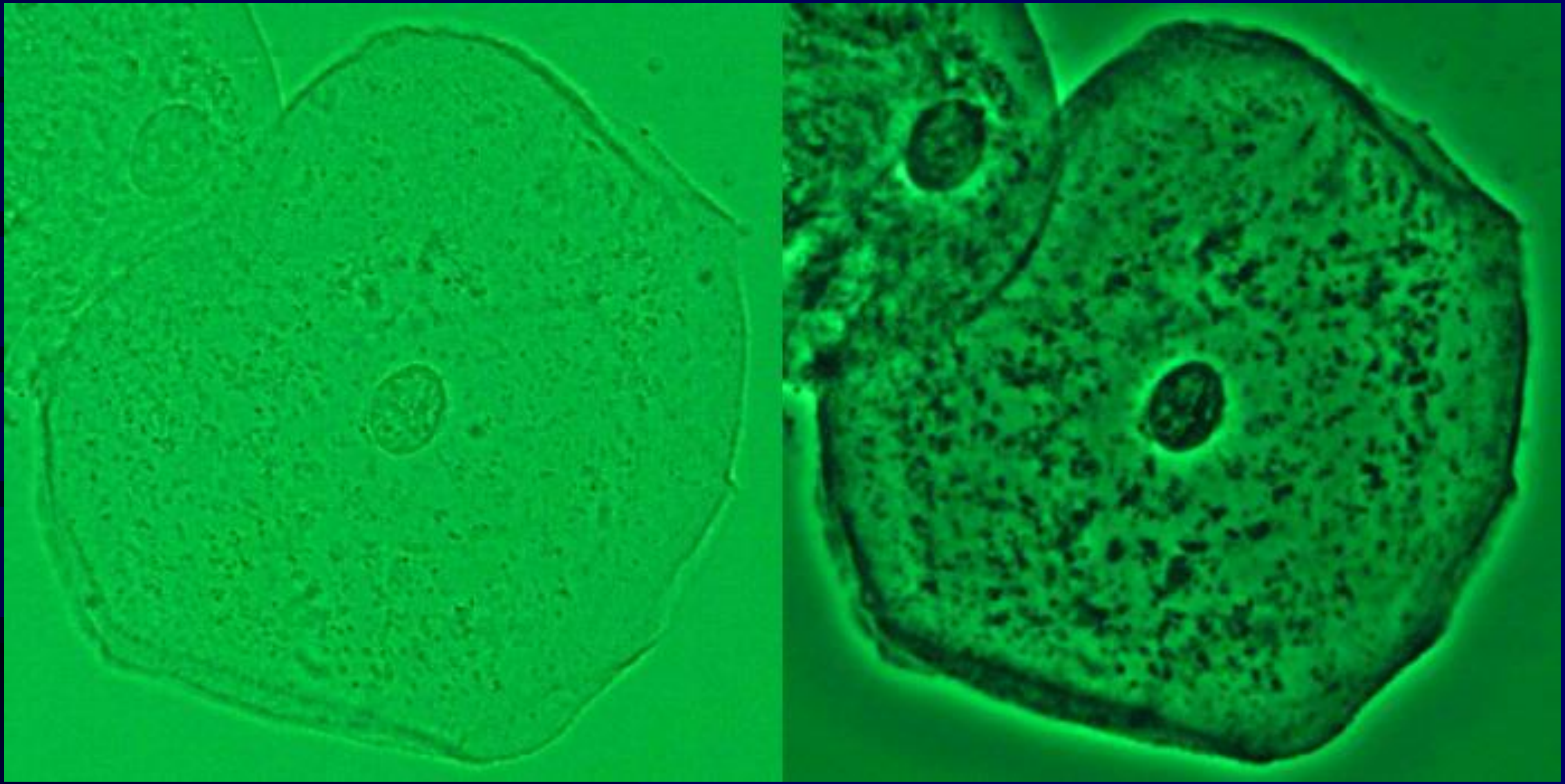
Review of Zernike polynomials and their use in describing the impact of misalignment in optical systems

Jim Schwiegerling, PhD
Ophthalmology & Vision Science
Optical Sciences
The University of Arizona

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

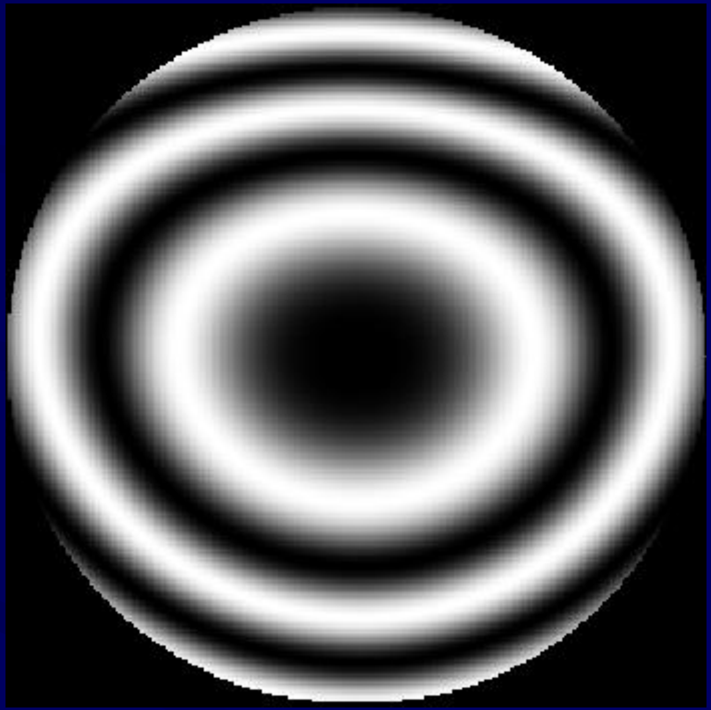


Transparent specimens leave the amplitude of the illumination virtually unchanged, but introduces a change in phase.

Applications

- Optical Design – describing complex shapes such as freeform surfaces and fabrication errors.

- Optical Testing - fitting reflected and transmitted wavefront data.

