Implementation and benchmarking of a crosstalk-free method for wavefront Zernike coefficients reconstruction using Shack-Hartmann sensor data

Biesheuvel, R. S.; Janssen, A. J.E.M.; Pozzi, P.; Pereira, S. F.

DOI
10.1364/OSAC.1.000581

Publication date
2018

Document Version
Final published version

Published in
OSA Continuum

Research Article | Vol. 1, No. 2 | 15 Oct 2018 | OSA CONTINUUM 581

Implementation and benchmarking of a crosstalk-free method for wavefront Zernike coefficients reconstruction using Shack-Hartmann sensor data

R. S. Biesheuvel,[1,4] A. J. E. M. Janssen,[2] P. Pozzi,[3] and S. F. Pereira[1,*]

[1] *Optics Research Group, ImPhys Department, Faculty of Applied Sciences, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, The Netherlands*
[2] *Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O.Box 513, 5600 MB Eindhoven, The Netherlands*
[3] *Delft Center of Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands*
[4] *Present address: DEMCON Focal BV, Institutenweg 25A, 7521 PH Enschede, The Netherlands*
[*] *S.F.Pereira@tudelft.nl*

Abstract: In wavefront characterization, often the combination of a Shack-Hartmann sensor and a reconstruction method utilizing the Cartesian derivatives of Zernike circle polynomials (the least-squares method, to be called here Method A) is used, which is known to introduce crosstalk. In [J. Opt. Soc. Am. A **31**, 1604 (2014], a crosstalk-free analytic expression of the LMS estimator of the wavefront Zersectnike coefficients is given in terms of wavefront partial derivatives (leading to what we call Method B). Here, we show an implementation of this analytic result where the derivative data are obtained using the Shack-Hartmann sensor and compare it with the conventional least-squares method.

1. Introduction

Determination of wavefronts is an important field in optics. In adaptive optics, for example, the wavefront is constantly monitored in order to feed the servo system that corrects distortions on it due to the optical system self or due to the environment. Adaptive optics started in the field of astronomy, but nowadays it has been extended, among others, to the fields of microscopy and retinal imaging. Apart from that, it is worth mentioning the application in accurate characterization of complex optical systems for aberrations.

In many applications involving wavefront reconstruction, Shack-Hartmann sensors are frequently used. In these sensors, an array of microlenses focuses the incoming beam into a camera, resulting in a grid of spots that are displaced from their center position (flat wave reference) because of aberrations contained in the wavefront. The displacement of each spot is then determined and used to calculate the local slope of the wavefront across each lens of the array. Once the slopes are known, an often-used polynomial basis to expand the phase of the scalar field (wavefront) is the set of Zernike circle polynomials. The wavefront Zernike coefficients can be estimated with for example, the least-squares method, to be called Method A henceforth. However, already since the 1980s, it is known that there is a fundamental problem in the combination of the least-squares method and Zernike circle polynomials: this method presents crosstalk between coefficients. In 2014, Janssen introduced an analytic result in the form of a new relation between the wavefront derivative data and the wavefront Zernike coefficients [1]. This relation theoretically solves the crosstalk problem that is present in the least-squares method. In particular, the wavefront coefficients, once estimated, do not change anymore when the number of Zernike circle polynomials involved in the fit is increased further.

#338367 https://doi.org/10.1364/OSAC.1.000581
Journal © 2018 Received 13 Jul 2018; accepted 7 Sep 2018; published 1 Oct 2018

OSA CONTINUUM

In this paper, we show an implementation of Janssen's analytic result in the reconstruction of several wavefronts that have been generated using a spatial light modulator. In particular, we show that this method, to be called Method B henceforth, does not suffer from cross-talk between the coefficients.

The paper is organized as follows. In Section 2 we present theoretical aspects of the problem, and in Section 3, the experimental setup and methods are explained. The actual results are presented and discussed in Section 4, and the conclusions from these measurements are presented in Section 5.

2. Theory

In this section, Zernike circle polynomials are discussed and used as a basis to describe the wavefront deviation with respect to a suitably defined reference sphere of an optical system. This wavefront deviation or wavefront aberration function will be loosely called 'wavefront aberration' or 'aberration' in what follows. We start this section by showing the definition of Zernike circle polynomials and a description of the data that is obtained with the Shack-Hartmann sensor is discussed. Next, we discuss two reconstruction methods: the well-known least-squares method (Method A) and the recently method introduced by Janssen [1] (Method B).

2.1. Zernike circle polynomials as a basis for wavefront expansion

In systems where aberrations are desired to be known, such as microscopes and telescopes, a circular aperture is usually present. In order to describe these aberrations, Zernike circle polynomials are commonly used. In this paper, the American National Standards Institute (ANSI) definition for Zernike circle polynomials is considered, which is commonly used to describe wavefront reconstruction from Shack-Hartmann data [2]. As defined in [3], the polynomials in polar coordinates (ρ, θ) are given by

$$z_n^m(\rho,\theta) = N_n^m R_n^{|m|}(\rho)\Theta_m(\theta), \tag{1}$$

where

$$N_n^m = \sqrt{(2-\delta_{m0})(n+1)}, \tag{2}$$

$$R_n^{|m|}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s!\left(\frac{n-|m|}{2}-s\right)!\left(\frac{n+|m|}{2}-s\right)!}\rho^{n-2s}, \tag{3}$$

$$\Theta_m(\theta) = \begin{cases} \cos(m\theta), & \text{if } m \geq 0, \\ -\sin(m\theta), & \text{if } m < 0, \end{cases} \tag{4}$$

such that $z_n^m(\rho,\theta)$ is a real-valued, orthonormal (on the unit disc) expression for Zernike circle polynomials, $n = 0, 1, 2, \ldots$ is the degree of the Zernike polynomial, and $m = 0, \pm 1, \pm 2, \ldots$ the azimuthal order satisfying:

$$n \geq 0, \tag{5}$$

$$n - |m| \text{ is even}, \tag{6}$$

$$|m| \leq n, \tag{7}$$

$\rho \leq 1$ being the radius on the unit disc, and $\delta_{nn'}$ the Kronecker delta function. Here, the n^{th} "order" polynomial is referred to as the n^{th} *degree* polynomial.

As mentioned above, these circle polynomials are orthonormal on the unit disc, i.e.,

OSA CONTINUUM

$$\frac{1}{\pi}\int_0^1\int_0^{2\pi} z_n^m(\rho,\theta)z_{n'}^{m'}(\rho,\theta)\rho\mathrm{d}\theta\mathrm{d}\rho = \delta_{nn'}\delta_{mm'}. \tag{8}$$

With this definition, any real-valued wavefront function (defined on the unit disc) can be described as a linear combination of the Zernike circle polynomials given by

$$W(\rho,\theta) = \sum_{m=-\infty}^{\infty}\sum_{n\in\eta_m} a_n^m z_n^m(\rho,\theta), \tag{9}$$

where η_m is the set of allowed n values dependent on m, namely $\eta_m = \{|m|, |m|+2, |m|+4, \ldots\}$, when $m \neq 0$, and $\eta_m = \{2, 4, \ldots, \infty\}$ when $m = 0$ and $z_0^0 = 1$. This set ensures that the constraints set in n and m are all met. Finally, a_n^m is the real-valued Zernike coefficient.

2.2. Complex-valued definition of the Zernike circle polynomials

Following the complex-valued definition of the Zernike circle polynomials as used in [1] in polar coordinates (ρ,θ) on the unit disc, we have that

$$Z_n^m(\rho,\theta) = R_n^{|m|}(\rho)e^{im\theta}. \tag{10}$$

The radial polynomial $R_n^{|m|}$ is given as in Eq. (1) above, and can also be given as

$$R_n^{|m|}(\rho) = \rho^{|m|}P_{\frac{n-|m|}{2}}^{(0,|m|)}(2\rho^2-1), \tag{11}$$

where $P_k^{(\alpha,\beta)}(x)$ is the Jacobi polynomial of degree k, which is orthogonal with respect to the weight $(1-x)^\alpha(1+x)^\beta$ on the interval $[-1,1]$. Note that the factor $\rho^{|m|}$ is missing in [1], Eq. (6), but it is merely a typo and has no consequences in the further developments in that reference. We set $Z_n^m = 0$ for all values of n and m where $n-|m|$ is odd or negative. There is the normalization condition

$$\int_0^1\int_0^{2\pi} Z_n^m(\rho,\theta)\left(Z_{n'}^{m'}(\rho,\theta)\right)^*\rho\mathrm{d}\theta\mathrm{d}\rho = \frac{\pi}{n+1}\delta_{nn'}\delta_{mm'}. \tag{12}$$

This orthogonality means that, like the real-valued Zernike polynomial, any sufficiently smooth (complex) wavefront can be described by

$$W(\rho,\theta) = \sum_{m=-\infty}^{\infty}\sum_{n\in\eta_m} \alpha_n^m Z_n^m(\rho,\theta), \tag{13}$$

where α_n^m are generally complex-valued coefficients corresponding to the complex Zernike polynomial and η_n is as in Eq. (9). Expanding the complex exponential in Eq. (10) leads to the following conversion between this definition of complex Zernike cirlce polynomials and the ANSI standard

$$N_n^m Z_n^m(\rho,\theta) = \begin{cases} z_n^{|m|}(\rho,\theta) + iz_n^{-|m|}(\rho,\theta), & \text{if } m>0,\\ z_n^{|m|}(\rho,\theta) - iz_n^{-|m|}(\rho,\theta), & \text{if } m<0,\\ z_n^m, & \text{if } m=0, \end{cases} \tag{14}$$

where N_n^m is defined in Eq. (2).

