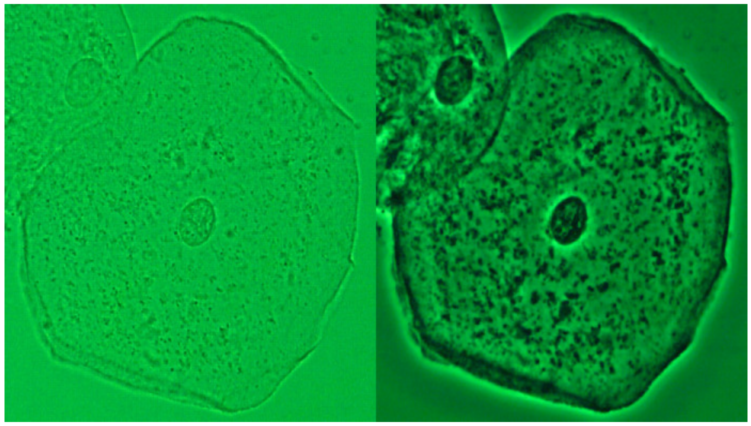
Zernike Polynomials

- Fitting irregular and non-rotationally symmetric surfaces over a circular region.
- Atmospheric Turbulence.
- Corneal Topography
- Interferometer measurements.
- Ocular Aberrometry

Background

- The mathematical functions were originally described by Frits Zernike in 1934.
- They were developed to describe the diffracted wavefront in phase contrast imaging.
- Zernike won the 1953 Nobel Prize in Physics for developing Phase Contrast Microscopy.

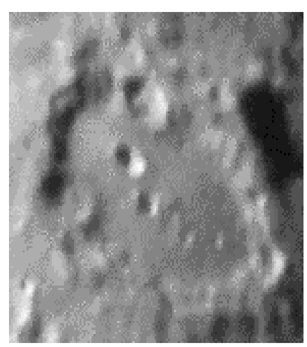
Phase Contrast Microscopy



Transparent specimens leave the amplitude of the illumination virtually unchanged, but introduces a change in phase.

Applications

- Typically used to fit a wavefront or surface sag over a circular aperture.
- Astronomy - fitting the wavefront entering a telescope that has been distorted by atmospheric turbulence.
- Diffraction Theory - fitting the wavefront in the exit pupil of a system and using Fourier transform properties to determine the Point Spread Function.

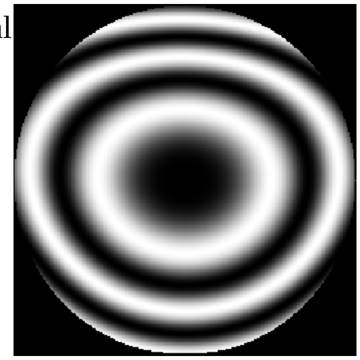


Source: http://salzgeber.at/astro/moon/seeing.html

Applications

- Ophthalmic Optics - fitting corneal topography and ocular wavefront data.

- Optical Testing - fitting reflected and transmitted wavefront data measured interferometically.

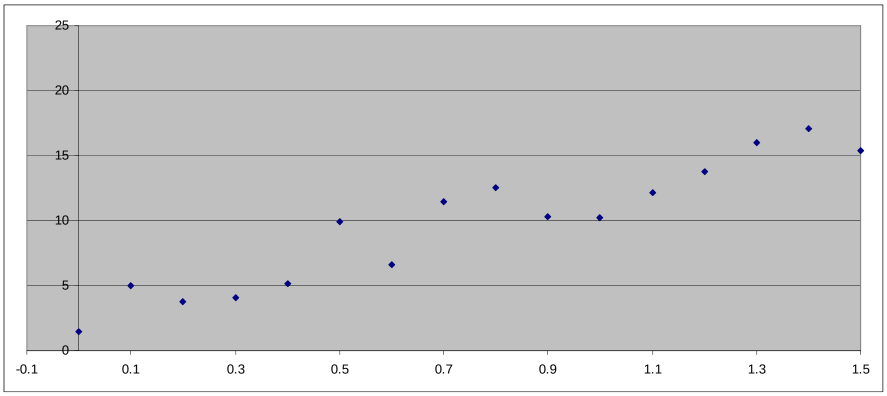


Surface Fitting

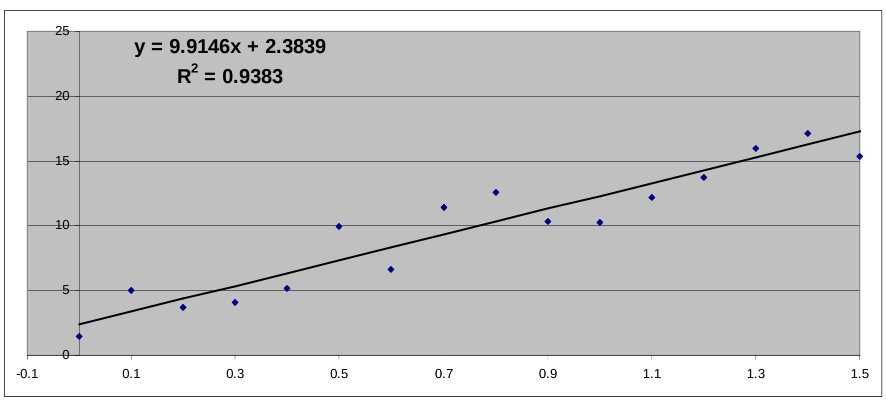
- Reoccurring Theme: Fitting a complex, non-rotationally symmetric surfaces (phase fronts) over a circular domain.