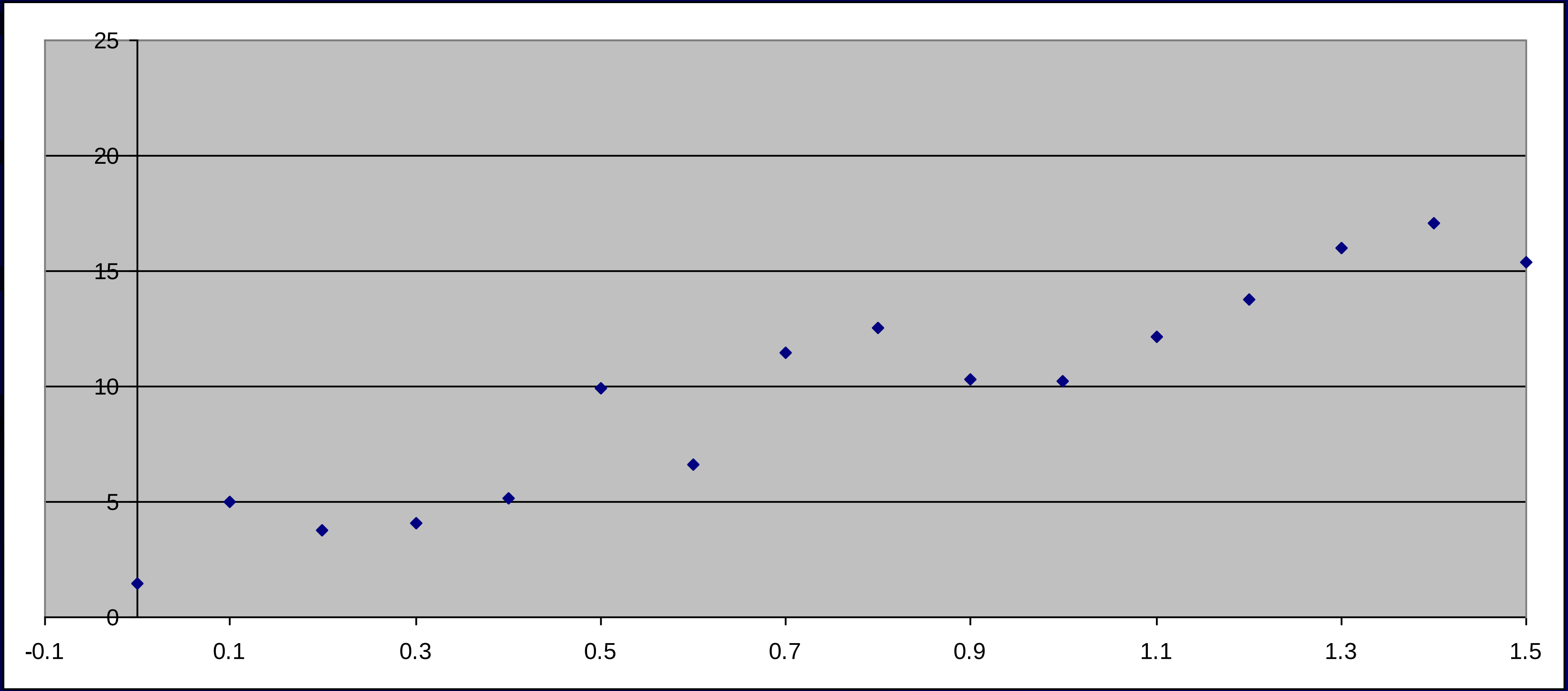


Surface Fitting

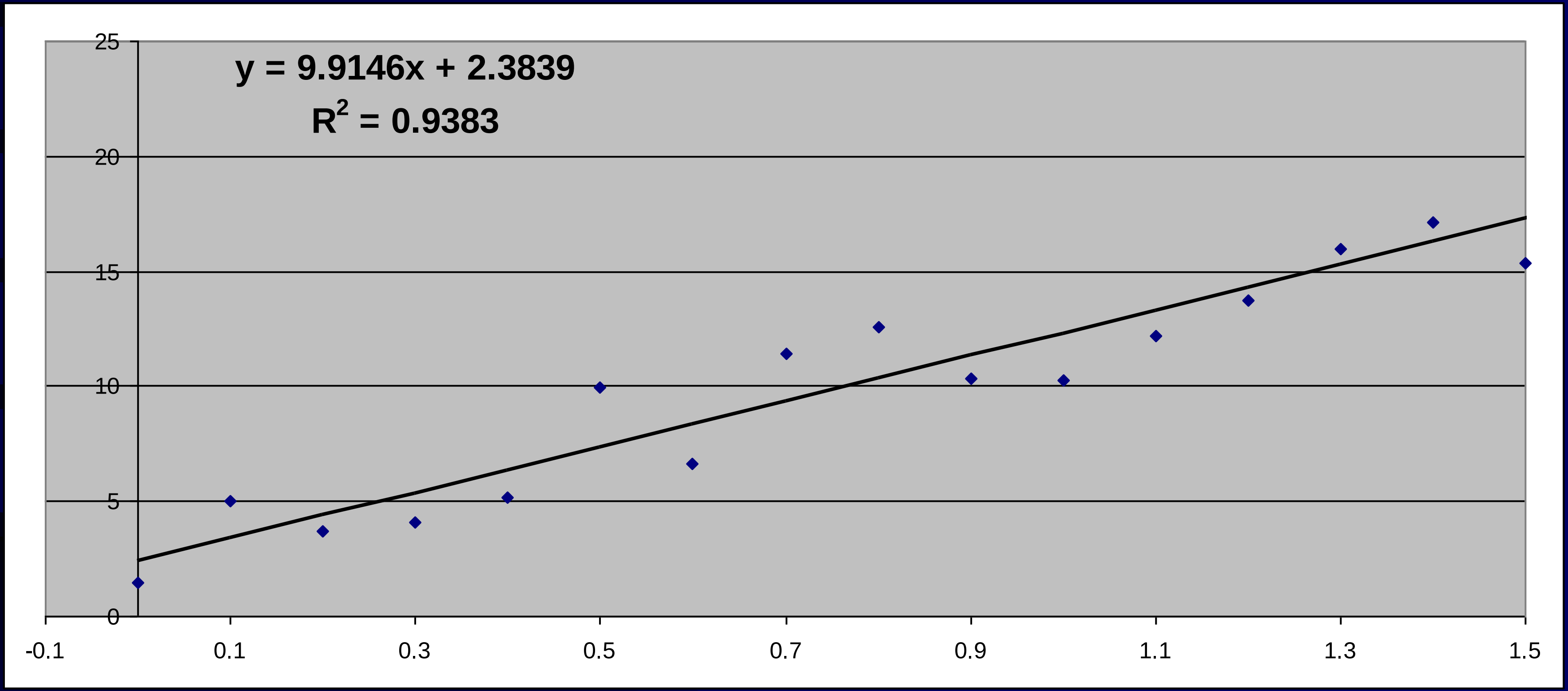
- Fitting a complex, non-rotationally symmetric surfaces (phase fronts) over a circular domain.

- 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

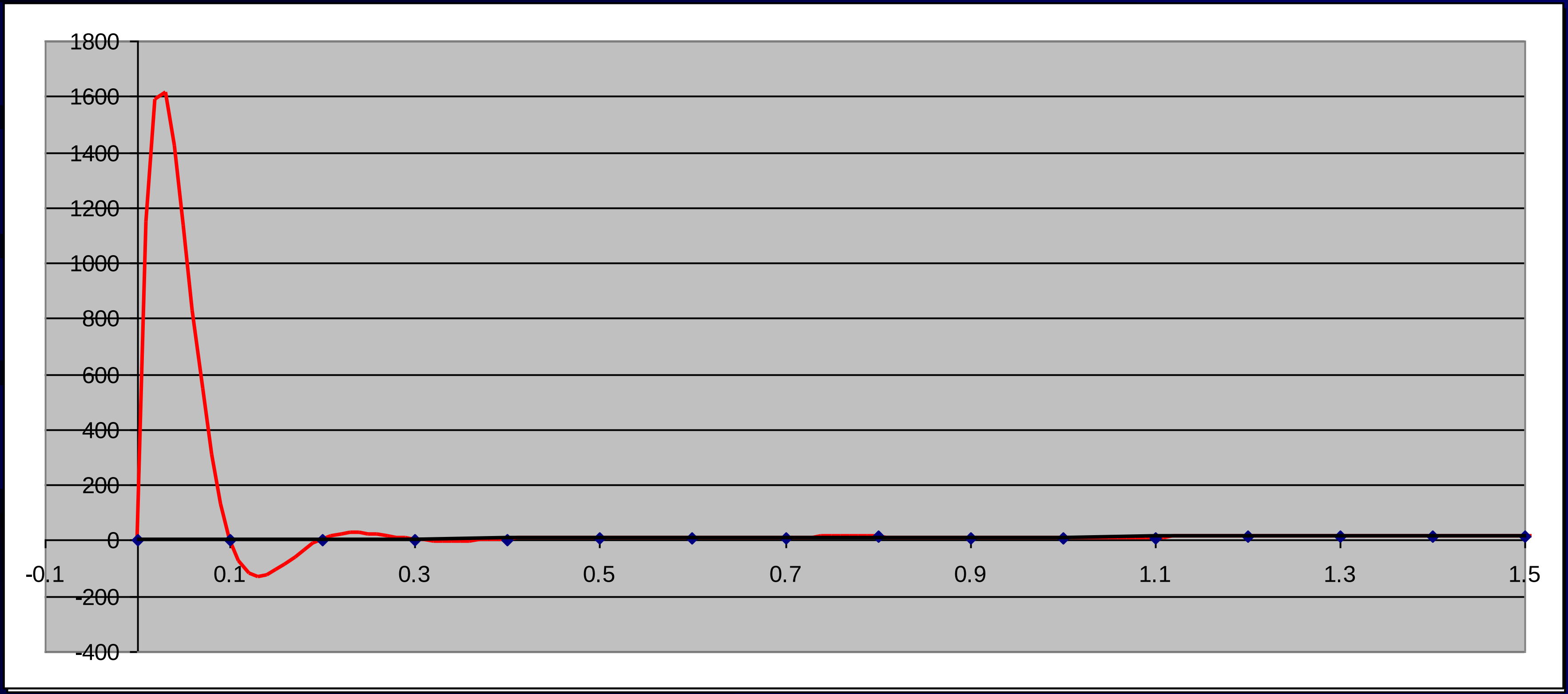


Low-order Polynomial Fit



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



$$y = a_0 + a_1x + a_2x^2 + \ldots a_{15}x^{15}$$

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.
- 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.

Orthogonality - Zernike

Orthogonality means we have an easy means of calculating expansion coefficients.

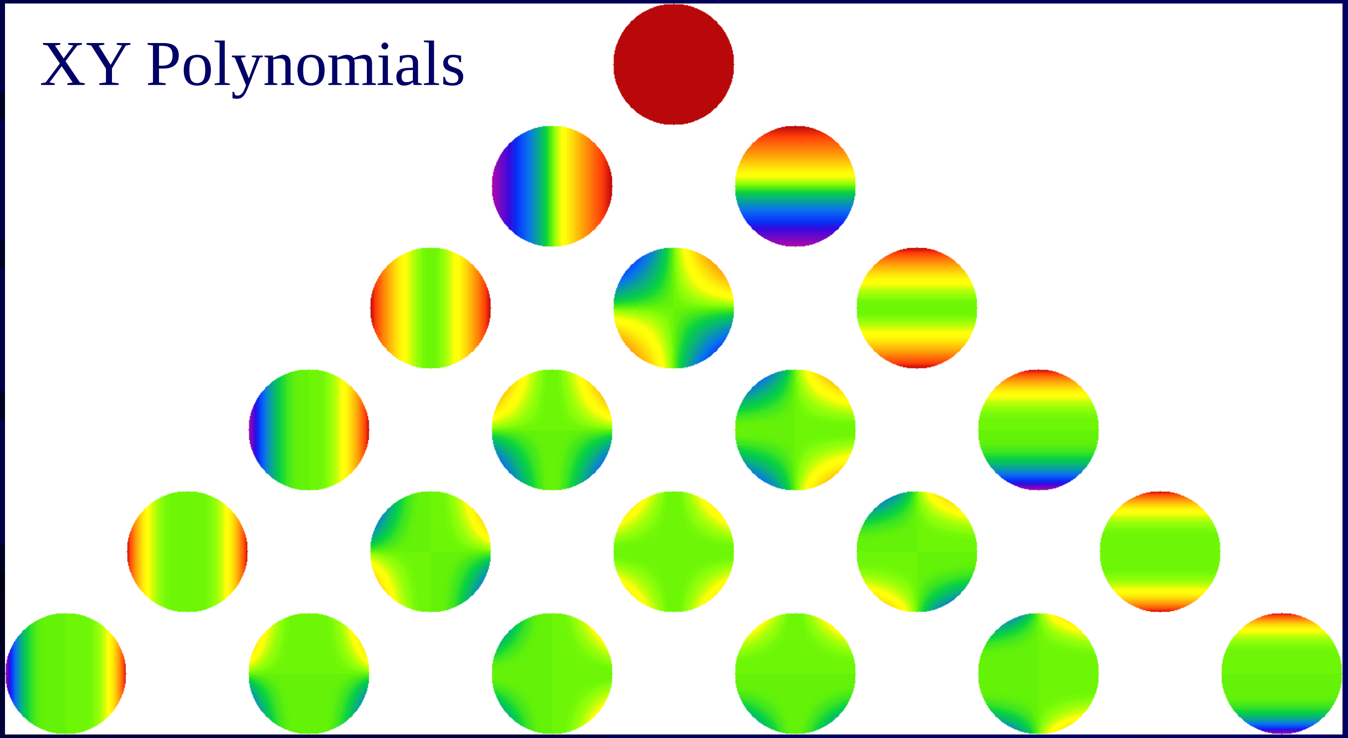
$$W(\rho,\theta)=\sum_{n,m} a_{nm} Z_n^m(\rho,\theta)$$

$$a_{nm}=\frac{1}{\pi}\int_0^{2\pi}\int_0^1 W(\rho,\theta) Z_n^m(\rho,\theta)\rho d\rho d\theta$$

