Chapter 1

Aberrations and Ray Tracing

The theory of optical aberrations presented earlier was derived as an approximate correction to the results of paraxial optics. This section treats optical aberrations as a way of characterizing the behavior of rays using information extracted from exact ray tracing. There will be a progression of raytracing methods presented, with increasing complexity. After each method is presented, a spreadsheet or computer program will be given for the method, with example results. Ray *fans* obtained using these methods are used to display the transverse ray aberrations of two example lenses. The method of least-squares will be introduced as a way of obtaining the optical aberration coefficients. This, in turn, leads into an introduction to the use of orthogonal polynomials in studying optical aberrations. There will be frequent numerical comparisons of the aberration coefficients obtained from ray fans to those from the optical aberration formulas presented earlier. Although the methods of this section lead naturally to *ray* aberrations, our analysis and discussion will be centered on the corresponding *wavefront* aberrations.

The general progression of raytracing methods is as follows. The first method is the trigonometric raytrace. It is an exact method of tracing meridional rays, using formulas appropriate for a hand calculator or a spreadsheet. It can be used to evaluate all orders of spherical aberration on axis and provides some information about the off-axis aberrations such as coma. The next method is the Coddington (or Young) raytrace, which uses a pair of rays very close to a single meridional ray traced from an off-axis position to evaluate astigmatism and field curvature. The third method is the general raytrace, which avoids trigonometric calculations, but does use two square-root evaluations per spherical surface. This is the preferred method for use in computer programs. The last method is the general differential raytrace, from which the paraxial raytrace and Coddington equations are special cases.

The methods presented here are also described by many other authors, although there is considerable variations in notation and sign conventions. I found the discussion by Kingslake in *Lens Design Fundamentals* on the methods of trigonometric raytracing to be very helpful. I have tried to emulate his use of numerical examples.

1.1 Trigonometric or (Q, U) raytrace

In this method of raytracing, a ray is characterized by its angle U to the optic axis (z-axis) and the distance of closest approach Q to the local origin which is always the vertex of an optical surface. The ray tracing process involves finding the angle U' and distance Q' of the refracted ray. Fig. 1.1 shows an incident ray intersecting the surface at the point P. The distance Q is found by constructing a line perpendicular to the ray and passing through the vertex V. A similar construction is made for the refracted ray which must also intersect the surface at the point P.

If we start the ray trace from a point on the optical axis, as shown in Fig. 1.2, we can find $Q = L\sin(U)$ where L is the distance from the starting point to the vertex of the first optical surface.

The angle of incidence I is needed to apply Snell's Law of Refraction. The relationship between Q and U and the angle of incidence I is shown in Fig. 1.3. The diagram shows a spherical surface centered at the point C and intersecting the optic axis at the vertex point V. The incident ray intersects the surface at the point P. The distance $\overline{\mathrm{VC}}$ is the radius of curvature R. The diagram shows the line AC constructed perpendicular to the ray and passing through the center of curvature. A line parallel to the ray and passing though the vertex V is also constructed. This line intersects the line AC at the point B. The distance $\overline{\mathrm{AB}}$ is the distance of closest approach Q. The line drawn through PC is the normal to the point of intersection P, and the angle $\angle$CPA is the angle of incidence. The triangle PCA is used to obtain the distance $\overline{\mathrm{AC}} = R\sin(I)$, and the triangle VBC is used to obtain the distance $\overline{\mathrm{BC}} = R\sin(U)$. Then the distance of closest approach Q is given by $\overline{\mathrm{AB}} = R\sin(I) - R\sin(U)$. The angle of incidence I is then given by

$$\sin(I) = \sin(U) + Qc \tag{1.1}$$

where $c = 1/R$. Eq. 1.1 corresponds to the paraxial equation $i = u + cy$. If $Q > R$ then the ray will certainly miss the spherical surface. Even if $Q \leq R$ the ray will still miss the surface if the ray angle is large enough. The general case of the ray missing the sphere is having $|\sin(I)| > 1$ in Eq. 1.1.

The angle of incidence and angle of refraction are related by

$$n'\sin(I') = n\sin(I). \tag{1.2}$$

If $\sin(I')$ is greater than one, we have total internal reflection. The ray angles are related by $I = U + \beta$ and $I' = U' + \beta$, from which we obtain

$$U' = U + I' - I \tag{1.3}$$

Eq. 1.1 could be used to find Q', but this equation is numerically unstable for small c and indeterminate in the limit $c \to 0$. Instead, we adopt the method shown in Fig. 1.4.

Given a ray intersecting a spherical surface at the point P, draw a line through P and the center of curvature C. This is the surface normal at P. Then construct two perpendicular

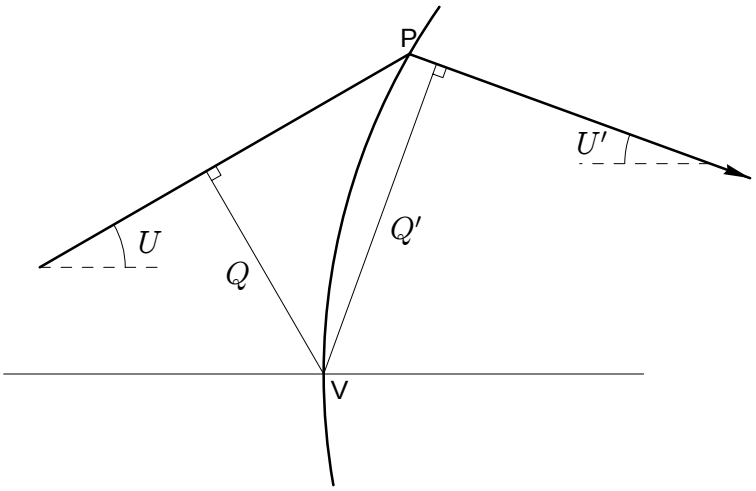


Figure 1.1: Ray refraction at spherical surface using distance of closest approach

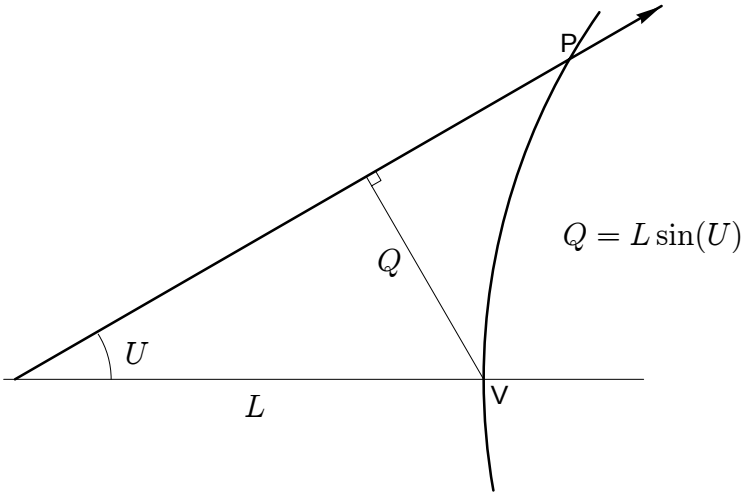
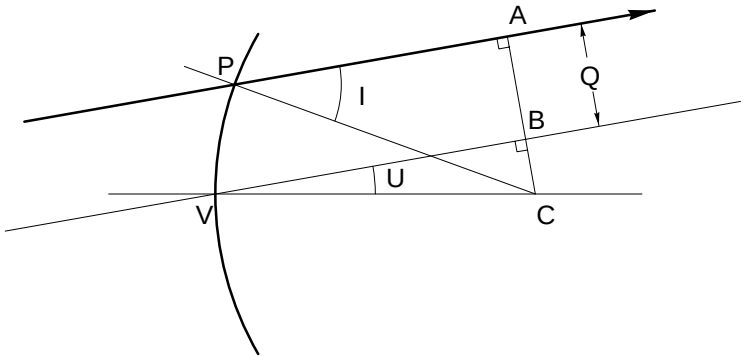


Figure 1.2: Starting the ray trace with an on-axis ray.



$$\overline{\mathrm{PC}} = \overline{\mathrm{VC}} = R$$
$$\overline{\mathrm{AC}} = R\sin(I)$$
$$\overline{\mathrm{BC}} = R\sin(U)$$
$$Q = \overline{\mathrm{AB}}$$
$$Q = R\sin(I) - R\sin(U)$$

Figure 1.3: Diagram showing the relationship between the distance of closest approach Q, the ray angle U, and the angle of incidence I for a spherical surface with radius of curvature R.

lines, one from the vertex V to the line PC and the other from P to the line VC. Let the point O be the intersection of these two perpendicular lines. Then the distance $\overline{\mathrm{VO}}$ will be equal to the distance $\overline{\mathrm{PO}}$, and we name this distance as G. Note that G depends only on the points P, V, and C and is therefore the same before and after refraction.

Now construct a line parallel to the ray and passing through the vertex V. Draw a line through O perpendicular to the ray (intersecting the ray at A, and the line parallel to the ray at B). Observe that $\angle$AOP is equal to U, and that $\angle$VOB is equal to I. The right triangle APO may be used to find $\overline{\mathrm{AO}} = G\cos(U)$, and the right triangle VBO may be used to find $\overline{\mathrm{OB}} = G\cos(I)$. Finally $Q = \overline{\mathrm{AB}}$. The final result is the pair of equations

$$G = \frac{Q}{\cos(U) + \cos(I)} \tag{1.4}$$

and

$$Q' = G(\cos(U') + \cos(I')) \tag{1.5}$$

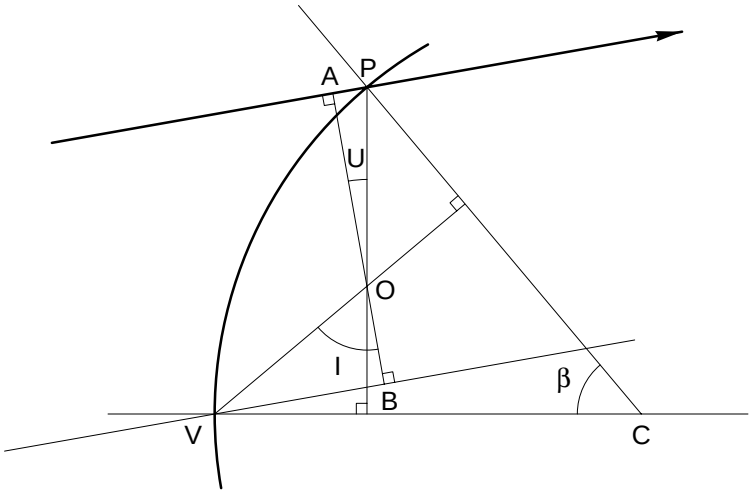
The procedure for ray transfer between spherical surfaces is shown in Fig. 1.5. Two surfaces are shown, with Q_1' for the refracted ray at surface 1 and Q_2 for the incident ray at surface 2. The construction in Fig. 1.5 shows that

$$Q_2 = Q_1' + d\sin(U) \tag{1.6}$$

where d is the axial distance between the two surface vertices.

