Zernike Circle Polynomials and Optical Aberrations of Systems with Circular Pupils

By Virendra N. Mahajan, The Aerospace Corporation, Los Angeles, California 90009

Abstract
Zernike circle polynomials, their numbering scheme, and relationship to balanced optical aberrations of systems with circular pupils are discussed.

Introduction
Zernike polynomials were introduced by Zernike for his phase contrast method for testing the figure of circular mirrors figures.[1] They were used by Nijboer[2,3] to balance the classical aberrations of a power-series expansion of the aberration function of an optical imaging system and to study the effects of small aberrations on the diffraction images formed by rotationally symmetric systems with circular pupils. Noll used them to describe the aberrations introduced by Kolmogorov atmospheric turbulence.[4] Today, they are in widespread use in optical design as well as in optical testing. However, there appears to be no standard for their form or their ordering, *i.e.*, numbering.[5–7] In this first of several notes, we outline the characteristics of Zernike circle polynomials and reemphasize the use of their orthonormal form and ordering first suggested by Noll. The use of orthonormal polynomials in the expansion of an aberration function has the advantage that the coefficients of the expansion terms represent their standard deviations. The ordering scheme discussed lends itself to easy calculation of the number of terms through a certain order in the expansion for general as well as rotationally symmetric optical systems. In future *Notes* we will discuss Zernike annular polynomials[8] appropriate for systems with annular pupils as well as Zernike-Gauss polynomials,[8,9] which are suitable for systems with Gaussian pupils. Unlike circle polynomials, these polynomials are not readily available, especially, in one place. It is hoped that these issues of *Notes* will fill this gap.

Zernike Circle Polynomials and the Aberration Function
Consider an optical system with a circular pupil of radius a. Let (r, θ) be the polar coordinates of a point on the pupil. Let $\rho = r/a$ so that $0 \le \rho \le 1$. Of course, $0 \le \theta < 2\pi$. The wave aberration function $W(\rho, \theta)$ of the system can be expanded in terms of a complete set of Zernike circle polynomials $R_n^m(\rho)\cos m\theta$ and $R_n^m(\rho)\sin m\theta$, which are orthogonal over a unit circle in the form:

$$W(\rho, \theta) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} [2(n+1)/(1+\delta_{m0})]^{1/2} R_n^m(\rho) \cdot (c_{nm} \cos m\theta + s_{nm} \sin m\theta), \quad (1)$$

where c_{nm} and s_{nm} are the expansion or the aberration coefficients, n and m are positive integers including zero, $n - m \ge 0$ and even, δ_{ij} is a Kronecker delta, and

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2} - s\right)!\left(\frac{n-m}{2} - s\right)!} \rho^{n-2s} \quad (2)$$

is a polynomial of degree n in ρ containing terms in ρ^n, ρ^{n-2}, . . ., and ρ^m. The radial polynomials $R_n^m(\rho)$ are even or odd in ρ depending on whether n (or m) is even or odd, respectively. Note that

$$R_n^m(0) = \begin{cases} \delta_{m0}, & n/2 \text{ even} \\ -\delta_{m0}, & n/2 \text{ odd} \end{cases}, \qquad R_n^m(1) = 1, \qquad R_n^n(\rho) = \rho^n. \quad (3)$$

The index n represents the radial *degree* or the *order* of the polynomial since it represents the highest power of ρ in the polynomial, and m may be called the azimuthal frequency.

The orthogonalities of the radial polynomials and the angular functions are:

$$\int_0^1 R_n^m(\rho) R_{n'}^m(\rho) \rho d\rho = \frac{1}{2(n+1)} \delta_{nn'}, \quad (4)$$

$$\int_0^{2\pi} \cos m\theta \cos m'\theta d\theta = \pi(1 + \delta_{m0})\delta_{mm'}, \quad (5a)$$

$$\int_0^{2\pi} \cos m\theta \sin m'\theta d\theta = 0, \quad (5b)$$

and

$$\int_0^{2\pi} \sin m\theta \sin m'\theta d\theta = \pi\delta_{mm'}. \quad (5c)$$

The expansion or the aberration coefficients are given by:

$$(c_{nm}, s_{nm}) = (1/\pi)[2(n+1)/(1+\delta_{m0})]^{1/2} \cdot \int_0^1 \int_0^{2\pi} W(\rho, \theta) R_n^m(\rho)(\cos m\theta, \sin m\theta) \rho d\rho d\theta, \quad (6)$$

as may be seen by substituting Eq. (1) into Eq. (6). It should be evident that $s_{n0} = 0$. We note that the angular dependence of an aberration term consists of cosine (or the sin) of an

