Invited Paper

Zernike polynomials and aberration balancing

Virendra N. Mahajan*
The Aerospace Corporation
El Segundo, CA 90245

ABSTRACT

For small aberrations, the Strehl ratio of an imaging system depends on the aberration variance. If the aberration function is expanded in terms of a complete set of polynomials that are orthogonal over the system aperture, then the variance is given by the sum of the square of the aberration coefficients. One such set is that of Zernike polynomials, which are orthogonal over a circular pupil. Its advantage lies in the fact that Zernike polynomials can be identified with the classical aberrations that are balanced to yield minimum variance, and thus a maximum Strehl ratio. We discuss classical aberrations, balanced aberrations, and Zernike polynomials for systems with circular pupils. How these polynomials change for an annular or a Gaussian pupil are also discussed.

Keywords: Aberrations, balanced aberrations, Zernike circle polynomials, Zernike annular polynomials, circular pupils, annular pupils, Gaussian pupils, Zernike-Gauss polynomials.

1. INTRODUCTION

In Gaussian optics, the rays diverging from a point object and transmitted by an imaging system converge to its Gaussian image point. In a real system, they generally do not converge to a point and their displacements in the image plane from the Gaussian image point represent the ray aberrations. A spherical wave incident on the system exits from it only as an approximately spherical wave and its deviations from being spherical represent the wave aberrations.

The optical imaging systems generally have an axis of rotational symmetry and their pupil is circular, or annular as in the case of systems with mirrors. The aberration function of such systems can be expanded in a power series or a complete set of orthogonal polynomials. Although introduced by Zernike for testing the figure of a circular mirror by a knife edge test or his phase contrast method,[1] Zernike polynomials were studied extensively by Nijboer to describe the diffraction theory of aberrations.[2, 3] We show that these polynomials, which are orthogonal over a circular pupil, represent balanced aberrations that yield minimum variance. For small aberrations, a minimum of aberration variance yields a maximum of Strehl ratio. Similarly, Zernike annular polynomials or Zernike-Gauss polynomials represent balanced aberrations with minimum variance for systems with annular or Gaussian pupils, respectively. Balanced aberrations that yield minimum root mean square spot radius are different from those that yield minimum variance. They are also discussed briefly and identified with polynomials derived from Zernike polynomials.

2. ABERRATIONS OF A ROTATIONALLY SYMMETRIC SYSTEM

2.1 Power Series Expansion

The aberration function of a rotationally symmetric system depends on three rotational invariants h^2, r^2, and $hr\cos(\theta - \theta_o)$, where (h, θ_o) and (r, θ) are the polar coordinates of the object and pupil points, respectively. A power-series expansion of the aberration function for a point object along the x_o axis, so that $\theta_0 = 0$, may be written

$$\begin{aligned} W(h';r,\theta) &= \sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\sum_{m=0}^{\infty} C_{lpm}\left(h^2\right)^l\left(r^2\right)^p (hr\cos\theta)^m \\ &= \sum_{l=0}^{\infty}\sum_{p=0}^{\infty}\sum_{m=0}^{\infty} C_{lpm} h^{2l+m} r^{2p+m}\cos^m\theta \quad , \\ &= \sum_{l=0}^{\infty}\sum_{n=1}^{\infty}\sum_{m=0}^{n} {}_{2l+m}a_{nm}\, h'^{2l+m} r^n \cos^m\theta \quad , \end{aligned} \tag{1}$$

*virendra.n.mahajan@aero.org; phone 1 310 336-1783

Proceedings of SPIE Vol. 5173 Current Developments in Lens Design and Optical Engineering IV,
edited by Pantazis Z. Mouroulis, Warren J. Smith, R. Barry Johnson
(SPIE, Bellingham, WA, 2003) · 0277-786X/03/$15.00

where C_{lpm} and ${}_{2l+m}a_{nm}$ are the expansion coefficients; and l, p, and m are positive integers, including zero, and $n = 2p + m$. The aberration function is defined with respect to a Gaussian reference sphere of radius of curvature R that passes through the center of the exit pupil and whose center of curvature lies at the Gaussian image point at a height h' from the optical axis. The magnification of the image is $M = h'/h$. Note that $n - m \geq 0$ and even. The order i of an aberration term, which is equal to its degree in the object and pupil coordinates, is given by $i = 2l + m + n$. The order i is always even, and the number of terms N_i of this order is given by

$$N_i \; = \; (i+2)(i+4)/8 \quad . \tag{2}$$

This number includes a term with $n = 0 = m$, called piston aberration, although such a term does not constitute an aberration (since it corresponds to the chief ray, which by definition has a zero aberration associated with it). The aberration terms are referred to as the classical aberrations.

If the aberration terms having different dependencies on the object coordinates but the same dependence on pupil coordinates are combined so that there is only one term for each pair of (n, m) values, Eq. (1) for the power-series expansion of the aberration function may be written

$$W(\rho, \theta) \; = \; \sum_{n=1}^{\infty} \sum_{m=0}^{n} a_{nm} \rho^n \cos^m \theta \quad , \tag{3}$$

where $\rho = r/a$ is a radial variable normalized by the radius a of the exit pupil, and an expansion coefficient

$$a_{nm} \; = \; a^n \sum_{l=0}^{\infty} {}_{2l+m}a_{nm} h'^{2l+m} \tag{4}$$

represents the peak value of an aberration term. The number of terms through a certain order i in the reduced power-series expansion of the aberration function is also given by Eq. (2). The terms of Eq. (1) through a certain order i correspond to those terms of Eq. (4) for which $n + m \leq i$.

2.2 Primary or Seidel Aberration Function

The primary or Seidel aberration function corresponds to $i = 4$ and consists of five terms given by

$$W_P(h'; r, \theta) \; = \; {}_0a_{40} r^4 + {}_1a_{31} h' r^3 \cos\theta + {}_2a_{22} h'^2 r^2 \cos^2\theta + {}_2a_{20} h'^2 r^2 + {}_3a_{11} h'^3 r \cos\theta \quad , \tag{5}$$

where ${}_0a_{40}$, ${}_1a_{31}$, ${}_2a_{22}$, ${}_2a_{20}$, and ${}_3a_{11}$ represent the coefficients of spherical aberration, coma, astigmatism, field curvature, and distortion. The primary aberration function $(n + m \leq 4)$ in the reduced form is given by

$$W_P(\rho, \theta) \; = \; a_{40}\rho^4 + a_{31}\rho^3 \cos\theta + a_{22}\rho^2 \cos^2\theta + a_{20}\rho^2 + a_{11}\rho\cos\theta \quad . \tag{6}$$

The field curvature term has the same dependence on the pupil coordinates as a wavefront defocus aberration, but its coefficient a_{20} varies with h' as h'^2. Similarly, the distortion term has the same dependence on the pupil coordinates as a wavefront tilt aberration, but its coefficient a_{11} varies with h' as h'^3.

To simplify the notation further, we write the primary aberration function in the form

$$W(r, \theta; h') \; = \; a_s r^4 + a_c h' r^3 \cos\theta + a_a h'^2 r^2 \cos^2\theta + a_d h'^2 r^2 + a_t h'^3 r \cos\theta \quad , \tag{7}$$

or

$$W(\rho, \theta) \; = \; A_s \rho^4 + A_c \rho^3 \cos\theta + A_a \rho^2 \cos^2\theta + A_d \rho^2 + A_t \rho \cos\theta \quad . \tag{8}$$

The various aberration coefficients are related to each other according to

$$A_s \equiv a_{40} = a_s a^4, \; A_c \equiv a_{31} = a_c h' a^3, \; A_a \equiv a_{22} \equiv a_a h'^2 a^2, \; A_d \equiv a_{20} = a_d h'^2 a^2, \; A_t \equiv a_{11} = a_t h'^3 a \, . \tag{9}$$

