2|f(\mathbf{x})| \left| \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k) \right| \\
 \leq & f(\mathbf{x})^2 + \sum_{k=1}^D |c_k|^2 + 2|f(\mathbf{x})| \sqrt{\sum_{k=1}^D |c_k|^2} \\
 \leq & f(\mathbf{x})^2 + m^2 + 2f(\mathbf{x})m \\
 \leq & (\|f\|_\infty + m)^2. \tag{4.52}
 \end{aligned}$$

Note that $E(\mathbf{x}, \mathbf{c})$ is continuous in \mathbf{c} and measurable in \mathbf{x} . Let now \mathbf{X}_n denote i.i.d. random vectors with distribution $p_{\mathbf{X}}$. Using Theorem [214, Thm. 2] we get, with probability one,

$$\lim_{N \rightarrow \infty} \sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| = 0. \tag{4.53}$$

Since almost sure convergence implies convergence in probability [217, Ch. 2], we also have:

$$\begin{aligned}
 \lim_{N \rightarrow \infty} P \left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 \right. \right. \\
 \left. \left. - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| > \epsilon \right) = 0 \quad \forall \epsilon > 0. \tag{4.54}
 \end{aligned}$$

We will need this result when considering the case with noise. For the case with noise, i.e. $y_n = f(\mathbf{x}_n) + \eta_n$, let

$$\begin{aligned}
 \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 &= \left(f(\mathbf{x}) + \eta - \sum_{k=1}^D c_k \cos(\boldsymbol{\omega}_k^T \mathbf{x} + b_k) \right)^2 \\
 &= E(\mathbf{x}, \mathbf{c})^2 + 2\eta E(\mathbf{x}, \mathbf{c}) + \eta^2. \tag{4.55}
 \end{aligned}$$

Using the properties of the noise η with p.d.f. p_H , this gives the following mean square error:

$$\begin{aligned}
 & \int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \\
 &= \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) \left(\int_{\mathbb{R}} p_H(\eta) d\eta \right) d\mathbf{x} \\
 &+ 2 \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c}) \left(\int_{\mathbb{R}} \eta p_H(\eta) d\eta \right) p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
 &+ \int_{\mathcal{X}} p_{\mathbf{X}}(\mathbf{x}) \left(\int_{\mathbb{R}} \eta^2 p_H(\eta) d\eta \right) d\mathbf{x}
 \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} + \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c}) \underbrace{\mathbb{E}[H_n]}_{=0} p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \\
&\quad + \mathbb{E}[H_n^2] \\
&= \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} + \sigma_H^2. \tag{4.56}
\end{aligned}$$

Here, H_n is a random variable with distribution p_H . For any choice of $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3 > 0$ such that $\epsilon_1 + \epsilon_2 + \epsilon_3 = \epsilon_0$, we have, following a similar proof as in [218, Thm. 3.3(a)]:

$$\begin{aligned}
&P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N \tilde{E}(\mathbf{X}_n, H_n, \mathbf{c})^2 - \int_{\mathcal{X}} \int_{\mathbb{R}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \right| > \epsilon_0\right) \\
&= P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 + \frac{2}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) + \frac{1}{N} \sum_{n=1}^N H_n^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} - \sigma_H^2 \right| > \epsilon_0\right) \\
&\leq P\left(\sup_{\mathbf{c} \in C_m} \left\{ \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| + \left| \frac{2}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) \right| + \left| \frac{1}{N} \sum_{n=1}^N H_n^2 - \sigma_H^2 \right| \right\} > \epsilon_0\right) \\
&\leq P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| + \sup_{\mathbf{c} \in C_m} \left| \frac{2}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) \right| + \left| \frac{1}{N} \sum_{n=1}^N H_n^2 - \sigma_H^2 \right| > \epsilon_0\right) \\
&\leq P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| > \epsilon_1 \right. \\
&\quad \text{or } \sup_{\mathbf{c} \in C_m} \left| \frac{2}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) \right| > \epsilon_2 \\
&\quad \left. \text{or } \left| \frac{1}{N} \sum_{n=1}^N H_n^2 - \sigma_H^2 \right| > \epsilon_3 \right) \\
&\leq P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N E(\mathbf{X}_n, \mathbf{c})^2 - \int_{\mathcal{X}} E(\mathbf{x}, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right| > \epsilon_1\right) \\
&\quad + P\left(\sup_{\mathbf{c} \in C_m} \left| \frac{2}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) \right| > \epsilon_2\right) \\
&\quad + P\left(\left| \frac{1}{N} \sum_{n=1}^N H_n^2 - \sigma_H^2 \right| > \epsilon_3\right).
\end{aligned}$$

Of these last three probabilities, the first one is proven to converge to zero in (4.54), while the last one converges to zero by the weak law of large numbers. For the second probability, we can make use of Theorem [214, Thm. 2] again, noting that $\eta_n E(\mathbf{x}_n, \mathbf{c})$ is continuous in \mathbf{c} . We use (4.52) to get

$$|\eta E(\mathbf{x}, \mathbf{c})| \leq |\eta| (\|f\|_\infty + m) \quad \forall \mathbf{x}, \eta, \mathbf{c}. \quad (4.57)$$

Again, since uniform convergence implies convergence in probability, and since $\mathbb{E}[H_n E(\mathbf{X}_n, \mathbf{c})] = \mathbb{E}[H_n] \mathbb{E}[E(\mathbf{X}_n, \mathbf{c})] = 0$ for all n , using Theorem [214, Thm. 2] gives the desired convergence in probability

$$\lim_{N \rightarrow \infty} P \left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N H_n E(\mathbf{X}_n, \mathbf{c}) \right| > \epsilon_2 \right) = 0 \quad \forall \epsilon_2. \quad (4.58)$$

4

Together with the other two convergences and (4.57) we get:

$$\begin{aligned} \lim_{N \rightarrow \infty} P \left(\sup_{\mathbf{c} \in C_m} \left| \frac{1}{N} \sum_{n=1}^N \tilde{E}(\mathbf{X}_n, H_n, \mathbf{c})^2 \right. \right. \\ \left. \left. - \int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \right| > \epsilon \right) = 0. \end{aligned} \quad (4.59)$$

The following bound follows from (4.52) and (4.56):

$$\begin{aligned} 0 &\leq \int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \\ &\leq (\|f\|_\infty + m)^2 + \sigma_H^2. \end{aligned} \quad (4.60)$$

In light of this bound, [215, Key Thm.] now implies that the mean square error between the output of the RFE with least squares weight vector and the noisy measurements is approaching its ideal value as the number of samples increases. More precisely, for any choice of $\epsilon_4 > 0$ and $\delta_1 > 0$, there exists an N_0 such that, for all $N > N_0$,

$$\begin{aligned} \left| \int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{C}_N)^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \right. \\ \left. - \int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{C}^0)^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta \right| < \epsilon_4 \end{aligned} \quad (4.61)$$

with probability at least $1 - \delta_1$. Here, \mathbf{C}_N denotes the vector \mathbf{c}_N as a random variable as it depends on the input and noise samples and on the samples $\omega_1, \dots, \omega_D, b_1, \dots, b_D$, and $\mathbf{C}^0 \in C_m$ minimizes $\int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c}) p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta$. Next, it is shown that the same holds for the mean square error between the least-squares RFE outputs and the unknown, noise-free function values.

According to [207, Thm 3.2], for any $\delta_2 > 0$, with probability at least $1 - \delta_2$ w.r.t. $\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_D$ and B_1, \dots, B_D , there exists a $\mathbf{c} \in C_m$ with the following bound¹:

$$\int_{\mathcal{X}} \left(f(\mathbf{x}) - \sum_{k=1}^D c_k \cos(\mathbf{\Omega}_k^T \mathbf{x} + B_k) \right)^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \frac{\gamma(\delta_2)^2}{D},$$

¹The weights found in the proof of the cited theorem satisfy $\mathbf{c} \in C_m$ if $m \geq M$, which was shown in the beginning of this appendix. Here we also made use of the result from Theorem 1 of this paper to get what is denoted with α in [207]. We have also used, with the notation of [207], that $\|f - \hat{f}\|_\mu \leq \|f - \hat{f}\|_\infty$.

$$\begin{aligned}\gamma(\delta_2) &= \sup_{\omega, b} \left| \frac{1}{(2\pi)^d} \frac{\tilde{c}(\omega, b)}{p_{\mathbf{\Omega}}(\omega) p_B(b)} \right| \left(\sqrt{\log \frac{1}{\delta_2} + 4r} \right), \\ r &= \sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x}\|_2 \sqrt{\sigma^2 d + \pi^2/3},\end{aligned}\quad (4.62)$$

with σ^2 denoting the variance of $p_{\mathbf{\Omega}}$. For this particular \mathbf{c} , (4.55), (4.56) and (4.62) imply that

$$\int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{c})^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta < \frac{\gamma(\delta_2)^2}{D} + \sigma_H^2. \quad (4.63)$$

Since $\mathbf{C}^0 \in C_m$ minimizes the left-hand in the equation above by definition, we also have that

$$\int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{C}^0)^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta < \frac{\gamma(\delta_2)^2}{D} + \sigma_H^2 \quad (4.64)$$

with probability at least $1 - \delta_2$. Since the event in (4.64) only depends on $\mathbf{\Omega}_1, \dots, \mathbf{\Omega}_D$ and B_1, \dots, B_D , while the event in (4.61) only depends on the input and noise samples, we can combine these two equations as follows. For any choice of $\epsilon_4 > 0$, $\delta_1 > 0$ and $\delta_2 > 0$, there exists an N_0 such that, for all $N > N_0$,

$$\int_{\mathbb{R}} \int_{\mathcal{X}} \tilde{E}(\mathbf{x}, \eta, \mathbf{C}_N)^2 p_{\mathbf{X}}(\mathbf{x}) p_H(\eta) d\mathbf{x} d\eta < \epsilon_4 + \frac{\gamma(\delta_2)^2}{D} + \sigma_H^2 \quad (4.65)$$

with probability at least $(1 - \delta_1)(1 - \delta_2)$. Using (4.56) now gives the following result. For any choice of $\epsilon_4 > 0$, $\delta_1 > 0$ and $\delta_2 > 0$, there exists an N_0 such that, for all $N > N_0$, we have

$$\int_{\mathcal{X}} E(\mathbf{x}, \mathbf{C}_N)^2 p_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} < \epsilon_4 + \frac{\gamma(\delta_2)^2}{D} \quad (4.66)$$

with probability at least $(1 - \delta_1)(1 - \delta_2)$.

Choosing $D_0, \epsilon_4, \delta_1$ and δ_2 such that $D_0 > \gamma(\delta_2)^2/(\epsilon - \epsilon_4)$ and $(1 - \delta_1)(1 - \delta_2) = \delta$ concludes the proof. \square

4.8. MINIMUM-VARIANCE PROPERTIES

The following theorem presents the probability density function for $\mathbf{\Omega}_k$ that minimizes the variance of a RFE at a fixed measurement location \mathbf{x} .

Theorem 7. *Given \mathbf{x} , the p.d.f. $p_{\mathbf{\Omega}}^*$ that minimizes the variance of the unbiased estimator $G(\mathbf{x}) = \sum_{k=1}^D C_k \cos(\mathbf{\Omega}_k^T \mathbf{x} + B_k)$ as defined in Theorem 1, with C_k as defined in Theorem 3, is equal to*

$$p_{\mathbf{\Omega}}^*(\omega) = \frac{|\hat{f}(\omega)| \sqrt{\cos(2\angle \hat{f}(\omega) + 2\omega^T \mathbf{x}) + 2}}{\int_{\mathbb{R}^d} |\hat{f}(\tilde{\omega})| \sqrt{\cos(2\angle \hat{f}(\tilde{\omega}) + 2\tilde{\omega}^T \mathbf{x}) + 2d\tilde{\omega}}}. \quad (4.67)$$

For this choice of p_{Ω} , the variance is equal to

$$\frac{1}{2D(2\pi)^{2d}} \left(\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| \sqrt{\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2d\boldsymbol{\omega}} \right)^2 - f(\mathbf{x})^2. \quad (4.68)$$

Proof. The proof is similar to the proof of [212, Thm. 4.3.1]. Let q_{Ω} be any p.d.f. of Ω_k that satisfies $q_{\Omega}(\boldsymbol{\omega}) > 0$ if $|\hat{f}(\boldsymbol{\omega})| > 0$. Let $\text{Var}_{q_{\Omega}, p_B}$ be the variance of $G(\mathbf{x})$ under the assumption that $p_{\Omega} = q_{\Omega}$, $p_B = \text{Uniform}(0, 2\pi)$, and $C_k = \frac{2}{D(2\pi)^d} \frac{|\hat{f}(\Omega_k)|}{q_{\Omega}(\Omega_k)} \cos(\angle \hat{f}(\Omega_k) - B_k)$. According to Theorem 3, this choice for C_k makes sure that $G(\mathbf{x})$ is an unbiased estimator, i.e., $f(\mathbf{x}) = \mathbb{E}[G(\mathbf{x})]$. The variance of $G(\mathbf{x})$ can be computed as:

4

$$\begin{aligned} & \text{Var}_{q_{\Omega}, p_B}[G(\mathbf{x})] \\ &= \text{Var}_{q_{\Omega}, p_B} \left[\sum_{k=1}^D C_k \cos(\Omega_k^T \mathbf{x} + B_k) \right] \\ &= D \text{Var}_{q_{\Omega}, p_B} [C_1 \cos(\Omega_1^T \mathbf{x} + B_1)] \\ &= \frac{D}{2\pi} \int_{\mathbb{R}^d} \int_0^{2\pi} \left(\frac{2}{D(2\pi)^d} \frac{|\hat{f}(\boldsymbol{\omega})|}{q_{\Omega}(\boldsymbol{\omega})} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b) \right)^2 \\ & \quad \cos(\boldsymbol{\omega}^T \mathbf{x} + b)^2 q_{\Omega}(\boldsymbol{\omega}) db d\boldsymbol{\omega} - f(\mathbf{x})^2. \end{aligned} \quad (4.69)$$

For the stated choice of p_{Ω}^* , using

$$\begin{aligned} & \int_0^{2\pi} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b)^2 \cos(\boldsymbol{\omega}^T \mathbf{x} + b)^2 db \\ &= \int_0^{2\pi} \frac{1}{4} (1 + \cos(2\angle \hat{f}(\boldsymbol{\omega}) - 2b))(1 + \cos(2\boldsymbol{\omega}^T \mathbf{x} + 2b)) db \\ &= \int_0^{2\pi} \frac{1}{4} db + \frac{1}{4} \int_0^{2\pi} \cos(2\angle \hat{f}(\boldsymbol{\omega}) - 2b) db \\ & \quad + \frac{1}{4} \int_0^{2\pi} \cos(2\boldsymbol{\omega}^T \mathbf{x} + 2b) db \\ & \quad + \frac{1}{4} \int_0^{2\pi} \cos(2\angle \hat{f}(\boldsymbol{\omega}) - 2b) \cos(2\boldsymbol{\omega}^T \mathbf{x} + 2b) db \\ &= \frac{2\pi}{4} + \frac{1}{8} \int_0^{2\pi} \cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) \\ & \quad + \cos(2\angle \hat{f}(\boldsymbol{\omega}) - 2\boldsymbol{\omega}^T \mathbf{x} - 4b) db \\ &= \frac{2\pi}{4} + \frac{2\pi}{8} \cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) \\ &= \frac{\pi}{4} (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) \end{aligned} \quad (4.70)$$

we get:

$$\text{Var}_{p_{\Omega}^*, p_B}[G(\mathbf{x})] + f(\mathbf{x})^2 = \mathbb{E}_{p_{\Omega}^*, p_B}[G(\mathbf{x})^2]$$

$$\begin{aligned}
&= \frac{D}{2\pi} \int_{\mathbb{R}^d} \int_0^{2\pi} \left(\frac{2}{D(2\pi)^d} \frac{|\hat{f}(\boldsymbol{\omega})|}{p_{\boldsymbol{\Omega}}^*(\boldsymbol{\omega})} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b) \right)^2 \\
&\quad \cos(\boldsymbol{\omega}^T \mathbf{x} + b)^2 p_{\boldsymbol{\Omega}}^*(\boldsymbol{\omega}) db d\boldsymbol{\omega} \\
&= \frac{D}{2\pi} \int_{\mathbb{R}^d} \frac{1}{p_{\boldsymbol{\Omega}}^*(\boldsymbol{\omega})} \left(\frac{2}{D(2\pi)^d} \right)^2 |\hat{f}(\boldsymbol{\omega})|^2 \\
&\quad \int_0^{2\pi} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b)^2 \cos(\boldsymbol{\omega}^T \mathbf{x} + b)^2 db d\boldsymbol{\omega} \\
&= \frac{D}{2\pi} \int_{\mathbb{R}^d} \frac{1}{p_{\boldsymbol{\Omega}}^*(\boldsymbol{\omega})} \left(\frac{2}{D(2\pi)^d} \right)^2 |\hat{f}(\boldsymbol{\omega})|^2 \\
&\quad \frac{\pi}{4} (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) d\boldsymbol{\omega} \tag{4.71}
\end{aligned}$$

$$\begin{aligned}
&\stackrel{(4.67)}{=} \frac{D}{2\pi} \left(\frac{2}{D(2\pi)^d} \right)^2 \\
&\quad \left(\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| \sqrt{\frac{\pi}{4} (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2)} d\boldsymbol{\omega} \right)^2 \\
&= \frac{1}{2D(2\pi)^{2d}} \left(\int_{\mathbb{R}^d} |\hat{f}(\boldsymbol{\omega})| \sqrt{(\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2)} d\boldsymbol{\omega} \right)^2 \tag{4.72}
\end{aligned}$$

This gives the value of the optimal variance. To show that the variance is indeed optimal, compare it with any arbitrary p.d.f. $q_{\boldsymbol{\Omega}}$ using Jensen's inequality:

$$\begin{aligned}
&\text{Var}_{p_{\boldsymbol{\Omega}}^*, p_B} [G(\mathbf{x})] + f(\mathbf{x})^2 \\
&= \frac{D}{2\pi} \left(\frac{2}{D(2\pi)^d} \right)^2 \\
&\quad \left(\int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|}{q_{\boldsymbol{\Omega}}(\boldsymbol{\omega})} \sqrt{\frac{\pi}{4} (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2)} q_{\boldsymbol{\Omega}}(\boldsymbol{\omega}) d\boldsymbol{\omega} \right)^2 \\
&\stackrel{\text{Jensen}}{\leq} \frac{D}{2\pi} \left(\frac{2}{D(2\pi)^d} \right)^2 \\
&\quad \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{q_{\boldsymbol{\Omega}}(\boldsymbol{\omega})^2} \frac{\pi}{4} (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) q_{\boldsymbol{\Omega}}(\boldsymbol{\omega}) d\boldsymbol{\omega} \\
&\stackrel{(4.70)}{=} \frac{D}{2\pi} \int_{\mathbb{R}^d} \int_0^{2\pi} \left(\frac{2}{D(2\pi)^d} \frac{|\hat{f}(\boldsymbol{\omega})|}{q_{\boldsymbol{\Omega}}(\boldsymbol{\omega})} \cos(\angle \hat{f}(\boldsymbol{\omega}) - b) \right)^2 \\
&\quad \cos(\boldsymbol{\omega}^T \mathbf{x} + b)^2 q_{\boldsymbol{\Omega}}(\boldsymbol{\omega}) db d\boldsymbol{\omega} \\
&\stackrel{(4.69)}{=} \text{Var}_{q_{\boldsymbol{\Omega}}, p_B} [G(\mathbf{x})] + f(\mathbf{x})^2. \tag{4.73}
\end{aligned}$$

This shows that the chosen p.d.f. $p_{\boldsymbol{\Omega}}^*$ gives the minimum variance. \square

The following theorem compares the second moments in real and complex RFEs for different probability distributions.

