Zernike-based matrix model of deformable mirrors: optimization of aperture size

Javier Alda and Glenn D. Boreman

The actuator influence functions of a typical deformable mirror are expanded in a Zernike polynomial decomposition. This expansion is then extended to a matrix formalism that describes the modal operation of the mirror. The size of the aperture over which the Zernikes are defined affects the accuracy of the expansion. The optimum size of this aperture is found by minimizing the variance of the wave-front error.

Key words: Deformable mirror, adaptive optics, Zernike polynomials, modal driving, zonal driving.

Introduction

Zernike polynomials have been used widely for describing optical wave fronts.[1,2] They are used when the aperture of an optical instrument shows a circular symmetry.[3] They are also directly related to wavelength-independent optical aberrations.

An orthogonal decomposition of the incoming wave front is a powerful tool for analyzing the performance and operation of deformable mirrors (DM's).[4–6] DM's can be operated in either a zonal or a modal fashion.[7] When they are operated in the zonal way, the surface of the mirror and the actuator stroke correct the topography of the incoming wave front. Modal driving is conceptually more involved. First, it is necessary to choose a proper set of polynomials to expand the mirror and the wave-front functions. Then the wave front is decomposed in terms of the polynomial set. Finally the mirror is driven so that the displacements of the actuators resemble the desired polynomial expansion.

In this paper we investigate the expansion of each of the actuator influence functions into a set of Zernike polynomial functions. This decomposition permits us to calculate the fitting between the Zernike polynomials and the surface of the DM. The size of the aperture over which the Zernikes are defined affects the accuracy of the expansion. The optimum size of this aperture is found by minimizing the variance of the wave-front error.

To treat a realistic case, we performed our calculations for a DM with 21 actuators arranged in a square grid as shown in Fig. 1.

Matrix Representation of Modal Driving

Two alternative approaches represent the operation of adaptive-optics systems. Zonal driving[8] has a simple matrix representation and permits in most cases a one-dimensional treatment because of the separability of the problem.[9] Modal driving describes the operation of the mirror in terms of functions related to the optical aberrations. In addition the orthogonality properties of the chosen basis functions also simplify the calculation required for the treatment of a two-dimensional (2-D) problem.

The surface of a DM with J actuators located in a 2-D grid can be described by a linear superposition of the influence functions of the mirror:

$$\omega_m(x, y) = \sum_{j=1}^{J} m_j(x, y) a_j, \tag{1}$$

where a_j is the signal that drives the actuator j whose influence function is $m_j(x, y)$ and ω_m is the surface displacement of the DM.

If the wave front is sampled at discrete locations, the previous continuous functions for ω_j and m_j also become discrete and the operation of the mirror can be described by means of a matrix expression

$$\omega_m = \mathscr{M}\mathbf{a}, \tag{2}$$

where $\mathbf{a}$ represents the vector containing the driving signal for each actuator, and $\mathscr{M}$ is the matrix contain-

J. Alda is with the Department of Optics, Faculty of Physics, Complutense University of Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain. G. D. Boreman is with the Department of Electrical Engineering, Center for Research in Electro-Optics and Lasers, University of Central Florida, Orlando, Florida 32816.

Received 31 January 1992.
0003-6935/93/132431-08$05.00/0.

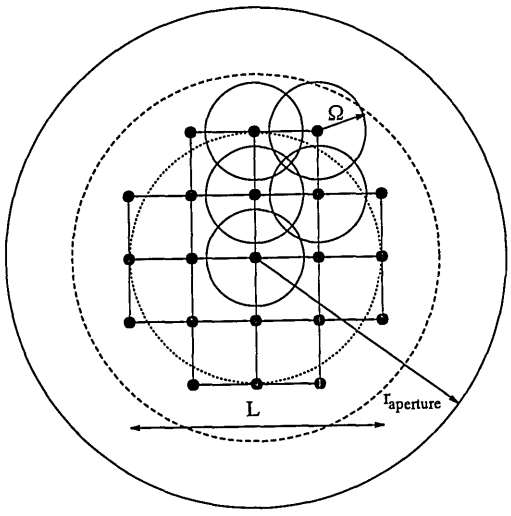

Fig. 1. Arrangement of the 21-actuator deformable mirror. The total radius of the mirror is approximately four times the distance between actuators. We have also represented the Gaussian size of the influence function Ω (where the influence function reaches a value of $1/e$ the maximum value). Three different aperture sizes have been plotted: solid curve, $s = 1.0$; dashed curve, $s = 1.5$; and dotted curve, $s = 2.0$.

ing the J influence functions sampled along a 2-D grid of N sampling points. Because the problem of interest is two dimensional, let us represent Eq. (2) as shown in Fig. 2, where we have assumed a circular aperture. Each 2-D vertical slice of the $\mathcal{H}$ matrix contains one influence function sampled along the 2-D grid. They are stacked to form the three-dimensional array representation of the J influence functions.