From the definition of the complex Zernike polynomial in Eq. (10) we can also see that the complex conjugate $\left(Z_n^{|m|}\right)^*$ is equal to $Z_n^{-|m|}$. This observation, together with the relations

OSA CONTINUUM

between the real and complex Zernike circle polynomials leads to the expression

$$z_n^m = \begin{cases} \frac{N_n^m}{2}\left(Z_n^{|m|} + Z_n^{-|m|}\right) = N_n^m \Re\left(Z_n^{|m|}\right), & \text{if } m > 0 \\ \frac{N_n^m}{2i}\left(Z_n^{|m|} - Z_n^{-|m|}\right) = N_n^m \Im\left(Z_n^{|m|}\right), & \text{if } m < 0 \\ N_n^m Z_n^{|m|}, & \text{if } m = 0. \end{cases} \tag{15}$$

When using complex polynomials to describe a real wavefront, the coefficients are usually complex valued. In order to transform them back to the coefficients of the real-valued Zernike z_n^m, the following relations can be used

$$a_n^m = \begin{cases} \frac{1}{N_n^m}\Re\left(\alpha_n^{|m|} + \alpha_n^{-|m|}\right), & \text{if } m > 0 \\ \frac{-1}{N_n^m}\Im\left(\alpha_n^{|m|} - \alpha_n^{-|m|}\right), & \text{if } m < 0 \\ \frac{1}{N_n^m}\Re\left(\alpha_n^m\right), & \text{if } m = 0. \end{cases} \tag{16}$$

Here, the Zernike circle polynomials are evaluated on the computer, and data is saved in vectors and matrices. To make it easy to loop over all polynomials, a single index is introduced. In the notation by [3], this single index $j = 0, 1, ...$ is given by

$$j = \frac{n(n+2)+m}{2}, \tag{17}$$

where n is the Zernike degree and m the azimuthal order. From now on, we will write $Z_j = z_n^m$. The reverse can also be done, that is finding n and m from j as follows

$$n = \left\lceil \frac{-3 + \sqrt{9+8j}}{2} \right\rceil, \tag{18}$$

$$m = 2j - n(n+2), \tag{19}$$

where $\lceil x \rceil$ denotes the "ceiling" function, that is the smallest integer greater or equal than x.

2.3. Data from the Shack-Hartmann sensor

It is well known that aberrations in an optical system degrade its imaging quality. The phase aberration is defined by a function $\varphi_W = arg(W)$. A typical way of measuring this function is by using a Shack-Hartmann sensor. A Shack-Hartmann sensor consists of a camera chip and lenslet array. The lenslets are placed at the focal distance from the camera chip such that an incoming plane wave will be focused as many spots on the camera. If the incoming wave is aberrated, the position of each spot on the sensor will changed.

It can be derived from the theory how much a spot is displaced due to an aberration. Because this displacement is linear, the reverse problem can be solved as well. In other words, one can retrieve the aberration given the spot displacement.

In Ref. [4], it has proven that for a varying wavefront over the subaperture, the slope needs to be averaged over that sub-aperture. The expressions for the average slopes in terms of displacement of the spot on the camera becomes

$$\begin{cases} \frac{1}{A_\Sigma}\int_\Sigma \frac{\partial W}{\partial x}\mathrm{d}x\mathrm{d}y &= r\frac{\Delta x}{f} \\ \frac{1}{A_\Sigma}\int_\Sigma \frac{\partial W}{\partial y}\mathrm{d}x\mathrm{d}y &= r\frac{\Delta y}{f} \end{cases} \tag{20}$$

where $\Delta x, \Delta y$ are the shift in the $x-$ and $y-$ positions of the spots, r is the radius of the incoming beam on the Shack-Hartmann sensor, and f is the focal length of the sensor lens array, Σ is the illuminated sub-aperture domain, with surface area A_Σ. Here Σ and A_Σ change when the sub-aperture is only partially illuminated (i.e., at the edge of the beam). This averaging of the slope is of great matter in recovering the wavefront.

OSA CONTINUUM

2.3.1. Method A

The least-squares (LSQ) fit is based on the real Zernike circle polynomials, and uses the fact that the coefficients are not dependent on x and y. The wavefront in x, y coordinates (with $\sqrt{x^2+y^2} \leq 1$) can be described as

$$W(x,y) = \sum_{m=-\infty}^{\infty} \sum_{n \in \eta_m} a_n^m z_n^m(x,y). \tag{21}$$

The measured quantity, however, is not W but $\frac{\partial W}{\partial x}$ and $\frac{\partial W}{\partial y}$. Taking the partial derivatives to x and y results in the over determined system of

$$\begin{cases} \frac{\partial W}{\partial x} = \sum_{m=-\infty}^{\infty} \sum_{n \in \eta_m} a_n^m \frac{\partial z_n^m}{\partial x} \\ \frac{\partial W}{\partial y} = \sum_{m=-\infty}^{\infty} \sum_{n \in \eta_m} a_n^m \frac{\partial z_n^m}{\partial y}. \end{cases} \tag{22}$$

This system can be solved for a finite set of polynomials. Using Eq. (20), given the displacements Δx and Δy, one can create a vector $\mathbf{s}$ containing the slopes as such

$$\mathbf{s} = \left[\left.\overline{\frac{\partial W}{\partial x}}\right|_1 \quad \left.\overline{\frac{\partial W}{\partial x}}\right|_2 \quad \cdots \quad \left.\overline{\frac{\partial W}{\partial x}}\right|_{n_{\text{spots}}} \quad \left.\overline{\frac{\partial W}{\partial y}}\right|_1 \quad \left.\overline{\frac{\partial W}{\partial y}}\right|_2 \quad \cdots \quad \left.\overline{\frac{\partial W}{\partial y}}\right|_{n_{\text{spots}}} \right]^T. \tag{23}$$

The partial derivatives of the Zernikes in the x- and y-direction can also be put in a matrix, called the geometry matrix. We recall the convention that $Z_j = z_n^m$ with j, m, n related as in Eqs. (17)–(19). The geometry matrix G can be built up $\left.\overline{\frac{\partial Z_j}{\partial x}}\right|_n$, the average gradient of Zernike mode j at the position of subaperture n. This averaging is done due to the fact that the spot displacement measured with the Shack-Hartmann sensor is proportional to the average slope of the wavefront, as expressed in Eq. (20). It should be noted that the positions over which the averaging is done is normalized to the unit disc. These windows are the same defined in Eq. (20). The matrix will have a size of $(2n_{\text{spot}} \times J)$, where J is the maximum index for the Zernike modes used to have a good approximation of the true wavefront. The expression for G becomes

$$G = \begin{bmatrix} \left.\overline{\frac{\partial Z_1}{\partial x}}\right|_1 & \left.\overline{\frac{\partial Z_1}{\partial x}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_1}{\partial x}}\right|_{n_{\text{spot}}} & \left.\overline{\frac{\partial Z_1}{\partial y}}\right|_1 & \left.\overline{\frac{\partial Z_1}{\partial y}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_1}{\partial y}}\right|_{n_{\text{spot}}} \\ \left.\overline{\frac{\partial Z_2}{\partial x}}\right|_1 & \left.\overline{\frac{\partial Z_2}{\partial x}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_2}{\partial x}}\right|_{n_{\text{spot}}} & \left.\overline{\frac{\partial Z_2}{\partial y}}\right|_1 & \left.\overline{\frac{\partial Z_2}{\partial y}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_2}{\partial y}}\right|_{n_{\text{spot}}} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \left.\overline{\frac{\partial Z_J}{\partial x}}\right|_1 & \left.\overline{\frac{\partial Z_J}{\partial x}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_J}{\partial x}}\right|_{n_{\text{spot}}} & \left.\overline{\frac{\partial Z_J}{\partial y}}\right|_1 & \left.\overline{\frac{\partial Z_J}{\partial y}}\right|_2 & \cdots & \left.\overline{\frac{\partial Z_J}{\partial y}}\right|_{n_{\text{spot}}} \end{bmatrix}^T. \tag{24}$$

The system of Eq. (22) can then be written as

$$\mathbf{s} \approx G \cdot \mathbf{a}, \tag{25}$$

where $\mathbf{a}$ is the vector containing the Zernike coefficients. The least-squares estimation of a_n^m becomes

$$\mathbf{a} \approx G^+ \cdot \mathbf{s}, \tag{26}$$

where G^+ the generalized inverse of the geometry matrix. This is an approximation as G only contains the information of a finite number of Zernike modes, and their contribution is averaged over the lenslet array.

OSA CONTINUUM

2.3.2. Method B

Method B relies on an analytical relation found between the local derivatives of the wavefront and Zernike circle polynomials. This is in contrast with method A, where there is a link between the local derivatives of the wavefront and the derivatives of the Zernike circle polynomials.

In Ref. [1], it has found that the LMS complex coefficients are given as (see Appendix)

$$\hat{\alpha}_n^m = C_n^m \varphi_n^m - C_{n+2}^m \varphi_{n+2}^m, \tag{27}$$

where

$$C_n^m = \frac{1 + \delta_{n|m|}}{2n} \tag{28}$$

$$\varphi_n^m = \frac{1}{2}(\beta_+)_{n-1}^{m+1} + \frac{1}{2}(\beta_-)_{n-1}^{m-1}, \tag{29}$$

and where $\delta_{nn'}$ is the Kronecker delta equal to 1 if $n = n'$ and 0 otherwise. Furthermore β_+ and β_- are the Zernike coefficients of $\frac{\partial W}{\partial x} \pm i\frac{\partial W}{\partial y}$ so that

$$\frac{\partial W}{\partial x} \pm i\frac{\partial W}{\partial y} = \sum_{m=-\infty}^{\infty} \sum_{n \in \eta_m} (\beta_\pm)_n^m Z_n^m. \tag{30}$$

Note that the LMS coefficients $\hat{\alpha}_n^m$ are analytically related to four β coefficients, namely $(\beta_+)_{n+1}^{m+1}$, $(\beta_+)_{n-1}^{m+1}$, $(\beta_-)_{n-1}^{m-1}$, and $(\beta_-)_{n+1}^{m-1}$. As a consequence of orthogonality of the Z_n^m in the right-hand side of Eq. (30), the LMS coefficients, once estimated, do not change anymore when the number of Zernike terms used in a finitized version of Eq. (30) is increased further.

The fact that $\hat{\alpha}_n^m$ is analytically related to β coefficients is desirable, because β-coefficients can directly be estimated (in a least-squares sense) from measurable quantities. This fit is made in the same way as the least-squares method. This means that also the complex Zernike circle polynomials need to be averaged over the lenslets. The fit to get the β coefficients, however, is done with a different basis than in the least-squares method to find the a-coefficients. The effects of this are discussed in the following section.

As a note for this method, when $n = |m|$, there will be non-existent combinations of n and m in Eq. (29). In that case the value of β will be set to 0. For instance, $\hat{\alpha}_1^1$ is among others dependent on $(\beta_+)_0^2$, which goes against the constraint $|m| \leq n$. To go from complex coefficient α_n^m to the real coefficient a_n^m, the relations in Eq. (16) can be used.