- Possible goals of fitting a surface:
  - Exact fit to measured data points?
  - Minimize “Error” between fit and data points?
  - Extract Features from the data?

1D Curve Fitting

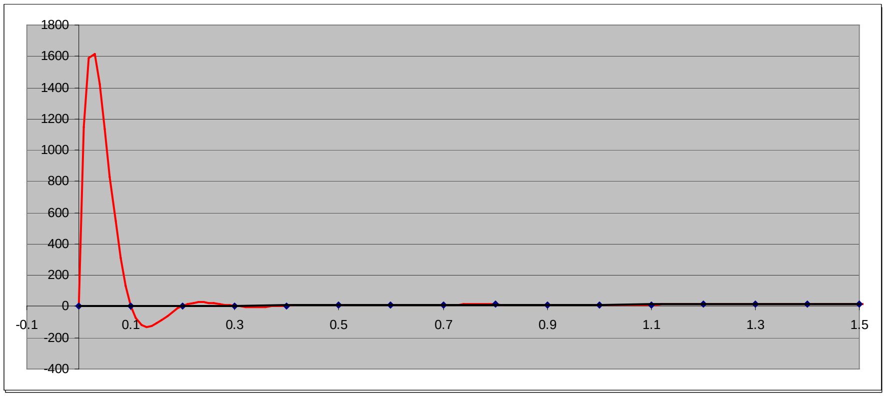


Low-order Polynomial Fit



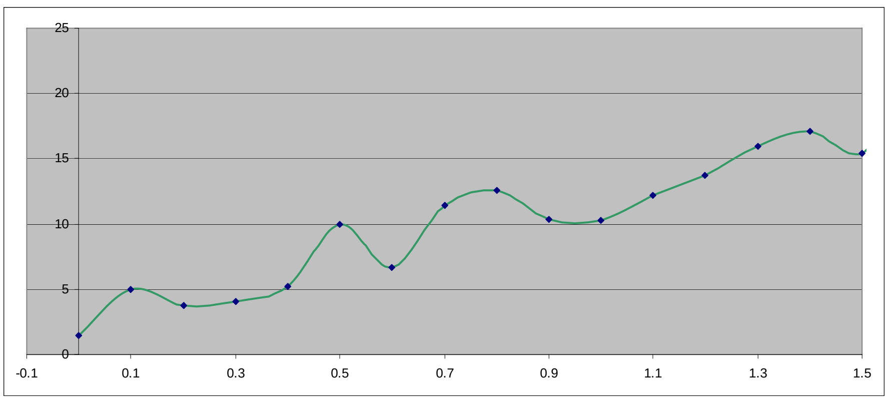
In this case, the error is the vertical distance between the line and the data point. The sum of the squares of the error is minimized.

High-order Polynomial Fit



$$y = a_0 + a_1x + a_2x^2 + \ldots a_{16}x^{16}$$

Cubic Splines



Piecewise definition of the function.

Fitting Issues

- Know your data. Too many terms in the fit can be numerically unstable and/or fit noise in the data. Too few terms may miss real trends in the surface.
- Typically want "nice" properties for the fitting function such as smooth surfaces with continuous derivatives. For example, cubic splines have continuous first and second derivatives.
- Typically want to represent many data points with just a few terms of a fit. This gives compression of the data, but leaves some residual error. For example, the line fit represents 16 data points with two numbers: a slope and an intercept.

Why Zernikes?

- Zernike polynomials have nice mathematical properties.
  - They are orthogonal over the continuous unit circle.
  - All their derivatives are continuous.
  - They efficiently represent common errors (e.g. coma, spherical aberration) seen in optics.
  - They form a complete set, meaning that they can represent arbitrarily complex continuous surfaces given enough terms.

Orthogonal Functions

- Orthogonal functions are sets of surfaces which have some nice mathematical properties for surface fitting.
- These functions satisfy the property

$$\iint_A V_i(x,y)V_j(x,y)dxdy = \begin{cases} C_j & i = j \\ 0 & \text{Otherwise} \end{cases}$$

where C_j is a constant for a given j

Orthogonality and Expansion Coefficients

$$W(x,y) = \sum_i a_i V_i(x,y) \quad \text{Linear Expansion}$$

$$W(x,y)V_j(x,y) = \sum_i a_i V_i(x,y)V_j(x,y)$$

$$\iint_A W(x,y)V_j(x,y)dxdy = \sum_i a_i \iint_A V_i(x,y)V_j(x,y)dxdy$$

$$a_j = \frac{1}{C_j}\iint_A W(x,y)V_j(x,y)dxdy$$

Orthogonality - 1D Example

Consider the integral

$$\int_0^{2\pi} \sin(mx)\cos(m'x)dx \quad \text{where m and m' are integers}$$

$$= \frac{1}{2}\int_0^{2\pi}[\sin(m+m')x + \sin(m-m')x]dx$$

$$= -\frac{1}{2}\left[\frac{\cos(m+m')x}{m+m'} + \frac{\cos(m-m')x}{m-m'}\right]_{x=0}^{x=2\pi}$$

$$= 0$$

Orthogonality - 1D Example

Two sine terms

$$\int_0^{2\pi} \sin(mx)\sin(m'x)dx \quad \text{where m and m' are integers}$$

$$= -\frac{1}{2}\int_0^{2\pi}[\cos(m+m')x - \cos(m-m')x]dx$$

$$= -\frac{1}{2}\left[\frac{\sin(m+m')x}{m+m'} - \frac{\sin(m-m')x}{m-m'}\right]_{x=0}^{x=2\pi}$$

$= 0$ if m ≠ m' !!!

