OPTICAL
Specification, Fabrication, and Testing

Jim Schwiegerling

Since this set of functions is orthogonal, Eq. (2.80) can be used to represent an arbitrary function $f(\theta)$ as a linear combination of the terms in the set such that

$$f(\theta) = \sum_m a_m \Theta_m(\theta), \tag{2.90}$$

for θ between 0 and 2π. The values of the expansion coefficients are given by Eq. (2.83) such that

$$a_m = \frac{1}{(1 + \delta_{m0})\pi} \int\limits_0^{2\pi} f(\theta)\Theta_m(\theta)d\theta. \tag{2.91}$$

It is left to the reader to verify that this definition is consistent with more conventional representations of the Fourier series.

2.3.4 Zernike polynomials

For fitting wavefront errors over a circular exit pupil or representing aspheric surface shapes over a circular aperture, a two-dimensional set of functions that are orthogonal over the unit circle is useful. In referring to the orthogonality condition in Eq. (2.79), normalized polar coordinates are the natural system for representing the position vector such that $\mathbf{r} \Rightarrow (\rho, \theta)$. The differential $d\mathbf{r} = \rho d\rho d\theta$. The domain of integration is the unit circle, so ρ ranges from 0 to 1, and θ ranges from 0 to 2π. Under these requirements, a set of functions that satisfy

$$\int\limits_0^{2\pi}\int\limits_0^1 V_i(\rho, \theta)V_j(\rho, \theta)\rho d\rho d\theta = \begin{cases} C_j & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{2.92}$$

is needed. One approach for finding a suitable set of functions is to use separation of variables to split the two-dimensional functions into a product of one-dimensional functions:

$$\int\limits_0^1 R_i(\rho)R_j(\rho)\rho d\rho \int\limits_0^{2\pi} \Theta_i(\theta)\Theta_j(\theta)d\theta = \begin{cases} C_j & \text{for } i = j \\ 0 & \text{otherwise.} \end{cases} \tag{2.93}$$

The θ integral is now exactly the orthogonality requirement seen in the discussion on Fourier series in the previous section. Consequently, sine and cosine terms with integer multiples of the angular variable are solutions to the azimuthal component of the preceding expression. The first integral containing the radial component, however, has an extra ρ from the polar differential and therefore requires alternate solutions to satisfy the orthogonality relationship. Fortunately, a variety of sets satisfying the requirements of the radial component are known. One such set of functions are the Zernike polynomials.

The Zernike polynomials are a set of functions that are orthogonal over the unit circle. This set has found extensive optical applications in areas such as representing non–rotationally symmetric higher-order aspheric optical surface shapes, fitting wavefront error emerging from an exit pupil, describing the phase errors introduced by atmospheric turbulence in astronomical viewing, and fitting interferograms in optical metrology. This set of functions was developed by Zernike[1] in the 1930s for application to phase contrast imaging. The widespread use of Zernike polynomials in these applications is due to their mathematical properties over a circular domain, which easily matches the typical shape of an exit pupil or optical element. These properties include orthogonality over the continuous unit circle, continuous derivatives, terms that are closely related to familiar aberrations such as coma, astigmatism, and spherical aberration; additionally, they form a complete set, which means that they can represent arbitrarily complex shapes if enough terms are used in the expansion.

The Zernike polynomials are defined as[2,3]

$$Z_n^m(\rho, \theta) = \begin{cases} N_n^m R_n^{|m|}(\rho)\cos m\theta & \text{for } m \geq 0, \\ -N_n^m R_n^{|m|}(\rho)\sin m\theta & \text{for } m < 0, \end{cases} \tag{2.94}$$

where N_n^m is a normalization constant, and $R_n^{|m|}(\rho)$ is a polynomial in ρ. Note that two indices are needed to uniquely define each Zernike polynomial. The first index, known as the *radial order n*, is the maximum polynomial order of $R_n^{|m|}(\rho)$. The second index, known as the *azimuthal frequency m*, is the integer multiplier of the angular variable θ. The index n is a positive integer or zero. The index m is also an integer and must satisfy $|m| \leq n$. Furthermore, $n-|m|$ must be even. The definition in Eq. (2.94) follows national and international standards on reporting aberrations in the human eye. The definitions of the Zernike polynomials vary in other areas of optics. These differences will be discussed in more detail in Section 2.3.8. The normalization constant is defined by

$$N_n^m = \sqrt{\frac{2n+2}{1+\delta_{m0}}}. \tag{2.95}$$

The radial polynomials are given by

$$R_n^{|m|}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s(n-s)!}{s!\left[\frac{n+|m|}{2}-s\right]!\left[\frac{n-|m|}{2}-s\right]!}\rho^{n-2s}. \tag{2.96}$$

Note that the absolute value of m in the definition means that the radial polynomial is identical for the corresponding sine and cosine terms. The radial polynomials themselves satisfy the following orthogonality condition:[4]

$$\int_0^1 R_n^{|m|}(\rho) R_{n'}^{|m|}(\rho)\rho d\rho = \frac{\delta_{nn'}}{2n+2}. \tag{2.97}$$

Finally, combining Eqs. (2.88), (2.95), and (2.97), the orthogonality relation for the Zernike polynomials is obtained such that

$$\int_0^{2\pi} \int_0^1 Z_n^m(\rho, \theta) Z_{n'}^{m'}(\rho, \theta) \rho d\rho d\theta = \pi \delta_{nn'} \delta_{mm'}. \qquad (2.98)$$

2.3.5 Examples of different orders of Zernike polynomials

Table 2.3 lists the Zernike polynomials through sixth order ($n = 6$). A total of 28 polynomials are listed. The final column of the table shows some common names for the various terms, although these names are not universal. These names are often preceded by "Zernike" to distinguish them from their Seidel equivalents. It is also sometimes useful to use a single index j to reference the individual polynomials. Extreme care should be taken when using a single index since it does not uniquely identify the specific polynomial, and a variety of numbering and ordering schemes exist. Here, the scheme from Refs. 2 and 3 is used. Other variations are described in

Table 2.3   The Zernike polynomials through sixth order.