Theorem 8. Let \tilde{p}_Ω , p_Ω^* , \tilde{G} and G be as in Theorems 5 and 7. Then

$$\frac{1}{\sqrt{3}} \mathbb{E}_{p_\Omega^*, p_B} [G(\mathbf{x})^2] \leq \mathbb{E}_{\tilde{p}_\Omega, p_B} [G(\mathbf{x})^2] \leq \sqrt{3} \mathbb{E}_{p_\Omega^*, p_B} [G(\mathbf{x})^2], \quad (4.74)$$

$$\frac{1}{2} \mathbb{E}_{\tilde{p}_\Omega, p_B} [\tilde{G}(\mathbf{x})^2] \leq \mathbb{E}_{\tilde{p}_\Omega, p_B} [G(\mathbf{x})^2] \leq \frac{3}{2} \mathbb{E}_{\tilde{p}_\Omega, p_B} [\tilde{G}(\mathbf{x})^2]. \quad (4.75)$$

Proof. From

$$1 \leq \sqrt{(\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2)} \leq \sqrt{3}, \quad (4.76)$$

and from (4.67) and (4.33) it follows that

$$\begin{aligned} \frac{1}{\sqrt{3}} p_\Omega^*(\boldsymbol{\omega}) &\leq \tilde{p}_\Omega(\boldsymbol{\omega}) \leq \sqrt{3} p_\Omega^*(\boldsymbol{\omega}), \\ \frac{1}{\sqrt{3}} \frac{1}{p_\Omega^*(\boldsymbol{\omega})} &\leq \frac{1}{\tilde{p}_\Omega(\boldsymbol{\omega})} \leq \sqrt{3} \frac{1}{p_\Omega^*(\boldsymbol{\omega})}. \end{aligned} \quad (4.77)$$

Combining the above with (4.71) yields:

$$\begin{aligned} &\frac{1}{\sqrt{3}} \mathbb{E}_{p_\Omega^*, p_B} [G(\mathbf{x})^2] \\ &= \frac{1}{\sqrt{3}} \frac{1}{2D(2\pi)^{2d}} \\ &\quad \int_{\mathbb{R}^d} \frac{1}{p_\Omega^*(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) d\boldsymbol{\omega} \\ &\leq \frac{1}{2D(2\pi)^{2d}} \\ &\quad \int_{\mathbb{R}^d} \frac{1}{\tilde{p}_\Omega(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) d\boldsymbol{\omega} \\ &= \mathbb{E}_{\tilde{p}_\Omega, p_B} [G(\mathbf{x})^2] \\ &\leq \sqrt{3} \frac{1}{2D(2\pi)^{2d}} \\ &\quad \int_{\mathbb{R}^d} \frac{1}{p_\Omega^*(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) d\boldsymbol{\omega} \\ &= \sqrt{3} \mathbb{E}_{p_\Omega^*, p_B} [G(\mathbf{x})]. \end{aligned} \quad (4.78)$$

Combining (4.76) with (4.34) yields:

$$\begin{aligned} &\frac{1}{2} \mathbb{E}_{\tilde{p}_\Omega} [\tilde{G}(\mathbf{x})^2] \\ &= \frac{1}{2D(2\pi)^{2d}} \int_{\mathbb{R}^d} \frac{1}{\tilde{p}_\Omega(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\ &\leq \frac{1}{2D(2\pi)^{2d}} \int_{\mathbb{R}^d} \frac{1}{\tilde{p}_\Omega(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 \\ &\quad (\cos(2\angle \hat{f}(\boldsymbol{\omega}) + 2\boldsymbol{\omega}^T \mathbf{x}) + 2) d\boldsymbol{\omega} \end{aligned}$$

$$\begin{aligned} &= \mathbb{E}_{\tilde{p}_{\Omega}, p_B} [G(\mathbf{x})^2] \\ &\leq \frac{3}{2D(2\pi)^{2d}} \int_{\mathbb{R}^d} \frac{1}{\tilde{p}_{\Omega}(\boldsymbol{\omega})} |\hat{f}(\boldsymbol{\omega})|^2 d\boldsymbol{\omega} \\ &= \frac{3}{2} \mathbb{E}_{\tilde{p}_{\Omega}} [\tilde{G}(\mathbf{x})^2]. \end{aligned} \tag{4.79}$$

□

5

WAVEFRONT SENSORLESS ADAPTIVE OPTICS OCT WITH THE DONE ALGORITHM FOR *in vivo* HUMAN RETINAL IMAGING

In this report, which is an international collaboration of OCT, adaptive optics, and control research, we demonstrate the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm to guide the image based optimization for wavefront sensorless (WFSL) AO OCT for in vivo human retinal imaging. The ocular aberrations were corrected using a multi-actuator adaptive lens after linearization of the hysteresis in the piezoelectric actuators. The DONE algorithm succeeded in drastically improving the OCT signal intensity and image quality, while achieving a computational time of 1 ms per iteration, making it applicable for many high speed applications. Data acquired from an imaging phantom and in vivo from human research volunteers are presented.

Parts of this chapter have been published in [219].
The visualizations in this chapter will be made available by the publisher of the article upon publication.
Before publication, please use the following URL:
<https://1drv.ms/f/s!AneHoxPZdGw0nEBQYfL80d72Kjpe>

5.1. INTRODUCTION

Since the inception of optical coherence tomography (OCT) in 1991 [2], it has contributed to significant advancements in clinical ophthalmic imaging. A particular strength of OCT is the ability to visualize the cross-sectional thickness of the retina, and the various cell layers that are organized by function. With the axial resolution dependent on the coherence length, but decoupled from the optics delivering light to the eye, commonly available OCT systems were designed to have a depth of focus that encapsulated the entire retinal thickness. This design goal is commonly achieved using of a probe beam diameter of ~ 1 mm incident on the cornea, resulting in a focused spot size of ~ 20 μm at the retina. In order to reveal the cellular structures of the retina, such as the cone photoreceptor mosaic and the nerve fibre bundles, the resolution of the imaging system needs to be increased. Given the fixed focal length of a representative eye, the focal waist can be reduced by imaging with a larger beam at the pupil. In the special case of healthy volunteers with good eye optics, general OCT imaging systems are capable of imaging parafoveal or perifoveal photoreceptor cones [220–222]. However, the image reliability and quality of the cone photoreceptor images deteriorate when imaging with a large incident beam because of wavefront aberrations present in the refractive elements of the eye.

For the majority of eyes, Adaptive Optics (AO) is essential to maximize the image quality for *in vivo* optical retinal imaging with a large pupil [223]. In particular, this is true for resolving the cone mosaic close to the fovea, where the cone photoreceptor density increases, as the diameter of the cones decrease to ~ 2 μm in the center of the fovea. Adaptive optics OCT has been reported to improve the quality of *in vivo* retinal images to such an extent that single cone photoreceptors and individual nerve fiber bundles are clearly resolved [96, 97, 99, 224–227]. Conventional AO-OCT imaging systems use a Shack-Hartmann (SH) wavefront sensor to directly quantify and reconstruct the ocular wavefront aberrations. An active component such as a deformable mirror or lens is then used to remove these aberrations. The image quality obtained with these systems are excellent. However, AO-OCT systems containing wavefront sensors are susceptible to wavefront reconstruction errors. If back-reflections and non-common path errors are not properly taken into account, they can lead to poor performance in the wavefront reconstruction [228]. Limitations of the SH wavefront sensor based AO systems occur in cases where subjects have irregularly shaped pupils, cloudy corneas, or other ocular opacities that scatter light back to the SH wavefront sensor, obscuring the detection of the ocular aberrations. Additional reasons for removing the wavefront sensor out of an AO system are to reduce the size, complexity, and cost of the AO system. These drawbacks of the SH wavefront sensor have led to the development of wavefront sensorless AO (WFSL-AO) algorithms.

In place of a direct measurement of the optical aberrations, WFSL-AO algorithms attempt to remove the optical aberrations by optimizing an image quality metric based on a sequence of images acquired with different test aberrations applied to the adaptive element [152, 169, 170, 229, 230]. Implementations of WFSL-AO for retinal imaging with scanning laser ophthalmoscopy (SLO) have been demonstrated using metrics determined by the brightness or sharpness of an image [231, 232]. WFSL-AO OCT was first demonstrated with a simulated annealing optimization algorithm [145]. A coordinate

search (CS) algorithm was also demonstrated to improve the OCT signal [144, 156]. More recently, *in vivo* WFSL-AO OCT has been demonstrated for retinal imaging of mice [88, 158, 233] and humans [234, 235], also using a CS algorithm.

A limitation of the CS algorithm for *in vivo* imaging is that it is susceptible to noise (i.e. motion artifacts) in the data used for the quality metric during the optimization process. In mice and especially in human subjects, involuntary eye movements such as microsaccades, tremor, and drift can significantly degrade the OCT volumes by causing motion artifacts. If the noise artifact causes the CS algorithm to select an improper coefficient for a particular Zernike mode, the algorithm cannot recover. For robust *in vivo* imaging with WFSL-AO, an image quality optimization algorithm that accounts for motion artifacts is essential.

The Data-based Online Nonlinear Extremum-seeker (DONE) algorithm was first described for WFSL-AO in OCT [172]. In contrast to the aforementioned algorithms that take the measurement with the lowest (or highest) metric value, DONE was explicitly designed to take all past measurements into account such that the robustness of the algorithm with respect to noise is increased. Figure 5.1 shows an example of optimization performed with the DONE algorithm generalized to a one dimensional case. Instead of

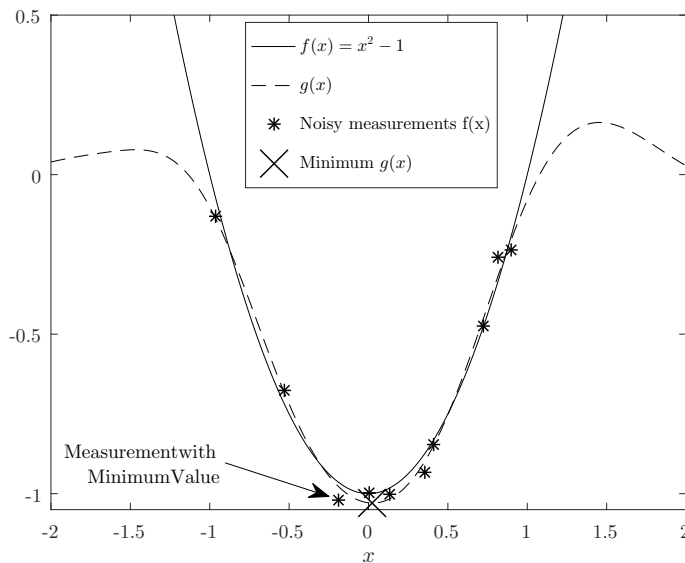


Figure 5.1: Example of optimization with DONE. The unknown function $f(x)$ is approximated by the random cosine model $g(x)$ with ten noisy measurements. The minimum of $g(x)$ is found and approximates the minimum of $f(x)$.

assuming a convex merit function (solid line), the DONE algorithm fits a random cosine model to the measurements (dashed line). With every new metric evaluation, the DONE algorithm updates the random cosine model of the merit function, improving its robustness with respect to noisy measurements.

The purpose of this report is to demonstrate the DONE algorithm for *in vivo* retinal imaging in humans with WFSL-AO OCT. Our choice of wavefront correcting element was the multi-actuator adaptive lens (MAL), described and demonstrated for mouse retinal imaging in [88]. An advantage of a transmissive adaptive lens is that it can be more readily integrated with a wide range of existing imaging systems by placing it in existing pupil planes, unlike a conventional deformable mirror, which requires optical setups to relay the pupil plane with folded optical paths [88]. The MAL was recently demonstrated for WFSL-AO OCT for human retinal imaging in a compact system using a CS algorithm [235]. However, the CS algorithm was sensitive to motion during the optimization process. Furthermore, the piezoelectric actuators in the MAL suffer from hysteresis, which was compensated by adding additional steps into the CS optimization algorithm in [235], increasing the time required for optimization.

This paper demonstrates that the DONE algorithm can be used for WFSL-AO OCT for *in vivo* human retinal imaging using an adaptive lens as the wavefront corrector. When performing *in vivo* imaging, high speeds are imperative for both image acquisition, as well as computation of the optimization algorithm. To meet the necessity of high speed imaging, we implemented a high speed DONE routine to calculate the random cosine model between the successive OCT volume acquisitions of the optimization process. In order to more accurately determine the wavefront aberration introduced by the lens, the hysteresis effect of the actuators was characterized, and then suppressed during *in vivo* optimization to increase the accuracy of the MAL without the presence of a wavefront sensor for feedback. We present WFSL-AO OCT images acquired in human subjects, that are not trained in visual fixation, and demonstrate successful aberration correction even in the presence of motion artifact.

5.2. MATERIALS AND METHODS

The measurements were performed with a compact clinical WFSL-AO OCT system described in [235] which contained two deformable transmissive optical elements. A double buffered 200 kHz effective line rate swept source laser was used as the imaging light source (1060 nm center wavelength, 80 nm FWHM bandwidth, Axsun, Inc.). The results of a Zemax simulation of the lens-based sample arm, shown in Fig. 5.2, indicated that the optical design was diffraction limited over a 1.5° scan range. The wavefront aberrations of the optical configuration were experimentally measured to be ~ 0.069 nm RMS, which is below the Maréchal criterion for diffraction limited imaging (wavelength divided by 14). The first deformable lens was a MAL with 18 actuators described in more detail in [88], placed adjacent to the collimator. The MAL was used to adjust fine focus, the vertical and oblique astigmatisms, and either two comas or two trefoils. The pupil plane was related to the second deformable lens, which was a variable focal length lens (Varioptics ARCTIC 316-AR850) and was used to manually correct the bulk of the defocus in the human eye. The beam size on the subject's pupil had a Gaussian profile with a $1/e^2$ diameter of 4.8 mm. The calculated $1/e^2$ waist was ~ 2.4 μm based on Gaussian beam calculations using the Gullstrand-LeGrand model of the human eye [236,237]. Real-time processing of the OCT data was done on a GPU as described in [238].

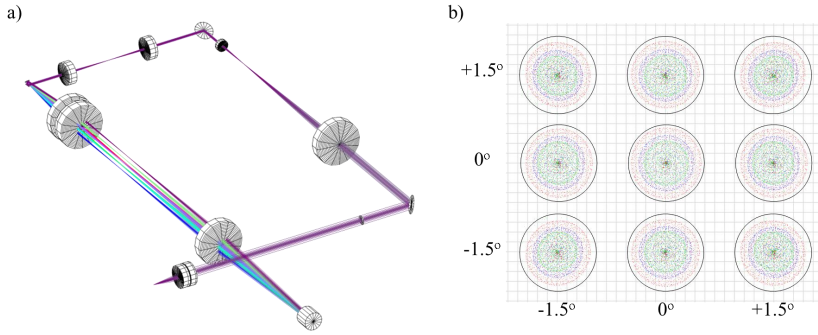


Figure 5.2: (a) Zemax 3D simulation of the optical aberrations of the lens-based sample arm. (b) Spot diagram for three wavelengths spanning the 80 nm bandwidth of the light source (green, 1020 nm; blue, 1060 nm; red, 1100 nm). The eye was modeled as a paraxial lens with 16 mm focal length in air.

5.2.1. HYSTERESIS CORRECTION OF THE MULTI-ACTUATOR ADAPTIVE LENS

The MAL has 18 piezoelectric actuators that suffer from hysteresis. The hysteresis error of a single actuator is approximately 14%. In order to achieve the maximum possible accuracy in depicting optical wavefront aberrations with the MAL without using a wavefront sensor, the hysteresis was characterized and compensated. Several methods exist to correct hysteresis in deformable mirrors, for example [239–241]. This information was used to linearize the response of the piezoelectric actuators in terms of the voltage input and displacement. The method used for hysteresis correction directly approximated the inverse hysteresis curve with a combination of a polynomial and a Prandtl-Ishlinskii (PI) model [242]. The model used was formed by the discrete-time elementary backlash operator and a polynomial:

$$H_r[S(\bar{u})](t) = \max\{S(\bar{u}(t)) - r, \min\{S(\bar{u}(t)) + r, H_r[S(\bar{u})](t - T)\}\}, \quad (5.1)$$

where $\bar{u}(t)$ is the input, r is the threshold, a sampling time T is used, and the polynomial $S(\bar{u}(t)) = c_m \bar{u}^m(t) + c_{m-1} \bar{u}^{m-1}(t) + \dots + c_1 \bar{u}^1(t)$. Assuming that the piezoelectric actuators started from the de-energized state, the initial condition $H_r[S(\bar{u})](0)$ was set to zero. The full model was then expressed as

$$\bar{\phi}^{-1}(\bar{u}(t)) = \sum_{i=1}^n w_i H_{r_i}[S(\bar{u})](t). \quad (5.2)$$

The weights w_i were fitted to match the model to the inverse hysteresis curve. The number of backlash operators was set to $n = 40$, the order of the polynomial was set to $m = 7$, and the thresholds $r_i = 0.02(i - 1)$. We fitted the hysteresis according to the following procedure. A Shack-Hartmann (SH) wavefront sensor was placed at a conjugate plane of the MAL such that the influence matrix of the lens could be measured. Each of the 18 piezoelectric actuators of the MAL were then sequentially actuated and the change

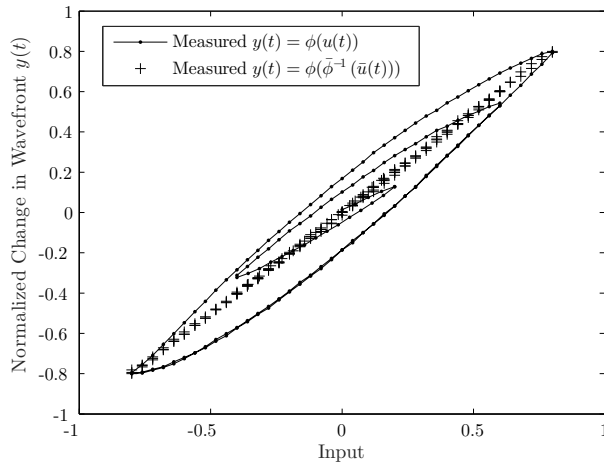


Figure 5.3: Hysteresis curve of piezoelectric actuator measured by Shack-Hartmann wavefront sensor before and after linearization.

5

in wavefront error was recorded by the SH sensor. The procedure was repeated separately for each actuator. Figure 5.3 shows a measured hysteresis curve identifying the relationship $y(t) = \phi(u(t))$ between the input $u(t)$ of one actuator of the MAL and the normalized change in the wavefront error $y(t)$. The inverse ϕ^{-1} of this measured hysteresis curve was approximated by $\bar{\phi}^{-1}$, a PI model combined with a polynomial fit. Assuming that the real input of the actuator was set to $u(t) = \bar{\phi}^{-1}(\bar{u}(t))$, then the model of the inverse hysteresis function was used to derive an input for the actuator that resulted in a linear relationship between the wavefront error and the linear desired input of the actuator $\bar{u}(t)$, $y(t) = \phi(\bar{\phi}^{-1}(\bar{u}(t))) \approx \bar{u}(t)$. The linear realization can be seen in Fig. 5.3. Using our approach the maximum hysteresis error after the linearization was below 2%. The procedure was repeated for all actuators. After linearization, the response of all actuators of the lens was considered to be linear. This resulted in an influence matrix that is valid over the entire stroke of the MAL and turns the generation of arbitrary wavefront aberrations into a linear problem [29]. The normalized Zernike coefficients depicted on the MAL can be calculated as $\mathbf{x} = \mathbf{M}\bar{\mathbf{u}}$, where \mathbf{M} is derived from the influence matrix and $\bar{\mathbf{u}} = [\bar{u}_1, \bar{u}_2, \dots, \bar{u}_{18}]^T$ is a vector of the desired linear inputs of all 18 actuators. The calibration for the linearization of the actuators was performed once and remained stable for all of the following imaging experiments.

5.2.2. THE DONE ALGORITHM

For the application of WFSL-AO OCT to *in vivo* human retinal imaging, minimization of the optimization time is essential to reduce the effects of motion artifact on the aberration correction process. In order to perform the DONE model computations in near real-time, fast algorithms are necessary. We present an improved and faster implementation of the DONE algorithm [172, 178] used to maximize the AO-OCT signal for human retinal imaging. The biggest difference between the DONE algorithm reported in [172] and

the current work is a faster implementation that allows the exploration and the bounds for each aberration to be set independently. Essentially, DONE maintains a model based on D random cosines of which the amplitudes are determined by a linear least squares problem. With the least squares solver implementation reported in [172], the computational time per iteration of the DONE algorithm was ~ 60 ms. For 70 iterations, the time for computation alone would exceed 4 s. Hence, we changed the least squares routine of the DONE algorithm into a recursive procedure, which resulted in a significant increase in computational speed. With the implementation of the recursive routine, the time per iteration of the DONE algorithm was reduced to ~ 1 ms, which is smaller than the settling time of the MAL.

