5 FUNDAMENTAL METHODS OF RAY TRACING

5.1 GENERAL

5.1.1 Basic optical system. Every optical system consists of one or more reflecting or refracting surfaces. The function of the system is to transform the diverging spherical wavefronts coming from object points in object space to converging spherical wavefronts going towards image points in image space. As mentioned in paragraph 2.1.2 the passage of the wavefronts through the optical system can be most easily discussed by utilizing the concept of rays. The passage of rays through an optical system may be determined by purely geometrical considerations, since it is correct to make the following assumptions:

(1) A ray travels in a straight line in a homogeneous medium.

(2) A ray reflected at an interface obeys the law of reflection.

(3) A ray refracted at an interface obeys the law of refraction.

Computing the passage of rays through an optical system is a purely geometric problem best solved by the techniques of analytic geometry.

5.1.2 Centered optical systems.

5.1.2.1 Fortunately, nearly every theoretical optical system consists of centered refracting or reflecting surfaces. In a centered optical system all surfaces are rotationally symmetrical about a single axis. A cross-section view of a typical photographic lens is shown in Figure 5.1. In this case all the surfaces are spherical surfaces and the centers are assumed to lie on the optical axis. Herein lies one of the differences between theory and practice. In the design phase, the system is assumed to have an axis of symmetry. In practice the lenses may not be lined up perfectly so it will not be a centered optical system. If the lens is to perform according to the design, the lenses must be adjusted until they are centered. Procedures to assure centering of the elements are a prime consideration in the mechanical design of optical instruments.

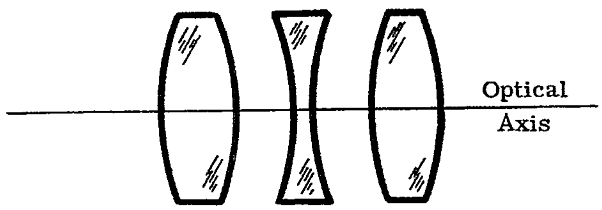

Figure 5.1 - A cross-section view of a photographic lens.

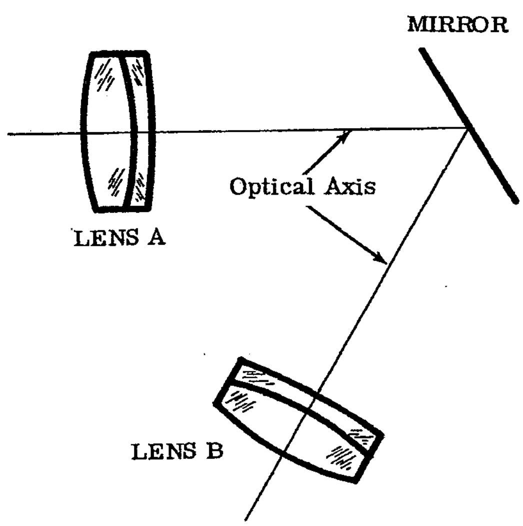

Figure 5.2–An optical system containing a mirror.

5.1.2.2 The optical system shown in Figure 5.2 may not appear at first glance to be a centered optical system. The optical axes of the two lenses do not coincide. However, if properly constructed this may be a centered optical system. To understand this consider Figure 5.3, which shows how a system involving plane mirrors can be thought of as folded out. These ideas are treated in detail in Section 13.

5.1.2.3 Consideration of the law of reflection shows that the ray of light traveling along the optical axis from lens A is actually deflected, but can be thought to continue straight through the mirror. If the axis of lens B lies on the extended axis of lens A, then the system is a centered optical system. One can see that if lens B of Figure 5.3 is shifted to the left or right, there will be a corresponding shifting of lens B' up or down; the system will become decentered and lose its axial symmetry.

5.1.2.4 The sections on geometrical optics in this handbook consider centered systems. Decentered systems usually, when carefully analyzed, are seen to be part of some centered system. Hence if a final design calls for a decentered system, the preliminary design considers the centered system as a basic starting point.

5.1.3 <u>Plane, spherical and aspheric surfaces.</u>

5.1.3.1 Production techniques for generating plane and spherical surfaces on optical materials are well established and thus these are most commonly used. Aspheric surfaces, however, offer certain advantages, and recent advances in the generation of this type of surface, coupled with the need for the design refinements they offer, have resulted in more frequent design application of this type. Aspheric surfaces are also usually considered to have rotational symmetry about the optical axis.

5.1.3.2 In ray tracing, plane surfaces will be considered to be special cases of spherical surfaces, having radii equal to infinity; hence no special technique for plane surfaces will be developed in detail in this section. In Section 13, reflection from plane surfaces is considered more fully. The technique for treating aspheric surfaces is developed by extending the technique for spherical surfaces. In both cases, the surfaces are considered to be centered.

5.1.4 <u>Ray tracing, the basic tool of optical design.</u>

5.1.4.1 In order to understand clearly the kind of image formed by a system, and what must be done to im-

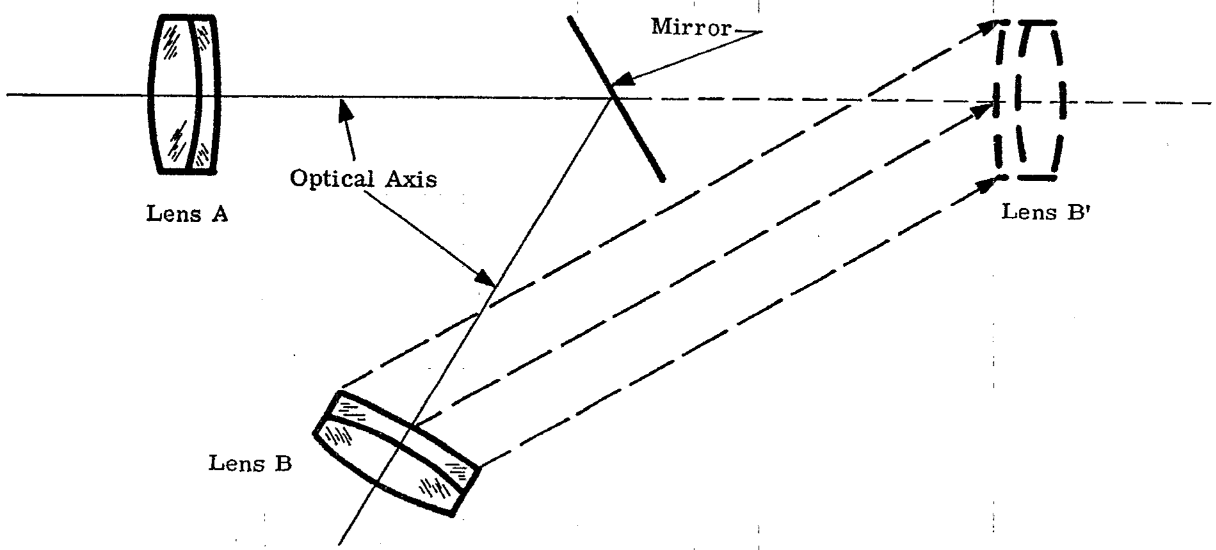

Figure 5.3 - Diagram showing "folding out" of an optical system containing a mirror.

prove this image, a certain number of rays must be determined in their passage through the system. This process of ray tracing involves the determination of the direction and location in space of each segment of a ray as it goes from object to image. Since the function of the system is to transfer light from an object surface to an image surface, the object surface and the image surface, although neither reflecting nor refracting, can be considered as surfaces of the optical system.

5.1.4.2 Figure 5.4 shows a cross-section view of a centered optical system. The ray, consisting of seven straight line segments, goes from the object point, O, on the object surface, to the image point, O', on the image surface, being refracted at six intermediate surfaces. The remainder of Section 5 will be concerned with numerical and graphical methods of determining the course of general and special rays through a general system.

5.2 DEFINITIONS AND CONVENTIONS

5.2.1 Need for specific conventions. The ray tracing formulae to be used for tracing a ray through a system involve parameters of more than a single surface or a single medium. Therefore, it is important to adopt a convention of notation which will clearly distinguish one surface from another and one medium from another. In addition, many optical systems employ mirrors, so that the rays sometimes proceed in a direction generally opposite to the incident rays. Our conventions should be such that a reflecting surface can be handled as any other general refracting surface. It is assumed that before applying these conventions the system has been folded out in the sense of Figure 5.3.

5.2.2 Statements of definitions and conventions. The following definitions and conventions, which are in agreement with those given in MIL-STD-34, will be used in Sections 2, and 5 through 15, inclusive. Reference to Figures 5.4 and 5.5 will indicate examples of some of these conventions.

(1) It will be assumed that light initially travels from left to right.

(2) An optical system will be regarded as a series of surfaces starting with an object surface and ending with an image surface. The surfaces will be numbered consecutively, in the order in which the light is incident on them, starting with zero for the object surface and ending with k for the image surface. A general surface will be called the j th surface.

(3) All quantities between surfaces will be given the number of the immediately preceding surface.

(4) A primed superscript will be used to denote quantities after refraction only when necessary.

(5) r_j is the radius of the jth surface. It will be considered positive when the center of curvature lies to the right of the surface.

(6) The curvature of the jth surface is $c_j = 1/r_j$. c_j has the same sign as r_j.

(7) t_j is the axial thickness of the space between the jth and the $j + 1$ surface. It is positive if the $j + 1$ surface physically lies to the right of the jth surface. Otherwise it is negative.

(8) n_j is the index of the material between the jth and the $j + 1$ surface. It is positive if the physical ray is traveling from left to right. Otherwise it is negative.

(9) K_j, L_j, M_j are the products of n_j and the direction cosines (with respect to the X, Y, Z axes respectively) of a ray in the space between the jth and the $j + 1$ surface. They will be called the optical direction cosines.

(10) The right-handed coordinate system shown in Figure 5.5 will be used. The optical axis will coincide with the Z axis. The light travels initially toward larger values of Z. Positive values of X are away from the reader in Figure 5.5.

(11) X_j, Y_j, Z_j are the position coordinates of a ray where it intersects the jth surface.

(12) In writing formulae where no confusion is likely to result, the j will be omitted from the subscript. Thus the curvature of the $j - 1$ surface will be written c_{-1}, the curvature of the j th surface will be written c and the curvature of the $j + 1$ surface will be written c_{+1}.

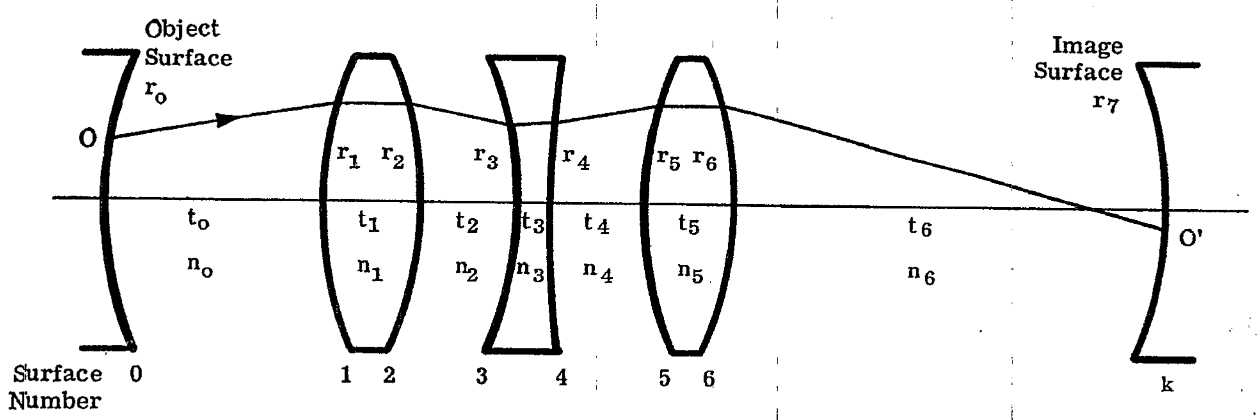

Figure 5.4—Cross-sectional view of a centered optical system.

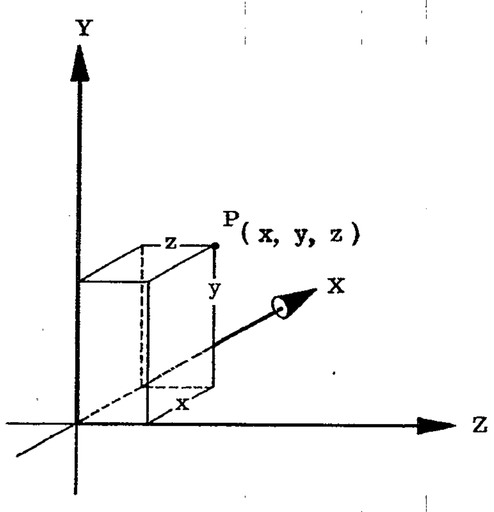

Figure 5.5- Right-handed coordinate axes.

5.3 BASIC RAY TRACE PROCEDURE

5.3.1 Transfer procedure. As can be seen from Figure 5.4 a ray travels in a straight line from a point on one surface to a point on the following surface. It is then refracted and proceeds to the next surface in a straight line. The ray tracing procedure then consists of two parts, the transfer procedure, and the refraction procedure. The transfer procedure involves computing the intersection point of the ray on the surface from the optical direction cosines and the intersection point data at the previous surface. That is, given K_{-1}, M_{-1}, L_{-1} and X_{-1}, Y_{-1}, Z_{-1}, compute X, Y, Z. The equations used are called the transfer equations.

5.3.2 Refraction procedure. The refraction procedure involves computing the optical direction cosines of a ray from the intersection point data and the optical direction cosines of the previous ray segment. That is, given X, Y, Z and K_{-1}, M_{-1}, L_{-1}, compute K, L, M. The equations used are called the refraction equations.