Discrete Data

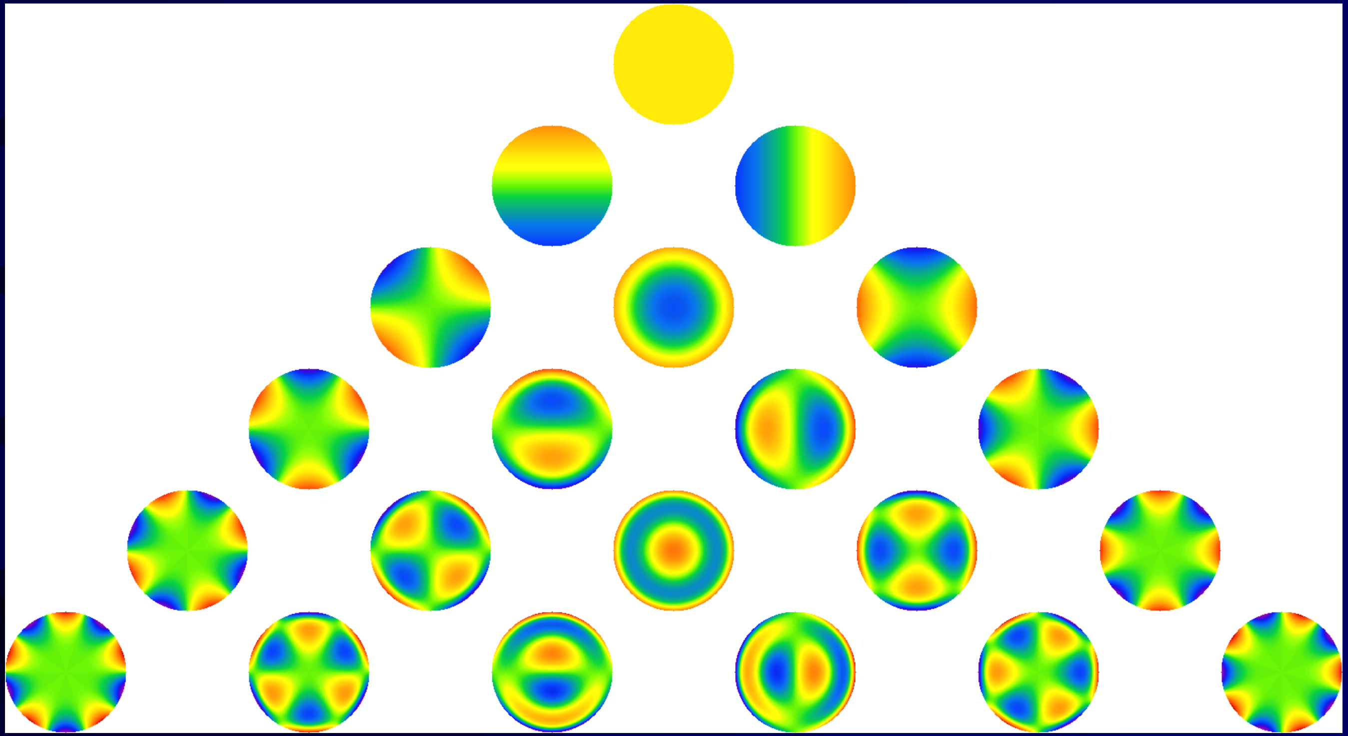
- Typically, we do not have a continuous description of $W(\quad)$, instead the data is discretely sampled (e.g. pixels on the digital sensor of an interferometer.
- In this case, we use matrix methods to find the expansion coefficients. This method is the same technique that we would use for fitting non-orthogonal functions.
- So why use orthogonal function?

XY Polynomials



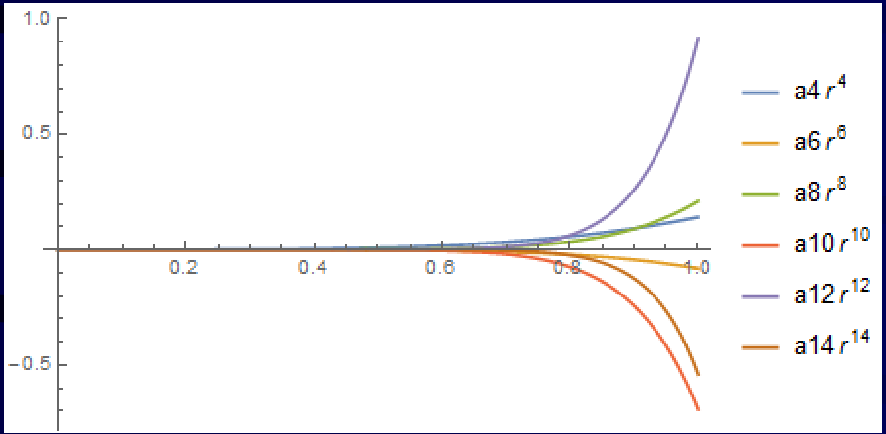
Most of the variation for the higher order terms occurs at the edges.

Zernike Polynomials

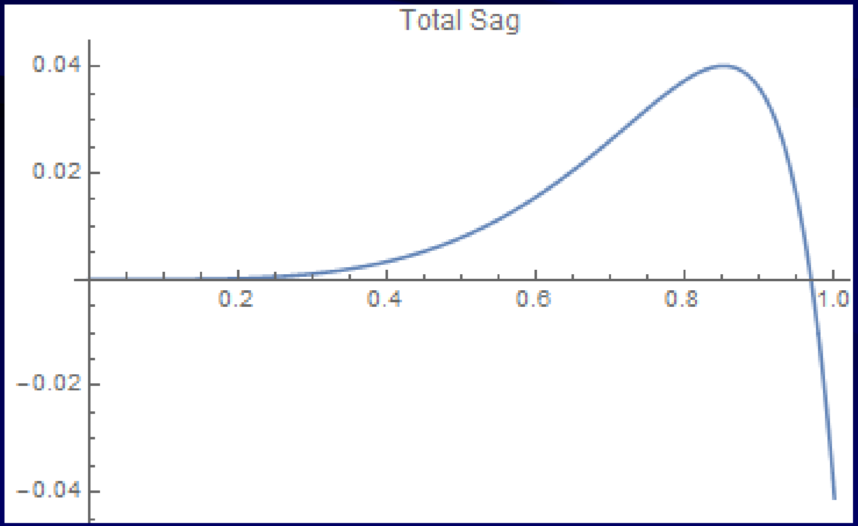


The variation oscillates in the radial and azimuthal direction.

Even Asphere

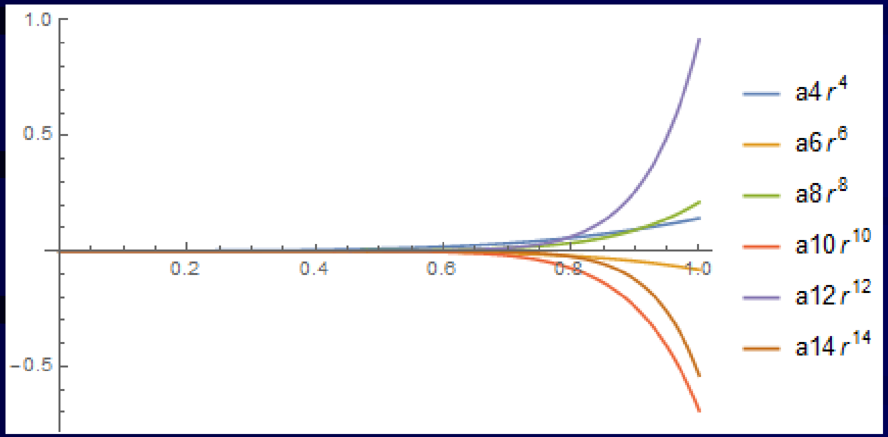


When the fitting functions have most of their change at the edge, then we need huge values of high order terms to represent small total sag.

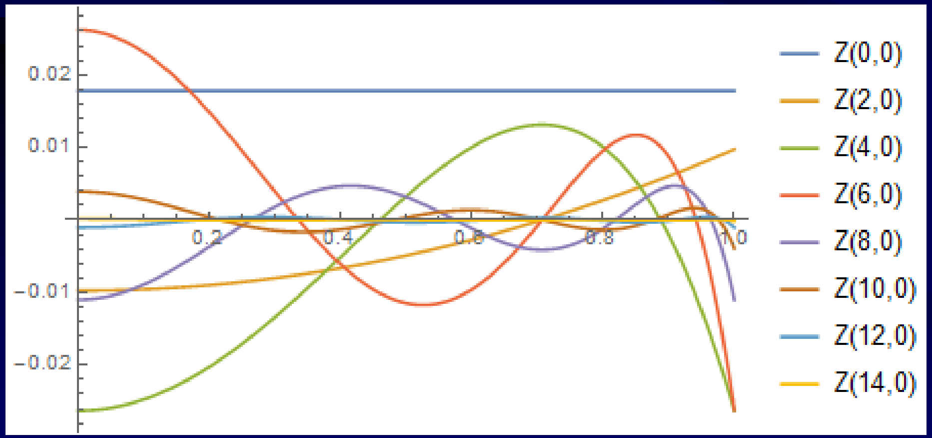


Numerical precision becomes an issue here as small changes to coefficients can cause large changes in total sag.

Zernike Equivalent



When the fitting functions have most of their change at the edge, then we need huge values of high order terms to represent small total sag.



Numerical precision becomes an issue here as small changes to coefficients can cause large changes in total sag.

