

Orthonormal vector general polynomials derived from the Cartesian gradient of the orthonormal Zernike-based polynomials

COSMAS MAFUSIRE*  AND TJAART P. J. KRÜGER

Department of Physics, Faculty of Natural and Agricultural Sciences, University of Pretoria, Private Bag X20, Hatfield, 0028, South Africa

*Corresponding author: cosmasmafusire@gmail.com

The concept of orthonormal vector circle polynomials is revisited by deriving a set from the Cartesian gradient of Zernike polynomials in a unit circle using a matrix-based approach. The heart of this model is a closed-form matrix equation of the gradient of Zernike circle polynomials expressed as a linear combination of lower-order Zernike circle polynomials related through a gradient matrix. This is a sparse matrix whose elements are two-dimensional standard basis transverse Euclidean vectors. Using the outer product form of the Cholesky decomposition, the gradient matrix is used to calculate a new matrix, which we used to express the Cartesian gradient of the Zernike circle polynomials as a linear combination of orthonormal vector circle polynomials. Since this new matrix is singular, the orthonormal vector polynomials are recovered by reducing the matrix to its row echelon form using the Gauss–Jordan elimination method. We extend the model to derive orthonormal vector general polynomials, which are orthonormal in a general pupil by performing a similarity transformation on the gradient matrix to give its equivalent in the general pupil. The outer form of the Gram–Schmidt procedure and the Gauss–Jordan elimination method are then applied to the general pupil to generate the orthonormal vector general polynomials from the gradient of the orthonormal Zernike-based polynomials. The performance of the model is demonstrated with a simulated wavefront in a square pupil inscribed in a unit circle. © 2018 Optical Society of America

OCIS codes: (000.3860) Mathematical methods in physics; (080.1005) Aberration expansions; (110.0110) Imaging systems; (120.5050) Phase measurement.

1. INTRODUCTION

Slope-based wavefront sensors characterize phase in three stages. First, the Cartesian derivative of the phase map is generated by measuring the distribution of the local wavefront tilts. This is followed by a numerical integration procedure to generate a phase map. Lastly, an appropriate polynomial set is fitted to the phase map to generate expansion coefficients. The process can be shortened into two stages, namely, generation of the wavefront slope map, followed by fitting of an appropriate vector polynomial set to the slope map, from which the coefficients can be derived. Traditionally, Zernike circle (ZC) polynomials are used to characterize the wavefront. These polynomials provide a clear advantage when working with circular pupils since their orthogonality guarantees minimum variance and they can be related to classical aberrations. However, it becomes a tedious task to use ZC polynomials for determining the wavefront slope in that vector polynomials calculated from their Cartesian gradient are not normalized across the slope map. A complete

orthonormal vector circle polynomial set spanning the wavefront slope map in a unit circle (OVC) was derived by Zhao and Burge [1,2]. Their model for an OVC set provided a step forward from Noll's landmark paper in 1976, where the Cartesian phase slope was presented in a unit circle as a linear combination of lower ZC polynomials [3]. Noll's model lacked both a closed-form equation and orthonormal polynomials. Rigorous investigations of the slope of the ZC polynomials with trigonometric azimuthal dependence were implemented by Stephenson [4] for real ZC polynomials with trigonometric azimuthal dependence, and by Janssen [5] using an unnormalized ZC set with an exponential azimuthal complex component. In a recent development, Mahajan and Acosta advanced a new analytical method for deriving vector polynomials orthogonal to the gradients of Zernike circle polynomials. What is special about these polynomials is that they propagate minimum uncorrelated random noise from the data to coefficients [6], the discussion of which is beyond the scope of this paper.

Wavefront slope is also critical in the study of intensity moments of weakly truncated aberrated TEM₀₀ Gaussian laser beams [7–9] where models have been developed showing the dependence of the moments on aberration coefficients. In fact, Alda *et al.* based a large part of their method on Noll's model [7]. Orthonormal Zernike-based (OZ) polynomials have been applied to pupils other than circular pupils [10–13]. Recent efforts to deal with orthonormality in wavefront slope include the work of Ye *et al.*, who devised a matrix-based numerical orthogonal transformation method for generating numerical orthogonal gradient polynomials [14], and that of Li *et al.*, who derived orthonormal vector polynomials orthogonal in a square pupil of unit half-length [15].

The above literature has dealt with the characterization of wavefront slope including the use of various polynomial systems. To our knowledge, the only study that has dealt with the wavefront slope in noncircular pupils was in [15]. It is our assertion that a procedure can be found to derive vector polynomials orthonormal in any pupil. In this study, we present an analytical method for the generation of OVC polynomials using matrices in row-reduced echelon form and extending the model to the general noncircular nonuniform pupil using the Cholesky decomposition and a matrix similarity transformation. The result is a model that would allow one to derive vector polynomials that are orthonormal in general pupils (OVG). A general pupil is noncircular and nonuniform, which would make a unit circle a special case. The model can therefore be used to analyze any problem involving wavefront slopes in any pupil shape. We revisit the existing OVC polynomial model as outlined by Zhao and Burge [1,2] but this time using matrices to facilitate the conversion to general pupil where appropriate similarity and change of basis transformations would be presented. We treat the pupil as a vector space in which the OZ set is one of the bases that span the pupil. The approach we use is purely analytical where the removal of the piston is not assumed but is built into the model and so the piston is removed naturally. The relationship between OZ coefficients and OVG coefficients in general pupils is provided just like the one relating ZC coefficients with OVC coefficients in a unit circle in the case of the wavefront. Matrices have been utilized selectively in some of the literature referred to so far so that we have decided to look at the possibility of performing wavefront analysis using matrices exclusively. The results for the case of the unit circle [1,3,4] serve as an acid test for the validity of our method before it is applied to the general pupil.

The paper is organized as follows. In the following section, the ZC Cartesian gradient field is introduced and expressed as a sum of lower-order ZC polynomials. From this matrix, the OVC polynomial set is derived using the Cholesky decomposition method from which a relationship between the ZC and the OVC coefficients sets is presented. We follow up with Section 3 in which we explain how the OVC polynomials can be used to characterize slope maps. In Section 4, a matrix formulation of the Gram–Schmidt (GS) procedure is introduced, which is used to derive an OZ set. In Section 5, the OZ set is used to derive the orthonormal vector general polynomial set starting with the appropriate gradient polynomial field. The efficacy of the model is applied to the analysis of

simulated wavefront slope data in a square pupil inscribed in a unit circle in Section 6. The paper is summarized in Section 7.

2. ZERNIKE CIRCLE SLOPE EXPANSION

Consider a circular pupil described by a pupil function of the form

$$E(\rho, \theta) = \begin{cases} \exp\{i\phi_c(\rho, \theta)\}, & \rho \leq 1 \\ 0, & \rho > 1 \end{cases}, \quad (1)$$

where a real unit amplitude described by a circular function and a real phase $\phi_c(\rho, \theta)$ are assumed, with (ρ, θ) being cylindrical spatial coordinates across the pupil and $i = \sqrt{-1}$. The phase can be written as a linear expansion of the complete ordered Zernike polynomial set, $\mathbf{z}(\rho, \theta) = \{Z_n^m(\rho, \theta)\}_{j=1}^J$, i.e.,

$$\phi_c(\rho, \theta) = \mathbf{c}^t \mathbf{z}(\rho, \theta), \quad (2)$$

where $\mathbf{c} = \{C_n^m\}_{j=1}^J$ is a column vector of expansion coefficients, with n and m being the order and azimuthal numbers, respectively, and the superscript t representing matrix transposition. The orthonormality of the circle polynomials is epitomized by the equation

$$\frac{1}{\pi} \int_0^{2\pi} \int_0^1 \mathbf{z}(\rho, \theta) \mathbf{z}^t(\rho, \theta) \rho d\rho d\theta = \mathbf{I}, \quad (3)$$

where \mathbf{I} is an identity matrix. Each coefficient, C_n^m , and its associated ZC polynomial, $Z_n^m(\rho, \theta)$, are therefore associated with a unique double index set (n, m) . For all polynomial sets we will follow the sequence of elements set by Noll [3], where each aberration is also indicated by a mode number j . We assume that the dimension is set at some finite number J . Note that all matrices are represented by bold capital letters and are of dimension $J \times J$, while vectors are represented by bold small letters and are of dimension $J \times 1$. In cylindrical geometry, each ZC polynomial is given by a product of a radial and an azimuthal term of the form

$$Z_n^m(\rho, \theta) = \sum_{k=0}^{\frac{1}{2}(n-|m|)} \frac{\sqrt{n+1}(-1)^k(n-k)! \rho^{n-2k}}{k! \left\{ \frac{1}{2}(n+m) - k \right\}! \left\{ \frac{1}{2}(n-m) - k \right\}!} \times \begin{cases} \sqrt{2 - \delta_{|m|,0}} \cos m\theta, & m \geq 0 \\ \sqrt{2} \sin |m|\theta, & m < 0 \end{cases}. \quad (4)$$

Note that $Z_n^m(\rho, \theta) = 0$ when $n - |m|$ is a negative or odd integer; otherwise, it is a nonnegative, even integer. Equation (3) is an inner product matrix referred to as the Gram matrix, which is constructed using the inner product of an orthonormal set, in this case the ZC polynomial set in the unit circle. The fact that the result is an identity matrix implies that the ZC polynomial set forms an orthonormal basis in this pupil. The determinant of this matrix is nonzero, meaning that the ZC polynomials are linearly independent [16]. The size of the matrix obviously depends on the number of basis functions used.