2.3 Secondary or Schwarzschild Aberration Function

The secondary or Schwarzschild aberration function corresponding to $i = 6$ and consists of nine terms given by

$$W_S(h';r,\theta) = {}_0a_{60}r^6 + {}_1a_{51}h'r^5\cos\theta + {}_2a_{42}h'^2r^4\cos^2\theta + {}_3a_{33}h'^3r^3\cos^3\theta + {}_2a_{40}h'^2r^4$$
$$+ {}_3a_{31}h'^3r^3\cos\theta + {}_4a_{22}h'^4r^2\cos^2\theta + {}_4a_{20}h'^4r^2 + {}_5a_{11}h'^5r\cos\theta \quad , \tag{10}$$

where the first four terms with $l = 0$ are referred to as the secondary aberrations [spherical, coma, astigmatism (also called wings), and arrows], and the last five are called the lateral aberrations [spherical, coma, astigmatism, field curvature, and distortion]. The lateral aberrations are similar to the primary aberrations except for their dependence on the image height h'. The lateral spherical aberration ${}_2a_{40}h'^2r^4$ is also called the oblique spherical aberration. The aberration function through the sixth order $(n + m \le 6)$ according to Eq. (3) may be written

$$W(\rho,\theta) = a_{11}\rho\cos\theta + a_{20}\rho^2 + a_{22}\rho^2\cos^2\theta + a_{31}\rho^3\cos\theta + a_{33}\rho^3\cos^3\theta$$
$$+ a_{40}\rho^4 + a_{42}\rho^4\cos^2\theta + a_{51}\rho^5\cos\theta + a_{60}\rho^6 \quad . \tag{11}$$

Since the dependence of a term on the image height h' is contained in the coefficient a_{nm}, the coefficients in Eqs. (6) and (11) that seem to be the same, are actually different from each other. For example, a_{11} is equal to ${}_3a_{11}h'^3a$ in Eq. (6) but equal to $\left({}_3a_{11}h'^3 + {}_5a_{11}h'^5\right)a$ in Eq. (11).

3. BALANCED ABERRATIONS FOR MINIMUM ABERRATION VARIANCE

3.1 Strehl Ratio

A measure of image quality of a system is its Strehl ratio. It represents the ratio of the central irradiances of the image of a point object with and without aberrations. Approximate expressions for the Strehl ratio when the aberration is small are given by[2-5]

$$S \simeq \left(1 - \sigma_\Phi^2/2\right)^2 \simeq 1 - \sigma_\Phi^2 \simeq \exp\left(-\sigma_\Phi^2\right) \quad , \tag{12}$$

where

$$\sigma_\Phi^2 = <\Phi^2> - <\Phi>^2 \tag{13}$$

is the variance of the phase aberration across the exit pupil. The phase and the wave aberration are related to each other according to $\Phi = (2\pi/\lambda)W$, where λ is the wavelength of the object radiation. It is evident from these formulas that the Strehl ratio is maximum when the aberration variance is minimum. Numerical analysis on primary aberrations shows that $\exp\left(-\sigma_\Phi^2\right)$ yields the best approximation.[5]

3.2 Circular Pupils

The mean and the mean square values of the aberration for a system with a uniformly illuminated circular exit pupil are given by

$$<W^n> = \pi^{-1}\int_0^1\int_0^{2\pi} W^n(\rho,\theta)\,\rho\,d\rho\,d\theta \quad , \tag{14}$$

with $n = 1$ and 2, respectively. The standard deviation σ_W of a primary (or a Seidel) aberration is listed in Table 1 in terms of its peak aberration coefficients A_i.

The variance of spherical aberration can be reduced, thereby improving the Strehl ratio, if it is combined with a defocus aberration. Thus, if R is the radius of curvature of the Gaussian reference sphere with respect to which the peak spherical aberration of the wavefront is A_s, and we observe the image in a slightly defocused image plane at a distance z, the aberration of the defocused image may be written

$$W(\rho) = A_s\rho^4 + B_d\rho^2 \quad , \tag{15}$$

where

Table 1. Primary aberrations and their standard deviation for systems with uniformly illuminated circular pupils

Aberration	$W(\rho,\theta)$	σ_W
Spherical	$A_s\rho^4$	$2A_s/3\sqrt{5}$
Coma	$A_c\rho^3\cos\theta$	$A_c/2\sqrt{2}$
Astigmatism	$A_a\rho^2\cos^2\theta$	$A_a/4$
Field Curvature	$A_d\rho^2$	$A_d/2\sqrt{3}$
Distortion	$A_t\rho\cos\theta$	$A_t/2$

$$B_d = \frac{1}{2}\left(\frac{1}{z}-\frac{1}{R}\right)a^2 \simeq -\frac{\Delta R}{8F^2} \quad . \tag{16}$$

Here, $z \simeq R$, $\Delta R = z - R$, and $F = R/2a$ is the focal ratio of the image-forming light cone. By calculating σ_W^2 and letting $\partial\sigma_W^2/\partial B_d = 0$, we find that the aberration variance is minimum when $B_d = -A_s$, or $z - R = 8F^2 A_s$. The corresponding value of the standard deviation is $A_s/6\sqrt{5}$, which is a factor of 4 smaller than when $B_d = 0$. It should be evident that the Strehl ratio in this case represents the normalized axial irradiance. It can be shown that the axial irradiance is symmetric about the point $B_d = -A_s$.[6]

Similarly, the variance of coma can be reduced if it is combined with a wavefront tilt aberration. Thus, we consider an aberration given by

$$W(\rho,\theta) = A_c\rho^3\cos\theta + B_t\rho\cos\theta \quad , \tag{17}$$

and determine the peak value B_t of the tilt aberration that minimizes its variance. This value is given by $B_t = -2A_c/3$ and the standard deviation is reduced by a factor of 3. It implies that the aberration variance is minimum when measured with respect to the point $(4FA_c/3, 0)$, or the Strehl ratio is maximum at this point. When we balance astigmatism with defocus, we find that its standard deviation is reduced by a factor of 1.225 when a defocus aberration of $B_d = -A_a/2$ is introduced. Thus, the balanced astigmatism is given by

$$W(\rho,\theta) = A_a\rho^2\left(\cos^2\theta - 1/2\right) = (1/2)A_a\rho^2\cos 2\theta \quad . \tag{18}$$

Table 2 lists the balanced primary aberrations, their standard deviation, and the diffraction focus, i.e., the point with respect to which the aberration variance is minimum and Strehl ratio is maximum.

Table 2. Balanced primary aberrations, their standard deviation, and diffraction focus for systems with uniformly illuminated circular pupils

Balanced Aberration	$W(\rho,\theta)$	σ_W	**Diffraction Focus***
Spherical	$A_s\left(\rho^4-\rho^2\right)$	$A_s/6\sqrt{5}$	$\left(0,0,8F^2A_s\right)$
Coma	$A_c\left(\rho^3-2\rho/3\right)\cos\theta$	$A_c/6\sqrt{2}$	$\left(4FA_c/3,0,0\right)$
Astigmatism	$A_a\rho^2\left(\cos^2\theta-1/2\right)$	$A_a/2\sqrt{6}$	$\left(0,0,4F^2A_a\right)$

*The diffraction focus coordinates are relative to the Gaussian image point.

As stated earlier, minimum aberration variance yields maximum Strehl ratio for only small aberrations. Thus, although the standard deviation of spherical aberration is reduced by a factor of 4 when combined with an equal and opposite amount of defocus, regardless of the value of A_s, the Strehl ratio improves only if $A_s \lesssim 2.3\lambda$. For larger values of A_s, a different value of B_d yields maximum Strehl ratio.[5] Similarly, balanced coma yields maximum irradiance only if $A_c \lesssim 0.7\lambda$, which corresponds to $S \gtrsim 0.76$. For larger values of A_c, the distance of the point of maximum irradiance does not increase linearly with its value, and even fluctuates as A_c increases. Moreover, for $A_c > 2.3\lambda$, the Seidel coma gives a larger Strehl ratio than the balanced coma; i.e., the irradiance at the Gaussian image point is larger than at the point with respect to which the aberration variance is minimum.[7]

When secondary spherical aberration and secondary coma are balanced with lower-order aberrations to minimize their variance, it is found that a maximum of Strehl ratio is obtained only if its value comes out to be greater than about 0.5. Otherwise, a mixture of aberrations yielding a larger-than-minimum possible variance gives a higher Strehl ratio than the one provided by a minimum variance mixture.[8]

3.3 Annular Pupils

For an annular pupil with an obscuration ratio of ϵ, the radial variable $\epsilon \le \rho \le 1$. Accordingly, the mean values of W^n are given by

$$< W^n > \; = \left[\pi\left(1-\epsilon^2\right)\right]^{-1} \int_{\epsilon}^{1} \int_{0}^{2\pi} W^n(\rho, \theta; \epsilon)\, \rho\, d\rho\, d\theta \quad . \tag{19}$$

Table 3 lists the standard deviation of primary aberrations. Table 4 lists the corresponding balanced aberrations, their standard deviation, and the diffraction focus. We note that the balancing defocus in the case of spherical aberration or the balancing wavefront tilt in the case of coma depends on the value of ϵ. They both increase as ϵ increases. The amount of balancing defocus is larger by a factor of $1+\epsilon^2$ compared to that for a circular pupil. The axial irradiance in the case of spherical aberration is symmetric about the point with respect to which the aberration variance is minimum.[6] The amount of balancing defocus in the case of astigmatism, is independent of ϵ.