Standards

ANSI Z80.28-2010

Methods for Reporting Optical Aberrations of Eyes

ISO 24157:2008

Ophthalmic optics and instruments — Reporting aberrations of the human eye

Normalized

ISO 14999-2:2005

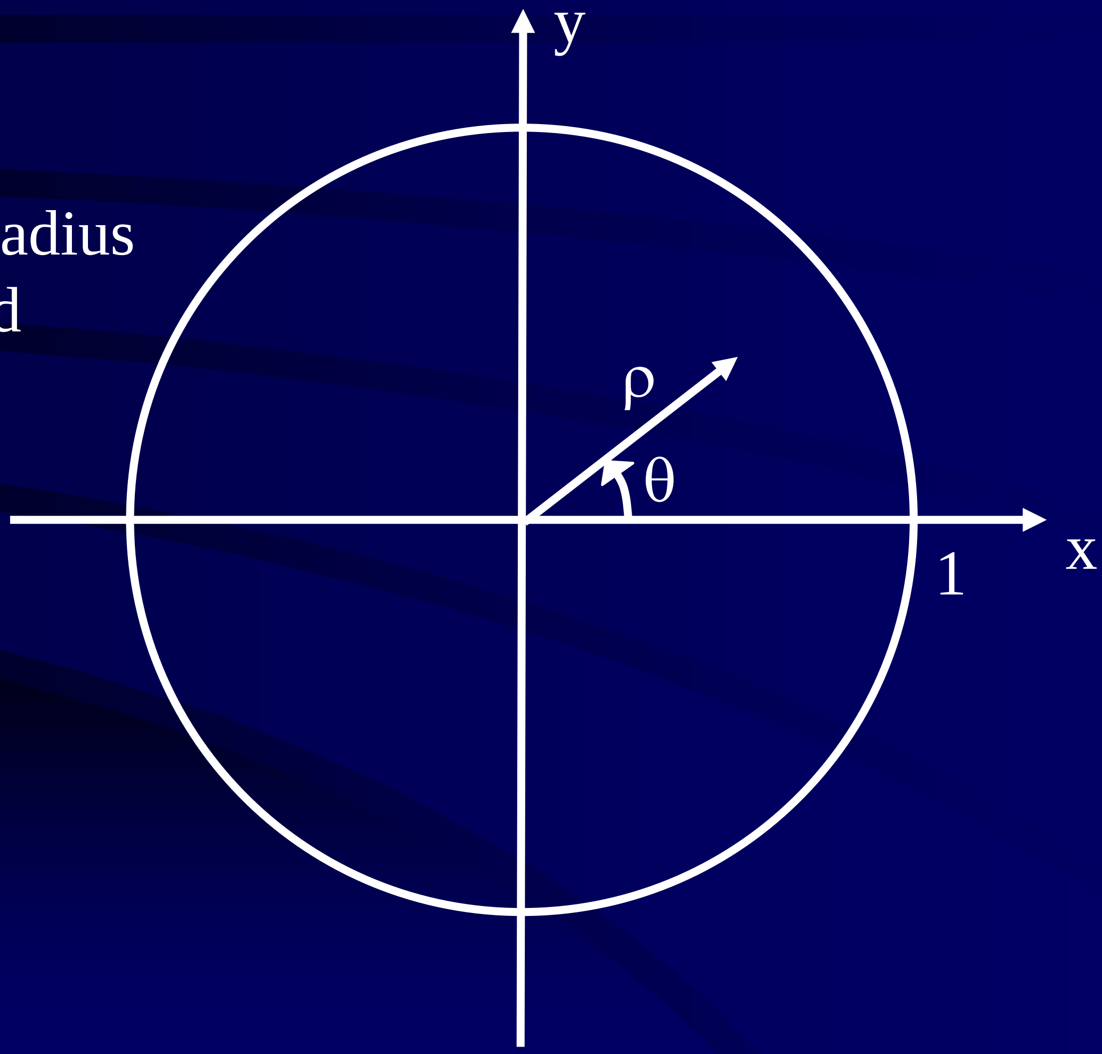
Optics and photonics — Interferometric measurement of optical elements and optical systems —

Part 2:
Measurement and evaluation techniques

Unnormalized

Unit Circle

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



ANSI Z80.28/ISO 24157 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/ISO 24157 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

ANSI Z80.28/ISO 24157 Zernikes

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

constant that
depends on n & m

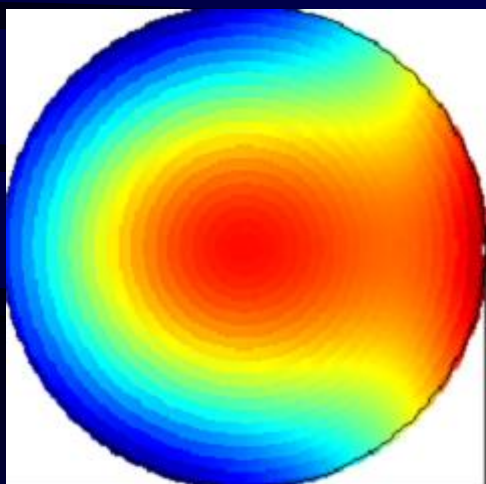
Wavefront Variance

- Wavefront variance and its square root RMS wavefront error are metrics of image quality.
- For the normalized Zernikes, the wavefront variance is trivial to calculate.

$$\sigma_W^2 = \sum_{n \geq 1} a_{nm}^2 - a_{00}$$

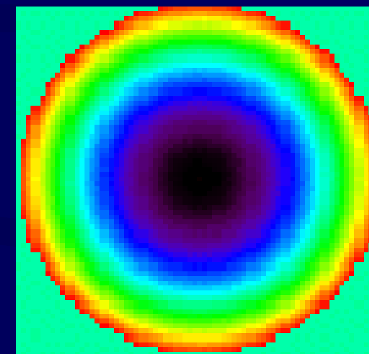
- Basically, the squared magnitude of each term describes its contribution to the variance.
- RMS Error is just square root of the variance.

