Zonal Matrix Iterative Method for Wavefront Reconstruction From Gradient Measurements

Sophia I. Panagopoulou, PhD; Daniel R. Neal, PhD

ABSTRACT

PURPOSE: To present an alternate method to Zernike decomposition (modal) of wavefront reconstruction using iterative implicit solution to the finite difference equations (zonal).

METHODS: Different reconstruction methods, modal and zonal, were compared and the advantages of each method were analyzed.

RESULTS: Although the modal or Zernike method allows for quantitative interpretation of some of the aberrations, it is cumbersome for use with fine details and may lead to errors for eyes with keratoconus or other rapidly varying aberration. The zonal method produces a very high-resolution map that can be used for identifying irregular structures.

CONCLUSIONS: The distinction between the two methods is useful to maintain, and the solution methods are generally different. In practice, both methods are useful and, with modern computers, both zonal and lower-order modal may be calculated rapidly. The difference between the wavefronts derived from the two methods may provide useful insight or interpretation of the information. [*J Refract Surg*. 2005;21:S563-S569.]

Although ocular wavefront aberrations have been measured since 1961,[1,2] only recently, with improved aberrometers, has it become possible to routinely obtain wavefront maps under clinical conditions.[3-5] In refractive surgery, apart from the long-established corneal topography, the wavefront analyzers are used for mapping the total aberrations of an eye through the pupil entrance. At this time, the preferred surface fitting method for characterizing the wavefront aberrations is the Zernike decomposition, which uses the Zernike polynomials to describe and quantify the various wavefront surfaces. The conventional refraction can be decomposed into prism, defocus, cylinder, and axis. In the same manner, the remaining optical errors can be described with individual aberrations, so-called Zernike modes, with the process of Zernike decomposition. The evaluation of individual Zernike modes reveals large differences in their subjective impact.[6,7] Studies from Applegate et al[6,7] and Williams[8] showed that this decomposition of the wave aberration in its fundamental components to approach the subjective image quality could cause misinterpretations. This occurs because Zernike modes can interact strongly with each other, and in complex ways, to determine the final image quality. Smolek and Klyce[9] measured the interactions between different Zernike coefficients and found that pairs of aberrations can sometimes increase the acuity from each individual component as well as leading to a larger decrease.

Although the Zernike approximation is effective for describing the individual lower and higher order aberrations, in most practical implementations they also have a smoothing effect that can significantly limit the aberrations that account for the wavefront error. Because of limitations with Zernike (modal) reconstruction, a different reconstruction method, such as zonal reconstruction, should be considered. This article presents the concept for modal and zonal reconstructions.

From the Department of Ophthalmology, University of Crete Medical School, Crete, Greece (Panagopoulou); and WaveFront Sciences Inc, Albuquerque, NM (Neal).

Dr Panagopoulou has no proprietary interest in the materials presented herein. Dr Neal has financial and proprietary interests in WaveFront Sciences Inc, Albuquerque, NM.

Correspondence: Sophia I. Panagopoulou, PhD, University of Crete Medical School, VEIC P.O. Box 1352, 71110 Heraklion, Crete, Greece. Tel: 306942558053; Fax: 302810300097; E-mail: spanagop@med.uoc.gr

WAVEFRONT RECONSTRUCTION METHODS

The wavefront *w*, at any point *(x,y)*, is related to the gradients through the vector gradient equation

$$\nabla w = \frac{\partial w(x,y)}{\partial x}\,\mathbf{i} + \frac{\partial w(x,y)}{\partial y}\,\mathbf{j} = S^X_{j,k}\,\mathbf{i} + S^Y_{j,k}\,\mathbf{j} \qquad (1)$$

where **i** and **j** are the unit vectors in the x and y directions and S^X_{jk} and S^Y_{jk} are wavefront slope measurements at the *(x, y)* location.

Southwell[10] describes several methods for determining a wavefront surface from slope data. He describes several different geometries in common use for obtaining slope data and introduces the concept of Modal and Zonal reconstructors. In the end, the objective of either method is to take a set of wavefront gradient measurements S^X and S^Y and to calculate, or reconstruct, the wavefront map from this data.