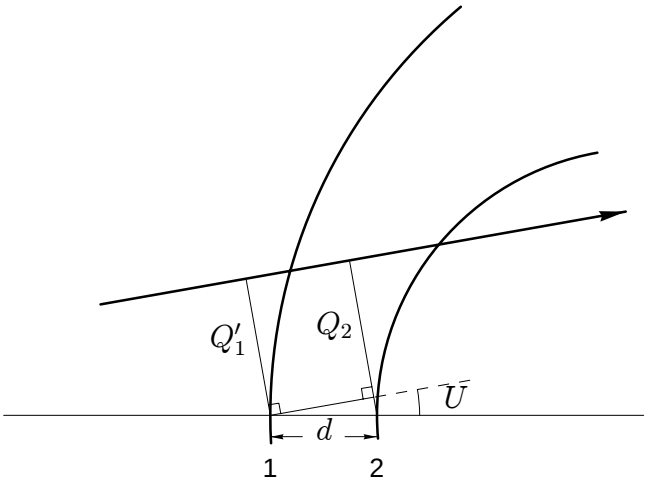
The method of trigonometric raytracing does not use the intersection point P directly. If the coordinates of P are desired, they may be obtained by inspection of Fig 1.4 as

$$\begin{aligned} x_p &= 0 \\ y_p &= G(1 + \cos(\beta)) \qquad (1.7) \\ z_p &= G\sin(\beta) \qquad (1.8) \end{aligned}$$



$$I = U + \beta$$
$$G = \overline{\mathrm{VO}} = \overline{\mathrm{PO}}$$
$$\overline{\mathrm{AO}} = G\cos(U)$$
$$\overline{\mathrm{OB}} = G\cos(I)$$
$$Q = \overline{\mathrm{AB}}$$
$$Q = G\cos(U) + G\cos(I)$$

Figure 1.4: Numerically stable method for trigonometric ray trace calculation.



$$Q_2 = Q_1' + d\sin(U)$$

Figure 1.5: Ray transfer between spherical surfaces.

Table 1.1: Spreadsheet layout for trigonometric raytrace through singlet lens.

	A	B	C	D	E	F	G
1	#	rd	cv	th	n	y_a	u_a
2	0			1.00E+20	1	0	0
3	1		0	0	1	6.25	0
4	2	29.96	0.033378	2	1.5168	6.25	-0.07108
5	3	-183.63	-0.00545	49.86274	1	6.07845	-0.125
6	4			1	0	-0.125	

	H	I	J	K	L	M	N	O
1	Y	U	Z	I	I'	Q	Q'	G
2	0	0						
3	6.25	0	0	0	0	6.25	6.25	3.125
4	6.25	-0.07218	0.65916	0.210155	0.13797	6.25	6.281264	3.15976
5	6.160519	-0.12722	-0.10337	-0.10574	-0.16078	6.137022	6.097616	3.081127
6	-0.1029	-0.12722	0	-0.12722	-0.12722	-0.10207	-0.10207	-0.05145

wavelength	0.00058756	f/no	4.00	half-field 4°
epsilon	0.0047	efl	50.00	

W_{040}	W_{131}	W_{222}	W_{220}	W_{311}
5.2547	0.8087	3.15133	2.64721	-0.047

A spreadsheet may be used to calculate a trigonometric raytrace. The singlet lens shown in Table 1.1 and the doublet lens in Table 1.2 are illustrations of the trigonometric raytrace method. Columns A through E are the specifications of the lens. Columns F and G are a paraxial raytrace for the axial ray. In these examples, the paraxial data are used to solve for the back focal distance in cell D5. Table 1.3 lists the formulas actually used in the trigonometric raytrace (Columns H through O). Cell M3 provides the starting value of Q and cell I2 the starting value for U. Note the increase in complexity required to implement a real raytrace. The paraxial raytrace uses two columns of formulas. The trigonometric raytrace uses eight columns of formulas.

The fourth-order aberrations listed for the lenses are based on a half-field angle of 4° for an infinite object plane. The doublet lens is a commercially available lens (Edmund Scientific Catalog 1993, stock no. N32,322). The trigonometric raytrace generates an exact meridional ray through the upper edge of the pupil. The Y intercept in cell H6 for both lenses gives the transverse spherical aberration of the axial ray.

Table 1.2: Spreadsheet layout for trigonometric raytrace through doublet lens.

	A	B	C	D	E	F
1	#	cv	th	n	y_a	u_a
2	0		1.00E+20	1	0	1.25E-19
3	1	0.0289603	9	1.67	12.5	-0.14524
4	2	-0.0454959	2.5	1.728	11.19288	-0.12327
5	3	-0.0046592	43.55158	1	10.88471	-0.24993
6	4			1	0	-0.24993

	G	H	I	J	K	L	M	N
1	Y	U	Z	I	I'	Q	Q'	G
2	0	0						
3	12.5	-0.15191	2.34194	0.37042	0.21850	12.5	12.71045	6.46939
4	12.02938	-0.1222	-3.58395	-0.73103	-0.70132	11.34848	11.50278	6.54858
5	11.31885	-0.25277	-0.29867	-0.17497	-0.30553	11.19803	10.88448	5.66336
6	-0.00746	-0.25277	0	-0.25277	-0.25277	-0.00722	-0.00722	-0.00373

wavelength	0.00055	f/no	2.00	half-field 4°
epsilon	0.0022	efl	50.01	

W_{040}	W_{131}	W_{222}	W_{220}	W_{331}
24.8090	-3.4336	12.5085	10.6534	0.4497

Table 1.3: Spreadsheet formulas for trigonometric raytrace.

col.	name	equation	example formula
G	Y	1.7	G3 = N3*(1+COS(J3-H2))
H	U	1.3	H3 = H2+K3-J3
I	Z	1.8	I3 = N3*SIN(J3-H2)
J	I	1.1	J3 = ASIN(L3*B3+SIN(H2))
K	I'	1.2	K3 = ASIN(D2*SIN(J3)/D3)
L	Q	1.6	L4 = M3+C3*SIN(H3)
M	Q'	1.5	M3 = N3*(COS(H3)+COS(K3))
N	G	1.4	N3 = L3/(COS(H2)+COS(J3))

y	y_e
0.0	0.0
0.2	-0.00083
0.4	-0.00671
0.6	-0.02283
0.8	-0.05476
1	-0.10864

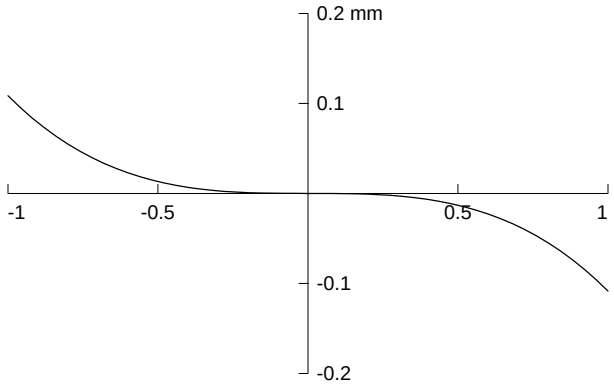


Figure 1.6: Meridional ray fan for singlet lens (on-axis, 0 field)

1.2 Spherical Aberration and the Meridional Ray Fan

Spherical aberration is the variation of focal length with aperture. In a symmetrical optical system, the spherical aberration can be evaluated using only meridional rays. By varying the height at which the starting ray intersects the pupil, we can use a trigonometric raytrace to completely characterize the spherical aberration of a lens. A meridional ray fan or *y-fan* is a plot of the transverse ray error y_e as a function of the relative height y at which the starting ray intersects the entrance pupil.

Fig. 1.6 shows a meridional y-fan from an on-axis object point for the singlet lens in Table 1.1. Only half of the ray fan really needs to be plotted because of the symmetry of the system. A full ray fan is shown here to emphasize the symmetry of the plot.

Fourth order spherical aberration is a wavefront of the form,

$$W(x,y) = a_4(x^2+y^2)^2 \tag{1.9}$$

The transverse ray errors in a meridional y-fan then have the form

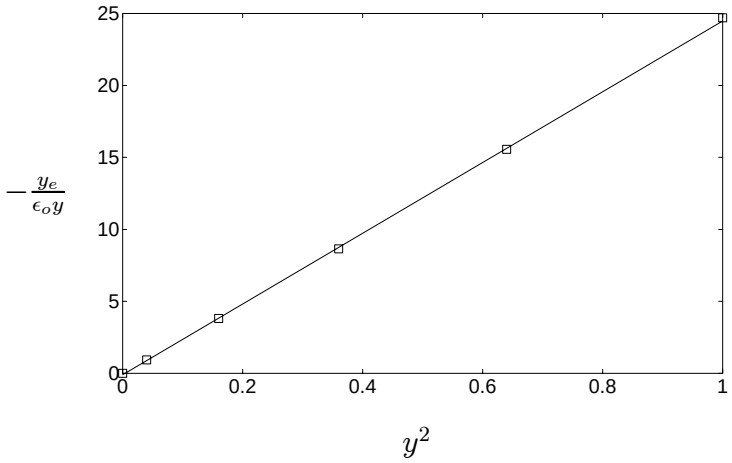
$$y_e = -4\epsilon_o a_4 y^3 \tag{1.10}$$

Fig. 1.6 certainly *looks* like a cubic curve. From Table 1.1 we find that $\epsilon_o = 0.0044$ and $a_4 = W_{040} = 5.9155$, so that the maximum tranverse error at the edge of the aperture should be $-4\epsilon_o a_4$ = -0.1041, which is in reasonably good agreement with the actual value of -0.1086.

