Use of Zernike polynomials for efficient estimation of orthonormal aberration coefficients over variable noncircular pupils

Hanshin Lee

McDonald Observatory, University of Texas at Austin, 1 University Station C1402, Austin, Texas 78712, USA (lee@astro.as.utexas.edu)

Received February 1, 2010; revised May 24, 2010; accepted May 26, 2010; posted June 9, 2010 (Doc. ID 123632); published June 22, 2010

An efficient way of estimating orthonormal aberration coefficients on variable noncircular pupils is proposed. The method is based on the fact that all necessary pieces of information for constructing orthonormal polynomials (via the Gram–Schmidt process) can be numerically obtained during a routine least-squares fit of Zernike polynomials to wavefront data. This allows the method to use the usual Zernike polynomial fitting with an additional procedure that swiftly estimates the desired orthonormal aberration coefficients without having to use the functional forms of orthonormal polynomials. It is also shown that the method naturally accounts for the pixelation effect of pupil geometries, intrinsic to recording wavefront data on imaging sensors (e.g., CCDs), making the coefficient estimate optimal over a given pixelated pupil geometry. With these features, the method can be ideal for real-time wavefront analysis over dynamically changing pupils, such as in the Hobby–Eberly Telescope (HET), which is otherwise inefficient with analytic methods used in past studies.

OCIS codes: 010.1080, 080.1010, 120.3180, 220.4840.

Aberration coefficients of optical wavefront are affected by the physical state of systems where the wavefront has propagated (e.g., the Earth atmosphere or a telescope). Thus, useful insights into the observed behaviors of the systems can be acquired by estimating the coefficients from wavefront measurements. In doing so, the estimated coefficients are desired to be unique, and this is enabled by fitting a set of orthonormal polynomials to the measurement [1]. Zernike polynomials, which constitute one particular such set over a unit disk and are used in this Letter, have been applied [2–4]. However, there are increasingly many optical systems whose pupils are noncircular [5–7]. In addition, some systems exhibit variability in pupil shape as a function of field or pointing angles [8,9]. On such pupils, Zernike polynomials lose orthogonality, and it is desirable to use the coefficients of new orthonormal aberration polynomials.

Studies show that orthonormal polynomials can be analytically constructed, via the Gram–Schmidt (GS) process, for noncircular pupils in simple shape without shape variability, i.e., annulus, hexagon, ellipse, and rectangle [10–13]. The studies overlook issues in applying the analytic method to wavefront data measured over more complex pixelated (due to pixel-based imaging sensors) pupil shapes with dynamic variability. The analytic method can be extended to general pupils, but the analytic calculations can be challenging over more complicated pixelated pupil geometries, such as that of the Hobby–Eberly Telescope (HET). Any dynamic change in pupil shape complicates the execution of the method even more, as the analytic forms of orthonormal polynomials have to be obtained whenever the change happens. This can also slow the orthonormal aberration coefficient estimation in real-time wavefront analysis, thereby diminishing the advantage of having orthonormal coefficients over variable noncircular pupils.

We propose an efficient way of estimating orthonormal aberration coefficients that can resolve the problems mentioned previously. The estimation uses the usual numerical fit of Zernike polynomials to wavefront data recorded in a pixelated format, but essentially without having to compute the functional forms of the orthonormal polynomials (which can be complex and time-consuming), thereby keeping simplicity and speed of the routine least-squares Zernike fit. The orthonormal coefficients obtained in this way, naturally account for the pupil pixelation effect and are, thus, optimal over a given pixelated pupil geometry. The details of the method are described as follows.