MODAL RECONSTRUCTOR

In the modal reconstructor method, the wavefront surface is described in terms of a set of smoothly varying modes. These may be polynomials or other functions, or they may be derived from some other system properties (eg, deformable mirror actuator influence functions are used in adaptive optics). The key property of these modes is that an analytic derivative can be obtained that may be fit to the measured slope data. Thus the wavefront surface may be written as expansion of polynomials $P_n(x,y)$

$$W(x,y) = \sum_n C_n P_n(x,y) \qquad (2)$$

with coefficients C_n. Because the functions $P_n(x,y)$ are continuous and analytic through the first derivative, the wavefront gradient may be written

$$\begin{bmatrix} \dfrac{\partial w(x,y)}{\partial x} \\ \dfrac{\partial w(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} \displaystyle\sum_n C_n \frac{\partial P_n(x,y)}{\partial x} \\ \displaystyle\sum_n C_n \frac{\partial P_n(x,y)}{\partial y} \end{bmatrix} \qquad (3)$$

This provides an analytic description of the wavefront slope at every point *(x,y)*, which may be compared to the measured gradient values to obtain the appropriate set of coefficients C_n by least squares fitting or other methods.

Given a set of gradient data β^x_k and β^y_k at measurement point (x_k,y_k): a least squares fit to data may be obtained by minimizing the error function

$$\chi^2 = \sum_k \left(\beta^x_k - \sum_{n=2}^{M} C_n \frac{\partial P_n(x_k,y_k)}{\partial x} \right)^2 + \sum_k \left(\beta^y_k - \sum_{n=2}^{M} C_n \frac{\partial P_n(x_k,y_k)}{\partial y} \right)^2 \qquad (4)$$

with respect to the coefficients C_n. Once these coefficients are determined, the wavefront surface is known through Eq.(2). Equation 4 may also be used to calculate the residual fit error, which is useful for characterizing the quality of the fit or the appropriateness of the particular set of polynomials for describing the measured data.

This type of reconstructor provides a compact notation for describing the wavefront, as the detailed shape information is contained in the modes $P_n(x,y)$. If the modes are chosen to be polynomials that approximate optical properties, the coefficients may be interpreted to obtain a direct description of the physical system. For example, Zernike polynomials are often used in optics because they are closely related to the aberrations introduced in optical systems and they are defined over a circular pupil. Hence individual polynomials describe parameters such as defocus, astigmatism, coma, and spherical aberration. If orthogonal polynomials are used, it is possible to separate the various effects, as the presence (or absence) of one term does not affect the others.

ZONAL RECONSTRUCTOR

The term "zonal reconstructor" comes about because, rather than describing the wavefront in terms of a set of overall modes, each of which is described analytically over the whole surface, the wavefront is described only locally over a limited zone. These zones are usually chosen to match the area described by each gradient measurement. They can represent a regular grid of square or rectangular elements or some other local region such as a hexagonal or other shape. For Shack-Hartmann wavefront sensors, these zones are generally chosen to match the lenslets, although this is not strictly necessary.

ZONAL RECONSTRUCTION METHODS

There are several methods for solving for a point-by-point wavefront surface representing a self-consistent solution to the gradient equation (Eq.1).[11,12] These include the matrix iterative approach, direct least-square fitting for wavefront values, and spline integration.

This article concentrates on systems where the slope data were obtained on a regular set of square regions and where the average slope across each region is known. This is the typical arrangement for Shack-Hartmann wavefront sensor data. The same methods can be used for other geometries, but the appropriate nomenclature would need to be developed.

INTERPOLATING FUNCTIONS

For Shack-Hartmann gradient data, the average slope

across each lenslet is determined from focal spot position shift. Because the slope varies from lenslet to lenslet, it is reasonable to assume an interpolating function where the wavefront slope varies linearly between lenslets. Thus

$$S^X(x,0) = c_1 + c_2 x \text{ and } S^Y(0,y) = c_3 + c_4 y, \quad (5)$$

where c_1, c_2, c_3, and c_4 are coefficients used to model the slope variation.