Orthogonality - 1D Example

Two sine terms with m = m'

$$\int_0^{2\pi} \sin(mx)\sin(m'x)dx \quad \text{where m and m' are integers}$$

$$= \int_0^{2\pi} \sin^2(mx)dx$$

$$= \left[\frac{x}{2} - \frac{\sin(2mx)}{4m}\right]_{x=0}^{x=2\pi}$$

$$= \pi$$ Similar arguments for two cosine terms

Orthogonality - 1D Example

From the previous arguments, we can define

$$V_j = \begin{cases} \cos\left(\frac{j}{2}x\right) & j\,\text{even} \\ \sin\left(\frac{j+1}{2}x\right) & j\,\text{odd} \end{cases}$$

Note that when j = 0, V_0 = 1 and

$$\int_0^{2\pi} V_0(x)V_0(x)dx = \int_0^{2\pi} dx = 2\pi$$

so the constant $C_j = 2\pi$ for j = 0, and $C_j = \pi$ for all other values of j.

Orthogonality - 1D Example

The functions satisfy the orthogonality condition over [0, 2π]

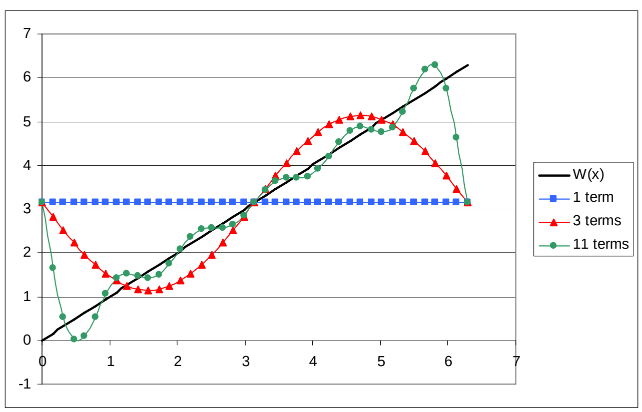
$$\int_0^{2\pi} V_i(x)V_j(x)dx = \begin{cases} (1+\delta_{i0})\pi & i=j \\ 0 & \text{Otherwise} \end{cases}$$

where δ is the Kronecker delta function defined as

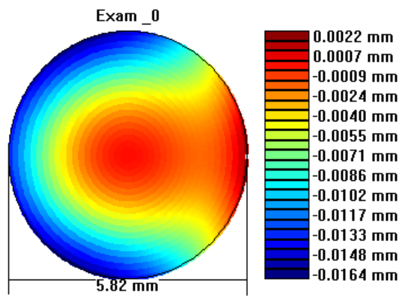
$$\delta_{ij} = \begin{cases} 1 & i=j \\ 0 & \text{Otherwise} \end{cases}$$

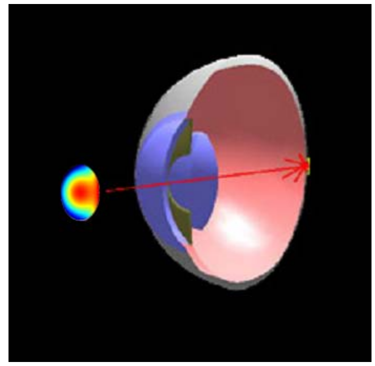
Fit to W(x) = x

j	a_j
0	π
1	-2
2	0
3	-1
4	0
5	-0.6666
6	0
7	-0.5
8	0
9	-0.4
10	0



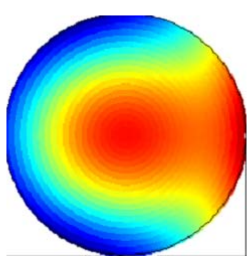
Extension to Two Dimensions



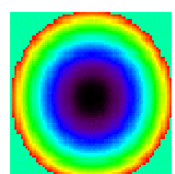


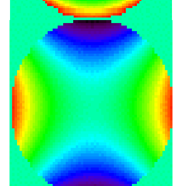
In many cases, wavefronts take on a complex shape defined over a circular region and we wish to fit this surface to a series of simpler components.