The usual rules of matrix multiplication apply here when the one-dimensional vector $\mathbf{a}$ multiplies the three-dimensional array $\mathcal{H}$ to obtain the 2-D slice ω_m. Actually this operation can be reduced to the conventional matrix operation if the elements on the slices are arranged as a long column vector. However, our representation preserves the geometry of the problem and allows clearer physical insight.

The wave-front error ϵ is a measure of how good the fit is between the incoming wave front ω and the DM

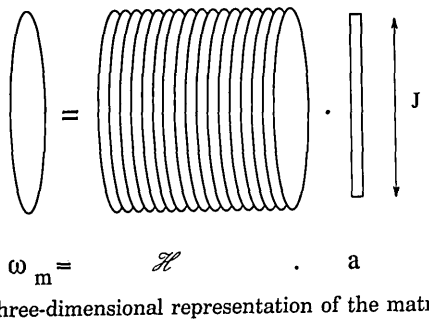

Fig. 2. Three-dimensional representation of the matrix relation $\omega_m = \mathcal{H}\mathbf{a}$. Each slice of the matrix $\mathcal{H}$ corresponds to the influence function of each actuator. Therefore the whole stack contains J slices.

surface ω_m:

$$\epsilon = \omega - \omega_m. \tag{3}$$

The total squared error over the DM aperture provides a convenient summary measure of the correction achieved by the system. When both the wave front ω and the surface of the mirror ω_m are sampled, they become discrete arrays. Following a standard notation in adaptive optics,[9] we define the total squared error as a wave-front error variance under the assumption that the mean wave-front error is zero:

$$\sigma_{wf}{}^2 = \epsilon^t\epsilon, \tag{4}$$

where superscript t denotes the transpose operation.

We can use Eq. (2) to find the driving signal that best fits an incoming wave front ω by minimizing the variance of the wave-front error[9,10]:

$$\mathbf{a} = (\mathcal{H}^t\mathcal{H})^{-1}\mathcal{H}^t\omega, \tag{5}$$

where superindex t denotes the transpose operation. Therefore Eq. (2) becomes

$$\omega_m = \mathcal{H}(\mathcal{H}^t\mathcal{H})^{-1}\mathcal{H}^t\omega. \tag{6}$$

This problem has been treated, and a transfer function representation has been found by Lawrence and Moore.[9,10] In Eq. (6) the product $\mathcal{H}^t\mathcal{H}$ is the interaction matrix between the individual influence functions described by Tyson.[6] Therefore, when the influence functions are orthogonal to each other, this interaction matrix is equal to the unity matrix, $\mathcal{H}^t\mathcal{H} = (\mathcal{H}^t\mathcal{H})^{-1} = \mathcal{J}$, and Eq. (6) transforms into

$$\omega_m = \mathcal{H}'\mathcal{H}'^t\omega, \tag{7}$$

where $\mathcal{H}'$ is the matrix of the orthonormalized influence functions.

The numerical evaluation of Eqs. (6) and (7) requires many arithmetic operations when a grid with a large number of sampling points is used, since the matrix products involved in those equations contain $N \times N$ elements. However, this matrix approach is well suited to describing the zonal driving mode.

For modal driving the Zernike polynomials are often used to expand the incoming wave front ω and also to represent the whole surface of the mirror.[6] We will use the same functions to decompose the individual influence functions for each actuator. If we assume a finite family of Zernike polynomials with K elements, the decomposition can be approximately described as

$$m_j(x, y) = \sum_{k=0}^{K-1} c_{jk}Z_k(x, y), \tag{8}$$

where the functional dependence on x and y requires a change of variables from the usual polar coordinates. The subscript k denotes the index number of the Zernike polynomial for a given numbering scheme of

the family.[1,2,4] Even for influence functions whose functional forms are identical, each influence function will have a different representation in terms of the Zernikes because each actuator has a different location with respect to the symmetry of the Zernike set. The coefficients of the expansion are given by

$$c_{jk} = \frac{1}{\pi}\int_S m_j(x, y)Z_k(x, y)\mathrm{d}x\mathrm{d}y, \tag{9}$$

where the area of integration is the circular aperture S over which the Zernikes are defined.