For a first order difference equation system, we will consider only those adjacent points that are either directly above or below, or right and left, of a central point. Integrating $S^X(x,0)$ and $S^Y(0,y)$ along the x and y axis results in the interpolating function for the wavefront:

$$w(x,0) = c_0 + c_1 x + \frac{1}{2} c_2 x^2 \quad \text{and}$$

$$w(0,y) = c_0' + c_3 y + \frac{1}{2} c_4 y^2. \quad (6)$$

Difference Equations

For a lenslet l, with center (x_l, y_l), the average slope over the lenslet is (S_l^X, S_l^Y). It is more convenient to represent the data on a grid with indices (j,k) and spacing h_x and h_y. Thus the slope data is known at points $S_{j,k}^X$ and $S_{j,k}^Y$. Evaluating Eq.(5) for adjacent points (eg, at points $[j,k]$ and $[j+1,k]$) allows the determination of the coefficients:

$$c_1 = S_{j,k}^X,\ c_2 = \frac{S_{j+1,k}^X - S_{j,k}^X}{h_x}\ c_3 = S_{j,k}^Y \text{ and } c_4 = \frac{S_{j,k+1}^Y - S_{j,k}^Y}{h_y}, \quad (7)$$

$$c_0 = c_0' = w_{j,k} \quad (8)$$

These coefficients can be determined for each direction from the central point (j,k), ie, $(j+1,k)$, $(j-1,k)$, $(j,k+1)$, $(j,k-1)$. This yields the following set of difference equations that describe the wavefront at a central point w_{jk} in terms of adjacent points and the measured gradient values.

$$w_{jk} = w_{j,k-1} + \frac{h_y}{2}(S_{j,k-1}^Y + S_{j,k}^Y) = H_Y^+\{w_{j,k-1}\} \quad (9)$$

$$w_{jk} = w_{j,k+1} - \frac{h_y}{2}(S_{j,k+1}^Y + S_{j,k}^Y) = H_Y^-\{w_{j,k+1}\} \quad (10)$$

$$w_{jk} = w_{j-1,k} + \frac{h_x}{2}(S_{j-1,k}^X + S_{j,k}^X) = H_X^+\{w_{j-1,k}\} \quad (11)$$

$$w_{jk} = w_{j+1,k} - \frac{h_x}{2}(S_{j+1,k}^X + S_{j,k}^X) = H_X^-\{w_{j+1,k}\} \quad (12)$$

where $H_X^-\{w_{j+1,k}\}$ is a predictor operator that predicts the wavefront of w_{jk} from the measured slope and the values at $w_{j+1,k}$.

Boundary Conditions

The boundary conditions must properly account for edge, corner, or interior points in the measurement. Generally, for a Shack-Hartmann or other slope sensor, there is also an independent measurement of the total amount of light that has been received by the lenslet. Thus we define a weighting function I_{jk} that defines these boundaries. This function has zero value everywhere that there is no incident light. It may have a value (1) in regions where the lenslet is fully illuminated or some other intermediate value for partially illuminated lenslets.

Matrix Iterative Solution

The weighted average of the predictions from the four different directions is given by

$$w_{jk} = \frac{I_{j-1,k}H_X^+\{w_{j-1,k}\} + I_{j+1,k}H_X^-\{w_{j+1,k}\} + I_{j,k-1}H_Y^+\{w_{j,k-1}\} + I_{j,k+1}H_Y^-\{w_{j,k+1}\}}{I_{j-1,k} + I_{j+1,k} + I_{j,k-1} + I_{j,k+1}} \quad (13)$$

This allows the value of the wavefront to be predicted from wavefront values in the four different directions. Gathering terms allows the definition of:

$$q_{jk} = \frac{h_x}{2}\left[S_{jk}^x\left(I_{j-1,k} - I_{j+1,k}\right)\right] + \frac{h_y}{2}\left[S_{jk}^y\left(I_{j,k-1} - I_{j,k+1}\right)\right] \quad (14)$$

$$b_{jk} = \frac{h_x}{2}\left[I_{j-1,k}S_{j-1,k}^X - I_{j+1,k}S_{j+1,k}^X\right] + \frac{h_y}{2}\left[I_{j,k-1}S_{j,k-1}^Y - I_{j,k+1}S_{j,k+1}^Y\right] \quad (15)$$

$$g_{jk} = I_{j-1,k} + I_{j+1,k} + I_{j,k-1} + I_{j,k+1} \quad (16)$$

and

$$\overline{w}_{j,k} = \frac{I_{j-1,k}w_{j-1,k} + I_{j+1,k}w_{j+1,k} + I_{j,k-1}w_{j,k-1} + I_{j,k+1}w_{j,k+1}}{g_{j,k}}. \quad (17)$$

This equation allows for iterative solutions for the wavefront as given by:

$$w_{j,k}^{[m+1]} = \overline{w}_{j,k}^{[m]} + \frac{(b_{j,k} + q_{jk})}{g_{j,k}}, \quad (18)$$

where m is the iteration number.