Engineering & Laboratory Notes

EDITOR
Robert R. Shannon
University of Arizona

STAFF
Catherine B. Kaufman, Production Editor
Alexine Moore, Manuscript Assistant
Andrea Pendleton, Director of Member Services

Please *mail* manuscripts to:
Engineering & Laboratory Notes
c/o Manuscript Office
2010 Massachusetts Ave., NW
Washington, D.C. 20036

For questions on manuscript submission and status, please call: 202/416-1916 or e-mail opn@osa.org

For author guidelines and questions on manuscripts in production, please call: 202/416-1420 or e-mail opn@osa.org

integral multiple of angle θ rather than the integral power of the cosine of the angle as in a power-series expansion of the aberration function in pupil coordinates. The aberration terms of a power-series expansion are called classical aberrations. Because of their orthogonality, the aberration terms of a Zernike-polynomial expansion are referred to as the orthogonal aberrations. The orthonormal Zernike aberrations and the names associated with some of them are also listed in Table 1 for $n \leq 8$. The number of aberration terms in the expansion of the aberration function through a certain order n is given by

$$N_n = (n+1)(n+2)/2. \tag{7}$$

Note that piston term with $n = m = 0$ does not constitute an aberration, although it is counted as such in Eq. 7. For example, the number of aberrations through the fourth order ($n \leq 4$) is 15.

We note that each Zernike or orthogonal aberration is made up of one or more classical aberrations. The classical aberration of the highest degree in pupil coordinates is optimally balanced with those of equal and lower degrees such that its variance across the pupil is minimized. Accordingly, a Zernike polynomial aberration may also be referred to as a balanced aberration. For example, the Zernike primary spherical aberration $R_4^0(\rho)$ consists of a classical primary spherical aberration (ρ^4 term) optimally balanced with defocus (ρ^2 term) to minimize its variance. It may be called balanced primary spherical aberration. Similarly, the Zernike secondary spherical aberration $R_6^0(\rho)$ consists of a classical secondary spherical aberration (ρ^6 term) optimally balanced with primary classical spherical aberration and defocus, and may be called balanced secondary spherical aberration. Inclusion of the constant term in these aberrations makes their mean value zero. Since unity $[R_0^0(\rho)]$ is one of the Zernike aberrations, the orthogonality of an aberration also implies that its mean value is zero. The Zernike primary coma $R_3^1(\rho)\cos\theta$ consists of classical primary coma ($\rho^3 \cos\theta$ term) optimally balanced with tilt ($\rho \cos\theta$ term) and may be called balanced coma. The Zernike primary astigmatism $R_2^2(\rho)\cos 2\theta$ consists of classical primary astigmatism ($\rho^2\cos^2\theta$ term) optimally balanced with defocus.

The Zernike polynomials are unique in that they are the only complete set of polynomials in two coordinate variables ρ and θ that are (a) orthogonal over a unit circle, (b) are invariant in form with respect to rotation of the axes about the origin, and (c) include a polynomial for each permissible pair of n and m values. They are used in optical design and testing for expressing an aberration function because of their association with optimally balanced classical aberrations.

The aberration function may also be written in terms of orthonormal Zernike circle polynomials $Z_j(\rho, \theta)$ in the form

$$W(\rho,\theta) = \sum_{j=1}^{\infty} a_j Z_j(\rho,\theta), \tag{8}$$

where the index j is a polynomial-ordering number, which is a function of both n and m, a_j is the expansion or aberration coefficient, and

$$Z_{\text{even}j}(\rho,\theta) = \sqrt{2(n+1)}\, R_n^m(\rho)\cos m\theta, \qquad m \neq 0, \tag{9a}$$

$$Z_{\text{odd}j}(\rho,\theta) = \sqrt{2(n+1)}\, R_n^m(\rho)\sin m\theta, \qquad m \neq 0, \tag{9b}$$

$$Z_j(\rho,\theta) = \sqrt{n+1}\, R_n^0(\rho), \qquad m = 0. \tag{9c}$$

The relationships among the indices j, n, and m are given in Table 1. The polynomials are ordered such that even j corresponds to a symmetric polynomial given by $\cos m\theta$, while odd j corresponds to an antisymmetric polynomial varying as $\sin m\theta$. For a given value of n, a polynomial with a lower value of m is ordered first. This ordering is different from those considered in recent publications.[10–12]