The standard deviation of spherical and balanced spherical aberrations and defocus decreases as ϵ increases. It increases in the case of coma, astigmatism, balanced astigmatism, and distortion. For balanced coma, it first slightly increases, achieves its maximum value at $\epsilon = 0.29$ and then decreases rapidly as ϵ increases. The factor by which the standard deviation of an aberration is reduced by balancing it with another aberration is reduced in the case of spherical aberration, but increases in the case of astigmatism, as ϵ increases. The tolerance for a certain aberration increases if its standard deviation decreases and vice versa. Thus, for example, the depth of focus of an unaberrated system increases as ϵ increases.[9]

Table 3. Primary aberrations and their standard deviation for systems with uniformly illuminated annular pupils

Aberration	$W(\rho, \theta)$	σ_W
Spherical	$A_s \rho^4$	$\left(4-\epsilon^2-6\epsilon^4-\epsilon^6+4\epsilon^8\right)^{1/2} A_s/3\sqrt{5}$
Coma	$A_c \rho^3 \cos\theta$	$\left(1+\epsilon^2+\epsilon^4+\epsilon^6\right)^{1/2} A_c/2\sqrt{2}$
Astigmatism	$A_a \rho^2 \cos^2\theta$	$\left(1+\epsilon^2\right)^{1/2} A_a/4$
Field curvature	$A_d \rho^2$	$\left(1-\epsilon^2\right) A_d/2\sqrt{3}$
Distortion	$A_t \rho \cos\theta$	$\left(1+\epsilon^2\right)^{1/2} A_t/2$

Table 4. Balanced primary aberrations, their standard deviation, and diffraction focus for systems with uniformly illuminated annular pupils

Balanced Aberration	$W(\rho, \theta)$	σ_W	Diffraction Focus
Spherical	$A_s\left[\rho^2 - \left(1+\epsilon^2\right)\rho^2\right]$	$\dfrac{1}{6\sqrt{5}}\left(1-\epsilon^2\right)^2 A_s$	$\left[0, 0, 8\left(1+\epsilon^2\right)F^2 A_s\right]$
Coma	$A_c\left(\rho^3 - \dfrac{2}{3}\dfrac{1+\epsilon^2+\epsilon^4}{1+\epsilon^2}\rho\right)\cos\theta$	$\dfrac{\left(1-\epsilon^2\right)\left(1+4\epsilon^2+\epsilon^4\right)^{1/2}}{6\sqrt{2}\left(1+\epsilon^2\right)^{1/2}} A_c$	$\left[\dfrac{4\left(1+\epsilon^2+\epsilon^4\right)}{3\left(1+\epsilon^2\right)} F A_c, 0, 0\right]$
Astigmatism	$A_a\rho^2\left(\cos^2\theta - 1/2\right)$	$\dfrac{1}{2\sqrt{6}}\left(1+\epsilon^2+\epsilon^4\right)^{1/2} A_a$	$\left(0, 0, 4F^2 A_a\right)$

3.4 Gaussian Pupils

The mean value of W^n for a Gaussian circular or annular pupil is given by

$$< W^n > \; = \int_0^1 \int_0^{2\pi} A(\rho)\left[W(\rho, \theta)\right]^n \rho\, d\rho\, d\theta \Bigg/ \int_0^1 \int_0^{2\pi} A(\rho)\, \rho\, d\rho\, d\theta \tag{20}$$

and

$$< W^n > \; = \left\{\gamma/\pi\left[\exp\left(-\gamma\epsilon^2\right) \; - \; \exp(-\gamma)\right]\right\} \int_\epsilon^1 A(\rho)\left[W(\rho; \theta)\right]^n \rho\, d\rho\, d\theta \quad , \tag{21}$$

respectively, where

$$A(\rho) \; = \; A_0 \exp\left(-\gamma\rho^2\right) \tag{22}$$

represents the variation of the amplitude across the pupil with A_0 being a constant. Table 5 gives the aberrations and their standard deviations for $\gamma = 1$, i.e., when the irradiance at the edge of the pupil is $1/e^2$ of its value at the center.[10] Comparing these results with those given in Tables 1 and 2 for uniform circular pupils, we note that the balancing defocus in the case of spherical aberration and the balancing tilt in the case of coma are slightly different for Gaussian pupils. However, the balancing defocus in the case of astigmatism is the same. For a Gaussian pupil, the balancing of an aberration reduces its standard deviation by a factor of 3.74, 2.64, and 1.27, for spherical aberration, coma, and astigmatism, respectively, compared to factors of 4, 3, and 1.22 for a uniform pupil.

Table 5. Primary aberrations and their standard deviations for optical systems with Gaussian circular pupils and $\gamma = 1$

Aberration	$W(\rho, \theta)$	σ_W
Spherical	$A_s\rho^4$	$A_s/3.67$
Balanced spherical	$A_s\left(\rho^4 - 0.933\rho^2\right)$	$A_s/13.71$
Coma	$A_c\rho^3\cos\theta$	$A_c/3.33$
Balanced coma	$A_c\left(\rho^3 - 0.608\rho\right)\cos\theta$	$A_c/8.80$
Astigmatism	$A_a\rho^2\cos^2\theta$	$A_a/4.40$
Balanced astigmatism	$A_a\rho^2\left(\cos^2\theta - 1/2\right)$	$A_a/5.61$
Defocus	$B_d\rho^2$	$B_d/3.55$
Tilt	$B_t\rho\cos\theta$	$B_t/2.19$

4. ZERNIKE CIRCLE POLYNOMIALS AS BALANCED ABERRATIONS

Nijboer used Zernike circle polynomials to describe the diffraction theory of aberrations.[2, 3] The aberration function of a system with a circular exit pupil for a point object at a certain angle from its axis can be expanded in terms of a complete set of Zernike circle polynomials $R_n^m(\rho)\cos m\theta$ that are orthogonal over a unit circle in the form[11]

$$W(\rho,\theta) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} c_{nm} Z_n^m(\rho,\theta) \quad , \tag{23}$$

where c_{nm} are the expansion coefficients that depend on the object angle, n and m are positive integers including zero, $n - m \geq 0$ and even, and

$$Z_n^m(\rho,\theta) = \left[2(n+1)/(1+\delta_{m0})\right]^{1/2} R_n^m(\rho)\cos m\theta \tag{24a}$$

is an orthonormal Zernike polynomial. Here, δ_{ij} is a Kronecker delta, and

$$R_n^m(\rho) = \sum_{s=0}^{(n-m)/2} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!}\rho^{n-2s} \tag{24b}$$

is a polynomial of degree n in ρ containing terms in ρ^n, ρ^{n-2}, …, and ρ^m. The radial circle polynomials $R_n^m(\rho)$ are even or odd in ρ, depending on whether n (or m) is even or odd. Also, $R_n^n(1) = 1$, $R_n^n(\rho) = \rho^n$, and $R_n^m(0) = \delta_{m0}$ for even $n/2$ and $-\delta_{m0}$ for odd $n/2$. The Zernike polynomials are orthogonal according to

$$\frac{1}{\pi}\int_0^1\int_0^{2\pi} Z_n^m(\rho,\theta) Z_{n'}^{m'}(\rho,\theta)\rho\, d\rho\, d\theta = \delta_{mm'}\delta_{nn'} \quad , \tag{25a}$$

$$\int_0^{2\pi} \cos m\theta \cos m'\theta = \pi(1+\delta_{m0})\delta_{mm'} \quad , \tag{25b}$$

and

$$\int_0^1 R_n^m(\rho) R_{n'}^m(\rho)\rho\, d\rho = \frac{1}{2(n+1)}\delta_{nn'} \quad . \tag{25c}$$

The Zernike expansion coefficients are given by

$$c_{nm} = (1/\pi)\left[2(n+1)(1+\delta_{m0})\right]^{1/2}\int_0^1\int_0^{2\pi} W(\rho,\theta) R_n^m(\rho)\cos m\theta\, \rho\, d\rho\, d\theta \quad , \tag{26}$$

as may be seen by substituting Eq. (23) and utilizing the orthogonality relations.