Equation 15 contains only those points needed for an interior (uniform weighted) calculation, whereas the boundary conditions are contained in Eq.(14). An important consideration is the effect of edges on the data. For an edge or corner point, the wavefront can be

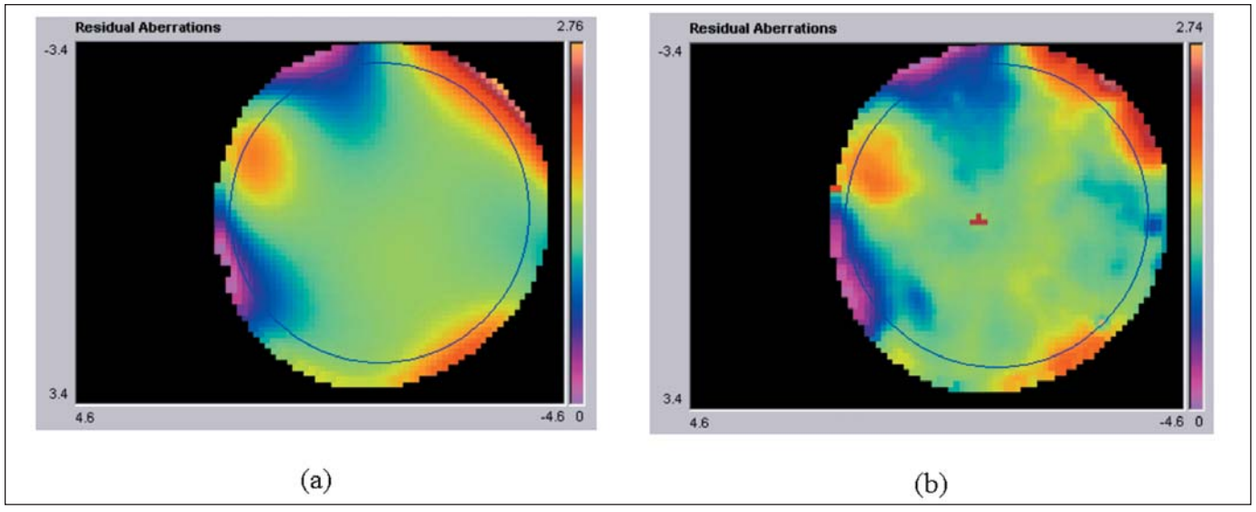

Figure 1. Comparison of modal and zonal methods for wavefront reconstruction. **A)** 6th order modal (Zernike) reconstruction and **B)** zonal reconstruction of the same data. Both data sets are from a patient with a normal eye and were recorded with the COAS HD with 105 μm resolution.

predicted from only one or two directions. However, if the data are placed into a larger array that has been initialized to zero, the intensity values at boundary points are always zero, and the boundary conditions will naturally be considered properly. It should be noted that for uniform intensity, where I=1 everywhere except at the boundaries, these equations converge to exactly the same equations as derived by Southwell.[10] It generally takes a number of iterations equal to the number of lenslets across the entire array to converge to a stable solution.

For laser beams or other wavefronts where the irradiance distribution varies rapidly, we have found that it is possible to replace the weighting function I_{jk} with the measured irradiance over each lenslet. This weighting serves to minimize the effect of increased noise as the irradiance decreases.

It should be noted that the five-point kernel of Eq.(14-17) results in a connection between adjacent data points. The redundancy of the data in these equations provides averaging of noise or other error, and the iterative solution produces the wavefront that is the most consistent with all of the slope measurements. The method produces a wavefront surface that is accurate at spatial frequencies up to the Nyquist limit. It attenuates frequencies that are higher than that limit.