The Cartesian gradient of the phase is defined by a transverse linear gradient operator, $\nabla_T = \hat{i}\partial_x + \hat{j}\partial_y$, on the phase of the field as follows:

$$\vec{\nabla}_T \phi_c(\rho, \theta) = \mathbf{c}^t \nabla_T \mathbf{z}(\rho, \theta), \quad (5)$$

where the Cartesian gradient of the phase is presented as an expansion of a field of polynomials represented by the vector field $\nabla_T \mathbf{z}(\rho, \theta)$ of size $J \times 1$ [1]:

$$\nabla_T \mathbf{z}(\rho, \theta) = \hat{i} \partial_x \mathbf{z}(\rho, \theta) + \hat{j} \partial_y \mathbf{z}(\rho, \theta) = \vec{M} \mathbf{z}(\rho, \theta). \quad (6)$$

We define all symbols with an arrow on top of the symbol to represent a transverse two-dimensional Euclidean vector of the standard basis. If the arrow is on top of a symbol representing a column vector or a matrix, then all the elements in the column vector or matrix would be transverse Euclidean vectors. Each aberration up to $j = J$ is arranged in the order of the ZC set as $\{Z_n^m(\rho, \theta)\}_{j=1}^J$.

In Subsection 2.A, we now outline a derivation of the ZC gradient matrix using Taylor monomials. We then follow up in Subsection 2.B with a derivation of a special singular matrix to represent the Cartesian gradient of the ZC polynomials as a linear combination of the OVC polynomials. We follow up in Subsection 2.C with an innovative procedure involving the use of the Gauss–Jordan elimination to reverse the matrix equation outlined in Subsection 2.B to express the OVC polynomials in terms of ZC polynomials.

A. Calculation of the ZC Polynomial Cartesian Gradient Matrix \vec{M} from Taylor Monomials

The matrix \vec{M} is a characteristic square matrix for representing the ZC polynomial Cartesian gradient as a linear combination of lower-order ZC polynomials. The method to calculate this matrix is to set the phase ϕ_c as a linear combination of the Taylor monomials. The reason for this is that the first-order Cartesian derivatives of Taylor monomials are easier to get than those of the ZC set. The procedure first establishes a relationship between the two expansion sets and then uses the Taylor monomial derivatives to derive the Cartesian derivatives of the ZC set. The Taylor monomials set can be represented by the vector

$$\begin{aligned} \boldsymbol{\tau}(\rho, \theta) &= \{T_p^q(\rho, \theta)\}_{l=1}^J = \{\rho^p \text{Cos}^{p-q} \theta \text{Sin}^q \theta\}_{l=1}^J \\ &= \{x^{p-q} y^q\}_{l=1}^J = \boldsymbol{\tau}(x, y), \end{aligned} \quad (7)$$

which is an ordered set arranged according to the order number l , each one associated with a unique double index set (p, q) , which can be related to l through [17,18]

$$\begin{aligned} p &= \text{int} \left\{ \left(\sqrt{8(l-1) + 1} - 1 \right) / 2 \right\}; \\ q &= l - p(p+1)/2 - 1. \end{aligned} \quad (8)$$

Unlike in [17,18], the expressions for p and q have been modified to allow the monomial ordering to start at $l = 1$ instead of 0.

The ZC and the Taylor monomial sets can be related through a matrix equation:

$$\mathbf{z}(\rho, \theta) = \mathbf{P} \boldsymbol{\tau}(\rho, \theta) = \mathbf{P} \boldsymbol{\tau}(x, y), \quad (9)$$

where \mathbf{P} is the nonsingular change-of-basis matrix from the Taylor monomials to the ZC set. For $J = 21$, it is illustrated in Fig. 1(a). The results match those shown in Table 5.10 of [18]. The gradient of the Taylor monomials can be expressed as a linear combination of lower-order Taylor monomials as follows:

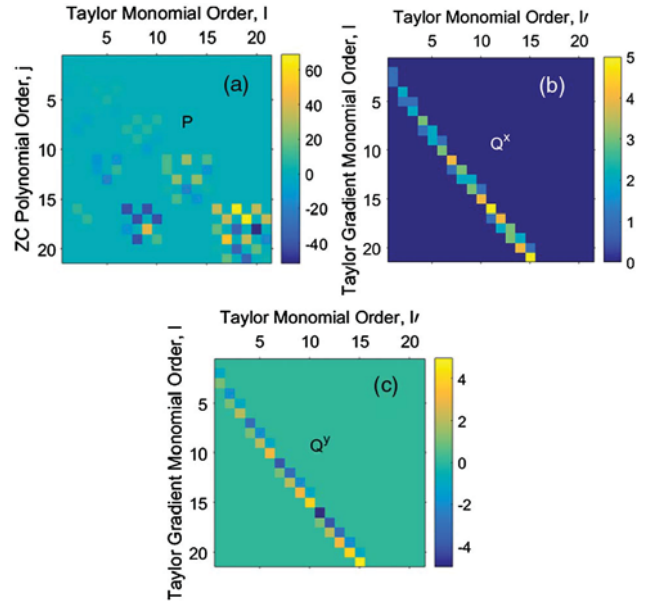


Fig. 1. Scaled color image plot of (a) \mathbf{P} , a square matrix for calculating the first 21 ZC polynomials from Taylor monomials. The two square matrices (b) \mathbf{Q}^x and (c) \mathbf{Q}^y for expressing the Cartesian gradient of Taylor monomials as linear combinations of the lower-order Taylor monomials. Intensity scale bar is shown on the right of each plot.

F1:1
F1:2
F1:3
F1:4
F1:5

$$\nabla_T \boldsymbol{\tau}(\rho, \theta) = \vec{Q} \boldsymbol{\tau}(\rho, \theta) = \hat{i} \mathbf{Q}_{ll'}^x \boldsymbol{\tau}(\rho, \theta) + \hat{j} \mathbf{Q}_{ll'}^y \boldsymbol{\tau}(\rho, \theta), \quad (10)$$

where the elements making up the vector $\vec{\nabla}_T \boldsymbol{\tau}(\rho, \theta)$ can be calculated from [17]

$$\begin{aligned} \nabla_T \boldsymbol{\tau}(\rho, \theta) &= \{\vec{\nabla}_T T_p^q(\rho, \theta)\}_{l=1}^J \\ &= \{\hat{i}(p-q) T_{p-1}^q(\rho, \theta) + \hat{j} q T_{p-1}^{q-1}(\rho, \theta)\}_{l=1}^J, \end{aligned} \quad (11)$$

where $\boldsymbol{\tau}(\rho, \theta)$ follows a slightly different sequence from Section 3.4.2 of [18]. These two matrices $\mathbf{Q}_{ll'}^x$ and $\mathbf{Q}_{ll'}^y$ are illustrated in Figs. 1(b) and 1(c). It is apparent that, as with the ZC polynomials, the Cartesian gradient of each Taylor monomial is a linear combination of Taylor monomials of lower order than that itself.