The orthonormality of Zernike polynomials implies that:

$$\int_0^1 \int_0^{2\pi} Z_j(\rho,\theta) Z_{j'}(\rho,\theta)\rho d\rho d\theta \Big/ \int_0^1 \int_0^{2\pi} \rho d\rho d\theta = \delta_{jj'}. \tag{10}$$

The expansion coefficients a_j are given by

$$a_j = \pi^{-1} \int_0^1 \int_0^{2\pi} W(\rho,\theta) Z_j(\rho,\theta)\rho d\rho d\theta. \tag{11}$$

Aberration Variance

An advantage of the orthogonal-polynomial expansion of the aberration function in the form of Eq. 1 is that each aberration coefficient c_{nm} or s_{nm} represents the standard deviation of the corresponding aberration term across the pupil, and, therefore, it is very easy to determine the standard deviation of the aberration function once the expansion coefficients are known. We note that the mean and mean square values of the aberration function are given by:

$$\begin{aligned}\langle W(\rho,\theta)\rangle &= \int_0^1 \int_0^{2\pi} W(\rho,\theta)\rho d\rho d\theta \Big/ \int_0^1 \int_0^{2\pi} \rho d\rho d\theta \\ &= c_{00}\end{aligned} \tag{12}$$

(since $\int_0^{2\pi} \cos m\theta d\theta = 2\pi\delta_{m0}$), and

$$\begin{aligned}\langle W^2(\rho,\theta)\rangle &= \int_0^1 \int_0^{2\pi} W^2(\rho,\theta)\rho d\rho d\theta \Big/ \int_0^1 \int_0^{2\pi} \rho d\rho d\theta \\ &= \sum_{n=0}^{\infty} \sum_{m=0}^{n} (c_{nm}^2 + s_{nm}^2),\end{aligned} \tag{13}$$

as may be seen by substituting Eq. (1) and using the orthogonality Eqs. (4) and (5). Hence, the variance of the aberration function is given by:

$$\begin{aligned}\sigma_w^2 &= \langle W^2(\rho,\theta)\rangle - \langle W(\rho,\theta)\rangle^2 \\ &= \sum_{n=1}^{\infty} \sum_{m=0}^{n} (c_{nm}^2 + s_{nm}^2),\end{aligned} \tag{14}$$

where σ_w is its standard deviation. The root mean square (rms) value of the aberration function is given by $\langle W^2(\rho, \theta)\rangle^{1/2}$. This is not equal to its standard deviation σ_w unless its mean value $c_{00} = 0$. Since the mean value of an aberration term (except the piston) is zero, its rms value is

Table 1. Orthonormal Zernike *circle* polynomials $Z_j(\rho, \theta)$. The indices j, n, and m are defined as the polynomial number, radial degree, and azimuthal frequency, respectively. The polynomials Z_j are ordered such that even j corresponds to a symmetric polynomial defined by $\cos m\theta$, while odd j corresponds to an antisymmetric polynomial given by $\sin m\theta$. For a given n, a polynomial with a lower value of m is ordered first. $x = \rho\cos\theta$, $y = \rho\sin\theta$, $0 \le \rho \le 1$, $0 \le \theta < 2\pi$.