The Zernike circle polynomials are unique in that they are the only polynomials in two variables ρ and θ, which (a) are orthogonal over a unit circle, (b) are invariant in form with respect to rotation of the coordinate axes about the origin, and (c) include a polynomial for each permissible pair of n and m values.[3, 12]

The orthonormal Zernike polynomials and the names associated with some of them when identified with aberrations are listed in Table 6 for $n \leq 8$. The number of Zernike (or orthogonal) aberration terms in the expansion of an aberration function through a certain order n is given by

$$N_n = \left(\frac{n}{2}+1\right)^2 \quad \text{for even } n \quad , \tag{27a}$$

$$= (n+1)(n+3)/4 \quad \text{for odd } n \quad . \tag{27b}$$

Table 6. Orthonormal Zernike *circle* polynomials and balanced aberrations

n	m	**Orthonormal Zernike Polynomial** $Z_n^m(\rho, \theta) = \left[\dfrac{2(n+1)}{1+\delta_{m0}}\right]^{1/2} R_n^m(\rho) \cos m\theta$	**Aberration Name***
0	0	1	Piston
1	1	$\rho \cos\theta$	Distortion (tilt)
2	0	$\sqrt{3}\left(2\rho^2 - 1\right)$	Field curvature (defocus)
2	2	$\sqrt{6}\,\rho^2 \cos 2\theta$	Primary astigmatism
3	1	$\sqrt{8}\left(3\rho^3 - 2\rho\right)\cos\theta$	Primary coma
3	3	$\sqrt{8}\,\rho^3 \cos 3\theta$	
4	0	$\sqrt{5}\left(6\rho^4 - 6\rho^2 + 1\right)$	Primary spherical
4	2	$\sqrt{10}\left(4\rho^4 - 3\rho^2\right)\cos 2\theta$	Secondary astigmatism
4	4	$\sqrt{10}\,\rho^4 \cos 4\theta$	
5	1	$\sqrt{12}\left(10\rho^5 - 12\rho^3 + 3\rho\right)\cos\theta$	Secondary coma
5	3	$\sqrt{12}\left(5\rho^5 - 4\rho^3\right)\cos 3\theta$	
5	5	$\sqrt{12}\,\rho^5 \cos 5\theta$	
6	0	$\sqrt{7}\left(20\rho^6 - 30\rho^4 + 12\rho^2 - 1\right)$	Secondary spherical
6	2	$\sqrt{14}\left(15\rho^6 - 20\rho^4 + 6\rho^2\right)\cos 2\theta$	Tertiary astigmatism
6	4	$\sqrt{14}\left(6\rho^6 - 5\rho^4\right)\cos 4\theta$	
6	6	$\sqrt{14}\,\rho^6 \cos 6\theta$	
7	1	$4\left(35\rho^7 - 60\rho^5 + 30\rho^3 - 4\rho\right)\cos\theta$	Tertiary coma
7	3	$4\left(21\rho^7 - 30\rho^5 + 10\rho^3\right)\cos 3\theta$	
7	5	$4\left(7\rho^7 - 6\rho^5\right)\cos 5\theta$	
7	7	$4\rho^7 \cos 7\theta$	
8	0	$3\left(70\rho^8 - 140\rho^6 + 90\rho^4 - 20\rho^2 + 1\right)$	Tertiary spherical

*The words “orthonormal Zernike” are to be associated with these names, e.g., *orthonormal Zernike primary astigmatism*.

Consider a typical Zernike aberration term in Eq. (23):

$$W_n^m(\rho,\theta) = c_{nm} Z_n^m(\rho,\theta) \quad . \tag{28}$$

Unless $n = m = 0$, its mean value is zero; i.e.,

$$\left\langle W_n^m(\rho,\theta) \right\rangle = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} W_n^m(\rho,\theta) \rho \, d\rho d\theta = 0 \quad , \quad n \neq 0, m \neq 0 \quad . \tag{29}$$

For $m = 0$, this may be seen with the help of Eq. (25b) and the fact that $R_0^0(\rho) = 1$ is a member of the polynomial set. The orthogonality Eq. (25c) yields the result that the mean value of $R_n^0(\rho)$ is zero. The presence of the constant term in $R_n^0(\rho)$ yields a zero mean value. When $m \neq 0$, the average value of $\cos m\theta$ is zero. Similarly, the mean square value of the aberration is given by

$$\left\langle \left[W_n^m(\rho,\theta) \right]^2 \right\rangle = \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \left[W_n^m(\rho,\theta) \right]^2 \rho \, d\rho d\theta = c_{nm}^2 \quad . \tag{30}$$

Hence, its variance is given by

$$\sigma_{nm}^2 = \left\langle \left(W_n^m \right)^2 \right\rangle - \left\langle W_n^m \right\rangle^2 = c_{nm}^2 \quad , \quad n \neq 0, m \neq 0 \quad . \tag{31}$$

Thus, each expansion coefficient, with the exception of c_{00}, represents the standard deviation of the corresponding aberration term. The variance of the aberration function is accordingly given by

$$\sigma_W^2 = \left\langle W^2(\rho,\theta) \right\rangle - \left\langle W(\rho,\theta) \right\rangle^2 = \sum_{n=1}^{\infty} \sum_{m=0}^{n} c_{nm}^2 \quad . \tag{32}$$

Unless the mean value of the aberration $< W > = 0$, $\sigma_W \neq W_{rms}$, where $W_{rms} = < W^2 >^{1/2}$ is the root-mean-square (rms) value of the aberration.

A balanced aberration represents an aberration of a certain order in the power series expansion of the aberration function in pupil coordinates mixed with aberrations of lower order such that the variance of the net aberration is minimized. The balanced primary aberrations can be identified easily with the corresponding Zernike polynomials. For example, for $n = 4$ and $m = 0$, Eq. (28) becomes

$$W_4^0(\rho) = \sqrt{5} c_{40} R_4^0(\rho) = \sqrt{5} c_{40} \left(6\rho^4 - 6\rho^2 + 1 \right) \quad . \tag{33}$$

Comparing this with the balanced spherical aberration given in Table 2,

$$W_{bs}(\rho) = A_s \left(\rho^4 - \rho^2 \right) \quad , \tag{34}$$

we note the following. The aberration $W_4^0(\rho)$ contains a constant (independent of ρ and θ) term. This term does not change the standard deviation of the balanced aberration or the Strehl ratio corresponding to it. In Eq. (33), as in Eq. (34), the spherical aberration is balanced with an equal and opposite amount of defocus. Comparing the coefficients of ρ^4 term in these equations, we find immediately that the standard deviation of the balanced spherical aberration is given by

$$\sigma_{bs} = c_{40} = A_s / 6\sqrt{5} \quad , \tag{35}$$

in agreement with the result given in Table 2. Similarly, we can identify balanced coma with $Z_3^1(\rho,\theta)$ and balanced astigmatism with $Z_2^2(\rho,\theta)$, respectively. Moreover, the field curvature (or defocus) and distortion (or wavefront tilt) can be identified with $Z_2^0(\rho)$ and $Z_1^1(\rho,\theta)$, respectively. These results are summarized in Table 7.