The gradient of the ZC set can then be calculated by taking the gradient of Eq. (9), and then using Eq. (10) and again Eq. (9) to get a matrix representation of the ZC Cartesian gradient in terms of the ZC set:

$$\nabla_T \mathbf{z}(\rho, \theta) = \mathbf{P} \nabla_T \boldsymbol{\tau}(\rho, \theta) = \mathbf{P} \vec{Q} \boldsymbol{\tau}(\rho, \theta) = \mathbf{P} \vec{Q} \mathbf{P}^{-1} \mathbf{z}(\rho, \theta). \quad (12)$$

Comparing Eq. (6) with Eq. (12), one finds a similarity transformation on the Taylor monomial gradient matrix,

$$\vec{M} = \mathbf{P} \vec{Q} \mathbf{P}^{-1} = \hat{i} \boldsymbol{\gamma}_{jj'}^x + \hat{j} \boldsymbol{\gamma}_{jj'}^y, \quad (13)$$

that facilitates the calculation of the ZC gradient matrix. Similar matrices, like \vec{M} and \vec{Q} , share certain properties such as rank, trace, determinant and eigenvalues, which remain invariant during coordinate transformation [16]. Its components are given by $\boldsymbol{\gamma}_{jj'}^x$ and $\boldsymbol{\gamma}_{jj'}^y$, respectively, shown in Tables II and III in [3], which are x and y components. Equation (6) is a representation of the equations listed in

2

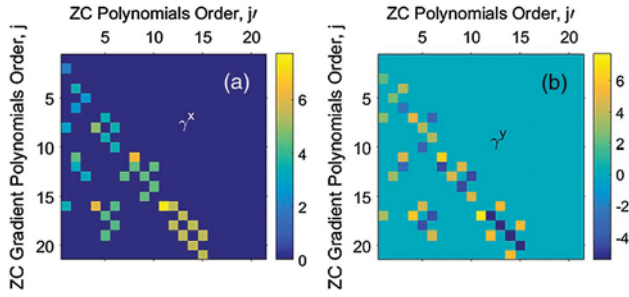


Fig. 2. Matrix components of \vec{M} , (a) γ_{jj}^x , and (b) γ_{jj}^y . The matrices illustrate that the Cartesian gradient of each ZC polynomial depends on linear combinations of ZC polynomials of orders lower than itself.

Table 1 of [1]. The elements of γ_{jj}^x and γ_{jj}^y for $J = 21$ are shown as scaled color images in Figs. 2(a) and 2(b), respectively.

B. Orthonormal Vector Circle Polynomials

The ZC Cartesian gradient set, $\nabla_T \mathbf{z}(\rho, \theta)$, is not orthonormal in a circular pupil. A new set of OVC polynomials, which possesses this property, was therefore developed using the Gram–Schmidt orthogonalization procedure [1,2]. We are revisiting the same problem here, but using matrices instead. The Zernike gradient set can be represented as a linear combination of the OVC polynomials:

$$\nabla_T \mathbf{z}(\rho, \theta) = \mathbf{K} \vec{s}(\rho, \theta), \quad (14)$$

where \mathbf{K} , which is to be determined, is a square matrix of expansion coefficients of the polynomials. To ensure that the OVC polynomial set, $\vec{s}(\rho, \theta) = \hat{i}s_x(\rho, \theta) + \hat{j}s_y(\rho, \theta)$, is orthonormal in a circle, the following Gram matrix equation should be valid:

$$\frac{1}{\pi} \int_0^{2\pi} \int_0^1 \vec{s}(\rho, \theta) \circ \vec{s}(\rho, \theta)^t \rho d\rho d\theta = \mathbf{I}, \quad (15)$$

where \circ represents scalar multiplication. Comparing Eqs. (6) and (14), we find a relation between the OVC and the Zernike set, as shown in the first equation of Eq. (16) below. Each side of the resultant equation is then post scalar multiplied by the transposed version of itself resulting in $\vec{M} \mathbf{z}(\rho, \theta) \circ \mathbf{z}^t(\rho, \theta) \vec{M}^t$ on the left-hand side and $\mathbf{K} \vec{s}(\rho, \theta) \circ \vec{s}^t(\rho, \theta) \mathbf{K}^t$ on the right-hand, respectively. The inner product of both sides is then calculated in a unit circle and then simplified using both Gram matrices, Eqs. (3) and (15), eliminating both $\mathbf{z}(\rho, \theta)$ and $\vec{s}(\rho, \theta)$, thereby generating the second equation of Eq. (16):

$$\vec{M} \mathbf{z}(\rho, \theta) = \mathbf{K} \vec{s}(\rho, \theta) \Rightarrow \mathbf{K} \mathbf{K}^t = \vec{M} \circ \vec{M}^t. \quad (16)$$

Since $\vec{M} \circ \vec{M}^t$ is not positive definite, the Cholesky decomposition procedure cannot be applied to solve for \mathbf{K} . Neglecting the first row and first column of $\vec{M} \circ \vec{M}^t$, which are all zeros, a principal submatrix conformable with this procedure is obtained. This problem can be handled by partitioning \mathbf{K} into the form

$$\mathbf{K} = \begin{bmatrix} \mathbf{0} & \vec{\mathbf{0}}^t \\ \vec{\mathbf{0}} & \vec{\mathbf{K}} \end{bmatrix}, \quad (17)$$

where the vector $\vec{\mathbf{0}} = \{\vec{\mathbf{0}}\}$ is of size $(J - 1) \times 1$, with $\vec{\mathbf{K}}$ being of size $(J - 1) \times (J - 1)$. The method is referred to as the outer

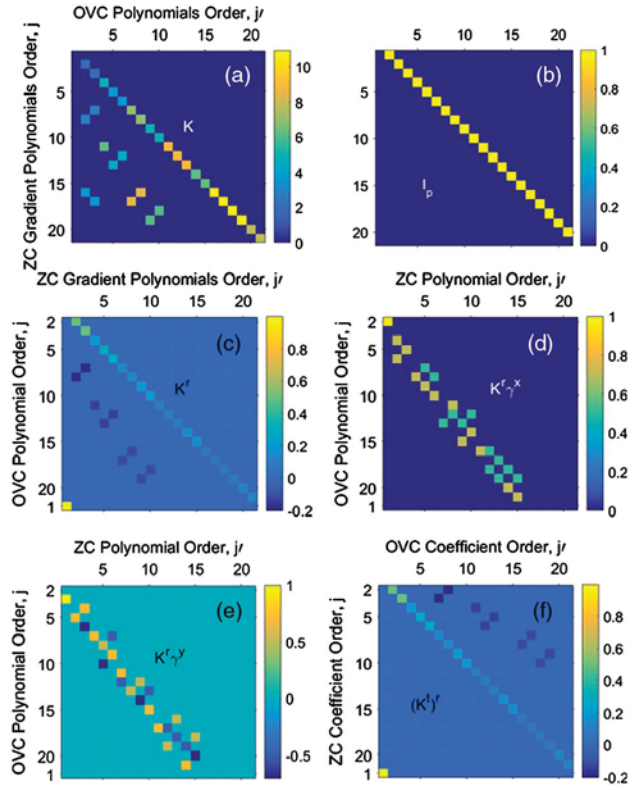


Fig. 3. (a) Matrices for expressing the ZC Cartesian gradient in terms of OVC polynomials, \mathbf{K} ; (b) a superdiagonal unit matrix, \mathbf{I}_p ; (c) expressing OVC polynomials in terms of ZC gradient polynomials, \mathbf{K}^t ; (d), (e) expressing OVC polynomials in terms of ZC polynomials, $\mathbf{K}^t \gamma^x$ and $\mathbf{K}^t \gamma^y$; (f) polynomials in terms of the OVC polynomials, $(\mathbf{K}^t)^r$.

product form of the Cholesky decomposition [19]. Applying this method to Eq. (16) gives the equation $\vec{\mathbf{K}} \vec{\mathbf{K}}^t = \vec{M} \circ \vec{M}^t$, which is conformable with the decomposition procedure. After solving for $\vec{\mathbf{K}}$, it is reinstated in Eq. (17). Figure 3(a) shows the scaled color image plot of \mathbf{K} for $J = 21$ where it can be verified that it is also a lower triangular matrix, which is even sparser than \vec{M} , with the first row and column being zeroes.

C. Orthonormal Vector Circle Polynomials in Terms of the Zernike Circle Polynomials

The OVC polynomials, $\vec{s}(\rho, \theta)$, can be calculated from $\nabla_T \mathbf{z}(\rho, \theta)$ by solving Eq. (14). This is not a straightforward task considering that \mathbf{K} is a singular matrix. In this study, we propose a method that will provide an analytical solution. The matrix \mathbf{K} is appended with an identity matrix of the same size to get an augmented matrix $[\mathbf{K}:\mathbf{I}]$ and row-reducing the result using the Gauss–Jordan elimination method to get the reduced row-echelon form $[\mathbf{I}_p:\mathbf{K}^r]$ [16] such that Eq. (14) becomes

$$\mathbf{I}_p \vec{s}(\rho, \theta) = \mathbf{K}^r \nabla_T \mathbf{z}(\rho, \theta). \quad (18)$$

The matrix \mathbf{I}_p is a superdiagonal unit square matrix in which all the elements are zeroes except those in the superdiagonal,

2

2

2

2

which is occupied by ones, as shown in Fig. 3(b). It is apparent that the effect of \mathbf{I}_p on $\vec{s}(\rho, \theta)$ is to create a left-hand side in which $\vec{S}_0^0(\rho, \theta)$ is effectively eliminated and plays no further part, leaving the last element as $\vec{0}$. The right-hand side is the product of \mathbf{K}^r and the Cartesian gradient of the ZC set. \mathbf{K}^r is a sparse matrix illustrated in Fig. 3(c) with the superdiagonal elements being nonzero. There are no nonzero elements above the superdiagonal with a 1, which is the largest element, at the bottom left-hand corner. The sequence of the OVC polynomials has been altered slightly than from that proposed by Noll with the first order moved to the bottom row, making the last element as $\nabla_T Z_0^0$. Note that the last elements on both sides are the same regardless of the value of J meaning that the result $\nabla_T Z_0^0 = 0$ is therefore a universal result. Equation (18) can be used to construct Table 3 in [1]. Consequently, the OVC polynomials can be derived from the Zernike set by combining Eqs. (6) and (18) by eliminating $\nabla_T \mathbf{z}(\rho, \theta)$, which leaves us with

$$\mathbf{I}_p \vec{s}(\rho, \theta) = \mathbf{K}^r \vec{\mathbf{M}} \mathbf{z}(\rho, \theta). \quad (19)$$

Equation (19) can also be acquired directly by row-reducing Eq. (14). Due to the construction of $\vec{\mathbf{M}}$, the matrix $\mathbf{K}^r \vec{\mathbf{M}}$ consists of two vector components, shown in Figs. 3(d) and 3(e) for the x and y components, respectively. The last row of $\mathbf{K}^r \vec{\mathbf{M}}$ contains all zeroes, which is also a universal result independent of J used in the analysis. Equation (19) can be used to construct Table 4 of [1].