j	n	m	$Z_n^m(\rho, \theta)$	Name
0	0	0	1	Piston
1	1	–1	$2\rho\sin\theta$	Vertical tilt
2	1	1	$2\rho\cos\theta$	Horizontal tilt
3	2	–2	$\sqrt{6}\rho^2\sin2\theta$	Oblique astigmatism
4	2	0	$\sqrt{3}(2\rho^2 - 1)$	Defocus
5	2	2	$\sqrt{6}\rho^2\cos2\theta$	Horizontal astigmatism
6	3	–3	$\sqrt{8}\rho^3\sin3\theta$	Oblique trefoil
7	3	–1	$\sqrt{8}(3\rho^3 - 2\rho)\sin\theta$	Oblique coma
8	3	1	$\sqrt{8}(3\rho^3 - 2\rho)\cos\theta$	Horizontal coma
9	3	3	$\sqrt{8}\rho^3\cos3\theta$	Horizontal trefoil
10	4	–4	$\sqrt{10}\rho^4\sin4\theta$	Oblique quatrefoil
11	4	–2	$\sqrt{10}(4\rho^4 - 3\rho^2)\sin2\theta$	Oblique secondary astigmatism
12	4	0	$\sqrt{5}(6\rho^4 - 6\rho^2 + 1)$	Spherical aberration
13	4	2	$\sqrt{10}(4\rho^4 - 3\rho^2)\cos2\theta$	Horizontal secondary astigmatism
14	4	4	$\sqrt{10}\rho^4\cos4\theta$	Horizontal quatrefoil
15	5	–5	$\sqrt{12}\rho^5\sin5\theta$	
16	5	–3	$\sqrt{12}(5\rho^5 - 4\rho^3)\sin3\theta$	
17	5	–1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho)\sin\theta$	Vertical secondary coma
18	5	1	$\sqrt{12}(10\rho^5 - 12\rho^3 + 3\rho)\cos\theta$	Horizontal secondary coma
19	5	3	$\sqrt{12}(5\rho^5 - 4\rho^3)\cos3\theta$	
20	5	5	$\sqrt{12}\rho^5\cos5\theta$	
21	6	–6	$\sqrt{14}\rho^6\sin6\theta$	
22	6	–4	$\sqrt{14}(6\rho^6 - 5\rho^4)\sin4\theta$	
23	6	–2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2)\sin2\theta$	
24	6	0	$\sqrt{7}(20\rho^6 - 30\rho^4 + 12\rho^2 - 1)$	
25	6	2	$\sqrt{14}(15\rho^6 - 20\rho^4 + 6\rho^2)\cos2\theta$	
26	6	4	$\sqrt{14}(6\rho^6 - 5\rho^4)\cos4\theta$	
27	6	6	$\sqrt{14}\rho^6\cos6\theta$	

Section 2.3.8. Using this format, the index j starts with zero and is related to the radial and azimuthal indices by

$$j = \frac{n(n+2)+m}{2}. \tag{2.99}$$

It is also possible to recover the radial index from j with

$$n = \left\lceil \frac{-3 + \sqrt{9+8j}}{2} \right\rceil, \tag{2.100}$$

where $\lceil\ \rceil$ is the ceiling operator, and the azimuthal index from

$$m = 2j - n(n+2). \tag{2.101}$$

For example, $Z_3^1(\rho, \theta)$ has $n = 3$ and $m = 1$. The corresponding single index $j = 8$. Reversing the process,

$$n = \left\lceil \frac{-3 + \sqrt{73}}{2} \right\rceil = \lceil 2.772 \rceil = 3, \tag{2.102}$$

and $m = 2(8) - 3(3+2) = 1$.

Figure 2.24 shows the first three Zernike polynomials and represents the cases of $n = 0$ and $m = 0$, as well as $n = 1$ and $m = \pm 1$. The first term is constant over the unit circle and represents a piston-type shift in the surface being fit by the polynomials. Figures 2.24(a) and (b) are planes rotated about the ρ_y axis and ρ_x axis, respectively. These terms represent a tilt to the surface under fit. Suppose that a function $f(\rho, \theta)$ is represented by a linear combination of these terms such that

$$f(\rho, \theta) = 2a_{1,-1}\rho\sin\theta + 2a_{1,1}\rho\cos\theta. \tag{2.103}$$

Switching to normalized Cartesian coordinates gives

$$f(\rho_x, \rho_y) = 2a_{1,-1}\rho_y + 2a_{1,1}\rho_x. \tag{2.104}$$

Equation (2.104) describes a plane passing through the origin that is tilted about the line $a_{1,-1}\rho_y + a_{1,1}\rho_x = 0$. In other words, combining the two planes
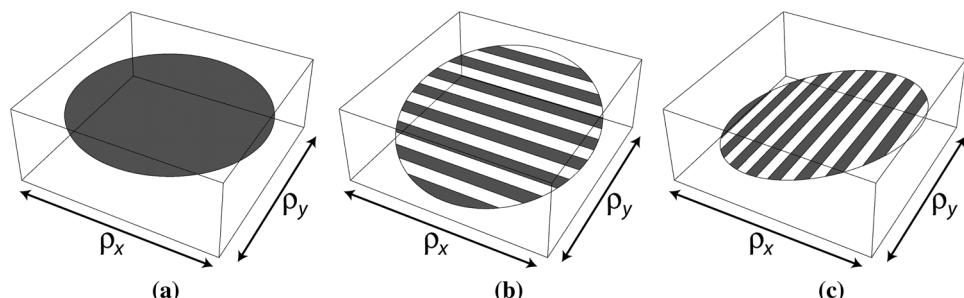
Figure 2.24 The zeroth and first-order Zernike polynomials (a) $Z_0^0(\rho, \theta)$, (b) $Z_1^{-1}(\rho, \theta)$, and (c) $Z_1^1(\rho, \theta)$.

described by these two Zernike terms results in a plane that tilted about an axis that is dependent on the coefficients $a_{1,-1}$ and $a_{1,1}$. This theme where the sine and cosine pairs of Zernike polynomials combine to form the same-shaped surface rotated in the θ direction reoccurs throughout the set of polynomials. For example, consider the general pair

$$f(\rho, \theta) = a_{n,-|m|} N_n^m R_n^{|m|}(\rho)\sin|m|\theta + a_{n,|m|} N_n^m R_n^{|m|}(\rho)\cos|m|\theta. \qquad (2.105)$$

This expression can be rewritten as

$$f(\rho, \theta) = \sqrt{a_{n,-|m|}^2 + a_{n,|m|}^2}\, N_n^m R_n^{|m|}(\rho)\cos|m|(\theta - \theta_0), \qquad (2.106)$$

where

$$\theta_0 = \frac{1}{|m|}\tan^{-1}\left[\frac{a_{n,-|m|}}{a_{n,|m|}}\right]. \qquad (2.107)$$

Equation (2.106) has the same form as the cosine term but is rotated by θ_0 in the θ direction.

Figure 2.25 shows the second-order ($n = 2$) Zernike polynomials. Figures 2.25(a) and (c) are a matched pair of saddle-shaped surfaces, with the $Z_2^{-2}(\rho, \theta)$ term rotated by $\pi/4$ in the θ direction relative to $Z_2^2(\rho, \theta)$. This pair of surfaces in known as *Zernike astigmatism* since, when combined, they form a surface with an equal and opposite quadratic dependence along two orthogonal meridians, in much the same manner that Seidel astigmatism and field curvature in Eq. (2.62) have different quadratic dependence in the ρ_x direction and ρ_y direction. Figure 2.25(b) shows the *Zernike defocus* term, which has a quadratic dependence on ρ.