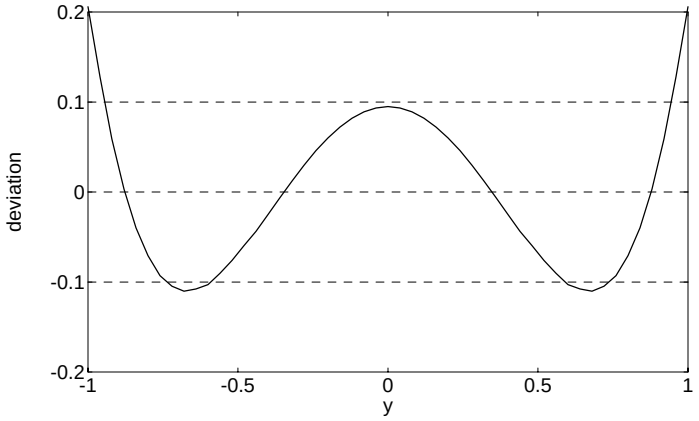
We can also linearize the ray aberrations by transforming them to the form

$$\left(-\frac{y_e}{\epsilon_o y}\right) = 4a_4\left(y^2\right) \tag{1.11}$$

and fitting a straight line to the resulting data. Fig. 1.7(a) shows the results of this transformation. The points marked are those tabulated in Fig. 1.6. Although a straight line seems



(a) transformed ray aberration data



(b) residuals from 4th order fit

Figure 1.7: Fitting the singlet ray fan data with a polynomial series.

to be a good fit to the data, if we use all of the data (51 points) to plot the deviation from the line, as shown in Fig. 1.7(b), we note that there are small systematic deviations between the actual ray data and the straight line. These deviations can be eliminated by adding another term to the polynomial series.

We can use the method of least-squares to fit a higher-order polynomial to the y-fan. If we model the wavefront as a Taylor series in r^2,

$$W(r) = a_2 r^2 + a_4 r^4 + a_6 r^6, \tag{1.12}$$

then terms through sixth-order spherical aberration will be included. There are $m = 3$ independent coefficients in this wavefront. The transverse ray errors then have the form

$$-y_e/(\epsilon_o y) = 2a_2 + 4a_4 y^2 + 6a_6 y^4 \tag{1.13}$$

Given n tabulated values (Fig. 1.6 contains 51 data values), we can construct a matrix $\mathbf{c}$ with n=51 rows, and m=3 columns such that each row is

$$c_i = (\ 2 \quad 4y_i^2 \quad 6y_i^4\ ). \tag{1.14}$$

We also construct a column vector $\mathbf{x}$ of length m such that

$$\mathbf{x} = \begin{pmatrix} a_2 \\ a_4 \\ a_6 \end{pmatrix} \tag{1.15}$$

and another column vector $\mathbf{d}$ of length n such that

$$d_i = -\frac{(y_e)_i}{\epsilon_o y_i}. \tag{1.16}$$

Then the linear equations to be satisfied are given by the single matrix equation

$$\mathbf{cx} = \mathbf{d} \tag{1.17}$$

with sizes

$$(51 \text{ x } 3)(3 \text{ x } 1) = (51 \text{ x } 1) \tag{1.18}$$

in which $\mathbf{x}$ is found by minimizing the mean squared deviation ϕ^2 defined by

$$\phi^2 = \frac{1}{n}|\mathbf{cx} - \mathbf{d}|^2 \tag{1.19}$$

with respect to $\mathbf{x}$, which has the nominal solution

$$\mathbf{x} = (\mathbf{c}^\dagger \mathbf{c})^{-1} \mathbf{c}^\dagger \mathbf{d} \tag{1.20}$$

with sizes

$$(3 \text{ x } 1) = ((3 \text{ x } 51)(51 \text{ x } 3))^{-1}(3 \text{ x } 51)(51 \text{ x } 1). \tag{1.21}$$

Table 1.4: Wavefront coefficients of Taylor series for singlet lens from on-axis infinite object.

n	rms	y^2	y^4	y^6
4	0.086	-0.0475	6.1427	
6	0.001	0.0002	5.9129	0.1721

y	y_e
0.0	0.0
0.1	-0.00022
0.2	-0.00171
0.3	-0.00564
0.4	-0.01285
0.5	-0.02364
0.6	-0.03735
0.7	-0.05161
0.8	-0.06091
0.9	-0.05392
1.0	-0.00746

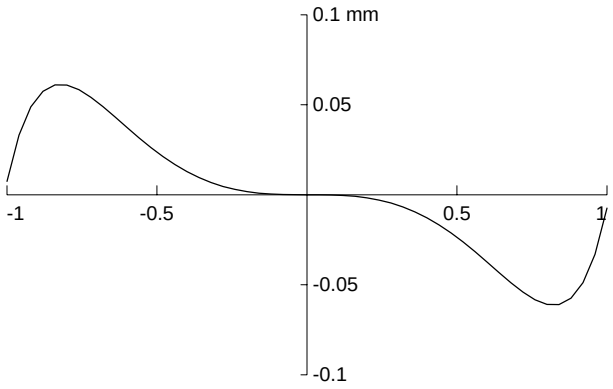


Figure 1.8: Meridional ray fan for doublet lens (on-axis, 0 field)

This can be evaluated directly with a software package such as MathCAD.

Table 1.4 shows the result of a least-squares fit to the data shown in Fig. 1.6. The first row shows the coefficients obtained using the first two terms in the Taylor series and the second row the coefficients obtained using three terms. Note that the fit is much better with the sixth order term than without, even though the value of the sixth-order coefficient is small. Furthermore, without the sixth-order term present, the values of the remaining terms are different than expected. At paraxial focus the second order term must be zero. Without the sixth-order term, the least-squares method changes the contributions of the lower order terms to partially compensate for the higher-order variations. Only after the sixth-order term is included do the lower-order coefficients take on their expected values.

A least-squares fit requires more data points than unknown coefficients. As a general rule there should be at least twice as many data points as coefficients to produce reliable results. The least-squares process should identify both the values of the coefficients and the goodness of fit.

The results for the singlet lens demonstrate that for this design, fourth-order spherical aberration is a adequate description of the wavefront.

In contrast to the singlet lens, the achromatic doublet lens can not be characterized by fourth-order aberration alone. Fig. 1.8 shows the on-axis y-fan for the achromatic doublet

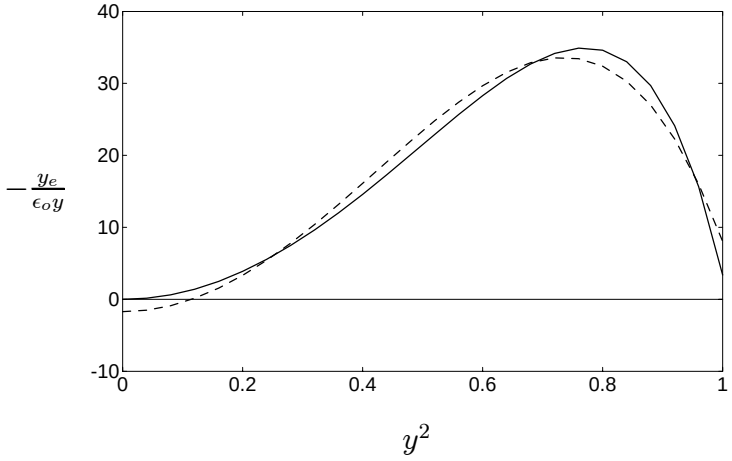


Figure 1.9: Fitting doublet ray fan data with polynomial. The solid line is the transformed ray data. The dashed line is the 6th order polynomial fit.

lens. The total aberration is almost half that of the singlet lens, even though the doublet lens is twice the diameter for the same focal length. The shape of the y-fan plot suggests that the fourth order spherical aberration is balanced against higher order spherical aberration.

Once again we try to fit the ray data by means of a Taylor expansion. A sixth-order fit produces the result shown in Fig. 1.9, which shows large deviations from the actual data. We must go to eighth-order to find an acceptable match to the ray fan data, and use several more terms to obtain a fit similar to that achieved for the singlet lens example. Table 1.5 shows the results of a series of fits of increasing order in wavefront. The final row contains coefficients to twelfth order for the following Taylor series

$$W(r) = a_2 r^2 + a_4 r^4 + a_6 r^6 + a_8 r^8 + a_{10} r^{10} + a_{12} r^{12} \tag{1.22}$$

Note that the coefficient $a_4 = 24.74$ in the bottom row of Table 1.5 compares to the value $W_{040} = 24.81$ in Table 1.2

There are several problems with the results in Table 1.5. First, they are sensitive to small changes in the input data, such as leaving out one of the data points, for example. The matrix inversion performed here is a classic example of ill-conditioning. Second, the values of the coefficients are unreliable. Although the values second and fourth order coefficients appear to be converging as the order of the fit increases, we have no indication that the values of the higher order terms are meaningful. Indeed, there is no reason to believe that we should stop at twelfth order even though the goodness of fit is reduced to near zero.

Both of the problems associated with using a Taylor series representation of the wavefront can be solved by using a different basis set of polynomials. In particular, the Zernike polynomials, which will be introduced shortly, are a suitable representation of wavefront aberrations for circular pupils.

Table 1.5: Coefficients for Taylor expansion of wavefront for doublet lens from on-axis infinite object.

n	rms	r^2	r^4	r^6	r^8	r^{10}	r^{12}
6	1.789	-0.85	32.67	-20.67			
8	0.258	0.12	22.81	-1.12	-10.10		
10	0.036	-0.02	25.25	-9.78	0.80	-4.52	
12	0.005	0.00	24.74	-6.89	-5.52	1.45	-2.05

Although the general principles of constructing a y-fan are fairly standard, there are differences in the details. For example, we elected to use the fractional height on the entrance pupil plane as the dependent variable. We could have equally as well have chosen the fractional height on the *exit* pupil plane. The results will be different if there is any distortion between the entrance and exit pupils. We elected to use paraxial values to determine pupil sizes. We could have used values obtained from an exact raytrace. The appearance of the y-fan certainly depends on how the pupil size is determined. Finally, we elected to use linear pupil coordinates (appropriate for a plane) for our ray fans. We could have used angular coordinates (direction cosines), which would be more appropriate for large numeric apertures (smaller f-numbers). Differences in implementation of ray fan calculations in different optical design codes means that different numerical results for aberrations may be obtained for the same lens specifications. Generally these differences are in degree rather than in kind. Spherical aberration does not disappear as a result of electing to use angular pupil coordinates rather than linear coordinates. However the magnitude of spherical aberration reported may differ between computer codes.

1.3 Spherical Aberration and Orthogonal Polynomials

Previously we expressed spherical aberration by means of a Taylor series of the form

$$W(r) = a_2 r^2 + a_4 r^4 + a_6 r^6 + a_8 r^8 + a_{10} r^{10} + a_{12} r^{12} \tag{1.23}$$

There are other set of polynomials which may represent a better choice of basis functions. We will present two such sets. The first is the set of Zernike polynomials, which are orthogonal over the unit circle. The second is the set of Chebyshev polynomials, which have uniform ripple properties.