The Zernike polynomials (Z_i) can be identified by radial and angular orders $[n,m]$ [1]. The polynomial index scheme used here follows: $Z_i[n1,m1]$ precedes $Z_j[n2,m2]$ if $n1<n2$; if $n1=n2$ and $|m1|<|m2|$, Z_i precedes Z_j; if $n1=n2$, $|m1|=|m2|$, and $m1<0$, Z_i precedes Z_j. Suppose a wavefront function $W(x,y)$ on a noncircular pupil E that is a subset of a unit disk S. Assume that W can be described in terms of M Zernike polynomials,

$$W=\alpha_1 Z_1+\alpha_2 Z_2+\ldots+\alpha_M Z_M. \tag{1}$$

The estimation of α_i starts with the following (A the area of E):

$$\sum_{j=1}^{M}\alpha_j F_{ij}=\sum_{j=1}^{M}\alpha_j\iint_E \frac{Z_i Z_j}{A}\mathrm{d}A=\iint_E \frac{Z_i W}{A}\mathrm{d}A. \tag{2}$$

Z_i are not orthonormal over E, and thus a $M\times M$ matrix $\mathbf{F}$, whose (i,j) element is F_{ij}, is neither unitary nor diagonal. The estimate of α_i can still be given by

$$\alpha_{i,e}=\sum_{j=1}^{M}G_{ij}\iint_E \frac{Z_i W}{A}\mathrm{d}A, \tag{3}$$

with G_{ij} being the (i,j) element of $\mathbf{G}=\mathbf{F}^{-1}$. $\alpha_{i,e}$ is neither α_i nor the desired orthonormal aberration coefficient (β_i) of W on E, but closely related to β_i as follows.

0146-9592/10/132173-03$15.00/0

It is widely known that new orthonormal polynomials, say U_i, can be determined by the GS orthonormalization of a known set of orthonormal polynomials (Z_j in our case) [14]. To utilize the GS orthonormalization, we allow $U_1 = Z_1$:

$$V_j = Z_j - \sum_{k=1}^{j-1} C_{jk} U_k, \qquad U_j = V_j \sqrt{A \Big/ \iint_E V_j V_j \mathrm{d}A}, \quad (4)$$

with $C_{jk} = A^{-1}\iint_E Z_j U_k \mathrm{d}A$ where $j = 1, 2, \ldots, M$. Using Eq. (4), the following algebraic relation is obtained:

$$\iint_E \frac{V_i V_j}{A}\mathrm{d}A = F_{ij} + \sum_{m=1}^{i-1}\left\{\sum_{n=1}^{j-1} C_{im}\delta_{mn}C_{nj}\right\} - \sum_{n=1}^{i-1} C_{in}C_{nj} - \sum_{m=1}^{j-1} C_{im}C_{mj}, \quad (5)$$

with δ_{mn} being the Kronecker delta. Given the orthogonality of V_i, the left-hand side of Eq. (5) equals C_{ii}^2 for $i = j$ or vanishes for $i \neq j$. This reduces Eq. (5) to the following algebraic expressions of the GS orthonormalization coefficient C_{ij}:

$$C_{ii} = \sqrt{F_{ii} - \sum_{k=1}^{i-1} C_{ik}^2}, \qquad C_{ij} = \frac{F_{ij}}{C_{jj}} - \sum_{k=1}^{j-1}\frac{C_{ik}C_{jk}}{C_{jj}}. \quad (6)$$

$C_{i,j}$ is essentially the (i,j) element of a $M \times M$ lower triangle matrix $\mathbf{C}$, which is the Cholesky factor of $\mathbf{F}$, as noted in a different way elsewhere [12]. From Eq. (4), $Z_i = \sum_{j=1}^{i} C_{ij}U_j$ by which W can be given as

$$W \approx \sum_{i=1}^{M} \alpha_{i,e} Z_i = \sum_{i=1}^{M}\sum_{j=1}^{i} \alpha_{i,e} C_{ij} U_j = \sum_{j=1}^{M} \beta_j U_j, \qquad \text{leading to} \quad \beta_j = \sum_{i=1}^{M} \alpha_{i,e} C_{ij}. \quad (7)$$