Figure 2.26 and 2.27 show the third-order Zernike terms. Figures 2.26(a) and (b) are the *Zernike coma* terms, named for the functional format ($\rho^3\cos\theta$) that is similar to the Seidel coma of Eq. (2.56). Figures 2.27(a) and (b) are known as *trefoil* aberration. This three-lobed pattern does not have an analogous Seidel aberration.
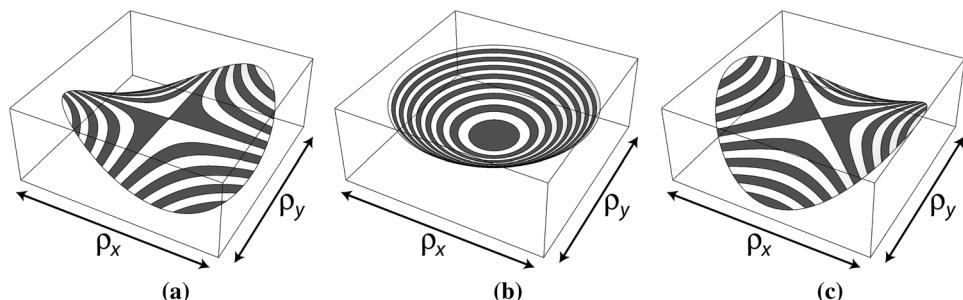
Figure 2.25 The second-order Zernike polynomials (a) $Z_2^{-2}(\rho, \theta)$, (b) $Z_2^0(\rho, \theta)$, and (c) $Z_2^2(\rho, \theta)$.
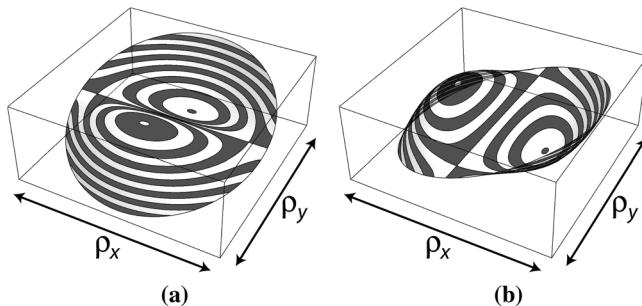
Figure 2.26 The Zernike coma terms (a) $Z_3^{-1}(\rho, \theta)$ and (b) $Z_3^1(\rho, \theta)$.
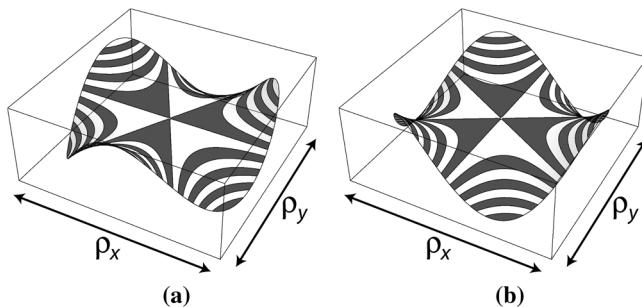
Figure 2.27 The Zernike trefoil terms (a) $Z_3^{-3}(\rho, \theta)$ and (b) $Z_3^3(\rho, \theta)$.

Finally, Fig. 2.28 shows the fourth-order ($n = 4$) *Zernike spherical aberration* term. This function has ρ^4 dependence, much like the Seidel spherical aberration in Eq. (2.5), but includes additional defocus and piston terms. The other fourth-order terms include secondary astigmatism with $n = 4$ and $m = \pm 2$, as well as a pair of four-lobed terms known as quatrefoil with $n = 4$ and $m = \pm 4$.
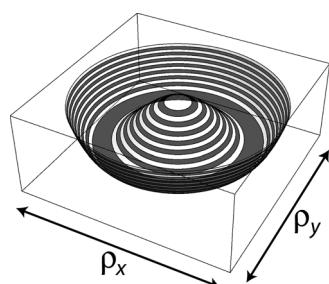
Figure 2.28 The Zernike spherical aberration term $Z_4^0(\rho, \theta)$.

2.3.6 Fitting Zernike polynomials to wavefront error

A two-dimensional function $W(\rho, \theta)$ can be represented as a linear combination of Zernike polynomials. Mathematically, this expansion is given by

$$W(\rho, \theta) = \sum_{n,m} a_{nm} Z_n^m(\rho, \theta), \qquad (2.108)$$

where only valid values of n and m are used in the sum. Since the Zernike polynomials are orthogonal over the unit circle, Eq. (2.83) can be used to determine the expansion coefficients such that

$$a_{nm} = \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 W(\rho, \theta) Z_n^m(\rho, \theta) \rho d\rho d\theta. \qquad (2.109)$$

As an example, the Zernike expansion of $W(\rho, \theta) = 2\rho^3 \cos\theta$ is determined with the following procedure. First, note that the largest exponent of ρ is 3. This suggests that the radial index satisfies $0 \le n \le 3$, and higher-order Zernike polynomials can be ignored. Second, the function to be fit only depends on $\cos\theta$, so the only valid value of m is 1. Equation (2.109) is then used to evaluate each coefficient that satisfies the requirements on n and m, which are in this case

$$a_{1,1} = \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 (2\rho^3 \cos\theta) \cdot (2\rho\cos\theta)\rho d\rho d\theta = \frac{4}{\pi} \int\limits_0^1 \rho^5 d\rho \int\limits_0^{2\pi} \cos^2\theta = \frac{2}{3}, \qquad (2.110)$$

and

$$a_{3,1} = \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 (2\rho^3 \cos\theta) \cdot \sqrt{8}(3\rho^3 - 2\rho)\cos\theta \rho d\rho d\theta$$

$$= \frac{2\sqrt{8}}{\pi} \int\limits_0^1 (3\rho^7 - 2\rho^5)d\rho \int\limits_0^{2\pi} \cos^2\theta = \frac{\sqrt{2}}{6}, \qquad (2.111)$$

so

$$2\rho^3 \cos\theta = \frac{2}{3} Z_1^1(\rho, \theta) + \frac{\sqrt{2}}{6} Z_3^1(\rho, \theta). \qquad (2.112)$$

The integrals for all of the other coefficients are zero.

A second technique for finding the Zernike expansion $W(\rho, \theta) = 2\rho^3 \cos\theta$ is through inspection of the definitions of the various Zernike terms. Examining Table 2.3, $Z_3^1(\rho, \theta)$ contains a $\rho^3 \cos\theta$ term. The definition of $Z_3^1(\rho, \theta)$ can be rearranged to give

$$\rho^3 \cos\theta = \frac{1}{3\sqrt{8}} [Z_3^1(\rho, \theta) + 2\sqrt{8}\rho\cos\theta]. \qquad (2.113)$$

Furthermore, the $\rho\cos\theta$ term is related to the $Z_1^1(\rho, \theta)$ definition by

$$\rho\cos\theta = \frac{1}{2}\left[Z_1^1(\rho, \theta)\right]. \tag{2.114}$$

Incorporating these last two results,

$$W(\rho, \theta) = 2\rho^3\cos\theta = \frac{2}{3\sqrt{8}}\left[Z_3^1(\rho, \theta) + \sqrt{8}Z_1^1(\rho, \theta)\right] = \frac{2}{3}Z_1^1(\rho, \theta) + \frac{\sqrt{2}}{6}Z_3^1(\rho, \theta), \tag{2.115}$$

which is identical to the previous expansion results.