The matrix version of Eq. (8) is

$$\mathscr{H} = Z\mathscr{C}, \tag{10}$$

where Z is the array obtained by stacking the K slices corresponding to the Zernike set. The matrix $\mathscr{C}$ contains $J \times K$ elements that describe the decomposition of the influence functions. Figure 3 is a graphical representation of Eq. (10) and its transpose version $\mathscr{H}^t = \mathscr{C}^t Z^t$. If the decomposition in Eq. (10) is used, Eq. (2) becomes

$$\omega_m = Z\mathscr{C}\mathbf{a}. \tag{11}$$

The input wave front can also be represented as a Zernike decomposition in the following matrix form:

$$\omega = Z\mathbf{I}, \tag{12}$$

where $\mathbf{I}$ is the vector containing the coefficients of the decomposition that can be obtained in the same way

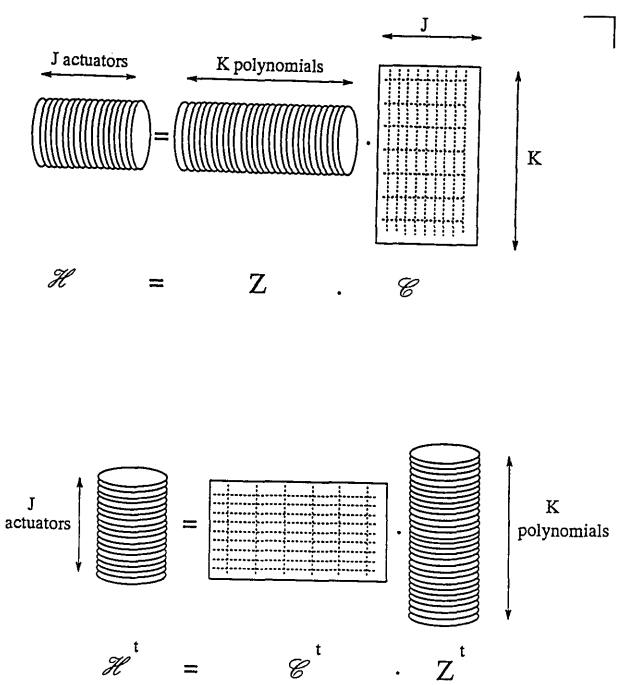

Fig. 3. Three-dimensional representation of the modal decomposition $\mathscr{H} = Z\mathscr{C}$. Two stacks of a different number of slices are related by means of the matrix $\mathscr{C}$. The transposed version of this matrix relation is also represented.

as was done for the coefficients of the influence functions [see Eq. (9)].

If Eqs. (10) and (12) are substituted into the matrix relation (6), we obtain the following expression:

$$\omega_m = Z\mathscr{C}(\mathscr{C}^t Z^t Z\mathscr{C})^{-1}\mathscr{C}^t Z^t Z\mathbf{I} = Z\mathscr{C}(\mathscr{C}^t\mathscr{C})^{-1}\mathscr{C}^t\mathbf{I}, \tag{13}$$

where the product Z^tZ is equal to the unity matrix $\mathscr{I}$ because of the orthonormality relations of the Zernike polynomial set. After we apply this property and use Eq. (11), the following relation gives the driving signal for each actuator as a function of the modal decomposition of the incoming wave front:

$$\mathbf{a} = (\mathscr{C}^t\mathscr{C})^{-1}\mathscr{C}^t\mathbf{I}. \tag{14}$$

Equation (14) is the modal version of the zonal relation given by Eq. (5). The use of one or the other depends on the type of driving, zonal or modal, that we choose for the operation of the mirror.

We can combine Eqs. (5) and (14) to find the modal decomposition in terms of the zonal map of the incoming wave front ω in the following form:

$$\mathbf{I} = \mathscr{C}(\mathscr{H}^t\mathscr{H})^{-1}\mathscr{H}^t\omega. \tag{15}$$

When Eq. (14) is inverted, the modal composition that describes the surface of the mirror for a given actuator driving signal $\mathbf{a}$ is given by

$$\mathbf{I} = \mathscr{C}\mathbf{a}. \tag{16}$$

By using the decomposition of the wave front and the surface of the mirror in terms of the Zernike polynomials [see Eqs. (11), (12), and (14)], the residual variance of the wave-front error can be expressed as

$$\sigma_{wf(\mathrm{modal})}{}^2 = \mathbf{I}^t[\mathscr{I} - \mathscr{C}(\mathscr{C}^t\mathscr{C})^{-1}\mathscr{C}^t]\mathbf{I}, \tag{17}$$

which corresponds with the zonal version of the variance expressed in terms of the matrix of influence functions $\mathscr{H}$ and the incoming wave front ω:

$$\sigma_{wf(\mathrm{zonal})}{}^2 = \omega^t[\mathscr{I} - \mathscr{H}(\mathscr{H}^t\mathscr{H})^{-1}\mathscr{H}^t]\omega. \tag{18}$$

Choice of the Optimum Aperture Size

Equations (8) and (9) show how any influence function can be expanded in terms of the Zernike set. The representation of the influence functions will be more accurate when more polynomials are used in the summation (see Fig. 4). However, for a fixed number of polynomials it is more difficult to represent a spatially narrow function than a spatially broad function. If the surface of the DM contains many actuators, the influence function will be narrow with respect to the total surface of the mirror. The expansion in terms of the Zernikes will then need a large number of terms to improve the accuracy of the restored influence function. However, when the DM has few actuators, the influence function will be broader relative to the whole mirror and can be