Wavefront Fitting

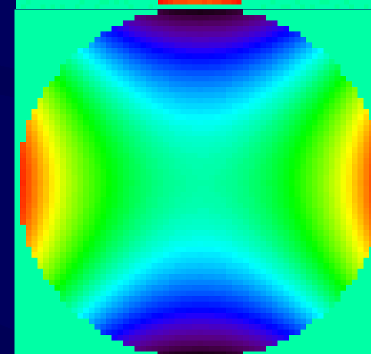


=

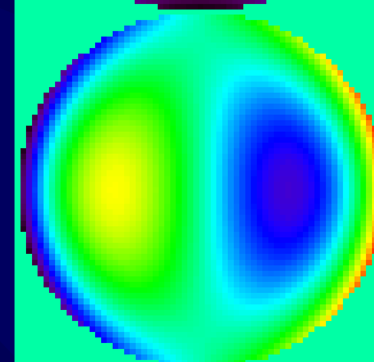
-0.003 x



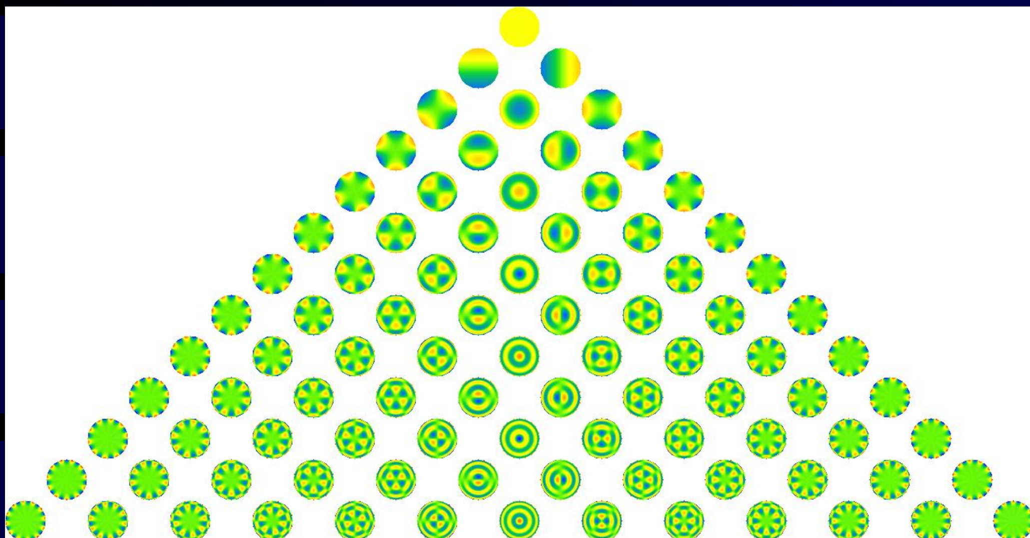
+ 0.002 x



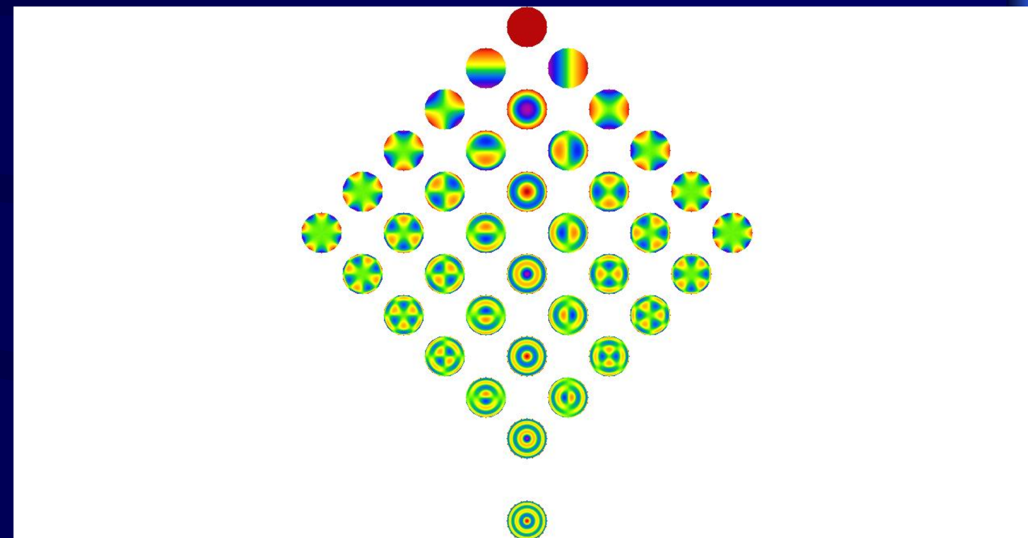
+ 0.001 x



Different Zernike Sets



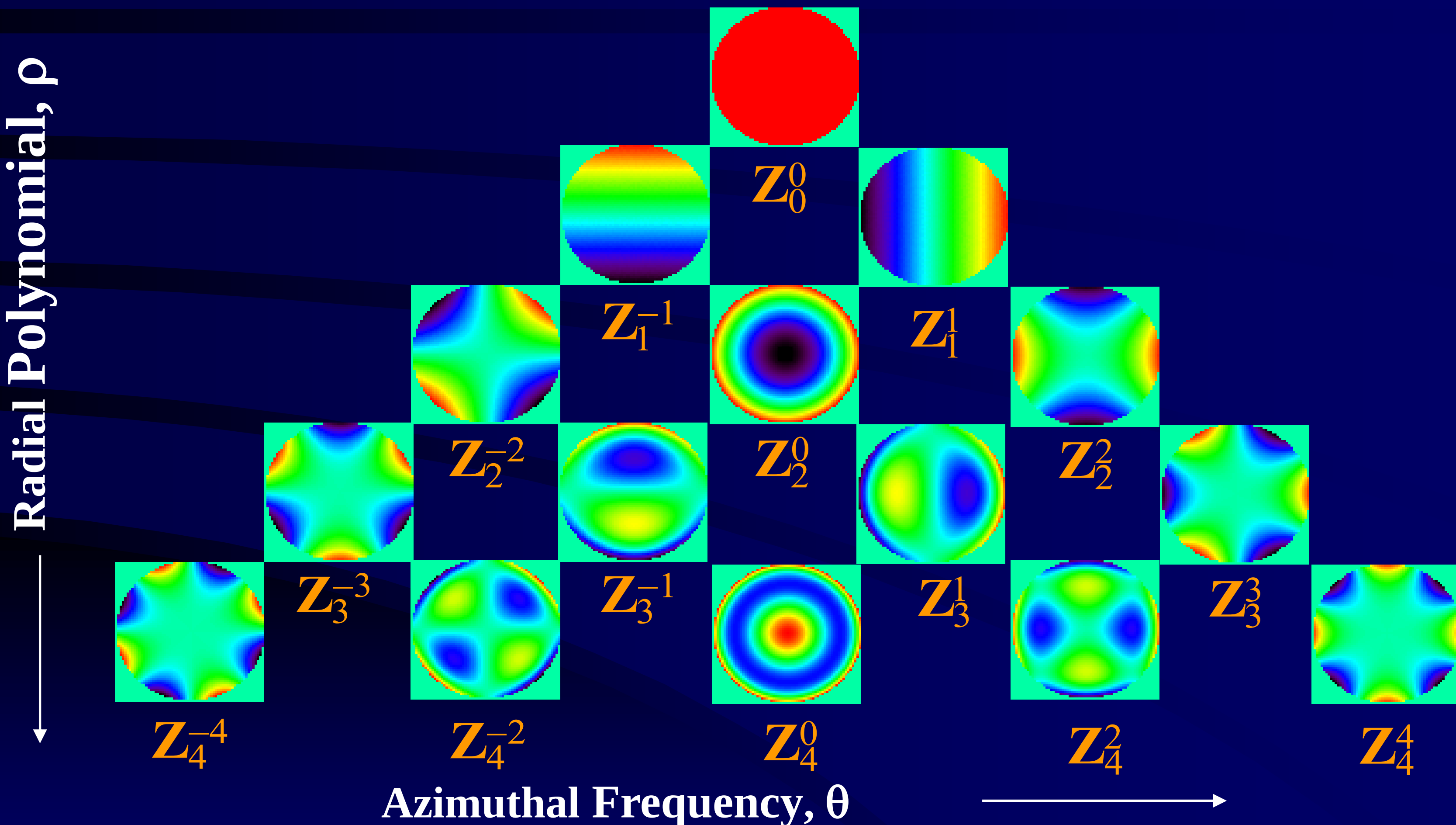
“Standard” or Noll Zernike



Fringe Zernike

The Fringe Zernike set is a subset of the Zernike polynomials.

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.
- Make sure which set is being given for a specific application.

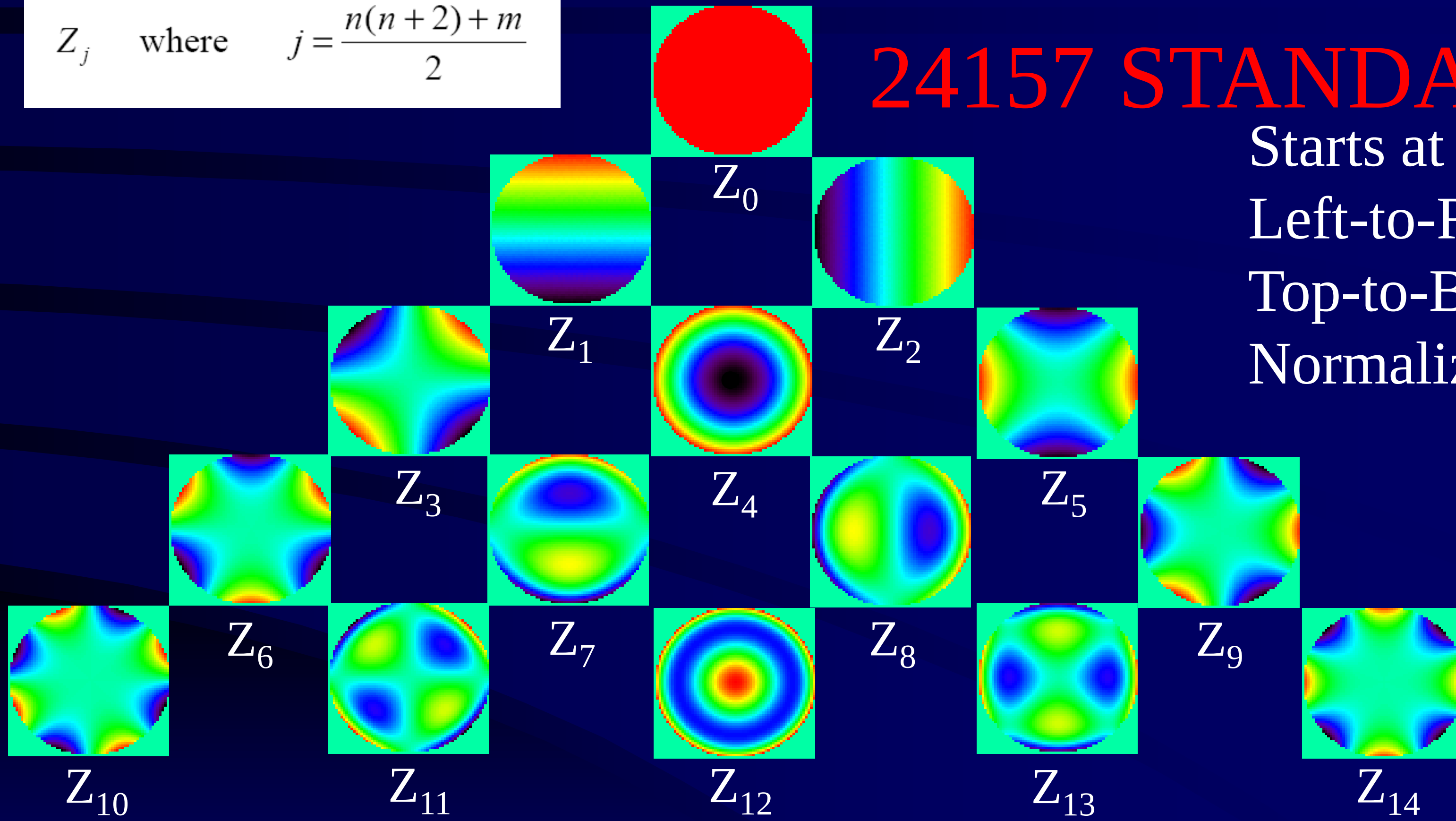
Zernike Polynomials - Single Index

$$Z_j \quad \text{where} \quad j = \frac{n(n+2)+m}{2}$$

ANSI Z80.28/ISO 24157 STANDARD

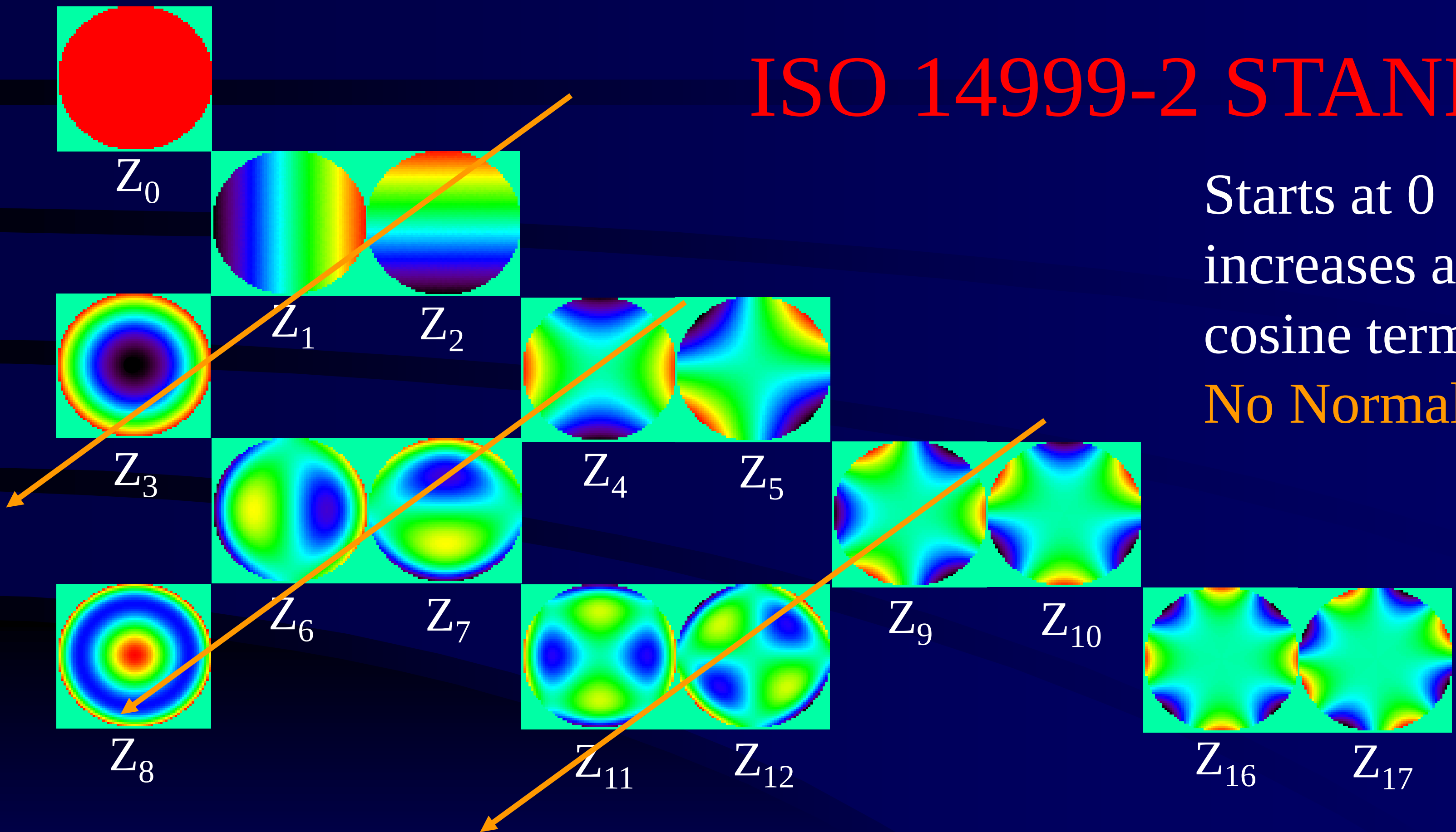
Starts at 0
Left-to-Right
Top-to-Bottom
Normalized