Often, a continuous mathematical description of the function to be fit is not available. Instead, a series of discretely sampled data needs to be fit. An example would be a set of wavefront error measurements at a series of exit pupil coordinates. In this case, the orthogonality relationship of the Zernike polynomials is no longer valid. However, if the function is well sampled over a circular region, many of the benefits of fitting to this polynomial set are retained. Calculation however becomes a matrix operation. Given a set of measurements $\{W(\rho_i, \theta_i)\}$ captured at the points (ρ_i, θ_i), where $i = 1...N$, a matrix equation

$$\mathbf{Ax} = \mathbf{b} \tag{2.116}$$

can be solved to determine the expansion coefficients. Here, the bold upper-case variables represent matrices, and the bold lower-case variables are column vectors. The matrix $\mathbf{A}$ has N rows corresponding to the number of sample points and typically $j + 1$ columns, where j is given by Eq. (2.99) with $n = n_{\max}$, the maximum radial order of the Zernike polynomials to be fit, and $m = n_{\max}$ as well. The matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{pmatrix} Z_0^0(\rho_1, \theta_1) & Z_1^{-1}(\rho_1, \theta_1) & Z_1^1(\rho_1, \theta_1) & \cdots & Z_{n_{\max}}^{n_{\max}}(\rho_1, \theta_1) \\ Z_0^0(\rho_2, \theta_2) & Z_1^{-1}(\rho_2, \theta_2) & Z_1^1(\rho_2, \theta_2) & \cdots & Z_{n_{\max}}^{n_{\max}}(\rho_2, \theta_2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Z_0^0(\rho_N, \theta_N) & Z_1^{-1}(\rho_N, \theta_N) & Z_1^1(\rho_N, \theta_N) & \cdots & Z_{n_{\max}}^{n_{\max}}(\rho_N, \theta_N) \end{pmatrix}, \tag{2.117}$$

so each column corresponds to a different Zernike polynomial in the expansion, and each row corresponds to the respective Zernike polynomial evaluated at the measurement point (ρ_i, θ_i). The vector $\mathbf{x}$ is a column vector containing the expansion coefficients that need to be determined in order to represent the measured data as a linear expansion, as in Eq. (2.108). The vector $\mathbf{x}$ has $j + 1$ elements, given by

$$\mathbf{x} = \begin{pmatrix} a_{0,0} & a_{1,-1} & a_{1,1} & \cdots & a_{n_{\max}n_{\max}} \end{pmatrix}^{\mathrm{T}}, \tag{2.118}$$

where T represents the transpose operation. The vector $\mathbf{b}$ is a column vector containing the known measured data $\{W(\rho_i, \theta_i)\}$. The vector $\mathbf{b}$ has N elements and is given by

$$\mathbf{b} = \begin{bmatrix} W(\rho_1, \theta_1) & W(\rho_2, \theta_2) & W(\rho_2, \theta_2) & \cdots & W(\rho_N, \theta_N) \end{bmatrix}^{\mathrm{T}}. \tag{2.119}$$

Equation (2.116) needs to be solved for the expansion coefficients defined in $\mathbf{x}$. If $N = j + 1$, then the matrix $\mathbf{A}$ is square, and both sides of Eq. (2.116) can be multiplied by the inverse matrix $\mathbf{A}^{-1}$ (assuming it exists) to recover $\mathbf{x}$. However, this case typically leads to solutions where the linear expansion oscillates wildly in the regions between sample points. Instead, typically, $N >> j + 1$, so the matrix $\mathbf{A}$ has many more rows than columns, and an inverse for a nonsquare matrix is not defined. In this case, both sides of Eq. (2.116) are multiplied by $\mathbf{A}^{\mathrm{T}}$ to give

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \mathbf{A}^{\mathrm{T}}\mathbf{b}. \tag{2.120}$$

The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ is now square, and multiplying both sides of the preceding equation by the inverse of this matrix gives

$$\mathbf{x} = [\mathbf{A}^{\mathrm{T}}\mathbf{A}]^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b}. \tag{2.121}$$

The elements of $\mathbf{x}$ contain the expansion coefficients a_{nm}, which represent a least-squares fit to the sampled data.

As an example of this calculation, suppose that the data shown in Table 2.4 needs to be fit to a set of Zernike polynomials. Here, the number of data points is $N = 13$. The second and third columns show the location of the samples. The sampled points are located on a Cartesian grid with spacing of 0.5 in the horizontal and vertical directions contained within the unit circle. The fourth and fifth columns of the table show these sample locations converted to polar coordinates. Finally, the last column shows the measured data at the sample locations. Since there are 13 data points, the number of Zernike polynomial terms to be fit should be less than this number. Here, the maximum radial order $n_{\max}$ will be chosen to be 2, so that $j + 1 = 6$, based on Eq. (2.99). Consequently, the matrix $\mathbf{A}$ will have 13 rows and 6 columns, and is given by

Table 2.4   Example data for fit to Zernike polynomials.

i	ρ_{xi}	ρ_{yi}	ρ_i	θ_i	$W(\rho_i, \theta_i)$
1	0.000000	1.000000	1.000000	1.570796	3.394057
2	−0.500000	0.500000	0.500000	2.356194	−0.577350
3	0.000000	0.500000	0.250000	1.570796	−1.952935
4	0.500000	0.500000	0.500000	0.785398	−0.577350
5	−1.000000	0.000000	1.000000	3.141593	5.843547
6	−0.500000	0.000000	0.250000	3.141593	−1.799842
7	0.000000	0.000000	0.000000	0.000000	−2.309401
8	0.500000	0.000000	0.250000	0.000000	−1.799842
9	1.000000	0.000000	1.000000	0.000000	5.843547
10	−0.500000	−0.500000	0.500000	−2.356194	−0.577350
11	0.000000	−0.500000	0.250000	−1.570796	−1.952935
12	0.500000	−0.500000	0.500000	−0.785398	−0.577350
13	0.000000	−1.000000	1.000000	−1.570796	3.394057

$$\mathbf{A} = \begin{pmatrix} 1.000000 & 2.000000 & 0.000000 & 0.000000 & 1.732051 & -2.449490 \\ 1.000000 & 0.707107 & -0.707107 & -0.612372 & -0.866025 & 0.000000 \\ 1.000000 & 0.500000 & 0.000000 & 0.000000 & -1.515544 & -0.153093 \\ 1.000000 & 0.707107 & 0.707107 & 0.612372 & -0.866025 & 0.000000 \\ 1.000000 & 0.000000 & -2.000000 & 0.000000 & 1.732051 & 2.449490 \\ 1.000000 & 0.000000 & -0.500000 & 0.000000 & -1.515544 & 0.153093 \\ 1.000000 & 0.000000 & 0.000000 & 0.000000 & -1.732051 & 0.000000 \\ 1.000000 & 0.000000 & 0.500000 & 0.000000 & -1.515544 & 0.153093 \\ 1.000000 & 0.000000 & 2.000000 & 0.000000 & 1.732051 & 2.449490 \\ 1.000000 & -0.707107 & -0.707107 & 0.612372 & -0.866025 & 0.000000 \\ 1.000000 & -0.500000 & 0.000000 & 0.000000 & -1.515544 & -0.153093 \\ 1.000000 & -0.707107 & 0.707107 & -0.612372 & -0.866025 & 0.000000 \\ 1.000000 & -2.000000 & 0.000000 & 0.000000 & 1.732051 & -2.449490 \end{pmatrix}.$$

