

## Modernizing the Teaching of Advanced Geometric Optics

Allen Nussbaum  
Electrical Engineering Department  
University of Minnesota  
Minneapolis MN 55455 USA

### ABSTRACT

Advanced geometric optics has been traditionally presented to students as subjects of great mathematical complexity. In addition, there is confusion and misunderstanding about the nature of aberrations. We show here how to improve understanding and make the material more enjoyable through the use of the following pedagogic devices:

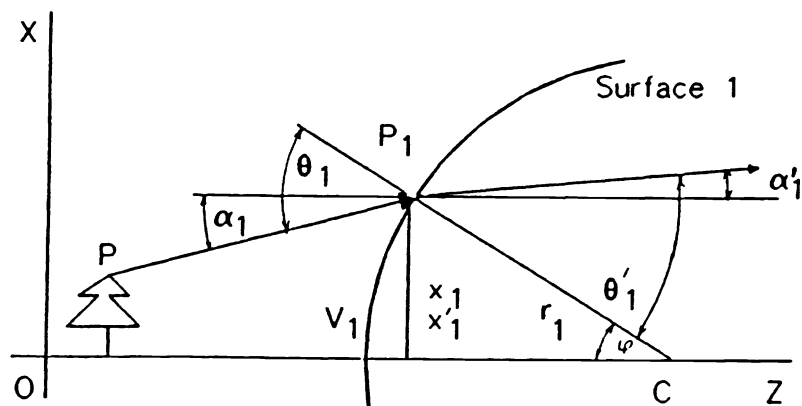
- (1) A matrix approach to paraxial optics.
- (2) A numerical treatment of non-paraxial optics and aberrations.

### 1. PARAXIAL MATRIX OPTICS

In many books<sup>1</sup>, the elementary lens equations are derived by starting with an object in front of a lens, locating the image produced at the first surface, and using this image as an object for the second surface. The algebra involved in this procedure is confusing, as are the typical sign conventions. We give a much simpler derivation of all of paraxial optics using matrices. Figure 1 shows a ray of light leaving point P on an object, striking the first surface of a lens at P<sub>1</sub>, and refracted there. Snell's law for small angles, with  $\sin\theta$  approximated by  $\theta$ , is  $n_1\theta_1 = n'_1\theta'_1$ , which is the paraxial ("close to the axis") form. Using  $\theta_1 = \alpha_1 + \phi$ ,  $\theta'_1 = \alpha'_1 + \phi$ , where  $\phi$  can be specified paraxially as  $x_1/r_1$  in place of the tangent, converts Snell's law to

$$n'_1\alpha'_1 = \frac{n_1 - n'_1}{r_1}x_1 + n_1\alpha_1 \quad (1)$$

Figure 1



Combining this with the trivial relation  $x_1 = x'_1$  gives

$$\begin{pmatrix} n'_1 \alpha'_1 \\ x'_1 \end{pmatrix} = \begin{pmatrix} 1 & -k_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n_1 \alpha_1 \\ x_1 \end{pmatrix} \quad (2)$$

where the constant  $k_1$  is the refracting power of surface 1 and is defined as  $k_1 = (n'_1 - n_1)/r_1$ . The square matrix is the refraction matrix  $R_1$  for surface 1, written as

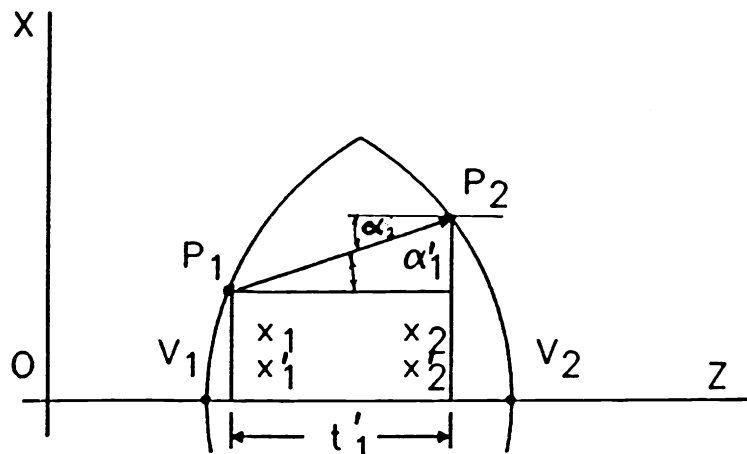
$$R_1 = \begin{pmatrix} 1 & -k_1 \\ 0 & 1 \end{pmatrix} \quad (3)$$