1 We chose this method as opposed to others, such as singular value decomposition and QR decomposition, because of its simplicity and ease of use. It appears that the matrix \mathbf{I}_p , which is central to our model, could best be derived using the Gauss–Jordan procedure. This is because the elements of the first row of $\vec{\mathbf{M}}$ are all zeroes. The resultant matrix, \mathbf{K} , has a zero first row and a zero first column. This makes using the least-squares approach impossible to implement since $\mathbf{K}^t \mathbf{K}$ is singular. Moreover, although the QR decomposition technique can successfully decompose \mathbf{K} into an orthogonal matrix and an upper triangular matrix, both are singular and hence it is not possible to use this method either.

3. ZERNIKE CIRCLE SLOPE RECONSTRUCTION

The OVC coefficients can now be calculated by fitting OVC polynomials, which span the wavefront slope in a unit circle, generating slope coefficients, $\mathbf{b} = \{B_n^m\}_{j=1}^J$, such that the wavefront slope can be reconstructed using

$$\vec{\nabla}_T \phi_c(\rho, \theta) = \mathbf{b}^r \vec{s}(\rho, \theta). \quad (20)$$

The slope maps concerned are generated from the Cartesian gradient of the phase, which has \hat{i} and \hat{j} components such that $\nabla_T \phi_c = \hat{i} \partial_x \phi_c(\rho, \theta) + \hat{j} \partial_y \phi_c(\rho, \theta)$. Therefore, the orthonormality of the set equation allows one to directly calculate these coefficients through a sum of two Cartesian components from fitting the respective OVC polynomials as shown:

$$\begin{aligned} \mathbf{b} &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \vec{s}(\rho, \theta) \circ \vec{\nabla}_T \phi_c(\rho, \theta) \rho d\rho d\theta \\ &= \frac{1}{\pi} \int_0^{2\pi} \int_0^1 \begin{pmatrix} s_x(\rho, \theta) \partial_x \phi_c(\rho, \theta) \\ + s_y(\rho, \theta) \partial_y \phi_c(\rho, \theta) \end{pmatrix} \rho d\rho d\theta = \mathbf{b}_x + \mathbf{b}_y. \end{aligned} \quad (21)$$

The method used in [5] avoided the derivation of the OVC set and thus avoids the use of Eq. (21). The Cartesian gradient of the phase in a unit circle was expressed in the form $\partial_x \phi_c \pm \partial_y \phi_c$, which reduces to $\mathbf{c}^t \mathbf{A}_\pm \mathbf{z}$ where the matrix \mathbf{A}_\pm is calculated from $\gamma_{jj}^x \pm \gamma_{jj}^y$. If we recast the Cartesian gradient into a linear combination of the ZC set, $\delta_\pm^t \mathbf{z}$, it is apparent that $\delta_\pm = \mathbf{A}_\pm^t \mathbf{c}$, which was then solved by finding the pseudo-inverse solution, $(\mathbf{A}_\pm \mathbf{A}_\pm^t)^{-1} \mathbf{A}_\pm$. The least-squares estimate of the expansion coefficients is a linear combination of the coefficients of $\partial_x \phi_c \pm \partial_y \phi_c$ of orders $n+1$, $n-1$ and azimuthal numbers $m \pm 1$ [5].

From Eqs. (5) and (14), it has been established that

$$\nabla_T \phi_c(\rho, \theta) = \mathbf{c}^t \mathbf{K} \vec{s}(\rho, \theta). \quad (22)$$

By comparing Eqs. (20) and (22), the relationship between the two coefficients is given by $\mathbf{b} = \mathbf{K}^t \mathbf{c}$. Applying the row-reduction method, we obtain

$$\mathbf{I}_p \mathbf{c} = \{\mathbf{K}^t\}^r \mathbf{b}. \quad (23)$$

This equation allows us to calculate ZC coefficients from slope coefficients in a unit circle if the OVC set is used to characterize a slope map without the need to add an extra step of integrating the slope maps to generate the phase. The matrix $\{\mathbf{K}^t\}^r$ shown in Fig. 3(f) is a mirror image to \mathbf{K}^r about the superdiagonal except for the element 1 at the bottom left-hand corner, which is common to both matrices.

The effect of \mathbf{I}_p is the elimination of the piston coefficient, C_0^0 , from the formalism. Pre-multiplying $\{\mathbf{K}^t\}^r$ with \mathbf{b} leaves us with a vector whose last entry is 0 after effectively eliminating B_0^0 . This is a very sparse matrix with nonzero elements along the superdiagonal and to the right, giving two nonzero elements in each row up to the ninth row. A cursory inspection tells us that in each row corresponding to C_n^m , the two nonzero terms indicate that each ZC coefficient is calculated from B_n^m and B_{n+2}^m .

4. ORTHONORMAL ZERNIKE-BASED POLYNOMIALS

The general pupils can be described mathematically by any function of any pupil shape with the circular function being the special case. This includes a nonuniform amplitude, a noncircular system aperture, or both. Here the assumption is that the field is defined by an amplitude function $A(\rho)$ inside the aperture, which obtains a value of zero outside, with ρ being a spatial coordinate pinpointing any location within the aperture. The phase distribution inside the aperture is given by $\phi(\rho)$. The overall result is a field in the aperture described by the pupil function

$$E(\rho) = \begin{cases} A(\rho) \exp\{i\phi(\rho)\}, & \text{inside the aperture} \\ 0, & \text{outside the aperture} \end{cases}. \quad (24)$$

The wavefront inside the pupil is expressed as an expansion of an infinite but discrete OZ polynomial set, $\underline{\mathbf{z}}(\rho) = \{\underline{Z}_n^m(\rho)\}_{j=0}^J$, spanning the pupil such that

$$\phi(\rho) = \underline{c}^t \underline{z}(\rho) = \chi^t \mathbf{z}(\rho, \theta), \quad (25)$$

where $\underline{c} = \{\underline{C}_n^m\}_{j=0}^J$ and $\chi = \{X_n^m\}_{j=0}^J$ are the respective expansion coefficients if the phase is expanded as linear combinations of the OC or ZC polynomials. The OZ set is orthonormal in the pupil and so its Gram matrix has the property given by the first equation shown below:

$$\frac{\iint \underline{z}(\rho) \underline{z}^t(\rho) A(\rho) d^2 \rho}{\iint A(\rho) d^2 \rho} = I. \quad (26)$$

Equation (26) allows the calculation of the vector containing the expansion coefficients, which can be derived using the first equation. The OZ polynomials in a general pupil can be written as a linear combination of the ZC polynomials and so can be calculated through the first matrix equation below:

$$\underline{z}(\rho) = (\rho, \theta) \Rightarrow \underline{c} = (\mathbf{H}^{-1})^t \chi, \quad (27)$$

which, incidentally, is a matrix form of the GS orthogonalization procedure [10–12], indicating the change of basis from the ZC to the OZ set. The right-hand side of Eq. (28) illustrates the vector relationship between the two coefficient sets if we select the same number of terms, J , for both expansions.