In iteration n of DONE, we define $\mathbf{a}_n = \cos(\boldsymbol{\omega}^T \mathbf{x}_n + \mathbf{B})$, where the vector \mathbf{x}_n contains the coefficients of the Zernike aberrations of the MAL and y_n is the corresponding OCT signal metric value. Moreover, the matrix $\boldsymbol{\omega}$ consists of d by D independent and identically distributed (i.i.d) random frequencies drawn from a normal distribution with zero mean and standard deviation σ and the column vector \mathbf{B} consists of D i.i.d phase offsets with values between 0 and 2π drawn from a uniform distribution. Also, the cos function operates element-wise on a vector. We implemented the inverse QR algorithm [209, Sec. 21] for the recursive update, which is known to be especially numerically reliable. The initial amplitudes of the cosines are set to $\mathbf{c}_0 = 0$ and the initial matrix $\mathbf{P}_0 = \lambda^{-1} \mathbf{I}_{D \times D}$, where λ is the regularization parameter and $\mathbf{I}_{D \times D}$ is an identity matrix with D columns. We find a rotation matrix $\boldsymbol{\Theta}_n$ that lower triangularizes the upper triangular matrix in Eq. (5.3) below and generates a post-array with positive diagonal entries:

$$\begin{bmatrix} 1 & \mathbf{a}_n^T \mathbf{P}_{n-1}^{1/2} \\ \mathbf{0} & \mathbf{P}_{n-1}^{1/2} \end{bmatrix} \boldsymbol{\Theta}_n = \begin{bmatrix} \gamma_n^{-1/2} & \mathbf{0} \\ \mathbf{g}_n \gamma_n^{-1/2} & \mathbf{P}_n^{1/2} \end{bmatrix}. \quad (5.3)$$

The rotation matrix $\boldsymbol{\Theta}_n$ can be found by performing a QR decomposition of the transpose of the matrix on the left hand side of (5.3), or by the procedure explained in [209, Sec. 21]. Then we update the amplitudes of the cosines \mathbf{c}_n as follows,

$$\mathbf{c}_n = \mathbf{c}_{n-1} + \mathbf{g}_n (y_n - \mathbf{a}_n^T \mathbf{c}_{n-1}). \quad (5.4)$$

One OCT metric evaluation takes place per iteration and is used to update the model. The model is then defined as $g(\mathbf{x}) = \mathbf{c}_n^T \cos(\boldsymbol{\omega}^T \mathbf{x} + \mathbf{B})$. After the model update, the minimum or maximum of the model $g(\mathbf{x})$ is found. At the end of each iteration the DONE algorithm proposes new values of the Zernike coefficients \mathbf{x}_{n+1} to add on the MAL and the next measurement for the OCT signal metric y_{n+1} is taken. This process is repeated until the algorithm has converged up to a pre-defined maximum number of iterations.

DONE can deal with arbitrary aberrations and is mainly limited by the correction capabilities of the MAL. DONE does not have a predetermined search pattern and will change the position of the measurements based on past metric evaluations. Therefore, the final accuracy of the aberration correction of the DONE algorithm is not limited by the step size as for example compared to the CS algorithm [172]. To improve the accuracy of the lens, hysteresis compensation of the actuators in the lens was necessary. Also, by choosing a small standard deviation for the frequencies of the cosines, the model will serve as a low-pass filter for the metric evaluations, while other algorithms often take the argument of the highest or lowest value of the evaluated metric, which can yield an

outlier due to noise as shown in Fig. 5.1. In this way, DONE mitigates the noise effects of motion on the merit function during the optimization.

5.2.3. HUMAN IMAGING AND ABERRATION CORRECTION

The performance of DONE in combination with the MAL was investigated by imaging the retinal photoreceptor layer of 10 healthy volunteers with refractive errors of less than 4 diopters. The mean age of the volunteers was 28.3 ± 7.6 years, consisting of 2 females and 8 males. The mean axial length and cylindrical refractive error of the research volunteers (measured with an IOL Master 500) was 24.19 ± 0.94 mm and 1.042 ± 0.52 D respectively. Human retinal imaging was performed in accordance with the research ethics approved by the Office for Research Ethics (ORE) at Simon Fraser University, the University of British Columbia, and Vancouver General Hospital. Written and informed consent was obtained prior to imaging from all imaging subjects. The average power of light incident on the cornea during imaging was limited to $900 \mu\text{W}$. Imaging was initiated with the eye dilated, and the subject seated comfortably with their head supported by a chin and forehead rest. An additional benefit of the pupil dilation was that it caused temporary paralysis of the eye's ciliary muscles, preventing accommodation. Prior to the optimization, the subject's eye was aligned to the imaging system and an OCT volume of 400 by 400 A-scans was acquired (pre-optimization). For the optimization procedure, small volumes (C-scans) consisting of 8,000 A-scans (400×20) were acquired at a 200 kHz line rate, corresponding to a volume acquisition rate of ~ 25 volumes per second. *En face* images of the operated selected retinal layer were extracted from the OCT volume in real-time during the optimization. The brightness of the *en face* OCT images was used as the signal metric for the DONE algorithm [156]. The axial position of the retinal layers in the OCT B-scan was dynamically tracked by the acquisition software [243] to account for the axial motion of the subjects [155] and ensure the OCT signal metric was calculated based on the same retinal layer throughout the optimization process. Different retinal layers could be chosen to perform the optimization, as shown in [235]. In this work we tracked and imaged the photoreceptors in the outer retina. The DONE algorithm was configured to take 70 metric evaluations for the optimization, which corresponded to a total time for the optimization (including actuation of the lens and processing of the small C-scans) of ~ 3 seconds. Immediately after the optimization a large volume of 400 by 400 A-scans was acquired for the final image (post-optimization).

For human eye imaging, the MAL was used to correct $d = 5$ Zernike aberrations, namely, the defocus, two astigmatisms, and two comas. The empirically determined parameters used for the DONE algorithm optimization in the human eye are described in Table 5.1. Setting the number of cosines to $D = 1,000$ provided good results for at least up to seven aberrations [172]. The standard deviation of the frequencies of the cosines was $\sigma = 1$, which was set to match the frequency content in the transfer function of the aberration [156]. The regularization parameter $\lambda = 5$ prevented over-fitting of the model. The upper and lower bound vectors \mathbf{ub} and \mathbf{lb} , respectively, consisted of the individual bounds for each aberration and matched the maximum capabilities of the lens with hysteresis correction. The vectors σ_η and σ_ξ determined the amount of exploration done by the DONE algorithm for each aberration. The exploration parameters define a trade-off between the number of steps used exploring and the final accuracy, a larger value could

result in faster convergence, while a smaller value could lead to a smaller wavefront error. The exploration factor for the comas was set lower than for the other aberrations because the upper bounds of the comas are lower.

Table 5.1: Parameter values for the DONE algorithm (wavefront aberrations are defined in μm)

DONE	
D	1000
σ	1
λ	5
σ_η, σ_ξ	[0.07, 0.07, 0.07, 0.025, 0.025]
ub	[1.8, 2, 2, 0.4, 0.4]
lb	− ub

5.3. RESULTS

The imaging performance of the DONE algorithm was compared against the hill climbing CS algorithm used in previous work reported using the MAL [158, 235]. In order to provide a comparison between the two optimization techniques, we imaged a tissue phantom with no motion artifact. The DONE algorithm used for phantom imaging was the same as that described for human imaging in the previous section. The CS algorithm was implemented as in the previous work using the MAL [158, 235], using a static look-up-table for the actuator voltages to generate a specific amplitude of a particular Zernike mode. For each Zernike mode, 10 OCT volumes were acquired with the aberration applied to the MAL using coefficient values that were uniformly distributed across the same range of upper and lower bounds (**ub** and **lb**) as used for the DONE algorithm optimization. The coefficient resulting in the highest value of the merit function was selected as the optimized value for that Zernike mode, and applied to the MAL. Subsequent Zernike modes used the optimized values of the preceding modes as the starting point. Because the CS algorithm did not take into account the hysteresis of the lens, the actuators were de-energized by applying a decreasing sinusoid to each of the actuators; this ‘relax’ procedure required ~ 100 ms. The CS was performed by relaxing the MAL, applying the previously found optimized values, and then stepping through look-up-table of coefficients. For each Zernike mode, the procedure first searched the positive coefficient values, relaxed the MAL actuators, and then searched the negative coefficient values. A Shack-Hartmann wavefront sensor optically conjugated to the MAL was used to measure the interaction of the Zernike modes while running the CS algorithm; the measurements are presented in [Visualization 1](#). The video results indicate that the Zernike modes have been generated by the MAL with minimal cross-talk to the other modes with the exception of coma, and even in that case the effects on the reconstructed wavefront map were minimal. The results of the DONE and CS optimization on a phantom (lens paper fibres) are presented in Fig. 5.4. The sequences of *en face* images acquired during the optimization process for CS and DONE are presented in [Visualization 2](#).

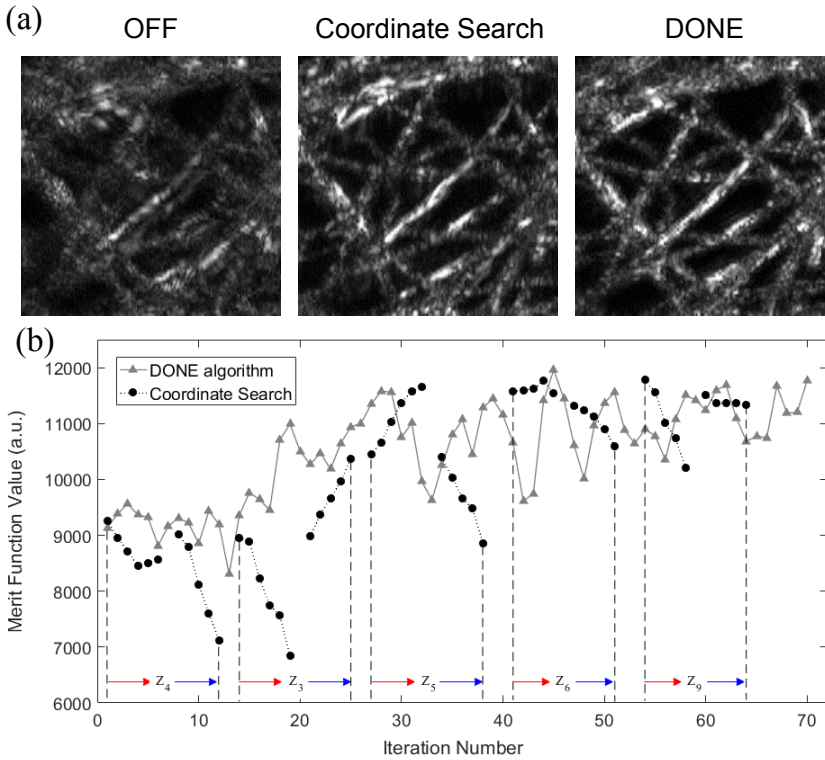


Figure 5.4: Comparison of Coordinate Search (CS) and DONE optimization on a stationary sample. (a) Unoptimized and final images of the phantom after aberration correction with the CS and DONE algorithms. (b) Merit function recorded as a function of iteration for CS and DONE.

5.3.1. IMAGING HUMAN PHOTORECEPTORS

Images of human retina acquired from the research subjects are presented before and after optimization with the DONE algorithm. The images were processed identically. In Fig. 5.5, a $450\ \mu\text{m}$ by $450\ \mu\text{m}$ *en face* image shows the retinal cones of a healthy subject centered at approximately 3 degrees from the center of the fovea. The *en face* image was

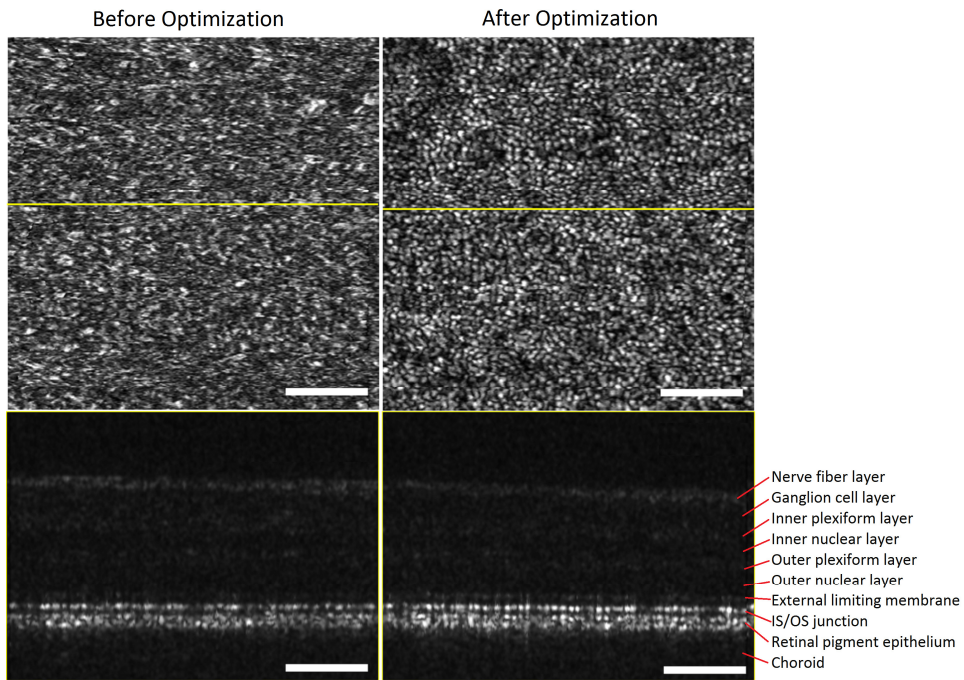


Figure 5.5: The top row shows the OCT *en face* image of human photoreceptors before and after optimization using 70 iterations of DONE. The bottom row shows the corresponding B-scans at the location shown in yellow on the *en face* image. All scale bars are $100\ \mu\text{m}$.

extracted from 400 by 400 A-scans. In the pre-optimization image, the individual cones are not resolvable. The contrast and structure of the cones is improved after the optimization; a mosaic pattern can be visualized, and the cones are readily distinguished. In Fig. 5.6, the OCT image metric values are plotted against the iteration number of the DONE algorithm for the optimization used in Fig. 5.5. The fluctuation of the metric is caused by noise, changing aberrations on the MAL and small movements of the subject's eye. The progress of the metric function during the optimization shows a large increase after the optimization. The WFSL-AO OCT signal was maximized at the IS-OS junction layer, which becomes approximately seven times higher after the optimization (from an initial value of 553 to a final value of 3761). The total root mean square (RMS) wavefront error of the corrected wavefront aberrations was $1.22\ \mu\text{m}$ with an oblique astigmatism that had a RMS wavefront error of $0.53\ \mu\text{m}$. The WFSL-AO optimized images from another representative research subject are presented in [Visualization 3](#). Although the focus was optimized at the photoreceptor layer, the full thickness of the retina can be vi-

sualized in the OCT volume (post DONE optimization, single acquisition, not averaged).

The images presented in the previous figures were acquired on research subjects that had relatively good fixation. Figure 5.7 shows the optimization of the OCT signal in a different subject, that is more representative of average to below-average fixation ability. This data demonstrates DONE's ability to function when the subject is not able to maintain a fixed gaze, as suggested from the large motion artifacts that are present both before and after optimization. These motion artifacts, which appear as discontinuities in the *en face* images, are caused by involuntary movements of the subject's eye and are taken into account by the DONE algorithm as noise. A similar improvement in image quality was observed despite the presence of the motion artifacts. Here, the total RMS wavefront error of the corrected wavefront aberrations was $0.58 \mu\text{m}$, including a vertical astigmatism with a RMS wavefront error of $-0.49 \mu\text{m}$.

We investigated the utility of the DONE algorithm optimization for imaging wider regions in the retina. In Figure 5.8, 12 WFSL-OCT volumes were acquired and manually aligned to generate a montage image. The different regions were acquired by asking the research subject to change the fixation to calibrated points in the field of view. The four regions immediately surrounding the central square are presented on a larger scale, demonstrating that the cone photoreceptor mosaics are clearly resolved. Since the size of the cone photoreceptors increases and the cone density decreases at larger eccentricities, the remaining images are also readily resolved.

5.4. DISCUSSION

We demonstrated WFSL-AO OCT aberration correction using the DONE algorithm and a multi-actuator adaptive lens (MAL) for human retinal imaging *in vivo*. This work represents the results of an international collaboration, and relative to our previous works, a significant advance towards WFSL-AO OCT imaging with a MAL in a clinical setting. In terms of the system hardware, the non-linearity and hysteresis of the MAL was removed. Algorithmically, an improved version of the DONE algorithm was developed and implemented to minimize the computational time required in between data points. In combination, the linearized e lens and the high-speed DONE algorithm were used to maximize the WFSL-AO OCT signal from the human retina in different subjects. The linearization of the hysteresis did not need to be repeated for the measurements of the ten subjects. Physical changes in the lens may cause a slow variation of the hysteresis fit, however, our results demonstrated that they remained valid during the entire period of data acquisition of all subjects. Similar to the linearization of the MAL actuators, the tuning of the parameters of the DONE algorithm was done beforehand.

The results acquired on the tissue phantom in Fig. 5.4 demonstrated that the DONE algorithm and the CS algorithm arrived to very similar results for the optimization of a static sample. The CS algorithm did not have hysteresis compensation, and required a 'relax' procedure, which added significantly to the optimization time. A key difference between the CS and DONE algorithms is apparent in the graph of merit function versus iteration number in Fig. 5.4 (b). The merit function measurements for the CS algorithm showed significant variations when scanning through each Zernike mode. A small motion artifact affecting the CS early in the optimization process could have disastrous results on the overall result, causing the selection of an incorrect value for a Zernike mode

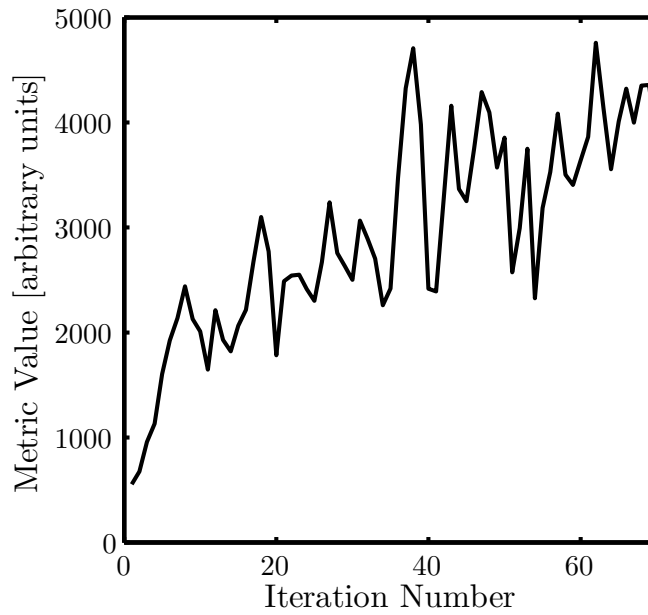


Figure 5.6: Values of the metric function during the optimization with the DONE algorithm.