The Zernike circle polynomials $R_n(r)$ are a set of even polynomials in r of order n. They are useful because they are orthogonal over the interior of the unit circle. In particular,

$$\int_0^1 R_n(r) R_m(r) r dr = \frac{1}{2[n+1]} \delta_{nm} \tag{1.24}$$

Table 1.6: Zernike Polynomials

$R_0(r)$	1
$R_2(r)$	$2r^2 - 1$
$R_4(r)$	$6r^4 - 6r^2 + 1$
$R_6(r)$	$20r^6 - 30r^4 + 12r^2 - 1$
$R_8(r)$	$70r^8 - 140r^6 + 90r^4 - 20r^2 + 1$
$R_{10}(r)$	$252r^{10} - 630r^8 + 560r^6 - 210r^4 + 30r^2 - 1$
$R_{12}(r)$	$924r^{12} - 2772r^{10} + 3150r^8 - 1680r^6 + 420r^4 - 42r^2 + 1$

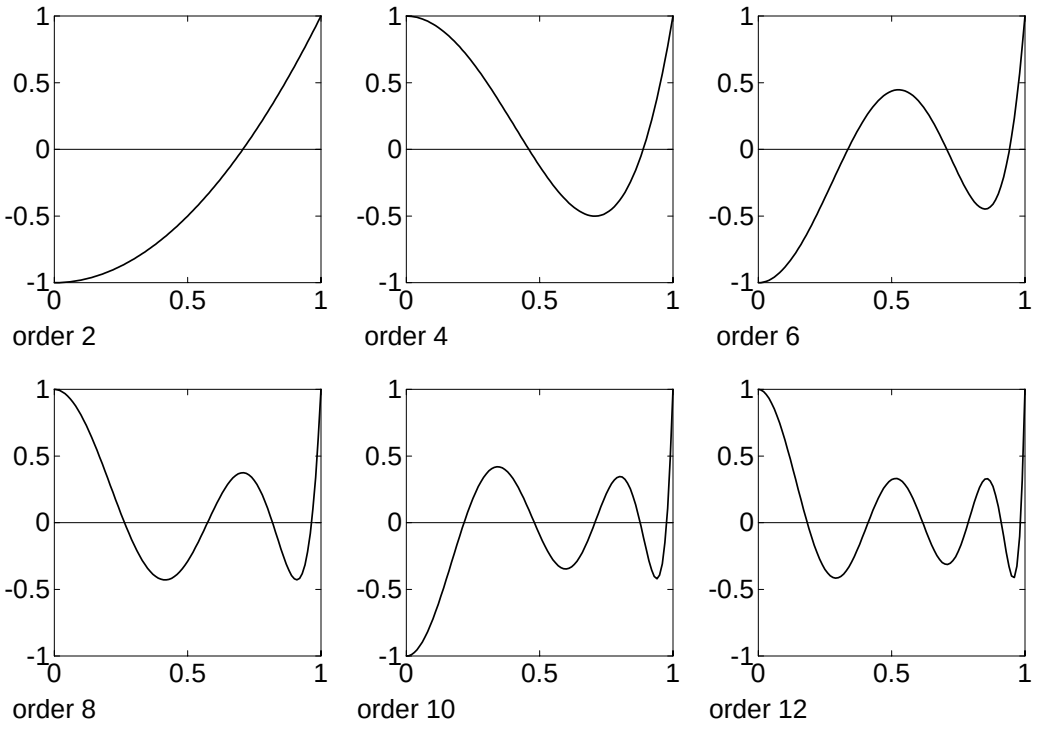


Figure 1.10: Zernike polynomials orders 2 though 12

Table 1.7: Coefficients for Zernike polynomial expansion of wavefront for doublet lens.

n	rms	$R_2(r)$	$R_4(r)$	$R_6(r)$	$R_8(r)$	$R_{10}(r)$	$R_{12}(r)$
6	1.789	6.841	0.406	-1.008			
8	0.258	6.922	0.635	-1.066	-0.144		
10	0.036	6.920	0.646	-1.037	-0.150	-0.018	
12	0.005	6.920	0.645	-1.036	-0.147	-0.019	-0.002

The first few Zernike circle polynomials are listed in Table 1.6 and plots of these polynomials are shown in Fig. 1.10.

Suppose we have a set of wavefront values, evenly distributed over the pupil. We wish to fit these values with Zernike polynomials using the method of least-squares. This requires that we minimize

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(W_i - \sum_{j=0}^{m} a_{2j}R_{2j}(r_i)\right)^2 \tag{1.25}$$

where n is the number of wavefront values and m is the number of terms in the polynomial series. If the pupil is indeed evenly sampled, then the summation over data points is almost equivalent to an integration over the area of the pupil. That is,

$$\langle(\cdots)\rangle = \frac{1}{n}\sum_{i=1}^{n}(\cdots) = 2\int_0^1(\cdots)rdr \tag{1.26}$$

This equivalence allows us to use the orthogonality of Zernike polynomials to simplify Eq. 1.25 to

$$\sigma^2 \approx \left\langle W_i^2\right\rangle - \sum_{j=0}^{m}\frac{a_{2j}^2}{2j+1} \tag{1.27}$$

where the coefficients a_{2j} are given by

$$a_{2j} = x\frac{\langle W_i R_{2j}(r_i)\rangle}{\langle(R_{2j}(r_i))^2\rangle} \tag{1.28}$$

This means that value of each coefficient is essentially independent of the other terms in the expansion. In particular, the value of the coefficient should not change as more terms are added.

To see how this works in practice, compare Table 1.7 to Table 1.5. Both tables are based on the same doublet lens data. The column values in Table 1.7 are much more stable as more terms are added. This very stability gives some reassurance that the pupil has been adequately sampled.

Furthermore the variations along the rows in Table 1.7 suggest that the terms of order r^{10} or higher could be ignored without significantly affecting any later analysis. We have no such confidence when we examine the coefficients in Table 1.5.

Table 1.8: Chebyshev Polynomials

$T_0(r)$	1
$T_2(r)$	$2r^2 - 1$
$T_4(r)$	$8r^4 - 8r^2 + 1$
$T_6(r)$	$32r^6 - 48r^4 + 18r^2 - 1$
$T_8(r)$	$128r^8 - 256r^6 + 160r^4 - 32r^2 + 1$
$T_{10}(r)$	$512r^{10} - 1280r^8 + 1120r^6 - 400r^4 + 50r^2 - 1$
$T_{12}(r)$	$2048r^{12} - 6144r^{10} + 6912r^8 - 3584r^6 + 840r^4 - 72r^2 + 1$

Finally compare the quality of the fits for the same number of terms, and observe that the rms deviation of the fit is independent of which polynomial set is used. The Zernike polynomial series is merely a rearrangement of the terms in the Taylor series and vice-versa. If a robust method of least-squares is used to obtain the coefficients, the quality of the fit will depend only on the number of terms used.

Zernike polynomials have the following optical significance. Suppose we have included enough terms in a Zernike series that we have fitted the sampled wavefront almost perfectly, meaning that σ^2 in Eq. 1.27 is very nearly zero. The first term in the Zernike series is the constant term, so that $a_0 = \langle W \rangle$. We can then rearrange Eq. 1.27 as

$$\left\langle (W - \overline{W})^2 \right\rangle = \sum_{j=1}^{m} \frac{a_{2j}^2}{2j+1}. \tag{1.29}$$

The Zernike coefficients are therefore related to the wavefront variance, which in turn is related to image quality via the Strehl ratio. Since the wavefront variance depends on the sum of the Zernike coefficients squared, each coefficient must be in dependently minimized to reduce the wavefront variance. The Zernike polynomials thus represent a set of orthogonal optical aberrations.

The Chebyshev polynomials $T_n(r)$ are another set of polynomials of order n. The first few Chebyshev polynomials are listed in Table 1.8 and plots of these polynomials are shown in Fig. 1.11. Successive terms in the sequence of polynomials can be found from

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x). \tag{1.30}$$

The following recurrence relation, less well known, is also useful

$$T_{2n}(x) = T_n(2x^2 - 1). \tag{1.31}$$

because it represents a more compact method of evaluating the even polynomials. Chebyshev polynomials are also terms in a Fourier series since

$$T_n(x) = \cos(n\theta) \tag{1.32}$$

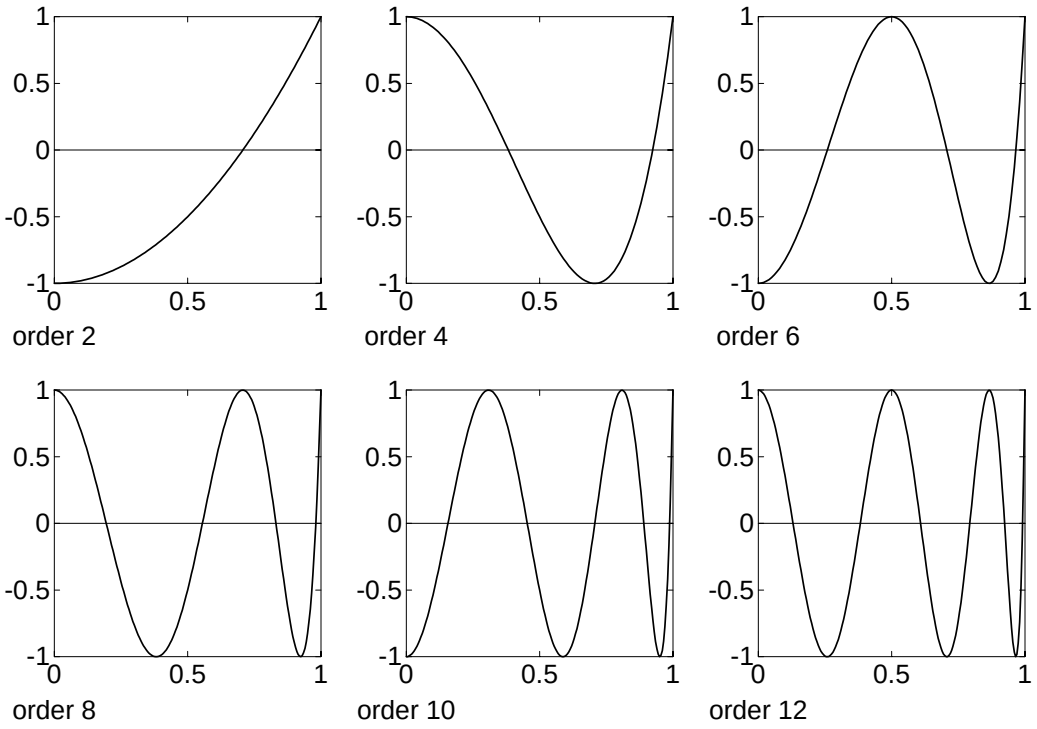


Figure 1.11: Chebyshev polynomials orders 2 though 12

Table 1.9: Coefficients for Chebyshev polynomial expansion of wavefront for doublet lens.

n	rms	$T_2(r)$	$T_4(r)$	$T_6(r)$	$T_8(r)$	$T_{10}(r)$	$T_{12}(r)$
6	1.789	6.463	0.305	-0.630			
8	0.258	6.522	0.431	-0.666	-0.079		
10	0.036	6.748	0.437	-0.653	-0.082	-0.009	
12	0.005	6.756	0.438	-0.653	-0.081	-0.009	-0.001

where $x = \cos(\theta)$. We may observe from the relation in Eq. 1.32 that the Chebyshev polynomials have a succession of maximums and minimums of alternating signs, each of magnitude one. This is the *uniform ripple* property of the Chebyshev polynomials. The Chebyshev polynomials do form an orthogonal set over the interval [-1, 1], but with a nonuniform weighting function, so that the orthogonality has no immediate relevance.