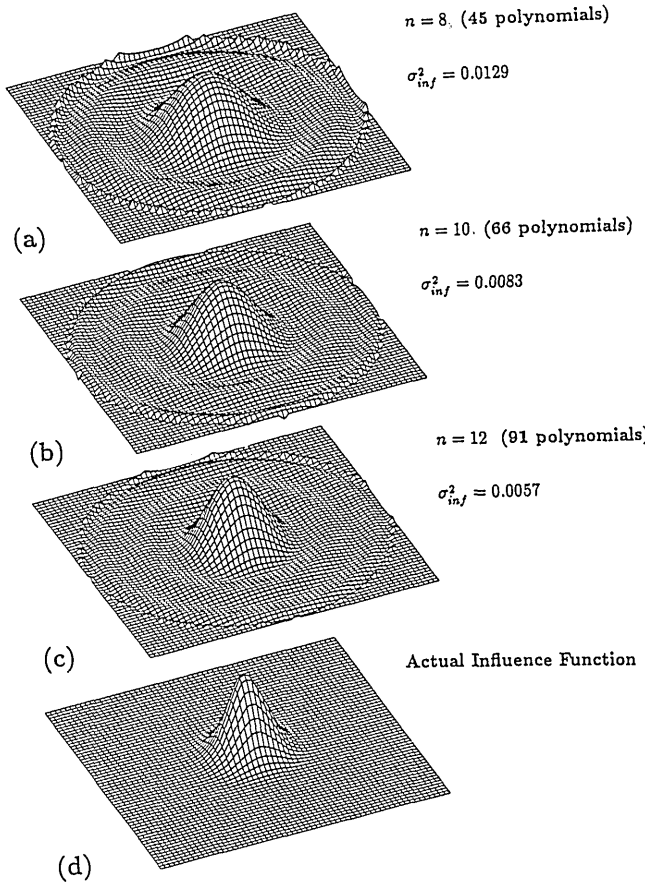

Fig. 4. Gaussian-type influence functions as represented by a finite set of Zernike polynomials. The representation is more accurate as more polynomials are used. (a), (b), (c) show the restoration of the influence function using 45 (up to the eighth degree, $n = 8$), 66 (10th degree, $n = 10$), and 91 (12th degree, $n = 12$) Zernike polynomials, respectively; (d) shows the actual influence function corresponding to the central actuator. These figures correspond to a scaling factor $s = 1$. If the scaling factor increases, the influence function becomes wider and the fitting will be better.

represented more accurately by means of a finite Zernike expansion, which is why the method proposed in this paper is well suited for a small mirror with few actuators. In particular the DM modeled in this paper has 21 actuators distributed along a square grid (see Fig. 1). The influence functions of the actuators are assumed to be equal and have a Gaussian shape with a coupling factor of 13% to the nearest neighbor.

The relationship between the size of the DM and the size of the aperture for the Zernike expansion will affect the accuracy of the fitting. The scaling factor s represents the relationship between the geometrical parameters that define the mirror and the aperture over which the Zernike polynomials are defined:

$$s = \frac{L}{r_{\text{aperture}}}, \tag{19}$$

where L is the side of the square grid where the actuators are located. In our case L is equal to 4

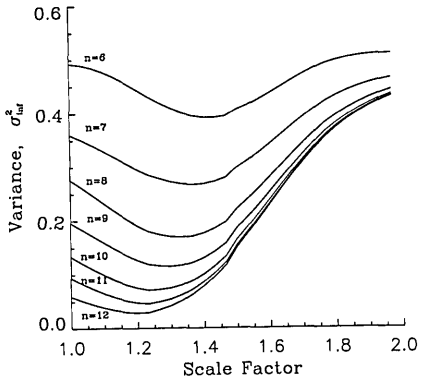

Fig. 5. Total variance of the whole actuator array σ_{inf}^2 is represented as a function of the scale factor s for several decompositions including 6th- to 12th-degree polynomials. The parameter n is the maximum degree of the polynomial used in the expansion, which varies from 6 to 12.

times the interactuator distance (see Fig. 1). We evaluated the fitting for values of s between 1 and 2. A scaling factor $s = 1$ indicates that the circular aperture radius is 4 times the interactuator spacing. The value $s = 2$ indicates that the circular aperture radius is only twice the interactuator separation.