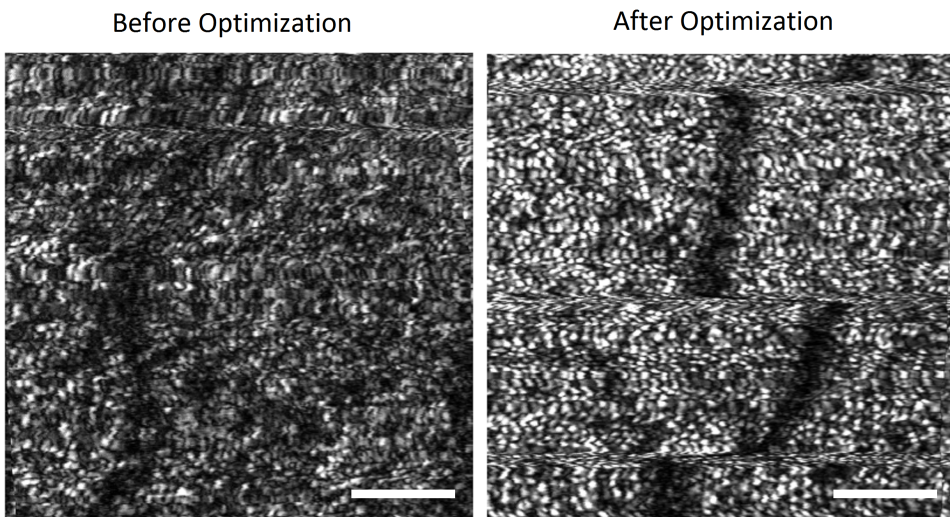


Figure 5.7: OCT *en face* image of human photoreceptor starting before and after optimization with 70 iterations of DONE. All scale bars are 100 μm .

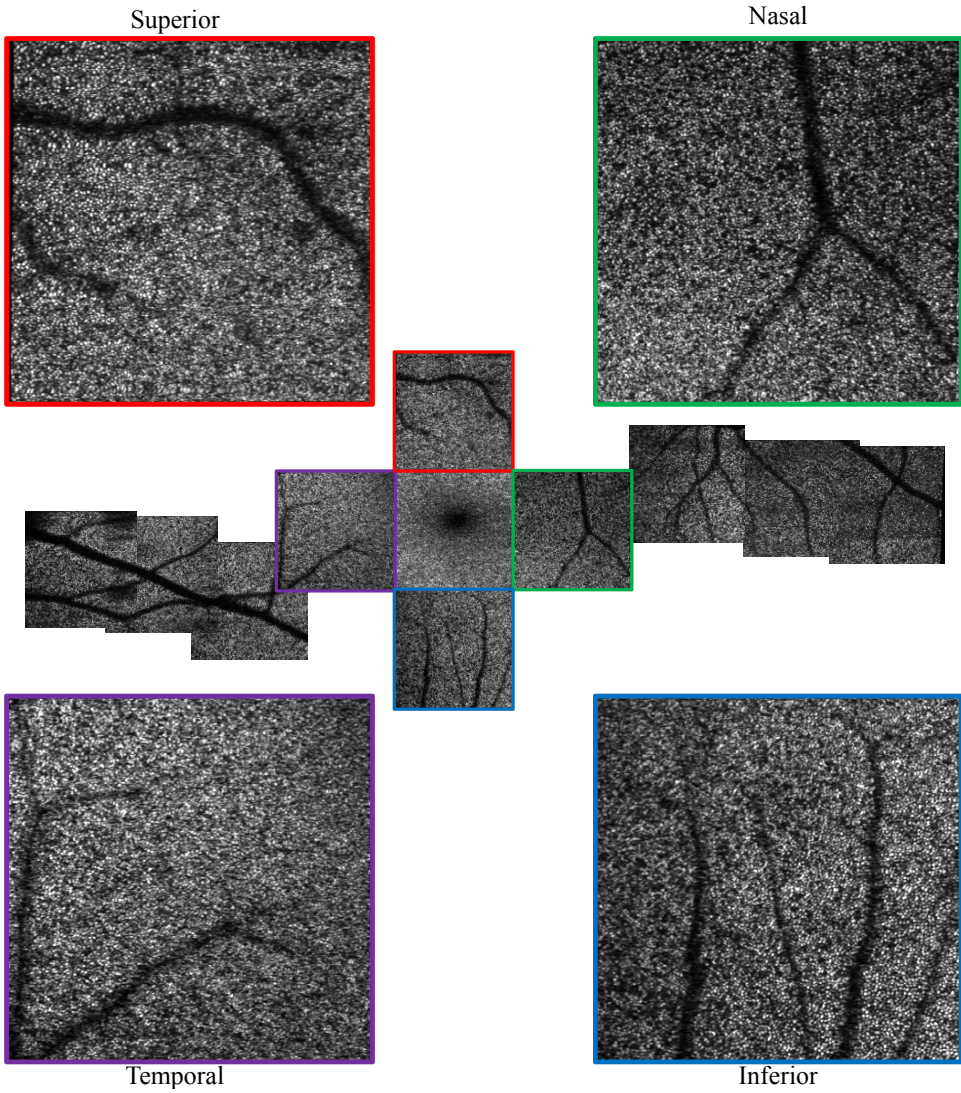


Figure 5.8: Mosaicked images of the retina acquired across a retina in a single imaging session. The position on the retina was controlled by asking the subject to fixate on different calibrated points in the visual field.

coefficient because of noisy outliers. In contrast, the merit function measurements for the DONE algorithm show a general upward trend, and due to the model fitting on the past measurements, is more resistant to the noise in the measurements.

Model based algorithms for WFSL-AO converge faster than stochastic approaches [148, 151, 172]; taking into account prior knowledge of the optical system to reduce the required number of measurements for the optimization [148, 151]. However, in retinal OCT imaging, the use of a static model of the refractive elements of the eye, the cornea and intraocular lens, is hampered by of significant variation among subjects. Using information of past measurements in a dynamic model can increase the accuracy and robustness of the optimization algorithm with respect to the noise level of the measurements. In [156] we showed that all wavefront aberrations can be removed in a wavefront sensorless manner by maximizing the signal intensity in an OCT system.

When comparing the images of the retina before and after optimization in Fig. 5.5, the images after wavefront correction show more contrast. The increase of the image quality metric shown in Fig. 5.6, which is the average WFSL-AO OCT signal at the IS-OS junction layer, confirms an increase in the OCT signal by a factor seven and, hence, an improvement in the signal to noise ratio. When comparing the images before and after optimization in Fig. 5.5 and Fig. 5.7, the images after wavefront correction clearly show that features such as the cones and blood vessels are sharper and have more contrast after optimization. Additionally, it should be noted that not only the intensity of the WFSL-AO OCT signal of the cones improves, but also the shape becomes more circular.

The DONE parameters are general for human retinal imaging with the MAL and the number of aberrations corrected, and were kept constant for all the measurements reported. The results demonstrate that the DONE algorithm has successfully corrected aberrations during *in vivo* measurements, which resulted in improved WFSL-AO OCT images. The circumstances under which the optimization algorithm has to perform have become clearer from the data acquisition. In Fig. 5.7 motion artifacts show small lateral movements of the eye within the C-scan. Throughout the optimization the sample is continuously moving by similar amounts and the metric was evaluated at slightly different lateral locations. The DONE algorithm was exhibited to be robust with respect to unexpected motion induced noise in the image quality metric. Further improvements to the optimization algorithm can be made by filtering out measurements when the signal disappears due to big movements of the eye. However, if relatively long lasting signal deviations that can no longer be considered noise occur, like blinking, the DONE algorithm could fail to find an improved WFSL-AO OCT signal.

Mitigation of the motion artifact during optimization is essential for successful aberration correction with WFSL-AO OCT. Eye movements, and blinking are more likely to occur if the duration becomes too long, which in turn will deteriorate the performance of the algorithm. However, as mentioned previously, the minimum number of iterations permitted will also impact the performance. Active tracking of eye movements could reduce the movement of the image [244,245], but it would lead to additional hardware and added costs for implementation. Fast optimization methods are imperative for application of WFSL-AO aberration compensation to a large group of people. The rate limiting factor of the present WFSL-AO OCT implementation was the time for acquisition of the OCT volumes with a 200 kHz A-scan light source (nominally 40 ms). However, with a

state-of-the-art 1.5 MHz Fourier domain mode locking source [246], the volume acquisition time could be reduced to ~ 5 ms. The second limiting factor of the current implementation was the settling time of the MAL, which was on the order of 5 ms. In this report, we presented a recursive implementation of the DONE algorithm with a computational time of only one millisecond per iteration. For the number of iterations performed in this report, and with incremental modifications to the system, the optimization speed could be in the order of a second with the current set of search parameters for the DONE algorithm.

The aberration correcting performance of DONE relative to the CS was previously presented in [172]. For trained fixators with normal eyes, there were no significant differences in the final image quality with DONE or with the CS algorithm. Empirically, our experience was that the DONE algorithm was more reliable in terms of obtaining an aberration correction for subjects with an average gaze fixation ability. This is emphasized in the graph of Fig. 5.6. The value of the metric function trends upwards even in the presence of motion artifact induced noise in the measurements.

The number of iterations used in the DONE algorithm is a trade-off between the amount of time available for searching, and the final aberration corrected performance. Whenever considering more wavefront aberration modes to correct, more measurements are needed. In contrast, the parameters of the DONE algorithm will remain largely the same, even for more modes. The maximum stroke of the MAL limits the maximum amplitudes of the modes, hence the exploration parameters and bounds for each mode should be carefully chosen. Previously, DONE has successfully been used in WFSL-AO OCT, light sheet microscopy, and simulations of an optical beam forming network [172, 178, 247]. It was shown that DONE outperforms other algorithms in residual wavefront error and convergence speed [172].

5.5. CONCLUSION

The improved and faster version of the DONE optimization algorithm was successfully applied to maximize the OCT signal during *in vivo* measurements of the human eye. The optical wavefront aberrations were corrected using a multi-actuator adaptive lens after linearizing the hysteresis of the actuators to improve the accuracy. This paper demonstrates that the improved version of the DONE algorithm succeeds in drastically improving the OCT signal while achieving a computational time of 1 ms per iteration, making it applicable for many high speed applications in optics. We have shown that the improved version of the DONE algorithm is fast enough for *in vivo* retinal imaging and robust towards small involuntary movements of the eye which it considers as noise in the change of the average OCT signal.

6

APPLYING DONE TO S-PAM

Science is the great antidote to the poison of enthusiasm and superstition.

Adam Smith

We apply the DONE algorithm to a Smart Programmable Array Microscope (S-PAM) to correct sample induced optical wavefront aberrations. The DONE algorithm was modified to work with a sliding-window principle. In this way older measurements are forgotten so that the algorithm can adapt to changing wavefront aberrations. We successfully remove sample induced aberrations. Compared to the coordinate search optimization algorithm the combination of S-PAM and DONE converges faster.

Parts of this chapter have been published in [248].

6.1. INTRODUCTION

The spatial resolution of optical imaging systems is severely influenced by phase aberrations [249]. Because of the spatial refractive index variations in biological samples wavefront aberrations are created when imaging inside tissue. In fluorescence microscopy, for example, these wavefront aberrations in combination with the high numerical aperture lead to a significant resolution loss. Particularly, in confocal microscopy, wavefront aberrations cause a loss in signal, because it is based on point-like excitation and detection through a confocal pinhole. The confocal image of the excitation point spread function becomes wider than the confocal pinhole, which results in a reduction of the measured fluorescence signal. Adaptive optics (AO) can be used to correct these aberrations [36].

Two approaches for aberration corrections are common, namely direct sensing of the wavefront [250] and iterative optimization of an image metric [170]. In the latter case, a metric is computed of the image and an optimization algorithm is applied to find the wavefront correction that optimizes the metric. The image metrics are typically based on fluorescence intensity, image sharpness, or spatial frequencies. This approach is easily applicable on most samples, but requires the acquisition of multiple images for the optimization. Because of the number of aberrations and the low frame rate of traditional confocal microscopes, the optimization can require several seconds or minutes. This correction time precludes dynamic correction of time dependent aberrations, such as those introduced during the acquisition of a stack of images, or by time dependent changes in the sample itself [251]. Moreover, it reduces the image quality due to the effect of photobleaching, which reduces the fluorescence over time, due to the light exposure during the optimization procedure.

Here, we present a Smart Programmable Array Microscope (S-PAM), an approach to AO in optical sectioning fluorescence microscopy. Instead of implementing AO as an add-on to traditional optical sectioning setups, a dedicated system was realized in order to achieve comparable image quality, however, with substantially faster aberration correction. The system uses rejected out of focus light from confocal apertures, to compute a performance metric at high frequency during a scanning procedure. This allows for optimization with the DONE algorithm at a much higher rate than the imaging speed of the system.

6.2. S-PAM

For high speed dynamic correction of aberrations in confocal microscopy, without a wavefront sensor, it is a prerequisite that the performance metric can be estimated at a frequency higher than the microscope frame rate. Since most optical sectioning techniques rely on sample scanning, the metric should be independent from the dynamics of the image scanning procedure. This is practically impossible in a traditional, single beam, laser scanning confocal microscope; as the only signal acquired, namely the fluorescence intensity, varies throughout the field of view. Consequently, a metric calculated over the full image scan is typically used, limiting the optimization frequency to the imaging frame rate. If the fluorescence signal is acquired simultaneously in multiple positions that are uniformly distributed throughout the field of view, the inhomogeneity of the sample is averaged out, and a metric can be computed faster than the full frame

time.

While never used for AO applications, the possibility of achieving optical sectioning in fluorescence imaging through the use of a digital micromirror device (DMD) [252] was first proven in 1998 [253, 254], under the name of Programmable Array Microscope (PAM).

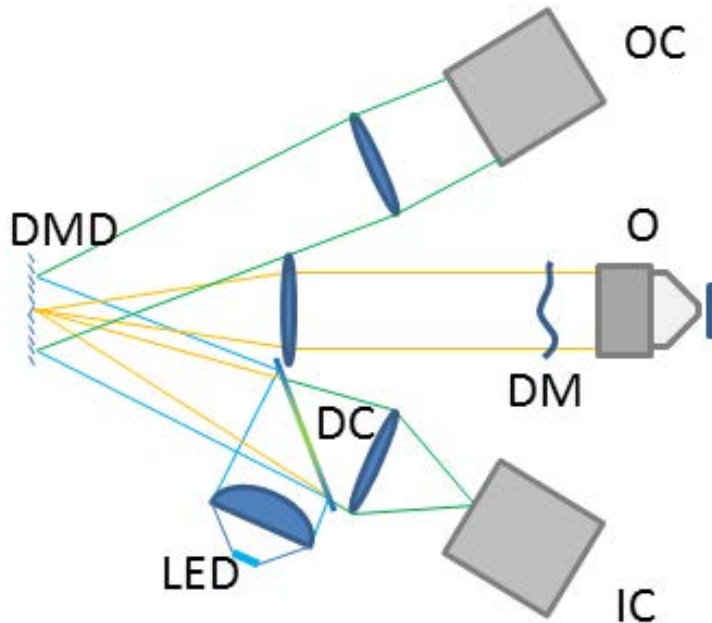


Figure 6.1: A simplified scheme of the S-PAM setup missing a 4f system between the deformable mirror (DM) and the objective (O). Blue lines are the excitation light, green lines are the fluorescence light, orange lines are the optical path shared by excitation and fluorescence light.

LED – light source, DC - Dichroic cube, DMD – Micromirror device, IC – Imaging camera, OC – Optimization camera.

6.3. METHODS

We built a Smart Programmable Array Microscope (S-PAM) as schematically shown in Fig. 6.1, in which a standard PAM is modified with a high speed camera detecting out of focus fluorescence, and with a deformable mirror in the pupil plane, used to correct aberrations in the system. The micromirrors of the DMD act as an array of tiny pinholes and hence, the set-up is essentially an adaptive, multi-aperture confocal microscope.

A DMD is a digitally controlled optical device, consisting of a bi-dimensional array of reflective elements, which can be individually tilted in a binary fashion at an angle of $+\alpha$ (“ON”) or $-\alpha$ (“OFF”). In S-PAM, the DMD is located in an image plane in the optical path shared by excitation and fluorescence light, forming an angle $+2\alpha$ between the normal of the mirror and the optical axis of the excitation light. In this position, the DMD acts as a binary intensity modulator, deflecting excitation light in the direction of the ob-

jective for elements that are “ON”, and out of the system, at an angle -4α , if elements are “OFF”. As such, the micromirrors act as point sources for excitation light, and as conjugated pinholes for the fluorescence light. S-PAM measures the distribution of signal intensity in the proximity of the confocal excitation spot during a parallelized scanning procedure with multiple programmable apertures. Hence, a performance metric related to the shape of the excitation spots can be used. An optically sectioned image can be acquired with a pixelated detector, if all micromirrors turn “ON” for an equally long period. The DMD is used to detect the out of focus fluorescence light rejected by the pinholes in a secondary imaging arm (optimization arm) at an angle -2α from the DMD surface. The optimization camera can operate at the DMD update rate and detects no light from the locations of the “ON” pixels, but only captures the out of focus light rejected by the “OFF” pixels. Even in a diffraction limited confocal setup with the “ON” pixels organized in clusters of size equal to the Airy diameter of the system, part of the fluorescent light is reflected by the “OFF” mirrors adjacent to the “ON” clusters due to diffraction. In the presence of optical aberrations, the area of fluorescence light on the adjacent “OFF” mirrors becomes wider. The width of this area is used as an image quality metric for wavefront sensorless adaptive optics (WFSL-AO). Multiple confocal apertures or pinholes are generated with the DMD concurrently and are used to calculate the metric at different lateral points in the sample simultaneously.

6

The optimization procedure minimizes the second moment of the fluorescence intensity distribution at each of the confocal aperture images reflected from the DMD on the optimization camera. The metric is computed as the average second moment of the fluorescence distribution along all visible confocal apertures in the image. To perform the optimization of the metric the DONE algorithm is used [172, 178]. The DONE algorithm was run with a total of $D = 600$ cosine functions, and the model was fit on a set of the last 50 measurements of the metric. The optimization procedure is performed on the Zernike coefficients, excluding piston, tip, tilt and defocus, up to the 5th order, for a total of 18 degrees of freedom. The S-PAM aberration correction with DONE is compared to a standard confocal microscopy imaged-based coordinate search optimization.

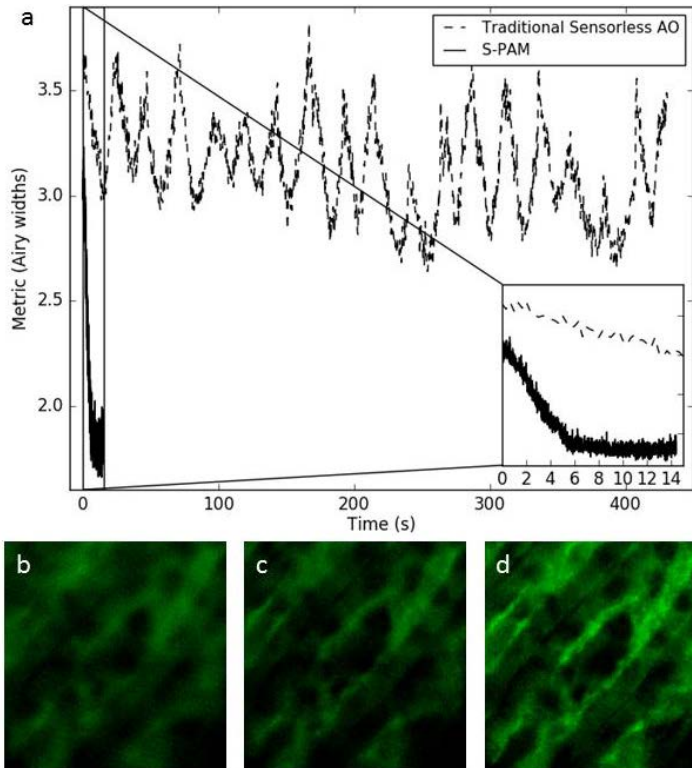
For this application the DONE algorithm is modified to run indefinitely and correct aberrations with slow dynamics. A sliding-window principle is implemented in the recursive least squares (RLS) procedure of the DONE algorithm to only take the last N function or metric evaluations into account. The oldest measurement is removed by performing a RLS downdate and the latest measurement is added with a RLS update, ensuring that at all times a maximum of N measurements are taken into account for the fit of the random Fourier expansion. The RLS downdate is similar to the RLS update, the difference is that it removes a measurement from the current model. In this way DONE adapts to dynamic variations in the aberrations that are varying at a slower rate than the convergence speed of the algorithm. In our implementation of the DONE algorithm the computation of the metric and one iteration take a time shorter than the exposure time of the optimization camera, which runs at 135 Hz. The measurements and the model of DONE are up- and downdated at this rate of 135 Hz and the model is based on the last 50 measurements. Therefore, the model updates its entire set of measurements at a rate of 2.7 Hz and slow changes in the 18 Zernike aberrations are taken into account.