Note also $U_i = \sum_{j=1}^{i} D_{ij} Z_j$, where D_{ij} is the (i,j) element of $\mathbf{D} = \mathbf{C}^{-1}$. The above orthonormalization process clearly indicates that knowing F_{ij} is central to computing C_{ij} and constructing analytic forms of U_i, via D_{ij}, which are fitted to W to determine β_i. However, analytically computing F_{ij} over a complex pupil can be out of the question. Although computer algebra software or numerical quadrature could be helpful, it is still required to extract (or approximate) the pupil geometry from a measured pupil illumination or to divide the geometry into a number of integrable simpler subdomains, which would make the whole analytic approach increasingly complicated and vastly inefficient when large sets of orthonormal polynomials need to be obtained for variable noncircular pupils. In addition, it would be much harder to incorporate the pixelation effect, intrinsic to digital wavefront data, into analytic derivations of F_{ij} and U_i, and ultimately into estimating β_i.

However, realizing that the wavefront is recorded on N discrete pixels on E ($\vec{W}$ a $N \times 1$ vector), Eq. (1) is given as a least-squares Zernike fit in a matrix form:

$$\vec{W} = \mathbf{Z}\vec{\alpha} \Rightarrow \hat{\vec{\alpha}}_e = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\vec{W} \approx \vec{\alpha}, \quad (8)$$

where $\vec{\alpha}$ is the column vectors of α_i and $\mathbf{Z}$ is a $N \times M$ matrix of M Zernike polynomials evaluated on *the pixels* (E_{pix}). $\hat{\alpha}_{i,e}$ (the ith element of $\hat{\vec{\alpha}}_e$) is the estimate of α_i on E_{pix}. Realizing $\mathbf{F} \approx \hat{\mathbf{F}} = (\mathbf{Z}^T\mathbf{Z})/N$ leads to

$$\hat{\beta}_j = \sum_{i=1}^{M} \hat{\alpha}_{i,e}\hat{C}_{ij} \approx \beta_j, \quad (9)$$

where $\hat{C}_{ij}$, the estimate of C_{ij}, is given by $\hat{\mathbf{F}}$ in Eq. (6). The significance of Eq. (8) and (9) is threefold: (i) all estimates are natural by-products of the least-squares fit of Zernike polynomials to the wavefront data, (ii) the estimated coefficients are obtained over the given pixelated pupil geometry but essentially without knowing the functional forms of the orthonormal polynomials on the true pupil E, and (iii) the optimality of $\hat{\beta}_j$ on E_{pix} while retaining simplicity and speed of the usual least-squares Zernike fit.

To test the proposed method, simple demonstrations are conducted. Two wavefronts are generated over a 256×256 square grid by two sets of the first 36 Zernike coefficients, given from the standard Kolmogorov turbulence model with $D/r_0 = 1.4$ [3]. The wavefronts are filtered through two pupils: Figs. 1(a) and 1(b), respectively.

The first example is performed on pupil A, for which we analytically derive $\mathbf{D}$ using Maxima [15] for a later comparison. The given Zernike coefficients are converted to β of which the first 15 are listed in the "True β_i" column in Table 1. These coefficients are then estimated by fitting the known orthonormal polynomials, built by using $\mathbf{D}$, to the wavefront (method X) and by the proposed method (method Y). The difference between the two methods is shown in the "$|X - Y|$" column. The proposed method reproduces the coefficients given by the analytic method with a relative difference of less than 1% in most cases. The relative difference between the wavefront variances (σ^2) given by the two methods is negligible (less than 0.2%) as compared to the true value on pupil A. The proposed method also gives $\hat{\mathbf{F}}$, from which $\hat{\mathbf{D}}$ (the estimate of $\mathbf{D}$) is given. In Table 2, $\hat{D}_{ij} \neq 0$ for i,j up to 11 are compared to those in $\mathbf{D}$. We omit $\hat{D}_{ij}$, whose absolute values are less than the expected estimation error of $\sim 3 \times 10^{-3}$. Method Y identifies all coefficients important for analytically expressing the