$$(2.122)$$

The vector **b** corresponds to the last column of Table 2.4, or

$$\mathbf{b} = \begin{pmatrix} 3.394057 \\ -0.577350 \\ -1.952935 \\ -0.577350 \\ 5.843547 \\ -1.799842 \\ -2.309401 \\ -1.799842 \\ 5.843547 \\ -0.577350 \\ -1.952935 \\ -0.577350 \\ 3.394057 \end{pmatrix}.$$

$$(2.123)$$

The matrix **A** and its transpose are then used to calculate

$$[\mathbf{A}^{\mathrm{T}}\mathbf{A}]^{-1} = \begin{pmatrix} 0.081232 & 0.000000 & 0.000000 & 0.000000 & 0.012938 & 0.000000 \\ 0.000000 & 0.095238 & 0.000000 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 0.095238 & 0.000000 & 0.000000 & 0.000000 \\ 0.000000 & 0.000000 & 0.000000 & 0.666667 & 0.000000 & 0.000000 \\ 0.012938 & 0.000000 & 0.000000 & 0.000000 & 0.038842 & 0.000000 \\ 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.000000 & 0.041505 \end{pmatrix}.$$

$$(2.124)$$

Finally, Eq. (2.121) is used to calculate the vector **x** containing the expansion coefficients such that

$$\mathbf{x} = \begin{pmatrix} 1.154701 \\ 0.000000 \\ 0.000000 \\ 0.000000 \\ 2.000000 \\ 0.500000 \end{pmatrix}. \tag{2.125}$$

Based on this result, the Zernike expansion for this example is given by

$$W(\rho, \theta) = 1.154701 Z_0^0(\rho, \theta) + 2Z_2^0(\rho, \theta) + 0.5Z_2^2(\rho, \theta). \tag{2.126}$$

2.3.7 Pupil size conversion

Often it useful to scale a Zernike fit to a smaller diameter. An example of such an instance is the case in which, given the wavefront over a large exit pupil, the system is then stopped down to a smaller pupil. The emerging wavefront is unchanged, but the peripheral regions are clipped by the smaller pupil. If the Zernike polynomial fit over the large pupil is known, the coefficients of the Zernike polynomial fit over the smaller pupil can be calculated directly from the original expansion coefficients. The wavefronts are identical over the region of the smaller pupil, but the expansion coefficients are different to account for the clipping. A variety of techniques have been developed to scale Zernike polynomials to new pupil sizes.[5–8] These derivations vary in their complexity, but the most elegant result is provided by Janssen and Dirksen.[8] First, the Zernike polynomial expansion over the original pupil with diameter $2r_1$ is given by

$$W(\rho, \theta) = \sum_{n, m} a_{nm} Z_n^m(\rho, \theta), \tag{2.127}$$

where $\rho = r/r_1$, the sum over n ranges from $n = 0, 1, 2 \ldots n_{\max}$, and the sum over m ranges from $-n, -n+2, \ldots n-2, n$. If the pupil is scaled to a new radius r_2, then the function $W(\rho, \theta)$ can be represented as a Zernike expansion over this smaller pupil as well. Defining $\varepsilon = r_2/r_1$, the function over the new pupil is given by

$$W(\varepsilon\rho, \theta) = \sum_{n, m} b_{nm} Z_n^m(\rho, \theta). \tag{2.128}$$

The goal of pupil rescaling is to relate the expansion coefficients b_{nm} to the expansion coefficients a_{nm}. In this manner, once the expansion is known for a given pupil radius r_1, the expansion can be easily determined for arbitrary

pupils of radius r_2. Janssen and Dirksen[8] showed that the new expansion coefficients are given by

$$b_{nm} = \frac{1}{N_n^m} \sum_{n'} a_{n'm} N_{n'}^m [R_{n'}^n(\varepsilon) - R_{n'}^{n+2}(\varepsilon)], \qquad (2.129)$$

where the sum ranges over $n' = n, \ n+2, \ \ldots n_{\max}$.

2.3.8 Different variations found in the literature

There are several caveats to using Zernike polynomials, especially when comparing results from multiple sources. A variety of variations in the definitions of the polynomials are found in the literature. When comparing different sources, care should be taken in understanding the fundamental definition used by the author so as to avoid errors and confusion. The definition of Zernike polynomials presented here is based on ANSI and ISO standards for reporting aberrations in the human eye.[3,4] Several variations of these definitions are presented here.

One such variation is the definition of the polar angle. Some texts, usually those related to aberration theory, use the polar angle ψ measured clockwise from the ρ_y axis, whereas the definition of polar angle θ used here is defined as the angle measured counterclockwise from the ρ_x axis. See Section 2.1.5 for further details on these definitions. In comparing these polar angles, $\theta = (\pi/2) - \psi$. This relationship has the effect of swapping some of the sine and cosine terms in the polynomial definitions and/or changing the sign of the azimuthal term. For example, the $\cos\theta$ terms (i.e., $m = 1$) in the definition here are equivalent to the $\sin\psi$ terms when the other polar angle is used. Similarly, the $\cos2\theta$ terms (i.e., $m = 2$) in the definition here are equivalent to $-\cos2\psi$ terms when the other polar angle is used, with the minus sign typically being absorbed into the expansion coefficient.

Another variation found in the literature is the use of the normalization factor N_n^m. Some definitions of the Zernike polynomials, such as the one used here, incorporate the normalization factor. Other definitions ignore the normalization factor and in effect absorb it into the expansion coefficients. Using the normalization factor provides a straightforward relationship between the variance of the wavefront error, as will be described in further detail in the next chapter. Ignoring the normalization factor relates the individual expansion coefficients to the peak-to-valley error of each of the individual Zernike terms. Comparing expansion into sets with and without the normalization factor,

$$\sum_{n,\,m} a_{nm} N_n^m R_n^{|m|}(\rho) \begin{Bmatrix} \cos m\theta & \text{for } m \geq 0 \\ -\sin m\theta & \text{for } m < 0 \end{Bmatrix}$$

$$= \sum_{n,\,m} b_{nm} R_n^{|m|}(\rho) \begin{Bmatrix} \cos m\theta & \text{for } m \geq 0 \\ -\sin m\theta & \text{for } m < 0 \end{Bmatrix}, \qquad (2.130)$$

it is evident that the coefficients are simply related by

$$a_{nm}N_n^m = b_{nm},\tag{2.132}$$

where the a_{nm} are the normalized expansion coefficients, and the b_{nm} are the unnormalized expansion coefficients. An example of an unnormalized definition of the Zernike polynomials is found in Annex A of the ISO 14999$-$2 standard.[9] Both the normalized and unnormalized forms are widely used.