For the ray traveling from surface 1 to surface 2 (Figure 2), the distance from the axis becomes (again paraxially)

$$x_2 = x'_1 + t'_1 \alpha'_1 \quad (4)$$

Another approximation is to regard  $t'_1$  as being equal to the distance between the lens vertices  $V_1$  and  $V_2$ . Combining this with another trivial identity  $\alpha'_1 = \alpha_2$  gives the matrix equation

Figure 2



$$\begin{pmatrix} n_2 \alpha_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ t'_1/n'_1 & 1 \end{pmatrix} \begin{pmatrix} n'_1 \alpha'_1 \\ x'_1 \end{pmatrix} \quad (5)$$

where the translation matrix  $T_{21}$  is

$$T_{21} = \begin{pmatrix} 1 & 0 \\ t'_1/n'_1 & 1 \end{pmatrix} \quad (6)$$

The equation which combines a refraction, a translation, and a second refraction is then

$$\begin{pmatrix} n'_2 \alpha'_2 \\ x'_2 \end{pmatrix} = R_2 T_{21} R_1 \begin{pmatrix} n_1 \alpha_1 \\ x_1 \end{pmatrix} \quad (7)$$

This process can obviously be extended to any number of refracting (or reflecting) surfaces. The product  $S_{21}$  of the three square matrices is the system matrix and is also written as

$$R_2 T_{21} R_1 = \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \quad (8)$$

where the Gaussian constants  $a$ ,  $b$ ,  $c$ , and  $d$  can be obtained by combining the above equations to obtain

$$a = k_1 + k_2 - \frac{k_1 k_2 t'_1}{n'_1}, \quad d = -\frac{t'_1}{n'_1} \quad (9)$$

and

$$b = 1 - \frac{k_2 t'_1}{n'_1}, \quad c = 1 - \frac{k_1 t'_1}{n'_1} \quad (10)$$

For an object at  $t$  and an image at  $t'$ , we can connect the initial and final rays with the equation

$$\begin{pmatrix} n'_2 \alpha'_2 \\ x'_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ t'/n'_2 & 1 \end{pmatrix} \begin{pmatrix} b & -a \\ -d & c \end{pmatrix} \begin{pmatrix} 1 & 0 \\ t/n_1 & 1 \end{pmatrix} \begin{pmatrix} n_1 \alpha_1 \\ x_1 \end{pmatrix} \quad (11)$$

where the three  $2 \times 2$  matrices on the right-hand side combine as an object-image matrix  $S_{P'P}$

$$S_{P'P} = \begin{pmatrix} b - \frac{at}{n_1} & -a \\ \frac{bt'}{n'_2} - \frac{att'}{n_1 n'_2} - d + \frac{ct}{n_1} & c - \frac{at'}{n'_2} \end{pmatrix} \quad (12)$$

This matrix implies that the value of the magnification  $m = x'/x$  will depend on  $\alpha_1$ , which leads to an imperfect image. This means the lower left-hand element in the matrix must vanish, so that

$$\frac{t'}{n'_2} = \frac{d - ct/n_1}{b - at/n_1} \quad (13)$$

and, in addition,  $m = c - at'/n'_2$ . The positions of the unit planes  $H$  and  $H'$ , the locations for which object and image are identical, are found by letting  $m = 1$ . Then

$$t' = l'_H = n'_2 (c - 1) / a \quad (14)$$

with a similar expression for  $l_H$ . The focal point  $F'$  is located by letting  $x' = 0$ , so that  $l'_P = n'_2 c/a$ , and similarly for  $l_P$ . The focal lengths  $f$  and  $f'$  are defined as the distances between the respective focal and unit points. It then follows that  $1/a = f' = -f$ . Let a ray leave a point on the axis at an angle  $\alpha_1$  and emerge at an angle  $\alpha'_2$ . Since  $x = 0$  at the starting point, then  $n'_2 \alpha'_2 = n_1 \alpha_1/m$ , where  $\alpha'_2/\alpha_1$  is the angular magnification  $\mu$ , so that  $\mu = 1/m$  in air. The respective locations of the object and image for unit angular magnification are the nodal points  $N$  and  $N'$ . The six points  $F, F', H, H', N, N'$  comprise the cardinal points. The equal angle condition shows that the nodal point locations are

