Wavefront fitting with discrete orthogonal polynomials in a unit radius circle

Daniel Malacara, FELLOW SPIE
The Institute of Optics
University of Rochester
Rochester, New York 14627

J. Martin Carpio-Valadéz
J. Javier Sánchez-Mondragón
Centro de Investigaciones en Optica, A. C.
Apdo. Post. No. 948
Leön, Gto., México 37000

Abstract. Zernike polynomials have been used for some time to fit wavefront deformation measurements to a two-dimensional polynomial. Their orthogonality properties make them ideal for this kind of application. The typical procedure consists of first obtaining the fitting using x-y polynomials and then transforming them to Zernike polynomials by means of a matrix multiplication. Here, we present a new method for making this fitting faster by using a set of orthogonal polynomials on a discrete base of data points on a unitary circle.

Subject terms: optical testing; interferometry; polynomials.

Optical Engineering 29(6), 672-675 (June 1990).

CONTENTS

1. INTRODUCTION

Zernike polynomials[1] are ideal for fitting the measured data points in an interferogram wavefront to a two-dimensional polynomial, due to their orthogonality properties.[2] In order to do this, the common procedure[3–5] is to fit the measured points to an x-y polynomial and then transform it to a linear combination of Zernike polynomials by means of a linear transformation, performing a matrix multiplication.

The disadvantages of this method are as follows: (a) We go through an unnecessary intermediate step, increasing the number of arithmetic operations and therefore the round-off error. (b) The x-y polynomial is orthogonal over a unit square and not a unit circle. Most lens apertures are circular; thus, a square domain for the polynomial is not the logical choice.[6] (c) Alone, each Zernike polynomial does not represent a best least squares approximation to the measured data points, because they are not orthogonal on this discrete base. This latter fact has many disadvantages, as we show in this paper. Here, we describe a method to perform a least squares fitting using orthogonal polynomials over a discrete set of data points on a circle.

2. WAVEFRONT REPRESENTATION

Let us start by defining the real (not complex) Zernike polynomials, following Rimmer and Wyant,[4] as

$$U_n^l(\rho,\theta) = \begin{cases} R_n^l(\rho)\cos l\theta\ , & \text{for } l \leqslant 0\ , \\ R_n^l(\rho)\sin l\theta\ , & \text{for } l > 0\ , \end{cases} \tag{1}$$

where l is any positive or negative integer and n is any integer greater than or equal to zero. The numbers n and l are both even or both odd. The polynomials $R_n^l(\rho)$ are the radial components of the Zernike polynomials. The polynomials $U_n^l(\rho,\theta)$ satisfy the orthogonality condition

$$\int_0^1\int_0^{2\pi} U_n^l(\rho,\theta)U_{n'}^{l'}(\rho,\theta)\rho d\rho d\theta = \frac{\pi}{2(n+1)}\delta_{nn'}\delta_{ll'}\ , \tag{2}$$

and their radial components $R_n^l(\rho)$ also satisfy the following orthogonality condition:

$$\int_0^1 R_n^l(\rho)R_{n'}^l(\rho)\rho d\rho = \frac{\pi}{2(n+1)}\delta_{nn'}\ . \tag{3}$$

The wavefront $W(\rho,\theta)$ may then be represented by

$$W(\rho,\theta) = \sum_{n=0}^{K}\sum_{l=-n}^{n} A_{nl}U_n^l = \sum_{n=0}^{K}\sum_{l=-n}^{n} A_{nl}R_n^l\begin{bmatrix}\sin\\ \\ \cos\end{bmatrix}l\theta\ , \tag{4}$$

where K is the degree of the fitted polynomial. The sine function is used when $l > 0$ and the cosine function is used when $l \leqslant 0$. Now, if a positive number m is defined as

$$m = \frac{n-l}{2}\ , \tag{5}$$

the wavefront $W(\rho,\theta)$ may be written

$$W(\rho,\theta) = \sum_{n=0}^{K}\sum_{m=0}^{n} A_{nm}R_n^{n-2m}\begin{bmatrix}\sin\\ \\ \cos\end{bmatrix}(n-2m)\theta\ . \tag{6}$$

As explained before, the most common method to make the data fitting to the measured points in an interferogram is to use

---

Paper 2784 received Aug. 20, 1989; revised manuscript received Feb. 2, 1990; accepted for publication Feb. 8, 1990. This paper is a revision of a paper presented at the 14th Congress of the International Commission for Optics, Aug. 24–28, 1987, Quebec, Canada. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 813.

an x-y polynomial for this fitting and then convert the result to a linear combination of Zernike polynomials by means of a matrix multiplication, as explained by Malacara et al.[5]

During the least squares procedure, orthogonal x-y polynomials are generated to eliminate the need for the inversion of a matrix. The Zernike polynomials combination is obtained by a linear transformation using a well-known transformation matrix.