Radial Polynomial, ρ



Azimuthal Frequency, θ

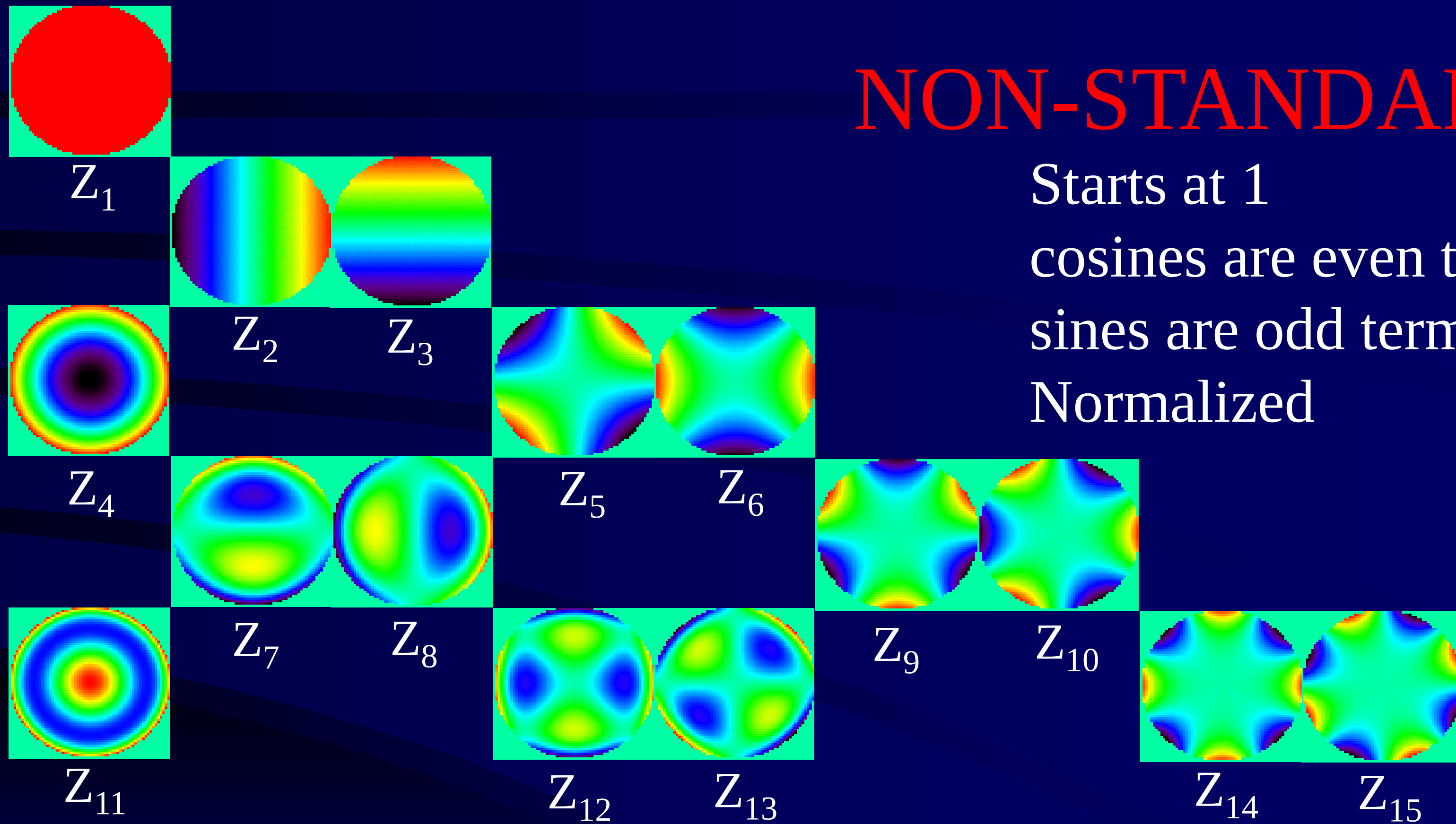
Other Single Index Schemes



ISO 14999-2 STANDARD

Starts at 0
increases along diagonal
cosine terms first
No Normalization

Other Single Index Schemes



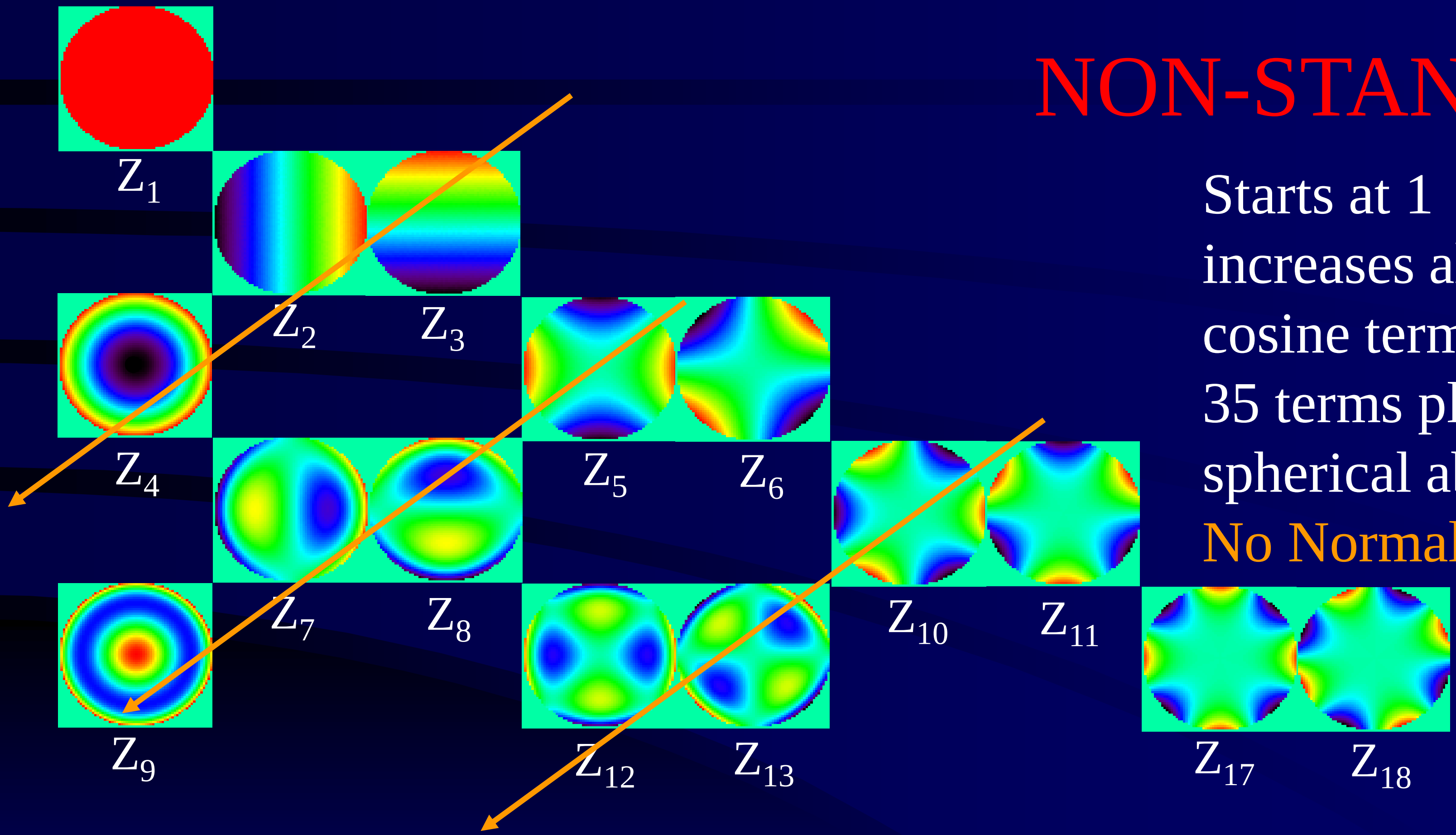
NON-STANDARD

Starts at 1
cosines are even terms
sines are odd terms
Normalized

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"
Code V Zernikes
Also, Air Force or University of Arizona

Other Single Index Schemes

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

NON-STANDARD

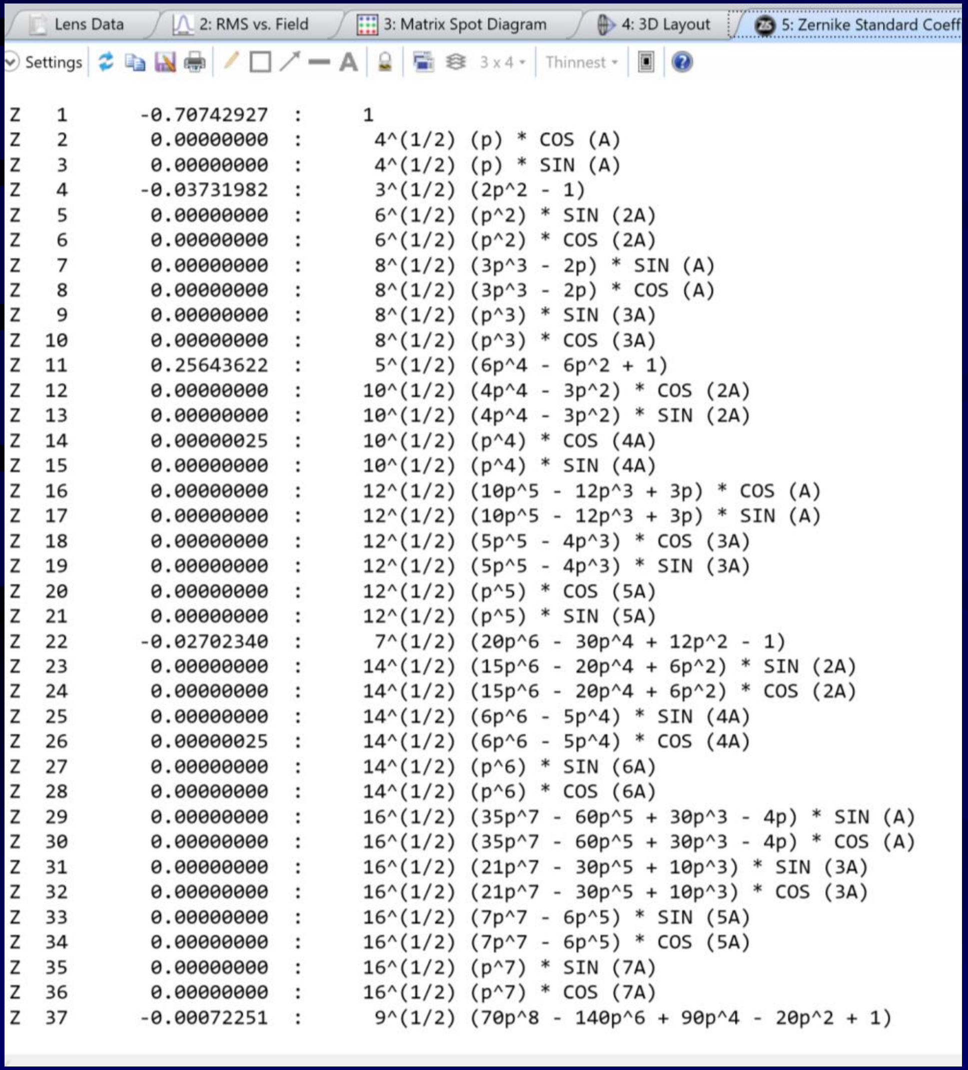
Use two indices n, m to unambiguously define polynomials.
Use a single standard index only if needed to avoid confusion.