Table 7. Comparison of balanced primary aberrations with the corresponding Zernike polynomials

Aberration	$W(\rho,\theta)$	Orthonormal Zernike Polynomial	σ_W
Balanced Spherical	$A_s\left(\rho^4-\rho^2\right)$	$\sqrt{5}c_{40}\,R_4^0(\rho)$ $=\sqrt{5}c_{40}\left(6\rho^4-6\rho^2+1\right)$	$c_{40}=\dfrac{A_s}{6\sqrt{5}}$
Balanced Coma	$A_c\left(\rho^3-2\rho/3\right)\cos\theta$	$2\sqrt{2}c_{31}R_3^1(\rho)\cos\theta$ $=2\sqrt{2}c_{31}\left(3\rho^3-2\rho\right)\cos\theta$	$c_{31}=\dfrac{A_c}{6\sqrt{2}}$
Balanced Astigmatism	$A_a\rho^2\left(\cos^2\theta-1/2\right)$ $=\left(A_a/2\right)\rho^2\cos 2\theta$	$\sqrt{6}c_{22}R_2^2(\rho)\cos 2\theta$ $=2\sqrt{6}c_{22}\rho^2\left(\cos^2\theta-1/2\right)$	$c_{22}=\dfrac{A_a}{2\sqrt{6}}$
Field Curvature	$A_d\rho^2$	$\sqrt{3}c_{20}\,R_2^0(\rho)$ $=\sqrt{3}c_{20}\left(2\rho^2-1\right)$	$c_{20}=\dfrac{A_d}{2\sqrt{3}}$
Distortion	$A_t\rho\cos\theta$	$2c_{11}R_1^1(\rho)\cos\theta$ $=2c_{11}\,\rho\cos\theta$	$c_{11}=\dfrac{A_t}{2}$

Thus, we see that Zernike polynomials can be identified with balanced aberrations; that, in fact, is their advantage over another complete set that is orthogonal over a circular pupil. Although we have discussed only the primary aberrations, the aberration function of an optical system will generally consist of higher-order aberrations. The higher-order aberrations can be balanced in a similar manner. For example, when secondary spherical aberration varying as ρ^6 is combined with primary spherical aberration and defocus to minimize its variance, the balanced aberration obtained is the same as the Zernike polynomial $Z_6^0(\rho)$ (except for a constant term). Similarly, when secondary coma varying as $\rho^5\cos\theta$ is combined with primary coma and tilt to minimize its variance, the balanced aberration obtained is the same as the Zernike polynomial $Z_5^1(\rho,\theta)$.[13]

5. ZERNIKE ANNULAR POLYNOMIALS AS BALANCED ABERRATIONS

The aberration function of a system with an annular pupil for a point object at a certain angle from its axis can be expanded in terms of a complete set of Zernike annular polynomials $R_n^m(\rho;\epsilon)\cos\theta$ that are orthogonal over the annulus in the form[9]

$$W(\rho,\theta;\epsilon) = \sum_{n=0}^{\infty}\sum_{m=0}^{n} c_{nm}\left[2(n+1)/\left(1+\delta_{m0}\right)\right]^{1/2} R_n^m(\rho;\epsilon)\cos m\theta \quad , \tag{36}$$

where c_{nm} are the expansion coefficients, n and m are positive integers, $n-m \geq 0$ and even, the expansion coefficients c_{nm} depend on the object angle and the radial annular polynomials $R_n^m(\rho;\epsilon)$ obey the orthogonality relation

$$\int_{\epsilon}^{1} R_n^m(\rho;\epsilon)\,R_{n'}^m(\rho;\epsilon)\,\rho\,d\rho = \frac{1-\epsilon^2}{2(n+1)}\delta_{nn'} \quad . \tag{37}$$

The annular polynomials are similar to the circle polynomials, except that they are orthogonal over an annular pupil. Thus, $R_n^m(\rho;\epsilon)$ is a radial polynomial of degree n in ρ containing terms in ρ^n, $\rho^{n-2},\ldots,$ and ρ^m with coefficients that depend on ϵ. The Zernike annular polynomials can be obtained from the corresponding circle polynomials by the Gram-Schmidt orthogonalization process.[9, 14] The annular polynomials are unique in the same manner as the circle polynomials. The radial Zernike annular polynomials for $n \leq 6$ are listed in Table 8.

Once again, the balanced primary aberrations can be identified with the corresponding Zernike annular polynomials. For example, consider the aberration term with $n=4$ and $m=0$:

Table 8. Radial Zernike *annular* polynomials

n	m	$R_n^m(\rho, \epsilon)$
0	0	1
1	1	$\rho \big/ \left(1+\epsilon^2\right)^{1/2}$
2	0	$\left(2\rho^2 - 1 - \epsilon^2\right) \big/ \left(1-\epsilon^2\right)$
2	2	$\rho^2 \big/ \left(1+\epsilon^2+\epsilon^4\right)^{1/2}$
3	1	$\dfrac{3\left(1+\epsilon^2\right)\rho^3 - 2\left(1+\epsilon^2+\epsilon^4\right)\rho}{\left(1-\epsilon^2\right)\left[\left(1+\epsilon^2\right)\left(1+4\epsilon^2+\epsilon^4\right)\right]^{1/2}}$
3	3	$\rho^3 \big/ \left(1+\epsilon^2+\epsilon^4+\epsilon^6\right)^{1/2}$
4	0	$\left[6\rho^4 - 6\left(1+\epsilon^2\right)\rho^2 + 1 + 4\epsilon^2 + \epsilon^4\right] \big/ \left(1-\epsilon^2\right)^2$
4	2	$\dfrac{4\rho^4 - 3\left[\left(1-\epsilon^8\right)/\left(1-\epsilon^6\right)\right]\rho^2}{\left\{\left(1-\epsilon^2\right)^{-1}\left[16\left(1-\epsilon^{10}\right) - 15\left(1-\epsilon^8\right)^2/\left(1-\epsilon^6\right)\right]^{1/2}\right\}}$
4	4	$\rho^4 \big/ \left(1+\epsilon^2+\epsilon^4+\epsilon^6+\epsilon^8\right)^{1/2}$
5	1	$\dfrac{10\left(1+4\epsilon^2+\epsilon^4\right)\rho^5 - 12\left(1+4\epsilon^2+4\epsilon^4+\epsilon^6\right)\rho^3 + 3\left(1+4\epsilon^2+10\epsilon^4+4\epsilon^6+\epsilon^8\right)\rho}{\left(1-\epsilon^2\right)^2\left[\left(1+4\epsilon^2+\epsilon^4\right)\left(1+9\epsilon^2+9\epsilon^4+\epsilon^6\right)\right]^{1/2}}$
5	3	$\dfrac{5\rho^5 - 4\left[\left(1-\epsilon^{10}\right)/\left(1-\epsilon^8\right)\right]\rho^3}{\left\{\left(1-\epsilon^2\right)^{-1}\left[25\left(1-\epsilon^{12}\right) - 24\left(1-\epsilon^{10}\right)^2/\left(1-\epsilon^8\right)\right]\right\}^{1/2}}$
5	5	$\rho^5 \big/ \left(1+\epsilon^2+\epsilon^4+\epsilon^6+\epsilon^8+\epsilon^{10}\right)^{1/2}$
6	0	$\left[20\rho^6 - 30\left(1+\epsilon^2\right)\rho^4 + 12\left(1+3\epsilon^2+\epsilon^4\right)\rho^2 - \left(1+9\epsilon^2+9\epsilon^4+\epsilon^6\right)\right] \big/ \left(1-\epsilon^2\right)^3$
6	2	$\dfrac{15\left(1+4\epsilon^2+10\epsilon^4+4\epsilon^6+\epsilon^8\right)\rho^6 - 20\left(1+4\epsilon^2+10\epsilon^4+10\epsilon^6+4\epsilon^8+\epsilon^{10}\right)\rho^4 + 6\left(1+4\epsilon^2+10\epsilon^4+20\epsilon^6+10\epsilon^8+4\epsilon^{10}+\epsilon^{12}\right)\rho^2}{\left(1+\epsilon^2\right)^2\left[\left(1+4\epsilon^2+10\epsilon^4+4\epsilon^6+\epsilon^8\right)\left(1+9\epsilon^2+45\epsilon^4+65\epsilon^6+45\epsilon^8+9\epsilon^{10}+\epsilon^{12}\right)\right]^{1/2}}$
6	4	$\dfrac{6\rho^6 - 5\left[\left(1-\epsilon^{12}\right)/\left(1-\epsilon^{10}\right)\right]\rho^4}{\left\{\left(1-\epsilon^2\right)^{-1}\left[36\left(1-\epsilon^{14}\right) - 35\left(1-\epsilon^{12}\right)^2/\left(1-\epsilon^{10}\right)\right]\right\}^{1/2}}$
6	6	$\rho^6 \big/ \left(1+\epsilon^2+\epsilon^4+\epsilon^6+\epsilon^8+\epsilon^{10}+\epsilon^{12}\right)^{1/2}$

$$W_4^0(\rho;\epsilon) \;=\; \sqrt{5}\,c_{40}\,R_4^0(\rho;\epsilon) \;=\; \sqrt{5}\,c_{40}\left[6\rho^4 - 6\left(1+\epsilon^2\right)\rho^2 + \left(1+4\epsilon^2+\epsilon^4\right)\right]\Big/\left(1-\epsilon^2\right)^2 \quad . \tag{38}$$