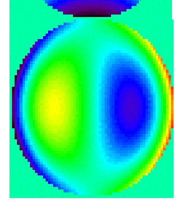
Wavefront Fitting



=

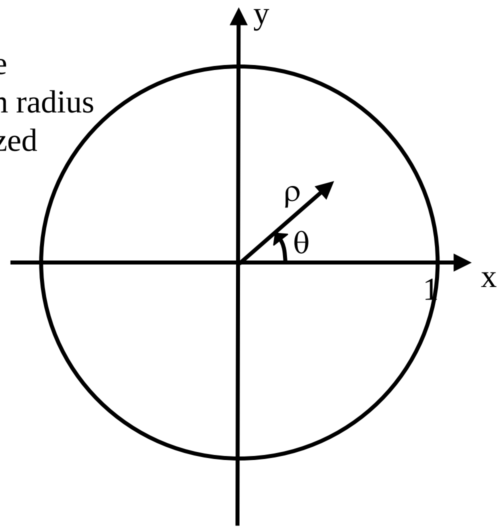
-0.003 x 

+ 0.002 x 

+ 0.001 x 

Unit Circle

Divide the real radial coordinate by the maximum radius to get a normalized coordinate ρ



Orthogonal Functions on the Unit Circle

- Taylor polynomials (i.e. 1, x, y, x^2, xy, y^2,….) are not orthogonal on the unit circle.

$$\int_0^{2\pi}\int_0^1 V_i(\rho,\theta)V_j(\rho,\theta)\rho\, d\rho\, d\theta = \begin{cases} C_j & i = j \\ 0 & \text{Otherwise} \end{cases}$$

where C_j is a constant for a given j

Many solutions, but let's try something with the form

$$V_i(\rho,\theta) = R_i(\rho)\Theta_i(\theta)$$

Orthogonal Functions on the Unit Circle

- The orthogonality condition separates into the product of two 1D integrals

$$\left[\int_0^1 R_i(\rho)R_j(\rho)\rho d\rho\right]\left[\int_0^{2\pi}\Theta_i(\theta)\,\Theta_j(\theta)d\theta\right]=\begin{cases} C_j & i=j \\ 0 & \text{Otherwise}\end{cases}$$

This has extra ρ in it, so we need different functions

This looks like the 1D Example, so sin(mθ) and cos(mθ) a possibility

ANSI Standard Zernikes

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

Double Index
n is radial order
m is azimuthal frequency

Normalization

Radial Component

Azimuthal Component

ANSI Z80.28-2004 Methods for Reporting Optical Aberrations of Eyes.

ANSI Standard Zernikes

$$R_n^{|m|}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s!\left[0.5(n+|m|)-s\right]!\left[0.5(n-|m|)-s\right]!}\rho^{n-2s}$$

only depends on |m| (i.e. same for both sine & cosine terms)

Constant that depends on n and m

Powers of ρ

The Radial polynomials satisfy the orthogonality equation

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

ANSI Standard Zernikes

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

constant that depends on n & m

Orthogonality

$$\int_0^{2\pi}\int_0^1 Z_{n'}^{m'}(\rho,\theta) Z_n^m(\rho,\theta)\rho\, d\rho\, d\theta = \begin{cases} C_{n,m} & \text{for } n=n'; m=m' \\ 0 & \text{Otherwise} \end{cases}$$

$$N_{n'}^{m'} N_n^m \int_0^{2\pi} \begin{Bmatrix} \cos(m'\theta) \\ \sin(-m'\theta) \end{Bmatrix} \begin{Bmatrix} \cos(m\theta) \\ \sin(-m\theta) \end{Bmatrix} d\theta$$

$$\times \int_0^1 R_{n'}^{m'}(\rho) R_n^m(\rho)\rho\, d\rho = \begin{cases} C_{n,m} & \text{for } n=n'; m=m' \\ 0 & \text{Otherwise} \end{cases}$$

Orthogonality

$$N_{n'}^{m'} N_n^m \left[\underbrace{\delta_{m'm}\left(1+\delta_{m0}\right)\pi}_{\text{This equals zero unless } m=m'}\right]\left[\underbrace{\frac{\delta_{n'n}}{2n+2}}_{\text{This equals zero unless } n=n'}\right] = \begin{cases} C_{n,m} & \text{for } n=n'; m=m' \\ \boxed{0 \quad \text{Otherwise}} \end{cases}$$

So this portion is satisfied

What happens when n=n’ and m=m’?

Orthogonality

When n=n' and m=m'

$$\left(N_n^m\right)^2\left[\left(1+\delta_{m0}\right)\pi\right]\left[\frac{1}{2n+2}\right]=C_{n,m}$$

$$\left[\frac{2n+2}{1+\delta_{m0}}\right]\left[\left(1+\delta_{m0}\right)\pi\right]\left[\frac{1}{2n+2}\right]=C_{n,m}$$