6.4. RESULTS AND DISCUSSION

Wavefront sensorless aberration correction tests were performed on a 16 μm thick tissue slice (Fluocells prepared slide #3, mouse kidney with Alexa Fluor 488, Invitrogen, U.S.A.). While the tissue slice is thin, our experiments show that some sample induced aberrations are present, and the use of AO can increase the image quality. The convergence of DONE with S-PAM is faster than the traditional image based method as shown in Fig. 6.2. There are two reasons for the fast convergence observed in Fig. 6.2(a). The first reason is that S-PAM does not have to wait for a full image to calculate a value for the metric. The second reason is that DONE converges faster than the coordinate search algorithm as it searches more efficiently for the optimum, as previously demonstrated in [172]. In less than 8 seconds, 18 Zernike aberrations have been corrected by DONE. The images after the optimization with DONE in Fig. 6.2(c) are compared with the image before optimization in Fig. 6.2(a) and the image after a more than 400 s long optimization with the CS algorithm in Fig. 6.2(b). The images optimized with DONE in Fig. 6.2(c) have more contrast and are sharper than the other two images. In Fig. 6.3 a larger field of view shows the sample before and after optimization with the DONE algorithm. There is a clear improvement of the image quality and contrast after the optimization with DONE.

6.5. CONCLUSION

In this work, we have shown the successful removal of sample aberrations using the DONE algorithm in S-PAM. At an update rate of 135 Hz, DONE optimizes the image quality of S-PAM images by correcting for 18 Zernike aberrations. DONE succeeds in correcting 18 Zernike aberrations in less than 8 seconds.



6

Figure 6.2: (a) Convergence speed for an artificially induced wide amplitude aberration. A traditional coordinate search procedure, performed on sharpness data of the imaging camera, (dashed line) is compared to the result of S-PAM optimization with DONE (solid line). Image details (b) for uncorrected image, (c) for the image after coordinate search image optimization, and (d) for S-PAM optimization.

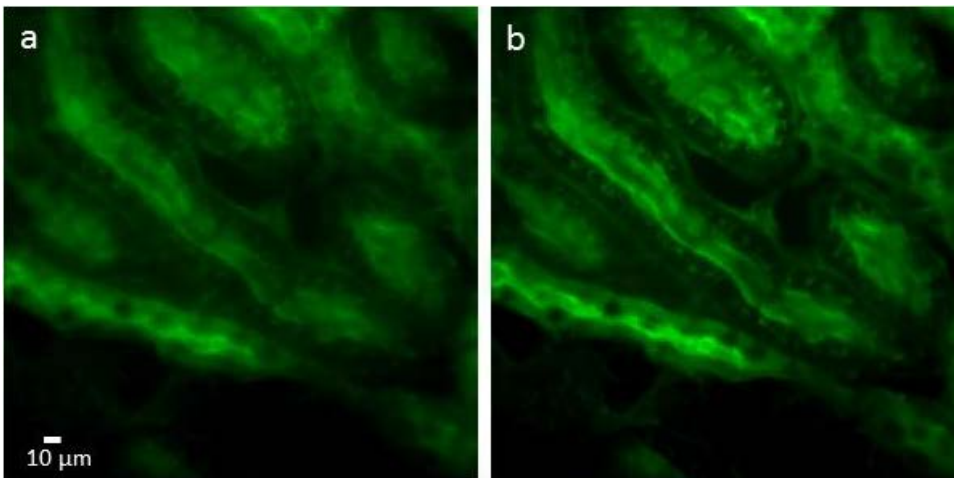


Figure 6.3: Sample (a) without aberration correction, and (b) after optimization with DONE.

7

CONCLUSIONS

Conclusions and recommendations are given for the optical coherence tomography aberration model, the DONE algorithm and the application of DONE in wavefront sensorless aberration correction. Furthermore, the current limitations of the contributions are presented and further research questions that address the current limitations are proposed.

7.1. CONCLUSIONS

The conclusions in this thesis are divided into three different subjects: modeling aberrations in optical coherence tomography, the DONE algorithm, and sensorless aberration correction with DONE.

7.1.1. MODELING ABERRATIONS IN OPTICAL COHERENCE TOMOGRAPHY

Two OCT transfer functions, modeling the effect of lateral aberrations for a mirror and a scattering medium on the OCT signal, have been derived using Fresnel propagation, overlap integrals and Gaussian fields. This is the first time that various wavefront aberrations other than the defocus have been modeled for a mirror and a scattering medium in single-mode fiber based OCT. Measurements on a mirror and a scattering medium with an AO-OCT system closely resembled the calculated transfer functions. This leads us to conclude that the proposed models correctly predict the effect of aberrations in OCT. Furthermore, a one-step defocus correction method has been derived based on these models and was applied successfully. The pseudo-convex nature predicted by the theoretical transfer function is validated by the 100% convergence success of a hill-climbing algorithm both numerically and experimentally. The OCT aberrations models were then used for the development of a better optimization algorithm for wavefront sensorless adaptive optics (WFSL-AO) in OCT. When performing OCT imaging on the human retina, many parameters of the OCT transfer function are not known. To make the WFSL-AO algorithm robust towards this uncertainty, it has to be able to find the optimum when the parameters of the transfer function are unknown. From the measured transfer functions it is concluded that the optimization algorithm has to accurately determine large aberrations in OCT, while performing in the presence of extremely noisy signals. The combination of an unknown transfer function and the noise in the OCT measurements have led to the development of the DONE algorithm.

7.1.2. THE DONE ALGORITHM

We presented the Data-based Online Nonlinear Extremum-seeker (DONE) algorithm that is used to find the minimum of an unknown function using measurements that are costly to obtain and corrupted by noise. DONE maintains a surrogate model in the form of a random Fourier expansion (RFE), which is updated whenever a new measurement is available. After updating the surrogate function, it finds the minimizing argument of this surrogate function with standard derivative-based methods. The next measurement is chosen near the last optimum and this strategy allows the DONE algorithm to measure only in regions of interest, reducing the overall number of measurements required. DONE is comparable to Bayesian optimization algorithms. Both methods rely on the fitting of a surrogate function to the measured data and, subsequently, use other well-known optimization methods to find the optimum of the surrogate function. However, DONE has the distinctive advantage that the computational complexity of one iteration does not grow with the number of measurements that have already been taken. As a theoretical result, we have shown that a RFE surrogate function that is trained by solving a linear least squares can approximate square integrable functions arbitrarily well. Tikhonov regularization is used in this linear least squares problem and an upper bound on the regularization parameter was given. Additionally, a more practical,

well-performing probability distribution for the randomly chosen frequencies of the RFE model was given. By choosing the parameters of the RFE surrogate function of the DONE algorithm in this way, the performance and accuracy of DONE improves. Throughout the thesis, we have applied the DONE algorithm to a benchmark problem and to three simulated applications: optical coherence tomography, optical beam-forming network tuning, and a robot arm. We compared the algorithm to BayesOpt, a Bayesian optimization library. From the numerical simulations, we conclude that the DONE algorithm gave comparable or more accurate results on these applications while being quicker than the Bayesian optimization algorithm, due to the fixed computational complexity per iteration. We have demonstrated that DONE is a powerful optimization algorithm for smooth noisy and costly unknown objective functions with up to 150 input dimensions. After successful application of DONE to the numerical simulations, DONE was applied to experimental optical applications.

7.1.3. SENSORLESS ABERRATION CORRECTION WITH DONE

We have implemented the DONE algorithm in WFSL-AO-OCT imaging and have shown a significantly improved image quality on a roll of Scotch tape and a biological sample. The experiments also showed that the DONE algorithm succeeds in maximizing a noisy signal with high accuracy even for large aberrations. The performance of DONE was compared experimentally to two different methods for WFSL-AO in OCT. DONE outperformed both NEWUOA and a coordinate search algorithm in terms of final root mean square wavefront error and it converged up to ten times quicker than the coordinate search algorithm.

Subsequently, the DONE optimization algorithm was successfully applied to optimize the OCT signal and image quality during *in vivo* measurements of the human eye. The optical wavefront aberrations were corrected using a multi-actuator deformable lens after linearizing the hysteresis of the actuators. We demonstrated that the DONE algorithm succeeded in drastically improving the OCT signal. The computation time was less than 1 ms per iteration, showing that it is suitable for high speed applications in optics. Throughout the experiments the convergence of the DONE algorithm was found to be robust with respect to small involuntary movements of the eye during the optimization.

We have successfully applied DONE to S-PAM, a confocal microscopy technique. At a measurement rate of 135 Hz, DONE optimized the image quality of S-PAM images by simultaneously correcting 18 Zernike aberrations. DONE succeeded in correcting these 18 aberrations in less than 8 seconds. The application of DONE to S-PAM shows but the tip of the iceberg of all the applications DONE can be used for.

To conclude, the DONE algorithm is a powerful tool that can be used in various applications. For all applications considered, DONE has outperformed other derivative-free algorithms in terms of final convergence error. Furthermore, it should be stressed that DONE was quicker than the BayesOpt optimization library, due to the fixed computational complexity per iteration.

7.2. RECOMMENDATIONS FOR FUTURE WORK

In Chapter 2 we developed models for aberrations in optical coherence tomography. However, the validity of the aberration model has to be verified under more extreme conditions such as an increased numerical aperture or larger wavefront aberrations. In future work, improved models for scattering in OCT, such as presented in [255], could be combined with a model for aberrations. A desirable goal would be the creation of a full simulation of an OCT system that includes both wavefront aberrations and the effect of scattering in tissue on the wavefront. This would help in the development of specialized post-processing WFSL-AO-OCT aberration correction methods, such as described in [103], because in that case a full AO-OCT setup is no longer needed to test the algorithms. Additionally, advanced classification algorithms for diagnosis, functional imaging methods and their performance could be simulated in the presence of aberrations.

Throughout this thesis we have described the use of various adaptive and deformable components. Part of the motivation for this thesis was to find a cheaper wavefront correction device for retinal OCT imaging. However, an ideal, affordable and compact deformable device for correcting most wavefront aberrations in OCT introduced by the human eye does not yet exist when using a large pupil. For retinal imaging, deformable mirrors require lots of stroke and actuators to correct all aberrations. Nowadays, these deformable mirrors are still quite expensive and require a lot of optical path length in the imaging system. On the other hand, deformable lenses are still limited by their degrees of freedom or the maximum aberrations they can introduce. The past development of deformable mirrors, deformable lenses and their current prices show great promise for the future. It is recommended to research the tunable lens in more detail and improve on its number of degrees of freedom and magnitude of the aberration correction. In the near future, a more suitable affordable commercial WFSL-AO-OCT device based on these lenses can become available.

A limitation of DONE algorithm is that the computational complexity of the algorithm increases quadratically with the number of random Fourier basis functions. For higher dimensional optimization problems this can become an obstacle, as many basis functions are often required to accurately approximate the cost function. To improve the computation time for larger scale applications of DONE we recommend the use of graphics processing unit (GPU), a field-programmable gate array (FPGA), or another dedicated processing unit. Many computational aspects of DONE, such as the linear least squares procedure to fit the RFE model, can benefit from parallel computations.

In Chapter 6 an option to use the sliding window principle was added to the DONE algorithm such that it can cope with dynamic optical aberrations as applied to the S-PAM microscope. However, it should be further studied how fast and how well DONE with a sliding window can track changing aberrations. For many imaging applications, such as astronomy, dynamic aberrations are a problem and the system has to be able to correct these aberrations at a certain frequency rate. With the sliding window principle added to DONE, the first step towards these applications has been taken, but further research is required to determine what the maximum rate is at which DONE can track dynamic aberrations, such as turbulence. The tracking speed will of course depend on the computation time of DONE, but also on the time needed for one metric evaluation. Another open research question is whether the sliding window leads to quicker local op-

timization convergence for larger scale problems in contrast with the situation in which all the measurements are used in the model.

DONE has successfully been applied to sensorless wavefront aberration correction *in vivo*, but large movements of the patients disrupt any optimization algorithm. Hence, improvements are possible for *in vivo* imaging. These improvements are not related to the DONE optimization algorithm but more to the removal of measurements during the movement of patients or animal subjects. Fault detection mechanisms can be implemented to reject false measurements and if necessary automatically restart the optimization procedure if too little static measurements are acquired. If the eye moves or blinks during the optimization such that the retina is no longer imaged, this faulty measurement should not be passed to the optimization algorithm as a valid measurement or metric evaluation. Furthermore, fault detection mechanisms can be built in together with a segmentation procedure or layer tracking when the algorithm occasionally fails to calculate the metric at the correct layer due to movement and noisy measurements. This kind of fault detection can be added as a software extension to the (DONE) optimization algorithm, however, hardware based solutions to compensate for the eye movement such as eye trackers also exist.

In addition to wavefront sensorless aberration correction, the DONE algorithm has been applied to simulated data for several applications, such as the tuning of an OBFN network, and the positioning of a robot arm. Furthermore, DONE has been applied to applications such as aberration correction in S-PAM, and the extension of the field of view of a light-sheet microscope [247]. DONE is an optimization algorithm designed to optimize arbitrary derivative-free, noisy and costly functions online. Hence, DONE could also be used in other applications such as WFSL-AO in a scanning laser ophthalmoscope, astronomy and other types of microscopy, such as two-photon microscopy, stochastic optical reconstruction microscopy (STORM) or stimulated emission depletion (STED) microscopy. However, DONE is not limited to correcting aberrations in optical beams as demonstrated with the robot arm example. The limitations of the DONE algorithm still have to be further studied in terms of global convergence with many local minima and discontinuous functions. Similar to Bayesian optimization that has been applied to solve a wide range of problems [136], automatic machine learning toolboxes [187], robotics [256], sensor networks [257], interactive animation [258], big data analysis [259] etc. DONE could be applied to these problems as well. Therefore, another recommendation of this thesis is to uncover other applications of DONE.

BIBLIOGRAPHY

- [1] H. R. G. W. Verstraete, M. Almasian, P. Pozzi, R. Bilderbeek, J. Kalkman, D. J. Faber, and M. Verhaegen, “Feedforward operation of a lens setup for large defocus and astigmatism correction,” in *SPIE Photonics Europe*. International Society for Optics and Photonics, 2016, pp. 98 960T–98 960T.
- [2] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, and J. G. Fujimoto, “Optical coherence tomography,” *Science*, vol. 254, no. 5035, pp. 1178–1181, 1991.
- [3] J. G. Fujimoto, M. E. Brezinski, G. J. Tearney, S. A. Boppart, B. Bouma, M. R. Hee, J. F. Southern, and E. A. Swanson, “Optical biopsy and imaging using optical coherence tomography,” *Nat. Med.*, vol. 1, no. 9, pp. 970–972, 1995.
- [4] G. Hanneschläger, A. Nemeth, C. Hofer, C. Goetzloff, J. Reussner, K. Wiesauer, and M. Leitner, “Optical coherence tomography as a tool for non destructive quality control of multi-layered foils,” *Proc. of the 6th NDT in progress*, 2011.
- [5] S. Zhong, Y.-C. Shen, L. Ho, R. K. May, J. A. Zeitler, M. Evans, P. F. Taday, M. Pepper, T. Rades, K. C. Gordon *et al.*, “Non-destructive quantification of pharmaceutical tablet coatings using terahertz pulsed imaging and optical coherence tomography,” *Opt. Laser Eng.*, vol. 49, no. 3, pp. 361–365, 2011.
- [6] J. Veilleux, C. Moreau, D. Lévesque, M. Dufour, and M. Boulos, “Optical coherence tomography for the inspection of plasma-sprayed ceramic coatings,” *J. Therm. Spray Techn.*, vol. 16, no. 3, pp. 435–443, 2007.
- [7] J. Czajkowski, T. Prykäri, E. Alarousu, J. Palosaari, and R. Myllylä, “Optical coherence tomography as a method of quality inspection for printed electronics products,” *Opt. Rev.*, vol. 17, no. 3, pp. 257–262, 2010.
- [8] D. Stifter, K. Wiesauer, M. Wurm, E. Schlotthauer, J. Kastner, M. Pircher, E. Götzinger, and C. Hitzenberger, “Investigation of polymer and polymer/fibre composite materials with optical coherence tomography,” *Meas. Sci. Technol.*, vol. 19, no. 7, p. 074011, 2008.
- [9] P. Targowski, B. Rouba, M. Góra, L. Tymińska-Widmer, J. Marczak, and A. Kowalczyk, “Optical coherence tomography in art diagnostics and restoration,” *Appl. Phys. A*, vol. 92, no. 1, pp. 1–9, 2008.
- [10] V. M. Kodach, D. J. Faber, J. van Marle, T. G. van Leeuwen, and J. Kalkman, “Determination of the scattering anisotropy with optical coherence tomography,” *Opt. Express*, vol. 19, no. 7, pp. 6131–6140, Mar 2011.

- [11] H. M. Subhash, J. N. Hogan, and M. J. Leahy, "Multiple-reference optical coherence tomography for smartphone applications," *SPIE Newsroom*, 2015.
- [12] R. Huber, D. C. Adler, V. J. Srinivasan, and J. G. Fujimoto, "Fourier domain mode locking at 1050 nm for ultra-high-speed optical coherence tomography of the human retina at 236,000 axial scans per second," *Opt. Lett.*, vol. 32, no. 14, pp. 2049–2051, Jul 2007.
- [13] B. E. Bouma and G. J. Tearney, *Handbook of optical coherence tomography*. Marcel Dekker, 2002.
- [14] T. Li, A. Wang, K. Murphy, and R. Claus, "White-light scanning fiber Michelson interferometer for absolute position–distance measurement," *Opt. Lett.*, vol. 20, no. 7, pp. 785–787, 1995.
- [15] H. Maruyama, S. Inoue, T. Mitsuyama, M. Ohmi, and M. Haruna, "Low-coherence interferometer system for the simultaneous measurement of refractive index and thickness," *Appl. Opt.*, vol. 41, no. 7, pp. 1315–1322, 2002.
- [16] M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, and A. F. Fercher, "In vivo human retinal imaging by Fourier domain optical coherence tomography," *J. Biomed. Opt.*, vol. 7, no. 3, pp. 457–463, 2002.
- [17] W. Drexler and J. G. Fujimoto, *Optical coherence tomography: technology and applications*. Springer Science & Business Media, 2008.
- [18] A. F. Fercher, K. Mengedoht, and W. Werner, "Eye-length measurement by interferometry with partially coherent light," *Opt. Lett.*, vol. 13, no. 3, pp. 186–188, 1988.
- [19] A. F. Fercher, "Ophthalmic interferometry," in *Proc. of the International Conference on Optics in Life Sciences, Garmisch-Partenkirchen, Germany*, 1990, pp. 12–16.
- [20] S. Chiba and N. Tanno, "Backscattering optical heterodyne tomography," in *prepared for the 14th Laser Sensing Symposium*, 1991.
- [21] A. Fercher, C. Hitzenberger, G. Kamp, and S. El-Zaiat, "Measurement of intraocular distances by backscattering spectral interferometry," *Opt. Comm.*, vol. 117, no. 1–2, pp. 43 – 48, 1995.
- [22] G. Häusler and M. W. Lindner, "'coherence radar' and 'spectral radar' new tools for dermatological diagnosis," *J. Biomed. Opt.*, vol. 3, no. 1, pp. 21–31, 1998.
- [23] J. F. De Boer, B. Cense, B. H. Park, M. C. Pierce, G. J. Tearney, and B. E. Bouma, "Improved signal-to-noise ratio in spectral-domain compared with time-domain optical coherence tomography," *Opt. Lett.*, vol. 28, no. 21, pp. 2067–2069, 2003.
- [24] R. Leitgeb, C. Hitzenberger, and A. Fercher, "Performance of Fourier domain vs. time domain optical coherence tomography," *Opt. Express*, vol. 11, no. 8, pp. 889–894, 2003.