Comparing this with the balanced spherical aberration given in Table 4, namely,

$$W_{bs}(\rho) \;=\; A_s\left[\rho^4 - \left(1+\epsilon^2\right)\rho^2\right] \quad , \tag{39}$$

we note the following. The aberration $W_4^0(\rho;\epsilon)$ contains a constant (independent of ρ and θ) term. This term does not change the standard deviation of the balanced aberration or the Strehl ratio corresponding to it. In Eq. (38), as in Eq.(39), the spherical aberration is balanced with an amount of defocus that is $-\left(1+\epsilon^2\right)$ times the amount of the aberration. Comparing the coefficients of the ρ^4 term, we find immediately that the standard deviation of the balanced spherical aberration is given by

$$\sigma_{bs} \;=\; c_{40} \;=\; A_s\left(1-\epsilon^2\right)^2\Big/6\sqrt{5} \quad , \tag{40}$$

in agreement with the result given in Table 4. Similarly, we can compare balanced coma and balanced astigmatism with $Z_3^1(\rho,\theta;\epsilon)$ and $Z_2^2(\rho,\theta;\epsilon)$, respectively. Moreover, the field curvature (and defocus) and distortion (and wavefront tilt) can be identified with $Z_2^0(\rho;\epsilon)$ and $Z_1^1(\rho;\theta,\epsilon)$, respectively.

6. ZERNIKE-GAUSS POLYNOMIALS AS BALANCED ABERRATIONS

The polynomials that are orthonormal over an exit pupil with a Gaussian illumination can be obtained from the polynomials for a uniform illumination by the Gram-Schmidt orthogonalization process.[10, 15] The aberration function of a system with a circular exit pupil can be expanded in terms of a complete set of Zernike-Gauss circle polynomials $R_n^m(\rho;\gamma)\cos m\theta$ that are orthogonal over a unit circle in the form

$$W(\rho,\theta;\gamma) \;=\; \sum_{n=0}^{\infty}\sum_{m=0}^{n} c_{nm}\left[2(n+1)\big/\left(1+\delta_{m0}\right)\right]^{1/2} R_n^m(\rho;\gamma)\cos m\theta \quad , \tag{41}$$

where the radial polynomials $R_n^m(\rho;\gamma)$ obey the orthogonality relation

$$\int_0^1 R_n^m(\rho;\gamma)\,R_{n'}^m(\rho;\gamma)\,A(\rho)\,\rho\,d\rho\bigg/\int_0^1 A(\rho)\,\rho\,d\rho \;=\; \frac{1}{n+1}\delta_{nn'} \quad . \tag{42}$$

Except for a normalization constant, the radial polynomial $R_n^n(\rho;\gamma)$ is identical to the corresponding polynomial $R_n^n(\rho)$ for a uniformly illuminated circular pupil. The radial polynomial $R_n^m(\rho;\gamma)$ is a polynomial of degree n in ρ containing terms in ρ^n, ρ^{n-2}, ..., and ρ^m, whose coefficients depend on the Gaussian amplitude through γ; i.e., it has the form

$$R_n^m(\rho;\gamma) \;=\; a_n^m\rho^n + b_n^m\rho^{n-2} + \ldots + d_n^m\rho^m \quad , \tag{43}$$

where the coefficients a_n^m, etc., depend on γ. The number of Zernike (or orthogonal) aberration terms in the expansion of an aberration function through a certain order n is given by Eq. (27), as in the case of circle polynomials. The Zernike-Gauss expansion coefficients are given by

$$c_{nm} = (1/\pi)\left[2(n+1)\left(1+\delta_{m0}\right)\right]^{1/2}\int_0^1\int_0^{2\pi} W(\rho,\theta;\gamma)R_n^m(\rho;\gamma)\,A(\rho)\cos m\theta\,\rho\,d\rho\,d\theta\,, \tag{44}$$

as may be seen by substituting Eq. (41) and utilizing the orthogonality of the polynomials. The variance of the aberration function is accordingly given by

$$\sigma_w^2 \;=\; \left\langle W^2(\rho,\theta)\right\rangle - \left\langle W(\rho,\theta)\right\rangle^2 \;=\; \sum_{n=1}^{\infty}\sum_{m=0}^{n} c_{nm}^2 \quad . \tag{45}$$

A classical aberration of a certain order when combined with aberrations of lower order to minimize its variance across the Gaussian-weighted circular exit pupil may be represented by a corresponding Zernike-Gauss circle polynomial.

The radial polynomials corresponding to balanced primary aberrations are listed in Table 9. Thus, for $\gamma = 1$, the balancing defocus for spherical aberration $A_s\rho^4$ is given by $B_d\rho^2$, where

$$B_d \; = \; \left(b_4^0 / a_4^0\right) A_s \; = \; -\,0.933 A_s \quad . \tag{46a}$$

Accordingly, the balanced spherical aberration becomes

$$W_{bs}(\rho; 1) \; = \; A_s\left(\rho^4 \, - \, 0.933\rho^2\right) \quad , \tag{46b}$$

in agreement with the result given in Table 5. Similarly, balanced coma and balanced astigmatism listed in Table 5 can be identified with the corresponding Zernike-Gauss polynomials given in Table 9. The orthogonal polynomials for weekly truncated Gaussian beam, i.e., for large values of γ are also given in Table 9.

For systems with annular Gaussian pupils, the orthonormal set of Zernike-Gauss annular polynomial can be obtained from Zernike annular polynomials by the Gram-Schmidt orthogonalization process[10, 15] in exactly the same manner as the Zernike-Gauss circle polynomials are obtained from the Zernike circle polynomials, except that now the lower limit on the radial variable ρ of integration is ϵ instead of zero. The radial polynomials in this case may be written $R_n^m(\rho;\gamma;\epsilon)$. The aberration function for such a system can be written

$$W(\rho, \theta; \gamma; \epsilon) \;\; = \; \sum_{n=0}^{\infty} \; \sum_{m=0}^{n} c_{nm}\left[2(n+1)\big/\left(1+\delta_{m0}\right)\right]^{1/2} R_n^m(\rho; \gamma; \epsilon) \cos m\theta \quad , \tag{47}$$

where the radial polynomials obey the orthogonality relation

$$\int_{\epsilon}^{1} R_n^m(\rho;\gamma;\epsilon)\, R_{n'}^m(\rho;\gamma;\epsilon)\, A(\rho)\, \rho\, d\rho \bigg/ \int_{\epsilon}^{1} A(\rho)\, \rho\, d\rho \; = \; \frac{1}{n+1}\delta_{nn'} \quad . \tag{48}$$

The radial polynomial $R_n^n(\rho;\gamma;\epsilon)$ is identical to the corresponding polynomial for a uniformly illuminated annular pupil $R_n^n(\rho)$, except for a normalization constant. As in the case of the Zernike polynomials for uniformly illuminated circular and annular pupils, n and m are positive integers, including zero, and $n - m \geq 0$ and even. The radial polynomial $R_n^m(\rho;\gamma)$ is a polynomial of degree n in ρ containing terms in ρ^n, ρ^{n-2}, ..., and ρ^m whose coefficients depend on the Gaussian amplitude through γ, i.e., it has the form

$$R_n^m(\rho;\gamma;\epsilon) \; = \; a_n^m\rho^n \; + \; b_n^m\rho^{n-2} \; + \ldots + \; d_n^m\rho^m \quad , \tag{49}$$

where the coefficients a_n^m, etc., depend on both γ and ϵ. (50)

The coefficients for $\gamma = 1$ and $\epsilon = 0$, 0.25, 0.50, 0.75, and 0.90 are given in Table 10. For comparison, the coefficients for a uniformly illuminated pupil, i.e., for $\gamma = 0$, are given in parentheses in this table. An increase (decrease)

Table 9. Radial polynomials representing balanced primary aberrations for Gaussian beams. Polynomials for special cases of $\gamma = 1$ and $\gamma = 0$ (corresponding to a uniform beam) and weakly truncated Gaussian beams are also given.