5.3.3 Repetition for successive surfaces. After having applied the two procedures, we have the initial data for the next application. The transfer equations will be used to compute X_{+1}, Y_{+1}, Z_{+1} and the refraction equations will be used to compute K_{+1}, L_{+1}, M_{+1}. It should be noted that it is often convenient to introduce fictitious or non-refracting surfaces to simplify the procedure. One example is the tangent plane, an XY plane tangent to a physical surface at the optical axis. Another example is a sphere, tangent to an aspheric surface at the optical axis. These fictitious surfaces are handled in exactly the same manner as a physical surface. Transfer equations are used to go to or from such a surface. The refraction equation reduces to $I = I'$, and the direction cosines of the refracted ray equal those of the incident ray, as would be expected at a non-refracting surface. Fictitious surfaces will be used in the next section.

5.4 SKEW RAY TRACE EQUATIONS FOR SPHERICAL SURFACES

5.4.1 Types of rays.

5.4.1.1 A general ray is any ray passing from any object point through the optical system to its image point on the image surface. A special ray that lies in a plane containing the optical axis and the object point is called a meridional ray. Any non-meridional ray is a skew ray. A ray close to the optical axis is a paraxial ray. Because of the approximation involved, a paraxial ray is a special type of meridional ray. A skew ray is considered to be non-paraxial since it is non-meridional. These distinctions will become apparent as the subject is developed.

5.4.1.2 Corresponding to the three types of rays, skew, meridional, and paraxial, we will develop three sets of ray trace equations and procedures. Because the three rays, in the order given here, become less general and more specialized, the equations relating to these types of rays become simpler as we proceed from skew through meridional to paraxial. One method of developing the subject would be to discuss the simplest case first (paraxial), then proceed to the more complicated (meridional) and finally to the most general (skew). This procedure would have the advantage of beginning with the simplest derivation. However, it would necessitate three separate derivations.

5.4.1.3 We will proceed in the other direction, beginning with the most general case, the skew ray trace. From this the meridional and paraxial equations follow by simplification; hence only one derivation is necessary, instead of three. The particular equations derived in Section 5.4 are set up in a form for an electronic computer. However they are completely satisfactory for use with a desk calculator, and represent a good starting point for the human computer who has not yet worked out his own equations.

5.4.2 Initial data for a skew ray. Figure 5.6 shows the skew ray as it traverses the space between two surfaces. At the right hand surface it is refracted, and a drawing corresponding to Figure 5.6 could show this ray as it traverses the space between the j th and j_{+1} spherical surfaces. Similarly, another drawing could show the ray before refraction at the j_{-1} surface. The initial data for the ray we are considering will consist of the emergence point with the left surface, and the direction of the ray in space. Hence we specify X_{-1}, Y_{-1}, and Z_{-1}, the coordinates on the j_{-1} surface, and K_{-1}, L_{-1}, and M_{-1}, the optical direction cosines of the ray. From these data we will determine the intersection of the ray with the next surface, and the optical direction cosines of the refracted ray. These values then become the initial data for the new ray, and the process is repeated until the image point is reached.

5.4.3 Transfer procedure, physical surface to next tangent plane.

5.4.3.1 The first part of the problem, namely the determination of the intersection of the ray with the j th spherical surface, will be divided into two parts: first, the intersection of the ray with a non-physical surface, the plane tangent to the spherical surface, and, second, the final intersection with the spherical sur-

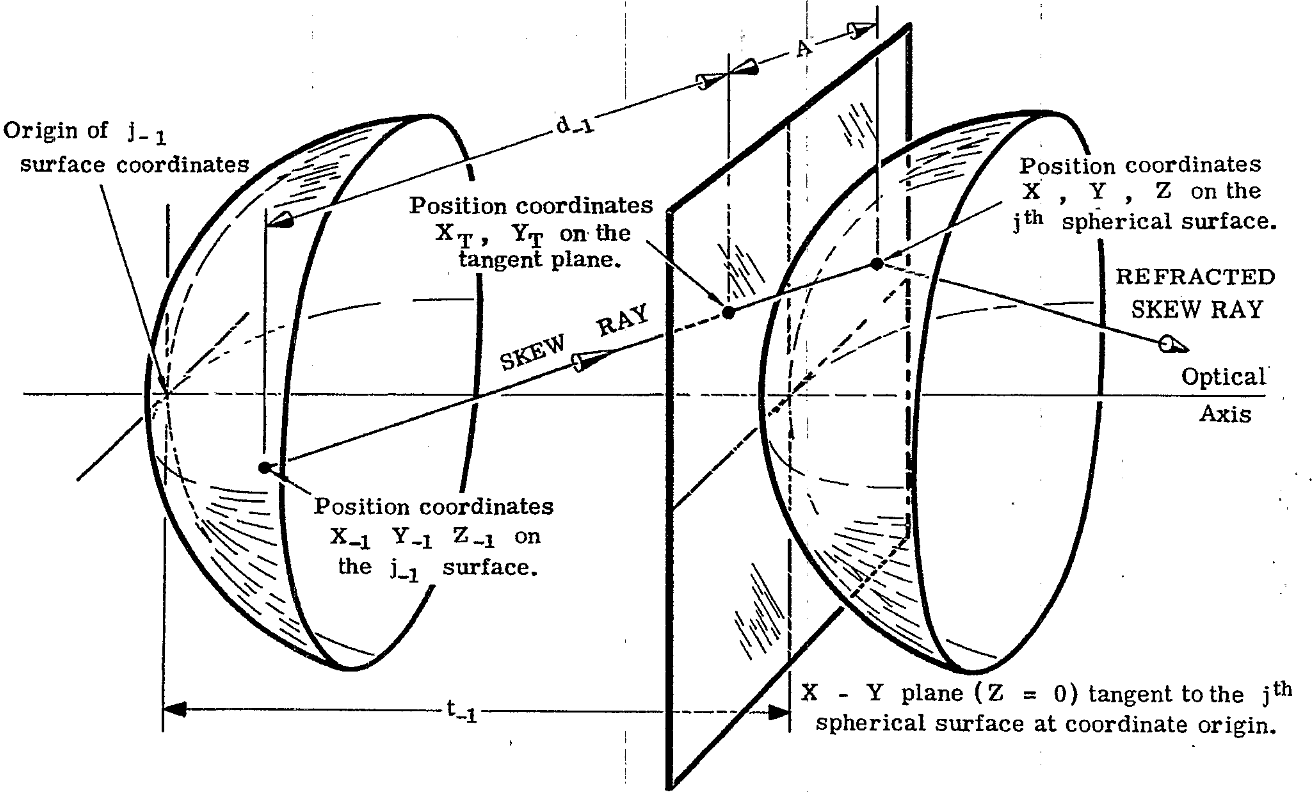

Figure 5.6 - Diagram of a skew ray in space between the j_{-1} surface and the j th surface.

face. In Section 5.4.3 we consider only the first part.

5.4.3.2 The origin of the position coordinates for points on the tangent plane is at the point of tangency, the optical axis. Hence $Z_T = 0$ for all points in the plane. The new value of X, X_T, is the old value, X_{-1}, plus the change in X, ΔX. The latter is the projection of the skew ray, of length d_{-1}, onto the X axis. Hence

$$X_T = X_{-1} + \Delta X = X_{-1} + d_{-1} \frac{K_{-1}}{n_{-1}},$$

since K_{-1}/n_{-1} is the direction cosine of the ray with respect to the X axis. There is a corresponding equation for Y_T.

5.4.3.3 The length of the ray, d_{-1}, between the left-hand surface and the tangent plane is not given; it must be calculated from the initial data. From Figure 5.6 the change in Z is given by

$$\Delta Z = t_{-1} - Z_{-1},$$

and this equals the projection of the ray along the Z axis. Therefore

$$\Delta Z = d_{-1} \frac{M_{-1}}{n_{-1}}.$$

5.4.3.4 It is now possible to summarize the three equations which are used to calculate the intersection of the ray with the tangent plane.

$$\frac{d_{-1}}{n_{-1}} = (t_{-1} - Z_{-1}) \frac{1}{M_{-1}}, \quad (1)$$

$$Y_T = Y_{-1} + \frac{d_{-1}}{n_{-1}} L_{-1}, \quad (2)$$

and

$$X_T = X_{-1} + \frac{d_{-1}}{n_{-1}} K_{-1}. \quad (3)$$

It should be pointed out that in addition to the initial data for the ray, we must be given the value t_{-1}, the distance between the surfaces measured along the optical axis. It is not necessary, however, to know explicitly the value of n_{-1} at this time. The specific procedure followed is first, to use Equation (1) to calculate the numerical value of d_{-1}/n_{-1}; second, to use the value thus obtained in Equations (2) and (3) to calculate Y_T and X_T respectively.

5.4.4 Transfer procedure, tangent plane to spherical surface.

5.4.4.1 The discussion in Section 5.4.3 treated the first part of the transfer problem. The following discussion treats the second part, transferring the ray coordinates on the tangent plane to those on the spherical surface.

5.4.4.2 Referring to Figure 5.6, since the tangent plane is not a refracting plane, the ray continues on to the sphere, for a distance A. The segment A has the same optical direction cosines as the segment d_{-1}. Therefore the new values of the coordinates, X, Y, and Z on the sphere, are determined from the values on the tangent plane, X_T, Y_T, and Z_T, by the process that was used to set up Equations (2) and (3). Remembering that Z_T is zero, we have

$$X = X_T + \frac{A}{n_{-1}} K_{-1}, \tag{4}$$

$$Y = Y_T + \frac{A}{n_{-1}} L_{-1}, \tag{5}$$

and

$$Z = \frac{A}{n_{-1}} M_{-1}. \tag{6}$$

5.4.4.3 In order to use Equations (4), (5), and (6), it is necessary to calculate the value of A. It is clear from Figure 5.6 that this value depends on the curvature of the jth spherical surface, the coordinates of the ray at the tangent plane, and the direction cosines of the ray. We will use a relation between X, Y, Z and c which depends on the properties of a sphere. This equation can be used with Equations (4), (5), and (6) to eliminate X, Y, and Z. The result will be an expression for A/n_{-1} in terms of known data.

5.4.4.4 Figure 5.7 shows a plane containing the optical axis and the intersection point (X, Y, Z) of the

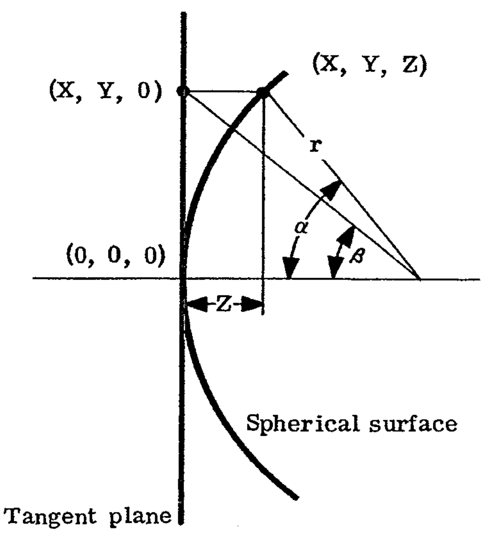

Figure 5.7 - Some properties of a spherical surface.

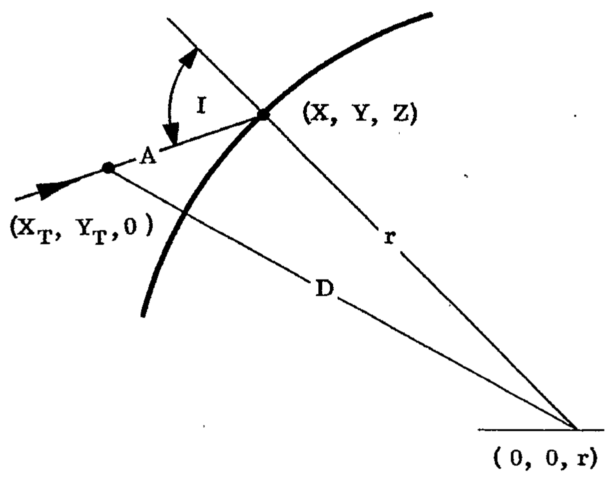

Figure 5.8 - Determination of $n_{-1} \cos I$.

ray on the spherical surface. From the figure, and recalling that $c = 1/r$, we have

$$Z = r - \left[r^2 - (X^2 + Y^2) \right]^{1/2} = \frac{1}{c} - \frac{1}{c} \left[1 - c^2 (X^2 + Y^2) \right]^{1/2},$$

which can be simplified, by transposing and squaring, to

$$c^2 (X^2 + Y^2 + Z^2) - 2cZ = 0 .$$

Substituting into this equation the expressions for X, Y, and Z from Equations (4), (5), and (6). The result, on collecting terms, is

$$\left(\frac{A}{n_{-1}}\right)^2 c (K_{-1}^2 + L_{-1}^2 + M_{-1}^2) - 2\left(\frac{A}{n_{-1}}\right)\left[M_{-1} - c(Y_T L_{-1} + X_T K_{-1}) \right] + c(X_T^2 + Y_T^2) = 0 .$$

5.4.4.5 In this last simplification it was assumed that $c \neq 0$; the case of $c \to 0$ will now be considered. Since the sum of the squares of the direction cosines is unity, the coefficient of $(A/n_{-1})^2$ is $c\, n_{-1}^2$. Calling the other coefficients 2 B and H respectively, we have

$$c\, n_{-1}^2 \left(\frac{A}{n_{-1}}\right)^2 - 2B\left(\frac{A}{n_{-1}}\right) + H = 0,$$

which has the solutions,

$$\frac{A}{n_{-1}} = \frac{B \pm n_{-1}\left[\left(\frac{B}{n_{-1}}\right)^2 - cH \right]^{1/2}}{c\, n_{-1}^2} .$$

As $c \to 0$, that is as the spherical surface approaches a plane surface, $A \to 0$ as can be seen from Figure 5.6. To insure this we can use only the negative sign in the above solutions. A has the same sign as c; this can be seen either by considering the expression for A/n_{-1}, or from Figure 5.6. When A is negative, the tangent plane lies to the right of the surface. The coefficients B and H were introduced for convenience in calculation. Their physical significance is not difficult to understand. From the definition of H, and from Figure 5.7, it is seen that

$$H = c(X_T^2 + Y_T^2) = r\left[\frac{X_T^2 + Y_T^2}{r^2}\right] = r \tan^2 \beta ,$$

where β is the angle between the optical axis and a line drawn from the center of curvature to the intersection of the ray with the tangent plane. From this expression for H, and the result derived in paragraph 5.4.4.4, an expression for B in terms of n_{-1}, and angles I and β can be found.