- [25] E. A. Swanson, D. Huang, C. P. Lin, C. A. Puliafito, M. R. Hee, and J. G. Fujimoto, "High-speed optical coherence domain reflectometry," *Opt. Lett.*, vol. 17, no. 2, pp. 151–153, 1992.
- [26] W. Drexler, U. Morgner, F. X. Kärtner, C. Pitris, S. A. Boppart, X. D. Li, E. P. Ippen, and J. G. Fujimoto, "In vivo ultrahigh-resolution optical coherence tomography," *Opt. Lett.*, vol. 24, no. 17, pp. 1221–1223, 1999.
- [27] H. Kogelnik and T. Li, "Laser beams and resonators," *Appl. Opt.*, vol. 5, no. 10, pp. 1550–1567, Oct 1966.
- [28] S. A. Self, "Focusing of spherical gaussian beams," *Appl. Opt.*, vol. 22, no. 5, pp. 658–661, Mar 1983.
- [29] R. K. Tyson, *Principles of adaptive optics*. CRC press, 2015.
- [30] R. S. Jonnal, O. P. Kocaoglu, R. J. Zawadzki, Z. Liu, D. T. Miller, and J. S. Werner, "A review of adaptive optics optical coherence tomography: Technical advances, scientific applications, and the future review of adaptive optics optical coherence tomography," *Invest. Ophthalm. Vis. Sci.*, vol. 57, no. 9, pp. OCT51–OCT68, 2016.
- [31] R. K. Luneburg and M. Herzberger, *Mathematical theory of optics*. Univ of California Press, 1964.
- [32] M. Born and E. Wolf, *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. CUP Archive, 2000.
- [33] J. Goodman, *Introduction to Fourier optics*. McGraw-hill, 2008.
- [34] V. Mahajan, "Optical imaging and aberrations, part ii. wave diffraction optics." SPIE, 2011.
- [35] F. Roddier, *Adaptive optics in astronomy*. Cambridge university press, 1999.
- [36] M. J. Booth, "Adaptive optics in microscopy," *Philos T. Roy. Soc. A*, vol. 365, no. 1861, pp. 2829–2843, 2007.
- [37] J. A. Kubby, *Adaptive Optics for Biological Imaging*. CRC press, 2013.
- [38] P. Artal and A. Guirao, "Contributions of the cornea and the lens to the aberrations of the human eye," *Opt. Lett.*, vol. 23, no. 21, pp. 1713–1715, 1998.
- [39] J. Liang, B. Grimm, S. Goelz, and J. F. Bille, "Objective measurement of wave aberrations of the human eye with the use of a Hartmann–Shack wave-front sensor," *J. Opt. Soc. Am. A*, vol. 11, no. 7, pp. 1949–1957, 1994.
- [40] J. Porter, A. Guirao, I. G. Cox, and D. R. Williams, "Monochromatic aberrations of the human eye in a large population," *J. Opt. Soc. Am. A*, vol. 18, no. 8, pp. 1793–1803, 2001.

- [41] P. Jacquinot and B. Roizen-Dossier, "II Apodisation," *Prog. Optics*, vol. 3, pp. 29–186, 1964.
- [42] F. Zernike, "Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica*, vol. 1, no. 7-12, pp. 689–704, 1934.
- [43] V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," *Appl. Opt.*, vol. 33, no. 34, pp. 8121–8124, 1994.
- [44] F. Zernike, "Phase contrast, a new method for the microscopic observation of transparent objects," *Physica*, vol. 9, no. 7, pp. 686–698, 1942.
- [45] S. N. Bezdid'ko, "The use of Zernike polynomials in optics," *Sov. J. Opt. Technol.*, vol. 41, pp. 425–429, 1974.
- [46] R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.*, vol. 66, no. 3, pp. 207–211, 1976.
- [47] H. H. van Brug, "Efficient Cartesian representation of Zernike polynomials in computer memory," in *Fifth International Topical Meeting on Education and Training in Optics*. International Society for Optics and Photonics, 1997, pp. 382–392.
- [48] V. N. Mahajan, "Zernike polynomials and aberration balancing," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2003, pp. 1–17.
- [49] V. N. Mahajan and G.-m. Dai, "Orthonormal polynomials in wavefront analysis: analytical solution," *J. Opt. Soc. Am. A*, vol. 24, no. 9, pp. 2994–3016, 2007.
- [50] T. S. Ross, "Limitations and applicability of the Maréchal approximation," *Appl. Opt.*, vol. 48, no. 10, pp. 1812–1818, 2009.
- [51] L. N. Thibos, X. Hong, A. Bradley, and X. Cheng, "Statistical variation of aberration structure and image quality in a normal population of healthy eyes," *J. Opt. Soc. Am. A*, vol. 19, no. 12, pp. 2329–2348, Dec 2002.
- [52] J. S. McLellan, S. Marcos, and S. A. Burns, "Age-related changes in monochromatic wave aberrations of the human eye," *Invest. Ophthalm. Vis. Sci.*, vol. 42, no. 6, pp. 1390–1395, 2001.
- [53] J. Feinleib, S. G. Lipson, and P. F. Cone, "Monolithic piezoelectric mirror for wavefront correction," *Appl. Phys. Lett.*, vol. 25, no. 5, pp. 311–313, 1974.
- [54] G. Vdovin and P. M. Sarro, "Flexible mirror micromachined in silicon," *Appl. Opt.*, vol. 34, no. 16, pp. 2968–2972, 1995.
- [55] R. K. Tyson, M. Scipioni, and J. Viegas, "Generation of an optical vortex with a segmented deformable mirror," *Appl. Opt.*, vol. 47, no. 33, pp. 6300–6306, 2008.

- [56] O. Cugat, S. Basrour, C. Divoux, P. Mounaix, and G. Reyne, "Deformable magnetic mirror for adaptive optics: technological aspects," *Sensor. Actuat. A Phys.*, vol. 89, no. 1, pp. 1–9, 2001.
- [57] C. R. Vogel and Q. Yang, "Modeling, simulation, and open-loop control of a continuous facesheet MEMS deformable mirror," *J. Opt. Soc. Am. A*, vol. 23, no. 5, pp. 1074–1081, 2006.
- [58] L. Zhu, P.-C. Sun, D.-U. Bartsch, W. R. Freeman, and Y. Fainman, "Adaptive control of a micromachined continuous-membrane deformable mirror for aberration compensation," *Appl. Opt.*, vol. 38, no. 1, pp. 168–176, 1999.
- [59] C. Schwartz, E. Ribak, and S. G. Lipson, "Bimorph adaptive mirrors and curvature sensing," *J. Opt. Soc. Am. A*, vol. 11, no. 2, pp. 895–902, 1994.
- [60] D. Brousseau, E. F. Borra, and S. Thibault, "Wavefront correction with a 37-actuator ferrofluid deformable mirror," *Opt. express*, vol. 15, no. 26, pp. 18 190–18 199, 2007.
- [61] R. H. Freeman and J. E. Pearson, "Deformable mirrors for all seasons and reasons," *Appl. Opt.*, vol. 21, no. 4, pp. 580–588, 1982.
- [62] G. Vdovin, S. Middelhoek, and P. M. Sarro, "Technology and applications of micro-machined silicon adaptive mirrors," *Opt. Eng.*, vol. 36, no. 5, pp. 1382–1390, 1997.
- [63] G. Rousset, J. C. Fontanella, P. Kern, P. Gigan, and F. Rigaut, "First diffraction-limited astronomical images with adaptive optics," *Astron. Astrophys.*, vol. 230, pp. L29–L32, 1990.
- [64] E. J. Fernández, I. Iglesias, and P. Artal, "Closed-loop adaptive optics in the human eye," *Opt. Lett.*, vol. 26, no. 10, pp. 746–748, 2001.
- [65] A. Dubra, J. Massa, and C. Paterson, "Preisach classical and nonlinear modeling of hysteresis in piezoceramic deformable mirrors," *Opt. Express*, vol. 13, no. 22, pp. 9062–9070, 2005.
- [66] H. Song, G. Vdovin, R. Fraanje, G. Schitter, and M. Verhaegen, "Extracting hysteresis from nonlinear measurement of wavefront-sensorless adaptive optics system," *Opt. Lett.*, vol. 34, no. 1, pp. 61–63, 2009.
- [67] D. Brousseau, E. F. Borra, M. Rochette, and D. B. Landry, "Linearization of the response of a 91-actuator magnetic liquid deformable mirror," *Opt. express*, vol. 18, no. 8, pp. 8239–8250, 2010.
- [68] N. Devaney, E. Dalimier, T. Farrell, D. Coburn, R. Mackey, D. Mackey, F. Laurent, E. Daly, and C. Dainty, "Correction of ocular and atmospheric wavefronts: a comparison of the performance of various deformable mirrors," *Appl. Opt.*, vol. 47, no. 35, pp. 6550–6562, 2008.
- [69] T. G. Bifano, J. Perreault, R. K. Mali, and M. N. Horenstein, "Microelectromechanical deformable mirrors," *IEEE J. Sel. Topics Quantum Electron.*, vol. 5, no. 1, pp. 83–89, 1999.

- [70] J. W. Hardy, J. E. Lefebvre, and C. L. Koliopoulos, "Real-time atmospheric compensation," *J. Opt. Soc. Am.*, vol. 67, no. 3, pp. 360–369, 1977.
- [71] C. Paterson, I. Munro, and J. C. Dainty, "A low cost adaptive optics system using a membrane mirror," *Opt. Express*, vol. 6, no. 9, pp. 175–185, 2000.
- [72] G. Vdovin, O. Soloviev, A. Samokhin, and M. Loktev, "Correction of low order aberrations using continuous deformable mirrors," *Opt. Express*, vol. 16, no. 5, pp. 2859–2866, 2008.
- [73] M. Blum, M. Büeler, C. Grätzel, and M. Aschwanden, "Compact optical design solutions using focus tunable lenses," in *SPIE Optical Systems Design*. International Society for Optics and Photonics, 2011, pp. 81 670W–81 670W.
- [74] L. W. Alvarez, "Two-element variable-power spherical lens," Feb. 21 1967, uS Patent 3,305,294.
- [75] G.-J. Wang, O. Pomerantzeff, and M. M. Pankratov, "Astigmatism of oblique incidence in the human model eye," *Vision Res.*, vol. 23, no. 10, pp. 1079–1085, 1983.
- [76] D. A. Atchison, A. Bradley, L. N. Thibos, and G. Smith, "Useful variations of the Badal optometer." *Optometry Vision Sci.*, vol. 72, no. 4, pp. 279–284, 1995.
- [77] L. N. Thibos, "Propagation of astigmatic wavefronts using power vectors," *S. Afr. Optom.*, vol. 62, no. 3, pp. 111–113, 2003.
- [78] C. Huang, L. Li, and Y. Y. Allen, "Design and fabrication of a micro Alvarez lens array with a variable focal length," *Microsyst. Technol.*, vol. 15, no. 4, pp. 559–563, 2009.
- [79] E. A. Villegas, E. Alcón, and P. Artal, "Optical quality of the eye in subjects with normal and excellent visual acuity," *Invest. Ophth. Vis. Sci.*, vol. 49, no. 10, pp. 4688–4696, 2008.
- [80] J. Arines and E. Acosta, "Low-cost adaptive astigmatism compensator for improvement of eye fundus camera," *Opt. Lett.*, vol. 36, no. 21, pp. 4164–4166, 2011.
- [81] H. Zappe, *Fundamentals of Micro-optics*. Cambridge University Press, 2010.
- [82] A. Wilson. (2010) Tunable optics. [Online]. Available: http://www.vision-systems.com/articles/2010/07/Tunable_Optics.html
- [83] B. Berge and J. Peseux, "Variable focal lens controlled by an external voltage: An application of electrowetting," *The European Phys. J. E*, vol. 3, no. 2, pp. 159–163, 2000.
- [84] F. Mugele and J.-C. Baret, "Electrowetting: from basics to applications," *J. Phys. Conden. Mat.*, vol. 17, no. 28, p. R705, 2005.
- [85] H. Ren and S.-T. Wu, *Introduction to adaptive lenses*. John Wiley & Sons, 2012, vol. 75.

- [86] U. Wallrabe, "Axicons et al.-Highly aspherical adaptive optical elements for the life sciences," in *2015 Transducers-2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)*. IEEE, 2015, pp. 251–256.
- [87] L. Rizzotto, S. Bonora, Y. Jian, P. Zhang, A. Zam, E. N. Pugh, F. Mammano, R. J. Zawadzki, and M. V. Sarunic, "Comparison of a novel adaptive lens with deformable mirrors and its application in high-resolution in-vivo OCT imaging," vol. 9335, 2015, pp. 933 514–933 514–5.
- [88] S. Bonora, Y. Jian, P. Zhang, A. Zam, E. N. Pugh, R. J. Zawadzki, and M. V. Sarunic, "Wavefront correction and high-resolution in vivo OCT imaging with an objective integrated multi-actuator adaptive lens," *Opt. Express*, vol. 23, no. 17, pp. 21 931–21 941, Aug 2015.
- [89] S. Sato, "Liquid-crystal lens-cells with variable focal length," *Jpn. J. Appl. Phys.*, vol. 18, no. 9, p. 1679, 1979.
- [90] W. Drexler, U. Morgner, R. K. Ghanta, F. X. Kärtner, J. S. Schuman, and J. G. Fujimoto, "Ultrahigh-resolution ophthalmic optical coherence tomography," *Nat. Medicine*, vol. 7, no. 4, pp. 502–507, 2001.
- [91] R. Leitgeb, W. Drexler, A. Unterhuber, B. Hermann, T. Bajraszewski, T. Le, A. Stingl, and A. Fercher, "Ultrahigh resolution fourier domain optical coherence tomography," *Opt. Express*, vol. 12, no. 10, pp. 2156–2165, 2004.
- [92] B. Cense, N. Nassif, T. Chen, M. Pierce, S.-H. Yun, B. Park, B. Bouma, G. Tearney, and J. de Boer, "Ultrahigh-resolution high-speed retinal imaging using spectral-domain optical coherence tomography," *Opt. Express*, vol. 12, no. 11, pp. 2435–2447, 2004.
- [93] M. Wojtkowski, V. Srinivasan, T. Ko, J. Fujimoto, A. Kowalczyk, and J. Duker, "Ultrahigh-resolution, high-speed, Fourier domain optical coherence tomography and methods for dispersion compensation," *Opt. Express*, vol. 12, no. 11, pp. 2404–2422, 2004.
- [94] D. T. Miller, J. Qu, R. S. Jonnal, and K. Thorn, "Coherence gating and adaptive optics in the eye," in *Proc. SPIE*, vol. 4956, 2003, pp. 65–72.
- [95] B. Hermann, E. Fernández, A. Unterhuber, H. Sattmann, A. F. Fercher, W. Drexler, P. M. Prieto, and P. Artal, "Adaptive-optics ultrahigh-resolution optical coherence tomography," *Opt. Lett.*, vol. 29, no. 18, pp. 2142–2144, 2004.
- [96] Y. Zhang, J. Rha, R. S. Jonnal, and D. T. Miller, "Adaptive optics parallel spectral domain optical coherence tomography for imaging the living retina," *Opt. Express*, vol. 13, no. 12, pp. 4792–4811, Jun 2005.
- [97] R. J. Zawadzki, S. M. Jones, S. S. Olivier, M. Zhao, B. A. Bower, J. A. Izatt, S. Choi, S. Laut, and J. S. Werner, "Adaptive-optics optical coherence tomography for high-resolution and high-speed 3D retinal in vivo imaging," *Opt. Express*, vol. 13, no. 21, pp. 8532–8546, Oct 2005.

- [98] E. J. Fernández, B. Považay, B. Hermann, A. Unterhuber, H. Sattmann, P. M. Prieto, R. Leitgeb, P. Ahnelt, P. Artal, and W. Drexler, “Three-dimensional adaptive optics ultrahigh-resolution optical coherence tomography using a liquid crystal spatial light modulator,” *Vision Res.*, vol. 45, no. 28, pp. 3432–3444, 2005.
- [99] 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*, vol. 14, no. 10, pp. 4380–4394, May 2006.
- [100] M. Pircher, R. J. Zawadzki, J. W. Evans, J. S. Werner, and C. K. Hitzenberger, “Simultaneous imaging of human cone mosaic with adaptive optics enhanced scanning laser ophthalmoscopy and high-speed transversal scanning optical coherence tomography,” *Opt. Lett.*, vol. 33, no. 1, pp. 22–24, 2008.
- [101] W. Drexler and J. G. Fujimoto, “State-of-the-art retinal optical coherence tomography,” *Prog. Retin. Eye Res.*, vol. 27, no. 1, pp. 45–88, 2008.
- [102] A. Kumar, W. Drexler, and R. A. Leitgeb, “Subaperture correlation based digital adaptive optics for full field optical coherence tomography,” *Opt. Express*, vol. 21, no. 9, pp. 10 850–10 866, May 2013.
- [103] S. G. Adie, B. W. Graf, A. Ahmad, P. S. Carney, and S. A. Boppart, “Computational adaptive optics for broadband optical interferometric tomography of biological tissue,” *Proc. of the National Academy of Sciences*, vol. 109, no. 19, pp. 7175–7180, 2012.
- [104] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
- [105] P. M. Pardalos, *Approximation and Complexity in Numerical Optimization: Continuous and Discrete Problems*. Springer Science & Business Media, 2013, vol. 42.
- [106] J. Snyman, *Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms*. Springer Science & Business Media, 2005, vol. 97.
- [107] C. G. Broyden, “Quasi-Newton methods and their application to function minimisation,” *Math. Comput.*, vol. 21, no. 99, pp. 368–381, 1967.
- [108] J. E. Dennis, Jr and J. J. Moré, “Quasi-Newton methods, motivation and theory,” *SIAM Rev.*, vol. 19, no. 1, pp. 46–89, 1977.
- [109] R. Fletcher, “Conjugate direction methods,” *Pract. Meth. Optimiz., Second Edition*, pp. 80–94, 1987.
- [110] M. R. Hestenes, *Conjugate direction methods in optimization*. Springer Science & Business Media, 2012, vol. 12.
- [111] A. R. Conn, N. I. Gould, and P. L. Toint, *Trust region methods*. Siam, 2000, vol. 1.

- [112] Z.-J. Shi, "Convergence of line search methods for unconstrained optimization," *Appl. Math. Comp.*, vol. 157, no. 2, pp. 393–405, 2004.
- [113] R. Chartrand, "Numerical differentiation of noisy, nonsmooth data," *ISRN Appl. Math.*, vol. 2011, 2011.
- [114] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*. Siam, 2009, vol. 8.
- [115] Y. Nesterov, "Efficiency of coordinate descent methods on huge-scale optimization problems," *SIAM J. Optimiz.*, vol. 22, no. 2, pp. 341–362, 2012.
- [116] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.
- [117] M. J. D. Powell, "An efficient method for finding the minimum of a function of several variables without calculating derivatives," *The Comp. J.*, vol. 7, no. 2, pp. 155–162, 1964.
- [118] —, "The NEWUOA software for unconstrained optimization without derivatives," in *Large-scale nonlinear optimization*. Springer, 2006, pp. 255–297.
- [119] —, "The BOBYQA algorithm for bound constrained optimization without derivatives," *Cambridge NA Report NA2009/06*, University of Cambridge, Cambridge, 2009.
- [120] —, "On fast trust region methods for quadratic models with linear constraints," *Math. Program. Comp.*, vol. 7, no. 3, pp. 237–267, 2015.
- [121] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík, "Comparing results of 31 algorithms from the black-box optimization benchmarking bbob-2009," in *Proc. of the 12th annual conference companion on Genetic and evolutionary computation*. ACM, 2010, pp. 1689–1696.
- [122] G. Deng and M. C. Ferris, "Adaptation of the uobyqa algorithm for noisy functions," in *Proc. of the 38th conference on Winter simulation*. Winter Simulation Conference, 2006, pp. 312–319.
- [123] V. Torczon, "On the convergence of pattern search algorithms," *SIAM J. Optimiz.*, vol. 7, no. 1, pp. 1–25, 1997.
- [124] J. E. Dennis and V. Torczon, "Derivative-free pattern search methods for multidisciplinary design problems," in *The Fifth AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 1994, pp. 922–932.
- [125] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi *et al.*, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.
- [126] L. Davis, "Handbook of genetic algorithms," 1991.