$$C_{n,m}=\pi$$

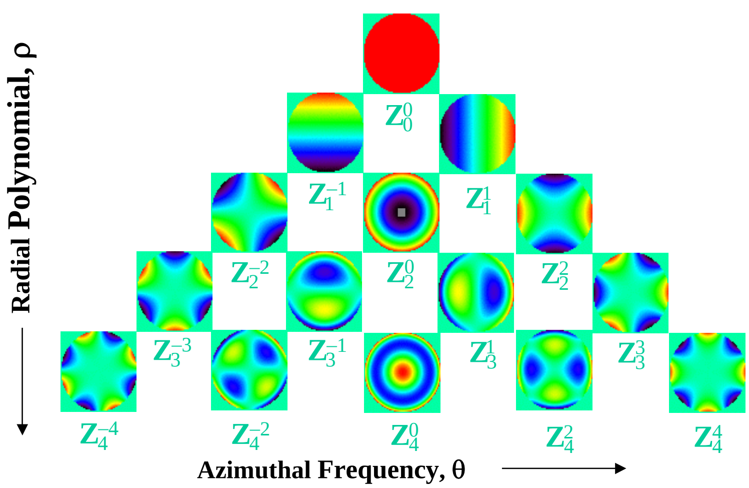
Orthogonality

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

The First Few Zernike Polynomials

$$Z_0^0(\rho,\theta) = 1$$
$$Z_1^{-1}(\rho,\theta) = \rho\sin\theta$$
$$Z_1^{1}(\rho,\theta) = \rho\cos\theta$$
$$Z_2^{-2}(\rho,\theta) = \sqrt{6}\left(\rho^2\sin 2\theta\right)$$
$$Z_2^{0}(\rho,\theta) = \sqrt{3}\left(2\rho^2 - 1\right)$$
$$Z_2^{2}(\rho,\theta) = \sqrt{6}\left(\rho^2\cos 2\theta\right)$$

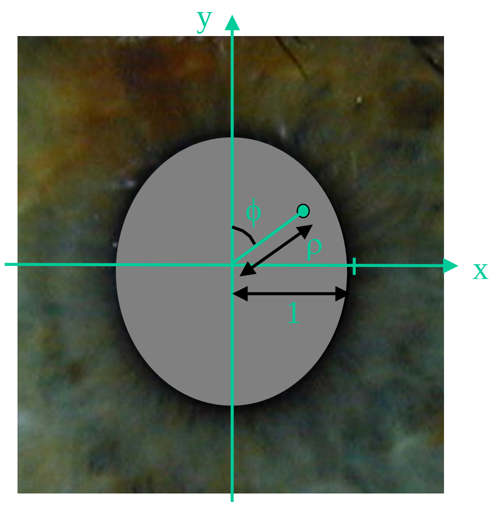
Zernike Polynomials



Caveats to the Definition of Zernike Polynomials

- At least six different schemes exist for the Zernike polynomials.
- Some schemes only use a single index number instead of n and m. With the single number, there is no unique ordering or definition for the polynomials, so different orderings are used.
- Some schemes set the normalization to unity for all polynomials.
- Some schemes measure the polar angle in the clockwise direction from the y axis.
- The expansion coefficients depend on pupil size, so the maximum radius used must be given.
- Some groups fit OPD, other groups fit Wavefront Error.
- Make sure which set is being given for a specific application.

Another Coordinate System



NON-STANDARD

Normalized Polar Coordinates:

$$\rho = \frac{r}{r_{max}}$$

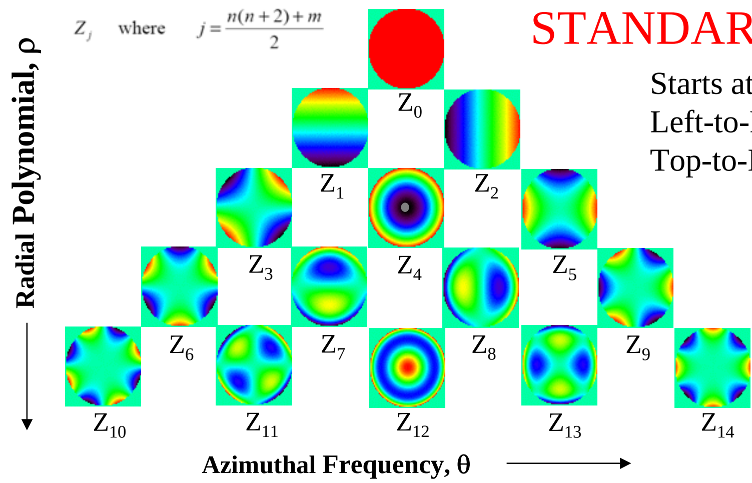
$$\phi = \tan^{-1}\left(\frac{x}{y}\right)$$

ρ ranges from [0, 1]
ϕ ranges from [-180°, 180°]