In [5], a seemingly different approach, based on vector polynomials, is used to express the wavefront Zernike coefficients in terms of wavefront slope data. However, we show in the appendix that this method is essentially equivalent to Janssen's algorithm in [1].

2.3.3. Wavefront reconstruction using Method A and Method B

The main difference between Method A and Method B for finding the coefficients is in how the fitting is done. Both methods use a least-squares fit using a geometry matrix, but the matrix elements are constructed differently. In Method A, the geometry matrix elements are evaluations of the average gradient of the real-valued Zernike circle polynomials over certain subdisks, while for Method B, the average of the complex-valued Zernike circle polynomials over the same windows [see Eq. (30)].

The gradients of the Zernike circle polynomials are known not to be orthogonal. This can cause problems called crosstalk when fitting the coefficients, especially when there are more aberrations present in the system than are being fit.

If $\mathbf{a}$ is an M-dimensional vector containing the coefficients of the aberrations present in the system, the slopes on the Shack-Hartmann sensor can be determined as $\mathbf{s} = G\mathbf{a}$, where $\mathbf{s}$ is an

OSA CONTINUUM

$2n_{\text{spot}}$ long vector containing the x- and y-displacement on the Shack-Hartmann sensor and G an $2n_{\text{spot}} \times J$ geometry matrix defined in Eq. (24).

The notation $Z_j = z_n^m$ is used since j, n, m are related [see Eq. (17)–(19)]. Z_0 is not included in the matrix since $Z_0 = z_0^0 = 1$, and its partial derivatives are equal to 0. Note that the first column contains the x- and y-derivatives of the first Zernike polynomial evaluated in all n_{spot} points. When a least-squares estimation of the coefficients $\hat{\mathbf{a}}$ (where the hat means to indicate that it is an estimated parameter) is made using less Zernike circle polynomials, up to Zernike polynomial $J < M$, crosstalk will occur. The estimator $\hat{\mathbf{a}}$ can be expressed as

$$\begin{aligned} \hat{\mathbf{a}} &\approx G_l^+ \mathbf{s} \\ \hat{\mathbf{a}} &\approx G_l^+ G \mathbf{a}, \end{aligned} \tag{31}$$

where G_l is the geometry matrix containing the columns of the first J Zernike circle polynomials. The estimator will estimate the lower-order values of the coefficients with influence of the higher-order values, because the matrix $G_l^+ G$ will not be an identity matrix.

When estimating the coefficient a_n^m, a higher-order aberration $a_{n'}^{m'}$ will influence the estimation if it is not accounted for in G_l (i.e., the single index of $a_{n'}^{m'}$ $j > J$) and if

$$\begin{aligned} \{(n, m, n', m') \in \mathbb{Z} \mid n \in \eta_m,\ m = m' \text{ or } m = m' \pm 2, \\ n' > n,\ n \geq m' + 2,\ n' \in \eta_{m'},\ m' \neq 0\}, \end{aligned} \tag{32}$$

or if

$$\begin{aligned} \{(n, m, n', m') \in \mathbb{Z} \mid n \in \eta_m,\ m = 0 \text{ or } m = 2, \\ n' > n,\ n \geq 2,\ n' \in \eta_{m'},\ m' = 0\}, \end{aligned} \tag{33}$$

where in both cases η_m and $\eta_{m'}$ are the sets of allowed values for n and n' dependent on m and m' such that Eqs. (5)–(7) are all met.

Because of this dependence of an expansion coefficient on the maximum degree of the system of equations, it is expected that methods such as the least-squares method will incorrectly estimate the coefficients when there are higher-order aberrations present that are not accounted for in the geometry matrix G_l. For Method B, the geometry matrix contains the Zernike circle polynomials themselves, and therefore it is not expected to present any crosstalk. This is experimentally verified and shown in this paper.

3. Experimental setup and methods

In Fig. 1, a schematic view of the experimental setup is shown. The phase of an uniform, collimated laser beam (HeNe laser) is modified by a spatial light modulator (SLM, Holoeye PLUTO-2-VIS-056) in such a way that known aberrations are added to the beam. In order to remove the zeroth- order reflected light from the SLM, a phase ramp is added to all SLM patterns so that the unaberrated spot is blocked by the iris after being focused by lens L3. The lenses L1 and L2 and L3 and L4 form two $4f$ systems that projects the phase imposed on the SLM onto the Shack Hartmann sensor (SHS). For each set of aberrations added in the SLM we obtain experimentally the raw data of the SHS, i.e., the image of the x-y displaced spots detected by the camera of the SHS. This data is used to find the average slopes of the wavefront as explained in Section 2.3.

Initially, the system is tested to identify its intrinsic aberrations (from lenses, SLM, etc). After calibration of the system response, theses aberrations are canceled by the SLM (flattening). Afterwards, additional and controlled aberrations are applied to the SLM with specific strength of a few Zernike coefficients. The raw measurements made with the SHS are further processed and used to compare the reconstruction methods A and B. In Algorithm 1 we describe all steps of the experiments.

OSA CONTINUUM

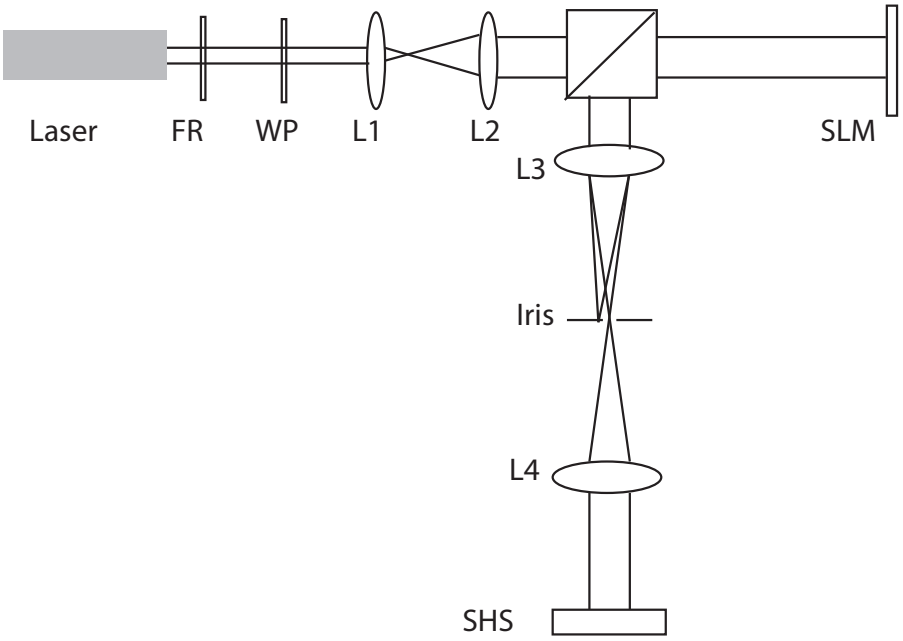

Fig. 1. Scheme of the experimental setup. FR is a Faraday isolator, WP is a halfwave plate, SLM is a spatial light modulator, SHS is the Shack Hartmann wavefront sensor. The zeroth order light of the SLM is blocked by the iris.

Algorithm 1 Complete measurement and comparison of Shack-Hartmann phase retrieval algorithms

1: Remove initial aberrations of the entire system
2: Add controlled aberration **a** to the SLM
3: Gather flat and aberrated wavefront Hartmannograms, i.e., the image of the spot pattern generated by the Shack-Hartmann sensor
4: Find the optimal center and radius position of the pupil for both methods separately such that the RMS value is minimized
5: Determine the Zernike coefficients using optimized center and radius of the pupil with a desired maximum degree. With the obtained Zernike coefficient values, reconstruct the wavefront for both Method A and Method B.
6: For each set of added aberrations, we also calculate the "reference" wavefront, which is the wavefront that is constructed with the nominal values of the added controlled aberrations
7: Calculate error between reference and estimate wavefront

3.1. Addressing a phase pattern to the SLM

3.1.1. Canceling the intrinsic aberrations

In order to be sure that only the aberration added on the SLM is measured, first all initial aberrations are eliminated. These aberrations can include alignment errors and the surface of the SLM itself, which might not be completely flat [6]. Removing the initial aberrations has been done by using the Shack-Hartmann sensor and retrieving the wavefront using the least-squares method (Method A).

OSA CONTINUUM

3.1.2. Calibration of the SLM response

The way a phase pattern is addressed to the SLM is as follows. First, an aperture on the SLM is defined. In the current research this is a circular aperture, but the same method is valid if an annular aperture is used. All pixels i within this aperture are used. A "Zernike matrix" can be set up, such that for each pixel i within the aperture the value of all necessary Zernike circle polynomials can be computed. In matrix form this would be

$$Z = \begin{bmatrix} Z_1|_1 & Z_2|_1 & \cdots & Z_J|_1 \\ Z_1|_2 & Z_2|_2 & \cdots & Z_J|_2 \\ \vdots & \vdots & \ddots & \vdots \\ Z_1|_I & Z_2|_I & \cdots & Z_J|_I \end{bmatrix}, \tag{34}$$

where J is the total number of Zernike circle polynomials evaluated, and I is the total number of pixels within the aperture. Note that here only the real valued polynomials were added in the wavefront, since this is compatible with the experiment. Due to the cyclic nature of the phase pattern and the limits of the SLM, the phase difference assigned to the SLM should be between 0 and 2π. This can be done using the modulo (or mod) operation: $a \bmod n = a - n\lfloor a/n \rfloor$.

If $\mathbf{p}$ is a vector containing the values of the individual pixels of the SLM, it can be constructed from the vector $\mathbf{a}$ containing the coefficients for the to-be-added aberration by

$$\mathbf{p} = (Z\mathbf{a}) \bmod 2\pi. \tag{35}$$

Therefore, in the case of updating the SLM pattern we use

$$\mathbf{p}_{new} = (\mathbf{p}_{old} - \gamma Z\hat{\mathbf{a}}) \bmod 2\pi, \tag{36}$$

where γ is the gamma function of the SLM, calibrated for the used wavelength.

After the correction is done and the wavefront from the SLM is flattened, the phase pattern of the "flat" phase is saved. To this phase pattern, a phase ramp is added to separate the 0^{th} and 1^{st} order of the SLM together with the desired aberration, in the same fashion as the correction is added. After this, it is necessary to check if the maximum phase change over 4 pixels is not being exceeded. Such a check is necessary to avoid aliasing of the SLM-phase.