$$\frac{l_N}{n_1} = \frac{(n'_2/n_1) - b}{a} \quad (15)$$

and

$$\frac{l'_N}{n'_2} = \frac{c - (n_1/n'_2)}{a} \quad (16)$$

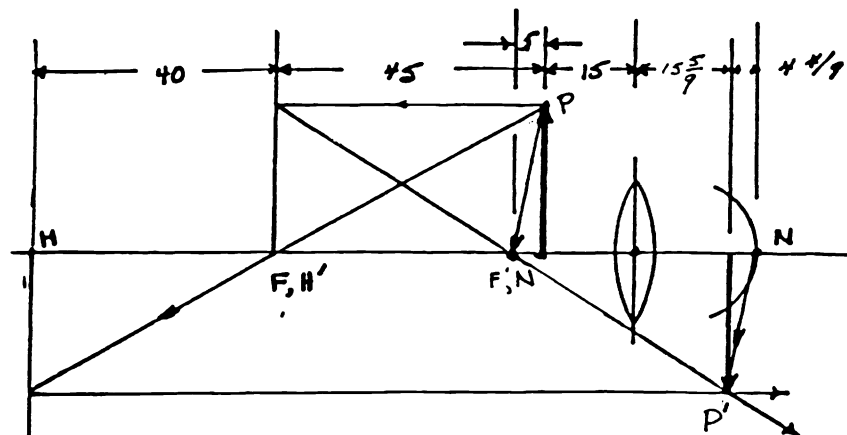
The sign conventions used (and summarized by O'Neill<sup>3</sup>) are mostly the familiar Cartesian rules; to these are added the stipulations that a surface opening to the right(left) has a positive(negative) curvature and that a reversal of direction due to a reflection reverses the index.

The power of paraxial matrix optics is indicated by combining the definition of  $a$ , or  $1/f'$ , in (9) with the definitions of  $k_1$  and  $k_2$  to obtain

$$\frac{1}{f'} = (n'_1 - 1) \left[ \frac{1}{r_1} - \frac{1}{r_2} + \frac{(n'_1 - 1) t'_1}{n'_1 r_1 r_2} \right] \quad (17)$$

which is the lensmakers' equation. This derivation is far more direct than usually found, and can easily be extended to an arbitrary combination of components. As another example of the ease of use of matrices, an object is 15 units to the left of a converging thin lens of focal length 10 units. For such a lens, with  $t'_1$  negligible, (9) and (10) show that  $a = 1/f'$ ,  $b = c = 1$ ,  $d = 0$ . A concave mirror, with radius of 16 units, is 20 units to the right of the lens. The Gaussian constants are then obtained by matrix multiplication. The cardinal points follow from the equations above, and are located to scale in Figure 3. We find the image by ray tracing.

Figure 3



A ray from P is extended backwards to the unit plane H must then proceed to the right (the positive direction) and parallel to the axis. The converse of this is a ray extended backwards and parallel until it reaches H'; it then passes through F' and the two rays intersect at P'. A check is provided by a third ray from P to N; it emerges at N' and goes to P' while remaining parallel to its original direction. The values of t and t' shown are confirmed by the matrix calculations. This situation can be readily extended to an arbitrary number of thick lenses.

Yet another example, and a truly remarkable one, is the Schwarzschild reflecting microscope objective, which has of two concentric mirrors of radii  $r_1 = -(\sqrt{5} + 1)$  and  $r_2 = -(\sqrt{5} - 1)$  (Figure 4). An object on the axis at a distance  $w = 1$  from the common center will send a parallel beam up the tube. This may be confirmed by multiplying the two translation and two refraction (actually, reflection) matrices involved and then using the fact that if the initial position and the final angle are zero, the upper left-hand corner of this product matrix must vanish. Although this is a paraxial calculation, it is found that when the exact ray tracing is done by the method given below, the angle of acceptance can be of the order of  $25^\circ$  (Figure 5).