j	n	m	$Z_j(\rho, \theta)$	Name
1	0	0	1	Piston
2	1	1	$2\rho\cos\theta$	x tilt
3	1	1	$2\rho\sin\theta$	y tilt
4	2	0	$\sqrt{3}(2\rho^2 - 1)$	Defocus
5	2	2	$\sqrt{6}\rho^2\sin 2\theta$	
6	2	2	$\sqrt{6}\rho^2\cos 2\theta$	Astigmatism
7	3	1	$\sqrt{8}(3\rho^3 - 2\rho)\sin\theta$	Primary y coma
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho)\cos\theta$	Primary x coma
9	3	3	$\sqrt{8}\rho^3\sin 3\theta$	
10	3	3	$\sqrt{8}\rho^3\cos 3\theta$	
11	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	Primary spherical
12	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2)\cos 2\theta$	Secondary astigmatism
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2)\sin 2\theta$	
14	4	4	$\sqrt{10}\rho^4\cos 4\theta$	
15	4	4	$\sqrt{10}\rho^4\sin 4\theta$	
16	5	1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho)\cos\theta$	Secondary x coma
17	5	1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho)\sin\theta$	Secondary y coma
18	5	3	$\sqrt{12}(5\rho^5 - 4\rho^3)\cos 3\theta$	
19	5	3	$\sqrt{12}(5\rho^5 - 4\rho^3)\sin 3\theta$	
20	5	5	$\sqrt{12}\rho^5\cos 5\theta$	
21	5	5	$\sqrt{12}\rho^5\sin 5\theta$	
22	6	0	$\sqrt{7}(20\rho^6 - 30\rho^4 + 12\rho^2 - 1)$	Secondary spherical
23	6	2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2)\sin 2\theta$	
24	6	2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2)\cos 2\theta$	
25	6	4	$\sqrt{14}(6\rho^6 - 5\rho^4)\sin 4\theta$	
26	6	4	$\sqrt{14}(6\rho^6 - 5\rho^4)\cos 4\theta$	
27	6	6	$\sqrt{14}\rho^6\sin 6\theta$	
28	6	6	$\sqrt{14}\rho^6\cos 6\theta$	
29	7	1	$4(35\rho^7 - 60\rho^5 + 30\rho^3 - 4\rho)\sin\theta$	Tertiary y coma
30	7	1	$4(35\rho^7 - 60\rho^5 + 30\rho^3 - 4\rho)\cos\theta$	Tertiary x coma
31	7	3	$4(21\rho^7 - 30\rho^5 + 10\rho^3)\sin 3\theta$	
32	7	3	$4(21\rho^7 - 30\rho^5 + 10\rho^3)\cos 3\theta$	
33	7	5	$4(7\rho^7 - 6\rho^5)\sin 5\theta$	
34	7	5	$4(7\rho^7 - 6\rho^5)\cos 5\theta$	
35	7	7	$4\rho^7\sin 7\theta$	
36	7	7	$4\rho^7\cos 7\theta$	
37	8	0	$3(70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1)$	Tertiary spherical

equal to its standard deviation, which in turn is simply equal to its aberration coefficient as defined in Eq. (1).

Considering the expansion of the aberration function given by Eq. (8), we find that

$$\langle W(\rho,\theta)\rangle = a_1 \tag{15a}$$

and

$$\langle W^2(\rho, \theta)\rangle = \sum_{j=1}^{\infty} a_j^2. \tag{15b}$$

Thus, except for a_1, the expansion coefficients a_j's represent the standard deviation of the corresponding jth term. The variance of the aberration function is given by

$$\sigma_w^2 = \sum_{j=2}^{\infty} a_j^2. \tag{16}$$

Rotationally Symmetric Systems

The aberration function of an optical imaging system that is rotationally symmetric about its optical axis (*i.e.*, the line joining the vertices of its surfaces) must be symmetric about the tangential plane (which contains the optical axis and the point object for which the aberration function is being considered). Hence, if the angle θ is measured from the tangential plane; for example, if the x axis lies in the tangential plane, then sin $m\theta$ terms of Eqs. 1 and 8 must be zero; *i.e.*, the aberration coefficients s_{nm} must be zero. Thus, in the design of rotationally symmetric optical systems, only $\cos m\theta$ terms need to be considered. In that case, the number of aberration terms through a certain order n is given by

$$N_n = (n + 2)(n + 4)/8. \tag{17}$$

The number of aberration terms through the fourth order is now equal to 6. They consist of piston and the terms that correspond to the five Seidel or primary classical aberrations.

Discussion

In the fabrication and testing of rotationally symmetric optical elements, the fabrication errors will generally consist of both the cos $m\theta$ and sin $m\theta$ terms, even though the design aberrations of a rotationally symmetric system will consist of only the former. Similarly, aberrations introduced by thermal distortion of optical elements may consist of both types of terms. The ordering of Zernike polynomials as in Table 1 does not imply that the aberration coefficients decrease as n increases. It is quite possible, for example, that a_{11} is larger than a_9 or a_{10}. The random aberrations introduced by Kolmogorov atmospheric turbulence, on the other hand, are such that the time-averaged mean square value of the aberration coefficients decreases as n increases and, for a given value of n, it is independent of the corresponding values of m.[4]