Aberration	Radial Polynomial	Gaussian	Gaussian $\gamma = 1$	Uniform $\gamma = 0$	Weakly Truncated Gaussian
Piston	R_0^0	1	1	1	1
Distortion (tilt)	R_1^1	$a_1^1\rho$	1.09367ρ	ρ	$\sqrt{\gamma/2}\rho$
Field curvature (defocus)	R_2^0	$a_2^0\rho^2 + b_2^0$	$2.04989\rho^2 - 0.85690$	$2\rho^2 - 1$	$(\gamma\rho^2 - 1)/\sqrt{3}$
Astigmatism	R_2^2	$a_2^2\rho^2$	$1.14541\rho^2$	ρ^2	$(\gamma/\sqrt{6})\rho^2$
Coma	R_3^1	$a_3^1\rho^3 + b_3^1\rho$	$3.111213\rho^3 - 1\cdot 89152\rho$	$3\rho^3 - 2\rho$	$\sqrt{\gamma/2}\left(\frac{\gamma}{2}\rho^3 - \rho\right)$
Spherical aberration	R_4^0	$a_4^0\rho^4 + b_4^0\rho^2 + c_4^0$	$6.12902\rho^4 - 5.71948\rho^2 + 0.83368$	$6\rho^4 - 6\rho^2 + 1$	$(\gamma^2\rho^4 - 4\gamma\rho^2 + 2)/2\sqrt{5}$

Table 10. Coefficients of terms in Zernike-Gauss radial polynomials $R_n^m(\rho;\gamma;\epsilon)$. for $\gamma = 1$. The numbers given in parentheses are the corresponding coefficients for uniform illumination.

ϵ	a_1^1	a_2^0	b_2^0	a_2^2	a_3^1	b_3^1	a_4^0	b_4^0	c_4^0
0.00	1.09367	2.04989	– 0.85690	1.14541	3.11213	– 1.89152	6.12902	– 5.71948	0.83368
	(1.00000)	(2.00000)	(– 1.00000)	(1.00000)	(3.00000)	(– 2.00000)	(6.00000)	(– 6.00000)	(1.00000)
0.25	1.04364	2.18012	– 1.00080	1.08940	3.01573	– 1.84513	6.95563	– 6.98197	1.25153
	(0.97014)	(2.13333)	(– 1.13333)	(0.96836)	(2.94566)	(– 1.97099)	(6.82667)	(– 7.25333)	(1.42667)
0.50	0.92963	2.70412	– 1.56449	0.93620	3.14319	– 2.06618	10.79549	– 13.08900	3.46706
	(0.89443)	(2.66667)	(– 1.66667)	(0.87287)	(3.11400)	(– 2.17980)	(10.66667)	(– 13.33333)	(3.66667)
0.75	0.80827	4.59329	– 3.51548	0.74439	4.55179	– 3.57767	31.47560	– 48.77879	18.39840
	(0.80000)	(4.57143)	(– 3.57143)	(0.72954)	(4.53877)	(– 3.63858)	(31.34694)	(– 48.97959)	(18.63265)
0.90	0.74453	10.53581	– 9.50324	0.63890	9.60573	– 8.69629	166.33359	– 300.66342	135.36926
	(0.74329)	(10.52632)	(– 9.52632)	(0.63679)	(9.60023)	(– 8.72012)	(166.20500)	(– 300.83102)	(135.62604)

in the value of a coefficient a_n^m of an orthogonal aberration $R_n^m(\rho;\gamma;\epsilon)\cos m\theta$ implies a decrease (increase) in the value of σ_W for a given amount of the corresponding classical aberration. This, in turn, implies that for small aberrations, the system performance as measured by Strehl ratio is less (more) sensitive to that classical aberration when balanced with other classical aberrations to form an orthogonal aberration. Thus, as ϵ increases, irrespective of the value of γ, the system becomes less sensitive to field curvature (defocus) and spherical aberration but more sensitive to distortion (tilt) and astigmatism. In the case of coma, it first becomes slightly more sensitive but is much less sensitive for larger values of ϵ. As γ increases, i.e., as the width of the Gaussian illumination becomes narrower, the system becomes less sensitive to all classical primary aberrations. Although the results for $\gamma = 0$ and $\gamma = 1$ only are given in Table 10, the coefficients for $0 \le \gamma \le 3$ show that the differences between the coefficients for uniform and Gaussian illumination are small, and they decrease as ϵ increases and increase as γ increases. This is understandable because as ϵ increases or γ decreases, the differences between the two illuminations decreases.

7. BALANCED ABERRATIONS FOR MINIMUM RMS SPOT RADIUS

The wave aberration $W(x, y)$ and the ray aberration (x_i, y_i) with respect to the Gaussian image point are related to each other according to

$$(x_i, y_i) = R\left(\frac{\partial W}{\partial x}, \frac{\partial W}{\partial y}\right) \quad . \tag{51}$$

For an image lying at infinity, its ray aberrations are described in angular coordinates by dividing both sides by R. The centroid (or the center of gravity) of the ray spot diagram is given by

$$(x_c, y_c) = (2F/\pi)\int\int\left(\frac{\partial W}{\partial x}, \frac{\partial W}{\partial \mathrm{y}}\right)dx\,dy \quad . \tag{52}$$

The root-mean-square (rms) radius (or the radius of gyration) of the image distribution for a uniformly illuminated pupil is given by

$$r_{irms} = \left[\frac{1}{\pi}\int_0^1\int_0^{2\pi} r_i^2\,\rho\,d\rho\,d\theta\right]^{1/2} \quad . \tag{53}$$

where $r_i^2 = x_i^2 + y_i^2$. The balanced primary aberrations for minimum rms radius are listed in Table 11. The value of rms radius in this table is multiplied by the spot radius in the Gaussian image plane in the case of spherical aberration,

Table 11. Balanced aberrations for minimum rms spot radius

Balanced Aberration	$W(\rho, \theta)$	RMS Radius
Spherical	$A_s\left(\rho^4 - \frac{4}{3}\rho^2\right)$	$0.167(8FA_s)$
Coma	$A_c\left(\rho^3 - \rho\right)\cos\theta$	$0.272(6FA_c)$
Astigmatism	$A_a\rho^2\left(\cos^2\theta - 1/2\right)$	$0.177(8FA_a)$

tangential coma in the case of coma, and the full length of tangential or sagittal line image in the case of astigmatism. The balancing defocus in the case of spherical aberration and the balancing tilt in the case of coma are different from their corresponding values that minimize the aberration variance. The balancing defocus in the case of astigmatism is the same in both cases.

A balanced aberration, giving the smallest rms radius, in terms of Zernike circle polynomial $R_n^m(\rho)\cos m\theta$, is given by $B_n^m(\rho)\cos m\theta$, where[16]

$$B_n^m(\rho) = R_n^m(\rho) - R_{n-2}^m(\rho) \quad . \tag{54}$$

These polynomials are listed in Table 12 and may be obtained from the Zernike polynomials given in Table 5. If the aberration function is written in terms of these polynomials, e.g.,

$$W(\rho, \theta) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} b_n^m B_n^m(\rho)\cos m\theta \quad , \tag{55}$$

then the rms radius of the image spot is given by

$$r_{irms} = 2F\left\{\sum_{n/2=1}^{\infty} 4n\left(b_n^0\right)^2 + \sum_{m=1}^{\infty}\left[m\left(b_m^m\right)^2 + \sum_{i=1}^{\infty} 2(2i+m)\left(b_{2i+m}^m\right)^2\right]\right\}^{1/2} \quad . \tag{56}$$

The polynomials $B_4^0(\rho)$, $B_3^1(\rho)\cos\theta$, and $B_2^2(\rho)\cos 2\theta$ represent balanced spherical aberration, coma, and astigmatism giving a minimum rms spot radius.