- [127] D. V. Arnold and H.-G. Beyer, "A comparison of evolution strategies with other direct search methods in the presence of noise," *Comput. Optim. Appl.*, vol. 24, no. 1, pp. 135–159, 2003.
- [128] J. Kennedy, "Particle swarm optimization," in *Encyclopedia of machine learning*. Springer, 2011, pp. 760–766.
- [129] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.
- [130] C.-F. Juang, "A hybrid of genetic algorithm and particle swarm optimization for recurrent network design," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 34, no. 2, pp. 997–1006, 2004.
- [131] S. Tsutsui and A. Ghosh, "Genetic algorithms with a robust solution searching scheme," *IEEE Trans. Evol. Comput.*, vol. 1, no. 3, pp. 201–208, 1997.
- [132] W. J. Gutjahr and G. C. Pflug, "Simulated annealing for noisy cost functions," *J. Global Optim.*, vol. 8, no. 1, pp. 1–13, 1996.
- [133] A. J. Booker, J. E. Dennis Jr, P. D. Frank, D. B. Serafini, V. Torczon, and M. W. Trosset, "A rigorous framework for optimization of expensive functions by surrogates," *Struct. Optimization*, vol. 17, no. 1, pp. 1–13, 1999.
- [134] Y. S. Ong, P. B. Nair, and A. J. Keane, "Evolutionary optimization of computationally expensive problems via surrogate modeling," *AIAA journal*, vol. 41, no. 4, pp. 687–696, 2003.
- [135] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.
- [136] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint arXiv:1012.2599*, 2010.
- [137] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015.
- [138] R. Martinez-Cantin, N. Freitas, E. Brochu, J. Castellanos, and A. Doucet, "A Bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot," *Autonomous Robots*, vol. 27, no. 2, pp. 93–103, 2009.
- [139] R. Martinez-Cantin, "BayesOpt: a Bayesian optimization library for nonlinear optimization, experimental design and bandits," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3735–3739, 2014.
- [140] 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.*, vol. 67, no. 1, pp. 36–44, 2005.

- [141] P. N. Marsh, D. Burns, and J. M. Girkin, "Practical implementation of adaptive optics in multiphoton microscopy," *Opt. Express*, vol. 11, no. 10, pp. 1123–1130, May 2003.
- [142] M. A. Vorontsov and V. P. Sivokon, "Stochastic parallel-gradient-descent technique for high-resolution wave-front phase-distortion correction," *J. Opt. Soc. Am. A*, vol. 15, no. 10, pp. 2745–2758, 1998.
- [143] G. V. Vdovin, "Optimization-based operation of micromachined deformable mirrors," in *Astronomical Telescopes & Instrumentation*. International Society for Optics and Photonics, 1998, pp. 902–909.
- [144] S. Bonora and R. J. Zawadzki, "Wavefront sensorless modal deformable mirror correction in adaptive optics: optical coherence tomography," *Opt. Lett.*, vol. 38, no. 22, pp. 4801–4804, Nov 2013.
- [145] M.-R. Nasiri-Avanaki, S. Hojjatoleslami, H. Paun, S. Tuohy, A. Meadway, G. Dobre, and A. Podoleanu, "Optical coherence tomography system optimization using simulated annealing algorithm," *Proce. of Math. Meth. and Appl. Comp., (WSEAS, 2009)*, pp. 669–674, 2009.
- [146] 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.*, vol. 25, no. 1, pp. 52–54, 2000.
- [147] N. Ji, D. E. Milkie, and E. Betzig, "Adaptive optics via pupil segmentation for high-resolution imaging in biological tissues," *Nat. Methods*, vol. 7, no. 2, pp. 141–147, 2010.
- [148] M. J. Booth, "Wave front sensor-less adaptive optics: a model-based approach using sphere packings," *Opt. Express*, vol. 14, no. 4, pp. 1339–1352, Feb 2006.
- [149] 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*, vol. 18, no. 23, pp. 24 070–24 084, 2010.
- [150] H. Linhai and C. Rao, "Wavefront sensorless adaptive optics: a general model-based approach," *Opt. Express*, vol. 19, no. 1, pp. 371–379, 2011.
- [151] 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*, vol. 29, no. 11, pp. 2428–2438, Nov 2012.
- [152] H. Yang, O. Soloviev, and M. Verhaegen, "Model-based wavefront sensorless adaptive optics system for large aberrations and extended objects," *Opt. Express*, vol. 23, no. 19, pp. 24 587–24 601, Sep 2015.
- [153] J. R. P. Angel, P. Wizinowich, M. Lloyd-Hart, and D. Sandler, "Adaptive optics for array telescopes using neural-network techniques," *Nature*, vol. 348, no. 6298, pp. 221–224, 1990.

- [154] B. C. Csáji, "Approximation with artificial neural networks," *Faculty of Sciences, Etsz Lornd University, Hungary*, vol. 24, p. 48, 2001.
- [155] R. de Kinkelder, J. Kalkman, D. J. Faber, O. Schraa, P. H. Kok, F. D. Verbraak, and T. G. van Leeuwen, "Heartbeat-induced axial motion artifacts in optical coherence tomography measurements of the retina," *Invest. Ophthalm. Vis. Sci.*, vol. 52, no. 6, pp. 3908–3913, 2011.
- [156] H. R. G. W. Verstraete, B. Cense, R. Bilderbeek, M. Verhaegen, and J. Kalkman, "Towards model-based adaptive optics optical coherence tomography," *Opt. Express*, vol. 22, no. 26, pp. 32 406–32 418, Dec 2014. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-22-26-32406>
- [157] B. Povazay, K. Bizheva, A. Unterhuber, B. Hermann, H. Sattmann, A. F. Fercher, W. Drexler, A. Apolonski, W. J. Wadsworth, J. C. Knight, P. S. J. Russell, M. Vetterlein, and E. Scherzer, "Submicrometer axial resolution optical coherence tomography," *Opt. Lett.*, vol. 27, no. 20, pp. 1800–1802, Oct 2002.
- [158] Y. Jian, J. Xu, M. A. Gradowski, S. Bonora, R. J. Zawadzki, and M. V. Sarunic, "Wave-front sensorless adaptive optics optical coherence tomography for in vivo retinal imaging in mice," *Biomed. Opt. Express*, vol. 5, no. 2, pp. 547–559, Feb 2014.
- [159] J. A. Izatt, M. D. Kulkarni, H.-W. Wang, K. Kobayashi, and M. V. Sivak Jr, "Optical coherence tomography and microscopy in gastrointestinal tissues," *IEEE J. Sel. Topics Quantum Electron.*, vol. 2, no. 4, pp. 1017–1028, 1996.
- [160] J. M. Schmitt, A. Knüttel, M. Yadlowsky, and M. A. Eckhaus, "Optical-coherence tomography of a dense tissue: statistics of attenuation and backscattering," *Phys. Med. Biol.*, vol. 39, no. 10, p. 1705, 1994.
- [161] J. A. Izatt, E. A. Swanson, J. G. Fujimoto, M. R. Hee, and G. M. Owen, "Optical coherence microscopy in scattering media," *Opt. Lett.*, vol. 19, no. 8, pp. 590–592, 1994.
- [162] T. G. Van Leeuwen, D. J. Faber, and M. C. Aalders, "Measurement of the axial point spread function in scattering media using single-mode fiber-based optical coherence tomography," *IEEE J. Sel. Topics Quantum Electron.*, vol. 9, no. 2, pp. 227–233, March 2003.
- [163] J. Y. Wang and D. E. Silva, "Wave-front interpretation with zernike polynomials," *Appl. Opt.*, vol. 19, no. 9, pp. 1510–1518, May 1980.
- [164] M. Gu, C. Sheppard, and X. Gan, "Image formation in a fiber-optical confocal scanning microscope," *J. Opt. Soc. Am.. A*, vol. 8, no. 11, pp. 1755–1761, 1991.
- [165] D. G. Voelz, *Computational Fourier optics: a MATLAB tutorial*. SPIE Press, 2011.
- [166] L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, "Standards for reporting the optical aberrations of eyes," *J. Refract. Surg.*, vol. 18, no. 5, pp. S652–S660, 2002.

- [167] G.-M. Dai, "Modified Hartmann–Shack wavefront sensing and iterative wavefront reconstruction," in *Proc. SPIE*, vol. 2201, 1994, pp. 562–573.
- [168] H. R. G. W. Verstraete, M. Verhaegen, and J. Kalkman, "Modeling the effect of wave-front aberrations in fiber-based scanning optical microscopy," in *Imaging and Applied Optics*. OSA, 2013, p. JTU4A.13. [Online]. Available: <http://www.opticsinfobase.org/abstract.cfm?URI=COSI-2013-JTu4A.13>
- [169] J. R. Fienup and J. J. Miller, "Aberration correction by maximizing generalized sharpness metrics," *J. Opt. Soc. Am. A*, vol. 20, no. 4, pp. 609–620, Apr 2003.
- [170] 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.*, vol. 34, no. 16, pp. 2495–2497, Aug 2009.
- [171] S. A. Rahman and M. J. Booth, "Adaptive optics for high-resolution microscopy: wave front sensing using back scattered light," in *SPIE MOEMS-MEMS*. International Society for Optics and Photonics, 2012, pp. 82 530I–82 530I.
- [172] H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Model-based sensor-less wavefront aberration correction in optical coherence tomography," *Opt. Lett.*, vol. 40, no. 24, pp. 5722–5725, Dec 2015. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-40-24-5722>
- [173] J. Porter, N. Sredar, H. Queener, C. Li, and H. Hofer, "Wavefront sensorless adaptive optics imaging in the living eye," in *Imaging and Applied Optics*. OSA, 2013, p. OTu1A.1.
- [174] M. J. D. Powell, *Large-Scale Nonlinear Optimization*. Boston, MA: Springer US, 2006, ch. The NEWUOA software for unconstrained optimization without derivatives, pp. 255–297.
- [175] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Adv. Neur. In.*, 2007, pp. 1177–1184.
- [176] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optimiz.*, vol. 9, no. 4, pp. 877–900, 1999.
- [177] R. A. Waltz, J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Math. Program.*, vol. 107, no. 3, pp. 391–408, 2006.
- [178] L. Bliëk, H. R. G. W. Verstraete, M. Verhaegen, and S. Wahls, "Online optimization with costly and noisy measurements using random Fourier expansions," *IEEE Trans. Neural Netw. Learn. Syst.* [Accepted], 2016.
- [179] L. M. Rios and N. V. Sahinidis, "Derivative-free optimization: a review of algorithms and comparison of software implementations," *J. Global Optim.*, vol. 56, no. 3, pp. 1247–1293, 2013.

- [180] D. R. Jones, C. D. Perttunen, and B. E. Stuckman, "Lipschitzian optimization without the Lipschitz constant," *J. Optimiz. Theory App.*, vol. 79, no. 1, pp. 157–181, 1993.
- [181] P. Gilmore and C. T. Kelley, "An implicit filtering algorithm for optimization of functions with many local minima," *SIAM J. Optimiz.*, vol. 5, no. 2, pp. 269–285, 1995.
- [182] A. L. Custódio and L. N. Vicente, "Using sampling and simplex derivatives in pattern search methods," *SIAM J. Optimiz.*, vol. 18, no. 2, pp. 537–555, 2007.
- [183] D. G. Kbiob, "A statistical approach to some basic mine valuation problems on the Witwatersrand," *Journal of Chemical, Metallurgical, and Mining Society of South Africa*, 1951.
- [184] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Adv. Neur. In.*, 2011, pp. 2546–2554.
- [185] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *LION*. Springer, 2011, pp. 507–523.
- [186] O. Roustant, D. Ginsbourger, and Y. Deville, "Dicekriging, Diceoptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization," *J. Stat. Softw.*, vol. 51, no. 1, p. 54p, 2012.
- [187] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Adv. Neur. In.*, 2012, pp. 2951–2959.
- [188] S. ur Rehman and M. Langelaar, "Efficient global robust optimization of unconstrained problems affected by parametric uncertainties," *Struct. Multidiscip. O.*, pp. 1–18, 2015.
- [189] R. C. Hansen, *Phased array antennas*. John Wiley & Sons, 2009, vol. 213.
- [190] C. G. H. Roeloffzen, L. Zhuang, R. G. Heideman, A. Borreman, and v. W. Etten, "Ring resonator-based tunable optical delay line in LPCVD waveguide technology," 2005.
- [191] A. Meijerink, C. G. H. Roeloffzen, R. Meijerink, L. Zhuang, D. A. Marpaung, M. J. Bentum, M. Burla, J. Verpoorte, P. Jorna, A. Hulzinga *et al.*, "Novel ring resonator-based integrated photonic beamformer for broadband phased array receive antennas—part i: Design and performance analysis," *J. Lightwave Technol.*, vol. 28, no. 1, pp. 3–18, 2010.
- [192] L. Zhuang, C. G. H. Roeloffzen, R. G. Heideman, A. Borreman, A. Meijerink, and W. Van Etten, "Single-chip optical beam forming network in lpcvd waveguide technology based on optical ring resonators," in *Microwave Photonics, 2006. MWP'06. International Topical Meeting on*. IEEE, 2006, pp. 1–4.
- [193] L. Zhuang, *Ring resonator-based broadband photonic beam former for phased array antennas*. University of Twente, 2010.

- [194] L. Bliet, M. Verhaegen, and S. Wahls, "Data-driven minimization with random feature expansions for optical beam forming network tuning," *IFAC*, vol. 48, no. 25, pp. 166 – 171, 2015, 16th {IFAC} Workshop on Control Applications of Optimization CAO'2015Garmisch-Partenkirchen, Germany, 6–9 October 2015.
- [195] J. de Lope, M. Santos *et al.*, "A method to learn the inverse kinematics of multi-link robots by evolving neuro-controllers," *Neurocomputing*, vol. 72, no. 13, pp. 2806–2814, 2009.
- [196] T. Hofmann, B. Schölkopf, and A. J. Smola, "Kernel methods in machine learning," *Ann. Stat.*, pp. 1171–1220, 2008.
- [197] J. A. K. Suykens and J. P. L. Vandewalle, *Nonlinear Modeling: advanced black-box techniques*. Springer Science & Business Media, 2012.
- [198] A. Rahimi and B. Recht, "Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning," in *Adv. Neur. In.*, 2009, pp. 1313–1320.
- [199] A. Singh, N. Ahuja, and P. Moulin, "Online learning with kernels: Overcoming the growing sum problem," in *2012 IEEE International Workshop on Machine Learning for Signal Processing*. IEEE, 2012, pp. 1–6.
- [200] C. J. C. Burges *et al.*, "Simplified support vector decision rules," in *ICML*, vol. 96. Citeseer, 1996, pp. 71–77.
- [201] D. A. F. M. B. Schölkopf, "Sampling techniques for kernel methods," in *Adv. Neur. In.*, vol. 1. MIT Press, 2002, p. 335.
- [202] J. Quinonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *J. Mach. Learn. Res.*, vol. 6, pp. 1939–1959, 2005.
- [203] B. Chen, S. Zhao, P. Zhu, and J. C. Principe, "Quantized kernel recursive least squares algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 9, pp. 1484–1491, 2013.
- [204] L. Zhang and P. Suganthan, "A comprehensive evaluation of random vector functional link networks," *Inform. Sciences*, 2015.
- [205] F. Girosi and G. Anzellotti, "Convergence rates of approximation by translates," DTIC Document, Tech. Rep., 1992.
- [206] A. R. Barron, "Universal approximation bounds for superpositions of a sigmoidal function," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 930–945, 1993.
- [207] A. Rahimi and B. Recht, "Uniform approximation of functions with random bases," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. IEEE, 2008, pp. 555–561.

- [208] L. K. Jones, "A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and neural network training," *Ann. Stat.*, pp. 608–613, 1992.
- [209] A. H. Sayed and T. Kailath, "Recursive least-squares adaptive filters," in *Digital Signal Processing Handbook*, V. Madisetti and D. Williams, Eds. CRC, 1999, ch. 21.
- [210] J. Nocedal, "Updating quasi-Newton matrices with limited storage," *Math. Comp.*, vol. 35, no. 151, pp. 773–782, 1980.
- [211] M. Pogu and J. E. S. De Cursi, "Global optimization by random perturbation of the gradient method with a fixed parameter," *J. of Global Optim.*, vol. 5, no. 2, pp. 159–180, 1994.
- [212] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo method*. John Wiley & Sons, 2011, vol. 707.
- [213] H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Numerical evaluation of advanced optimization algorithms for wavefront aberration correction in OCT," in *Imaging and Applied Optics 2015*. OSA, 2015, p. AOM3E3. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=AOMS-2015-AOM3E3>
- [214] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *Ann. Math. Stat.*, pp. 633–643, 1969.
- [215] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, 1999.
- [216] G. H. Golub and C. F. Van Loan, *Matrix computations*. JHU Press, 2012, vol. 3.
- [217] A. W. Van der Vaart, *Asymptotic statistics*. Cambridge university press, 2000, vol. 3.
- [218] A. Beitollahi and P. Azhdari, "Convergence in probability and almost surely convergence in probabilistic normed spaces," *Math. Sci.*, vol. 6, no. 1, pp. 1–5, 2012.
- [219] H. R. G. W. Verstraete, M. Heisler, M. J. Ju, D. Wahl, L. Bliet, J. Kalkman, S. Bonora, M. Sarunic, Y. Jian, and M. Verhaegen, "In vivo wavefront sensorless adaptive lens OCT with the DONE algorithm," [Submitted], 2016.
- [220] M. Pircher, B. Baumann, E. Götzinger, and C. K. Hitzenberger, "Retinal cone mosaic imaged with transverse scanning optical coherence tomography," *Opt. Lett.*, vol. 31, no. 12, pp. 1821–1823, Jun 2006.
- [221] B. Potsaid, I. Gorczynska, V. J. Srinivasan, Y. Chen, J. Jiang, A. Cable, and J. G. Fujimoto, "Ultra-high speed spectral / Fourier domain OCT ophthalmic imaging at 70,000 to 312,500 axial scans per second," *Opt. Express*, vol. 16, no. 19, pp. 15 149–15 169, Sep 2008.

- [222] M. Pircher, E. Götzinger, H. Sattmann, R. A. Leitgeb, and C. K. Hitzenberger, "In vivo investigation of human cone photoreceptors with SLO/OCT in combination with 3D motion correction on a cellular level," *Opt. Express*, vol. 18, no. 13, pp. 13 935–13 944, Jun 2010.
- [223] D. R. Williams, "Imaging single cells in the living retina," *Vision Res.*, vol. 51, no. 13, pp. 1379–1396, 2011.
- [224] B. Cense, E. Koperda, J. M. Brown, O. P. Kocaoglu, W. Gao, R. S. Jonnal, and D. T. Miller, "Volumetric retinal imaging with ultrahigh-resolution spectral-domain optical coherence tomography and adaptive optics using two broadband light sources," *Opt. Express*, vol. 17, no. 5, pp. 4095–4111, Mar 2009.
- [225] K. Kurokawa, K. Sasaki, S. Makita, M. Yamanari, B. Cense, and Y. Yasuno, "Simultaneous high-resolution retinal imaging and high-penetration choroidal imaging by one-micrometer adaptive optics optical coherence tomography," *Opt. Express*, vol. 18, no. 8, pp. 8515–8527, 2010.
- [226] O. P. Kocaoglu, S. Lee, R. S. Jonnal, Q. Wang, A. E. Herde, J. C. Derby, W. Gao, and D. T. Miller, "Imaging cone photoreceptors in three dimensions and in time using ultrahigh resolution optical coherence tomography with adaptive optics," *Biomed. Opt. Express*, vol. 2, no. 4, pp. 748–763, Apr 2011.
- [227] F. Felberer, J.-S. Kroisamer, B. Baumann, S. Zotter, U. Schmidt-Erfurth, C. K. Hitzenberger, and M. Pircher, "Adaptive optics SLO/OCT for 3D imaging of human photoreceptors in vivo," *Biomed. Opt. Express*, vol. 5, no. 2, pp. 439–456, Feb 2014.
- [228] Y. N. Sulai and A. Dubra, "Non-common path aberration correction in an adaptive optics scanning ophthalmoscope," *Biomed. Opt. Express*, vol. 5, no. 9, pp. 3059–3073, Sep 2014.
- [229] 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*, vol. 19, no. 2, pp. 356–368, Feb 2002.
- [230] S. Zommer, E. N. Ribak, S. G. Lipson, and J. Adler, "Simulated annealing in ocular adaptive optics," *Opt. Lett.*, vol. 31, no. 7, pp. 939–941, Apr 2006.
- [231] H. Hofer, N. Sredar, H. Queener, C. Li, and J. Porter, "Wavefront sensorless adaptive optics ophthalmoscopy in the human eye," *Opt. Express*, vol. 19, no. 15, pp. 14 160–14 171, Jul 2011.
- [232] D. J. Wahl, Y. Jian, S. Bonora, R. J. Zawadzki, and M. V. Sarunic, "Wavefront sensorless adaptive optics fluorescence biomicroscope for in vivo retinal imaging in mice," *Biomed. Opt. Express*, vol. 7, no. 1, pp. 1–12, Jan 2016.
- [233] Y. Jian, R. J. Zawadzki, and M. V. Sarunic, "Adaptive optics optical coherence tomography for in vivo mouse retinal imaging," *J. Biomed. Opt.*, vol. 18, no. 5, pp. 056 007–056 007, 2013.