A final variation on the definition is in the use of a single index for describing the various Zernike polynomial terms. It is important to understand that a single index is insufficient to uniquely describe Zernike polynomial definitions, and that the double-index notation should be used for clarity and to avoid confusion. The single-index scheme is useful for ordering the Zernike polynomials for computation such as the way in which the columns of the **A** matrix in Eq. (2.117) were ordered according to the single index j defined in Eq. (2.99). However, a variety of single index schemes appear in the literature and will be summarized here. Lens design packages such as Zemax (Radiant-Zemax, Redmond, WA) and Code V (Synopsys, Mountain View, CA) use the Fringe set of Zernike polynomials. This set, sometimes called the "Air Force" or "University of Arizona" set, is a subset of 37 Zernike terms that are unnormalized. The Fringe set uses a single-index scheme for ordering the terms. Zemax also provides the Zernike standard coefficients, which are a normalized set of Zernike polynomials that use a different single-index scheme for ordering. Note, the Zernike standard coefficients do not appear to be based on any ANSI or ISO standard but instead are based on a paper by Noll.[10] Table 2.5 describe the normalization and single-index schemes for these various Zernike polynomial definitions.

Table 2.5 Various definitions of Zernike polynomials, their normalization, and single-index ordering schemes.

Source	ANSI Z80.28-2004 ISO 24157:2008	ISO 14999-2:2005	Fringe	Zemax "Standard" Noll
Normalized	Yes	No	No	Yes
$Z_0^0(\rho,\theta)$	0	0	1	1
$Z_1^{-1}(\rho,\theta)$	1	2	3	3
$Z_1^1(\rho,\theta)$	2	1	2	2
$Z_2^{-2}(\rho,\theta)$	3	5	6	5
$Z_2^0(\rho,\theta)$	4	3	4	4
$Z_2^2(\rho,\theta)$	5	4	5	6
$Z_3^{-3}(\rho,\theta)$	6	10	11	9
$Z_3^{-1}(\rho,\theta)$	7	7	8	7
$Z_3^1(\rho,\theta)$	8	6	7	8
$Z_3^3(\rho,\theta)$	9	9	10	10
$Z_4^{-4}(\rho,\theta)$	10	17	18	15
$Z_4^{-2}(\rho,\theta)$	11	12	13	13
$Z_4^0(\rho,\theta)$	12	8	9	11
$Z_4^2(\rho,\theta)$	13	11	12	12
$Z_4^4(\rho,\theta)$	14	16	17	14

Note that, for example, Zernike spherical aberration is given by $Z_4^0(\rho, \theta)$. This same spherical aberration term is described by $Z_{12}(\rho, \theta)$, $Z_8(\rho, \theta)$, $Z_9(\rho, \theta)$ and $Z_{11}(\rho, \theta)$ in the various single-index definitions of the polynomials. The double-index scheme removes this ambiguity.

References

1. F. Zernike, "Beugungstheorie des Schneidenverfahrens und seiner verbesserten Form, der Phasekontrastmethode," *Physica* **1**, 689–704 (1934).

2. ANSI Z80.28-2004. Method for reporting optical aberrations of eyes.

3. ISO 24157:2008. Ophthalmic optics and instruments — Reporting aberrations of the human eye.

4. B. R. A. Nijboer, "The Diffraction Theory of Aberrations," Ph.D. Dissertation, University of Groningen, The Netherlands (1942).

5. J. Schwiegerling, "Scaling Zernike expansion coefficients to different pupil sizes," *J. Optical Society of America A* **19**, 1937–1945 (2002).

6. C. E. Campbell, "Matrix method to find a new set of Zernike coefficients from an original set when the aperture radius is changed," *J. Optical Society of America A* **20**, 209–217 (2003).

7. G.-M. Dai, "Scaling Zernike expansion coefficients to smaller pupil sizes: a simpler formula," *J. Optical Society of America A* **23**, 539–543 (2006).

8. A. J. E. M. Janssen and P. Dirksen, "Concise formula for the Zernike coefficients of scaled pupils," *J. Microlithography, Microfabrication, and Microsystems* **5**(3) 030501 (2006) [doi: 10.1117/1.2345672].

9. ISO 14999-2:2005. Optics and photonics — Interferometric measurement of optical elements and optical systems — Part 2: Measurement and evaluation techniques.

10. R. Noll, "Zernike polynomials and atmospheric turbulence," *J. Optical Society of America* **66**, 207–211 (1976).

Bibliography

J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley & Sons, New York (1978).

J. W. Goodman, *Introduction to Fourier Optics*, 3rd edition, Roberts & Company, Greenwood Village, CO (2005).

J. Sasián, *Introduction to Aberrations in Optical Imaging Systems*, Cambridge University Press, Cambridge (2013).

Table 3.1   Resolution (lp/mm) of the 1951 USAF resolution target.

	Group Number											
Element	−2	−1	0	1	2	3	4	5	6	7	8	9
1	0.250	0.500	1.00	2.00	4.00	8.00	16.00	32.00	64.00	128.0	256.0	512.0
2	0.281	0.561	1.12	2.24	4.49	8.98	17.96	35.92	71.84	143.7	287.4	574.7
3	0.315	0.630	1.26	2.52	5.04	10.08	20.16	40.32	80.63	161.3	322.5	645.1
4	0.354	0.707	1.41	2.83	5.66	11.31	22.63	45.25	90.51	181.0	362.0	—
5	0.397	0.794	1.59	3.17	6.35	12.70	25.40	50.80	101.6	203.2	406.4	—
6	0.445	0.891	1.78	3.56	7.13	14.25	28.51	57.02	114.0	228.1	456.1	—

used for evaluating the resolution limit as defined by the maximum resolvable spatial frequency in image space. The object space and image space spatial frequencies are related by

$$\text{image space spatial frequency} = \frac{\text{object space spatial frequency}}{|m|}, \quad (3.14)$$

where m is the system magnification. For example, the Group 1, Element 1 pattern has a spatial frequency of 2.00 lp/mm in object space. For an optical system with a magnification of −0.5, the image of this pattern has a spatial frequency of 4.00 lp/mm.