In the current research, this aliasing constraint is simplified to the constraint that the difference between two neighboring pixels should not exceed 0.5π. This is evaluated by letting $p_{i,j}$ be the value of the pixel located at position i, j on the SLM electrode matrix. Then, first two matrices are constructed:

$$\begin{aligned} \Delta p_x &= p_{i,j+1} - p_{i,j}, \\ \Delta p_y &= p_{i+1,j} - p_{i,j}. \end{aligned} \tag{37}$$

Afterward, the element-wise minimum is taken between Δp and $2\pi - |\Delta p|$ for both x and y, in order to account for the modulated phase. If any of the values of this piecewise minimum is above 0.5π, it is said to break the aliasing constraint.

3.2. Constructing the matrix G

Recalling from the theory described in the previous section, in order to recover the coefficients describing the aberrated wavefront, $\mathbf{s}$ and G have to be constructed. Here, $\mathbf{s}$ relies on the two Hartmannograms, one of the flat wavefront and one of the aberrated wavefront from the SLM, and the radius r_{SH} of beam hitting the Shack-Hartmann sensor. Two G matrices will then be constructed for the respective Methods A and B.

OSA CONTINUUM

3.2.1. Initial wavefront reconstruction pupil

To construct G, it is necessary to average the gradients of the Zernike circle polynomials for method A or the complex valued Zernike circle polynomials for method B. The window over which the polynomial has to be averaged can be seen as a scaled version of the lenslet, scaled so that all illuminated lenslets fit the unit disc. In order to compute the Zernike circle polynomials in these windows, the center position $\mathbf{c}$ and the radius r_{SH} of the beam on the Shack-Hartmann sensor have to be known. The window size is estimated by the average distance between the nearest neighbor spots on the Shack-Hartmann sensor. The center position $\mathbf{c}$ follows the average position of all the midpoints of the spots on the Shack-Hartmann sensor. The radius is estimated by calculating the length between the center and the furthest spot from the center.

3.2.2. Wavefront pupil boundaries optimization (with RMS optimization)

After this first estimation of the center position and radius is made, a better estimation can be found by optimization. For this optimization, an error has been defined that can be minimized. In this research, a root-mean-square, or RMS type error is chosen. Using the known added aberration and the measured aberration coefficients, an RMS error can be defined. To this purpose, a new Zernike matrix similar to Eq. (34) is constructed, this time with N points on a grid within the unit disc. The reference phase $\mathbf{p}_{\mathrm{ref}}$ and the recovered phase $\mathbf{p}_{\mathrm{rec}}$ can be constructed as

$$\mathbf{p} = Z\mathbf{a}, \tag{38}$$

where for $\mathbf{p}_{\mathrm{ref}}$, the reference vector $\mathbf{a}_{\mathrm{ref}}$ is used, while for $\mathbf{p}_{\mathrm{rec}}$ the estimated coefficient vector $\hat{\mathbf{a}}$ is used. The RMS error is then defined

$$\epsilon = \left\| \frac{\mathbf{p}_{\mathrm{ref}} - \mathbf{p}_{\mathrm{rec}}}{N} \right\|_2, \tag{39}$$

where $\|\mathbf{x}\|_2$ is the Euclidean vector norm of vector $\mathbf{x}$. This RMS error is then minimized for center position and radius of the beam on the Hartmannogram. A limited memory bound Broyden-Fletcher-Goldfarb-Shanno (L-BFGS-B) minimization algorithm is applied to find the center position and radius [7]. The termination conditions for this optimization are:

$$\frac{f^k - f^{k+1}}{\max\{|f^k|, |f^{k+1}|, 1\}} \leq 10^{-5}, \tag{40}$$

$$\max\{|\mathrm{proj}(g_i)| \; i = 1, ..., n\} \leq 10^{-5}, \tag{41}$$

$$k \geq 10^3, \tag{42}$$

where f^k is the value of the RMS error of the k^{th} iteration of the minimization algorithm, and $\mathrm{proj}(g_i)$ is i^{th} component of the projected gradient where n projections are made. If any of these statements were true, the optimization was terminated.

After looking at the RMS error landscapes, it was found that not every minimum found was a global minimum. If this was the case, the global minimum coordinates were estimated using the RMS landscape graphs, and a brute force optimization was run around those coordinates. This brute force optimization calculates the RMS error value in a grid of points. From the coordinates with the lowest RMS error, a new downhill simplex minimization algorithm is started. This way the global minimum was attempted to be found, and the optimal center and radius positions were determined.

These optimized parameters can then be used to find the RMS value fitting any degree of Zernike circle polynomials, and can also be used to determine the error landscape by calculating the RMS values when the center and radius differ slightly from the optimal value.

OSA CONTINUUM

3.2.3. Application to the Zernike coefficients determination

Following the rule of thumb defined in [8], the amount of Zernike circle polynomials that can be fit given k spots on the Shack-Hartmann sensor is $k/3$. In the current research, there were about 144 spots on the Shack-Hartmann sensor, so, in this way, the optimum amount of polynomials was determined to be 45, i.e., Zernike expansion with a maximum degree of eight. In Figs. 5 and 6 other fitting orders are shown for some Zernike coefficients, and as one can see, this fitting order is good for all cases.

For method A, the amount of polynomials fit is equal to the amount of retrieved coefficients, while for Method B, if a fit is made with up to maximum degree of eight, the first seven degrees can be retrieved.

It should be noted that this optimization can take long due to the fact that the geometry matrix needs to be calculated in every iteration, as the values in the matrix depend on the center position and radius.

4. Results

The two methods have been implemented for three different test cases: (a) single Zernike aberrations, (b) a combination of three aberrations, assumed not to show crosstalk, and (c) four cases of aberrations where crosstalk is present. The aberrations are listed in Table 1. We remind that the values of the Zernike coefficients presented in the table are the nominal values that are added to the SLM, in addition to the intrinsic aberrations of the system. These nominal values have been used to calculate the reference wavefront in order to compare it with the estimated wavefront from experimental data using Method A and B. In Figs. 5 and 6 one can see the nominal values (dotted red line) and the values that are estimated using Method A (triangles) and Method B (dots) for several maximum fitted Zernike degree when some of the added Zernike coefficients (Experiment sub_zerns_1 and sub_zers_3).

The reconstructions using the recovered coefficients are shown in Figs. 2–4. For these reconstructions, the first eight degrees in the Zernike expansion were used. The left column shows the reference wavefront that is constructed by using the nominal aberrations that have been added on the SLM, while the middle and right column show the wavefront reconstruction starting from the Shack Hartmann data and using our own implementation of the least-squares method (Method A) and Method B, respectively, as explained in the previous two sections. The RMS error, defined in Eq. (39) is the error between the reconstructed wavefronts using Method A and B (middle and right columns) and the reference (left column) is indicated below each row in the figures. From these results, it can be seen that the Method B compares very well with Method A and the input reference wavefront.

Based on the errors in the reconstruction found in Figs. 2, 3, and 4, one can see that both methods provide comparable accuracy in reconstructing the wavefront. However, it is known (and also observed here) that Method A presents crosstalk of coefficients when less Zernike powers are fit than there are aberrations present in the system. In the following, we show experimentally that this is not the case for Method B.

Using the same center position and radius, different amount of Zernike degrees can be fit in order to see the convergence behavior of gathered coefficients. Due to the fact that there are only 1, 2, or 3 Zernike modes present in the specific Zernike experiment, the convergence behavior of these experiments can be visualized.

Based on analysis of the aberrations listed in Table 1, the single Zernike experiments `5_1` and `6_4` and 3 random Zernike experiments `3_zerns_1` and `3_zerns_3`, one can see in Figs. 2 and 3 that the coefficients seem to be measured independently from each other, and there is a little difference between the two methods. However, looking at the subsequent Zernike experiments `sub_zerns_1` and `sub_zerns_3`, a large difference can be seen between the two reconstruction methods. The initial guesses of Method A over- or underestimate the presence

OSA CONTINUUM

Table 1. Coefficients Used in Specific Zernike Experiments, Rounded Off to 3 Significant Numbers

Experiment name	Values of the applied Zernike coefficients (Units of waves)		
5_1	a_5^1		
	0.750		
5_5	a_5^5		
	1.250		
6_2	a_6^2		
	0.500		
6_4	a_6^4		
	1.000		
6_6	a_6^6		
	2.000		
3_zerns_1	a_2^0	a_4^4	a_6^{-2}
	2.000	1.500	0.500
3_zerns_2	a_2^0	a_3^{-1}	a_5^5
	2.000	3.000	1.500
3_zerns_3	a_2^{-2}	a_4^0	a_6^{-4}
	2.000	1.500	0.750
sub_zerns_1	a_2^0	a_4^0	a_6^0
	4.000	1.500	0.750
sub_zerns_2	a_1^1	a_3^1	a_5^1
	4.000	1.500	0.750
sub_zerns_3	a_3^3	a_5^3	
	2.500	0.750	
sub_zerns_4	a_2^2	a_4^2	a_6^2
	2.500	1.000	0.500

of the aberration when not enough powers are fit. From Fig. 5, it can be seen that defocus is overestimated by more than 50% until 4 degrees are fit. Until a maximum degree of four, the spherical aberration coefficient a_4^0 is overestimated. All values seem to be within normal range when fitting a maximum degree of eight. The same can be seen in Fig. 6, where the coefficient a_3^3 is overestimated at maximum degree of three and four. Method B does not present this over-estimation. This behavior is also observed in the sets sub_zerns_2 and sub_zerns_4 (not shown in the figures). We also note that in the case of sub_zerns_3, both methods present some overestimation for large fitted degree. We don't know the exact reason for that, but