- [234] K. S. K. Wong, Y. Jian, M. Cua, S. Bonora, R. J. Zawadzki, and M. V. Sarunic, "In vivo imaging of human photoreceptor mosaic with wavefront sensorless adaptive optics optical coherence tomography," *Biomed. Opt. Express*, vol. 6, no. 2, pp. 580–590, Feb 2015.
- [235] Y. Jian, S. Lee, M. J. Ju, M. Heisler, W. Ding, R. J. Zawadzki, S. Bonora, and M. V. Sarunic, "Lens-based wavefront sensorless adaptive optics swept source OCT," *Sci. Rep.*, vol. 6, pp. 27 620–27 620, 2015.
- [236] A. Gullstrand, "Appendix 2," in *Handbuch der Physiologischen Optik*, 3rd ed., H. Von Helmholtz, Ed. Opt. Soc. Am. 1924 **1**, 1909, pp. 351–352.
- [237] Y. Le Grand and S. G. El Hage, "Physiological optics," *Springer Ser. Opt. Sci.*, 1980.
- [238] Y. Jian, K. Wong, and M. V. Sarunic, "Graphics processing unit accelerated optical coherence tomography processing at megahertz axial scan rate and high resolution video rate volumetric rendering," *J. Biomed. Opt.*, vol. 18, no. 2, pp. 026 002–026 002, 2013.
- [239] M. P. J. L. Chang, A. Zadrozny, D. F. Buscher, C. N. Dunlop, and D. J. Robinson, "Hysteresis correction of a piezoelectrically actuated segmented mirror," in *Astronomical Telescopes & Instrumentation*. International Society for Optics and Photonics, 1998, pp. 864–871.
- [240] Q. Yang, C. Ftaclas, M. Chun, and D. Toomey, "Hysteresis correction in the curvature adaptive optics system," *J. Opt. Soc. Am. A*, vol. 22, no. 1, pp. 142–147, 2005.
- [241] A. Dubra, J. Massa, and C. Paterson, "Preisach classical and nonlinear modeling of hysteresis in piezoceramic deformable mirrors," *Opt. Express*, vol. 13, no. 22, pp. 9062–9070, 2005.
- [242] Y. Qin, Y. Tian, D. Zhang, B. Shirinzadeh, and S. Fatikow, "A novel direct inverse modeling approach for hysteresis compensation of piezoelectric actuator in feed-forward applications," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 3, pp. 981–989, June 2013.
- [243] M. Cua, S. Lee, D. Miao, M. J. Ju, P. J. Mackenzie, Y. Jian, and M. V. Sarunic, "Retinal optical coherence tomography at 1 μm with dynamic focus control and axial motion tracking," *J. Biomed. Opt.*, vol. 21, no. 2, p. 026007, 2016.
- [244] M. Pircher, B. Baumann, E. Götzinger, H. Sattmann, and C. K. Hitzenberger, "Simultaneous slo/oct imaging of the human retina with axial eye motion correction," *Opt. Express*, vol. 15, no. 25, pp. 16 922–16 932, Dec 2007.
- [245] C. K. Sheehy, Q. Yang, D. W. Arathorn, P. Tiruveedhula, J. F. de Boer, and A. Roorda, "High-speed, image-based eye tracking with a scanning laser ophthalmoscope," *Biomed. Opt. Express*, vol. 3, no. 10, pp. 2611–2622, 2012.

- [246] T. Klein, W. Wieser, C. M. Eigenwillig, B. R. Biedermann, and R. Huber, "Megahertz OCT for ultrawide-field retinal imaging with a 1050 nm fourier domain mode-locked laser," *Opt. Express*, vol. 19, no. 4, pp. 3044–3062, Feb 2011.
- [247] D. Wilding, P. Pozzi, O. Soloviev, G. Vdovin, C. J. Sheppard, and M. Verhaegen, "Pupil filters for extending the field-of-view in light-sheet microscopy," *Opt. Lett.*, vol. 41, no. 6, pp. 1205–1208, Mar 2016.
- [248] P. Pozzi, D. Wilding, O. Soloviev, H. R. G. W. Verstraete, L. Bliet, G. Vdovin, and M. Verhaegen, "Real time wavefront sensorless aberration correction in digital micromirror based confocal microscopy," *Opt. Express [Accepted]*, 2016.
- [249] W. T. Welford, *Aberrations of optical systems*. CRC Press, 1986.
- [250] 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.*, vol. 36, no. 7, pp. 1062–1064, 2011.
- [251] N. Ji, T. R. Sato, and E. Betzig, "Characterization and adaptive optical correction of aberrations during in vivo imaging in the mouse cortex," *Proc. of the National Academy of Sciences*, vol. 109, no. 1, pp. 22–27, 2012.
- [252] L. J. Hornbeck, "Bistable DMD addressing circuit and method," Aug. 25 1992, uS Patent 5,142,405.
- [253] P. J. Verveer, Q. S. Hanley, P. W. Verbeek, L. J. Van Vliet, and W. M. Jovin, "Theory of confocal fluorescence imaging in the programmable array microscope (PAM)," *J. Microsc.*, vol. 189, no. 3, pp. 192–198, 1998.
- [254] Q. S. Hanley, P. J. Verveer, M. J. Gemkow, D. Arndt-Jovin, and T. M. Jovin, "An optical sectioning programmable array microscope implemented with a digital micromirror device," *J. Microsc.*, vol. 196, pp. 317–331, 1999.
- [255] M. Almasian, N. Bosschaart, T. G. van Leeuwen, and D. J. Faber, "Validation of quantitative attenuation and backscattering coefficient measurements by optical coherence tomography in the concentration-dependent and multiple scattering regime," *J. Biomed. Opt.*, vol. 20, no. 12, pp. 121 314–121 314, 2015.
- [256] D. J. Lizotte, T. Wang, M. H. Bowling, and D. Schuurmans, "Automatic gait optimization with Gaussian process regression." in *IJCAI*, vol. 7, 2007, pp. 944–949.
- [257] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.
- [258] E. Brochu, T. Brochu, and N. de Freitas, "A Bayesian interactive optimization approach to procedural animation design," in *Proc. of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*. Eurographics Association, 2010, pp. 103–112.

- [259] A. Klein, S. Bartels, S. Falkner, P. Hennig, and F. Hutter, "Towards efficient Bayesian optimization for big data," in *NIPS 2015 workshop on Bayesian Optimization (BayesOpt 2015)*, 2015.

SUMMARY

Optical coherence tomography (OCT) is a technique for non-invasive imaging based on low coherence interferometry. Its main application is found in ophthalmology, where it is used for 3D *in vivo* imaging of the cornea and the retina. OCT has evolved over the past decade as one of the most important ancillary tests in ophthalmic practice, providing great diagnostic value for disease screening and monitoring. In retinal OCT imaging, the lateral resolution is not determined by the pupil size, but instead it is limited by optical wavefront aberrations of the cornea and lens. These aberrations reduce the OCT image resolution and lower the signal to noise ratio. To obtain high quality OCT images the optical aberrations can be removed using adaptive optics (AO).

In general, AO consists of an adaptive optical element and a wavefront sensor. The adaptive element, such as a deformable mirror, is used to reshape the wavefront and remove the undesired aberrations. The wavefront sensor measures the aberrations by reconstructing the phase of the wavefront, which is used to determine the correction on the wavefront applied by the deformable mirror. However, the use of a wavefront sensor has some disadvantages. It requires light being directed out of the imaging path onto the wavefront sensor. This leads to a loss of signal in the imaging path and can result in non-common optical path errors in the aberrations estimation procedure. Additionally, the use of a deformable mirror and a wavefront sensor leads to a bulky and expensive OCT setup.

The work presented in this thesis has the goal of reducing the cost and bulkiness of an AO-OCT system. First, we investigate the influence of optical wavefront aberrations to the OCT signal strength. The establishment of the relation between aberrations and the OCT signal strength is key to estimating and correcting the aberrations based on single OCT scans. By using Fresnel optical wave propagation and determining the fiber coupling efficiency, we find that the OCT transfer function, i.e. the function that expresses the relation between the aberrations and OCT signal strength, is quasi-convex. We determine both analytically and experimentally the transfer function for both reflective and scattering media, such as a mirror and Scotch tape sample. Additionally, if the OCT system and its optical properties are well-known we demonstrate a method to correct a defocus aberration in one step.

Second, we use the OCT transfer function to develop and determine an efficient wavefront sensorless (WFSL) AO optimization procedure. WFSL-AO methods aim to correct the aberrations without using a wavefront sensor, but instead base the determination of the wavefront on the imaging signal itself. This eliminates the use of the wavefront sensor, its extra cost and its disadvantages from an AO-OCT setup. To keep up with the OCT imaging rate, which is of the order of several tens of kHz, the algorithm has to be computationally efficient. Furthermore, there are no analytic derivatives available for the optimization and the OCT signal is very noisy. Finally, the derivative-free optimization algorithm also has to be able to determine the aberrations accurately when

dealing with a minimum number of noisy measurements. We developed the Data-based Online Nonlinear extremum-seeker (DONE) algorithm. Every iteration, the DONE algorithm updates a surrogate function, which is based on random Fourier expansions (RFE) of the OCT transfer function, with a new OCT signal measurement. The optimum of the RFE surrogate function is then found with a well-known (quasi-Newton) optimization method. We demonstrate the effectiveness of the DONE algorithm compared to other optimization algorithms for WFSL-AO on biological and non-biological samples. We conclude that DONE has a smaller convergence error, while maintaining similar or faster convergence speeds compared to the other algorithms.

Third, we demonstrate a fully functional WFSL-AO OCT setup for retinal imaging. We use a state-of-the-art deformable lens with 18 actuators, rather than a deformable mirror, which leads to a smaller and more integrated WFSL-AO setup. The WFSL-AO OCT setup is successfully used for *in vivo* retinal OCT imaging and demonstrates that the DONE algorithm can remove the ocular wavefront aberrations with the deformable lens during *in vivo* OCT imaging. By developing a new algorithm and exploring the options for adaptive components, we have succeeded in retinal WFSL-AO OCT.

In a broader perspective, we show that the DONE algorithm is suitable for other applications than WFSL-AO OCT. We demonstrate that the DONE derivative-free optimization algorithm is robust towards noisy measurements for applications in robotics, microscopy and optical beam forming networks.

SAMENVATTING

Optische coherentie tomografie (OCT) is een techniek voor niet-invasieve beeldvorming, gebaseerd op laag coherente interferometrie. De hoofdtoepassing van OCT is in de oogheelkunde, waar het wordt gebruikt voor 3D *in vivo* beeldvorming van het hoornvlies en het netvlies. OCT heeft zich het afgelopen decennium tot één van de meest belangrijke aanvullende tests getoond voor de oogheekundige praktijk, waar het een toegevoegde waarde heeft bij het maken van diagnoses voor het screenen van ziektes. In OCT beeldvorming van het netvlies wordt de laterale resolutie gelimiteerd door optische golffront aberraties. Deze aberraties verslechteren de resolutie en de signaal-ruisverhouding van OCT beelden aanzienlijk. Om beelden van betere kwaliteit te maken kunnen de aberraties verwijderd worden met behulp van adaptieve optiek (AO).

In het algemeen bestaat AO uit een adaptief element en een golffrontsensor. Het adaptieve element, zoals een vervormbare spiegel, wordt gebruikt om het golffront te vervormen en verwijdert de ongewenste aberraties. De golffrontsensor meet de aberraties door de fase van het golffront te reconstrueren en wordt gebruikt om de correctie van het golffront die door de deformeerbare spiegel wordt geactueerd te bepalen. De golffrontsensor heeft een aantal nadelen. De sensor vereist licht dat uit het OCT beeldvormingspad wordt gehaald en wordt gemeten door de golffrontsensor. Dit leidt tot een verlies van signaal in het beeldvormingspad en kan leiden tot niet-gemeenschappelijke fouten in het optische pad die de schattingsmethode voor de aberraties bemoeilijken. Bovendien kan het gebruik van een vervormbare spiegel en een golffrontsensor leiden tot een omvangrijke en dure OCT setup.

Het werk dat gepresenteerd wordt in deze dissertatie heeft het doel om de kosten en omvang van een AO-OCT systeem te reduceren. Ten eerste onderzoeken we de invloed van optische golffrontaberraties op de OCT signaalsterkte. Het vinden van een relatie tussen aberraties en de OCT signaalsterkte is de sleutel tot het schatten en het corrigeren van de aberraties op basis van enkele OCT scans. Door het gebruik van Fresnel propagatie en het bepalen van de glasvezel-koppelingsefficiëntie, vinden we dat de OCT overdrachtsfunctie, d.w.z. de functie die de invloed van aberraties op het OCT signaal uitdrukt, een quasi-convexe functie is. We bepalen, zowel analytisch als experimenteel, de overdrachtsfunctie voor zowel reflectieve en verstrooiende media, zoals een spiegel en Scotch tape monster. Bovendien, als het OCT systeem en zijn optische eigenschappen bekend zijn, demonstreren wij een methode voor om een defocus aberratie te corrigeren in één stap.

Ten tweede gebruiken we de overdrachtsfunctie voor het ontwikkelen en het bepalen van een efficiënte golffrontsensorloze (WFSL) AO optimalisatie procedure. WFSL-AO methoden hebben als doel de aberraties te corrigeren zonder het gebruik van een golffrontsensor en baseren de bepaling van het golffront enkel op het af te beelden signaal zelf. Dit elimineert het gebruik van een golffrontsensor, de bijkomende kosten en

de nadelen van het gebruik van een golffrontsensor in een AO-OCT setup. Om de OCT beeldvormingsnelheid, die in de orde van enkele tientallen kHz is, bij te kunnen houden, moet het algoritme computationeel efficiënt zijn. Daarnaast zijn er geen analytische afgeleide beschikbaar voor de optimalisatie en zit er ruis op het OCT signaal. Het afgeleide vrije optimalisatiealgoritme moet ook in staat zijn om de aberraties nauwkeurig te bepalen uit een minimaal aantal metingen met ruis. Wij stellen het op gegevens gebaseerde online niet-lineaire extremum vinder (DONE) algoritme voor. In elke iteratie werkt het DONE algoritme een surrogaat functie, die gebaseerd is op willekeurige Fourier expansies (RFE), bij met een nieuwe meting van het OCT signaal. Het optimum van de RFE surrogaat functie wordt dan gevonden met een bekende (quasi-Newton) optimalisatie methode. We tonen de effectiviteit van het DONE algoritme aan op biologische en niet-biologische monsters en vergelijken verschillende algoritmen voor WFSL-AO optimalisatie. We concluderen dat DONE een kleinere convergentiefout behaalt en een soortgelijke of snellere convergentie snelheid behoudt in vergelijking met de andere algoritmen.

Ten derde, richten we ons op een volledig functionele WFSL-AO OCT setup voor beeldvorming van het netvlies. We maken gebruik van een state-of-the-art vervormbare lens met 18 actuators, in plaats van een vervormbare spiegel. Het gebruik van de vervormbare lens kan leiden tot kleinere WFSL AO-opstellingen. Daarnaast gebruiken we de WFSL-AO OCT setup voor *in vivo* OCT beeldvorming van het netvlies. De demonstratie toont het succes van het DONE algoritme voor het verwijderen van de oculaire golffront aberraties met de vervormbare lens tijdens *in vivo* beeldvorming. Door het ontwikkelen van een nieuw algoritme en het verkennen van de mogelijkheden voor de adaptieve componenten, zijn we erin geslaagd het *in vivo* netvlies in beeld te brengen met WFSL-AO OCT.

In een breder perspectief tonen wij aan dat het DONE algoritme ook geschikt is voor andere toepassingen dan WFSL-AO OCT. Het is een optimalisatiealgoritme dat robuust is met betrekking tot metingen met ruis en is geschikt voor applicaties in robotica, microscopie, en optische bundelvormingsnetwerken.

LIST OF PUBLICATIONS

JOURNAL PAPERS

- [1] H. R. G. W. Verstraete, B. Cense, R. Bilderbeek, M. Verhaegen, and J. Kalkman, "Towards model-based adaptive optics optical coherence tomography," *Opt. Express*, vol. 22, no. 26, pp. 32406–32418, Dec 2014. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-22-26-32406>
- [2] H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Model-based sensor-less wavefront aberration correction in optical coherence tomography," *Opt. Lett.*, vol. 40, no. 24, pp. 5722–5725, Dec 2015. [Online]. Available: <http://ol.osa.org/abstract.cfm?URI=ol-40-24-5722>
- [3] H. R. G. W. Verstraete, L. Blik, M. Verhaegen, and S. Wahls, "Online optimization with costly and noisy measurements using random Fourier expansions," *IEEE Trans. Neural Netw. Learn. Syst.* [Accepted], 2016.
- [4] H. R. G. W. Verstraete, M. Heisler, M. J. Ju, D. Wahl, L. Blik, J. Kalkman, S. Bonora, M. Sarunic, Y. Jian, and M. Verhaegen, "In vivo wavefront sensorless adaptive lens OCT with the DONE algorithm," [Submitted], 2016.
- [5] P. Pozzi, D. Wilding, O. Soloviev, H. R. G. W. Verstraete, L. Blik, G. Vdovin, and M. Verhaegen, "Real time wavefront sensorless aberration correction in digital micromirror based confocal microscopy," *Opt. Express* [Accepted], 2016.

CONFERENCE PAPERS

- [1] H. R. G. W. Verstraete, M. Verhaegen, and J. Kalkman, "Modeling the effect of wave-front aberrations in fiber-based scanning optical microscopy," in *Imaging and Applied Optics*. Optical Society of America, 2013, p. JT4A.13. [Online]. Available: <http://www.opticsinfobase.org/abstract.cfm?URI=COSI-2013-JT4A.13>
- [2] H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Numerical evaluation of advanced optimization algorithms for wavefront aberration correction in OCT," in *Imaging and Applied Optics 2015*. Optical Society of America, 2015, p. AOM3E3. [Online]. Available: <http://www.osapublishing.org/abstract.cfm?URI=AOMS-2015-AOM3E3>
- [3] H. R. G. W. Verstraete, S. Wahls, J. Kalkman, and M. Verhaegen, "Wavefront Sensorless Algorithms for Wavefront Correction in Model-Based OCT," in *AOIM Padova. Adaptive Optics for Industry and Medicine*, 2015.

-
- [4] H. R. G. W. Verstraete, R. Bilderbeek, J. Kalkman, and M. Verhaegen, "Feedforward Operation of a Lens Setup for Large Defocus and Astigmatism Correction," in *AOIM Padova. Adaptive Optics for Industry and Medicine*, 2015.
 - [5] H. R. G. W. Verstraete, M. Almasian, P. Pozzi, R. Bilderbeek, J. Kalkman, D. J. Faber, and M. Verhaegen, "Feedforward operation of a lens setup for large defocus and astigmatism correction," in *SPIE Photonics Europe*. International Society for Optics and Photonics, 2016, pp. 98 960T–98 960T.

CURRICULUM VITÆ

Hans Roeland Geert Wim VERSTRAETE

28-06-1990 Born in Brasschaat, Belgium.

EDUCATION

2008–2011 Bachelor of Science Electrical Engineering
Minor Finance and Honours programme
Delft University of Technology

2011–2012 Master of Science Control Systems
Imperial College London

2012–2016 PhD. Electrical Engineering & Control Systems
Delft University of Technology
Thesis: Optimization-based adaptive optics for optical co-
herence tomography
Promotor: Prof. dr. ir. M. Verhaegen

AWARDS

2011 Best High Tech Start-up Business plan - IEEE Benelux section

2012 Hertha Ayrton Centenary Prize, best MSc project - Imperial College London