The expansion coefficients in a lower triangular matrix, \mathbf{H} , are calculated from Eq. (28), which is derived by combining Eq. (26) and the first equation of Eq. (27). The matrix \mathbf{H} is evaluated through the Cholesky decomposition considering that $\mathbf{H}^t \mathbf{H}$ is a symmetric positive definite matrix [10–13]:

$$\mathbf{H}^t \mathbf{H} = \left(\frac{\iint \mathbf{z}(\rho, \theta) \mathbf{z}^t(\rho, \theta) A(\rho) d^2 \rho}{\iint A(\rho) d^2 \rho} \right)^{-1}. \quad (28)$$

5. ORTHONORMAL VECTOR GENERAL POLYNOMIALS

Now that a formulation for the generation of the OZ polynomials has been obtained, the model of the OVC polynomials introduced in Sections 2 and 3 can be extended to include general pupils. These are referred to as the OVG polynomials. The wavefront gradient in this pupil is then represented by the expression as a linear combination of the OVG set:

$$\nabla_T \phi(\rho) = \underline{c}^t \nabla_T \underline{z}(\rho). \quad (29)$$

Identical to the case of the gradient of a circular pupil, OVG polynomials are generated through a GS orthogonalization of $\nabla_T \underline{z}(\rho, \theta) = \{\nabla \underline{Z}_n^m(\rho)\}_{j=1}^J$. Exploiting the linearity of the operator, ∇_T , it can be demonstrated that

$$\nabla_T \underline{z}(\rho) = \nabla_T \mathbf{H} \mathbf{z}(\rho, \theta) = \mathbf{H} \nabla_T \mathbf{z}(\rho, \theta). \quad (30)$$

In terms of the ZC gradient as defined in Eq. (6) into Eq. (31) we eliminate $\nabla_T \mathbf{z}(\rho, \theta)$ to obtain

$$\nabla_T \underline{z}(\rho) = \mathbf{H} \vec{\mathbf{M}} \mathbf{z}(\rho, \theta) = \mathbf{H} \vec{\mathbf{M}} \mathbf{H}^{-1} \underline{z}(\rho) = \vec{\mathbf{M}} \underline{z}(\rho). \quad (31)$$

The third term in Eq. (31) is a result of the matrix form of the Gram–Schmidt procedure [Eq. (27)], in which $\nabla_T \underline{z}(\rho)$ is expressed in terms of a new gradient matrix $\vec{\mathbf{M}}$ derived by a similarity transformation of \mathbf{M} :

$$\vec{\mathbf{M}} = \mathbf{H} \vec{\mathbf{M}} \mathbf{H}^{-1}, \quad (32)$$

where \mathbf{H} acts as a change of basis matrix of the transformation. Using the method used in Eq. (9), the OZ Cartesian gradient can also be expressed as follows:

$$\nabla_T \underline{z}(\rho) = \mathbf{K}_H \vec{\underline{z}}(\rho), \quad (33)$$

where \mathbf{K}_H is a matrix to be determined as was done with \mathbf{K} but this time by solving the equation

$$\mathbf{K}_H \mathbf{K}_H^t = \vec{\mathbf{M}} \circ \vec{\mathbf{M}}^t. \quad (34)$$

One can therefore surmise that OVG polynomials can be calculated from OVC polynomials by a Gram–Schmidt transformation given by the first question below:

$$\vec{\underline{z}}(\rho) = \mathbf{H} \vec{\mathbf{s}}(\rho, \theta), \quad \frac{\iint \vec{\underline{z}}(\rho) \circ \vec{\underline{z}}^t(\rho) A(\rho) d^2 \rho}{\iint A(\rho) d^2 \rho} = I. \quad (35)$$

The second equation of Eq. (35) is a Gram matrix of this set in the general pupil. Equating Eqs. (31) and (33) and row-reducing the result gives the OVG polynomials

$$\mathbf{I}_p \vec{\underline{z}}(\rho) = \mathbf{K}_H^r \& \vec{\mathbf{M}} \underline{z}(\rho). \quad (36)$$

As described above, the acquisition of wavefront information by an appropriate measurement device is usually achieved using ZC polynomials, even for general pupils. In the case of wavefront slope, the results are acquired by fitting OVC polynomials to the wavefront in the general aperture to give results of the form

$$\nabla_T \phi(\rho) = \underline{\mathbf{b}}^t \vec{\underline{z}}(\rho) = \beta^t \vec{\mathbf{s}}(\rho, \theta) = \beta^t \mathbf{H}^{-1} \vec{\underline{z}}(\rho). \quad (37)$$

The elements of the vectors β and $\underline{\mathbf{b}}$ are a result of fitting the OVC set, $\vec{\mathbf{s}}(\rho, \theta)$, or the OVG set, $\vec{\underline{z}}(\rho)$, respectively, to a given slope map in a general pupil. The set $\underline{\mathbf{b}}$ is given by

$$\begin{aligned} \underline{\mathbf{b}} &= \frac{\iint \vec{\underline{z}}(\rho) \circ \vec{\nabla}_T \phi(\rho) A(\rho) d^2 \rho}{\iint A(\rho) d^2 \rho} \\ &= \frac{\iint [\underline{\mathbf{s}}_x(\rho) \partial_x \phi(\rho) + \underline{\mathbf{s}}_y(\rho) \partial_y \phi(\rho)] A(\rho) d^2 \rho}{\iint A(\rho) d^2 \rho} = \underline{\mathbf{b}}_x + \underline{\mathbf{b}}_y. \end{aligned} \quad (38)$$

To characterize the general pupil properly, it is necessary to derive an expression relating β and the OZ coefficients, \underline{c} . This will provide a way to restore the advantages due to normalization. Combining Eqs. (29) and (33) gives us an alternative definition of the wavefront gradient:

$$\nabla_T \phi(\rho) = \underline{c}^t \mathbf{K}_H \vec{\underline{z}}(\rho). \quad (39)$$

Comparing Eqs. (37) and (39) results in a matrix equation relating all three coefficient vectors $\underline{\mathbf{b}} = \mathbf{K}_H^t \underline{c} = (\mathbf{H}^{-1})^t \beta$. Since what is needed is to acquire \underline{c} from the measured values, this vector needs to be expressed as a function of $\underline{\mathbf{b}}$ and β . This is resolved by a row-reduction procedure:

$$\mathbf{I}_p \underline{c} = (\mathbf{K}_H^t)^r \underline{\mathbf{b}} = (\mathbf{K}_H^t)^r (\mathbf{H}^{-1})^t \beta. \quad (40)$$

Equation (40) is used to generate OZ coefficients after measurement of wavefront slope in general pupils. It is a generalization that reduces to Eq. (23) in a unit circle in that \mathbf{H} becomes an identity matrix, following a similar procedure as with other expressions introduced in this section.

6. WAVEFRONT RECONSTRUCTION IN A SQUARE PUPIL INSCRIBED IN A UNIT CIRCLE

The model will now be used to analyze the aberrations in a square pupil inscribed in a unit circle [10–12], as often used

in high-power laser systems [13]. The model for the pupil will be based on a slope-based wavefront sensor with a 69×69 microlens grid. The coordinate system used is a Cartesian coordinate system, where $\rho = (x, y)$ such that $x \in [-1/\sqrt{2}; 1/\sqrt{2}]$ and $y \in [-1/\sqrt{2}; 1/\sqrt{2}]$. The Zernike square (ZSq) polynomial set, which we represent by the symbol $\underline{z}(x, y)$, has been discussed previously [10,11,13]. These polynomials can be derived by generating the appropriate change of basis matrix, \mathbf{H} , as outlined in Section 4. The matrix for this case is illustrated by Fig. 4(a). The ZSq polynomials calculated in this case are listed in [10,11,13]. This gives us all the information we need to express the Cartesian gradient of the ZSq polynomials as linear combinations of the lower-order ZSq polynomials. Table 1 shows the set $\nabla_T \underline{z}(x, y)$ up to the fifth order expressed as linear combinations of the ZC set. As expected, the Cartesian gradient of the piston is a zero vector, which is a trivial result.

The expansion matrix used for the derivation, $\mathbf{H}\vec{\mathbf{M}}\underline{z}$, expresses the ZSq polynomials as a linear combination of the ZC polynomials. For the aberrations selected for this study, the x and y parts of the ZSq gradient polynomials are shown in Figs. 4(b) and 4(c), respectively. The resultant polynomials are listed in Table 1.

The OVG set for the inscribed square pupil was derived next from the results in Table 1. The results are shown in Table 2, where the OVG set is expressed as a function of the Cartesian coordinates up to the fifth order. Equation (36) was used replacing $\vec{\mathbf{M}}\underline{z}(\rho)$ with $\mathbf{H}\vec{\mathbf{M}}\underline{z}(\rho, \theta)$. Since the pupil is a square, it is prudent to use Cartesian coordinates such that the ZC polynomials $\underline{z}(\rho, \theta)$ are replaced by $\mathbf{P}\boldsymbol{\tau}(x, y)$, where \mathbf{P} is a $J \times J$ matrix and $\boldsymbol{\tau}(x, y)$ is a vector containing the first $J \times 1$ Taylor monomials. The matrix \mathbf{K}_H^T is illustrated in Fig. 4(d). The complete matrix, $\mathbf{K}_H^T \mathbf{H}\vec{\mathbf{M}}\mathbf{P}$, has the two x and y components, illustrated in Figs. 4(e) and 4(f), respectively.