5.4.4.6 Before simplifying the expression for $\frac{A}{n_{-1}}$ a discussion of the physical meaning of the square root, $\left[(B/n_{-1})^2 - cH\right]^{1/2}$, is in order. This term will be used by itself in the refraction procedure; it is convenient to put it in another form here. Consider Figure 5.8; all the lines are in the plane of incidence. Using the cosine law, it can be stated that

$$D^2 = X_T^2 + Y_T^2 + r^2 = A^2 + r^2 + 2Ar\cos I .$$

Solving for $\cos I$, and substituting H for $c(X_T^2 + Y_T^2)$ produces

$$n_{-1}\cos I = \frac{H - c\, n_{-1}^2 \left(\frac{A}{n_{-1}}\right)^2}{2\,\frac{A}{n_{-1}}} .$$

Finally, substituting the expression for $\frac{A}{n_{-1}}$, with the negative sign, given in paragraph 5.4.4.5, gives

$$n_{-1}\cos I = n_{-1}\left[\left(\frac{B}{n_{-1}}\right)^2 - cH \right]^{1/2} . \qquad (7)$$

5.4.4.7 Returning to the solution for $\frac{A}{n_{-1}}$ in paragraph 5.4.4.5, and using the expression for $n_{-1} \cos I$, we have

$$\frac{A}{n_{-1}} = \frac{B - n_{-1} \cos I}{c\, n_{-1}^2} .$$

But by using Equation (7)

$$c\, n_{-1}^2 = \frac{B^2 - n_{-1}^2 \cos^2 I}{H} = \frac{(B + n_{-1} \cos I)(B - n_{-1} \cos I)}{H},$$

and the final expression for $\frac{A}{n_{-1}}$ becomes,

$$\frac{A}{n_{-1}} = \frac{H}{B + n_{-1} \cos I} . \qquad (8)$$

5.4.4.8 The four equations, then, which are used to calculate $\frac{A}{n_{-1}}$ are, in the order used,

$$H = c(X_T^2 + Y_T^2), \qquad (9)$$

$$B = M_{-1} - c(Y_T L_{-1} + X_T K_{-1}), \qquad (10)$$

$$n_{-1} \cos I = n_{-1} \left[\left(\frac{B}{n_{-1}}\right)^2 - c H \right]^{1/2}, \qquad (7)$$

and

$$\frac{A}{n_{-1}} = \frac{H}{B + n_{-1} \cos I} . \qquad (8)$$

Equations (4), (5), and (6) are then used to calculate X, Y, and Z.

5.4.5 Refraction procedure at the spherical surface.

5.4.5.1 Now that X, Y and Z have been calculated, these values together with initial data K_{-1}, L_{-1}, and M_{-1}, can be used to determine K, L, and M, which specify the direction of the ray after refraction. The basic equations which will be employed are Equations 2-(3) and 2-(4).

5.4.5.2 In Section 2 it was shown that Equation 2-(3) has the following meaning: if vectors are drawn (refer to Figure 2.3) from the intersection point, in the direction of the incident and refracted rays respectively, and these vectors have lengths equal to n_o and n_1, then the closing side of the triangle is parallel to the normal to the surface, and is of length Γ.

5.4.5.3 We now redraw this figure considering the surface as the jth surface. This is shown in Figure 5.9, which is drawn in the plane of incidence. Thus, the radius of curvature of the surface is also in this plane. The line of length Γ is parallel to r. The unit vector $\vec{M}_1$ is the quotient of the vector parallel to the normal divided by r. Hence

$$\vec{M}_1 = c\left[(0 - X)\vec{i} + (0 - Y)\vec{j} + (r - Z)\vec{k} \right]$$

$$= c\left[-X\vec{i} - Y\vec{j} + (r - Z)\vec{k} \right],$$

where $\vec{i}, \vec{j}, \vec{k}$ are unit vectors along the coordinate axes. Using Equation 2-(3),

$$\vec{S}_1 - \vec{S}_o = -cX\Gamma\vec{i} - cY\Gamma\vec{j} + c(r - Z)\Gamma\vec{k} .$$

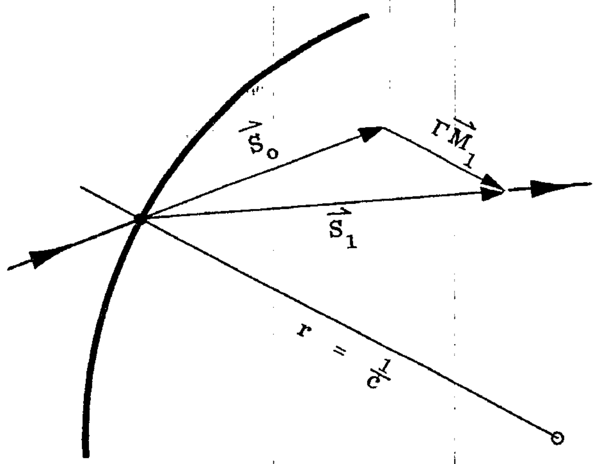

Figure 5.9 - Triangle for the law of refraction.

Now

$$\vec{S}_o = n_{-1}\vec{Q}_o = K_{-1}\vec{i} + L_{-1}\vec{j} + M_{-1}\vec{k},$$

and a similar equation holds for $\vec{S}_1$. Hence

$$\vec{S}_1 - \vec{S}_o = (K - K_{-1})\vec{i} + (L - L_{-1})\vec{j} + (M - M_{-1})\vec{k}.$$

Equating like coefficients of $\vec{i}$, $\vec{j}$, and $\vec{k}$, we have relations between the old and new optical direction cosines.

5.4.5.4 There remains the calculation of Γ. This is done by using Equation 2-(4). We can now write down the five equations which are used in the order given to calculate K, L, and M from the initial data or from previously calculated results.

$$n \cos I' = n\left[\left(\frac{n_{-1}}{n}\cos I\right)^2 - \left(\frac{n_{-1}}{n}\right)^2 + 1\right]^{1/2}, \tag{11}$$

$$\Gamma = n \cos I' - n_{-1} \cos I, \tag{12}$$

$$K = K_{-1} - X c \Gamma, \tag{13}$$

$$L = L_{-1} - Y c \Gamma, \tag{14}$$

and

$$M = M_{-1} - (Zc - 1)\Gamma. \tag{15}$$

5.4.6 Summary of ray trace equations.

5.4.6.1 In the previous sections there were derived the equations used to trace a skew ray from one surface through the following one. For convenience, the equations are now listed in the order of use. The initial ray data are X_{-1}, Y_{-1}, Z_{-1}, K_{-1}, L_{-1} and M_{-1}. The initial system data are t_{-1}, n_{-1} and c. Final values to be determined are X, Y, Z, K, L, and M.

$$\frac{d_{-1}}{n_{-1}} = (t_{-1} - Z_{-1}) \frac{1}{M_{-1}} , \tag{1}$$

$$Y_T = Y_{-1} + \frac{d_{-1}}{n_{-1}} L_{-1} , \tag{2}$$

$$X_T = X_{-1} + \frac{d_{-1}}{n_{-1}} K_{-1} , \tag{3}$$

$$H = c (X_T^2 + Y_T^2) , \tag{9}$$

$$B = M_{-1} - c (Y_T L_{-1} + X_T K_{-1}) , \tag{10}$$

$$n_{-1} \cos I = n_{-1} \left[\left(\frac{B}{n_{-1}}\right)^2 - c H \right]^{1/2} , \tag{7}$$

$$\frac{A}{n_{-1}} = \frac{H}{B + n_{-1} \cos I} , \tag{8}$$

$$X = X_T + \frac{A}{n_{-1}} K_{-1} , \tag{4}$$

$$Y = Y_T + \frac{A}{n_{-1}} L_{-1} , \tag{5}$$

$$Z = \frac{A}{n_{-1}} M_{-1} , \tag{6}$$

$$n \cos I' = n \left[\left(\frac{n_{-1}}{n} \cos I\right)^2 - \left(\frac{n_{-1}}{n}\right)^2 + 1 \right]^{1/2} , \tag{11}$$

$$\Gamma = n \cos I' - n_{-1} \cos I , \tag{12}$$

$$K = K_{-1} - X c \Gamma , \tag{13}$$

$$L = L_{-1} - Y c \Gamma , \tag{14}$$

and

$$M = M_{-1} - (Z c - 1) \Gamma . \tag{15}$$

5.4.6.2 The final calculated values, X , Y , Z , K , L , and M now become the initial ray data for the next calculation. The new system data, t , n , and c_{+1} must be given. These ray and system data are used with the above ray trace equations; in this way a given skew ray from any object surface can be traced through any number of spherical surfaces to the spherical image surface.

5.4.6.3 The equations listed in paragraph 5.4.6.1 are general, in that they also hold for plane surfaces. Referring to Figure 5.6, the physical result is that the j^{th} surface coincides with the tangent plane, hence the coordinates X_T , Y_T , Z_T equal X , Y , Z , and $A = 0$. These results follow mathematically by using $c = 0$ in the equations given in paragraph 5.4.6.1. Refraction at plane surfaces will be discussed in detail in Section 13.

5.4.7 <u>Step by step ray tracing procedure.</u>

5.4.7.1 The following table, Table 5.1 , shows how these calculations can be made in a compact systematic manner. The surfaces are numbered 0 , 1 , 2 , 3 beginning with 0 as the object surface. The initial system data are the values of the c , t , and n quantities indicated above the double line. In a numerical example (see Table 5.2) the values of these quantities are written in the places indicated. The letters in the left hand column have been defined in Section 5.2.2 , or by the equations in Section 5.4.6 .

5.4.7.2 The initial ray data are numerical values of X_o, Y_o , Z_o , K_o, L_o , and M_o , which would be written at the place indicated. Note that quantities pertaining to surfaces are written within the column for the corresponding surface; quantities pertaining to the space between surfaces are written in a break in the corresponding vertical line. The numbers running from 1 to 17 are the steps in the calculation in the order they are made. The steps, except (7) and (14) , correspond to the 15 equations, listed in order of steps, in

Section 5.4.6.1 . "Next step" indicates step No. 1 for the next ray segment. The table entries have been so chosen that a person using a desk calculator does not have to write down any number except those to be entered in the table.

SURFACE	0		1		2		3
c	c_o		c_1		c_2		
t		t_o		t_1		t_2	
n		n_o		n_1		n_2	
X	X_o		(9)				
Y	Y_o		(10)				
Z	Z_o		(11)				
K		K_o		(15)			
L		L_o		(16)			
M		M_o		(17)			
d_{-1}/n_{-1}		(1)		Next Step			
X_T			(2)				
Y_T			(3)				
H			(4)				
B			(5)				
n_{-1} cos I			(6)				
B + n_{-1}cos I			(7)				
A/n_{-1}			(8)				
n cos I'			(12)				
Γ			(13)				
c Γ			(14)				

Table 5.1-Skew ray trace computing sheet.

SURFACE	0		1		2		3
c	0		0.25284872		-0.01473947		
t		-2.2		0.6			
n		1.0		1.62		1.0	
X	1.48		1.48		1.43679417		
Y	0		-0.33445977		-0.29386784		
Z	0		0.30264162		-0.01585220		
K		0		-0.24330257		-0.25700617	
L		0.17360000		0.22858306		0.23138586	
M		0.98481625		1.58522985		0.93830084	
d_{-1}/n_{-1}		-2.23391927		0.18758061			
X_T			1.48		1.43436116		
Y_T			-0.38780839		-0.29158202		
H			0.59186710		-0.03157802		
B			1.00183892		1.57910362		
n_{-1} cos I			0.92413654		1.57871680		
B + n_{-1}cos I			1.92597546		3.15782041		
A/n_{-1}			0.30730770		-0.00999994		
n cos I'			1.57430250		0.93163659		
Γ			0.65016596		-0.64708020		
c Γ			0.16439363		0.00953762		

Table 5.2-Skew ray trace for three surfaces.

5.4.8 Numerical example.

5.4.8.1 Throughout the discussion of geometrical optics, lengthy explanations have been avoided by the inclusion of numerical examples showing the actual calculations. Table 5.2 is such an illustration. The calculations shown in this table can be made by an experienced person with a modern desk calculator without undue labor. In order for the calculations to be useful, at least six significant figures must be carried throughout. Since the introduction of the modern electronic computing machines, there is really very little justification for a human computer to carry out these calculations unless ray tracing is only done occasionally. The above equations can be programmed in a modern machine to make these calculations in less than one second per surface, with at least eight significant figures. The calculations shown in this and other numerical examples may not offer complete consistency in the number of significant figures for two reasons: (1) some were prepared from automatic computer results where intermediate values were not available and had to be developed by hand computing; (2) others were prepared from a designer's work sheets where the aim was not eight-figure accuracy but only three or four-figure accuracy in which case the designer had merely entered results as they appeared on the hand calculator. No units appear in this and other numerical examples, because the equations are valid for any set of consistent units. As long as all lengths are in the same units, the numerical example will be correct for any units.