Noll or Zemax “Standard” is closest to ANSI Z80.28/ISO 24157
Fringe set is closest to ISO 14999-2, but has limited terms.

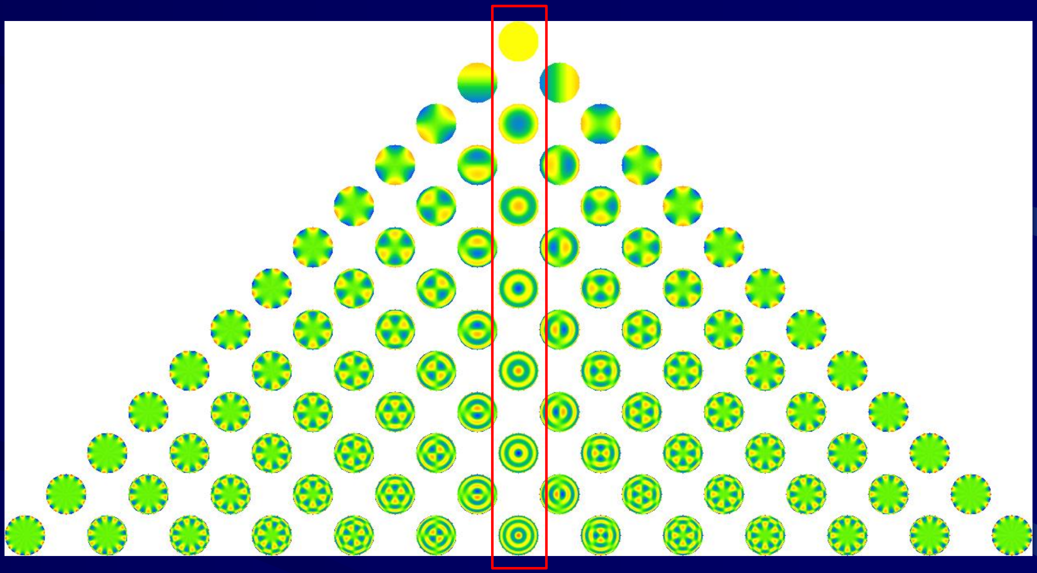
Summary

- Zernike polynomials are a useful set of functions for representing surface form and wavefronts on circular domains.
- The normalized version of the Zernikes gives a direct quality metric in the form of variance.
- Many different schemes and definitions exists, so be careful when comparing results from different sources.
- Two-index scheme is always unambiguous.

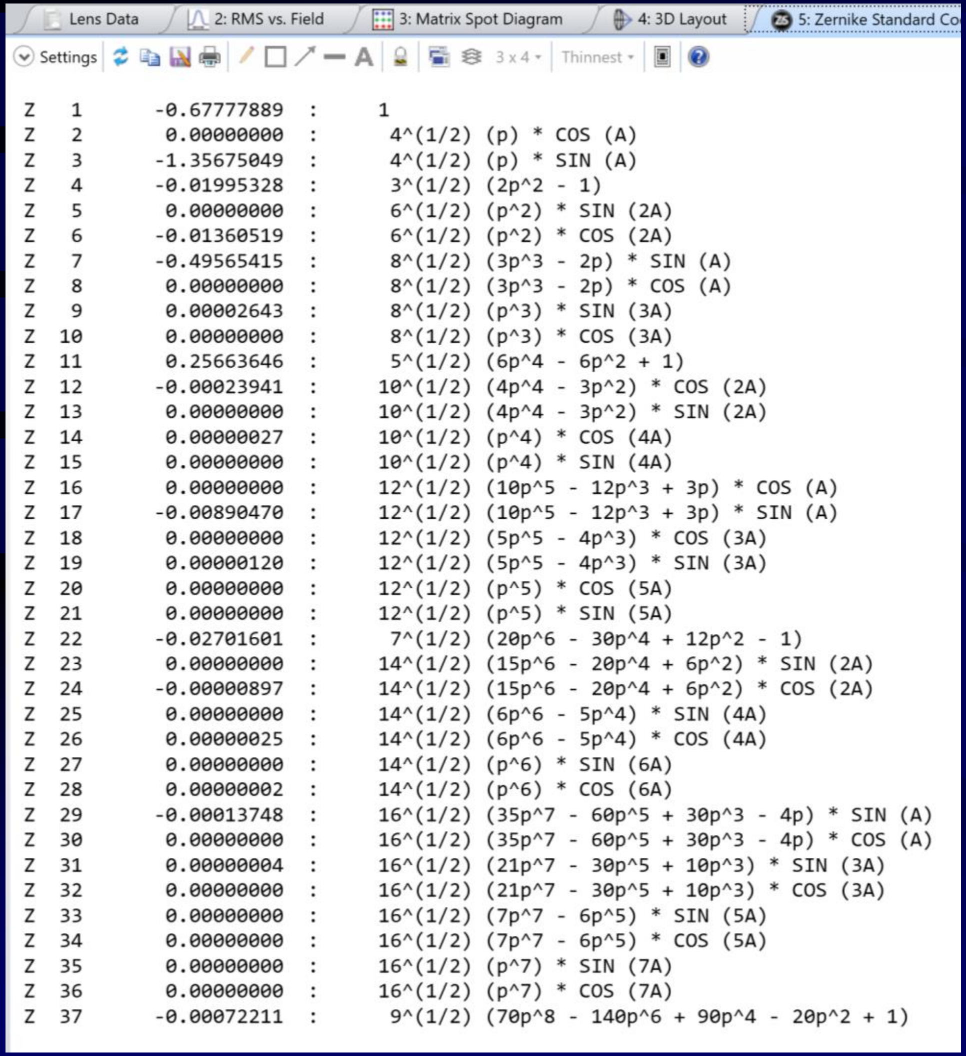
Aligned



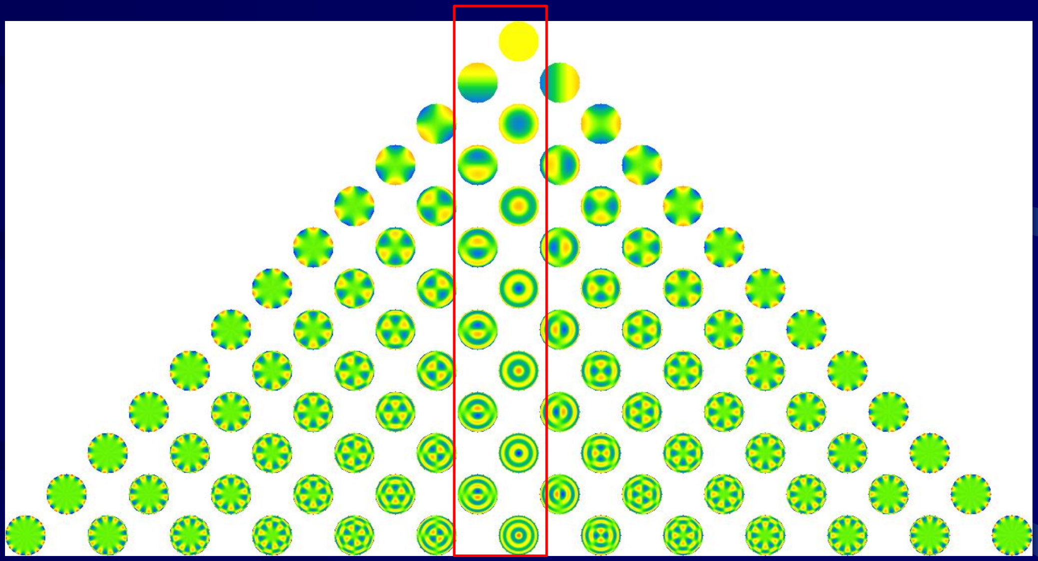
Z	1	-0.70742927	:	1
Z	2	0.00000000	:	4^(1/2) (p) * COS (A)
Z	3	0.00000000	:	4^(1/2) (p) * SIN (A)
Z	4	-0.03731982	:	3^(1/2) (2p^2 - 1)
Z	5	0.00000000	:	6^(1/2) (p^2) * SIN (2A)
Z	6	0.00000000	:	6^(1/2) (p^2) * COS (2A)
Z	7	0.00000000	:	8^(1/2) (3p^3 - 2p) * SIN (A)
Z	8	0.00000000	:	8^(1/2) (3p^3 - 2p) * COS (A)
Z	9	0.00000000	:	8^(1/2) (p^3) * SIN (3A)
Z	10	0.00000000	:	8^(1/2) (p^3) * COS (3A)
Z	11	0.25643622	:	5^(1/2) (6p^4 - 6p^2 + 1)
Z	12	0.00000000	:	10^(1/2) (4p^4 - 3p^2) * COS (2A)
Z	13	0.00000000	:	10^(1/2) (4p^4 - 3p^2) * SIN (2A)
Z	14	0.00000025	:	10^(1/2) (p^4) * COS (4A)
Z	15	0.00000000	:	10^(1/2) (p^4) * SIN (4A)
Z	16	0.00000000	:	12^(1/2) (10p^5 - 12p^3 + 3p) * COS (A)
Z	17	0.00000000	:	12^(1/2) (10p^5 - 12p^3 + 3p) * SIN (A)
Z	18	0.00000000	:	12^(1/2) (5p^5 - 4p^3) * COS (3A)
Z	19	0.00000000	:	12^(1/2) (5p^5 - 4p^3) * SIN (3A)
Z	20	0.00000000	:	12^(1/2) (p^5) * COS (5A)
Z	21	0.00000000	:	12^(1/2) (p^5) * SIN (5A)
Z	22	-0.02702340	:	7^(1/2) (20p^6 - 30p^4 + 12p^2 - 1)
Z	23	0.00000000	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * SIN (2A)
Z	24	0.00000000	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * COS (2A)
Z	25	0.00000000	:	14^(1/2) (6p^6 - 5p^4) * SIN (4A)
Z	26	0.00000025	:	14^(1/2) (6p^6 - 5p^4) * COS (4A)
Z	27	0.00000000	:	14^(1/2) (p^6) * SIN (6A)
Z	28	0.00000000	:	14^(1/2) (p^6) * COS (6A)
Z	29	0.00000000	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * SIN (A)
Z	30	0.00000000	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * COS (A)
Z	31	0.00000000	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * SIN (3A)
Z	32	0.00000000	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * COS (3A)
Z	33	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * SIN (5A)
Z	34	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * COS (5A)
Z	35	0.00000000	:	16^(1/2) (p^7) * SIN (7A)
Z	36	0.00000000	:	16^(1/2) (p^7) * COS (7A)
Z	37	-0.00072251	:	9^(1/2) (70p^8 - 140p^6 + 90p^4 - 20p^2 + 1)