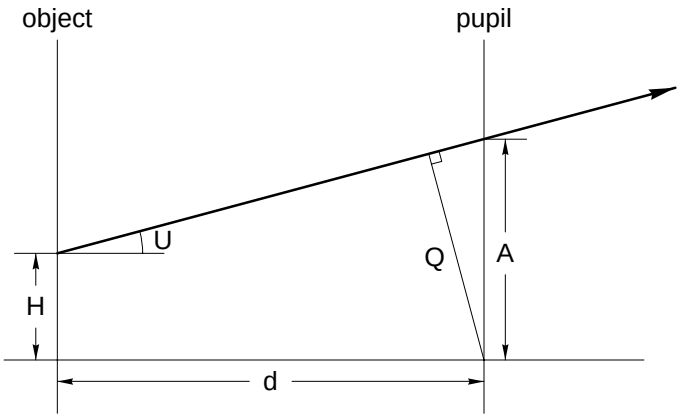
In comparing Fig. 1.10 to Fig. 1.11, one may observe that the two sets of polynomials are very similar in appearance. Table 1.9 shows the coefficients for the doublet lens. The table shows that the similarity between Zernike and Chebyshev polynomials applies to the values of the coefficients along the rows and to the stability of the coefficient values along the columns as the number of fitted terms increases. The stability of Zernike polynomial coefficients depends on how closely the summations using the sampled wavefront match the corresponding integrals over the circular pupil. Variations due to sampling are to be expected in the values of a coefficient as the number of fitted terms is increased. Zernike polynomials are not orthogonal over the sampled data points. In practice the variations in Zernike coefficients due to sampling may easily be as large as those encountered in the Chebyshev coefficients due to non-orthogonality over the unit circle. Certainly in the specific example presented here there is no reason to choose Zernike polynomials over Chebyshev polynomials merely because of the orthogonality property of the Zernike polynomials.

Similarly, although the Zernike polynomials do not have the uniform ripple property of the Chebyshev polynomials, there is sufficient uniformity to the Zernike polynomials to enable a decision on where to truncate the expansion. In practice there is no reason to choose Chebyshev polynomials merely because of the uniform ripple property of the Chebyshev polynomials.

The primary advantages of the Taylor polynomial expansion over either the Zernike or Chebyshev expansions are the increased precision near the center of the pupil and the ease which the Taylor coefficients may be scaled with changes in aperture size.

1.4 Off-axis Aberrations and the Meridional Ray Fan

A meridional ray for an off-axis object point can be specified as shown in Fig. 1.12 by the object height H and the height A at which the ray intersects the entrance pupil plane. Then



$$\tan(U) = (A - H)/d$$
$$Q = A\cos(U)$$

Figure 1.12: Diagram showing the relationships used to define an off-axis meridional ray.

the ray angle U is given by

$$\tan(U) = \frac{A - H}{d} \tag{1.33}$$

where d is the axial distance between the object and entrance pupil planes. The distance of closest approach Q to the origin of the pupil plane is given by

$$Q = A\cos(U) \tag{1.34}$$

For an infinite object distance, the angles of all entering rays are now the same and equal to the central angle hU_c where h is the fractional object height and U_c the chief ray angle. The distance of closest approach to the pupil is still given by Eq. 1.34.

To use the spreadsheet in Table 1.1 for an off-axis raytrace, make the following changes. Change cell H2 to the chief ray angle, using the formula =4*PI()/180 for a half-field angle of 4°. The calculated value of H2 should then be -0.06981. Change cell L3 to zero for a ray passing through the center of the entrance pupil. The Y value at the image plane (cell G5) should change to -3.49623. A paraxial raytrace of the chief ray, using $u_c = U_c$ = -0.06981, should give an image height of -3.49066. The height difference between the exact chief ray height and the paraxial chief ray height measures the distortion of the lens.

To use the same spreadsheet for a meriodinal fan, select an unused cell such as G7 as the fractional aperture height, and enter the value 1. Change cell L3 to the formula =G7*6.25*COS(H2) (See Eq. 1.34). The Y height at the image (G5) should change to -3.67064. The difference between this value and the value for the central ray (-3.49623) is the tranverse ray error (-0.1744). This difference can be calculated using the spreadsheet by setting cell H7 to the formula =G5+3.49623. Then cell H7 gives the transverse ray error for a specified fractional pupil coordinate in cell G7. These results can be tabulated and plotted as shown in Fig. 1.13.

Astigmatism and field curvature produce a quadratic variation in the wavefront, of the

y	y_e
-1.0	0.1581
-0.8	0.09494
-0.6	0.05357
-0.4	0.02769
-0.2	0.01162
0.0	0.0
0.2	-0.01229
0.4	-0.03039
0.6	-0.0596
0.8	-0.1055
1.0	-0.1744

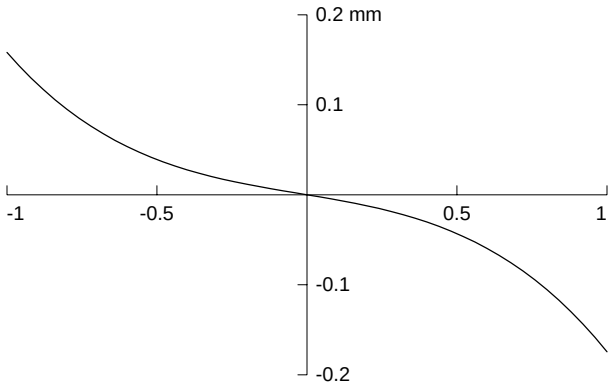


Figure 1.13: Meridional ray fan for singlet lens (off-axis, 4° half-field).

form

$$W(x,y) = W_{220}(x^2+y^2) + W_{222}y^2. \tag{1.35}$$

It is not possible to distinguish between these two aberrations using only a meridional ray fan ($x = 0$). A method for doing so will be given in the next section. However, we can characterize the *sum* of these aberrations with a meridional fan. To isolate the quadratic wavefront variation from higher order terms, we need to examine rays very close to the chief ray. The quadratic wavefront has the form

$$W(y) = a_2 y^2 \tag{1.36}$$

and the corresponding transverse ray error is

$$y_e = -2\epsilon_o a_2 y \tag{1.37}$$

Using the modified version of the spreadsheet from Table 1.1, with cell H7 set equal to the transverse ray error and G7 to the pupil height, we set an unused cell (A7) to $\epsilon_o = \lambda/u_a$ in image space and another unused cell (H8) to the formula =-H7/(2*G7*A7). As the value in G7 is set smaller, the value in H8 should approach 6.31. The accuracy of this process will depend on using a precise value for the chief ray height in the formula for cell H7. Note that the sum

$$W_{222} + W_{220} = 3.43 + 2.87 = 6.30, \tag{1.38}$$

using the results in Table 1.1.

Coma produces a wavefront variation of the form

$$W(x,y) = W_{131}y(x^2+y^2) \tag{1.39}$$

Table 1.10: Wavefront coefficients for Taylor series for singlet lens from off-axis infinite object (4° half-field angle).

rms	y^1	y^2	y^3	y^4	y^5	y^6
0.00034	0.0	6.310	0.644	6.021	-0.016	0.180

and spherical aberration has the form

$$W(x,y) = W_{040}(x^2+y^2)^2 \tag{1.40}$$

A meridional cross-section of a wavefront containing all the monochromatic fourth-order aberrations has the form

$$W(y) = a_2y^2 + a_3y^3 + a_4y^4 \tag{1.41}$$

where a_2 is $W_{220}+W_{222}$, $a_3 = W_{131}$ and $a_4 = W_{040}$. The corresponding transverse aberration is given by

$$y_e = -\epsilon_o(2a_2y + 3a_3y^2 + 4a_4y^3) \tag{1.42}$$

The contribution due to coma can be isolated because it is the only term in the above expression with even symmetry. Thus,

$$a_3 = \frac{y_e(+1) + y_e(-1)}{-6 * \epsilon_o} = \frac{-0.1744 + 0.1581}{-6 * 0.0044} = 0.617 \tag{1.43}$$

using the data from Fig. 1.13. This is close to the value $W_{131} = 0.615$ from Table 1.1.

Finally, we can evaluate the y-fan by using the method of least-squares to fit a Taylor series to the fan. The wavefront has the form

$$W(y) = a_1y + a_2y^2 + a_3y^3 + a_4y^4 + a_5y^5 + a_6y^6. \tag{1.44}$$

Table 1.10 shows the coefficients obtained using the data from Fig. 1.13. The value for coma is slightly different from that calculated earlier because of a small y^5 contribution. The values for a_4 and a_6 are consistent with those obtained earlier in Table 1.4, but we can not tell from a meridional y-fan whether the differences are due to oblique spherical aberration (spherical aberration that depends on field) or to higher-order astigmatism.