EXAMPLES

In Figure 1, we analyzed the same data from the measured wavefront of a human eye using both the modal and zonal methods. The modal method (a) produces very smooth surfaces, whereas much greater detail is evident in the zonal method (b). The overall wavefront peak-to-valley (P-V) was approximately the same in both cases. Although the additional structure could be construed as noise, the sensor design produces excellent focal spots, which can be located accurately.[13] For this case, the structure is a combination of tear film and fixed structure in the ocular wavefront measurement. In both cases, the sensor noise was low enough not to contribute significantly to the wavefront image.[14] The structure represents a real wavefront structure, present in the measurement.

Clinically, this becomes significant if the information needed to make a diagnosis cannot be seen in the data. For example, consider the eye in Figure 2 that underwent conductive keratoplasty. Figure 2A represents a 10th order Zernike fit and shows a very smooth wavefront. Except for some roughness at the edge, we would assume the vision was not impaired and significant higher order aberrations were not present. Figure 2C is a slit-lamp picture showing the aberrated cornea. These aberrations are evident in the zonal map (Fig 2B), which now has enough detail that the real aberration can be diagnosed. The fact that the structure is a feature of the cornea is also evident in Figure 2D, which shows the corneal topography elevation map. The zonal reconstructor does not artificially limit the available spatial frequencies in the data (except through the lenslet sampling and Nyquist limits).

Figure 3 presents another example where the detail needed to make a diagnosis was missing from the modal wavefront maps. In this case, the zonal wavefront map could be correlated to corneal topography, with the same features evident in both images. No such correlation can be obtained from the modal re-

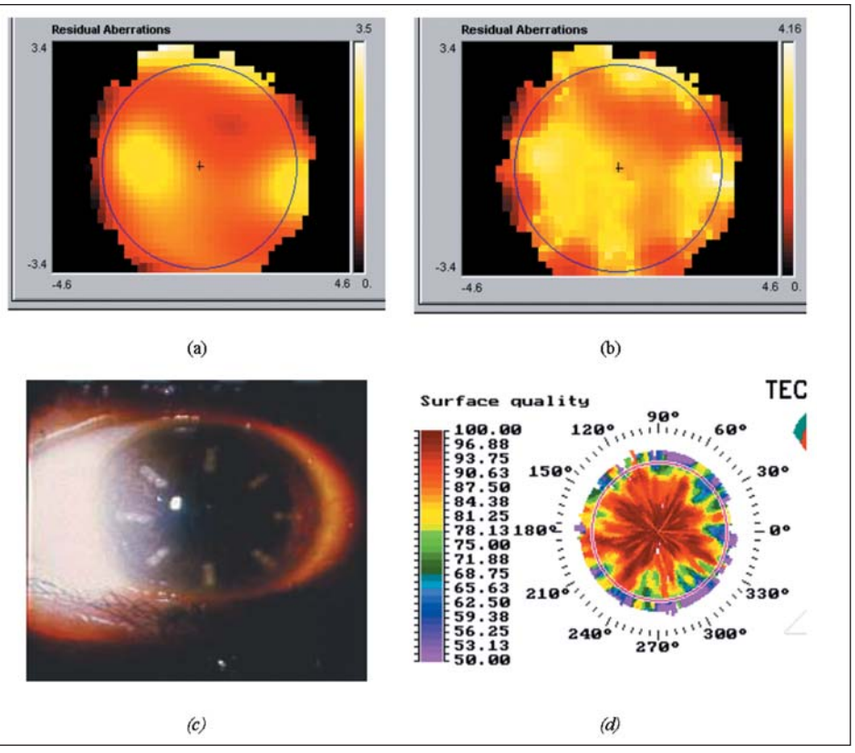

Figure 2. Presentation of the higher order wavefront map of an eye 1 month after conductive keratoplasty. **A)** Zernike 10th order aberration map and **B)** zonal wavefront map from the same data. **C)** Slit-lamp photograph of the eye. **D)** Results from a Technomed corneal topography. The modal data, even at 10th order, does not contain sufficient information for diagnosis or correlation with other clinical instruments.