References

1. F. Zernike, "Diffraction theory of knife-edge test and its improved form, the phase contrast method," Mon. Not. R. Astron. Soc. **94,** 377–384 (1934); "Beugungstheorie des Schneidenverfahrens und Seiner Verbesserten Form, der Phasenkontrastmethode," Physica **1,** 689–704 (1934).
2. B. R. A. Nijboer, "The diffraction theory of aberrations," Ph.D. thesis (University of Groningen, The Netherlands, 1942). Also, Physica **23,** 605–620 (1947.
3. M. Born and E. Wolf, *Principles of Optics*, 5th ed. Pergamon, New York, N.Y. (1975), Chapter 9.
4. R. J. Noll, "Zernike polynomials and atmospheric turbulence," J. Opt. Soc. Am. **66,** 207–211 (1976).
5. Bill Swantner, "Comparison of Zernike polynomial sets in commercial software," Opt & Phot. News **9** (1992), pp. 42–43.
6. R. K. Tyson, "Using Zernike polynomials," Opt. & Phot. News **12,** 3 (1992).
7. J. R. Rogers, "Zernike polynomials," Opt. & Phot. News **8,** 2–3 (1993).
8. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," J. Opt. Soc. Am. **71,** 75–85, 1408 (1981), **A1,** 685 (1984).
9. V. N. Mahajan, "Uniform versus Gaussian beams: A comparison of the effects of diffraction, obscuration, and aberrations," J. Opt. Soc. Am. **A3,** 470–485 (1986).
10. C.-J. Kim and R. R. Shannon, "Catalog of Zernike polynomials," in *Applied Optics and Optical Engineering*, Vol. X, pp. 193–221, R. R. Shannon and J. Wyant eds. Academic Press, New York, N.Y. (1987).
11. S. Zhang and R. R. Shannon, "Catalog of spot diagrams," in *Applied Optics and Optical Engineering*, Vol. XI, pp. 201–238, R. R. Shannon and J. Wyant eds. Academic Press, New York, N.Y. (1992).
12. J. Wyant and K. Creath, "Basic wavefront aberration theory for optical metrology," in *Applied Optics and Optical Engineering*, Vol. XI, pp. 1–53, R. R. Shannon and J. Wyant Eds. Academic Press, New York, N.Y. (1992).

Specifying Dispersion in the Design of Diffractive Optics

By Donald C. O'Shea, School of Physics and Center for Optical Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0430

Abstract

The chromatic aberration of a diffractive optical surface can be specified within a lens design program using the Sweatt model[1] by choosing the fictitious high refractive index equal to its corresponding wavelength in an appropriate unit.

One of the most startling facts about diffractive optical lenses is their extremely high dispersion compared to ordinary glass. The power of a diffractive lens is proportional to the wavelength of the illuminating light. This means that in the visible the V-number for a diffractive lens is

$$V_d = \phi/\Delta\phi = \lambda_d/(\lambda_F - \lambda_C) = -3.45. \quad (1)$$

This strong negative dispersion can be used with conventional lenses to produce a hybrid achromat. Because of the strong dispersion, only a modest amount of optical power in the diffractive surface is required to produce the correction.

Sweatt[1] has shown that holographic optical elements (HOEs) can be modeled using conventional lens design programs by treating the HOE as a thin medium of high refractive index (n = 100 to 10,000). This technique can also be applied to diffractive optical elements. The index must be high enough to avoid significant errors in the ray trace, but not too high as to slow down program calculations significantly. Farn[2] has derived two criteria that must be satisfied to assure accuracy of the model.

Once these criteria are satisfied, the exact value of the index is still arbitrary. If one wishes to study the chromatic aberrations of a diffractive lens or a refractive-diffractive hybrid, three indices must be chosen. Because the power of the surface is proportional to wavelength, it is both correct and convenient to choose the indices to be equal to their corresponding wavelengths in some appropriate unit.

For example, in the visible region of the spectrum, choosing the ultra-high refractive indices to be equal to the wavelength in nanometers might not provide sufficient accuracy according to Farn.[2] In this case the nearly defunct Ångstrom unit may have a modern day role to play. In the infrared region across the 8 to 12 μm band, the refractive indices could be specified in nanometers. This easily remembered algorithm should be useful to designers who are analyzing chromatic effects of diffractive surfaces when using the Sweatt model with conventional lens design programs.

References

1. W. C. Sweatt, "Mathematical equivalence between a holographic optical element and an ultra-high index lens," J. Opt. Soc. Am. **69,** 486–487 (1979).
2. M. W. Farn, "Quantitative comparison of the general Sweatt model and the grating equation," Appl. Opt. **31,** 5312–5316 (1992).