OSA CONTINUUM

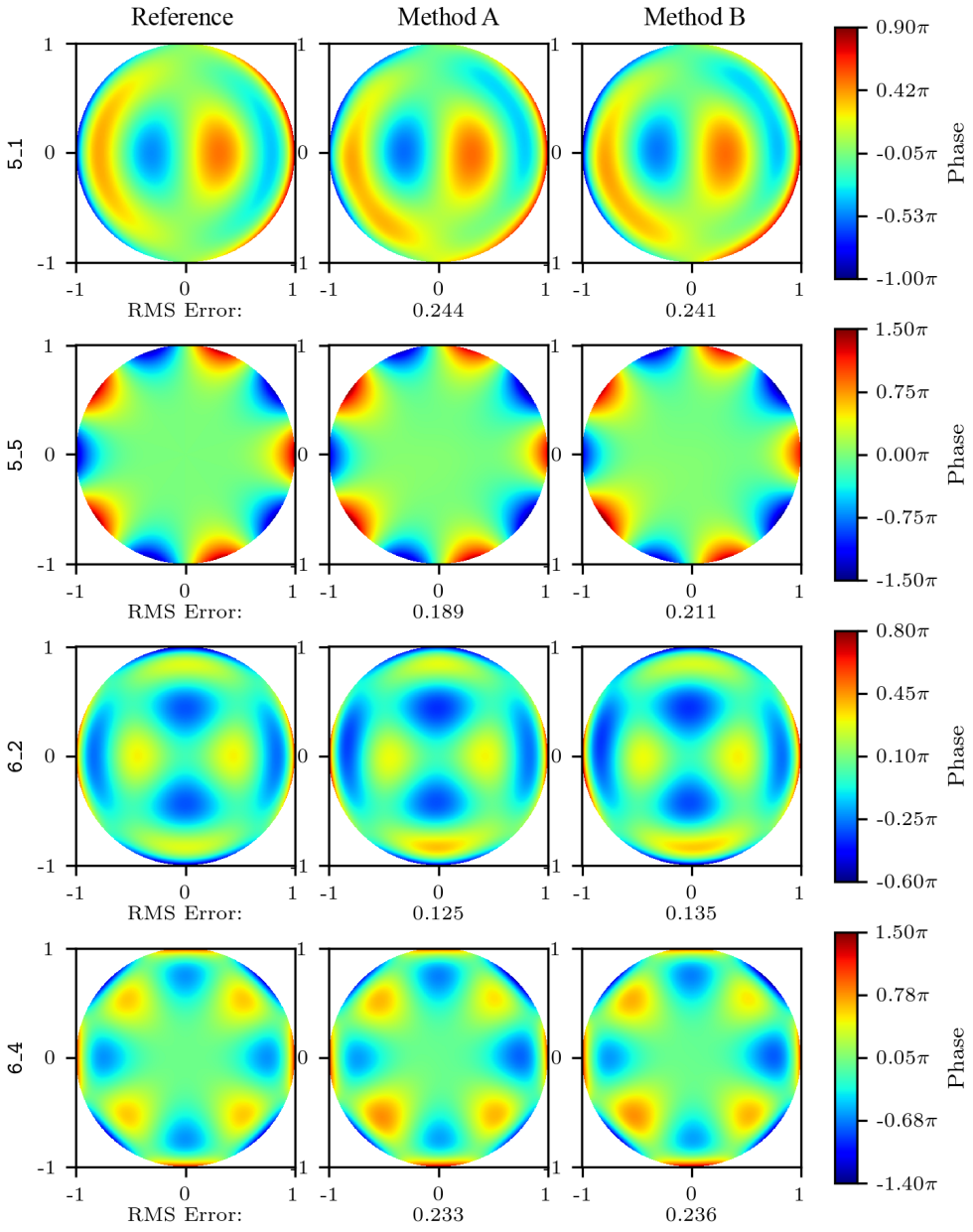

Fig. 2. Reference wavefronts (left column) for the first four experiments of Table 1. Reconstructed wavefronts from experimental data using Method A (middle column) and Method B (right column). The RMS errors under each row have been obtained by comparing the reconstructed (middle and right column) with reference wavefront (left column). For the reconstructions, the first 8 degrees in the Zernike expansion has been used.

we observe that for this experiment, the RMS error between both methods and the reference wavefront was higher than for the case of `sub_zerns_1`, `sub_zerns_2` and `sub_zerns_4` (see Fig. 4).

OSA CONTINUUM

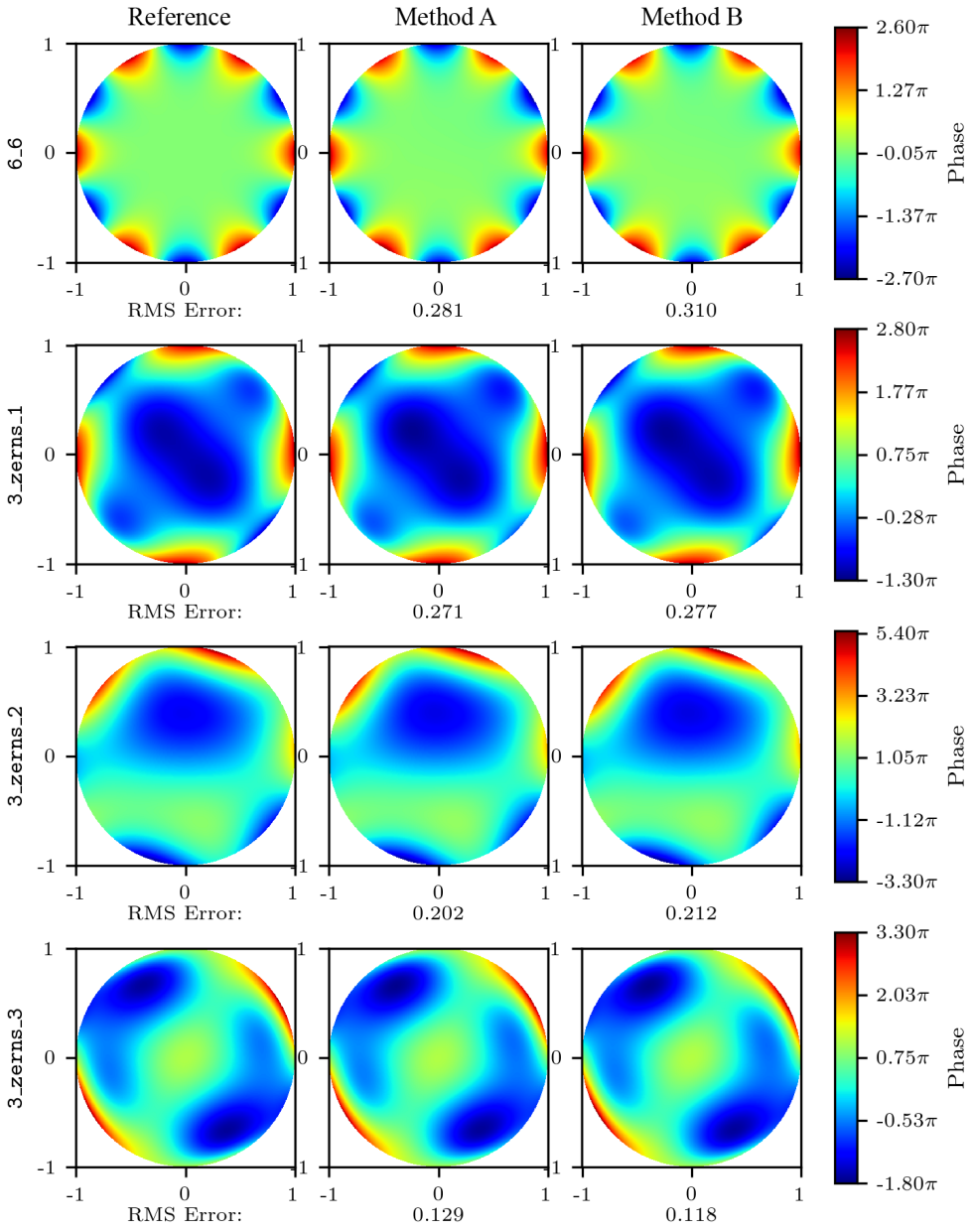

Fig. 3. Reference wavefronts (left column) for the second four experiments of Table 1. Reconstructed wavefronts from experimental data using Method A (middle column) and Method B (right column). The RMS errors under each row have been obtained by comparing the reconstructed (middle and right column) with reference wavefront (left column). For the reconstructions, the first 8 degrees in the Zernike expansion has been used.

OSA CONTINUUM

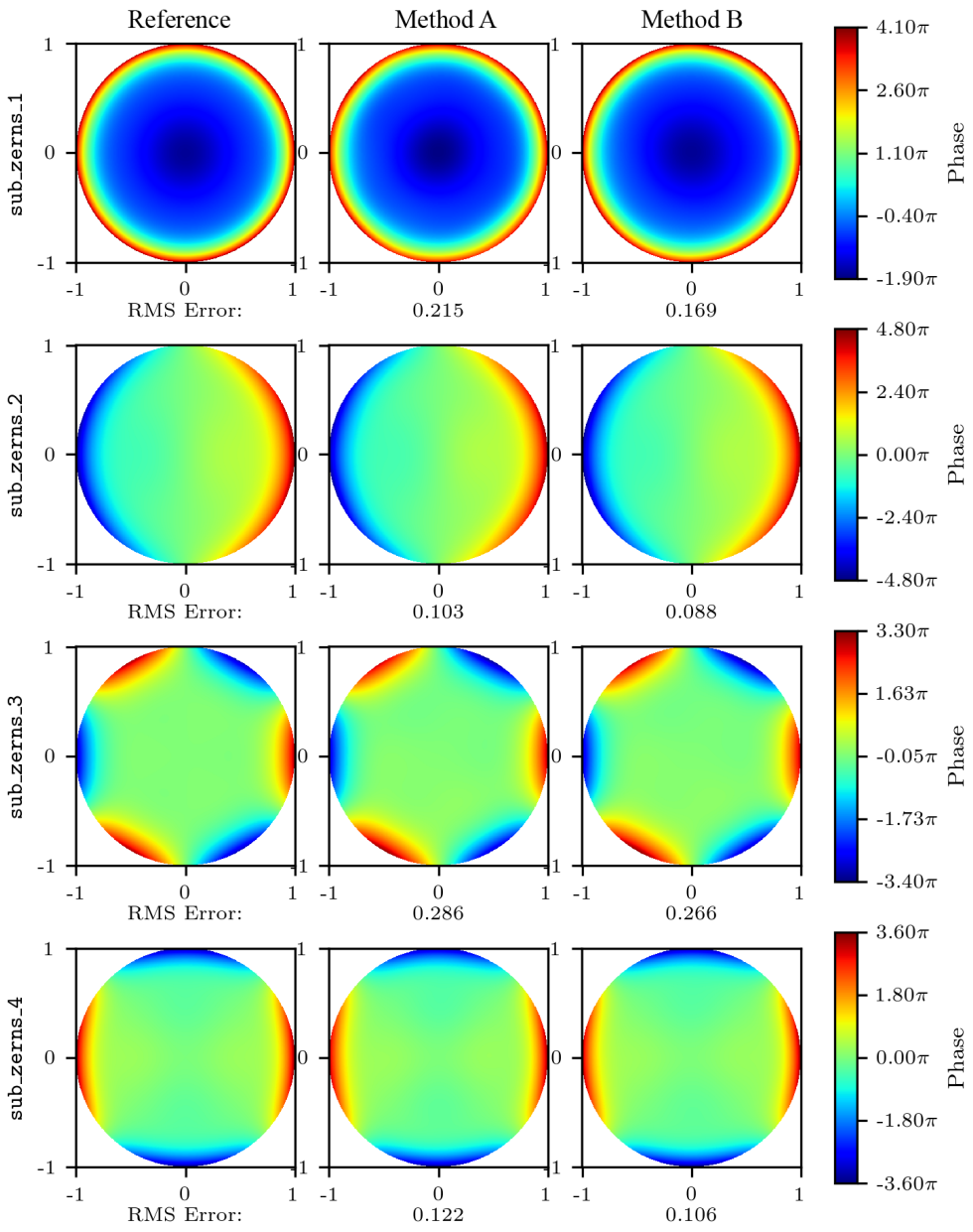

Fig. 4. Reference wavefronts (left column) for the last four experiments of Table 1. Reconstructed wavefronts from experimental data using Method A (middle column) and Method B (right column). The RMS errors under each row have been obtained by comparing the reconstructed (middle and right column) with reference wavefront (left column). For the reconstructions, the first 8 degrees in the Zernike expansion has been used.