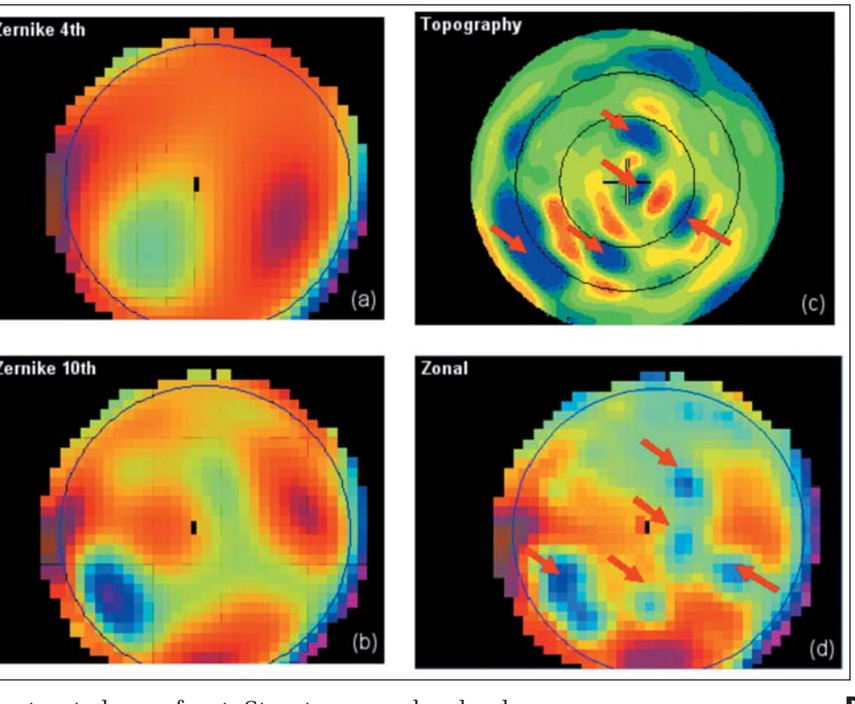

Figure 3. Eye with corneal scars. Zernike wavefront map of **A)** 4th and **B)** 10th order analysis. **C)** Corneal topography of the eye and **D)** corresponding zonal wavefront map. These maps show clinically relevant features that are beyond the ability of a modal reconstructor to properly represent, but still less than the Nyquist limit.

constructed wavefront. Structures can be clearly seen that are beyond the ability of the modal methods to accurately capture, yet are still well within the Nyquist limit.

DISCUSSION

Zernike modes can interact with each other to determine final image quality, and the complexity of the interactions between modes means that Zernike de-

TABLE

Computation Times Measured in Seconds for Modal (Zernike) and Zonal Reconstruction as a Function of Fit Order

		Modal		
Order	No. of Coefficients	Compute Design Matrix	Saved Design Matrix	Zonal
4	15	.031	.015	.016
6	28	.094	.031	.016
8	45	.281	.031	.016
10	66	.766	.047	.016
12	105	1.89	.078	.016
14	120	4.27	.109	.016
16	171	8.84	.156	.016
40	861	1632	.219	.016
54	1540	9360	87	.031

Note. The design matrix computation can be a significant portion of the total computation time. Below 12th order, the times are measured using the COAS analysis module. Above 12th order, the times are extrapolations based on the trends. This is necessary because COAS switches to the recursion relationship after 12th order, which is much slower.

composition is not necessarily effective as a metric of subjective image quality. It has been well established that the total root-mean-square wavefront error, by itself, is not a good predictor of image quality.[5] Modal representation has the advantage in that it can provide quantization of clinical data. In modal reconstruction, the fewer the number of Zernike orders that are used in a fit, the smoother the representation. On the other hand, when the data is over-smoothed, the information is mostly lost, especially for highly aberrated eyes.[5]

The advantage of the zonal representation lies in the lack of smoothing effect due to the use of the entire raw data set, resulting in representations with higher spatial frequencies. Zonal reconstruction may provide a better evaluation method for calculating customized ablation profiles in eyes dominated by higher order aberrations, as well as for diagnostic applications. Combined information from both methods could provide advantages in clinical evaluation of vision quality.