1.5 Astigmatism and the differential raytrace

Astigmatism results from a difference in wavefront curvature (quadratic variation) in the meridional and sagittal directions. The meridional variation was calculated in the previous section by examining the transverse ray error for a ray very close to the central ray through the entrance pupil. Such a ray is called a *differential* ray. Paraxial optics involved differential rays about the optic axis, because the optic axis is the path taken by the central ray from an

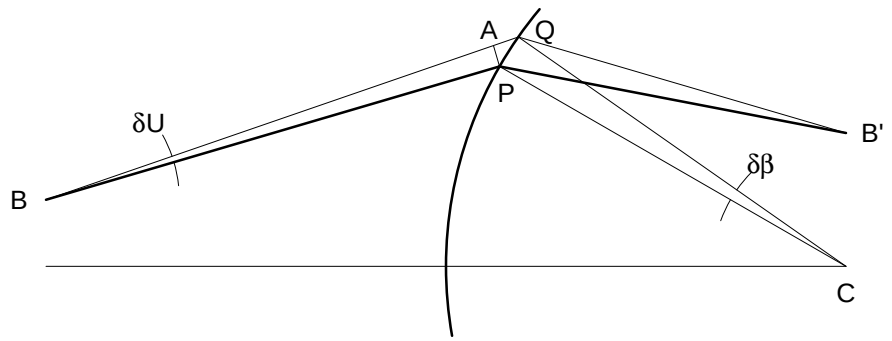


Figure 1.14: Differential ray through a spherical surface, taken with respect to an oblique central ray.

on-axis object. Astigmatism and field curvature can be calculated from a pair of differential rays, one in the meridional plane and the other in the sagittal plane.

Fig. 1.14 shows an oblique central ray from the point B, intersecting a spherical surface at the point P. The refracted ray then passes through the point B′. The center of the spherical surface is the point C. There is a differential ray intersecting the surface at the point Q, which in Fig. 1.14 is shown in the meridional plane. The short arc PQ is tangent to the sphere.

Starting with the differential of Snell's Law,

$$n' \cos(I')\delta I' = n \cos(I)\delta I \tag{1.45}$$

and the differential relationship $\delta I = \delta U + \delta\beta$, we can obtain

$$n \cos(I)\delta U - n' \cos(I')\delta U' = (n' \cos(I') - n \cos(I))\delta\beta \tag{1.46}$$

The differential change in the normal vector is $\delta\beta$, given by

$$\delta\beta = c\delta q \tag{1.47}$$

where δq is the length $\overline{\mathrm{PQ}}$. Then

$$n \cos(I)\delta U - n' \cos(I')\delta U' = \Phi\delta q \tag{1.48}$$

where

$$\Phi = (n' \cos(I') - n \cos(I))c. \tag{1.49}$$

The parameter Φ is called the oblique power of the surface and is a generalization of the paraxial power ϕ.

If the point Q lies in the meridional plane, then the differential vector δy from P to A, perpendicular to the central ray, is given by

$$\delta y = t\delta U \tag{1.50}$$

where t is the distance from B to P. Furthermore δy is related to δq by

$$\delta y = \cos(I)\delta q \tag{1.51}$$

from the right triangle PAQ in Fig. 1.14. Similarly for the refracted ray,

$$\delta y' = -t'\delta U' = \cos(I')\delta q \tag{1.52}$$

Finally Eq. 1.48 can be written as

$$\frac{n\cos^2(I)}{t} + \frac{n'\cos^2(I')}{t'} = \Phi \tag{1.53}$$

If Q lies in the sagittal plane, then $\delta y = \delta y' = \delta q$. The sagittal component of the refraction equation,

$$n'\mathbf{k}' = n\mathbf{k} + (n'\cos(I') - n\cos(I))\mathbf{N}, \tag{1.54}$$

can then be written as

$$\frac{n'}{s'} + \frac{n}{s} = \Phi, \tag{1.55}$$

using $k_x = (\delta q)/s$, $k'_x = -(\delta q)/s'$ and $N_x = -c(\delta q)$.

At the start of the raytrace, t and s are both equal to the oblique distance from the object to the first spherical surface. The oblique distance D_i from surface i to surface $i+1$ is obtained from

$$D_i = (d_i + Z_{i+1} - Z_i)/\cos(U_i) \tag{1.56}$$

The image distances are given by

$$s' = \frac{n'}{\Phi - (n/s)} \tag{1.57}$$

$$t' = \frac{n'\cos^2(I')}{\Phi - (n\cos^2(I)/t)} \tag{1.58}$$

The transfer equations from one surface to the next are

$$s_i = D_{i-1} - s'_{i-1} \tag{1.59}$$

$$t_i = D_{i-1} - t'_{i-1} \tag{1.60}$$

At the end of the raytrace, we generally want to know the axial distances of the tangential and sagittal focal lines from the paraxial image plane. These are given by

$$Z_s = s'\cos(U') + Z - \ell' \tag{1.61}$$

$$Z_t = t'\cos(U') + Z - \ell' \tag{1.62}$$

frc	fang	Y	sfs	tfs
0.00	0.00	0.00	-0.0000	-0.0000
0.10	0.40	0.35	-0.0019	-0.0041
0.20	0.80	0.70	-0.0075	-0.0163
0.30	1.20	1.05	-0.0168	-0.0366
0.40	1.60	1.40	-0.0299	-0.0650
0.50	2.00	1.75	-0.0467	-0.1015
0.60	2.40	2.10	-0.0673	-0.1462
0.70	2.80	2.45	-0.0916	-0.1988
0.80	3.20	2.80	-0.1196	-0.2595
0.90	3.60	3.15	-0.1513	-0.3282
1.00	4.00	3.50	-0.1867	-0.4049

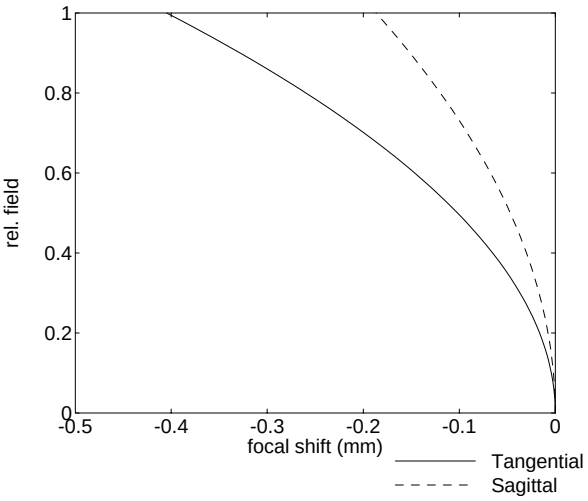


Figure 1.15: Field-curvature plot for doublet lens.

where Z is the sag of the final surface, ℓ' is the distance to the paraxial image and U' is the exit angle of the central ray.

Table 1.11 shows a spreadsheet layout that uses a differential raytrace to locate the tangential and sagittal astigmatic foci. The spreadsheet is modified from Table 1.2 for the achromatic doublet we are using as an example. The trigonometric raytrace is changed as described in the previous section to an off-axis ray at 4° half-field entering the lens at the center of the first surface. Columns O through V were added to generate the differential raytraces. Table 1.12 lists the formulas used in these columns. Columns O and P are the oblique power and thickness, respectively, for each surface. Columns Q and R handle the sagittal differential ray. Columns U and V do the tangential differential ray. Columns S and T are auxiliary variables used to simplify the formula in column V. The starting values for the differential raytrace were entered manually into cells Q3 and U3. The final results appear in cells U8 and U9.

A field curvature profile for this lens can be generated by varying the angle of the central ray and recording the corresponding field sags. The table in Fig. 1.15 contains three columns that relate to the field height, the fractional field height (frc), field angle (fang) in degrees, and image height (Y). The field sags are the tangential field sag (tfs) and sagittal field sag (sfs).

Since the shape of the field curvature profile in Fig. 1.15 looks basically quadratic in the field dependence, we may expect that the behavior is dominated by the fourth-order wavefront terms. At the edge of the field, we can calculate the wavefront coefficients as

$$W_{222} = -\frac{Z_T - Z_s}{\zeta} = \frac{0.21821}{0.0176} = 12.40 \tag{1.63}$$

$$W_{220} = -\frac{Z_s}{\zeta} = \frac{0.18669}{0.0176} = 10.61 \tag{1.64}$$

Table 1.11: Spreadsheet layout for differential raytrace through doublet lens.

	A	B	C	D	E	F
2	#	cv	th	n	y_a	u_a
3	0		1.00E+20	1	0	1.25E-19
4	1	0.0289603	9	1.67	12.5	-0.14524
5	2	-0.0454959	2.5	1.728	11.19288	-0.12327
6	3	-0.0046592	43.55158	1	10.88471	-0.24993
7	4			1	0	-0.24993

	G	H	I	J	K	L	M	N
2	Y	U	Z	I	I'	Q	Q'	G
3	0	-0.06981						
4	0	-0.04178	0	-0.06981	-0.04178	0	0	0
5	-0.37613	-0.04095	-0.00322	-0.02467	-0.02384	-0.37593	-0.37594	-0.18808
6	-0.47868	-0.06918	-0.00053	-0.03872	-0.06695	-0.4783	-0.47757	-0.23934
7	-3.49637	-0.06918	0	-0.06918	-0.06918	-3.48801	-3.48801	-1.74819

	O	P	Q	R	S	T	U	V
2	POWER	THK	S	S'	$n\cos^2(I)$	$n'\cos^2(I')$	T	T'
3								
4	0.019432	9.0046	1.0E+20	85.9419	0.995134	1.667086	1.0E+20	85.7919
5	-0.00264	2.5048	-76.9372	90.6303	1.668984	1.727018	-76.7873	90.4405
6	0.003396	43.6565	-88.1255	43.4694	1.72541	0.995525	-87.9358	43.2507
7								
8								
9						ZS	-0.18669	
10						ZT	-0.40490	

Table 1.12: Spreadsheet formulas for differential raytrace.

col.	name	equation	example formula
O	POWER	1.49	O4 =(D4*COS(K4)-D3*COS(J4))*B4
P	THK	1.56	P4 =(C4+I5-I4)/COS(H4)
Q	S		Q4 =C3
		1.59	Q5 =(P4-R4)
R	S'	1.57	R4 =D4/(O4-D3/Q4)
S	$n\cos^2(I)$		S4 =D3*COS(J4)^2
T	$n'\cos^2(I')$		T4 =D4*COS(K4)^2
U	T		U4 =C3
		1.60	U5 =P4-V4
V	T'	1.58	V4 = T4/(O4-S4/U4)
	ZS	1.61	U9 =R6*COS(H6)+I6-C6
	ZT	1.62	U10 =V6*COS(H6)+I6-C6

where $\zeta = 2\lambda/u_a^2$. The fourth-order coefficients based on paraxial optics, from Table 1.2, are $W_{222} = 12.51$ and $W_{220} = 10.65$.

1.6 General Raytrace

A general or *skew* ray is defined with a three-dimensional starting point in space and a three-dimensional direction of propagation. The meridional raytrace and differential raytrace are special cases of the general raytrace. The meridional raytrace is characterized by its use of trigonometric formulas, which is avoided in the general raytrace. The differential raytrace is an approximate raytrace, valid for rays close to a central ray.