Figure 4

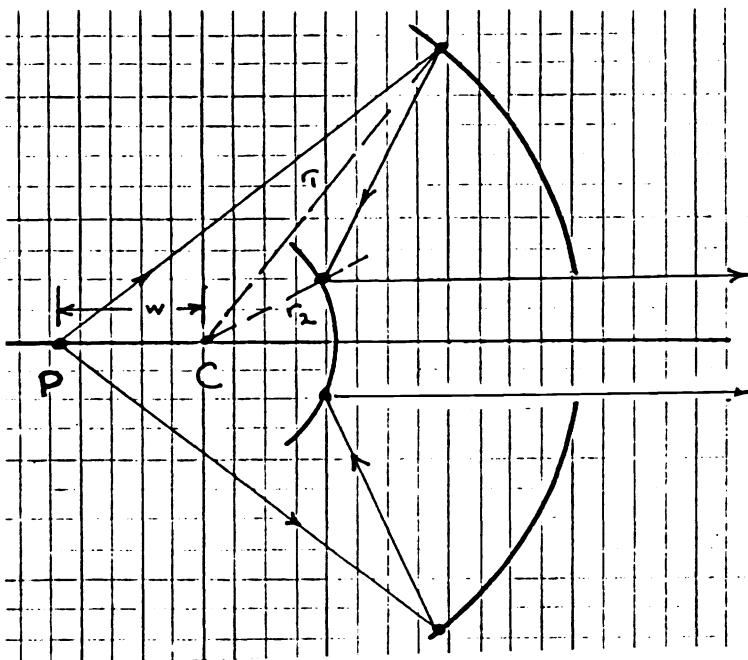
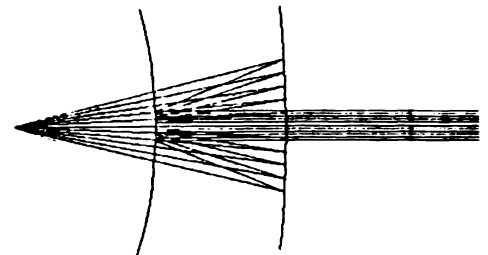


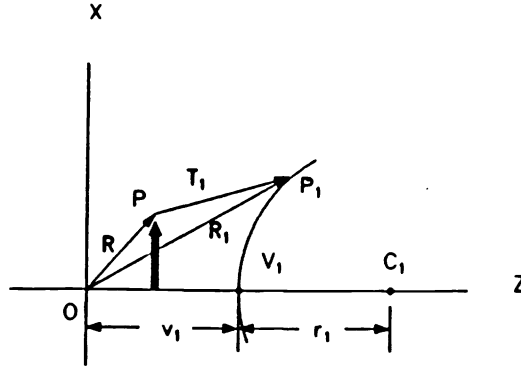
Figure 5



## 2. MERIDIONAL NON-PARAXIAL OPTICS

When the paraxial approximation is no longer valid, then the concepts of Part 1 lose their meaning, and we can no longer use terms such as cardinal points, focal length, and magnification. Instead, we must go to exact ray tracing. Figure 6 shows an object point P and a ray traveling to the first surface of a lens. A meridional ray leaves the object point P and strikes surface 1 at the point  $P_1$ . The translation is taken as a vector  $T_1$ , direction cosines L, M and N with respect to the x, y and z axes (note that OZ is taken as the symmetry axis). The direction cosines L and N are defined in terms of the components of  $T_1$  as

**Figure 6**



$$L = T_{1x}/T_1, \quad N = T_{1z}/T_1 \quad (18)$$

The location of the object point P and the intersection point P<sub>1</sub> are also designated as vectors, and the three vectors shown in the figure are related by

$$\mathbf{R}_1 = \mathbf{R} + \mathbf{T}_1 \quad (19)$$

Taking the dot or scalar product of each side with itself, we obtain

$$R_1^2 = R^2 + T_1^2 + 2T_1(zN + xL) \quad (20)$$

The equation of a circle in the coordinate system we are using may be written as

$$z_1^2 + x_1^2 = 2z_1(r_1 + v_1) - v_1^2 - 2v_1r_1 = R_1^2 \quad (21)$$

Combining this with equation (20) and eliminating z<sub>1</sub> gives

$$T_1^2 + 2T_1[N(z-r_1-v_1) + Lx] + (z-v_1)^2 + x^2 + 2r_1(v_1-z) = 0 \quad (22)$$

Using the curvature c<sub>1</sub> = 1/r<sub>1</sub>, this quadratic equation becomes

$$T_1 = \frac{F}{-E + \sqrt{E^2 - c_1 F}} \quad (23)$$

where

$$E = BC_1/2 = c_1 [(z - v_1)N + xL] - N \quad (24)$$

and

$$F = CC_1 = c_1 [(z-v_1)^2 + x^2] - 2(z-v_1) \quad (25)$$

(Full details of the derivation have been given previously<sup>4</sup>.) It is customary to shift the origin

after each translation, so that

$$z_1 = z + T_1 N - v_1 \quad (26)$$

and the other equation is

$$x_1 = x + T_1 L \quad (27)$$

Since the angles or the direction cosines remain unchanged during a translation, these two equations can be put into matrix form as