Zernike Polynomials - Single Index

ANSI STANDARD

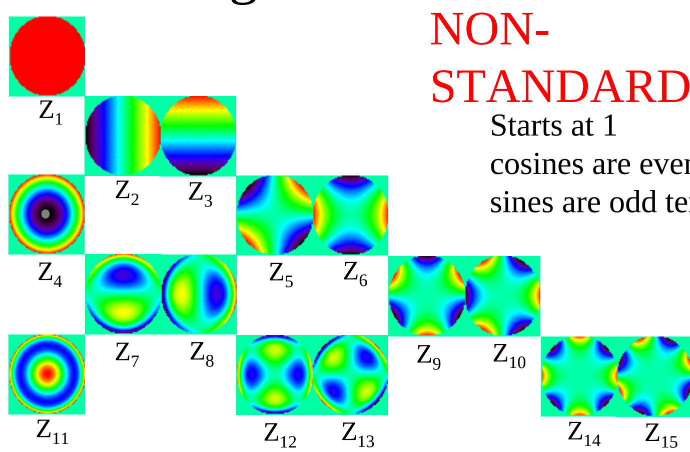
Starts at 0
Left-to-Right
Top-to-Bottom



Other Single Index Schemes

NON-STANDARD

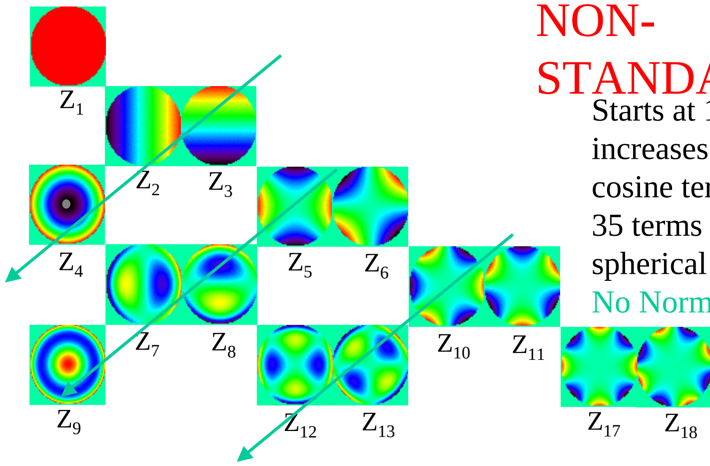
Starts at 1
cosines are even terms
sines are odd terms



Noll, RJ. Zernike polynomials and atmospheric turbulence. J Opt Soc Am 66; 207-211 (1976).

Also Zemax “Standard Zernike Coefficients”

Other Single Index Schemes



NON-STANDARD

Starts at 1
increases along diagonal
cosine terms first
35 terms plus two extra spherical aberration terms
No Normalization!!!

Zemax "Zernike Fringe Coefficients"

Also, Air Force or University of Arizona

Other Single Index Schemes

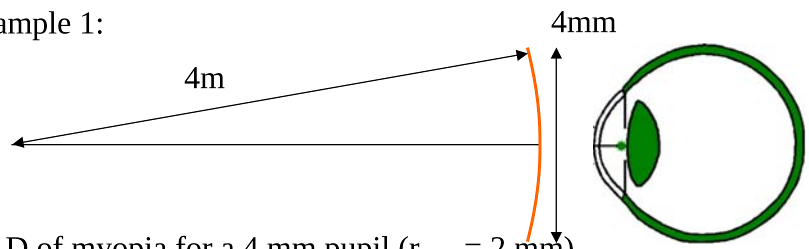
NON-STANDARD

- Born & Wolf
- Malacara
- Others??? Plus mixtures of non-normalized, coordinate systems.

Use two indices n, m to unambiguously define polynomials.
Use a single standard index if needed to avoid confusion.

Examples

Example 1:

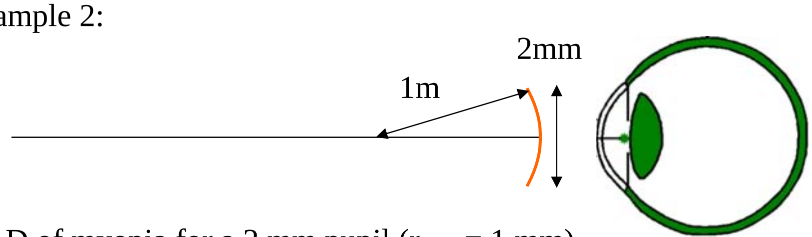


0.25 D of myopia for a 4 mm pupil (r_{max} = 2 mm)

$$W = \frac{r^2}{8000} = \frac{(2\rho)^2}{8000} = \frac{\rho^2}{2000} = \frac{1}{4000} Z_0^0(\rho,\theta) + \frac{1}{4000\sqrt{3}} Z_2^0(\rho,\theta)$$

Examples

Example 2:



1.00 D of myopia for a 2 mm pupil (r_{max} = 1 mm)

$$W = \frac{r^2}{2000} = \frac{\rho^2}{2000} = \frac{1}{4000} Z_0^0(\rho,\theta) + \frac{1}{4000\sqrt{3}} Z_2^0(\rho,\theta)$$

Same Zernike Expansion as Example 1, but different r_{max}.

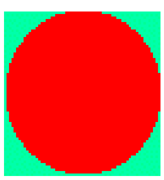
Always need to give pupil size with Zernike coefficients!!

RMS Wavefront Error

- RMS Wavefront Error is defined as

$$RMS_{WFE} = \sqrt{\frac{\iint (W(\rho,\phi))^2 \rho d\rho d\phi}{\iint \rho d\rho d\phi}} = \sqrt{\sum_{n>1,\text{all } m} \left(a_{n,m}\right)^2}$$

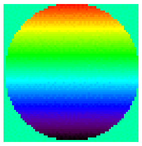
Zeroth Order Zernike Polynomials



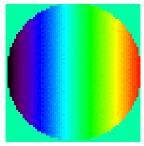
Z_0^0

This term is called Piston and is usually ignored. The surface is constant over the entire circle, so no error or variance exists.

First Order Zernike Polynomials



$\mathbf{Z}_1^{-1}$



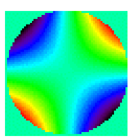
$\mathbf{Z}_1^{1}$

These terms represent a tilt in the wavefront.