The problem with this method is that the x-y data fitting was made with a square domain, while the points are on a circular aperture. This method is slower and has more round-off errors than the method described here. Some errors frequently appear near the corners of the square domain.

Zernike polynomials would be ideal because of their circular symmetry but are not used because they are orthogonal only on the continuous unit circle domain, not over the discrete and finite number of data points.

3. GENERATION OF ORTHOGONAL POLYNOMIALS ON A DISCRETE SET OF DATA POINTS

Here, we present a data fitting procedure that generates and uses some polynomials that are orthogonal on the finite and discrete set of irregularly distributed data points. Let us write the wavefront expression in Eq. (4) with a single index j, instead of the two n and m indices, defined by

$$j = \frac{n(n+1)}{2} + m + 1 \ . \tag{7}$$

The maximum value of j is the total number of Zernike polynomials used in the fit and given by $L = (K+1)(K+2)/2$, where K is the degree of the polynomial. Now we have four different indices that change as shown in Table I. This is the ordering of the polynomials found in many sources, but, of course, any other ordering might be used. Thus, the Zernike polynomials $U_n^m(\rho,\theta)$ may be considered to have a single integer and positive index j, as follows:

$$W(\rho,\theta) = \sum_{j=1}^{L} A_j U_j(\rho,\theta) \ . \tag{8}$$

As we have pointed out, the Zernike polynomials $U_j(\rho,\theta)$ are orthogonal on the continuous unit radius circle, but they are not orthogonal on the discrete set of data points. Now, we are going to find a set of orthogonal polynomials defined in a discrete set of data points within the circle, represented by $V_j(\rho,\theta)$. We would expect that these polynomials will approach the Zernike polynomials when the number of points are uniformly distributed and tend to infinity. Thus, the wavefront may be represented by

$$W(\rho,\theta) = \sum_{j=1}^{L} B_j V_j(\rho,\theta) \ , \tag{9}$$

where these polynomials $V_j(\rho,\theta)$ satisfy the discrete orthogonality condition on the set of N data points with coordinates (ρ_i,θ_i):

$$\sum_{i=1}^{N} V_j(\rho_i,\theta_i) V_p(\rho_i,\theta_i) = F\delta_{jp} \ , \tag{10}$$

where F is a factor whose value depends on the data points.

TABLE I. Values of indices *n, l, m,* and *j* for some Zernike polynomials.

n	*l*	*m*	*j*	Zernike polynomial	Meaning
0	0	0	1	1	Constant term
1	0	1	2	$\rho \sin\theta$	Tilt in x direction
1	1	−1	3	$\rho \cos\theta$	Tilt in y direction
2	0	2	4	$\rho^2 \sin 2\theta$	Astigmatism, axis at 45°
2	1	0	5	$2\rho^2 - 1$	Focus shift
2	2	−2	6	$\rho^2 \cos 2\theta$	Astigmatism, axis at 0° or 90°
3	0	3	7	$\rho^3 \sin 3\theta$	
3	1	1	8	$(3\rho^3 - 2\rho) \sin\theta$	Third order coma, x axis
3	2	−1	9	$(3\rho^3 - 2\rho) \cos\theta$	Third order coma, y axis
3	3	−3	10	$\rho^3 \cos 3\theta$	
4	0	4	11	$\rho^4 \sin 4\theta$	
4	1	2	12	$(4\rho^4 - 3\rho^2) \sin 2\theta$	
4	2	0	13	$6\rho^4 - 6\rho^2 + 1$	Third order spherical aberration
4	3	−2	14	$(4\rho^4 - 3\rho^2) \cos 2\theta$	
4	4	−4	15	$\rho^4 \cos 4\theta$	