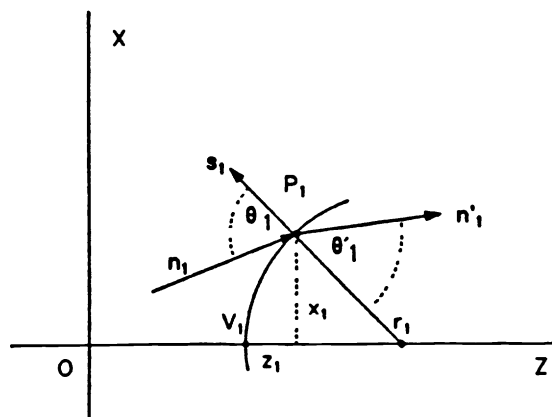
$$\begin{pmatrix} n_1 N \\ z_1 + v_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ T_1/n_1 & 1 \end{pmatrix} \begin{pmatrix} n_1 M \\ z \end{pmatrix} \quad (28)$$

and

$$\begin{pmatrix} n_1 L \\ x_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ T_1/n_1 & 1 \end{pmatrix} \begin{pmatrix} n_1 L \\ x \end{pmatrix} \quad (29)$$

The 2x2 matrix in these equations is the non-paraxial or exact translation matrix. The form of the exact refraction matrices can be obtained from Figure 7, which shows the incoming ray at surface 1. This ray is taken to be a vector  $\mathbf{n}_1$  with magnitude equal to the index of refraction on the left. Then the refracted ray is designated as  $\mathbf{n}'_1$  in a similar way.

Figure 7



Define a quantity  $K_1$ , the refracting power or skew power, in terms of these two rays by the relation

$$K_1 \mathbf{s}_1 = c_1 (\mathbf{n}_1 - \mathbf{n}'_1) \quad (30)$$

where  $\mathbf{s}_1$  is a unit normal vector at the surface. The scalar product of both sides of this equation with this vector gives

$$K_1 = c_1 (n_1' \cos \theta_1' - n_1 \cos \theta_1) \quad (31)$$

We can express the scalar product of the incoming ray and the normal in two ways; these are

$$\mathbf{n}_1 \cdot \mathbf{s}_1 = n_{1z} s_{1z} + n_{1x} s_{1x} = n_1 N_1 \frac{z_1 - r_1}{r_1} + n_1 L_1 \frac{x_1}{r_1} \quad (32)$$

and

$$\mathbf{n}_1 \cdot \mathbf{s}_1 = n_1 \cos (180^\circ - \theta_1) = -n_1 \cos \theta_1 \quad (33)$$

Combining these equations and using the direction cosines for the unit normal as indicated by the figure, we obtain

$$\cos \theta_1 = N_1 (1 - c_1 z_1) - L_1 x_1 c_1 \quad (34)$$

where  $z_1$  is measured from  $V_1$ , as mentioned just above. Having found the incident angle, we can find the angle after refraction with Snell's law, obtaining

$$K_1 = c_1 [n_1' \sqrt{1 - (n_1/n_1')^2 (1 - \cos^2 \theta_1)} - n_1 \cos \theta_1] \quad (35)$$

Write (30) as the two scalar equations in matrix form

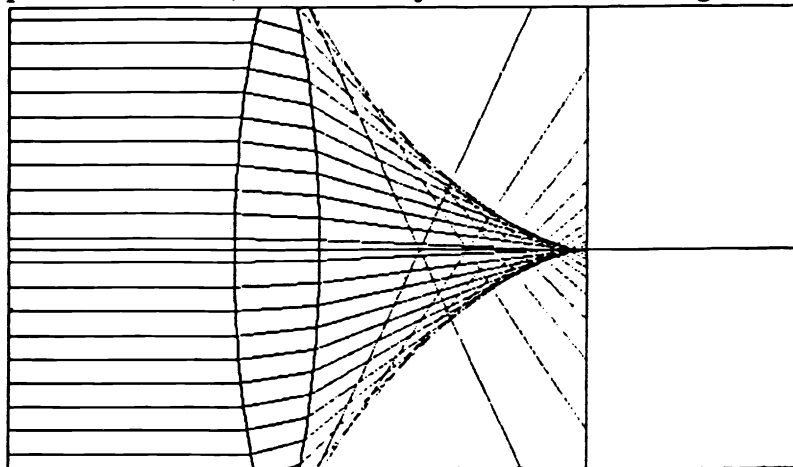
$$\begin{pmatrix} n_1' N_1' - K_1 / c_1 \\ z_1' \end{pmatrix} = \begin{pmatrix} 1 & -K_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n_1 N_1 \\ z_1 \end{pmatrix} \quad (36)$$

and

$$\begin{pmatrix} n_1' L_1' \\ x_1' \end{pmatrix} = \begin{pmatrix} 1 & -K_1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} n_1 L_1 \\ x_1 \end{pmatrix} \quad (37)$$

These equations show that the non-paraxial translation and refraction matrices are generalizations of the paraxial matrices, to which they reduce for small angles.