5.4.8.2 Some specific remarks should be made about Table 5.2 . The initial data which are given to one, two, or three significant figures are assumed exact. From the initial system data it is apparent that we are considering a double convex lens, index 1.62, surrounded by air. The incident light first intersects the convex face of the lens. The lens thickness is about one quarter of the distance between lens and object surface, but no information is given (or needed for a ray trace) concerning the absolute magnitude of any distance. The object surface is to the right of the lens; therefore the object is virtual.

5.4.8.3 From the initial ray data we see that the (virtual) object point, that is the point towards which the ray is heading, is on the X axis, but not in the Y - Z plane. The initial ray is parallel to the Y - Z plane, hence $X_1 = X_0$. The ray is inclined upwards at an angle with the Z axis of 10°. The calculations indicate that the ray intersects both surfaces of the lens below the X - Z plane (because Y is negative), and intersects both surfaces at points "away from the reader" with respect to the Y - Z plane (because X is positive). The Z value at the first surface is positive because the curvature is positive ; likewise the Z value at the second surface is negative.

5.5 SKEW RAY TRACE EQUATIONS FOR ASPHERIC SURFACES

5.5.1 General.

5.5.1.1 The discussion in Section 5.4 developed equations for, and demonstrated their use in, ray tracing procedures through spherical surfaces. Although spherical surfaces are still much easier to make, and hence are preferred by the lens maker, aspheric surfaces are readily handled by the lens designer who has access to an electronic computer. Aspheric surfaces afford the designer a great deal more latitude in the design, and in addition often permit better correction of aberrations. Aspheric surfaces are being used more and more, and their widespread use depends on inexpensive methods of production.

5.5.1.2 In the skew ray trace for spherical surfaces, it was convenient to effect the transfer from one physical surface to the next by introducing a non-physical tangent plane, and effecting the transfer in two steps. In the case of aspheric surfaces we introduce two non-physical surfaces, a plane and a sphere, both tangent to the physical aspheric surface at the optical axis. See Figure 5.10. The transfer between physical surfaces is now effected in three steps:

(1) first surface to next tangent plane;

(2) tangent plane to tangent sphere;

(3) tangent sphere to physical (second) surface.

Steps (1) and (2) are carried out using the procedure already developed in Section 5.4 .

5.5.2 Mathematical description of an aspheric surface.

5.5.2.1 We need to describe the aspheric surface in a way that indicates clearly its departure from the tangent sphere. This kind of description will not only be easily handled by the ray trace equations, but will also quickly and quantitatively show how close in form the aspheric is to the sphere.

5.5.2.2 In Paragraph 5.4.4.4 there is given an equation for Z ; this quantity is called the sag of the sphere, an abbreviation of sagitta. Using $S^2 = X^2 + Y^2$, this equation is

$$Z = \frac{1}{c} \left[1 - (1 - c^2 S^2)^{1/2} \right] .$$

$c = \frac{1}{R}$

By multiplying and dividing by $\left[1 + (1 - c^2 S^2)^{1/2} \right]$, we have

$$Z = \frac{c S^2}{1 + \sqrt{1 - c^2 S^2}} .$$

Because the shape of an aspheric surface (which is assumed to have rotational symmetry about the Z axis) differs from that of the tangent sphere, the sag (Z) of the aspheric at any distance S from the axis may differ from the sag of the tangent sphere. This is indicated by expressing the difference in these two sags by a power series in S^2 . (The series is in powers of S^2 , and hence only even powers of S appear, because the aspheric has rotational symmetry about the Z axis.) The final expression for the sag is

$$Z = \frac{c\,S^2}{1 + \sqrt{1 - c^2 S^2}} + eS^4 + fS^6 + gS^8 + hS^{10} + O(S^{12})$$

5.5.2.3 Each of the numerical coefficients e , f , g , and h may be positive or negative. The term $O(S^{12})$ stands for the rest of the series, that is terms of order 12 and higher. In a numerical calculation, if the sag is given by this expression, $O(S^{12})$ would be assumed zero, and the calculations would involve only e , f , g , and h . The terms eS^4 , fS^6 , etc., are called deformation terms.

5.5.3 Initial data, and transfer from physical surface to next tangent sphere. Part of the transfer from one physical surface to the next has already been solved in Section 5.4 . The initial ray data for the skew ray between aspheric surfaces is the same as given in Section 5.4.2 , namely X_{-1} , Y_{-1} , Z_{-1} , K_{-1} , L_{-1} and M_{-1} . We determine the intersection of this ray with the non-physical sphere, tangent to the j th aspheric surface, by the procedure given in Sections 5.4.3 and 5.4.4 . In other words we apply Equations (1), (2), (3), (9), (10), (7), (8), (4), (5) and (6) in that order. The only difference so far between this ray trace and the former is that in the previous case the sphere was a physical surface, while in the present case it is a purely fictitious surface. The equations do not know the difference between physical and non-physical surfaces; hence the same equations are used for both cases.

5.5.4 Transfer procedure, tangent sphere to aspheric surface.

5.5.4.1 In Paragraphs 5.4.4.4 and 5.4.4.5 an expression for $\frac{A}{n_{-1}}$ was derived using four equations, namely the equation for the sag, Z , of the sphere and Equations (4), (5), and (6) . This value of $\frac{A}{n_{-1}}$ was then used in Equations (4), (5), and (6) to transfer from tangent plane to sphere. It would be perfectly possible to proceed similarly here. We would set up three equations, corresponding to (4), (5), and (6) , but replacing A by A + A' . (See Figure 5.10) . Using these three equations, and the equation for the

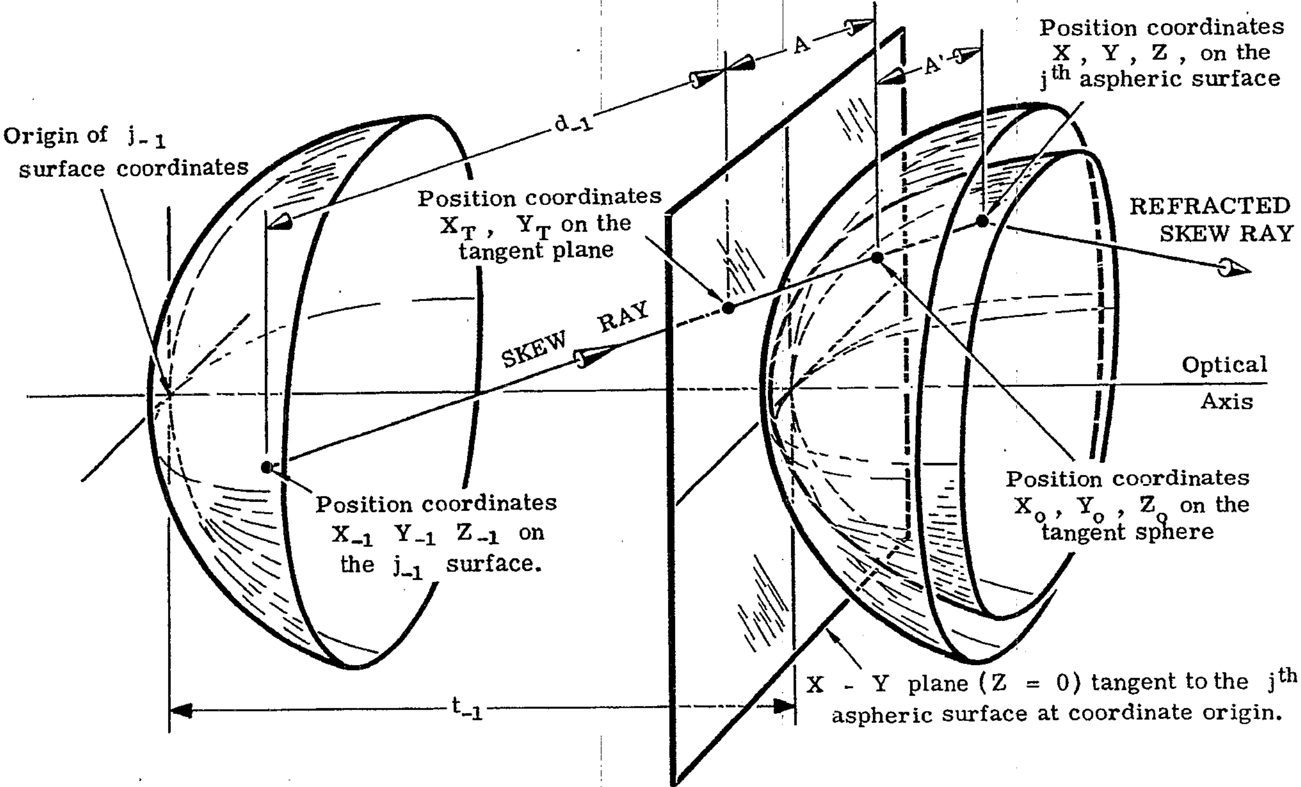

Figure 5.10 - Diagram of a skew ray in space between the j_{-1} surface and the j^{th} aspheric surface.

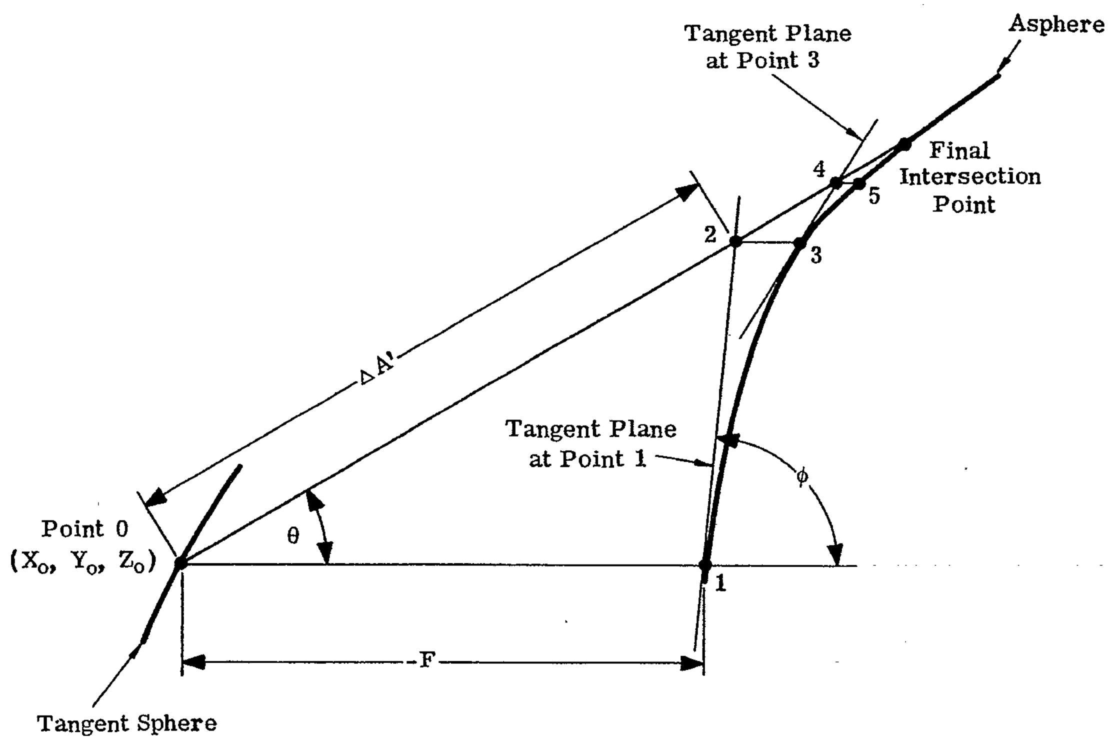

Figure 5.11- Step-wise approximations from tangent sphere intersection, point 0, to final intersection point.

sag of an aspheric surface, Paragraph 5.5.2.2, an expression for $\frac{A + A'}{n_{-1}}$ could be found, and used to transfer directly from tangent sphere to aspheric. The resulting calculations are extremely involved and it is preferable to proceed otherwise.

5.5.4.2 The procedure to be employed makes use of the fact that transfer to the tangent sphere is fairly simple. The remaining transfer from tangent sphere to aspheric is effected in a step-wise procedure approaching the final intersection by successive approximations. The physical procedure is indicated in Figure 5.11 ; this figure represents the plane determined by the skew ray, and a line through this ray parallel to the Z axis.

5.5.4.3 Beginning at point 0 , the intersection of the ray with the tangent sphere, the first approximation to the final point is point 1 . Point 1 has the same X and Y values as point 0 ; its Z value differs from that of point 0 by the deformation terms evaluated at these particular values of X and Y . The second approximation is point 2 , the intersection of the ray with a line tangent to the aspheric at 1 . The tangent line is determined from the known coordinates of point 1 and the calculated curvature of the aspheric at point 1 . Since the ray direction is known, its intersection with the tangent line, point 2 , is determined. The procedure is now repeated. Point 3 has the same X and Y as point 2 , and its Z value can be found from the deformation terms and the Z value at point 2 . The fourth and fifth approximations are points 4 and 5 , respectively. (The point 0 , on the sphere, is correctly called the zeroth approximation to the final point.)

5.5.4.4 The various values of X , Y , and Z for points 0, 1, 2,··· will be referred to as X_n , Y_n , and Z_n where the n will stand for the order of approximation. Let us begin at any even-numbered point, that is a point on the ray; in practice, the calculations begin with point 0 , but we wish to make the equations general so that n will stand for any even-numbered point. The next point, on the aspheric, will have coordinates X_n , Y_n , Z_m . Note that the X and Y values are the same as for the previous point. The S_n value now used to calculate the sag, Z_m , is

$$S_n^2 = X_n^2 + Y_n^2 . \qquad (16)$$

The change in Z , that is $Z_m - Z_n$, is the distance parallel to the Z axis between an even-numbered point and an odd-numbered point. Calling this distance $-F$ (see Figure 5.11), we have

$$F = Z_n - \left[\frac{c\,S^2}{1 + \sqrt{1 - c^2 S^2}} + eS^4 + fS^6 + gS^8 + hS^{10} \right]. \tag{17}$$

(The subscript n has been omitted. Henceforth all values of S are rigorously S_n.)