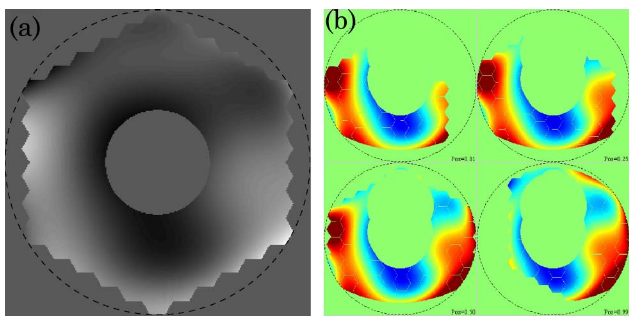

Fig. 1. (Color online) Two sample wavefronts on two pupils with a unit circle (dashed line): left, pupil A; right, pupil B (Media 1).

Table 1. $\hat{\beta}_i$ by Method X and Y on Pupil A

i	True β_i	Method X	Method Y	|X – Y|
1	−0.0076	−0.0076	−0.0072	0.434×10^{-3}
2	−0.1133	−0.1134	−0.1134	0.051×10^{-3}
3	−0.0233	−0.0234	−0.0235	0.055×10^{-3}
4	0.2741	0.2740	0.2741	0.103×10^{-3}
5	−0.0501	−0.0502	−0.0502	0.012×10^{-3}
6	0.1706	0.1706	0.1706	0.023×10^{-3}
7	−0.1108	−0.1110	−0.1107	0.247×10^{-3}
8	−0.0409	−0.0411	−0.0411	0.025×10^{-3}
9	−0.1064	−0.1062	−0.1062	0.070×10^{-3}
10	−0.0369	−0.0365	−0.0365	0.001×10^{-3}
11	−0.0648	−0.0648	−0.0645	0.310×10^{-3}
12	−0.0057	−0.0058	−0.0058	0.063×10^{-3}
13	−0.1070	−0.1071	−0.1072	0.109×10^{-3}
14	−0.0750	−0.0748	−0.0749	0.111×10^{-3}
15	0.0464	0.0466	0.0466	0.018×10^{-3}
σ_W^2	0.1704	0.1703	0.1704	0.034×10^{-3}

new polynomials. Also, $\hat{D}_{ij}$ are close to D_{ij} except for a few with small contributions. Because of analytically deriving **D**, method X took ~72 s longer than method Y did on the same computer. Thus, the efficiency gain by using method Y in dealing with variable noncircular pupils can be immense.

In the second example, the variance of the wavefront on pupil B [Fig. 1(b)] is given in three ways: the true variance of the data (σ^2), the sum of squares of the first 15 of $\hat{\beta}_i(\sigma^2(\beta))$, and the sum of squares of the first 15 of $\hat{\alpha}_{i,e}(\sigma^2(\alpha))$. Note the dynamic change in the pupil B shape as a function of the pupil position parameter. These variances are plotted in Fig. 2. As expected, $\sigma^2(\beta)$ closely approximates σ^2, while $\sigma^2(\alpha)$ differs from the true value. As the pupil shape becomes closer to the unit circle (i.e., position ~0.7), $\sigma^2(\alpha)$ asymptotes the other two curves, but quickly diverts away as the pupil shape starts departing from the unit disk again. The last 21 of $\hat{\beta}_i$ are also estimated independently, and the sum of squares of these $(\Delta\sigma^2(\beta))$ matches the true difference $\Delta\sigma^2 = \sigma^2(\beta) - \sigma^2$.