Figure 8





The procedure just developed has been applied to parallel meridional rays going through a double convex lens with radii of 50 units, thickness of 15 units, and an index of 1.5. The graphics capability of the language QuickBASIC 4.5 is demonstrated by Figure 8, which shows the point

defect called spherical aberration, with the associated caustic surface and circle of least confusion. This figure emphasizes that spherical aberration is the failure of meridional rays to obey the paraxial approximation, a definition unlike like what most optics books use<sup>1</sup>. This program has been extended<sup>5</sup> to show how spherical aberration can be reduced by altering the radii while holding the thickness, index and focal length constant, a process called bending the lens. It clearly demonstrates the advantage of numerical modeling, since the experiment cannot be done on an optical bench.

### 3. NON-MERIDIONAL NON-PARAXIAL OPTICS

By adding a term  $yM$  to (24) and a term  $y^2$  to (25), as well as  $-M_1L_1c_1$  to (34), the above meridional procedure can be applied to non-meridional or skew rays. These changes imply the existence of three translation and three refraction matrices, one pair for each of the coordinates. A straightforward way of observing the aberration associated with skew rays is to consider a cylinder of rays concentric to the axis and striking the first surface of a lens; it will form a sharp image on axis somewhere to the left of the paraxial focus. Now incline this cylinder at an angle so that the rays strike the original circle on the surface, which means that the cylinder will have an elliptical cross-section. The top and bottom rays, which are meridional, will define an image plane. All the other rays, which are skew, will produce two approximately circular images, not quite coincident. This effect can be demonstrated using the ray tracing procedure just developed and is shown in Figure 9. One circle in each pair is produced by incident rays covering either

Figure 9

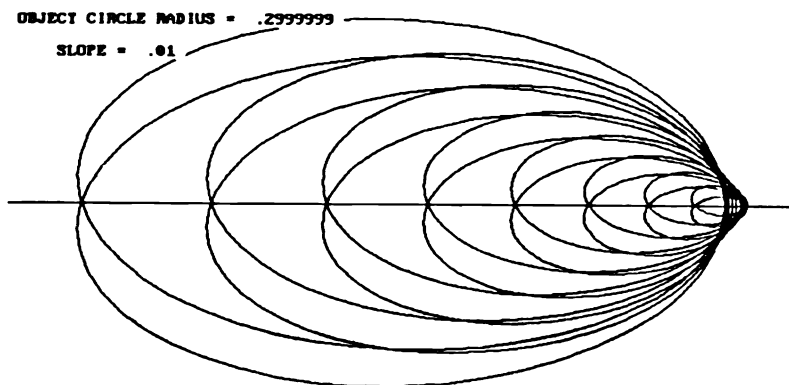
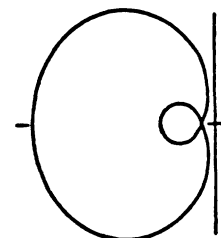


Figure 10



the left-hand or the right-hand half of the circle on the first surface of the lens. The combination of the two image patterns produces cardioid patterns, as shown in Figure 10, which represents coma as it should occur for an uncorrected system. Hence, we define coma--in parallel with our definition of spherical aberration--as the failure of skew rays to obey the paraxial approximation. This definition contradicts the customary one<sup>1</sup> and the patterns of Figure 10 are not found in the standard texts. However, this effect is known to professional photographers<sup>6</sup>. Let us now

assume that an optical system has been corrected both for spherical aberration and for coma. There is no guarantee that these separate corrections will bring the meridional rays and the skew rays from an off-axis point to a common focus. We thus see the source of astigmatism, which is the failure of the two corrections to agree. This explains the classical illustration of this aberration, found in all textbooks<sup>1</sup>. A meridional fan of rays leaving an off-axis object point forms a sharp image. A second fan, at right angles to the first one, then represents one which is completely non-meridional (except for a single, central ray) and it also forms a sharp, but separate, image. The blur between these two points would be eliminated if they could be made identical.