Generally, zonal reconstructions lead to a unique description of the wavefront surface at every measurement point. The zonal reconstruction supports rapid (point by point) variation in the wavefront surface and thus will provide for very high-resolution wavefront description. It does not provide for a direct interpretation in optical terms. An additional computation step is required to determine defocus, astigmatism, or other parameters from zonal data.

In theory, it is possible to fit very large basis sets with the modal such that the modal description can be used to describe very rapid local variations. However, for most solution methods, this would require a greater amount of computation time and could result in numerical instabilities that would lead to an inaccurate result. For practical wavefront sensors with high resolution lenslet arrays (800 to 1500 samples across a 6-mm pupil), the computation time can be significant for Zernike polynomial fitting at greater than 10th order. In the Table, the computation times are compared for modal vs zonal reconstruction for the data shown in Figure 1. The least-squares fitting process involves first computing a "design matrix," which represents the product of all modes evaluated at all points, and then inverting this matrix and performing a vector matrix multiply to obtain the solution for the coefficients. The expansion is then evaluated at each point to obtain the wavefront map. By pre-computing the design matrix, much of the computation can be performed in advance so long as the basis set or the coordinate space does not change.[12] However, it is difficult to anticipate the exact space that would be needed for a particular measurement, therefore it is difficult to pre-compute the correct matrix. Although other methods can be used for determining the polynomial coefficients, the least squares approach is generally considered the most accurate.[15] Even the saved design matrix approach results in significantly greater computation time for the modal reconstructor compared to the zonal method.

The distinction between the two methods is useful to maintain, and the solution methods are generally different. In practice, both methods are useful and, with modern computers, both zonal and lower-order modal

can be calculated rapidly. The difference between the wavefronts derived from the two methods may provide useful insight or interpretation of the information.

REFERENCES

1. Smirnov MS. Measurement of the wave aberration of the human eye. *Biofizika*. 1961;6:766-795.
2. Ivanoff J. About the spherical aberration of the eye. *Opt Soc Am*. 1956;46:901-903.
3. MacRae S, Krueger R, Applegate RA. *Customized Corneal Ablation: The Quest for Supervision*. Thorofare, NJ: SLACK Inc; 2001.
4. Krueger R, Chalita M. *Ophthalmology Clinics of North America: Wavefront Technology*. Philadelphia, Pa: Saunders; 2004.
5. Krueger R, Applegate RA, MacRae S. *Wavefront Customized Visual Correction: The Quest for Supervision*. 2nd ed. Thorofare, NJ: SLACK Inc; 2004.
6. Applegate RA, Sarver EJ, Khemsara V. Are all aberrations equal? *J Refract Surg*. 2002;18:S556-S562.
7. Applegate RA, Ballentine C, Gross H, Sarver EJ, Sarver CA. Visual acuity as a function of Zernike mode and level of root mean square error. *Optom Vis Sci*. 2003;80:97-105.
8. Williams DR. What adaptive optics can do for the eye. *Review of Refractive Surgery*. 2002;3:14-20.
9. Smolek MK, Klyce SD. Zernike polynomials fitting fails to represent all visually significant corneal aberrations. *Invest Ophthalmol Vis Sci*. 2003;44:4676-4681.
10. Southwell WH. Wavefront estimation from wave-front slope measurements. *Journal of the Optical Society of America*. 1980;70:998-1006.
11. Roddier F, Roddier C. Wavefront reconstruction using iterative Fourier transforms. *Appl Opt*. 1991;30:1325-1327.
12. Press WH, Teukolsky S, Vetterling WT, Flannery BP. *Numerical Recipes in C*. Cambridge, Mass: Cambridge University Press; 1992.
13. Neal DR, Topa DM, Copland J. The effect of lenslet resolution on the accuracy of ocular wavefront measurements. *Society for Photo-Optical Instrumentation Engineers*. 2001;4245:78-91.
14. Neal DR, Copland RJ, Neal DA. Shack-Hartmann wavefront sensor precision and accuracy. *Society for Photo-Optical Instrumentation Engineers*. 2002;4779:148-160.
15. Neal DR, Baer CD, Topa DM. Errors in Zernike transformations and non-modal reconstruction methods. *J Refract Surg*. 2005;21:S558-S562.