8. SUMMARY AND CONCLUSIONS

In this paper, we have identified balanced aberrations with the appropriate Zernike polynomials for systems with circular, annular, and Gaussian pupils. The aberrations are balanced for minimum variance, which yields maximum Strehl ratio for small aberrations. The balanced aberrations for minimum rms spot radius are also considered and identified with polynomials derived from Zernike polynomials. Aberrations can also be balanced for maximizing the Hopkins ratio[17-19] for the modulation transfer function (MTF) or the Struve ratio for the line-spread function (LSF).[20] The lower-order balancing aberrations for a certain higher-order aberration are the same in each case, but their amounts vary depending on the type of pupil (circular, annular, or Gaussian) or the criterion used (e.g., Strehl ratio, rms spot radius, Hopkins ratio, or the Struve ratio).

As a point object moves away from the optical axis, the pupil is vignetted for sufficiently large field angles. The vignetted pupil can be approximated by an ellipse.[21] The polynomials that are orthogonal over the elliptical pupil can be obtained from the circle polynomials by scaling the x and y coordinates of a circular pupil such that their ratio is equal to the ratio of the major and minor axes of the ellipse. The elliptical polynomials thus obtained, while orthogonal over the elliptical pupil, can not be identified with the classical aberrations. The radially symmetric aberrations cannot be

Table 12. Balanced polynomials for minimum rms spot radius

n	m	$B_n^m(\rho)\cos m\theta$	**Balanced Aberration**
0	0	1	Piston
1	1	$\rho\cos\theta$	Tilt
2	0	$2(\rho^2-1)$	Defocus
2	2	$\rho^2\cos 2\theta$	Primary astigmatism
3	1	$3(\rho^3-\rho)\cos\theta$	Primary coma
3	3	$\rho^3\cos 3\theta$	
4	0	$2(3\rho^4-4\rho^2+1)$	Primary spherical
4	2	$4(\rho^4-\rho^2)\cos 2\theta$	Secondary astigmatism
4	4	$\rho^4\cos 4\theta$	
5	1	$5(2\rho^5-3\rho^3+\rho)\cos\theta$	Secondary coma
5	3	$5(\rho^5-\rho^3)\cos 3\theta$	
5	5	$\rho^5\cos 5\theta$	
6	0	$2(10\rho^6-18\rho^4+9\rho^2-1)$	Secondary spherical
6	2	$3(5\rho^6-8\rho^4+3\rho^2)\cos 2\theta$	Tertiary astigmatism
6	4	$6(\rho^6-\rho^4)\cos 4\theta$	
6	6	$\rho^6\cos 6\theta$	
7	1	$7(5\rho^7-10\rho^5+6\rho^3-\rho)\cos\theta$	Tertiary coma
7	3	$7(3\rho^7-5\rho^5+2\rho^3)\cos 3\theta$	
7	5	$7(\rho^7-\rho^5)\cos 5\theta$	
7	7	$\rho^7\cos 7\theta$	
8	0	$2(35\rho^8-80\rho^6+60\rho^4-16\rho^2+1)$	Tertiary spherical

represented by the polynomials with $m = 0$. The defocus aberration, for example, can not be identified with the elliptical polynomial corresponding to $n = 2$ and $m = 0$. Hence, it would seem that the user will want to use the circle or annular polynomials even when the pupil deviates from being circular or annular, respectively.

The aberration function for a system without an axis of rotational symmetry will consist of terms not only in cos $m\theta$ but in sin $m\theta$ as well. This is true also of aberrations resulting from fabrication errors as well as those introduced by atmospheric turbulence.[22] Given a wavefront as measured by interferometry, it is quite common to determine its aberration coefficients in terms of the Zernike polynomials.[23] The reader interested in more details on the content of this paper may refer to the author's books on *Optical Imaging and Aberrations*.[24, 25]

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**, 689–794 (1934).
2. B. R. A. Nijboer, "The diffraction theory of optical aberrations. Part I: General discussion of the geometrical aberrations," *Physica* **10**, 679–692 (1943); "The diffraction theory of optical aberrations. Part II: Diffraction pattern in the presence of small aberrations," *Physica* **13**, 605–620 (1947); and K. Nienhuis and B. R. A. Nijboer, "The diffraction theory of optical aberrations. Part III: General formulae for small aberrations: experimental verification of the theoretical results," *Physica* **14**, 590–608 (1949).
3. M. Born and E. Wolf, *Principles of Optics*, 7th ed., Oxford, New York (1999).
4. A. Maréchal, "Etude des effets combines de la diffraction et des aberrations geometriques sur l'image d'un point lumineux," *Revue d'Optique* **26**, 257–277 (1947).
5. V. N. Mahajan, "Strehl ratio for primary aberrations: some analytical results for circular and annular pupils," *J. Opt. Soc. Am*. **72**, 1258–1266 (1982), Errata, **10**, 2092 (1993); "Strehl ratio for primary aberrations in terms of their aberration variance," *J. Opt. Soc. Am*. **73**, 860–861 (1983).
6. V. N. Mahajan, "Symmetry properties of aberrated point-spread functions," *J. Opt. Soc. Am.* **A11**, 1993–2003 (1994).
7. V. N. Mahajan, "Line of sight of an aberrated optical system," *J. Opt. Soc. Am*. **2**, 833–846 (1985).
8. W. B. King, "Dependence of the Strehl ratio on the magnitude of the variance of the wave aberration," *J. Opt. Soc. Am*. **58**, 655–661 (1968).
9. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," *J. Opt. Soc. Am.* **71**, 75–85 (1981); **71,** 1408 (1981); **1**, 685 (1984); "Zernike annular polynomials and optical aberrations of systems with annular pupils," *Appl. Opt.* **33**, 8125–8127 (1994).
10. 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).
11. V. N. Mahajan, "Zernike circle polynomials and optical aberrations of systems with circular pupils," *Appl. Opt.* **33**, 8121–8124 (1994); and "Zernike polynomials and optical aberrations," *Appl. Opt.* **34**, 8060–8062 (1995).
12. A. B. Bhatia and E. Wolf, "On the circle polynomials of Zernike and related orthogonal sets," *Proc. Cambridge Philos. Soc.* **50**, 40–48 (1954).
13. S. Szapiel, "Aberration-variance-based formula for calculating point-spread functions: rotationally symmetric aberrations," *J. Opt. Soc. Am.* **25**, 244–251 (1986).
14. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1968), p. 454.
15. V. N. Mahajan, "Zernike-Gauss polynomials for optical systems with Gaussian pupils," *Appl. Opt*. **34**, 8057–8059 (1995).
16. J. Braat, "Polynomial expansion of severely aberrated wave fronts," *J. Opt. Soc. Am*. **A4**, 643–650 (1987).
17. H. H. Hopkins, "The aberration permissible in optical systems," *Proc. Phys. Soc*. (London) **B52**, 449–470 (1957)
18. W. B. King, "Correlation between the relative modulation function and the magnitude of the wave aberration difference function," *J. Opt. Soc. Am*. **59**, 692–697 (1969).
19. S. Szapiel, "Hopkins variance formula extended to low relative modulations," *Optica Acta* **33**, 981–999 (1986).
20. H. H. Hopkins and B. Zalar, "Aberration tolerances based on the line spread function," *J. Mod. Opt*. **34**, 371–406 (1987).
21. W. B. King "The approximation of vignetted pupil shape by an ellipse," *Appl. Opt*. **7**, 197–201 (1968).
22. R. J. Noll, "Zernike polynomials and atmospheric turbulence," *J. Opt. Soc. Am.* **66**, 207–211 (1976).
23 J. Y. Wand and D. E. Silva, "Wavefront interpretation with Zernike polynomials," *Appl. Opt*. **19,** 1510–1518 (1980).
24. V. N. Mahajan, *Optical Imaging and Aberrations*, Part I: *Ray Geometrical Optics*, SPIE Press, Bellingham, Washington (1998).
25. V. N. Mahajan, *Optical Imaging and Aberrations*, Part II: *Wave Diffraction Optics*, SPIE Press, Bellingham, Washington (2001).