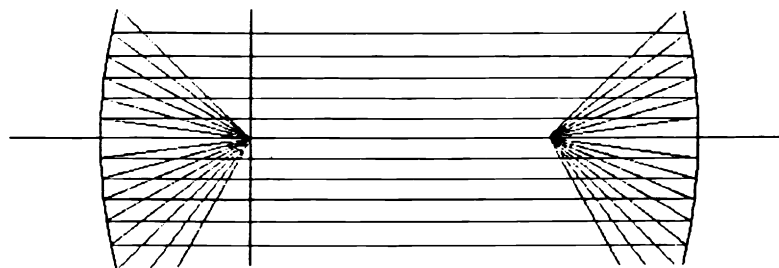
#### 4. SECOND-ORDER ASPHERIC SURFACES

Aspheric surfaces have been regarded as beyond the scope of undergraduate optics courses, and they are treated by Smith<sup>1</sup> with complicated iterative procedures. Restricting ourselves to second-order surfaces (paraboloids, ellipsoids, and hyperboloids), the above procedure can be extended easily once more by replacing the equation of the sphere--the three-dimensional form of (21)--with the conic section equation in the form

$$\frac{C(x^2 + y^2 - Sz^2)}{2} - z = 0 \quad (38)$$

where C is the vertex curvature and S is a shape factor which is negative for ellipsoids (S = -1 specifies a sphere), zero for paraboloids, and positive for hyperboloids. As a typical example, consider the design of an arc furnace composed of two parabolic mirrors of radius 100 units. The contaminating arc is at one focus and the sample to be heated is at the second focus, which can be placed at an arbitrary distance. The program which does the ray tracing is given below as an appendix and Figure 11 shows the resulting graphics. All the other examples given in this paper were handled with this program, or simplified versions for meridional or non-meridional spherical surfaces.

Figure 11



Other examples which students find interesting (and for which listings are available from the author<sup>5</sup>) are a keyboard-interactive lens-bending program (Figure 12), the formation of a virtual image by a double concave lens (Figure 13), and the Hubble Space Telescope (Figure 14); this instrument uses two hyperbolic mirrors, and its parameters were taken from a popular article.

Figure 12

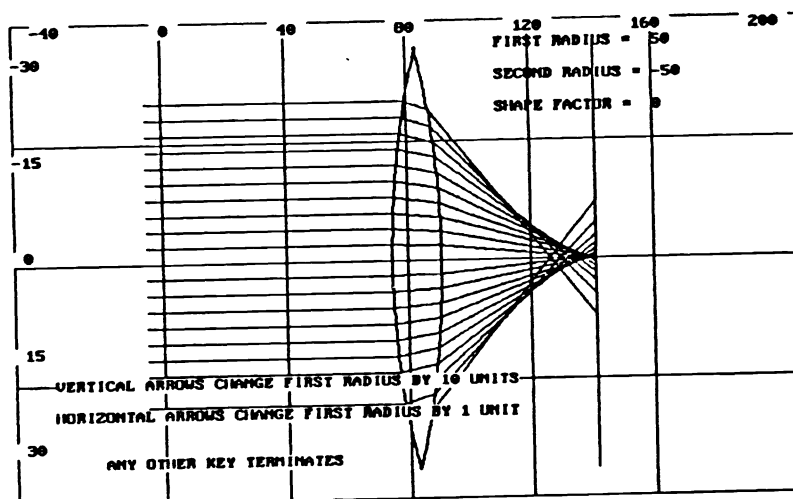
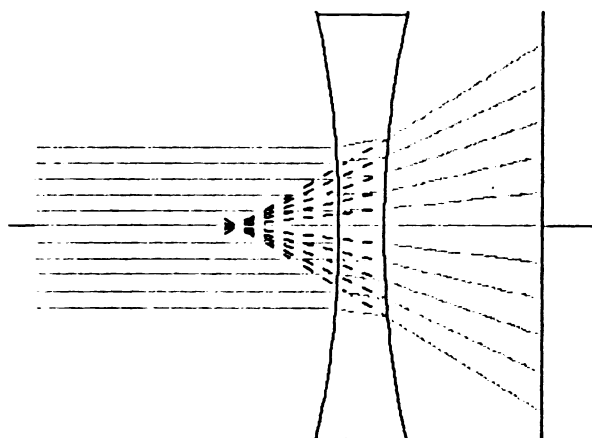
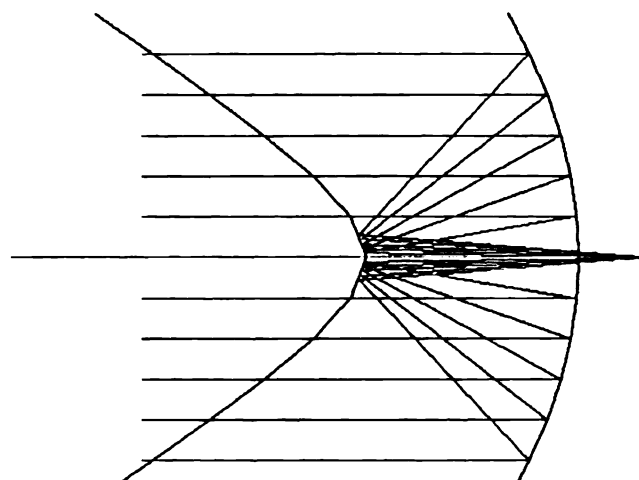