As usual, a Gram-Schmidt orthogonalization is carried out in order to find these polynomials. We begin by writing

$$\begin{aligned} V_1 &= U_1 \ , \\ V_2 &= U_2 + D_{21}V_1 \ , \\ V_3 &= U_3 + D_{31}V_1 + D_{32}V_2 \ , \\ &\vdots \\ V_j &= U_j + D_{j1}V_1 + D_{j2}V_2 + \ldots + D_{j,j-1}V_{j-1} \ , \end{aligned} \tag{11}$$

or in a general way,

$$V_j = U_j + \sum_{s=1}^{j-1} D_{js}V_s \ , \tag{12}$$

where $j = 1,2,3,\ldots,L$. Since $V_j(\rho,\theta)$ has to be orthogonal with $V_p(\rho,\theta)$, we multiply this expression by V_p and then sum for all data points, from $i = 1$ to N. If the orthogonality condition shown in Eq. (10) is then used, we obtain for values of j different from p

$$\sum_{i=1}^{N} V_jV_p = \sum_{i=1}^{N} U_jV_p + D_{jp}\sum_{i=1}^{N} V_p^2 = 0 \ . \tag{13}$$

Then, D_{jp} may be written

$$D_{jp} = \frac{\sum_{i=1}^{N} U_jV_p}{\sum_{i=1}^{N} V_p^2} \ , \tag{14}$$

with $j = 2,3,4,\ldots,L$ and $p = 1,2,\ldots,j-1$.

4. LEAST SQUARES FITTING

Now that we have defined the method to find the orthogonal polynomials on the discrete set of data points, let us consider the usual procedure to make the least squares fitting. We start by defining a quantity S, which is directly proportional to the square of the standard deviation, or variance, as follows:

$$S = \sum_{i=1}^{N}[W_i' - W(\rho_i,\theta_i)]^2 , \tag{15}$$

where W_i' is the measured wavefront deviation for data point i and $W(\rho_i,\theta_i)$ is the computed wavefront deviation after the polynomial fitting. If the wavefront is represented by a linear combination of the polynomials $V(\rho,\theta)$, we may represent the wavefront by

$$W(\rho_i,\theta_i) = \sum_{j=1}^{L} B_j V_j(\rho_i,\theta_i) . \tag{16}$$

Now, the following conditions are imposed in order to make S as small as possible:

$$\frac{\partial S}{\partial B_p} = 0 , \tag{17}$$

for $p = 1,2,3,\ldots,L$. Thus, the following system of L equations with L unknowns is obtained:

$$\sum_{j=1}^{L} B_j \sum_{i=1}^{N} V_j V_p - \sum_{i=1}^{N} W_i' V_p = 0 , \tag{18}$$

where $p = 1,2,3,\ldots,L$. Now, using the orthogonality condition, we obtain

$$B_p = \frac{\sum_{i=1}^{N} W_i' V_p}{\sum_{i=1}^{N} V_p^2} . \tag{19}$$

We now have the values of the coefficients that define the linear combination of the polynomials $V_j(\rho,\theta)$.

5. COMPUTING WAVEFRONT COEFFICIENTS

The next step is to determine the values of the coefficients C_{js} that define the orthogonal polynomials on the discrete set of measured data points V_j as a linear combination of the Zernike polynomials U_j. We begin by writing

$$\begin{aligned} V_1 &= U_1 , \\ V_2 &= U_2 + C_{21}U_1 , \\ V_3 &= U_3 + C_{31}U_1 + C_{32}U_2 , \\ &\vdots \\ V_j &= U_j + C_{j1}U_1 + C_{j2}U_2 + \ldots + C_{j,j-1}U_{j-1} , \end{aligned} \tag{20}$$

or in general,

$$V_j = U_j + \sum_{i=1}^{j-1} C_{ji}U_i , \tag{21}$$

where $j = 2,3,\ldots,L$, $C_{jj} = 1$, and $V_1 = U_1$. We may now find a few coefficients C_{ji} as follows:

$$\begin{aligned} C_{21} &= D_{21} , \\ C_{31} &= D_{32}C_{21} + D_{31} , \\ C_{32} &= D_{32} , \\ C_{41} &= D_{43}C_{31} + D_{42}C_{21} + D_{41} , \\ C_{42} &= D_{43}C_{32} + D_{42} , \\ C_{43} &= D_{43} , \end{aligned} \tag{22}$$

etc.

These results may be written in a general form as

$$C_{ji} = \sum_{s=1}^{j-i} D_{j,j-s}C_{j-s,i} , \tag{23}$$

where $i = 1,2,3,4,\ldots,j-1$ and $C_{jj} = 1$.