The last row in both matrices contains only zeroes. The resultant polynomials are listed in Table 2, showing that the model eliminates $\vec{\mathbf{S}}_1$. The orthonormality of the set was tested using Eq. (35) and was found to conform. This is the set used to characterize the slope maps such as the one generated by slope-based wavefront sensors. These sensors generate two maps, the x and y gradient maps, to which we can fit the OVG polynomials $\underline{s}_x(x, y)$ and $\underline{s}_y(x, y)$, which are additions of the \hat{i} and \hat{j} components, respectively, for each aberration. The resultant coefficients, \underline{b}_x and \underline{b}_y , are added to produce \underline{b} . The process can be repeated using the OVC polynomials $\underline{s}_x(x, y)$ and $\underline{s}_y(x, y)$ to get $\underline{\beta}_x$ and $\underline{\beta}_y$, which are also added together to produce $\underline{\beta}$. The orthonormal coefficients, \underline{c} , are then calculated using Eq. (40) from which the wavefront is reconstructed.

We demonstrate the efficacy of the model in reconstructing a simulated wavefront. The map consists of random coefficients of the first 21 ZCs, which we set in the range $[-0.1; 0.1]\lambda$. Coefficients belonging to the vector $\boldsymbol{\chi}$ are shown in Fig. 5(a) and the resultant phase map is shown in Fig. 5(b). The square pupil inscribed in this pupil is superimposed on the phase map such that the square's vertices touching the circular pupil such that the square's edges are vertical and horizontal. The expansion coefficients for the part of the phase in the square pupil are to be expressed as an expansion of the first 21 ZSq polynomials, \underline{c} , calculated using Eq. (28). The result is illustrated in Fig. 5(c). The phase thus reconstructed is shown in Fig. 5(d), with the rms wavefront error indicated on the graph. The agreement between Figs. 5(b) and 5(d) is apparent. The Cartesian gradient of the slope maps was calculated using Eq. (30) with $\nabla_T \underline{z}(\rho)$ replaced by $\mathbf{H}\vec{\mathbf{M}}\underline{z}(\rho, \theta)$. The result of this effort is shown in Figs. 5(e) and 5(f) in the form of the x and y maps, respectively.

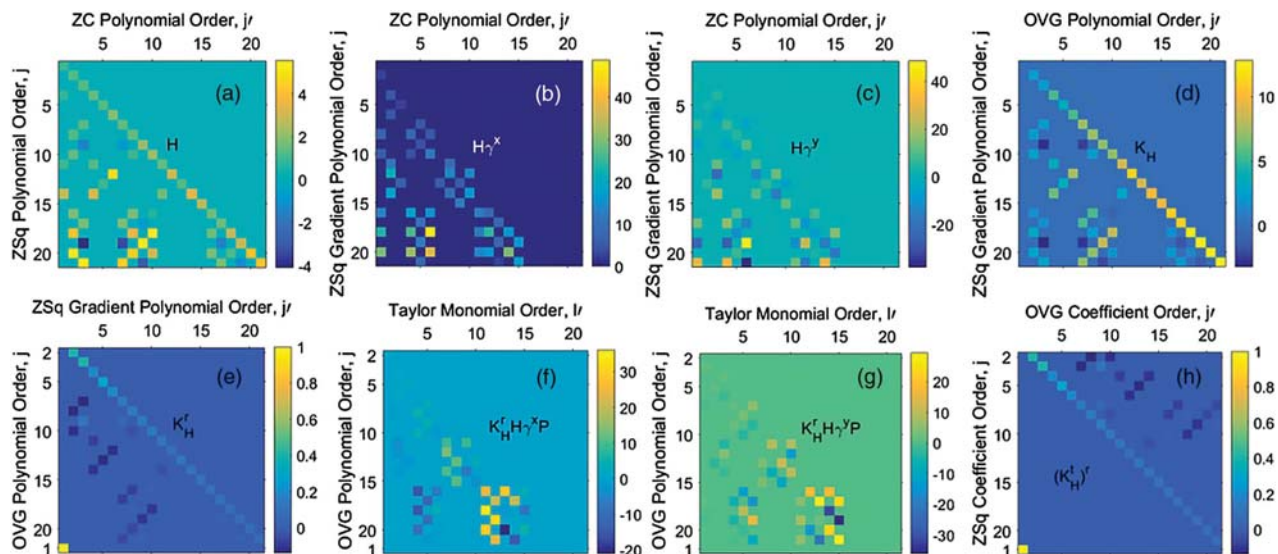


Fig. 4. (a) Transformation matrix for the change of basis matrix in a square pupil for calculating the first 21 Zernike square polynomials from Zernike circle polynomials, \mathbf{H} . This is used to calculate the other matrices such as those representing the Zernike square polynomials Cartesian gradient in terms of (b),(c) ZC polynomials, $\mathbf{H}\boldsymbol{\gamma}^x$ and $\mathbf{H}\boldsymbol{\gamma}^y$; (d) in terms of OVG polynomials, \mathbf{K}_H ; or (e) in terms of OVG polynomials, \mathbf{K}_H^T . (f), (g) OVG polynomials in terms of Taylor monomials, $\mathbf{K}_H^T \mathbf{H}\boldsymbol{\gamma}^x \mathbf{P}$ and $\mathbf{K}_H^T \mathbf{H}\boldsymbol{\gamma}^y \mathbf{P}$; or (h) Zernike square coefficients in terms of OZ Cartesian gradient, $(\mathbf{K}_H^T)^r$.

Table 1. List of the First 21 Cartesian Gradients of the Zernike Square Polynomials in a Square Pupil Inscribed in a Unit Circle Expressed as Functions of ZC Polynomials

$$\begin{aligned} \nabla_T \underline{Z}_1 &= \vec{0}; \nabla_T \underline{Z}_2 = \sqrt{6}Z_1\hat{i}; \nabla_T \underline{Z}_3 = \sqrt{6}Z_1\hat{j}; \nabla_T \underline{Z}_4 = 3\sqrt{5/2}(Z_2\hat{i} + Z_3\hat{j}); \nabla_T \underline{Z}_5 = 3(Z_3\hat{i} + Z_2\hat{j}); \nabla_T \underline{Z}_6 = 3\sqrt{5/2}(Z_2\hat{i} - Z_3\hat{j}); \\ \nabla_T \underline{Z}_7 &= (1/2)\sqrt{7/31}(-15\sqrt{2}Z_5\hat{i} + (15\sqrt{2}Z_6 + 30Z_4 + 16\sqrt{3}Z_1)\hat{j}); \\ \nabla_T \underline{Z}_7 &= (1/2)\sqrt{7/31}((15\sqrt{2}Z_6 + 30Z_4 + 16\sqrt{3}Z_1)\hat{i} + 15\sqrt{2}Z_5\hat{j}); \nabla_T \underline{Z}_8 \\ &= (1/2)\sqrt{5/31}((15\sqrt{2}Z_6 + 30Z_4 + 16\sqrt{3}Z_1)\hat{i} + 9\sqrt{6}Z_5\hat{j}); \\ \nabla_T \underline{Z}_{10} &= (1/2)\sqrt{5/31}(9\sqrt{6}Z_5\hat{i} + (22\sqrt{6}Z_6 - 13\sqrt{3}Z_4 - 21Z_1)\hat{j}); \nabla_T \underline{Z}_{10} = (1/2)\sqrt{5/31}((22\sqrt{6}Z_6 + 13\sqrt{3}Z_4 + 21Z_1)\hat{i} - 9\sqrt{6}Z_5\hat{j}); \\ \nabla_T \underline{Z}_{11} &= 15((7\sqrt{2}Z_8 + 12Z_2)\hat{i} + (7\sqrt{2}Z_7 + 12Z_3)\hat{j})/2\sqrt{67}; \nabla_T \underline{Z}_{12} = 15(7(Z_{10} + Z_8) + 8\sqrt{2}Z_2)\hat{i} + (7(Z_9 - Z_7) - 8\sqrt{2}Z_2)\hat{j}/8; \\ \nabla_T \underline{Z}_{13} &= \sqrt{21}(5(Z_9 + Z_7) + 4\sqrt{2}Z_2)\hat{i} + (5(Z_9 + Z_7) + 4\sqrt{2}Z_2)\hat{j}/4; \\ \nabla \underline{Z}_{14} &= ((67Z_{10} + 43Z_8 + 56\sqrt{2}Z_2)\hat{i} - (67Z_9 - 43Z_7 - 56\sqrt{2}Z_3)\hat{j})/8\sqrt{67}; \vec{\nabla} \underline{Z}_{15} = 5\sqrt{21}(Z_9\hat{i} + Z_{10}\hat{j})/2; \\ \nabla_T \underline{Z}_{16} &= \sqrt{11/983}((315Z_{12} + 315\sqrt{2}Z_{11} + 143\sqrt{15}Z_6 + 121\sqrt{30}Z_4 + 162\sqrt{10}Z_1)\hat{i} + 9(35Z_{13} + 11\sqrt{15}Z_5)\hat{j})/4; \\ \nabla_T \underline{Z}_{17} &= \sqrt{11/983}(9(35Z_{13} + 11\sqrt{15}Z_5)\hat{i} + (-315Z_{12} + 315\sqrt{2}Z_{11} - 143\sqrt{15}Z_6 + 121\sqrt{30}Z_4 + 162\sqrt{10}Z_1)\hat{j})/4; \\ \nabla_T \underline{Z}_{18} &= \frac{3}{16\sqrt{844397}} \left(\frac{(13762\sqrt{30}Z_{14} + 14882\sqrt{15}Z_{11} + 21203\sqrt{30}Z_{12} + 169285\sqrt{2}Z_6 + 127690Z_4 + 65792\sqrt{3}Z_1)\hat{i} +}{\sqrt{2}(13762\sqrt{15}Z_{15} - 6321\sqrt{15}Z_{13} - 41595Z_5)\hat{j}} \right); \\ \nabla_T \underline{Z}_{19} &= \frac{3}{16\sqrt{844397}} \left(\frac{\sqrt{2}(13762\sqrt{15}Z_{15} + 6321\sqrt{15}Z_{13} + 41595Z_5)\hat{i} -}{(13762\sqrt{30}Z_{14} + 14882\sqrt{15}Z_{11} - 21203\sqrt{30}Z_{12} - 169285\sqrt{2}Z_6 + 127690Z_4 + 65792\sqrt{3}Z_1)\hat{j}} \right); \\ \nabla_T \underline{Z}_{20} &= \frac{1}{16}\sqrt{\frac{7}{859}} \left(\frac{(1647\sqrt{10}Z_{14} + 882\sqrt{10}Z_{12} + 1047\sqrt{5}Z_{11} + 2750\sqrt{6}Z_6 + 3275\sqrt{3}Z_4 + 5250Z_1)\hat{i} +}{15(-62\sqrt{10}Z_{15} + 11\sqrt{10}Z_{13} + 35\sqrt{6}Z_5)\hat{j}} \right); \\ \nabla_T \underline{Z}_{21} &= \frac{1}{16}\sqrt{\frac{7}{859}} \left(\frac{15(62\sqrt{10}Z_{15} + 11\sqrt{10}Z_{13} + 35\sqrt{6}Z_5)\hat{i} +}{(1647\sqrt{10}Z_{14} - 882\sqrt{10}Z_{12} + 1047\sqrt{5}Z_{11} - 2750\sqrt{6}Z_6 + 3275\sqrt{3}Z_4 + 5250Z_1)\hat{j}} \right) \end{aligned}$$