Figure 13



Press any key to continue

Figure 14



Press any key to continue

## 6. REFERENCES

1. W. J. Smith, *Modern Optical Engineering, 2nd ed*, McGraw-Hill Inc, (1990)
2. A. Nussbaum and R. J. Phillips, *Contemporary Optics*, Prentice-Hall, (1976)
3. E. L. O'Neill, *Introduction to Statistical Optics*, Addison-Wesley, (1963)
4. A. Nussbaum, Advanced Geometrical Optics on a Programmable Pocket Calculator, Amer J Phys **47**, 351 (1979)
5. A. Nussbaum, Course Notes for Optical System Design (Unpublished)
6. B. Sherman, Photographs of Optical Aberrations, Modern Photography **32**, 118 (1968)

## APPENDIX: ASPHERIC SKEW RAY TRACING

Parabolic Mirrors File A-PAR2 20 Oct 90

Set window

```
1 CLS : SCREEN 9
WL = -180: WR = 80: WD = -60: WU = -WD
WINDOW (WL, WD)-(WR, WU)
```

Specify constants

```
T(1) = 50: T(2) = -200: T(3) = 50
C(1) = -.01: C(2) = .01: C(3) = 0
N(1) = 1: N(2) = -1: N(3) = 1
NP(1) = -1: NP(2) = 1: NP(3) = 1
S(1) = 0: S(2) = 0: S(3) = 0
```

Ray tracing

```
ELO = -.8
10 X = 0: Y = 0: Z = 0
Z1 = Z: X1 = X: T1 = 0
EL = ELO: EM = 0: EN = SQR(1 - EL * EL - EM * EM)
FOR J = 1 TO 3
ZV = Z - T(J)
D = 1 - (S(J) + 1) * EN * EN
E = C(J) * (X * EL + Y * EM - S(J) * ZV * EN) - EN
F = C(J) * (X * X + Y * Y - S(J) * ZV * ZV) - 2 * ZV
SD = F / (SQR(E * E - D * F * C(J)) - E)
X = X + SD * EL
Y = Y + SD * EM
Z = Z + SD * EN - T(J)
T1 = T1 + T(J)
Z2 = Z + T1: X2 = X
LINE (Z1, X1)-(Z2, X2), 9
Z1 = Z2: X1 = X2
OP = 1 + S(J) * C(J) * Z
SQ = SQR((X * X + Y * Y) * C(J) * C(J) + OP * OP)
COSTHETA = (EN * OP - C(J) * (EL * X + EM * Y)) / SQ
BC = N(J) * N(J) * (1 - COSTHETA * COSTHETA) / NP(J) / NP(J)
KOVERC = NP(J) * SQR(1 - BC) - N(J) * COSTHETA
K = C(J) * KOVERC
EL = (N(J) * EL - K * X / SQ) / NP(J)
EM = (N(J) * EM - K * Y / SQ) / NP(J)
EN = (N(J) * EN + KOVERC * OP / SQ) / NP(J)
NEXT J
ELO = ELO + .1
IF ELO < .7 THEN 10
```

Show mirrors

```
T1 = 0
FOR Q = 1 TO 3
T1 = T1 + T(Q)
X8 = 0: Z8 = T1
FOR H = 1 TO 30
X = X8 + 3: Z = X * X * C(Q) / (1 + SQR(C(Q) * C(Q) * S(Q) * X * X + 1))
X9 = X: Z9 = Z + T1
LINE (Z8, X8)-(Z9, X9)
X8 = X9: Z8 = Z9
NEXT H
X8 = 0: Z8 = T1
FOR H = 1 TO 30
X = X8 - 3: Z = X * X * C(Q) / (1 + SQR(C(Q) * C(Q) * S(Q) * X * X + 1))
X9 = X: Z9 = Z + T1
LINE (Z8, X8)-(Z9, X9)
X8 = X9: Z8 = Z9
NEXT H
NEXT Q
```