Since we now know the coefficients B_j and C_{ji}, the coefficients A_j necessary to find the wavefront $W(\rho,\theta)$ in Eq. (8) may be found if we substitute Eq. (21) into Eq. (16), obtaining

$$W(\rho,\theta) = B_1U_1 + \sum_{j=2}^{L} B_j\left(U_j + \sum_{i=1}^{j-1} C_{ji}U_i\right) , \tag{24}$$

where the coefficients C_{ji} are given in Eq. (23). Then, by rearranging the order of the terms in the sums we may find

$$W(\rho,\theta) = \sum_{j=1}^{L-1}\left(B_j + \sum_{i=j+1}^{L} B_iC_{ij}\right)U_j + B_LU_L . \tag{25}$$

Comparing this expression with Eq. (8) we may see that the coefficients A_j are given by

$$A_j = B_j + \sum_{i=j+1}^{L} B_iC_{ij} , \tag{26}$$

with $j = 1,2,3,\ldots,(L-1)$ and $A_L = B_L$.

6. SUMMARY

The procedure just presented for the fitting of some measured wavefront deviations to a linear combination of orthogonal polynomials on the discrete set of data points may be summarized as follows:

(1) Choose a polynomial degree K and then arrange in order the first $L = (K + 1)(K + 2)/2$ Zernike polynomials, assigning them the index $j = n(n + 1)/2 + m + 1$ (see Table I).

(2) Compute the Zernike polynomials U_j at each of the N measured data points on the wavefront.

(3) Making $j = 2,3,\ldots,L$ and $p = 1,2,3,\ldots,(j-1)$, use Eqs. (14) and (12) alternately until all coefficients D_{jp} and the polynomials V_j are computed.

(4) By means of Eq. (19), with $p = 1,2,3,\ldots,L$ and the measured values W_i' of the wavefront deviations, calculate all coefficients B_p.

(5) Apply Eq. (24), with $i = 1,2,3,4,\ldots,(j-1)$ and $C_{jj} = 1$, in order to calculate all coefficients C_{ji}.

(6) Use Eq. (26), with $j = 1,2,3,..,(L-1)$ and $A_L = B_L$, to calculate all aberration coefficients A_j.

(7) Finally, use Eq. (8) to compute the polynomial fitted wavefront deviations.

7. CONCLUSIONS

When the number of measured data points is very large and they are evenly distributed over the aperture, the coefficients C_{ji} tend to a zero value, making the polynomials V_j equal to the Zernike polynomials U_j. For the same reason, if the Zernike polynomials U_j are used to make the fitting, instead of the polynomials V_j, the matrix of the system of Eq. (18) will not be diagonal; rather, it will approach this condition when the number of data points is very large. Since the round-off errors when inverting a matrix are small, if this matrix is nearly diagonal, then this approach could also be taken.

If the number of data points is not very large and they are evenly distributed, the Zernike polynomials U_j are not orthogonal over the measured data points. Thus, individually they do not represent a best fit to the data. However, the polynomials V_j are orthogonal over the data points, and therefore they individually represent a best fit to the wavefront. Hence, it is advantageous to consider the aberrations in terms of the coefficients B_j instead of the coefficients A_j. In other words, if the Zernike polynomials are used, we have to make the complete least squares fit procedure any time an aberration is added or subtracted, because they are not orthogonal. However, the aberrations V_j are orthogonal, and therefore any term may be added or subtracted in the results without losing the least squares fit.

The procedure presented here is faster and more accurate than the traditional method passing through an intermediate x-y orthogonal polynomial. However, the main advantage is not this, but the fact that the aberrations V_j have the same physical meaning as the Zernike polynomials, but with the great advantage of being orthogonal over the discrete set of measured points.

8. REFERENCES

1. F. Zernike, "Beugungstheorie des Schneidenver-Eahrens und Seiner Verbesserten Form, der Phasenkontrastmethode," Physica 1, 689 (1934).
2. M. Born and E. Wolf, *Principles of Optics,* p. 464, Pergamon Press, New York (1964).
3. M. P. Rimmer, "Polynomial fitting of interferograms," ITEK Tech. Rept., Appendix A (1972).
4. M. P. Rimmer and J. C. Wyant, "Evaluation of large aberrations using a lateral shear interferometer having variable shear," Appl. Opt. 14, 142 (1975).
5. D. Malacara, A. Cornejo, and A. Morales, "Computation of orthogonal polynomials in optical testing," Bol. Inst. Tonantzintla 2, 121 (1976).
6. D. Malacara, J. M. Carpio-Valadéz, and J. J. Sánchez-Mondragón, "Interferometric data fitting on Zernike like orthogonal basis," in *14th Congress of the International Commission for Optics,* H. H. Arsenault, ed., Proc. SPIE 813, 35 (1987).