5. Conclusion and outlook

In conclusion, we have implemented a new wavefront reconstruction method (Method B) for Shack-Hartmann sensors based on Zernike expansion of derivatives of the Zernike circle

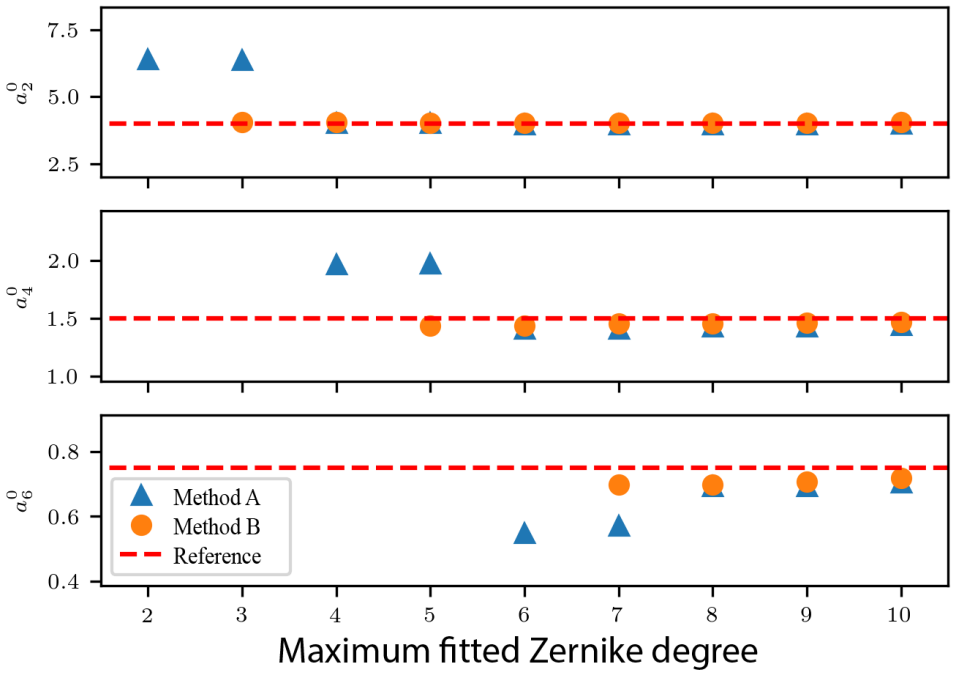

Fig. 5. Convergence of coefficients for the `sub_zerns_1` experiment.

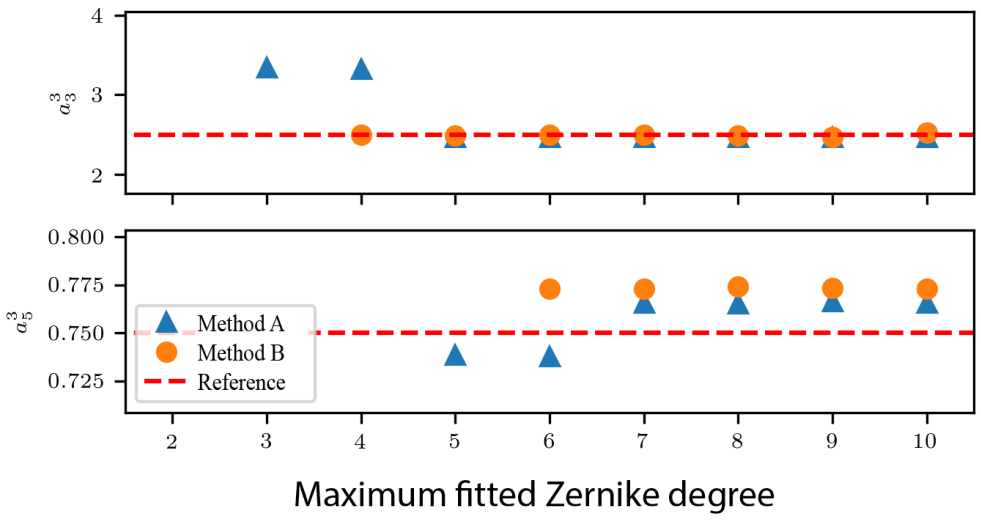

Fig. 6. Convergence of coefficients for the `sub_zerns_3` experiment.

polynomials that was introduced in [1]. We have shown with experiments that Method B is advantageous as compared to other known methods such as a least-squares method (Method

OSA CONTINUUM

A) due to the lack of cross-talk between the coefficients. We have also compared Method B to Method A for reconstructing the wavefront using a Shack-Hartmann sensor for some sets of aberrations.

Based on the error values of the reconstruction with optimal amount of Zernike coefficients, the quality of the fit is in general similar for Method A and Method B. When less than the optimal Zernike powers were fit, it was seen that Method B estimates the coefficients more accurately. For the single Zernike aberration case, it was shown that only Method A shows cross-coupling of higher order aberrations, while Method B does not.

As to computational load, we may point out that these are of comparable order, with Method B somewhat more demanding due to the complex arithmetics and the alignment needed for compatibility with the ANSI-format of the Shack-Hartmann sensor.

Finally, we would like to point out that there are also alternative expansions to treat this problem, namely the use of eigenfunctions of the Laplacian with Neumann boundary conditions [9, 10] that do have orthogonal gradients. It would be interesting as future work to compare this method with Method A and Method B considered in the present paper.

Appendix: Relation with orthogonal vector polynomials

In the course of our investigations, we became aware of a seemingly different approach to obtain the wavefront aberration coefficients from wavefront derivative data. We shall describe and relate this approach to our method in the framework and notations of [1], so that the Zernike circle polynomials are unnormalized and have exponential azimuthal dependence facilitating mathematical developments. The reader will have no particular problems in reformulating the main results of this appendix in terms of the ANSI-style circle polynomials using Eqs. (27)–(29). Thus, we have for integer n and m such that $n - |m|$ is even and non-negative:

$$Z_n^m(\nu, \mu) \equiv Z_n^m(\rho, \theta) = R_n^{|m|}(\rho) e^{im\theta}, \tag{43}$$

with real ν, μ such that $\nu^2 + \mu^2 \leq 1$ and

$$\nu + i\mu = \rho e^{i\theta};\ \nu = \rho\cos\theta;\ \mu = \rho\sin\theta, \tag{44}$$

and the radial polynomials $R_n^{|m|}(\rho)$ given by Eq. (11). We now sketch the approach in [5], where the notations and conventions differ from our present analysis. The approach [5] uses the notion of vector polynomials

$$\underline{G}_n^m = \underline{G}_n^m(\rho, \theta) = (G_{n,1}^m(\rho, \theta), G_{n,2}^m(\rho, \theta)) \in \mathbb{C}^2 \tag{45}$$

for integer n, m such that $n - |m|$ is even and non-negative with $n \neq 0$, that satisfy

$$\frac{1}{\pi}\int_0^1 \int_0^{2\pi} \nabla Z_n^m(\rho, \theta).\underline{G}_{n'}^{m'*}(\rho, \theta)\rho d\rho d\theta = \delta_{m,m'}\delta_{n,n'} \tag{46}$$

for integer n, m, n', m' such that $n - |m|$ and $n' - |m'|$ are even and non-negative while $n \neq 0 \neq n'$. In Eq. (46) we denote for $(z, w) \in \mathbb{C}^2$ and $(g, h) \in \mathbb{C}^2$

$$(z, w) \cdot (g, h) = zg + wh, \tag{47}$$

the $*$ denotes complex conjugation, and

$$(\nabla Z_n^m)(\nu, \mu) = \left(\frac{\partial Z_n^m}{\partial \nu}(\nu, \mu), \frac{\partial Z_n^m}{\partial \mu}(\nu, \mu)\right) \in \mathbb{C}^2. \tag{48}$$

OSA CONTINUUM

With α_n^m the Zernike coefficients of W (to be found) according to Eq. (13), where, due to orthogonality [see Eq. (12)],

$$\alpha_n^m = \frac{n+1}{\pi}\int_0^1\int_0^{2\pi} W(\rho,\theta)Z_n^{m*}(\rho,\theta)\rho d\rho d\theta, \tag{49}$$

we have

$$\nabla W = \sum_{n,m}\alpha_n^m \nabla Z_n^m, \tag{50}$$

and

$$\alpha_n^m = \frac{1}{\pi}\int_0^1\int_0^{2\pi} \nabla W(\rho,\theta)\underline{G}_n^{m*}(\rho,\theta)\rho d\rho d\theta. \tag{51}$$