Table 1.13 shows a spreadsheet layout for a general raytrace through a doublet lens. There are 18 columns used for the general raytrace compared to eight for the trigonometric raytrace and two for the paraxial raytrace. The general raytrace is sufficiently elaborate in both data storage and computation that we will discuss its implementation in a programming language as well as in a spreadsheet.

Consider a ray as a semi-infinite line consisting of a starting point in space $\mathbf{p} = (p_x, p_y, p_z)$ and a direction of propagation $\mathbf{k} = (k_x, k_y, k_z)$ through a medium of index of refraction n. A C-language data structure to describe a ray could be defined as:

```
typedef struct {
    double px,py,pz;  /* starting point */
    double kx,ky,kz;  /* direction of propagation */
    double n;         /* index of refraction */
```

```
}  RAY;
```

Because many of the computations in a general raytrace involve vector arithmetic, however, we prefer the alternative data structure below:

```
typedef struct {
    double p[3];  /* starting point */
    double k[3];  /* direction of propagation */
    double n;     /* index of refraction */
}  RAY;
```

where `p` and `k` are three-element arrays such that `p[0]` $= p_x$, `p[1]` $= p_y$, and `p[2]` $= p_z$.

We may declare in the C-language that `RAY *rp` defines `rp` as a *pointer* to a `RAY` data structure. Then the members of the structure are referenced using a notation such as `rp->n`, which refers to the value of the index of refraction for the ray `rp`.

The direction of propagation is expressed as optical cosines, such that

$$\mathbf{k} \cdot \mathbf{k} = k_x^2 + k_y^2 + k_z^2 = n^2 \tag{1.65}$$

The vector dot product is evaluated through the following C-language function:

```
double dot(double a[3], double b[3])
{
     return a[0]*b[0]+a[1]*b[1]+a[2]*b[2];
}
```

To obtain the value of the dot product between $\mathbf{p}$ and $\mathbf{k}$, for example, we would write a excerpt of code such as

```
    double p[3], k[3], result;
    double dot(double p[3], double k[3]);

    result = dot(p,k);
```

Ray translation The ray coordinates $\mathbf{p}'$ of a translated ray are given parametrically by

$$\begin{aligned} p'_x &= p_x + qk_x \\ p'_y &= p_y + qk_y \\ p'_z &= p_z + qk_z \end{aligned} \tag{1.66}$$

where $d = nq$ is the distance along the ray. The C-language code to implement ray translation is

```
    int i;
    double q;
    RAY *rp;

    for (i=0; i<3; i++) rp->p[i] += q * rp->k[i];
```

In this code, the translated coordinates *replace* the starting coordinates of the ray.

Surface normals A surface normal vector is located at the point of intersection of a ray with the surface specified by the equation:

$$F(x, y, z) = 0 \tag{1.67}$$

The surface normal $\mathbf{N}$ is given by

$$\mathbf{N} = \nabla F = \left\{ \begin{array}{ccc} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} & \frac{\partial F}{\partial z} \end{array} \right\} \tag{1.68}$$

Spherical Surfaces The equation for a sphere can be written as an implicit equation, using curvature instead of radius:

$$F(x, y, z) = c(x^2 + y^2 + z^2) - 2z = 0 \tag{1.69}$$

In this form, we obtain the equation of a plane, $z = 0$, as the curvature becomes zero.

The normal is given by

$$\mathbf{N} = \left\{ \begin{array}{ccc} -cx & -cy & 1 - cz \end{array} \right\} \tag{1.70}$$

where we have chosen the direction of the normal along the positive z-axis as $c \to 0$.

A C-language data structure for the parameters of an optical surface is

```
typedef struct {
    double cv;  /* surface curvature */
    double th;  /* distance to next surface */
    double n;   /* index of refraction following surface */
} SURF;
```

Surface intersection Given a ray starting at $\mathbf{p}$ and propagating in the direction $\mathbf{k}$, find the intersection of the ray $\mathbf{s}$ with a spherical optical surface of radius R with its vertex at the origin. The parametric equation of the ray is

$$\mathbf{s} = \mathbf{p} + q\mathbf{k} \tag{1.71}$$

We substitute the parametric equation for the ray into Eq. 1.69 and solve for the parameter q, which measures the oblique distance from the point $\mathbf{p}$ to the intersection point.

$$c(p_x + qk_x)^2 + c(p_y + qk_y)^2 + c(p_z + qk_z)^2 - 2(p_z + qk_z) = 0 \tag{1.72}$$

This then simplifies to the following quadratic equation

$$Aq^2 - 2qB + C = 0 \tag{1.73}$$

where

$$\begin{aligned} A &= (\mathbf{k}\cdot\mathbf{k})c = n^2 c && (1.74)\\ B &= k_z - (\mathbf{k}\cdot\mathbf{p})c && (1.75)\\ C &= (\mathbf{p}\cdot\mathbf{p})c - 2p_z && (1.76)\\ D &= B^2 - AC && (1.77) \end{aligned}$$

which has the solution

$$q = \frac{C}{B \pm \sqrt{D}} \tag{1.78}$$

There are two roots to the quadratic equation in q. The roots correspond to the two possible intersections of the ray with a sphere. Usually we want the smaller root, so we choose the sign of the discriminant to be the same as B. If $C = 0$, the point $\mathbf{p}$ is on the sphere, and $q = 0$. If the radical D is zero, the ray is tangent to the sphere, and if it is negative, the ray misses the sphere.

For the case of a plane $c = 0$ and the parameter q is given by

$$q = -p_z/k_z \tag{1.79}$$

Refraction at a spherical surface Given a ray incident on a spherical surface at the point $\mathbf{p}$ in the direction $\mathbf{k}$, find the direction of refraction $\mathbf{k}'$.

The general refraction equation is

$$\mathbf{k}' = \mathbf{k} + (\Gamma' - \Gamma)\mathbf{N} \tag{1.80}$$

where

$$\Gamma = n\cos I = \mathbf{k}\cdot\mathbf{N} \tag{1.81}$$

and I is the angle of incidence. From Snell's Law,

$$\Gamma' = \pm\sqrt{\Gamma^2 + n'^2 - n^2} \tag{1.82}$$

If the argument of the square root is negative, the ray is totally internally reflected. For refraction, the sign of Γ' must be the same as Γ.

Raytrace procedure Ray tracing is the procedure of calculating the ray intersections and refraction through a succession of optical surfaces. Start with the original ray description, $\mathbf{p}$ and $\mathbf{k}$.

The following steps trace the ray through a single optical surface.

1. Transfer to the local coordinates of an optical surface located a distance t along the z-axis (optical axis)
$$p'_z = p_z - t \tag{1.83}$$

2. Calculate the intersection with the spherical optical surface. This requires a square root calculation and yields $\mathbf{p}'$.

3. Find the surface normal vector using Eq. 1.70.

4. Calculate $\Gamma = \mathbf{k} \cdot \mathbf{N}$.

5. Calculate Γ'. This requires another square root calculation.

6. Calculate $\mathbf{k}' = \mathbf{k} + (\Gamma' - \Gamma)\mathbf{N}$

7. The final results are $\mathbf{p}'$ and $\mathbf{k}'$.

The final ray values are used as starting values for the next optical surface, and the steps listed above are repeated. The chain of calculations continues until the image surface is reached.

Table 1.14 shows how to construct the spreadsheet in Table 1.13. Fig 1.16 shows the output of a computer program used to trace the same ray shown in Table 1.13. A source listing of this program is given at the end of this chapter.

1.7 OPD Calculations

So far we have used raytracing to determine the transverse ray aberrations, and from those we have extracted the wavefront aberrations. Raytracing can also be used to determine the wavefront aberrations directly.

The actual distance traveled by the ray is given by the parameter q in Eq. 1.71, which is generated by Column M of the spreadsheet in Table 1.13. Since the magnitude of $\mathbf{k}$ is n, the parameter q must be a *reduced* path, D/n where D is the oblique physical path. A reference ray is needed for calculation of optical path differences. This is usually the central ray, but any ray could be chosen. Then OPD between two surfaces is calculated as

$$\text{OPD} = n^2(q_c - q) \tag{1.84}$$

where q_c is the reduced oblique distance for the reference ray and q is the reduced oblique distance for any other ray of interest. The total OPD is the sum of the OPD between surfaces. OPD is usually measured between a reference sphere centered on the object point and another reference sphere centered on the image point. The reference sphere reduces to a reference plane for an infinite object or image point. Table 1.13 could be used for OPD calculations by adding a dummy surface after the last real surface and setting the curvature equal to back focal distance. A dummy surface is one with the same index of refraction on

Table 1.13: Spreadsheet layout for general raytrace through doublet lens.

	A	B	C	D	E	F
1	#	cv	th	n	y_a	u_a
2	0		1.00E+20	1	0	1.25E-19
3	1	0.0000000	0	1	12.5	1.25E-19
4	2	0.0289603	9	1.67	12.5	-0.14524
5	3	-0.0454959	2.5	1.728	11.19288	-0.12327
6	4	-0.0046592	43.55158	1	10.88471	-0.24993
7	5	0.0000000	0	1	0	-0.24993

	G	H	I	J	K	L
1	X	Y	Z	KX	KY	KZ
2	0	0	0	0	1.25E-19	1
3	0	12.5	0	0	1.25E-19	1
4	0	12.5	2.341943	0	-0.25272	1.650767
5	0	12.029377	-3.58395	0	-0.21064	1.715113
6	0	11.318849	-0.29867	0	-0.25009	0.968223
7	0	-0.007461	0	0	-0.25009	0.968223

	M	N	O	P	Q	R	S	T
1	Q	POWER	ZT	NCOSI	NCOSR	NX	NY	NZ
2								
3								
4	2.341943	-0.69812	-6.6581	0.93218	1.630292	0	-0.36200	0.93218
5	1.862227	-0.07688	-6.084	1.24329	1.320172	0	0.54729	0.83694
6	3.373122	0.74793	-43.85	1.70162	0.953687	0	0.05274	0.99861
7	45.28939	0	0	0.96822	0.968223	0	0	1

	U	V	W	X
1	A	B	C	D
2				
3				
4	0.02896	1	4.52505	0.93218
5	-0.12688	1.00700	4.19056	1.24329
6	-0.01391	1.65469	11.32124	1.70162
7				

Table 1.14: Spreadsheet formulas for general raytrace.