The first 21 OVG and OVC polynomials were fitted to the slope maps to generate the sets \underline{b} and $\underline{\beta}$, using a least-squares approach from which we calculated \underline{c} . This was achieved by building a custom equation of the first 20 $\underline{s}_x(x, y)$ and

$\underline{s}_y(x, y)$ polynomials, which are the \hat{i} and \hat{j} components of the equations listed in Table 2 to the respective slope maps. The process was repeated using the first 20 $\underline{s}_x(x, y)$ and $\underline{s}_y(x, y)$ polynomials listed in Table 4 of [1], in which the poly-

Table 2. List of the First 20 Orthonormal Vector Polynomials in a Square Pupil Inscribed in a Unit Circle Expressed as Functions of Cartesian Coordinates

$$\begin{aligned} \vec{S}_2 &= \hat{i}; \vec{S}_3 = \hat{j}; \vec{S}_4 = \sqrt{3}(x\hat{i} + y\hat{j}); \vec{S}_5 = \sqrt{3}(y\hat{i} + x\hat{j}); \vec{S}_6 = \sqrt{3}(x\hat{i} - y\hat{j}); \\ \vec{S}_7 &= (6xy\hat{i} + (3x^2 + 9y^2 - 2)\hat{j})/\sqrt{3}; \\ \vec{S}_8 &= ((9x^2 + 3y^2 - 2)\hat{i} + 6xy\hat{j})/\sqrt{3}; \vec{S}_9 = (12xy\hat{i} + (6x^2 - 12y^2 + 1)\hat{j})/2\sqrt{2}; \\ \vec{S}_{10} &= ((12x^2 - 6y^2 - 1)\hat{i} - 12xy\hat{j})/\sqrt{8}; \\ \vec{S}_{11} &= \sqrt{21/62}(x\hat{i} + y\hat{j})(15x^2 + 15y^2 - 7); \vec{S}_{12} = \sqrt{7}(x(10x^2 - 3)\hat{i} + y(3 - 10y^2)\hat{j})/2; \\ \vec{S}_{13} &= \sqrt{21/38}(x(15x^2 + 5y^2 - 4)\hat{i} + y(5x^2 + 15y^2 - 4)\hat{j}); \\ \vec{S}_{14} &= (x(35x^2 - 27y^2 - 6)\hat{i} + y(-27y^2 - 3x^2 - 6)\hat{j})/\sqrt{5/62}; \vec{S}_{15} = \sqrt{35/3}(y(3x^2 - y^2)\hat{i} + x(x^2 - 3y^2)\hat{j}); \\ \vec{S}_{16} &= (14175(x^2 + y^2)(5x^2 + y^2) - 30(33x^2 + 13y^2) + 83)\hat{i} + 60xy(21(x^2 + y^2) - 13)\hat{j}/2\sqrt{1077}; \\ \vec{S}_{17} &= (60xy(21(x^2 + y^2) - 13)\hat{i} + 315(x^2 + y^2)(x^2 + 5y^2) - 30(33x^2 + 13y^2) + 83)\hat{j}/2\sqrt{1077}; \\ \vec{S}_{18} &= 3(140(860x^4 - 45x^2y^2 - 187y^4) - 30(1685x^2 - 522y^2) + 1279)\hat{i} - 40xy(105x^2 + 2618y^2 - 783)\hat{j}/2\sqrt{2489214}; \\ \vec{S}_{19} &= 3(40xy(2618x^2 + 105y^2 - 783)\hat{i} + 140(187x^4 + 45x^2y^2 - 860y^4) - 30(522x^2 - 1685y^2) - 1279)\hat{j}/2\sqrt{2489214} \\ \vec{S}_{20} &= \frac{1}{16}\sqrt{\frac{7}{13557143}} \left(\frac{(60(10948x^4 - 7830x^2y^2 + 2135y^4 - 3387x^2 - 350y^2) + 11171)\hat{i} -}{1200xy(261x^2 - 427y^2 + 35)\hat{j}} \right); \\ \vec{S}_{21} &= \frac{1}{16}\sqrt{\frac{7}{13557143}} \left(\frac{1200xy(427x^2 - 261y^2 - 35)\hat{i} +}{(60(2135x^4 - 7830x^2y^2 + 10948y^4 - 350x^2 - 3387y^2) + 11171)\hat{j}} \right) \end{aligned}$$

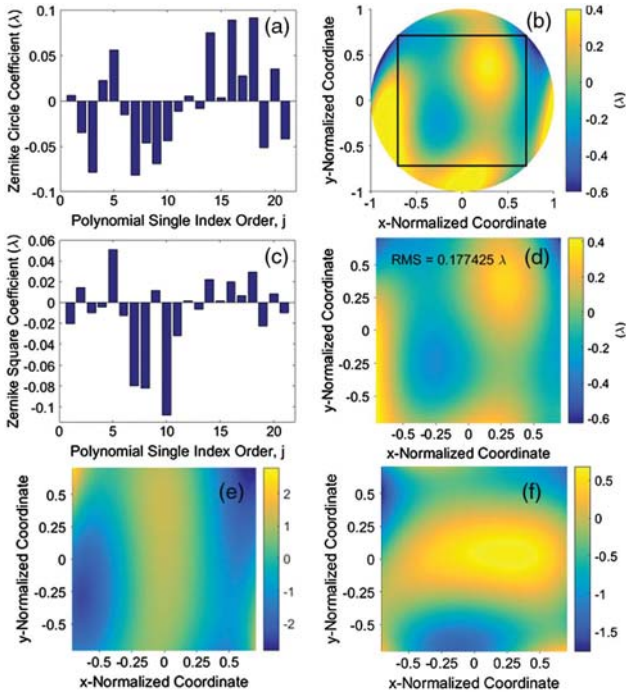


Fig. 5. (a) Set of 21 coefficients, each randomly generated in the range of -0.1 to 0.1 waves. A bar chart showing the generated Zernike circle coefficients, χ . (b) The reconstructed phase in a unit circle showing the inscribed square pupil. (c) A bar chart of the Zernike square coefficients, $\underline{\epsilon}$. (d) The phase map in a square pupil inscribed in a unit circle phase plot showing the calculated RMS error. The calculated (e) x - and (f) y -Cartesian gradients of the phase in the square pupil.

nomials are expressed in terms of ZC polynomials, which, incidentally, can be expressed in terms of Cartesian coordinates. The respective $\underline{\epsilon}$ coefficients were calculated from the result of the fits. The coefficients were then used to reconstruct the wavefronts shown in Figs. 6(a) and 6(b), with the respective wavefront errors of the reconstruction indicated on the graphs. As expected, Figs. 6(a) and 6(b) and their wavefront errors match very well with Fig. 5(d). The plots of the piston-removed residual error maps of both fits are shown in Figs. 6(c) and 6(d) and were acquired by the respective subtraction of both Figs. 6(a) and 6(b) from Fig. 5(d).