5.5.4.5 For notational purposes it is convenient to designate the square root in Equation (17) by W. Hence

$$W = \left[1 - c^2 S^2 \right]^{1/2}. \tag{18}$$

Referring to Figure 5.7, it can be seen that $W = \sqrt{1 - \sin^2 \alpha} = \cos \alpha$, where α is the angle between the normal to the surface and the optical axis. W therefore is the direction cosine, with respect to the Z axis, of a radius drawn from the sphere to the center.

5.5.4.6 From an odd-numbered point, whose coordinates we now know, we move along the tangent line to the ray. The coordinates of this new even-numbered point are X_{n+1}, Y_{n+1}, and Z_{n+1}. Calling the distance along the ray, between two even-numbered points, $\Delta A'$, we can write equations for the new coordinates similar to equations (4), (5), and (6). We have then

$$X_{n+1} = X_n + \frac{\Delta A'}{n_{-1}} K_{-1}, \tag{19}$$

$$Y_{n+1} = Y_n + \frac{\Delta A'}{n_{-1}} L_{-1}, \tag{20}$$

and

$$Z_{n+1} = Z_n + \frac{\Delta A'}{n_{-1}} M_{-1}. \tag{21}$$

We will consider the calculation of $\Delta A'$ presently. Once this is known, the new coordinates on the ray are known, and we repeat the calculations through two more steps until we get once again to the ray. This <u>iteration procedure</u> is continued until $\Delta A'/n_{-1}$ is less than any desired tolerance. In this manner we can approach the final point on the aspheric as closely as we choose.

5.5.4.7 The remaining problem in the transfer from tangent sphere to aspheric surface is the determination of $\Delta A'$. First we need the equation of the plane tangent to the aspheric surface at an odd-numbered point. From the equation for the sag of the surface, Paragraph 5.5.2.2, we can write

$$\psi(X, Y, Z) = Z - \left[\frac{c\,S^2}{1 + \sqrt{1 - c^2 S^2}} + eS^4 + fS^6 + gS^8 + hS^{10} \right] = 0,$$

where $\psi(X, Y, Z) = 0$ is the equation of the aspheric surface. Now a plane, tangent to the surface at the point X_n, Y_n, Z_m will coincide with the first approximation to the surface. Physically, if we restrict ourselves to points close to (X_n, Y_n, Z_m) the surface is a plane. To find the first approximation to the surface we expand $\psi(X, Y, Z)$ and keep only the zeroth and first order terms.

5.5.4.8 The equation of the tangent plane is then

$$\psi(X_n, Y_n, Z_m) + (X - X_n)\left[\frac{\partial \psi}{\partial X}\right]_{X_n, Y_n, Z_m} + (Y - Y_n)\left[\frac{\partial \psi}{\partial Y}\right]_{X_n, Y_n, Z_m} + (Z - Z_m)\left[\frac{\partial \psi}{\partial Z}\right]_{X_n, Y_n, Z_m} = 0,$$

where the first term is the zeroth order term, and the last three are the first order terms, in the expansion of $\psi(X, Y, Z)$. Using Equation (16) we have

$$\frac{\partial \psi}{\partial X} = \frac{\partial \psi}{\partial S}\frac{\partial S}{\partial X} = \frac{\partial \psi}{\partial S}\frac{X}{S},$$

$$\frac{\partial \psi}{\partial Y} = \frac{\partial \psi}{\partial S}\frac{Y}{S},$$

and

$$\frac{\partial \psi}{\partial Z} = 1.$$

Using the expression for $\psi(X, Y, Z)$ given in Paragraph 5.5.4.7, and Equation (18), we get

$$\frac{\partial\psi}{\partial S} = -\frac{S}{W}c - S\left[4eS^2 + 6fS^4 + 8gS^6 + 10hS^8\right],$$

which can be simplified to $\frac{\partial\psi}{\partial S} = -\frac{S}{W}E$ by defining

$$E = c + W\left[4eS^2 + 6fS^4 + 8gS^6 + 10hS^8\right]. \qquad (22)$$

(If the deformation coefficients are small, E is approximately the curvature of the aspheric surface at the distance S_n from the optical axis.)

5.5.4.9 The equation for the sag of the aspheric surface given in Paragraph 5.5.2.2 is an equation for Z_m if $S^2 = X_n^2 + Y_n^2$. Because of this, $\psi(X_n, Y_n, Z_m)$ is zero, using the equation for ψ in Paragraph 5.5.4.7. The zeroth order term in the expansion is therefore zero. The equation of the plane becomes, using the above expressions for the partial derivatives,

$$-(X - X_n)\frac{X_n}{W}E - (Y - Y_n)\frac{Y_n}{W}E + (Z - Z_n)\frac{W}{W} + Z_n - Z_m = 0,$$

where we have separated the term $Z - Z_m$ into two terms. By Equation (17), $F = Z_n - Z_m$. We define here two quantities,

$$U = -XE, \qquad (23)$$

$$V = -YE. \qquad (24)$$

5.5.4.10 With these substitutions the equation of the plane becomes

$$(X - X_n)U + (Y - Y_n)V + (Z - Z_n)W = -FW.$$

This equation holds for all values of X, Y, and Z, in particular X_{n+1}, Y_{n+1}, and Z_{n+1}. Instead of the difference $(X_{n+1} - X_n)$, we use $(\Delta A'/n_{-1})K_{-1}$ from Equation (19). Similarly, using Equations (20) and (21), and solving for $\Delta A'/n_{-1}$,

$$\frac{\Delta A'}{n_{-1}} = \frac{-FW}{K_{-1}U + L_{-1}V + M_{-1}W}. \qquad (25)$$

From Figure 5.10, it can be seen that the distance, D_{-1}, along the ray is

$$D_{-1} = d_{-1} + A + A'. \qquad (25a)$$

5.5.5 Refraction procedure at the aspheric surface.

5.5.5.1 Now that the intersection point, (X, Y, Z), of the ray and the aspheric surface has been found, the refraction equation is used to determine the new direction of the ray. The procedure is basically the same as that used for refraction at spherical surfaces, discussed in Section 5.4.5. In that section Equation (11) was used to calculate $n \cos I'$, because $n_{-1} \cos I$ had already been calculated using Equation (7).

5.5.5.2 In the present case there is not yet a value for $n_{-1} \cos I$. To calculate this we use the fact that the cosine of an angle between two directed lines is equal to the sum of the products of their corresponding direction cosines. Since we are calculating $\cos I$, the two lines in question are the ray whose optical direction cosines are K_{-1}, L_{-1}, and M_{-1}, and the normal to the aspheric surface. Now the normal to the surface is just the normal to the tangent plane. The equation of this tangent plane is given in Paragraph 5.5.4.10, where X_n, Y_n, and Z_n are the coordinates of the final point on the ray, the intersection with the surface.

5.5.5.3 Given the equation of a plane, the direction cosines of the normal are proportional to the corresponding coefficients of X, Y, and Z. Hence the direction cosines of the normal are, in the usual order, U/G, V/G, and W/G, where G is a proportionality constant. Because the sum of the squares of the direction cosines is unity, we have

$$G^2 = U^2 + V^2 + W^2. \qquad (26)$$

Using the direction cosines of the ray we get

$$\cos I = \frac{K_{-1}}{n_{-1}} \frac{U}{G} + \frac{L_{-1}}{n_{-1}} \frac{V}{G} + \frac{M_{-1}}{n_{-1}} \frac{W}{G} ,$$

which is rewritten in final form as

$$G\, n_{-1} \cos I = K_{-1}\, U + L_{-1}\, V + M_{-1}\, W . \quad (27)$$

5.5.5.4 Equation (11) is now used to determine $n \cos I'$. However, for calculation purposes, it is preferable to leave the G on both sides of the equation.

$$G\, n \cos I' = n \left[\left(G \frac{n_{-1}}{n} \cos I \right)^2 - G^2 \left(\frac{n_{-1}}{n} \right)^2 + G^2 \right]^{1/2} . \quad (28)$$

5.5.5.5 Returning to the equation in Paragraph 5.4.5.3, we write this vector equation as three scalar equations using the method of Paragraph 5.4.5.4. We get

$$K - K_{-1} = \Gamma \frac{U}{G} ,$$

$$L - L_{-1} = \Gamma \frac{V}{G} ,$$

and

$$M - M_{-1} = \Gamma \frac{W}{G} ,$$

because Γ is parallel to the normal to the surface and therefore has the same direction cosines. Introducing $P = \Gamma/G$, we have, using Equation (12),

$$P = (G\, n \cos I' - G\, n_{-1} \cos I)/G^2 . \quad (29)$$

Finally, K, L, and M are found from the equations,

$$K = K_{-1} + U P , \quad (30)$$

$$L = L_{-1} + V P , \quad (31)$$

and

$$M = M_{-1} + W P . \quad (32)$$

5.5.6 <u>Summary of ray trace equations.</u>

5.5.6.1 In the previous sections we have derived the equations used to trace a skew ray from a tangent sphere through the aspheric surface. For convenience we rewrite the equations in the order of use. The initial ray data are X_{-1}, Y_{-1}, Z_{-1}, K_{-1}, L_{-1}, and M_{-1}. The initial system data are t_{-1}, n_{-1}, c, and the deformation coefficients e, f, g, ··· . Final values to be determined are X, Y, Z, K, L, and M.

5.5.6.2 The position coordinates for the ray on the tangent sphere are calculated using the first ten equations listed in Section 5.4.6. Equations (16) through (32) are then used in the order listed below.

$$S_n^2 = X_n^2 + Y_n^2 , \quad (16)$$

$$W = \left[1 - c^2 S^2 \right]^{1/2} , \quad (18)$$

$$F = Z_n - \left[\frac{c\, S^2}{1 + \sqrt{1 - c^2 S^2}} + eS^4 + fS^6 + gS^8 + hS^{10} \right] , \quad (17)$$

$$E = c + W \left[4eS^2 + 6fS^4 + 8gS^6 + 10hS^8 \right] , \quad (22)$$

$$U = -XE , \quad (23)$$

$$V = -YE , \quad (24)$$

$$\frac{\Delta A'}{n_{-1}} = \frac{-FW}{K_{-1} U + L_{-1} V + M_{-1} W} , \qquad (25)$$

$$X_{n+1} = X_n + \frac{\Delta A'}{n_{-1}} K_{-1} , \qquad (19)$$

$$Y_{n+1} = Y_n + \frac{\Delta A'}{n_{-1}} L_{-1} , \qquad (20)$$

$$Z_{n+1} = Z_n + \frac{\Delta A'}{n_{-1}} M_{-1} , \qquad (21)$$

$$G^2 = U^2 + V^2 + W^2 , \qquad (26)$$

$$G n_{-1} \cos I = K_{-1} U + L_{-1} V + M_{-1} W , \qquad (27)$$

$$G n \cos I' = n \left[\left(G \frac{n_{-1}}{n} \cos I \right)^2 - G^2 \left(\frac{n_{-1}}{n} \right)^2 + G^2 \right]^{1/2} , \qquad (28)$$

$$P = (G n \cos I' - G n_{-1} \cos I)/G^2 , \qquad (29)$$

$$K = K_{-1} + UP , \qquad (30)$$

$$L = L_{-1} + VP , \qquad (31)$$

and

$$M = M_{-1} + WP . \qquad (32)$$

5.5.6.3 The first ten of these equations are used in an iterative process until $\Delta A'/n_{-1}$ becomes as small as desired. The final values of U, V, and W are then used in the last seven equations (26) through (32). The final calculated values of X, Y, Z, K, L, and M become the initial ray data for the next calculation. These values, together with new system data, t, n, c_{+1}, and deformation terms, are used in a reapplication of the ray trace equations.

SURFACE	0	1	2	3
c	0	0.25284872	-0.01473947	
e		-0.005		
f		0.00001		
g		-0.0000005		
h		0		
t	-2.2	0.6		
n	1.0	1.62	1.0	
X	1.48	1.48	1.44043943	
Y	0	-0.33905030	-0.29645624	
Z	0	0.27660001	-0.01594078	
K	0	-0.20481560		
L	0.17360000	0.22052072		
M	0.98481625	1.59179807		
d_{-1}/n_{-1}	-2.23391927			
X_T		1.48		
Y_T		-0.38780839		
H		0.59186710		
B		1.00183892		
$n_{-1} \cos I$		0.92413654		
$B + n_{-1} \cos I$		1.92597546		
A/n_{-1}		0.30730770		
$n \cos I'$				
Γ		Enter $X_n Y_n Z_n$		
c Γ		in Table 5.4		

Table 5.3 - Skew ray trace through an aspheric surface. Part of the calculations are shown in Table 5.4.

5.5.6.4 Because a spherical surface is a special case of an aspheric surface for which the deformation terms are zero, the ray trace equations for aspheric surfaces should easily reduce to those for spherical surfaces. We see, for the case of a sphere ($e = f = g = h = \ldots = 0$),

$$E = c ,$$

$$U = - X c ,$$

$$V = - Y c ,$$

$$W = - (Z c - 1) , \quad \text{(holds for aspheric also)}$$

$$G = 1 ,$$

$$n_{-1} \cos I = - c \left[X K_{-1} + Y L_{-1} + Z M_{-1} \right] ,$$

$$P = \Gamma ,$$

and equations (30) , (31) , and (32) become identical with equations (13) , (14) , and (15) .