As demonstrated, the proposed method can be used to efficiently estimate orthonormal aberration coefficients of wavefront data over noncircular pixelated pupils with

Table 2. D_{ij} by Analytic Integration and $\hat{D}_{ij}$ by Proposed Method on Pupil A

i,j	$D_{i,j}$	$\hat{D}_{ij}$	i,j	D_{ij}	$\hat{D}_{ij}$
01,01	1.000	1.000	08,08	1.214	1.214
02,02	1.025	1.025	09,02	−0.006	−0.003
03,03	1.022	1.022	09,07	−0.006	−0.003
04,01	0.109	0.108	09,09	1.068	1.068
04,04	1.355	1.355	10,03	−0.008	−0.011
05,05	1.106	1.106	10,08	−0.007	−0.009
06,01	−0.004	−0.004	10,10	1.385	1.385
06,04	−0.009	−0.008	11,01	0.827	0.827
06,06	1.105	1.104	11,04	0.098	0.096
07,02	0.495	0.493	11,06	−0.007	−0.006
07,07	1.221	1.221	11,11	1.632	1.634
08,03	0.484	0.484	—	—	—

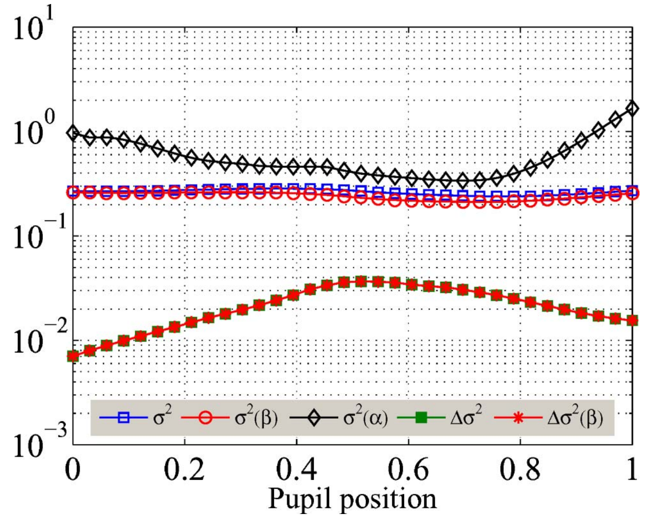

Fig. 2. (Color online) Variance estimates by different methods.

shape variability. It uses the routine least-squares fit of Zernike polynomials to wavefront measurements, but without having to know the functional forms of the orthonormal polynomials. This effectively eliminates the complex and time-consuming part from the estimation, thereby keeping the simplicity and speed of the usual least-squares fit, while ensuring the optimality of the resultant orthonormal aberration coefficients over a given pixelated pupil geometry. These features not only are ideal for real-time wavefront analysis over dynamically varying pixelated pupils but are also useful in extending the method to estimating orthonormal slope aberration coefficients, where *mean* wavefront slope is measured over a further complicated geometry formed by variable noncircular pupils and a coarse slope sampling grid [16].

References

1. M. Born and E. Wolf, *Principles of Optics*, 7th ed. (Oxford, 1999).
2. F. Zernike, Mon. Not. R. Astron. Soc. **94**, 377 (1934).
3. N. Roddier, Opt. Eng. **29**, 1174 (1990).
4. H. Lee, G. B. Dalton, I. A. Tosh, S.-W. Kim, Opt. Express **15**, 3127 (2007).
5. W. B. King, Appl. Opt. **7**, 197 (1968).
6. Hubble Space Telescope, http://hubblesite.org/.
7. Keck Telescopes, http://www.keckobservatory.org/.
8. Hobby–Eberly Telescope, http://www.as.utexas.edu/.
9. South African Large Telescope, http://www.salt.ac.za/.
10. C. F. Dunkl, SIAM J. Appl. Math. **47**, 343 (1987).
11. R. Upton and B. Ellerbroek, Opt. Lett. **29**, 2840 (2004).
12. G. Dai and V. Mahajan, Opt. Lett. **32**, 74 (2007).
13. V. N. Mahajan and G. Dai, J. Opt. Soc. Am. A **24**, 2994 (2007).
14. G. B. Arfken, H. J. Weber, and F. Harris, *Mathematical Methods for Physicists*, 5th ed. (Academic, 2001).
15. Maxima version 5.18.1 (a computer algebra system), http://maxima.sourceforge.net/.
16. H. Lee, M. Hart, and G. J. Hill are preparing a manuscript to be called "Optimal estimation of wavefront slope aberrations on variable non-circular pupils."