A parameter that can be used to quantify the accuracy of fitting of the whole array of actuators is the combined variance defined as

$$\sigma_{\text{inf}}^2 = \sum_{j=1}^{J} \sigma_j^2, \tag{20}$$

where σ_j is the variance of the error of the representation of an individual influence function by means of Zernike polynomials. To calculate σ_j we use the error of the representation given by the following

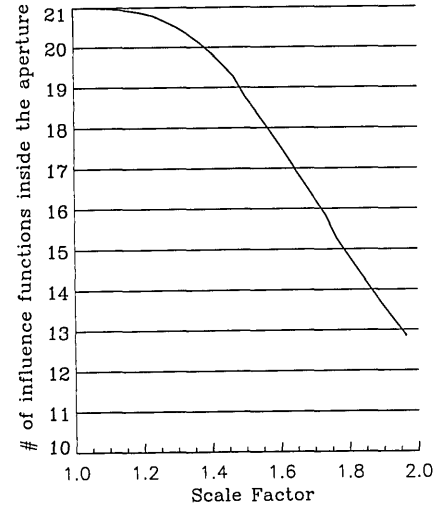

Fig. 6. When the circular aperture where the Zernikes are expanded is reduced, some portions of the outer influence functions are excluded. We have represented this decreasing in the characterization of the mirror in terms of the equivalent-area number of influence functions included inside the aperture.

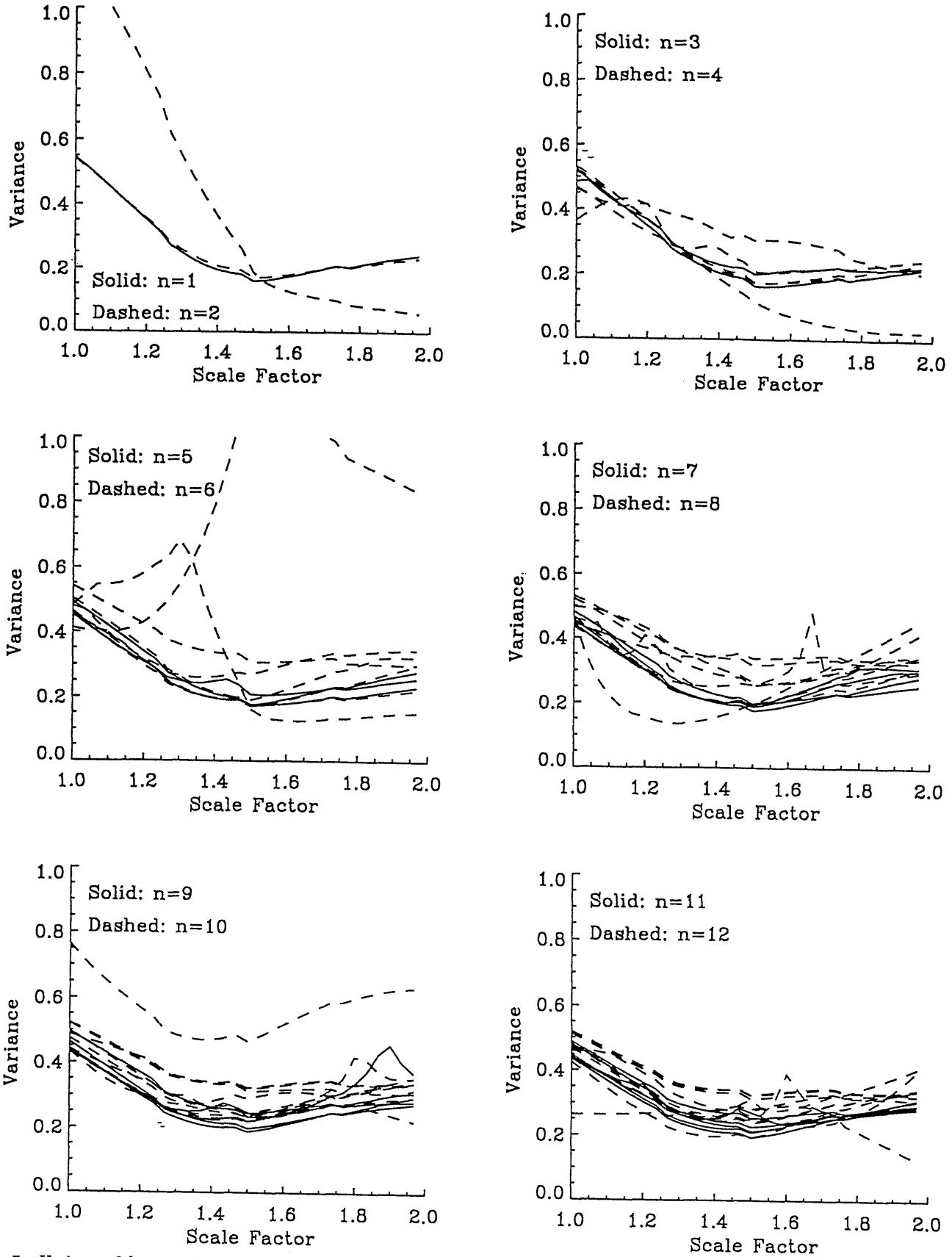

Fig. 7. Variance of the wave-front-error vector between the actual Zernikes and the restoration by the mirror σ_{wf}^2 is given in this figure for the Zernike polynomials for the modal driving procedure. This figure also represents qualitatively the behavior of the variance when the mirror is driven in the zonal mode.