ITERATION	1	2	3
X_n	1.48000000	1.48000000	1.48000000
Y_n	-0.33445977	-0.33905071	-0.33905030
Z_n	0.30264163	0.27659764	0.27659999
S_n^2	2.30226334	2.30535538	2.30535510
$1 - c^2 S^2$	0.85281060	0.85261290	0.85261290
W	0.92347745	0.92337040	0.92337040
$c / (1 + W)$	0.13145395	0.13146127	0.13146127
hS^2	0.00000000	0.00000000	0.00000000
$hS^4 + gS^2$	-0.00000115	-0.00000115	-0.00000115
$hS^6 + gS^4 + fS^2$	0.00002037	0.00002040	0.00002040
$hS^8 + gS^6 + fS^4 + eS^2$	-0.01146441	-0.01147976	-0.01147976
$hS^{10} + gS^8 + fS^6 + eS^4 + \frac{cS^2}{1 + W}$	0.27624751	0.27660004	0.27660000
$-F$	-0.02639412	-0.00000238	-0.00000002
$-10\ hS^2$	0.00000000	0.00000000	0.00000000
$-10\ hS^4 - 8\ gS^2$	0.00000921	0.00000922	0.00000922
$-10\ hS^6 - 8\ gS^4 - 6\ fS^2$	-0.00011693	-0.00011706	-0.00011706
$10\ hS^8 + 8\ gS^6 + 6\ fS^4 + 4\ eS^2$	-0.04577605	-0.04583724	-0.04583723
$-E$	-0.21057557	-0.21052397	-0.21052398
U	-0.31165184	-0.31157548	-0.31157549
V	0.07042906	0.07137830	0.07137822
$K_{-1}U + L_{-1}V + M_{-1}W$	0.92168208	0.92174144	0.92174143
$-FW$	-0.02437438	0.000002198	0.000000018
$\Delta A'/n_{-1}$	-0.02644553	0.000002384	0.000000020
X	1.48000000	1.48000000	1.48000000
Y	-0.33905071	-0.33905030	-0.33905030
Z	0.27659764	0.27659999	0.27660001
G^2			0.95478704
$Gn \cos I'$			1.54937517
P			0.65735466
K			-0.20481560
L			0.22052072
M			1.59179807

Table 5.4 - Skew ray trace iteration and refraction calculations. The table shows three iterations.

5.5.7 Numerical example.

5.5.7.1 A numerical example is shown in Tables 5.3 and 5.4. The system data is the same as the example shown in Table 5.2, except for the addition of three deformation coefficients e , f , and g . The coefficient h is specifically listed in both tables as zero. This avoids possible error in not being certain whether or not a coefficient was erroneously omitted. The initial ray data is identical with the previous example; hence the calculations and results for transfer to the tangent sphere are the same. Thus steps 1 through 11 (see Table 5.1) are identical, except for the location of the results of step 11. These are placed in Table 5.4 and are the initial data for the iteration process.

5.5.7.2 Table 5.4 shows the iteration process by which $(\Delta A'/n_{-1}) < 0.00001$; this represents the criterion, set up prior to the calculations, to determine when the iteration process is to be stopped. It is noticed that the first value of $\Delta A'/n_{-1}$ is negative, the second positive, the third almost zero. This oscillation about the target value (< 0.00001) is typical of the method of successive approximations. This method will be used in later sections where aberrations are discussed.

5.5.7.3 The final values of X , Y , and Z , shown just above the double line in Table 5.4 in the column 3 , are entered in Table 5.3 in the place for steps 9 , 10 , and 11 . (The entire iteration process gives the results for steps 9 , 10 , and 11 for an aspheric surface.) These values are now part of the initial ray data for the next surface. The refraction calculations at the aspheric surface are given in Table 5.4 below the double line, and use the final results found above. The values of K , L , and M are now entered in Table 5.3 as the results of steps 15 , 16 , and 17 . They will be used as initial data for the next surface.

5.6 MERIDIONAL RAYS

5.6.1 Definition. A meridional ray is any ray lying in a plane containing the optical axis. A meridional ray will remain in the same plane throughout an entire centered system. For this reason, the tracing of meridional rays is a two dimensional problem, while the tracing of skew rays, which do not lie in a plane containing the optical axis, is a three dimensional problem.

5.6.2 Use of skew ray trace equations. The skew ray formulae given in Sections 5.4 and 5.5 are designed for use on modern automatic computing machines. However, they are in a form which can be used with relative ease - for skew rays - on a desk calculator. Extensive skew ray tracing, which is essential in order to make a complete analysis of a lens system, should be done on a computing machine. In the preliminary design of a lens system it is usually convenient to trace a few selected meridional rays. These are often traced by hand. If the object point has coordinates $(X_0 = 0 , Y_0 , Z_0)$ and the ray pierces the first surface at coordinates $(X_1 = 0 , Y_1 , Z_1)$ the ray is meridional and will remain in the YZ-plane all the way to the image surface. Meridional rays can be traced using the skew ray formulae given in Sections 5.4 and 5.5 by setting $X = 0$ and $K = 0$.

5.6.3 Meridional ray trace, spherical surfaces.

5.6.3.1 Meridional ray tracing can be done for spherical surfaces by using Equations (1) through (10) , followed by either Equations (11) through (15) or Equations (16) through (32). For meridional rays, Equations (1) through (10) reduce to the following eight equations, in the order used:

$$\frac{d_{-1}}{n_{-1}} = (t_{-1} - Z_{-1}) \frac{1}{M_{-1}} , \qquad (1)$$

$$Y_T = Y_{-1} + \frac{d_{-1}}{n_{-1}} L_{-1} , \qquad (2)$$

$$H = c Y_T^2 , \qquad (9a)$$

$$B = M_{-1} - c Y_T L_{-1} , \qquad (10a)$$

$$n_{-1} \cos I = n_{-1} \left[\left(\frac{B}{n_{-1}} \right)^2 - cH \right]^{1/2} , \qquad (7)$$

$$\frac{A}{n_{-1}} = \frac{H}{B + n_{-1} \cos I} , \qquad (8)$$

$$Y = Y_T + \frac{A}{n_{-1}} L_{-1} , \qquad (5)$$

and

$$Z = \frac{A}{n_{-1}} M_{-1} . \qquad (6)$$

Only eight equations are needed, the other two being $X_T = X = 0$. These eight equations trace a meridional ray from any surface to the next spherical surface.

5.6.3.2 Refraction at the spherical surface may be calculated by applying Equations (11), (12), (14), and (15) as written. Equation (13) becomes $K = 0$. (This procedure is referred to as the short form). On the other hand Equations (16) through (32) may be used. These are reduced to the following seven equations, in the order used:

$$W = \left[1 - c^2 Y^2 \right]^{1/2}, \tag{18a}$$

$$V = -Yc, \tag{24a}$$

$$n_{-1} \cos I = L_{-1} V + M_{-1} W, \tag{27a}$$

$$n \cos I' = n \left[\left(\frac{n_{-1}}{n} \cos I \right)^2 - \left(\frac{n_{-1}}{n} \right)^2 + 1 \right]^{1/2}, \tag{11}$$

$$\Gamma = n \cos I' - n_{-1} \cos I, \tag{12}$$

$$L = L_{-1} - Yc\Gamma, \tag{14}$$

and

$$M = M_{-1} - W\Gamma. \tag{32a}$$

Only seven equations are needed, the other ten being $S_n = Y_n$, $E = c$, $G = 1$, $Y_{n+1} = Y_n$, $Z_{n+1} = Z_n$, and $F = U = \Delta A' = X_{n+1} = K = 0$.

5.6.4 Meridional ray trace, aspheric surfaces. For meridional rays and aspheric surfaces, after applying the eight equations given in Paragraph 5.6.3.1, the Equations (16) through (32) are used. These reduce to the following thirteen equations, in the order used:

$$W = \left[1 - c^2 Y^2 \right]^{1/2}, \tag{18a}$$

$$F = Z_n - \left[\frac{cY^2}{1+W} + eY^4 + fY^6 + gY^8 + hY^{10} \right], \tag{17a}$$

$$E = c + W \left[4eY^2 + 6fY^4 + 8gY^6 + 10hY^8 \right], \tag{22a}$$

$$V = -YE, \tag{24}$$

$$\frac{\Delta A'}{n_{-1}} = \frac{-FW}{L_{-1} V + M_{-1} W}, \tag{25b}$$

$$Y_{n+1} = Y_n + \frac{\Delta A'}{n_{-1}} L_{-1}, \tag{20}$$

$$Z_{n+1} = Z_n + \frac{\Delta A'}{n_{-1}} M_{-1}, \tag{21}$$

$$G^2 = V^2 + W^2, \tag{26a}$$

$$G n_{-1} \cos I = L_{-1} V + M_{-1} W, \tag{27a}$$

$$G n \cos I' = n \left[\left(G \frac{n_{-1}}{n} \cos I \right)^2 - G^2 \left(\frac{n_{-1}}{n} \right)^2 + G^2 \right]^{1/2}, \tag{28}$$

$$P = (G n \cos I' - G n_{-1} \cos I)/G^2, \tag{29}$$

$$L = L_{-1} + VP, \tag{31}$$

and

$$M = M_{-1} + WP. \tag{32}$$

Only 13 equations are needed, the other four being $S_n = Y_n$, and $U = X_{n+1} = K = 0$.

5.6.5 Simplified meridional ray trace, spherical surfaces.

5.6.5.1 There are many other methods, involving different parameters, which are commonly used to trace meridional rays. One such method specifies the angle the ray makes with the optical axis, and the perpendicular distance from the center of curvature of the surface to the ray. Figure 5.12 indicates the two

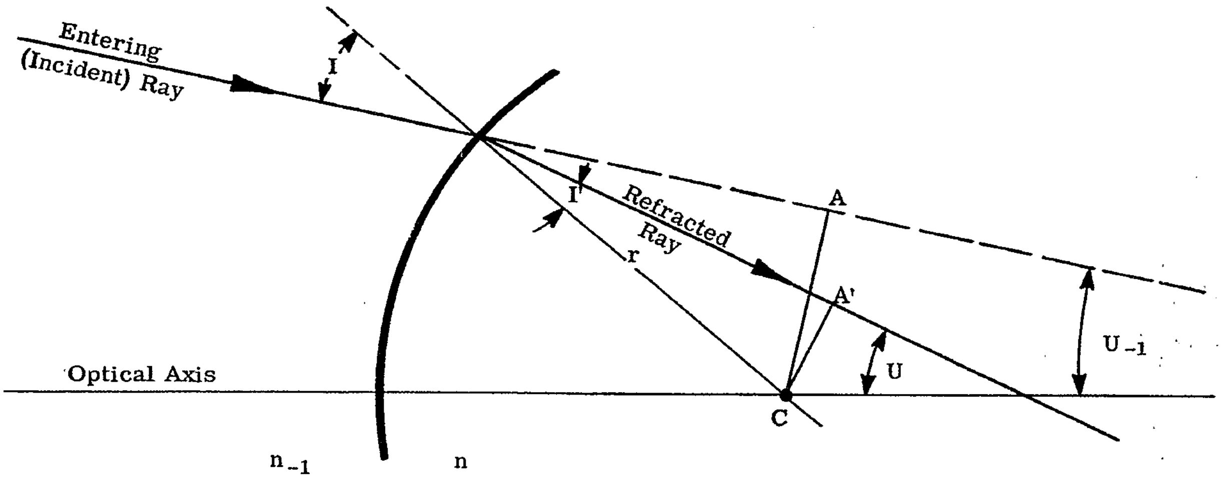

Figure 5. 12—Ray tracing by the PR method.

quantities specified, U_{-1} and CA . The corresponding quantities, U and CA', specify the refracted ray. This diagram involves the angles U and U_{-1} , called the slope angles. We will use a convention for the sign of a slope angle similar to that for incidence, reflection, and refraction angles. (See Section 2. 2. 2). If the ray must be rotated clockwise through the acute angle to bring it into coincidence with the optical axis the angle is called positive. Both U and U_{-1} are negative as drawn.

5. 6. 5. 2 The following equations are readily derived from the figure: *

$$\sin I = \frac{CA}{r},$$

and

$$\sin I' = \frac{CA'}{r}.$$

Therefore from Snell's law

$$n_{-1} \sin I = \frac{CAn_{-1}}{r} = \frac{CA'n}{r} = n \sin I'.$$

By definition:

$$P = CAn_{-1} = CA'n,$$

and

$$R = \frac{1}{n_{-1}\, r}, \quad R' = \frac{1}{nr}.$$

(Because of these two definitions, this method is referred to as the PR method.)

* The notation used in this simplified ray trace must not be confused with the skew ray formulae. There has been no attempt to avoid duplication of symbols.

The refraction equations then become

$$\sin I = PR , \quad (33)$$

$$\sin I' = PR' , \quad (34)$$

and

$$U = U_{-1} - (I - I') . \quad (35)$$

The value of P is transferred from one surface to the next by the following equation:

$$P_{+1} = P - (r - r_{+1} - t)\, n \sin U . \quad (36)$$

5.6.5.3 Equation (36) is seen to follow from Figure 5.13. We have

$$P_{+1} = C_{+1} A_{+1}\, n = C A'\, n + C\, C_{+1}\, n \sin U ,$$

because U is negative. The distance $C C_{+1} = t - r + r_{+1}$, and Equation (36) follows by rearrangement. The above ray tracing equations, (33) through (36), require a minimum of calculation and are ideal for hand computing. If several rays are to be calculated it is worth while to precalculate the lens constants R , R' , and $n(r - r_{+1} - t)$.