In [5], Section 2.1, the condition Eq. (46) is elaborated using Green's theorem, where it should be noted that in [5] the attention is restricted to the cases $n = n', m = m'$ in Eq. (46). It is shown that the problem of satisfying Eq. (46), with $n = n'$, $m = m'$ is solved when we can find $\underline{G}_n^m$ such that on the pupil

$$\nabla.\underline{G}_n^m = \frac{\partial G_{n,1}^m}{\partial \nu} + \frac{\partial G_{n,2}^m}{\partial \mu} = -(n+1)Z_n^m, \tag{52}$$

while the boundary condition

$$G_{n,1}^m(\rho=1,\theta)\cos\theta + G_{n,2}^m(\rho=1,\theta)\sin\theta = 0 \tag{53}$$

should be satisfied at the rim $\rho^2 = \nu^2 + \mu^2 = 1$ of the pupil. Next, in [5], Section 2.2.1, the $\underline{G}_n^m$ are required to be irrotational, meaning that the $\nabla \times \underline{G}_n^m = 0$, because of considerations of minimal noise propagation. This condition of irrotationality is satisfied when there is a scalar function U_n^m on the pupil such that

$$\underline{G}_n^m = \nabla U_n^m. \tag{54}$$

Combining Eq. (54) and Eq. (52), we see that we want U_n^m to satisfy

$$\nabla^2 U_n^m = \Delta U_n^m = -(n+1)Z_n^m \tag{55}$$

while the boundary condition Eq. (53)

$$\frac{\partial U_n^m}{\partial \nu}(\rho=1,\theta)\cos\theta + \frac{\partial U_n^m}{\partial \mu}(\rho=1,\theta)\sin\theta = 0 \tag{56}$$

should hold on the rim $\rho = 1$ of the pupil. In [5] Section 2.2.2–2.2.3, the U_n^m are found by writing the condition Eq. (55) in polar coordinates, with separate consideration of the cases $m = 0$ and $m \neq 0$, and explicitly using the series representation of radial polynomials. This yields the two components of G_n^m in trigonometric polynomial form, for which it can be shown that the boundary condition Eq. (53) is satisfied as well. In [1], Section 4, it is shown that Eq. (55) has a solution

$$U_n^m = -\left[\frac{Z_{n+2}^m}{4(n+2)} - \frac{(n+1)Z_n^m}{2n(n+2)} + \frac{Z_{n-2}^m}{4n}\right]. \tag{57}$$

We shall verify below that, when $n = |m| + 2, |m| + 4, \ldots$, this U_n^m also satisfies the boundary condition Eq. (56), and that a concise formula for $\underline{G}_n^m = \nabla U_n^m$ in terms of Zernike circle polynomials results. For the case that $n = |m|$, we shall show that

$$U_n^m = -\left[\frac{Z_{n+2}^m}{4(n+2)} - \frac{(3n+4)Z_n^m}{4n(n+2)}\right], n = |m|, \tag{58}$$

OSA CONTINUUM

satisfies Eqs. (55) and (56), and we find a concise formula in terms of the Zernike circle polynomials for $\underline{G}^m_{|m|}$ as well. This then can be used to show that α^m_n of Eq. (51) actually coincide with the LMS estimator found in [1], Section 3. The trigonometric/polynomial solution form of $\underline{G}^m_n$ found in [5] involves the coefficients in the series representation of $R^{|m|}_n(\rho)$ in Eq. (58), and these become awkward to use when the degree becomes large (n should be limited to ≤ 44 when using double precision). This problem is virtually absent when the representation of the $\underline{G}^m_n$ in terms of the Zernike circle polynomials is used since there are nowadays several methods for reliably computing (the radial parts of the) Zernike circle polynomials of arbitrary large degree n and azimuthal order m, [11], [12].

We shall now show for $m = 1, 2, \ldots$ and $n = m + 2, m + 4, \ldots$ that U^m_n of 57 satisfies Eq. (56). Recalling the convention in [1] that any $Z^{m'}_{n'}$ with $|m'| > n'$ is set to 0, we have

$$\left(\frac{\partial}{\partial \nu} \pm i\frac{\partial}{\partial \mu}\right) Z^{m'}_{n'}(\nu, \mu) = 2 \sum_{l=0}^{\frac{n'-|m'|}{2}} (n' - 2l) Z^{m'\pm 1}_{n'-1-2l}, \tag{59}$$

see [1], Eq. (13). Using Eq. (59) with $\nu + i\mu = e^{i\theta}$, noting that

$$Z^{m\pm 1}_{n-1-2l}(1, \theta) = e^{i(m\pm 1)\theta},\ l = 0, 1, \ldots \frac{1}{2}(n - |m| - 1), \tag{60}$$

while (as $m > 0$)

$$Z^{m-1}_{m-1}(1, \theta) = e^{i(m-1)\theta},\ Z^{m+1}_{m-1}(1, \theta) = 0, \tag{61}$$

we have

$$\left(\frac{\partial}{\partial \nu} + i\frac{\partial}{\partial \mu}\right) Z^m_n = 2 \sum_{l=0}^{\frac{n-m}{2}} (n - 2l) e^{i(m+1)\theta} - 2m e^{i(m+1)\theta}, \tag{62}$$

$$\left(\frac{\partial}{\partial \nu} - i\frac{\partial}{\partial \mu}\right) Z^m_n = 2 \sum_{l=0}^{\frac{n-m}{2}} (n - 2l) e^{i(m-1)\theta}. \tag{63}$$

Adding and subtracting Eqs. (62) and (63) from one another then gives

$$\frac{\partial Z^m_n}{\partial \nu} = 2e^{im\theta} \cos\theta \sum_{l=0}^{\frac{n-m}{2}} (n - 2l) - m e^{i(m+1)\theta}, \tag{64}$$

$$\frac{\partial Z^m_n}{\partial \mu} = 2e^{im\theta} \sin\theta \sum_{l=0}^{\frac{n-m}{2}} (n - 2l) - \frac{1}{i} m e^{i(m+1)\theta}. \tag{65}$$

We also observe that

$$\sum_{l=0}^{\frac{n-m}{2}} (n - 2l) = \frac{n + m}{2} \left(\frac{n + m}{2} + 1\right). \tag{66}$$

Then, from Eqs. (57), (64) and (66), we get

$$\begin{aligned} &\frac{\partial U^m_n}{\partial \nu}(\rho = 1, \theta) = \\ &= -\left[\frac{1}{4(n + 2)} \frac{\partial Z^m_n}{\partial \nu} - \frac{n + 1}{2n(n + 2)} \frac{\partial Z^m_n}{\partial \nu} + \frac{1}{4n} \frac{\partial Z^m_{n-2}}{\partial \nu}\right] \\ &= -2D^m_n e^{im\theta} \cos\theta + m E_n e^{i(m+1)\theta}, \end{aligned} \tag{67}$$

OSA CONTINUUM

where

$$D_n^m = \frac{\frac{n+2+m}{2}(\frac{n+2-m}{2}+1)}{4(n+2)} - \frac{(n+1)\frac{n+m}{2}(\frac{n-m}{2}+1)}{2n(n+2)} + \frac{\frac{n-2+m}{2}(\frac{n-2-m}{2}+1)}{4n} = 0, \tag{68}$$

and

$$E_n = \frac{1}{4(n+2)} - \frac{n+1}{2n(n+2)} + \frac{1}{4n} = 0. \tag{69}$$

Hence, $(\partial U_n^m/\partial\nu)(\rho,\theta) = 0$ for $\rho = 1$ and, similarly $(\partial U_n^m/\partial\mu)(\rho,\theta) = 0$ for $\rho = 1$, and this implies that Eq. (56) holds. This handles the case that $m = 1, 2, \ldots$ and $n = m+2, m+4, \ldots$. The case the that $m = -1, -2, \ldots$ and $n = |m|+2, |m|+4, \ldots$ follows from the case already handled by using that

$$Z_{n'}^{m'}(\nu,\mu) = (Z_{n'}^{-m'}(\nu,\mu))^*. \tag{70}$$

The case that $m = 0$ and $n = 2, 4, \ldots$ can be proved in the same way as the case $m = 1, 2, \ldots$ and $n = m+2, m+4, \ldots$, where now in Eqs. (64) and (65) the terms involving $e^{i(m+1)\theta}$ disappear.

We shall next show that for $n = |m|+2, |m|+4, \ldots$

$$\frac{\partial U_n^m}{\partial\nu} = -\frac{1}{4}\left[Z_{n+1}^{m+1} + Z_{n+1}^{m-1} - Z_{n-1}^{m+1} - Z_{n-1}^{m-1}\right], \tag{71}$$

$$\frac{\partial U_n^m}{\partial\mu} = -\frac{1}{4i}\left[Z_{n+1}^{m+1} - Z_{n+1}^{m-1} - Z_{n-1}^{m+1} + Z_{n-1}^{m-1}\right]. \tag{72}$$

Then Eqs. (71) and (72) yield the announced concise expression in terms of the Zernike circle polynomials for

$$\underline{G}_n^m = \left(\frac{\partial U_n^m}{\partial\nu}, \frac{\partial U_n^m}{\partial\mu}\right). \tag{73}$$

Using Eqs. (59) and (57), we have that

$$\begin{aligned}\frac{\partial U_n^m}{\partial\nu} = -[&\frac{1}{4(n+2)}\sum_{l=0}^{\frac{n+2-|m|}{2}}(n+2-2l)(Z_{n+1-2l}^{m+1} + Z_{n+1-2l}^{m-1})\\ &- \frac{n+1}{2n(n+2)}\sum_{l=0}^{\frac{n-|m|}{2}}(n-2l)(Z_{n-1-2l}^{m+1} + Z_{n-1-2l}^{m-1})\\ &+ \frac{1}{4n}\sum_{l=0}^{\frac{n-2-|m|}{2}}(n-2-2l)(Z_{n-3-2l}^{m+1} + Z_{n-3-2l}^{m-1})]\end{aligned} \tag{74}$$

We now observe that the three series in the last member of Eq. (74) have the same terms, except that the second series omits the term $l = 0$ from the first series and the third series omits the terms with $l = 0, 1$ from the first series. Using Eq. (69), we see that the terms in the first series with $l = 2, 3, \ldots (n+2-|m|)/2$ are canceled, and we get

$$\begin{aligned}\frac{\partial U_n^m}{\partial\nu} &= -[\frac{1}{4(n+2)}((n+2)(Z_{n+1}^{m+1} + Z_{n+1}^{m-1}) + n(Z_{n-1}^{m+1} + Z_{n-1}^{m-1}))\\ &\quad - \frac{n+1}{2n(n+2)}n(Z_{n-1}^{m+1} + Z_{n-1}^{m-1})]\\ &= -\left[\frac{1}{4}(Z_{n+1}^{m+1} + Z_{n-1}^{m+1}) - \frac{1}{4}(Z_{n-1}^{m+1} + Z_{n-1}^{m-1})\right],\end{aligned} \tag{75}$$

and this is Eq. (71). In a similar fashion, we get Eq. (72).