discrete array:

$$\epsilon_j = m_j - \sum_{k=0}^{K-1} c_{jk} Z_k, \tag{21}$$

where both m_j and Z_k have been sampled along the sampling grid. Thus the variance is defined as

$$\sigma_j^2 = \epsilon_j^t \epsilon_j. \tag{22}$$

In Fig. 5 the variance σ_{inf}^2 is plotted as a function of the scaling factor s. A family of curves is generated with the parameter n, which corresponds to the maximum radial number of the Zernikes used in the expansion. We note that for any particular n there is a value of s that minimizes the variance. As n increases the optimum scale factor decreases (larger aperture), which allows the narrower influence functions to be represented by the expansion. It should be noted that as the scale factor increases (the smaller aperture) a significant portion of the area of the outer influence functions will be outside the aperture, and therefore the effect of these actuators is diminished (see Fig. 6). At this point we could define the optimum size of the aperture over which the Zernikes are expanded as the size that minimizes the variance defined in Eq. (20). However, this gives only information about how well the influence function can be expanded in terms of the Zernikes when all the influence functions have equal weight. This would not be a measure of the behavior of the whole mirror when it is working in a given driving mode, because the DM is never operated with all the actuators displaced at their maximum stroke. Thus we should note that the variance defined in Eq. (20) σ_{inf}^2 does not represent the fitting in response to any incoming wave-front function as Eq. (4) does for σ_{wf}^2. It measures only the accuracy of the expansion of the influence functions in terms of the Zernikes.

To develop a wave-front-based criterion for the choice of the aperture size, we reverse the problem and represent the Zernike polynomials in terms of the influence functions. This procedure can be approached in three ways.

First, one applies a zonal fitting where the influence function tries to represent the polynomial around its area of influence. In this case we obtain the driving signal of the actuators by applying a least-squares fitting method whose general solution is Eq. (5). The variance of the fitting will be a function of the scaling factor s.

The second approach to representing the Zernike polynomials is based on a pure modal way, and the solution is given in Eq. (14) where all the elements of the **I** vector are zero except the one that corresponds to the Zernike polynomial that we want to represent. This value of this element of the **I** vector is equal to one.

The variance $\sigma_{wf(modal)}^2$ as a function of the scaling factor is shown in Fig. 7 for Zernike polynomials up to the 12th degree.

As we see in Eqs. (17) and (18), the behavior of $\sigma_{wf(modal)}^2$ and $\sigma_{wf(zonal)}^2$ is similar. Thus Fig. 7 shows only $\sigma_{wf(modal)}^2$. The difference between the two representations is usually small and occurs for only a few polynomials. This agreement is seen in Fig. 8, where we plot the scale factor that gives the minimum wave-front variance for both the zonal and modal cases as a function of the Zernike polynomial index k. It is clear the minimum is located, in the case studied, around the value $s = 1.5$ for most Zernike polynomials. For this value the aperture excludes the equivalent area of 2.24 influence functions (see Fig. 6). Referring to Fig. 1, we notice that this exclusion has to be shared, at least by the eight outer actuators, and therefore each of them has more than 70% of its influence function inside the clear aperture. The cases for which the variance is not minimum around $s = 1.5$ correspond to polynomials whose symmetry is different from the rectangular symmetry of the grid of actuators.

The value of the coefficients of the decomposition of the influence functions in terms of the Zernike polynomials is expressed by means of Eq. (9). When we use the shifting properties of the δ function, this equation is simpler when all the influence functions are equal:

$$c_{jk} = \frac{1}{\pi} \int \delta(x - x_j, y - y_j)[m(x, y) \bigstar Z_k(x, y)]\mathrm{d}x\mathrm{d}y, \tag{23}$$

where $\bigstar$ denotes the correlation operation.

The analysis of this expression gives us a third way to represent the Zernike polynomial. When the influence functions are narrow compared with the aperture size, the correlation gives the following approximate result:

$$Z_k \simeq m_j \bigstar Z_K, \tag{24}$$

and the coefficients of the expansion are approximately $\tilde{c}_{jk}$, where

$$\tilde{c}_{jk} = Z_k(x_j, y_j). \tag{25}$$

These coefficients permit the exact reproduction of

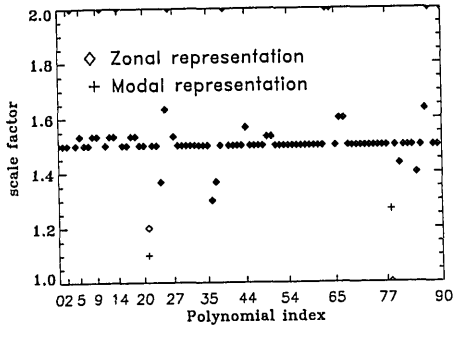

Fig. 8. Value of the scaling factor at which the minimum variance is reached is plotted in this diagram as a function of the polynomial index. The different symbols denote the different method of calculating the fitting: diamonds are the zonal representation; pluses are the modal representation.