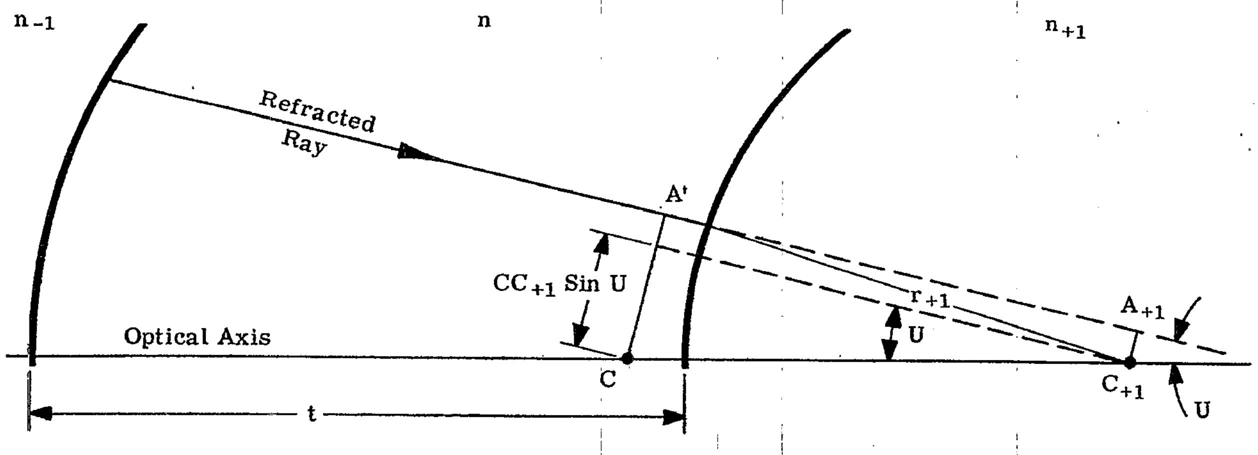

Figure 5.13 - Transfer procedure for the PR method.

5.6.5.4 A numerical example is shown in Table 5.5. The numbers above the double line are either given, such as r , t , and n , or are precalculated such as R , - R' , and $(r - r_{+1} - t)\, n$. The P below the line, surface 1, is calculated from initial ray data, CA . All other values in the table, below the double line, are calculated using Equations (33) through (36). The problem of finding angles I and I' from their sines, in order to use Equation (36), is discussed below.

SURFACE		1		2		3
r		19.23		-64.25		13.51
t			0.8		0.05	
n	1		1.51017		1	
R		0.0520021		-0.0103063		
-R'		-0.0344346		0.0155642		
$n(r - r_{+1} - t)$			124.861		-77.81	
P		3.330000		10.708028		1.703612
sin I		0.173167		-0.110360		
-sin I'		-0.114667		0.166662		
U	0		-0.059124		-0.115983	

Table 5.5 - Numerical example of ray tracing by the PR method.

5.6.5.5 One should note that the above formula, (36), cannot be used to transfer from a plane surface wherein $r \to \infty$, or to a plane surface wherein $r_{+1} \to \infty$. To deal with a plane, the procedure is to calculate the distance from the pole of the plane surface to the ray; see Figure 5.14.

Let

$$OA\, n_{-1} = Q \text{ for the entering ray, and}$$

$$OA'\, n = Q' \text{ for the refracted ray.}$$

Then, because $U_{-1} = I$, and $U = I'$, we have,

$$Q' = Q \frac{\tan U_{-1}}{\tan U} .$$

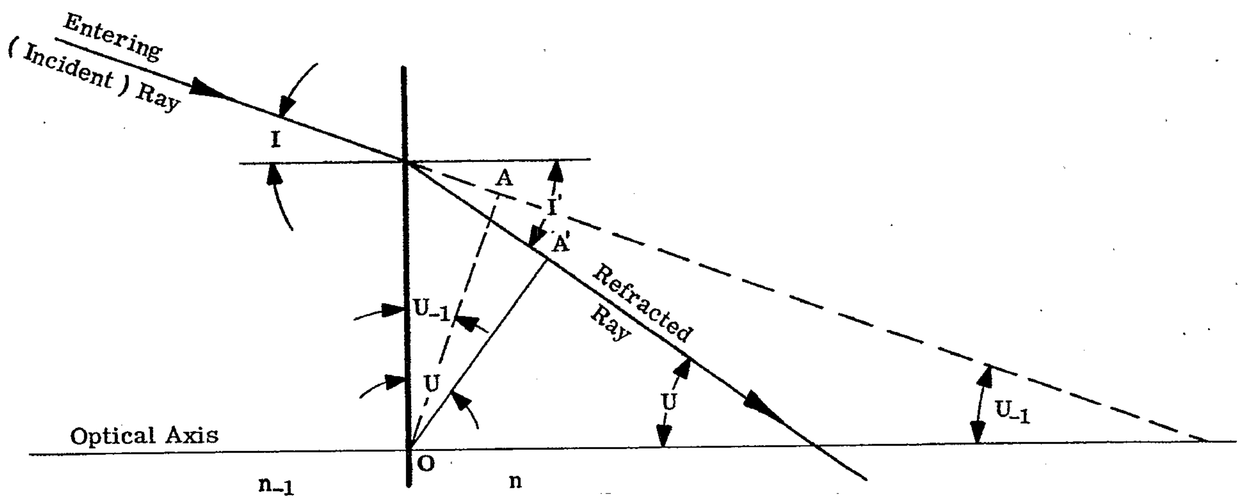

Figure 5.14 - Method of transfer for a plane surface.

To transfer from a spherical surface r , to a plane surface, $r_{+1} = \infty$, we use

$$Q_{+1} = P - (r - t)\, n \sin U .$$

To transfer from a plane surface, $r = \infty$, to a spherical surface, r_{+1} , the equation is

$$P_{+1} = Q' - (-r_{+1} - t)\, n \sin U .$$

5.6.5.6 To trace meridional rays through systems involving plane surfaces, Equations (33) through (36) are used until the plane surface is encountered. To transfer to a plane surface from a spherical one or vice versa, use one of the transfer equations in Paragraph 5.6.5.5. A transfer between two plane surfaces is calculated using

$$Q_{+1} = Q + t\, n \sin U .$$

Refraction at a plane surface is calculated using

$$\sin U = \frac{n_{-1}}{n} \sin U_{-1} ,$$

and

$$Q' = Q \frac{\tan U_{-1}}{\tan U} ,$$

where Q is either specified initially, calculated from initial data, or gotten by transfer from a previous surface. The calculations for plane surfaces are put into the same table (Table 5.5) as used for spherical surfaces. The values of $\sin U_{-1}$ and $\sin U$ are written opposite $\sin I$ and $\sin I'$ (which they equal respectively), and the values of Q and Q' are written opposite P . (The tangents need not be written down.)

5.6.5.7 One difficulty with the above formulation, Equation (36), is that if r_{+1} becomes large, but remains finite, P_{+1} becomes equal to the difference between a relatively small and a relatively large number. Hence unless a large number of significant figures are used for n , $\sin U$, and the coefficient of these terms, the value of P_{+1} will be independent of P . In doing hand computing one can readily notice this loss of precision. If this occurs, it is necessary to resort to other formulae, or reshape the lens so that the surface becomes plane. Another difficulty with Equation (36) arises if U becomes small, but remains finite. In this case the ray is almost parallel to the optical axis, and P_{+1} becomes equal to the difference of two nearly equal numbers. Hence unless both numbers are known to a large number of significant figures, the value of P_{+1} is quite inaccurate. In case the use of Equation (36) becomes difficult, the formulae given in Section 5.6.3 should be used.

5.6.5.8 In using the above equations it is necessary to convert sines to angles and to tangents, and to convert angles to sines. Tables are given in the Appendix. The tables convert from sine or tangent to the argument in radians and vice versa. They are designed for six place accuracy, and intervals are chosen for ease of interpolation. The first three digits of the function can always be found in the table and the last three digits are always multiplied by the interpolation constant and the product added to the tabular value. Interpolation therefore requires no mental arithmetic, and the process becomes completely automatic. By paying attention to such details a good human computer can trace rays through a lens at a speed of 40 to 60 seconds a surface. This method, in spite of its limitations, is an extremely useful method for hand computing meridional rays.

5.7 GRAPHICAL RAY TRACING PROCEDURE

5.7.1 Explanation of the method.

5.7.1.1 Rays may be traced graphically by means of a simple construction. The left side of Figure 5.15 shows a portion of two concentric circles whose radii are proportional to the indices n_{-1} and n . On the right side of the figure is shown the surface separating media of index n_{-1} and n . The angle of the refracted ray is determined from the diagram on the left. From this diagram $n_{-1} \sin I = n \sin I'$; thus, the construction solves Snell's law. Reference to Paragraph 5.4.5.3 will disclose that this is merely the graphical solution of the vector method.

5.7.1.2 The detailed procedure for tracing a ray is as follows. Draw a line through the center of the two circles parallel to the incident ray. Draw a line, parallel to the radius of curvature, through the intersection of the first line and the circle corresponding to the index of the object space. The line through the

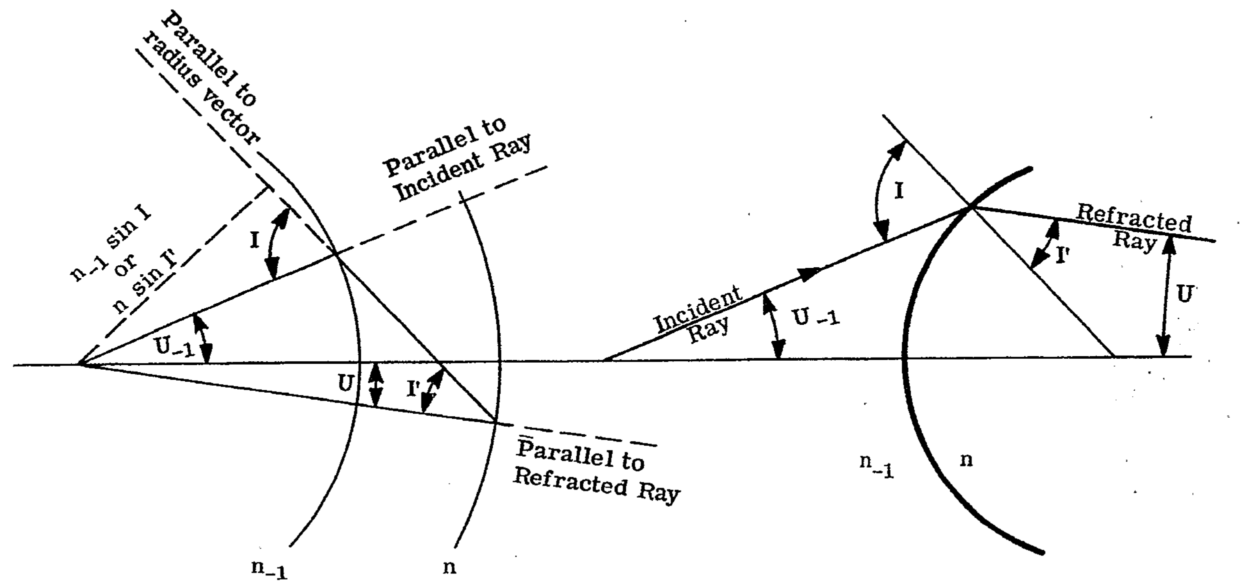

Figure 5.15- The method of tracing rays graphically.

center of the two concentric circles and the intersection of the second line with the other circle is the refracted ray. The incident and refracted rays can be drawn on the right hand diagram, but this is not necessary. The two diagrams may be superposed by placing the center of the concentric circles on the incident ray a distance n_{-1} (arbitrary units) to the left of the incidence point. This procedure makes unnecessary the drawing of the circle for n_{-1}, or the drawing of two lines each for the incident ray and the radius vector. The remainder of the construction is as given above.

5.7.2 Example using an air-spaced doublet. Figure 5.16 shows the graphical ray trace for a ray which is initially parallel to the axis (ray a). It is seen that the first surface of the second lens is a diverging surface; the other three surfaces are converging, because the ray is bent toward the optical axis. By measurement of the radii of the concentric circles, we see that $n_1 = 1.5$ and $n_2 = 1.7$. This combination of a converging crown lens, followed by a diverging flint lens is typical of a type of achromatic telescope objective. These lenses will be studied in detail in Section 11 .

5.8 DIFFERENTIAL RAY TRACING PROCEDURE

5.8.1 Meaning of a differentially traced ray.

5.8.1.1 In the previous sections equations have been developed for tracing a general ray (skew or meridional) through a general surface having rotational symmetry. Once such a ray has been traced through the system, we have a baseline from which to find the path of neighboring rays. A differentially traced ray, sometimes referred to as a close ray, is a ray differing from the originally traced ray by small, first order quantities. This means that the change in direction cosines, dK_{-1}, dL_{-1}, dM_{-1}, and the change in the coordinates of the intersection point, dX, dY, dZ, are first order differentials.

5.8.1.2 The tracing of one ray gives us information about the one intersection point of that ray with the image surface. The tracing of several neighboring rays gives us their intersection points and hence information about the structure of the image formed by these rays. In addition to this useful information, differentially traced rays are generally easier to calculate than a single, general ray. Because of these advantages, the concepts and procedures of differential ray tracing are important.