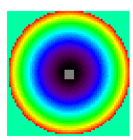
Combining these terms results in a general equation for a plane, thus by changing the coefficients, a plane at any orientation can be created. This rotation of the pattern is true for the sine/cosine pairs of Zernikes

$$a_{1-1}Z_1^{-1}(\rho,\theta)+a_{11}Z_1^{1}(\rho,\theta)$$
$$=a_{1-1}\rho\sin\theta+a_{11}\rho\cos\theta$$
$$=a_{1-1}\frac{y}{r_{max}}+a_{11}\frac{x}{r_{max}}$$

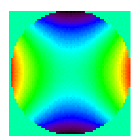
Second Order Zernike Polynomials



$\mathbf{Z}_2^{-2}$



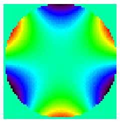
$\mathbf{Z}_2^{0}$



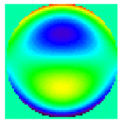
$\mathbf{Z}_2^{2}$

These wavefronts are what you would expect from Jackson crossed cylinder J0 and J45 and a spherical lens. Thus, combining these terms gives any arbitrary spherocylindrical refractive error.

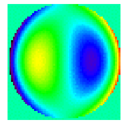
Third Order Zernike Polynomials



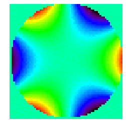
Z_3^{-3}



Z_3^{-1}



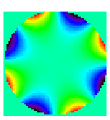
Z_3^{1}



Z_3^{3}

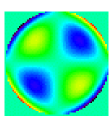
The inner two terms are coma and the outer two terms are trefoil. These terms represent asymmetric aberrations that cannot be corrected with convention spectacles or contact lenses.

Fourth Order Zernike Polynomials



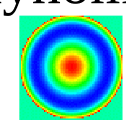
Z_4^{-4}

Quadroil



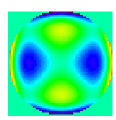
Z_4^{-2}

4th order Astigmatism



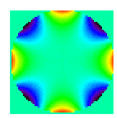
Z_4^{0}

Spherical Aberration



Z_4^{2}

4th order Astigmatism



Z_4^{4}

Quadroil

These terms represent more complex shapes of the wavefront. Spherical aberration can be corrected by aspheric lenses.

Discrete data

- Up to this point, the data has been continuous, so we can mathematically integrate functions to get expansion coefficients.
- Real-world data is sampled at discrete points.
- The Zernike polynomials are *not* orthogonal for discrete points, but for high sampling densities they are almost orthogonal.

Speed

- The long part of calculating Zernike polynomials is calculating factorial functions.

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

$$R_n^{|m|}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s!\left[0.5(n+|m|)-s\right]!\left[0.5(n-|m|)-s\right]!}\rho^{n-2s}$$

Speed

- Chong *et al.** developed a recurrence relationship that avoids the need for calculating the factorials.
- The results give a blazing fast algorithm for calculating Zernike expansion coefficents using orthogonality.

$$R_{p(q-4)}(r) = H_1 R_{pq}(r) + \left(H_2 + \frac{H_3}{r^2}\right) R_{p(q-2)}(r),$$

where the coefficients H_1, H_2 and H_3 are given by

$$H_1 = \frac{q(q-1)}{2} - qH_2 + \frac{H_3(p+q+2)(p-q)}{8},$$

$$H_2 = \frac{H_3(p+q)(p-q+2)}{4(q-1)} + (q-2),$$

$$H_3 = \frac{-4(q-2)(q-3)}{(p+q-2)(p-q+4)}.$$

*Pattern Recognition, 36;731-742 (2003).

Least Squares Fit

$$\begin{pmatrix} Z_0^0(x_1,y_1) & Z_1^{-1}(x_1,y_1) & Z_1^1(x_1,y_1) & \cdots & Z_n^m(x_1,y_1) \\ Z_0^0(x_2,y_2) & Z_1^{-1}(x_2,y_2) & Z_1^1(x_2,y_2) & \cdots & Z_n^m(x_2,y_2) \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ Z_0^0(x_N,y_N) & Z_1^{-1}(x_N,y_N) & Z_1^1(x_N,y_N) & \cdots & Z_n^m(x_N,y_N) \end{pmatrix} \begin{pmatrix} a_{00} \\ a_{1-1} \\ a_{11} \\ \vdots \\ a_{nm} \end{pmatrix} = \begin{pmatrix} f(x_1,y_1) \\ f(x_2,y_2) \\ \vdots \\ f(x_N,y_N) \end{pmatrix}$$

$$ZA = F$$

$$Z^T ZA = Z^T F$$

$$A = \left(Z^T Z\right)^{-1} Z^T F$$

Gram-Schmidt Orthogonalization

- Examines set of discrete data and creates a series of functions which are orthogonal over the data set.
- Orthogonality is used to calculate expansion coefficients.
- These surfaces can then be converted to a standard set of surfaces such as Zernike polynomials.

Advantages

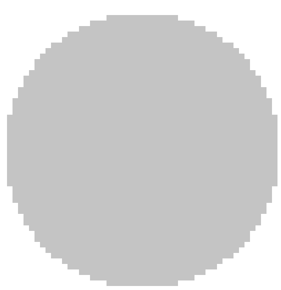
- Numerically stable, especially for low sampling density.

Disadvantages

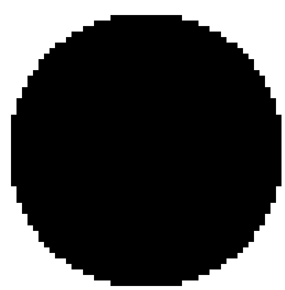
- Can be slow for high-order fits
- Orthogonal functions depend upon data set, so a new set needs to be calculated for every fit.