3.5 PSF and Wavefront-Based Metrics

3.5.1 Strehl ratio

The Airy pattern is the PSF for a diffraction-limited optical system. As aberrations are introduced into the optical system, the peak of the PSF will reduce, and the energy within the central portion of the PSF will be redistributed to cause a broadening of the PSF. The Strehl ratio is the ratio of the PSF irradiance value at the ideal image point of an aberrated optical system to the PSF irradiance value at the ideal image point for an equivalent diffraction-limited system. Based on Eq. (2.22),

$$\text{Strehl ratio} = \frac{1}{\pi^2} \left| \int_0^{2\pi} \int_0^1 \exp\left[\frac{i2\pi}{\lambda} W(h_x, h_y; \rho_x, \rho_y)\right] \rho \, d\rho \, d\theta \right|^2, \quad (3.15)$$

where (h_x, h_y) are the normalized field coordinates of the ideal image point. For example, consider the case where a quarter-wave of defocus is present in the system. In this case,

$$W(h_x, h_y; \rho_x, \rho_y) = \frac{\lambda}{4} \rho^2, \quad (3.16)$$

and the Strehl ratio = 0.81. Figure 3.12 shows a comparison of cross-sections through the PSFs of a diffraction-limited system and an F/10 system with a
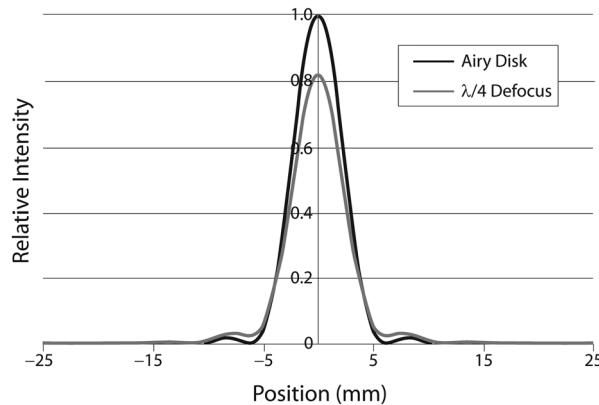
Figure 3.12 PSF comparison for a diffraction-limited system and an *F*/10 system with a quarter-wave of defocus.

quarter-wave of defocus and a wavelength of 0.5 μm. Based on Eq. (3.12), the resolution of this system is $1.22\lambda(f/\#) = 6.1$ μm, as can be seen in the figure as the location of the first zero of the Airy pattern. The peak of the defocused PSF is reduced to a relative value of 0.81, and the extra energy is distributed into the tails of the distribution. The Strehl ratio is a useful performance metric for well-corrected optical systems. The Maréchal criterion states that a system with a Strehl ratio greater than 0.8 is nearly indistinguishable from a diffraction-limited system. Care should also be taken when comparing the Strehl ratios for systems where significant aberrations occur. Off-axis, the peak of the aberrated PSF can shift relative to the peak of the Airy disk and/or multiple peaks can occur, leading to a poor metric of system quality.

3.5.2 Peak-to-valley, variance, and RMS wavefront error

The field of statistics defines a series of metrics to distill data into a small list of characteristics that describe the properties of the data. In this section, these metrics are applied to the wavefront error of an optical system to describe the performance of the system. In statistics, the simplest metric is most likely the range of the data being analyzed. The range is the difference between the maximum and minimum values of the data. When applied to the wavefront error of an optical system, the range is called the peak-to-valley error. The peak-to-valley (PV) error is the difference between the maximum and minimum wavefront error values in the exit pupil. For example, the maximum wavefront error for the case in Eq. (3.16) occurs at the edge of the pupil and equals $\lambda/4$. The minimum wavefront error value of zero occurs in the center of the pupil. The PV error for this example is $\lambda/4$. The PV error is a useful metric when the errors are distributed across the pupil. The Rayleigh criterion is approximately met when the PV error is less than $\lambda/4$. However,

this metric tends to underestimate system performance when the errors are localized to a small region of the pupil.

To better understand the shape of a distribution, the moments of the distribution are calculated. In regard to the wavefront error, the kth moment is given by

$$\langle w^k \rangle = \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 \left[W(h_x, h_y; \rho, \theta) \right]^k \rho d\rho d\theta, \tag{3.17}$$

where (h_x, h_y) are a fixed value of the normalized field coordinates. The first moment ($k = 1$) is simply the mean of the wavefront error. A positive mean suggests that on average the true wavefront is ahead of, or leads, the wavefront represented by the reference sphere, and the converse is true for a negative mean. The second moment ($k = 2$) is a measure of the spread of the wavefront error with regard to the reference sphere. An alternative and more common measure of this spread is the wavefront variance σ_w^2, defined as

$$\sigma_w^2 = \langle w^2 \rangle - \langle w \rangle^2 = \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 \left[W(h_x, h_y; \rho, \theta) \right]^2 \rho d\rho d\theta - \left[\frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 W(h_x, h_y; \rho, \theta) \rho d\rho d\theta \right]^2.$$

$$\tag{3.18}$$

For the wavefront error in Eq. (3.16), the wavefront variance is

$$\begin{aligned}
\sigma_w^2 &= \frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 \left[\frac{\lambda}{4} \rho^2 \right]^2 \rho d\rho d\theta - \left(\frac{1}{\pi} \int\limits_0^{2\pi} \int\limits_0^1 \frac{\lambda}{4} \rho^2 \rho d\rho d\theta \right)^2 \\
&= \frac{\lambda^2}{8} \int\limits_0^1 \rho^5 d\rho d\theta - \left(\frac{\lambda}{2} \int\limits_0^1 \rho^3 d\rho d\theta \right)^2 \\
&= \frac{\lambda^2}{48} - \frac{\lambda^2}{64} = \frac{\lambda^2}{192}.
\end{aligned} \tag{3.19}$$

Finally, a third alternative to measure the spread of the wavefront error is the *RMS wavefront error*, defined as

$$\text{RMS wavefront error} = \sqrt{\sigma_w^2}. \tag{3.20}$$

Based on the result in Eq. (3.19), the RMS wavefront error for a quarter-wave of defocus is

$$\text{RMS wavefront error} = \frac{\lambda}{8\sqrt{3}} \cong \frac{\lambda}{14}. \tag{3.21}$$

Based on a wavefront error of a quarter-wave of defocus, the metrics of a Strehl ratio of 0.8, a PV error of $\lambda/4$, and an RMS wavefront error of $\lambda/14$ are all equivalent and represent a nearly diffraction-limited system. When other

aberrations are present, these values diverge somewhat, but the preceding values for diffraction-limited performance are still used as a general rule.

3.5.3 Relationship to Zernike coefficients

The wavefront variance metric is closely related to the coefficients of a Zernike polynomial expansion of the wavefront error. As seen in Chapter 2, a wavefront error can be represented as

$$W(\rho,\theta) = \sum_{n,\,m} a_{nm} Z_n^m(\rho,\theta). \tag{3.22}$$

Note that the field coordinates have been suppressed here for clarity. In general, the wavefront error is dependent on field position. From Eq. (3.17), the mean wavefront error is given by

$$
\begin{aligned}
\langle w \rangle &= \frac{1}{\pi} \int\limits_0^{2\pi}\int\limits_0^1 \sum_{n,\,m} a_{nm} Z_n^m(\rho,\theta)\rho d\rho d\theta \\
&= \frac{1}{\pi} \sum_{n,\,m} a_{nm} N_n^m \int\limits_0^1 R_n^m(\rho)\rho d\rho \int\limits_0^{2\pi} \Theta_m(\theta)d\theta,
\end{aligned}
\tag{3.23}
$$

where the various elements were previously defined in Eqs. (2.85), (2.95), and (2.96). The final integral over θ will always be zero for $m \neq 0$. For the case where $m = 0$, the following relationship can be used for evaluating the mean wavefront error:

$$\int\limits_0^1 R_n^m(\rho)\rho d\rho = \frac{\delta_{m0}}{2}. \tag{3.24}$$

Combining these results shows that the mean wavefront error is given by $\langle w \rangle = a_{0,\,0}$. In other words, the Zernike piston coefficient describes the mean wavefront error.