Y-Decenter



Z	1	-0.67777889	:	1
Z	2	0.00000000	:	4^(1/2) (p) * COS (A)
Z	3	-1.35675049	:	4^(1/2) (p) * SIN (A)
Z	4	-0.01995328	:	3^(1/2) (2p^2 - 1)
Z	5	0.00000000	:	6^(1/2) (p^2) * SIN (2A)
Z	6	-0.01360519	:	6^(1/2) (p^2) * COS (2A)
Z	7	-0.49565415	:	8^(1/2) (3p^3 - 2p) * SIN (A)
Z	8	0.00000000	:	8^(1/2) (3p^3 - 2p) * COS (A)
Z	9	0.00002643	:	8^(1/2) (p^3) * SIN (3A)
Z	10	0.00000000	:	8^(1/2) (p^3) * COS (3A)
Z	11	0.25663646	:	5^(1/2) (6p^4 - 6p^2 + 1)
Z	12	-0.00023941	:	10^(1/2) (4p^4 - 3p^2) * COS (2A)
Z	13	0.00000000	:	10^(1/2) (4p^4 - 3p^2) * SIN (2A)
Z	14	0.00000027	:	10^(1/2) (p^4) * COS (4A)
Z	15	0.00000000	:	10^(1/2) (p^4) * SIN (4A)
Z	16	0.00000000	:	12^(1/2) (10p^5 - 12p^3 + 3p) * COS (A)
Z	17	-0.00890470	:	12^(1/2) (10p^5 - 12p^3 + 3p) * SIN (A)
Z	18	0.00000000	:	12^(1/2) (5p^5 - 4p^3) * COS (3A)
Z	19	0.00000120	:	12^(1/2) (5p^5 - 4p^3) * SIN (3A)
Z	20	0.00000000	:	12^(1/2) (p^5) * COS (5A)
Z	21	0.00000000	:	12^(1/2) (p^5) * SIN (5A)
Z	22	-0.02701601	:	7^(1/2) (20p^6 - 30p^4 + 12p^2 - 1)
Z	23	0.00000000	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * SIN (2A)
Z	24	-0.00000897	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * COS (2A)
Z	25	0.00000000	:	14^(1/2) (6p^6 - 5p^4) * SIN (4A)
Z	26	0.00000025	:	14^(1/2) (6p^6 - 5p^4) * COS (4A)
Z	27	0.00000000	:	14^(1/2) (p^6) * SIN (6A)
Z	28	0.00000002	:	14^(1/2) (p^6) * COS (6A)
Z	29	-0.00013748	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * SIN (A)
Z	30	0.00000000	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * COS (A)
Z	31	0.00000004	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * SIN (3A)
Z	32	0.00000000	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * COS (3A)
Z	33	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * SIN (5A)
Z	34	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * COS (5A)
Z	35	0.00000000	:	16^(1/2) (p^7) * SIN (7A)
Z	36	0.00000000	:	16^(1/2) (p^7) * COS (7A)
Z	37	-0.00072211	:	9^(1/2) (70p^8 - 140p^6 + 90p^4 - 20p^2 + 1)

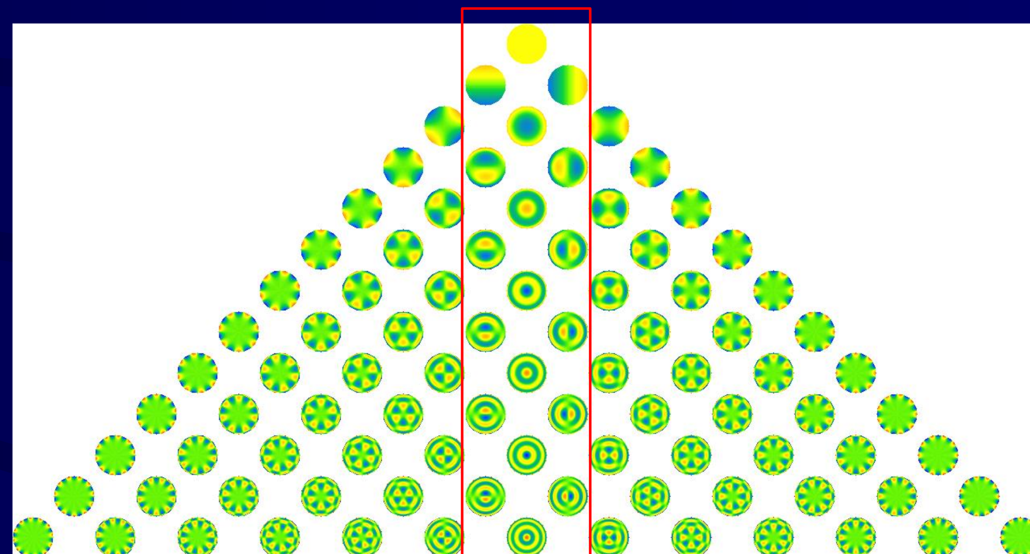


General Decenter

Lens Data | 2: RMS vs. Field | 3: Matrix Spot Diagram | 4: 3D Layout | 5: Zernike Standard Coefficients

Settings 3 x 4 Thinnest

Z	1	-0.69201164	:	1
Z	2	-0.54265947	:	4^(1/2) (p) * COS (A)
Z	3	-0.81398921	:	4^(1/2) (p) * SIN (A)
Z	4	-0.02828957	:	3^(1/2) (2p^2 - 1)
Z	5	0.00653016	:	6^(1/2) (p^2) * SIN (2A)
Z	6	-0.00272090	:	6^(1/2) (p^2) * COS (2A)
Z	7	-0.29736911	:	8^(1/2) (3p^3 - 2p) * SIN (A)
Z	8	-0.19824608	:	8^(1/2) (3p^3 - 2p) * COS (A)
Z	9	-0.00000190	:	8^(1/2) (p^3) * SIN (3A)
Z	10	0.00000972	:	8^(1/2) (p^3) * COS (3A)
Z	11	0.25654033	:	5^(1/2) (6p^4 - 6p^2 + 1)
Z	12	-0.00004788	:	10^(1/2) (4p^4 - 3p^2) * COS (2A)
Z	13	0.00011491	:	10^(1/2) (4p^4 - 3p^2) * SIN (2A)
Z	14	0.00000025	:	10^(1/2) (p^4) * COS (4A)
Z	15	0.00000000	:	10^(1/2) (p^4) * SIN (4A)
Z	16	-0.00356120	:	12^(1/2) (10p^5 - 12p^3 + 3p) * COS (A)
Z	17	-0.00534180	:	12^(1/2) (10p^5 - 12p^3 + 3p) * SIN (A)
Z	18	0.00000044	:	12^(1/2) (5p^5 - 4p^3) * COS (3A)
Z	19	-0.00000008	:	12^(1/2) (5p^5 - 4p^3) * SIN (3A)
Z	20	0.00000000	:	12^(1/2) (p^5) * COS (5A)
Z	21	0.00000000	:	12^(1/2) (p^5) * SIN (5A)
Z	22	-0.02701956	:	7^(1/2) (20p^6 - 30p^4 + 12p^2 - 1)
Z	23	0.00000431	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * SIN (2A)
Z	24	-0.00000179	:	14^(1/2) (15p^6 - 20p^4 + 6p^2) * COS (2A)
Z	25	0.00000000	:	14^(1/2) (6p^6 - 5p^4) * SIN (4A)
Z	26	0.00000025	:	14^(1/2) (6p^6 - 5p^4) * COS (4A)
Z	27	0.00000000	:	14^(1/2) (p^6) * SIN (6A)
Z	28	0.00000000	:	14^(1/2) (p^6) * COS (6A)
Z	29	-0.00008246	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * SIN (A)
Z	30	-0.00005497	:	16^(1/2) (35p^7 - 60p^5 + 30p^3 - 4p) * COS (A)
Z	31	0.00000000	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * SIN (3A)
Z	32	0.00000001	:	16^(1/2) (21p^7 - 30p^5 + 10p^3) * COS (3A)
Z	33	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * SIN (5A)
Z	34	0.00000000	:	16^(1/2) (7p^7 - 6p^5) * COS (5A)
Z	35	0.00000000	:	16^(1/2) (p^7) * SIN (7A)
Z	36	0.00000000	:	16^(1/2) (p^7) * COS (7A)
Z	37	-0.00072230	:	9^(1/2) (70p^8 - 140p^6 + 90p^4 - 20p^2 + 1)



Direction of decentration

$$tan^{-1}\left[\frac{a_{1,-1}}{a_{1,1}}\right]$$