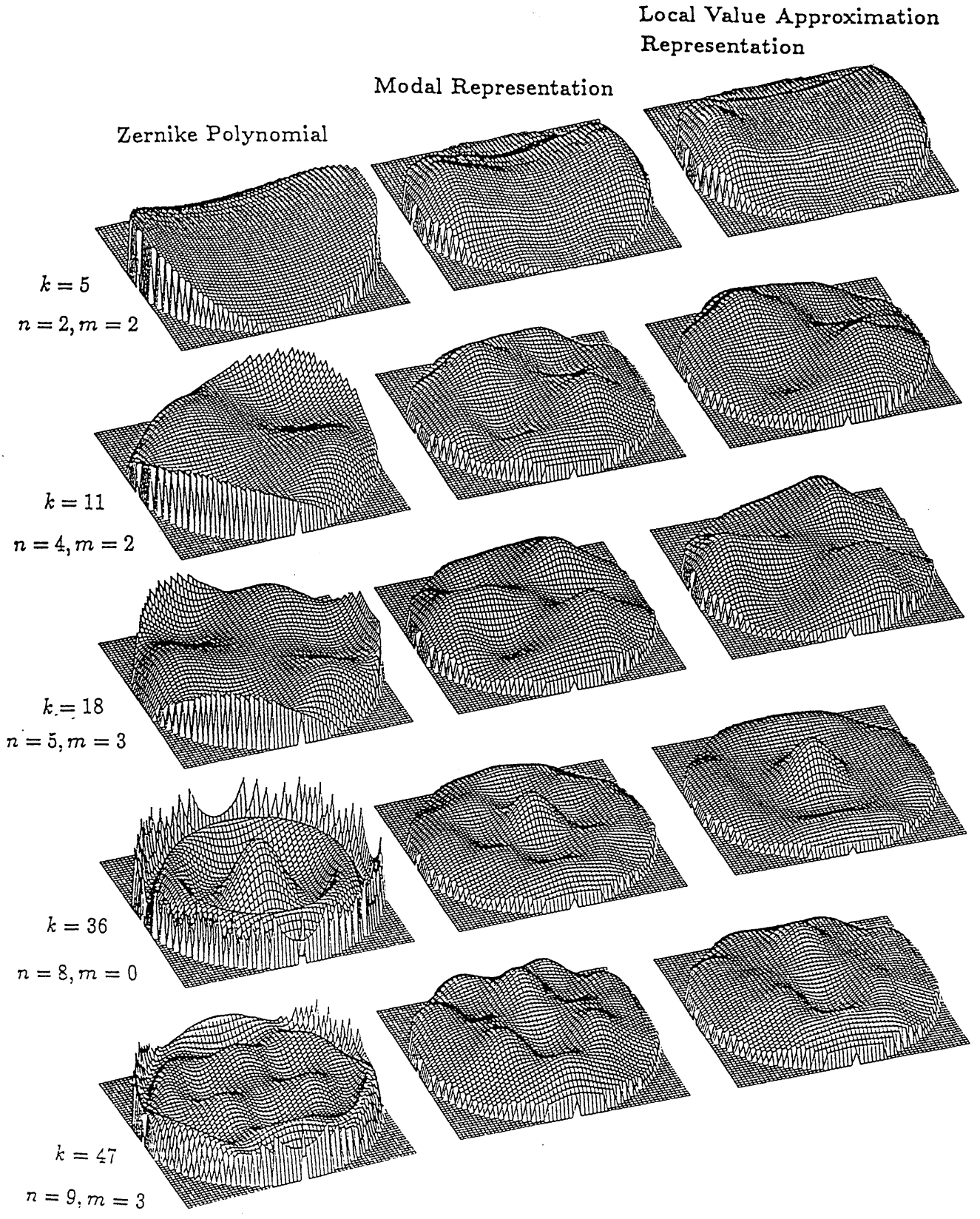

Fig. 9. DM compared with several Zernike polynomials whose radial numbers n, azimuthal numbers m, polynomial indices k are noted for each row. The first column is the polynomial that the DM represents. The second column corresponds to modal driving. The zonal representation has a similar appearance and is not shown separately. The third column is obtained from the local value approximation. A scale factor $s = 1.5$ is used throughout. The values for $\sigma_{wf}{}^2$ in each row are as follows from top to bottom: 0.1905 (modal), 0.1907 (zonal), 0.2085 (LVA); 0.1734 (modal), 0.1734 (zonal), 0.2118 (LVA); 0.2083 (modal), 0.2085 (zonal), 0.2477 (LVA); 0.1970 (modal), 0.1964 (zonal), 0.1260 (LVA); 0.1879 (modal), 0.1879 (zonal), 0.1876 (LVA).