OSA CONTINUUM

We now consider the case that $n = m = 1, 2, \ldots$. Then $Z^m_{n-2} = 0$ and $\Delta Z^m_n = 0$, and so it follows from Eq. (64) that $\Delta U = -(m+1)Z^m_m$ holds for any U of the form

$$U = -\left[\frac{Z^m_{m+2}}{4(m+2)} - CZ^m_m\right]. \tag{76}$$

We shall determine C such that Eq. (56) holds for $n = m$. Now

$$Z^m_m(\rho,\theta) = \rho^m e^{im\theta}, \quad Z^m_{m+2}(\rho,\theta) = ((m+2)\rho^{m+2} - (m+1)\rho^m)e^{im\theta}, \tag{77}$$

and so the condition Eq. (56) for $n = m$ and $\theta = 0$ yields

$$C = \frac{3m+4}{4m(m+2)}. \tag{78}$$

It can be shown that with this value of C the U of Eq. (76) satisfies Eq. (56) also for $\theta \neq 0$. This handles the case that $n = m = 1, 2, \ldots$. The case $n = m = 0$ is non-existent, and the case with $n = -m = 1, 2, \ldots$ follows from the case already handled by complex conjugation. We now also compute for this U, using Eq. (59), and $m > 0$

$$\begin{aligned}\frac{\partial U}{\partial \nu} &= -\left[\frac{1}{4(m+2)}\frac{\partial Z^m_{m+2}}{\partial \nu} - \frac{3m+4}{4m(m+2)}\frac{\partial Z^m_m}{\partial \nu}\right] \\ &= -\left[\frac{1}{4}Z^{m+1}_{m+1} + \frac{1}{4}Z^{m-1}_{m+1} + \frac{m}{4(m+2)}Z^{m-1}_{m-1} - \frac{3m+4}{4m(m+2)}mZ^{m-1}_{m-1}\right] \\ &= -\left[\frac{1}{4}Z^{m+1}_{m+1} + \frac{1}{4}Z^{m-1}_{m+1} - \frac{1}{2}Z^{m-1}_{m-1}\right],\end{aligned} \tag{79}$$

and similarly

$$\frac{\partial U}{\partial \mu} = -\frac{1}{i}\left[\frac{1}{4}Z^{m+1}_{m+1} - \frac{1}{4}Z^{m-1}_{m+1} + \frac{1}{2}Z^{m-1}_{m-1}\right]. \tag{80}$$

The results Eqs. (79) and (80) continue to hold when $m < 0$ and $n = |m|$. Hence, also in this case we get a concise result for $\underline{G}^m_{|m|}$ as in Eqs. (71)–(73).

We finally show that the α^m_n obtained in [1], Section 3 coincide with the α^m_n of Eq. (51). Observe that the vector polynomials $\underline{G}^m_m$ in Eq. (51) were derived in [5] from the conditions Eq. (46) and Eq. (53) under the assumption that they have vanishing curls, whereas the LMS estimator found in [1], section 3 has been derived under the condition that a natural mean-square error functional involving expansion coefficients is minimized. This LMS estimate uses the expansion coefficients $(\beta_\pm)^m_n$ in

$$\frac{\partial W}{\partial \nu} \pm i\frac{\partial W}{\partial \mu} = \sum_{n,m} \beta_{\pm n}{}^m Z^m_n. \tag{81}$$

That is, we have for $m \neq 0$

$$\alpha^m_n = C^m_n \varphi^m_n - C^m_{n+2}\varphi^m_{n+2}, n = |m|, |m|+2, \ldots, \tag{82}$$

with

$$\varphi^m_n = \frac{1}{2}(\beta_+)^{m+1}_{n+1} + (\beta_-)^{m-1}_{n-1}, \tag{83}$$

$$C^m_{|m|} = \frac{1}{|m|},\ C^m_n = \frac{1}{2n}, n = |m|+2, |m|+4, \ldots. \tag{84}$$

OSA CONTINUUM

(when $m = 0$ we only consider $n = 2, 4, \ldots$). In the case that $n = |m| + 2, |m| + 4, \ldots$, we see that Eq. (82) and (83) gives

$$\alpha_n^m = \frac{1}{4n}(\beta_+)_{n-1}^{m+1} + \frac{1}{4n}(\beta_-)_{n-1}^{m-1} - \frac{1}{4(n+2)}(\beta_+)_{n+1}^{m+1} - \frac{1}{4(n+2)}(\beta_-)_{n+1}^{m-1}. \tag{85}$$

Now we have from Eq. (81)

$$\frac{\partial W}{\partial \nu} = \sum_{n,m} \frac{1}{2}((\beta_+)_n^m + (\beta_-)_n^m) Z_n^m \tag{86}$$

$$\frac{\partial W}{\partial \mu} = \sum_{n,m} \frac{1}{2i}((\beta_+)_n^m - (\beta_-)_n^m) Z_n^m. \tag{87}$$

Therefore, from Eq. (71) and (72) and the orthogonality/normalization of the Z_n^m, see Eq. (12), we see that the integral expression at the right-hand side of Eq. (51) becomes

$$\begin{aligned}
&\frac{1}{\pi}\int_0^1\int_0^{2\pi}\left(\frac{\partial W}{\partial \nu}\frac{\partial U_n^{m*}}{\partial \nu} + \frac{\partial W}{\partial \mu}\frac{\partial U_n^{m*}}{\partial \mu}\right)\rho d\rho d\theta \\
&= -\frac{1}{8\pi}\int_0^1\int_0^{2\pi}\sum_{n',m'}((\beta_+)_{n'}^{m'} + (\beta_-)_{n'}^{m'})Z_{n'}^{m'} \\
&\times \left[Z_{n+1}^{m+1} + Z_{n+1}^{m-1} - Z_{n-1}^{m+1} - Z_{n+1}^{m-1}\right]^* \rho d\rho d\theta \\
&= -\frac{1}{8\pi}\int_0^1\int_0^{2\pi}\sum_{n',m'}((\beta_+)_{n'}^{m'} - (\beta_-)_{n'}^{m'})Z_{n'}^{m'} \\
&\times \left[Z_{n+1}^{m+1} - Z_{n+1}^{m-1} - Z_{n-1}^{m+1} + Z_{n+1}^{m-1}\right]^* \rho d\rho d\theta \\
&= -\frac{1}{8(n+2)}((\beta_+)_{n+1}^{m+1} + (\beta_-)_{n+1}^{m+1} + (\beta_+)_{n+1}^{m-1} + (\beta_-)_{n+1}^{m-1}) \\
&+ \frac{1}{8n}((\beta_+)_{n-1}^{m+1} + (\beta_-)_{n-1}^{m+1} + (\beta_+)_{n-1}^{m-1} + (\beta_-)_{n-1}^{m-1}) \\
&- \frac{1}{8(n+2)}((\beta_+)_{n+1}^{m+1} - (\beta_-)_{n+1}^{m+1} - (\beta_+)_{n+1}^{m-1} + (\beta_-)_{n+1}^{m-1}) \\
&+ \frac{1}{8n}((\beta_+)_{n-1}^{m+1} - (\beta_-)_{n-1}^{m+1} - (\beta_+)_{n-1}^{m-1} + (\beta_-)_{n-1}^{m-1}),
\end{aligned} \tag{88}$$

and this coincides with the right-hand side of Eq. (86) when the various cancellations are noted.

In a similar fashion the case $n = |m| > 0$ can be handled using Eqs. (79) and (80), and this yields that the $\alpha_{|m|}^m$ from Eqs. (82) and (83) coincide with the right-hand side integral expression in Eq. (51).

Funding

European Research Council (ERC), FP7/2007-2013, 339681.

Acknowledgments

The authors would like to acknowledge Prof. J. Braat for useful discussions and comments on this article. We also thank Dr. S. van Haver for discussions.

References

1. A. J. E. M. Janssen, "Zernike expansion of derivatives and laplacians of the zernike circle polynomials," J. Opt. Soc. Am. A **31**, 1604–1613 (2014).

OSA CONTINUUM

2. The choice for ANSI convention is motivated here by how the Shack-Hartmann sensor deals with the circle polynomials. We would like to point out that using the Born and Wolf convention (that uses exponential instead of cosine/sine functions), renders the results in [1] more transparent and concise.
3. L. N. Thibos, R. A. Applegate, J. T. Schwiegerling, and R. Webb, "Standards for reporting the optical aberrations of eyes," J. Refract. Surg. **18**, S652–S660 (2002).
4. G. M. Dai, *Wavefront Optics for Vision Correction*, vol. 179 (SPIE Press, 2008).
5. V. N. Mahajan and E. Acosta, "Vector polynomials for direct analysis of circular wavefront slope data," J. Opt. Soc. Am. **34**, 1908–1913 (2017).
6. N. Matsumoto, T. Ando, T. Inoue, Y. Ohtake, N. Fukuchi, and T. Hara, "Generation of high-quality higher-order Laguerre-Gaussian beams using liquid-crystal-on-silicon spatial light modulators," J. Opt. Soc. Am. A **25**, 1642–1651 (2008).
7. R. Fletcher, *Practical Methods of Optimization* (John Wiley & Sons, 1987).
8. O. Soloviev and G. Vdovin, "Estimation of the total error of modal wavefront reconstruction with zernike polynomials and hartmann-shack test," in *5th International Workshop on Adaptive Optics for Industry and Medicine,* vol. 6018 (International Society for Optics and Photonics, 2006), p. 60181D.
9. H. Shengyang, X. Fengjie, L. Changhai, and J. Zongfu, "Eigenfunctions of Laplacian for phase estimation from wavefront gradient or curvature sensing," Opt. Commun. **284**, 2781–2783 (2011).
10. H. Shengyang, N. Yu, X. Fengjie, and J. Zongfu, "Modal wavefront reconstruction with Zernike polynomials and eigenfunctions of the Laplacian," Opt. Commun. **288**, 7–12 (2013).
11. A. J. E. M. Janssen and P. Dirksen, "Computing zernike polynomials of arbitrary degree using the discrete fourier transform," J. Eur. Soc. Rapid Publ. **2**, 07012 (2007).
12. B. H. Shakibaei and R. Paramesran, "Recursive formula to compute Zernike radial polynomials," Opt. Lett. **14**, 2487–2489 (2013).