Elevation Fit Comparison

32 Orders or 560 total polynomials

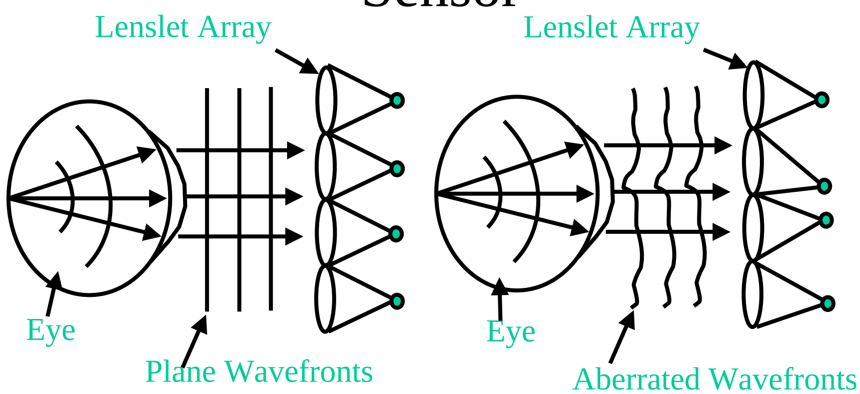


0.125 Seconds
Chong Algorithm



10 Seconds
Gram-Schmidt

Shack-Hartmann Wavefront Sensor



Perfect wavefronts give a uniform grid of points, whereas aberrated wavefronts distort the grid pattern.

Least Squares Fit

$$Z=\begin{bmatrix} dV_1(x_1,y_1)/dx & dV_2(x_1,y_1)/dx & \cdots & dV_J(x_1,y_1)/dx \\ dV_1(x_2,y_2)/dx & dV_2(x_2,y_2)/dx & \cdots & dV_J(x_2,y_2)/dx \\ \vdots & \vdots & \vdots & \vdots \\ dV_1(x_N,y_N)/dx & dV_2(x_N,y_N)/dx & \cdots & dV_J(x_N,y_N)/dx \\ dV_1(x_1,y_1)/dy & dV_2(x_1,y_1)/dy & \cdots & dV_J(x_1,y_1)/dy \\ dV_1(x_2,y_2)/dy & dV_2(x_2,y_2)/dy & \cdots & dV_J(x_2,y_2)/dy \\ \vdots & \vdots & \vdots & \vdots \\ dV_1(x_N,y_N)/dy & dV_2(x_N,y_N)/dy & \cdots & dV_J(x_N,y_N)/dy \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_J \end{bmatrix} = \begin{bmatrix} dW(x_1,y_1)/dx \\ dW(x_2,y_2)/dx \\ \vdots \\ dW(x_N,y_N)/dx \\ dW(x_1,y_1)/dy \\ dW(x_2,y_2)/dy \\ \vdots \\ dW(x_N,y_N)/dy \end{bmatrix}$$

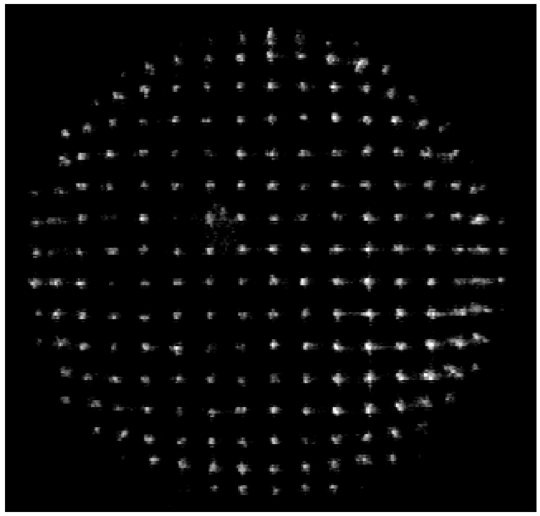
$$ZA = F$$

$$Z^T ZA = Z^T F$$

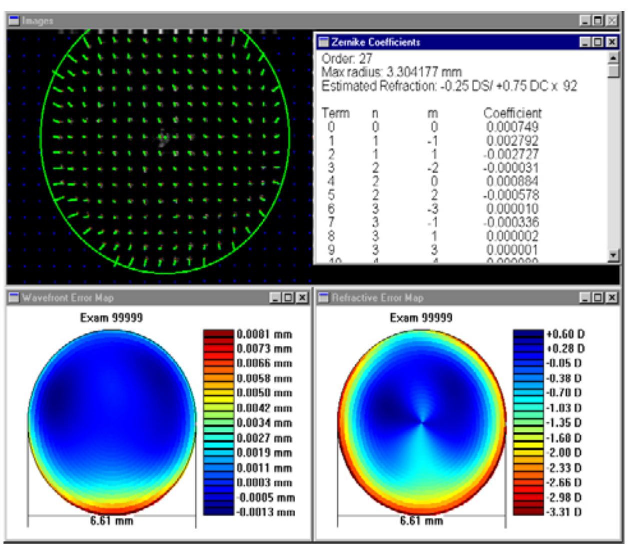
$$A = \left(Z^T Z\right)^{-1} Z^T F$$

Again, conceptually easy to understand, although this can be relatively slow for high order fits.

Example Image



Wavefront Reconstruction



PSF