the Zernike polynomial Z_k at the points where the actuators are located if we drive the actuator j with the driving signal $\tilde{c}_{jk}$. Then the Zernike Z_k can be represented by means of the following arrangement of influence functions:

$$Z_k \simeq \omega_m = \sum_{j=1}^{J} \tilde{c}_{jk} m_j, \tag{26}$$

or in a matrix form

$$\mathcal{Z} \simeq \mathcal{H}\{\tilde{c}_{jk}\}^t. \tag{27}$$

This behavior, which we call the local value approximation (LVA), can be useful in a first-order analysis when the influence functions are narrow. This occurs usually when the mirror has a large number of actuators.

The calculations above allow us to show how the different approaches work for a given polynomial term. Figure 9 shows the representation of a few Zernike polynomials (first column) when we use the zonal driving (second column) and the LVA (third column) for the optimum value, $s = 1.5$, obtained from the analysis of the modal and zonal driving. To obtain this representation, we calculated the driving signal of the actuators **a**. This driving signal has been normalized to have values between 0 and 1. The Zernike polynomial has also been modified so that its excursion ranges from 0 to 1. The surfaces of the DM and the Zernike are compared, and the variance is obtained. The maximum value of the variance is the unit circle area π, obtained for the case of the maximum error of unity occurring over the whole aperture. In the caption for Fig. 9 we noted the variance values for the modal, zonal, and LVA methods. The pictorial representation for modal driving also applies to zonal driving. We can see that the representation of the Zernike polynomials is much better in the center of the aperture where the spatial frequency of the Zernike is lower. The restoration degrades as we look to the edge of the aperture. However, the value of the variance does not change significantly for higher-degree polynomials because the high-frequency ripples are located in a narrow interval close to the edge of the aperture.[11]

Conclusion

Modal driving of a DM has been described by means of an expansion of the individual influence functions of the mirror in terms of the Zernike polynomials. This method is especially appropriate for representing DM's with a smaller number of actuators, because the accuracy of the expansion improves when the influence functions are relatively wide compared with the total size of the mirror. We optimize the size of the aperture over which the Zernike polynomials are defined by minimizing the variance of the reconstructed wave front. This optimum aperture is found to encroach just on the outermost actuators of the DM.

This study was done while one of the authors (J. Alda) was at the Center for Research in Electro-Optics and Lasers, University of Central Florida. The study was supported by the Programa de Perfeccionamiento y Movilidad del Personal Investigador, BE90-140, from the Dirección General de Investigación Científica y Tecnológia (Spain), and by a grant from the Center for Research in Electro-Optics and Lasers (Orlando, Fla.).

References

1. M. Born and E. Wolf, *Principles of Optics* (Pergamon, London, 1980), Sec. 9.2 and App. VII.
2. J. Y. Wang and D. E. Silva, "Wavefront interpretation with Zernike polynomials," Appl. Opt. **19,** 1510–1518 (1980).
3. V. N. Mahajan, "Zernike annular polynomials for imaging systems with annular pupils," J. Opt. Soc. Am. **71,** 75–85 (1981).
4. R. J. Noll, "Zernike polynomial and atmospheric turbulence," J. Opt. Soc. Am. **66,** 207–211 (1976).
5. D. E. Novoseller, "Zernike-ordered adaptive correction of thermal blooming," J. Opt. Soc. Am. A **5,** 1937–1942 (1988).
6. R. K. Tyson, "Using the deformable mirror as a spatial filter: application to circular beams," Appl. Opt. **21,** 787–793 (1982).
7. R. K. Tyson, *Principles of Adaptive Optics* (Academic, San Diego, Calif., 1991), and references therein.
8. T. R. O'Meara, "Theory of multidither adaptive optical systems operating with zonal control of deformable mirrors," J. Opt. Soc. Am. **67,** 318–325 (1977).
9. K. E. Moore and G. N. Lawrence, "Zonal model of an adaptive mirror," Appl. Opt. **29,** 4622–4628 (1990).
10. J. E. Harvey and G. M. Callahan, "Wavefront error compensation capabilities of multi-actuator deformable mirrors," in *Adaptive Optical Components I,* S. Holly and L. James, eds., Proc. Soc. Photo-Opt. Instrum. Eng. **141,** 50–57 (1978).
11. E. S. Claflin and N. Bareket, "Configuring an electrostatic membrane mirror by least-squares fitting with analytically derived influence functions," J. Opt. Soc. Am. A **3,** 1833–1839 (1986).