col.	name	equation	example formula
G	X	1.66	G5 =G4+M5*J4
H	Y	1.66	H5 =H4+M5*K4
I	Z	1.66	I5 =O4+M5*L4
J	KX	1.80	J5 =J4+N5*R5
K	KY	1.80	K5 =K4+N5*S5
L	KZ	1.80	L5 =L4+N5*T5
M	Q	1.78	M5 =W5/(V5+IF(V5>0,X5,-X5))
N	POWER	$\Gamma' - \Gamma$	N5 =IF(P5<0,-Q5,Q5)-P5
O	ZT	1.83	O5 =I5-C5
P	NCOSI	1.81	P5 =R5*J4+S5*K4+T5*L4
Q	NCOSR	1.82	Q5 =SQRT(P5^2+D5^2-D4^2)
R	NX	1.70	R5 =-B5*G5
S	NY	1.70	S5 =-B5*H5
T	NZ	1.70	T5 =1-B5*I5
U	A	1.74	U5 =B5*D4^2
V	B	1.75	V5 =L4-B5*(J4*G4+K4*H4+O4*L4)
W	C	1.76	W5 =B5*(G4^2+H4^2+O4^2)-2*O4
X	D	1.77	X5 =SQRT(V5^2-U5*W5)
	Q	1.79	M7 =-O6/L6

```
object point:     0.000000     0.000000     0.000000

Surface 1
surface intersection     0.000000    12.500000     2.341943
optical direction cosines     0.000000    -0.252721     1.650767
surface normal     0.000000    -0.362004     0.932177
q     1.862227 gcosi     0.932177 gcosr     1.630292

Surface 2
surface intersection     0.000000    12.029377    -3.583954
optical direction cosines     0.000000    -0.210644     1.715113
surface normal     0.000000     0.547287     0.836945
q     3.373122 gcosi     1.243290 gcosr     1.320172

Surface 3
surface intersection     0.000000    11.318849    -0.298668
optical direction cosines     0.000000    -0.250087     0.968223
surface normal     0.000000     0.052737     0.998608
q    45.289397 gcosi     1.701618 gcosr     0.953687

Surface 4
surface intersection     0.000000    -0.007461    -0.000000
optical direction cosines     0.000000    -0.250087     0.968223
surface normal     0.000000     0.000000     1.000000
q     0.000000 gcosi     0.968223 gcosr     0.968223
```

Figure 1.16: Axial raytrace of doublet lens using raytrace.c

Table 1.15: OPD calculation for ray at edge of aperture for doublet lens.

q_c	q	OPD
0	2.34194265	-2.341943
5.38922156	1.86222705	9.836435
1.44675926	3.37312229	-5.752089
0	1.73595218	-1.735952
Total OPD (mm)		0.006451
Total OPD (waves)		11.73

either side. Table 1.13 already includes a dummy entrance plane (Row 3). Table 1.15 shows the OPD for a ray at the edge of the aperture for the doublet lens used in Table 1.13. The total optical path length along the reference ray is 19.35 (9*1.67+2.5*1.728). A precision of .002 waves in OPD at a wavelength of 0.00055 is 10^{-6}, so for a total path of order 10^2, pathlengths must be calculated with a relative precision better than 10^{-8}. Double precision arithmetic on a computer has a precision of about 10^{-12}. Single precision arithmetic has a precision of only 10^{-6} and would not be adequate.

A meridional OPD yfan for the doublet lens is shown in Fig. 1.17. The wavefront on-axis is an even function of the aperture radius, as shown in the plot. By comparison, the transverse aberrations on-axis are odd functions of the aperture radius as shown in Fig. 1.8. Table 1.5 listed Taylor series coefficients for the wavefront such that the derivative of the wavefront was a least-squares best fit to the transverse ray aberrations. The wavefront at the edge of the aperture (y=1)is obtained by summing the coefficients along the rows of Table 1.5. This sum has the value 11.73, in agreement with the result shown in Table 1.15.

1.8 General Differential rays

We present here the equations for a general differential raytrace.

The differential ray is specified by a position differential $\delta\mathbf{p}$ and a direction differential $\delta\mathbf{k}$.

Surface intersection The equation for surface intersection is

$$\mathbf{s} = \mathbf{p} + q\mathbf{k} \tag{1.85}$$

The differential change in surface intersection is given by

$$\delta\mathbf{s} = (\delta\mathbf{p} + q\delta\mathbf{k}) + (\delta q)\mathbf{k} \tag{1.86}$$

subject to the condition that the differential ray lies in the tangent plane

$$\mathbf{N} \cdot (\delta\mathbf{s}) = 0 \tag{1.87}$$

y	OPD (waves)
0	0
0.1	0.0025
0.2	0.039
0.3	0.195
0.4	0.601
0.5	1.415
0.6	2.789
0.7	4.809
0.8	7.396
0.9	10.103
1	11.729

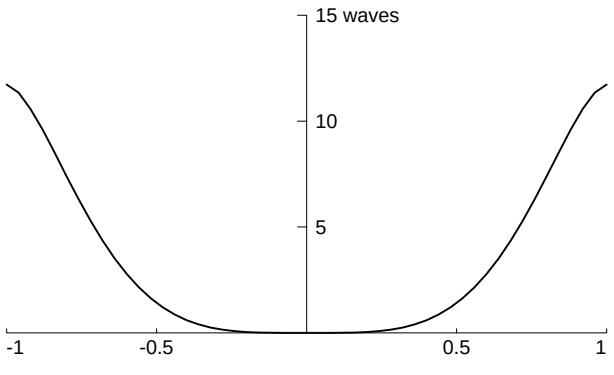


Figure 1.17: Meridional OPD fan for doublet lens (on-axis, 0 field).

where $\mathbf{N}$ is the surface normal vector. Then

$$\delta q = -\frac{1}{\Gamma}\left[\mathbf{N} \cdot (\delta \mathbf{p} + q\delta \mathbf{k})\right] \tag{1.88}$$

where

$$\Gamma = \mathbf{N} \cdot \mathbf{k} \tag{1.89}$$

Ray refraction The refracted ray is given by

$$\mathbf{k}' = \mathbf{k} + Q\mathbf{N} \tag{1.90}$$

where

$$Q = \Gamma' - \Gamma \tag{1.91}$$

and the differential refracted ray is given by

$$\delta \mathbf{k}' = (\delta \mathbf{k} + Q\delta \mathbf{N}) + (\delta Q)\mathbf{N} \tag{1.92}$$

subject to the condition that

$$\mathbf{k}' \cdot \delta \mathbf{k}' = 0 \tag{1.93}$$

since the magnitude of $\mathbf{k}'$ is fixed. Then

$$\delta Q = -\frac{1}{\Gamma'}\left[\mathbf{k}' \cdot (\delta \mathbf{k} + Q\delta \mathbf{N})\right] \tag{1.94}$$

The differential change in the surface normal $\delta \mathbf{N}$ is obtained from the Hessian matrix:

$$\begin{aligned}
\delta N_x &= \frac{\partial^2 F}{\partial x \partial x}\delta s_x + \frac{\partial^2 F}{\partial x \partial y}\delta s_y + \frac{\partial^2 F}{\partial x \partial z}\delta s_z \\
\delta N_y &= \frac{\partial^2 F}{\partial y \partial x}\delta s_x + \frac{\partial^2 F}{\partial y \partial y}\delta s_y + \frac{\partial^2 F}{\partial y \partial z}\delta s_z \\
\delta N_z &= \frac{\partial^2 F}{\partial z \partial x}\delta s_x + \frac{\partial^2 F}{\partial z \partial y}\delta s_y + \frac{\partial^2 F}{\partial z \partial z}\delta s_z
\end{aligned} \tag{1.95}$$

For a spherical surface, the differential normal vector simplifies to

$$\delta \mathbf{N} = -(\delta \mathbf{s})c \tag{1.96}$$

Given a set of n linear equations in m, where n is larger than m,

$$\mathbf{cx} = \mathbf{d} \tag{1.97}$$

with sizes

$$\text{(n x m)(m x 1) = (n x 1)}, \tag{1.98}$$

we seek a solution for $\mathbf{x}$ which minimizes the mean squared deviation ϕ^2 defined by

$$\phi^2 = \frac{1}{n}|\mathbf{cx} - \mathbf{d}|^2 \tag{1.99}$$

1.9 Normal Equations of Least-Squares

Given a set of n linear equations in m variables, where n is larger than m,

$$\mathbf{cx} = \mathbf{d}, \tag{1.100}$$

with sizes

$$\text{(n x m)(m x 1) = (n x 1)}, \tag{1.101}$$

we seek a solution for $\mathbf{x}$ which minimizes the mean squared deviation ϕ^2 defined by

$$\phi^2 = \frac{1}{n}|\mathbf{cx} - \mathbf{d}|^2. \tag{1.102}$$

We start by expressing ϕ^2 as

$$\phi^2 = \frac{1}{n}\sum_{i=1}^{n}|\sum_{j=1}^{m} c_{ij}x_j - d_i|^2 \tag{1.103}$$

and the expanding the square to give

$$\phi^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{m} c_{ij}x_j \sum_{k=1}^{m} c_{ik}x_k - 2\sum_{j=1}^{m}.c_{ij}x_j d_i + d_i^2\right) \tag{1.104}$$

Interchanging the order of summation gives

$$\phi^2 = \frac{1}{n}\sum_{j=1}^{m}\left(\sum_{k=1}^{m}\sum_{i=1}^{n} c_{ij}c_{ik}x_j x_k - 2x_j\sum_{i=1}^{n} c_{ij}d_i\right) + \frac{1}{n}\sum_{i=1}^{n} d_i^2. \tag{1.105}$$

Next we take the derivative of ϕ^2 with respect to x_j and set the result to zero. The result is

$$\frac{\partial \phi^2}{\partial x_j} = \frac{2}{n}\left(\sum_{k=1}^{m}\sum_{i=1}^{n} c_{ij}c_{ik}x_k - \sum_{i=1}^{n} c_{ij}d_i\right) = 0, \tag{1.106}$$

which can be simplified to

$$\sum_{k=1}^{m}\left(\mathbf{c}^\dagger\mathbf{c}\right)_{jk} x_k - \sum_{i=1}^{n} c_{ij}d_i = 0. \tag{1.107}$$

Using all values of j leads to the following linear system of equations

$$\mathbf{C}\mathbf{x} = \mathbf{c}^\dagger\mathbf{d} \tag{1.108}$$

where

$$\mathbf{C} = \mathbf{c}^\dagger\mathbf{c}. \tag{1.109}$$

The set of equations defined by Equation 1.108 are the normal equations of least squares. The solution has the nominal form

$$\mathbf{x} = \mathbf{C}^{-1}\mathbf{c}^\dagger\mathbf{d} \tag{1.110}$$