7. SUMMARY

We have developed orthogonal vector circle polynomials by deriving the set analytically, using matrix methods giving results which agree with those already in the literature in addition to the natural elimination of the piston. The novelty of our method lies in the fact that it is analytical and matrix based. The motivation for the matrix approach is that, apparently, in a number of important papers in wavefront analysis [4,5,10,11,13–15] it is possible to deal with wavefront analysis problems by manipulating the expansion coefficients in the form of matrices and acquire closed-form results similar to those already published in the literature. The basis of virtually all wavefront work is that the phase can be represented as a linear combination of a discrete infinite series, which can be

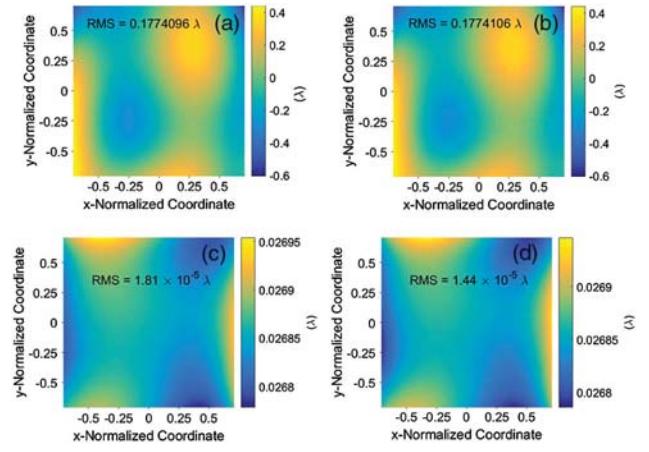


Fig. 6. Reconstructed wavefronts using different approaches outlined in this paper with the indicated rms wavefront error associated with each method. (a) Zernike circle polynomials fitting, (b) Zernike square polynomial fitting. (c) Piston-removed residual errors from OVC polynomial fitting and (d) OVG polynomial fitting.

F

expressed in vector form whose dimension can take any required size. This is based on the fact that the method can be used to derive polynomials with exact expansion coefficients that fulfill the orthonormality requirements in Eq. (15) exactly. The problem with round-off errors can be avoided if numbers under square root signs or those expressed as fractions of two integers are used without converting to their numerical approximations where rounding off becomes a negative factor.

Our model is a robust, elegant model, compatible with most programming languages. The matrices are, in general, sparse and therefore computationally considerably less expensive when working with large matrices. The relevant vector aspects were presented as Euclidean vectors in the form of standard basis vectors in two transverse dimensions, \hat{i} and \hat{j} , as elements making up the various vectors and matrices. The resultant products between these vectors and matrices are in the form of scalar products. The motivation behind selecting this method is that the matrix formalism allows an analytic transformation of the orthonormal vector set orthonormal in a unit circle, once found, to one orthonormal in a general pupil.

Importantly, our approach can be applied to any general pupil and any nonuniform laser beams. The matrix to facilitate this transformation is derived from the matrix form of the Gram–Schmidt procedure that is used to calculate a Zernike-based polynomial set orthonormal in a general pupil from the Zernike circle set. As long as this matrix can be found, then the derivation of the respective OVG set should follow through. These matrices have been derived for annular, hexagonal, rectangular, inscribed square and elliptical [10,11], Gaussian and annular-Gaussian [11,12], and square Gaussian [13] pupils, which implies that finding the OVG sets for these pupils should be implied. These results have shown that for these pupils, exact results have been acquired from the Cholesky decomposition or Gram–Schmidt procedure. The example discussed in this paper shows that the derived polynomials in Table 2 are exact and therefore fulfill the orthonormality condition of Eq. (35) exactly. This proves further that

although the method is matrix based, the outcome for a particular pupil can be analytical if the Cholesky decomposition of Eq. (28) in that pupil gives an analytical result [10].

Furthermore, the model can, potentially, be applied to fit wavefront slopes and analyze intensity moments of light beams in applications that require understanding the propagation of light beams. Consequentially, this has implications in the design of analytical tools for various research fields, such as for ground-based telescope systems and in the development of adaptive optics methods in microscopy.

3 Funding. University of Pretoria (A0W679); National Research Foundation of South Africa (94107).

Acknowledgment. The authors are grateful for funding from the following sources: The University of Pretoria's Postdoctoral Research Fellowship Programme (C. M.), the University of Pretoria's Research Development Programme (T. P. J. K.), and the Thuthuka Programme of the National Research Foundation of South Africa (T. P. J. K. and C. M.).

REFERENCES

1. C. Zhao and J. H. Burge, "Orthonormal vector polynomials in a unit circle Part I: basis set derived from gradients of Zernike polynomials," *Opt. Express* **15**, 18014–18024 (2007).
2. C. Zhao and J. H. Burge, "Orthonormal vector polynomials in a unit circle Part II: completing the basis set," *Opt. Express* **16**, 6586–6591 (2008).
3. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am. A* **66**, 207–211 (1976).
4. P. C. L. Stephenson, "Recurrence relations for the Cartesian derivatives of the Zernike polynomials," *J. Opt. Soc. Am. A* **31**, 708–715 (2014).
5. A. J. E. Janssen, "Zernike expansion of derivatives and Laplacians of the Zernike circle polynomials," *J. Opt. Soc. Am. A* **31**, 1604–1613 (2014).
6. V. N. Mahajan and E. Acosta, "Vector polynomials for direct analysis of circular wavefront data," *J. Opt. Soc. Am. A* **34**, 1908–1913 (2017).
7. J. Alda, J. Alonso, and E. Bernabeu, "Characterization of aberrated beams," *J. Opt. Soc. Am. A* **14**, 2737–2747 (1997).
8. C. Mafusire and A. Forbes, "Generalized beam quality factor of aberrated truncated Gaussian laser beams," *J. Opt. Soc. Am. A* **28**, 1372–1378 (2011).
9. C. Mafusire and A. Forbes, "Mean focal length of an aberrated lens," *J. Opt. Soc. Am. A* **28**, 1403–1409 (2011).
10. V. Mahajan and G.-M. Dai, "Orthonormal polynomials in wavefront analysis: analytical solution," *J. Opt. Soc. Am. A* **24**, 2994–3016 (2007).
11. V. Mahajan, *Optical Imaging and Aberrations, Part III: Wavefront Analysis* (SPIE, 2013).
12. V. Mahajan, "Zernike-Gauss polynomials and optical aberrations of systems with Gaussian pupils," *Appl. Opt.* **34**, 8057–8059 (1995).
13. C. Mafusire and T. P. J. Krüger, "Strehl ratio and the amplitude-weighted generalized orthonormal Zernike-based polynomials," *Appl. Opt.* **56**, 2336–2345 (2017).
14. J. Ye, W. Wang, Z. Gao, Z. Liu, S. Wang, P. Benítez, J. C. Miñano, and Q. Yuan, "Modal wavefront estimation from its slopes by numerical orthogonal transformation method over general shaped aperture," *Opt. Express* **23**, 26208–26220 (2015).
15. M. Li, D. Li, C. Zhang, K. E. Q. Wang, and H. Chen, "Modal wavefront reconstruction from slope measurements for rectangular apertures," *J. Opt. Soc. Am. A* **32**, 1916–1921 (2015).
16. J. T. Scheick, *Linear Algebra with Applications* (McGraw-Hill, 1997).
17. G.-M. Dai, *Wavefront Optics for Vision Correction* (SPIE, 2008).
18. G.-M. Dai, C. E. Campbell, L. Chen, H. Zhao, and D. Chernyak, "Wavefront propagation from one plane to another with the use of Zernike polynomials and Taylor monomials," *Appl. Opt.* **48**, 477–488 (2009).
19. D. S. Watkins, *Fundamentals of Matrix Computations*, 2nd ed. (Wiley, 2002).