The second moment of the wavefront error is

$$
\begin{aligned}
\langle w^2 \rangle &= \frac{1}{\pi} \int\limits_0^{2\pi}\int\limits_0^1 \left[\sum_{n,\,m} a_{nm} Z_n^m(\rho,\theta) \right] \left[\sum_{n',\,m'} a_{n'm'} Z_{n'}^{m'}(\rho,\theta) \right] \rho d\rho d\theta \\
&= \frac{1}{\pi} \sum_{n,\,m} a_{nm} \sum_{n',\,m'} a_{n'm'} \int\limits_0^{2\pi}\int\limits_0^1 Z_n^m(\rho) Z_{n'}^{m'}(\rho)\rho d\rho d\theta.
\end{aligned}
\tag{3.25}
$$

The double integral in Eq. (3.25) is exactly the orthogonality condition defined in Eq. (2.98). Using this result, the second moment is given by

$$\langle w^2 \rangle = \frac{1}{\pi} \sum_{n,\,m} a_{nm} \sum_{n',\,m'} a_{n'm'} \pi \delta_{nn'} \delta_{mm'} = \sum_{n,\,m} a_{nm}^2. \tag{3.26}$$

Finally, the wavefront variance is

$$\sigma_w^2 = \sum_{n,\,m} a_{nm}^2 - a_{0,\,0}^2 = \sum_{n\geq 1,\,m} a_{nm}^2. \tag{3.27}$$

Equation (3.27) illustrates the advantage of representing the wavefront error in terms of the orthogonal Zernike polynomials (along with the normalization factor N_n^m). The wavefront variance is simply the sum of the squares of the expansion coefficients, excluding the Zernike piston term. This means that the expansion coefficients with the largest magnitudes are the main contributors to the wavefront variance and are the main degraders of system performance. Since each term in the sum in Eq. (3.27) is positive, reducing the effects of one of the Zernike terms necessarily improves the system performance. Thus, the Zernike expansion provides insight into the relationship of the various aberrations found in an optical system and provides strategies for reducing their effects.

3.5.4 Relationship to Strehl ratio

The wavefront variance can also be related to the Strehl ratio, thus relating variations in the shape of the wavefront in the exit pupil to the broadening of the PSF in the image plane. Starting with the definition of the Strehl ratio in Eq. (3.15), the complex exponential is expanded in a power series, so that

$$\text{Strehl ratio} = \frac{1}{\pi^2} \left| \int_0^{2\pi} \int_0^1 \left(1 + \frac{i2\pi}{\lambda} W(\rho,\theta) - \frac{1}{2}\left(\frac{2\pi}{\lambda}\right)^2 W^2(\rho,\theta) + \dots \right) \rho d\rho d\theta \right|^2,$$

$$\tag{3.28}$$

where, again, the field coordinates are implicitly understood to be part of the wavefront error definition. Keeping only the first three terms of the series and comparing the integrals to the definitions of the moments of the wavefront error leads to

$$\text{Strehl ratio} \cong \frac{1}{\pi^2} \left| \pi + \frac{i2\pi^2}{\lambda}\langle w\rangle - \frac{\pi}{2}\left(\frac{2\pi}{\lambda}\right)^2 \langle w^2\rangle \right|^2. \tag{3.29}$$

Finally, carrying out the squared modulus and assuming that the square of the second moment is negligible gives

$$\text{Strehl ratio} \cong 1 - \left(\frac{2\pi}{\lambda}\right)^2 \sigma_w^2. \tag{3.30}$$

The approximation in Eq. (3.30) becomes increasingly inaccurate for wavefront variances $\sigma_w^2 \geq 0.0075 \ \mu m^2$. An improved approximation can be achieved by noting that the approximation is consistent with the first two

terms of the series expansion of the exponential function. Consequently, the improved Strehl ratio approximation is given by

$$\text{Strehl ratio} \cong \exp\left[-\left(\frac{2\pi}{\lambda}\right)^2 \sigma_w^2\right]. \tag{3.31}$$

3.5.5 Encircled and ensquared energy

The encircled energy (EE) is yet another metric of system performance based on the PSF. It describes the normalized amount of energy contained within a circular region of the PSF. As the diameter of the circular region increases, an increasing amount of energy is contained within the region. The EE therefore monotonically increases and approaches unity for large diameters. The more quickly the EE approaches unity, the more energy is concentrated within the central portion of the PSF. Based on Eq. (2.22), the EE is given by

$$EE(r_{img}) = \left[|E_0|^2 \frac{\pi D_{E'}^2}{4}\right]^{-1} \int_0^{2\pi} \int_0^{r_{img}} I(x'_{img}, y'_{img}) \ r'_{img} dr'_{img} d\theta'_{img}, \tag{3.32}$$

where $r'_{img} = \sqrt{x'^2_{img} + y'^2_{img}}$ and $\theta'_{img} = \tan^{-1}(y'_{img}/x'_{img})$. For a diffraction-limited system, the EE reduces to

$$EE(r_{img}) = 1 - J_0^2\left[\frac{\pi r_{img}}{\lambda(f/\#)}\right] - J_1^2\left[\frac{\pi r_{img}}{\lambda(f/\#)}\right], \tag{3.33}$$

where $J_0()$ and $J_1()$ are Bessel functions of the first kind. Ensquared energy is a concept similar to EE except that the domain of integration is a square region instead of a circular region. Ensquared energy describes the relative amount of energy found within a square region and is sometimes useful for evaluating systems where the image is sampled with square pixels, as is the case with CCD or CMOS sensors.

3.5.6 Example of optical quality metrics

The first-order properties of a cemented doublet were analyzed in Chapter 1. Table 1.2 provides the details of the dimensions and materials for this lens along with the paraxial raytrace details. Here, the metrics outlined above will be applied to this lens to further understand its optical performance. Figure 3.13 compares the PSFs of the lens for two different field angles with the Airy pattern. Figure 3.13(a) shows the Airy pattern for a diffraction-limited system with an entrance pupil diameter of 6 mm at a wavelength of 880 nm. Figure 3.13(b) shows the PSF of the example lens for an equivalent pupil diameter and wavelength for an on-axis object point. Clearly, the performance of the lens is nearly diffraction limited on axis. Figure 3.13(c) shows the PSF of the lens when the object point is moved to a field angle of 2 deg. The PSF shows significant astigmatism and some coma at this modest