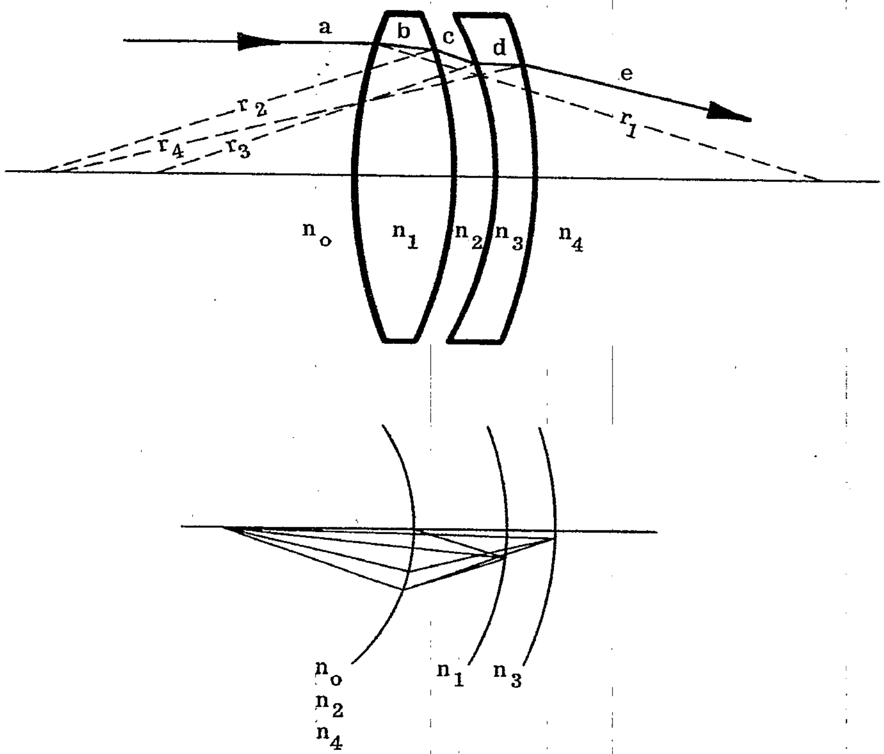

Figure 5.16 - Graphical ray trace of a doublet.

5.8.2 Differentially traced skew ray.

5.8.2.1 Once a skew ray has been traced through a lens system it is possible to trace the path of a ray differentially displaced from it. The skew ray trace provides the values of X, Y, Z on each surface, and K, L, M between surfaces. The values of X, Y, Z on adjacent surfaces are linked by the transfer equations

$$X = X_{-1} + \frac{D_{-1}}{n_{-1}} K_{-1} , \qquad (37)$$

$$Y = Y_{-1} + \frac{D_{-1}}{n_{-1}} L_{-1} , \qquad (38)$$

and

$$Z = Z_{-1} - t_{-1} + \frac{D_{-1}}{n_{-1}} M_{-1} , \qquad (39)$$

where D_{-1}, given by Equation (25a), is the geometrical distance along the skew ray between the two surfaces. These equations follow from Paragraph 5.4.3.2 applied to any two surfaces.

5.8.2.2 A neighboring ray, in the sense of Paragraph 5.8.1.1, will have slightly different coordinates on the jth surface. The differences, dX, dY, and dZ are found by differentiating Equations (37), (38), and (39). We have

$$dX = dX_{-1} + \frac{D_{-1}}{n_{-1}} dK_{-1} + K_{-1}\, d\left(\frac{D_{-1}}{n_{-1}}\right) , \qquad (40)$$

$$dY = dY_{-1} + \frac{D_{-1}}{n_{-1}} dL_{-1} + L_{-1}\, d\left(\frac{D_{-1}}{n_{-1}}\right) , \qquad (41)$$

and

$$dZ = dZ_{-1} + \frac{D_{-1}}{n_{-1}} dM_{-1} + M_{-1}\, d\left(\frac{D_{-1}}{n_{-1}}\right) . \qquad (42)$$

These equations may be referred to as differential transfer equations, in that they are used to calculate the change in coordinates. The changes in coordinates at the previous surface have been determined by the previous application of these equations; the changes in optical direction cosines are calculated by differential refraction equations discussed below. The last term, involving the change in total ray length, must also be calculated. The remaining equations will first be derived; then their order of use will be summarized.

5.8.2.3 The procedure used to derive an equation for $d\left(\frac{D_{-1}}{n_{-1}}\right)$ will be quite similar to that used to derive an equation for $\frac{A}{n_{-1}}$. (See Paragraphs 5.4.4.3 - 5.4.4.5). In that case we used four equations, Equations (4), (5), (6) and the equation for the sag, Paragraph 5.4.4.4. These four equations were solved simultaneously for $\frac{A}{n_{-1}}$. The equation for the sag, Z , is the equation for the surface, in that case the sphere. Because the intersection point must lie on the surface, this equation is called an equation of constraint. In the present case, the four equations to be used are Equations (40), (41), (42) and the differential equation of constraint.

5.8.2.4 Although the physical j th surface is a general surface of revolution, this surface is replaced by the plane, tangent at the intersection point. The reason this must be done is that we have restricted the change in coordinates to first order differentials; as one moves away from a point on a surface by distances of the order of first differentials, the motion is constrained to the plane tangent to the surface. The equation of the tangent plane is given in Paragraph 5.5.4.10. Differentiating this to obtain the differential equation of constraint we have

$$U dX + V dY + W dZ = 0 .$$

We now substitute into this equation the values of dX , dY , and dZ given by Equations (40), (41), and (42). Collecting terms, and using Equation (27), we have

$$d\left(\frac{D_{-1}}{n_{-1}}\right) = \frac{U\left(dX_{-1} + \frac{D_{-1}}{n_{-1}} dK_{-1}\right) + V\left(dY_{-1} + \frac{D_{-1}}{n_{-1}} dL_{-1}\right) + W\left(dZ_{-1} + \frac{D_{-1}}{n_{-1}} dM_{-1}\right)}{-G\, n_{-1} \cos I} . \quad (43)$$

5.8.2.5 Using Equation (43), and then Equations (40), (41), and (42), we will have completed the transfer of the differentially traced ray. The differential refraction equations, now to be derived, will be used to calculate dK , dL , and dM . Differentiating Equations (30) to (32) gives

$$dK = dK_{-1} + P dU + U dP , \quad (44)$$

$$dL = dL_{-1} + P dV + V dP , \quad (45)$$

and

$$dM = dM_{-1} + P dW + W dP . \quad (46)$$

5.8.2.6 In differentiating the equation for the tangent plane we kept U , V , and W constant and thereby obtained the differential equation of constraint. Physically this means that at any point on this tangent plane the ratio of the direction cosines of the normal, U : V : W , is the same as at any other point. (See Paragraph 5.5.5.3). Justifiably it may be asked why U , V , W were not held constant in differentiating Equations (30), (31), and (32). The answer is that though the tangent plane and surface differ by second order differentials, at the new intersection point, the normals to the two tangent planes, erected at the two intersection points, have direction cosines differing by first order differentials. Hence, since refraction involves the normal at the intersection point, dU , dV , and dW are not necessarily zero in Equations (44), (45), and (46).

5.8.2.7 Differentiating Equations (23), (24), and (18), we get

$$dU = -X dE - E dX , \quad (47)$$

$$dV = -Y dE - E dY , \quad (48)$$

and

$$dW = -\frac{c^2}{W} (X dX + Y dY) . \quad (49)$$

dE may be found by differentiating Equation (22), remembering that

$$dE = \frac{\partial E}{\partial X} dX + \frac{\partial E}{\partial Y} dY + \frac{\partial E}{\partial Z} dZ ,$$

thus

$$dE = - \left[\frac{c - E}{W} + \frac{2W^2}{c^2} (4e + 12fS^2 + 24gS^4 + 40hS^6) \right] dW . \qquad (50)$$

5.8.2.8 The one remaining problem is the determination of dP to be used in Equations (44), (45), and (46). This is done by the same method used to derive Equation (43). The four equations used are Equations (44), (45), (46), and a differential equation of constraint. In each case, the fourth equation involves the differentials on the left-hand side of the first three equations. Because the sum of the squares of the direction cosines of a given line is unity, we have

$$KdK + LdL + MdM = 0$$

as the differential equation of constraint. Substituting Equations (44), (45), and (46) into this constraint, and remembering Equation (27), we have

$$dP = \frac{K(dK_{-1} + PdU) + L(dL_{-1} + PdV) + M(dM_{-1} + PdW)}{- G n \cos I'} . \qquad (51)$$

5.8.2.9 We can now summarize the calculations, in the order made, used in tracing a differentially traced skew ray through aspheric surfaces. In addition to the results for the skew ray, available from tables such as 5.3 and 5.4, the initial ray data of the neighboring ray must be given. That is dX_{-1}, dY_{-1}, dZ_{-1}, dK_{-1}, dL_{-1}, and dM_{-1} must be specified. The following equations are used in the order given here: (43), (40), (41), (42), (49), (50), (47), (48), (51), (44), (45), and (46). With an automatic computer it does not seem to be worthwhile to trace close skew rays since the regular skew rays can be traced so rapidly. However for hand computing the close skew ray trace is a very valuable tool.

5.8.3 Differentially traced meridional ray. At first sight it might appear that it would take as much time to trace a differentially traced ray as a skew ray. Actually the equations are simple and no square roots are involved. An interesting application of the above equations occurs in connection with meridional rays. Assume a meridional ray has been traced from an object point ($X_o = 0$, Y_o, Z_o) through the lens system; let us now trace a ray from the same object point which will be differentially displaced by the amount dK_o. We also assume that $dL_o = 0$. Since

$$KdK + LdL + MdM = 0 ,$$

and the differential ray is to be traced around a meridional ray $K_o = 0$, $dM_o = 0$. (If originally we had assumed $dM_o = 0$, then it would follow that $dL_o = 0$). Equation (43) shows that $d(D_o/n_o) = 0$.

From Equations (40), (41) and (42),

$$dX_1 = \frac{D_o}{n_o} dK_o ,$$

$$dY_1 = 0 ,$$

and

$$dZ_1 = 0 .$$

Careful inspection of Equations (47) to (51) shows that,

$$dU = - EdX ,$$

$$dV = 0 ,$$

$$dW = 0 ,$$

$$dE = 0 ,$$

and

$$dP = 0 .$$

Substitution into Equation (44) gives the relation

$$dK_1 = dK_o - \left[\frac{G\, n_1 \cos I' - G\, n_o \cos I}{G^2} \right] E_1\, dX_1 .$$

It then follows, since $d\left(\frac{D_1}{n_1}\right) = 0$, that

$$dX_2 = dX_1 + \frac{D_1}{n_1}\, dK_1 ,$$

and

$$dK_2 = dK_1 - \left[\frac{G\, n_2 \cos I' - G\, n_1 \cos I}{G^2} \right] E_2\, dX_2 .$$

The close meridional ray may be traced through the system by successive application of the equations

$$dX = dX_{-1} + \frac{D_{-1}}{n_{-1}}\, dK_{-1} \tag{52}$$

and

$$dK = dK_{-1} - \left[\frac{G\, n \cos I' - G\, n_{-1} \cos I}{G^2} \right] E\, dX . \tag{53}$$

5.8.4 The Coddington equations.

5.8.4.1 The above two equations,(52) and (53), apply to a general surface having rotational symmetry. In case the surface is spherical, Equation (53) is simplified since $E = c$ and $G = 1$. If the close ray has $dL_o = dM_o = 0$, as in the above example, and if the traced meridional ray and the close ray intersect to form an image, these two rays obey one of the Coddington equations, namely,

$$\frac{n_o}{D_o} + \frac{n_1}{D_1} = c\,(n_1 \cos I' - n_o \cos I) .$$

Because the close ray was shifted in a way that resulted in $dY_1 = dZ_1 = 0$, the shift of the intersection point occurred parallel to the X axis, in other words in the sagittal plane. The resulting focus is referred to as the sagittal focus, or the skew focus, because the close ray is actually a skew ray.

5.8.4.2 The above Coddington equation can be derived from Equations (52) and (53) applied to a spherical surface, ($E = c$ and $G = 1$). Because we are dealing with a single object and single image point, $dX_o = dX_2 = 0$. Applying these two equations to Equation (52), we have

$$dX_1 = \frac{D_o}{n_o}\, dK_o = - \frac{D_1}{n_1}\, dK_1 .$$

Using Equation (53), for a spherical surface,

$$dK_1 = dK_o - (n_1 \cos I' - n_o \cos I)\, c\, dX_1 ,$$

and, expressing dK_o in terms of dK_1, and dX_1 in terms of dK_1, we get

$$dK_1 = - \frac{D_1}{n_1} \frac{n_o}{D_o}\, dK_1 + (n_1 \cos I' - n_o \cos I)\, c\, \frac{D_1}{n_1}\, dK_1 .$$

Simplification gives the above Coddington equation.

5.8.4.3 Instead of shifting the ray in a plane perpendicular to the meridional plane, the ray could have been shifted in the meridional plane. In this case, $dK = dX = 0$. In a manner similar to that used in Section 5.8.3, ray trace equations for dY and dL can be derived, corresponding to Equations (52) and (53). (We do not need specific equations for dZ and dM, because these are proportional to dY and dL respectively). For a single image to be formed by two close rays from a single object point, $dY_o = dY_2 = 0$. The final result is the second Coddington equation involving the meridional or tangential focus,

$$\frac{n_o \cos^2 I}{D_o} + \frac{n_1 \cos^2 I'}{D_1} = c\,(n_1 \cos I' - n_o \cos I) .$$

5.9 PARAXIAL RAYS

5.9.1 The paraxial ray concept. The previous section on differentially traced meridional rays provides a good way to introduce the concept of paraxial ray tracing and the meaning of paraxial rays. A ray passing directly along the optical axis of the system is a perfectly good ray to use as a base from which to trace a close, neighboring ray. Such a ray, differentially traced with respect to the optical axis, is a paraxial ray. Physically, paraxial rays are the rays that get through the system as the aperture of each lens, centered concentrically with respect to the optical axis, becomes very small. Because paraxial rays are fairly easy to visualize, and because the ray tracing equations become quite simple for these rays, the usefulness of paraxial rays in the preliminary design of optical elements cannot be overemphasized.

5.9.2 Ray trace equations.

5.9.2.1 For a ray coinciding with the optical axis, $\cos I$ and $\cos I'$ will be exactly equal to 1 on every surface and $D_{-1} \rightarrow t_{-1}$, so Equations (52) and (53) become

$$dX = dX_{-1} + \frac{t_{-1}}{n_{-1}} dK_{-1} \quad (54)$$

and

$$dK = dK_{-1} - (n - n_{-1}) c dX . \quad (55)$$

Therefore a ray may be traced differentially close to the optical axis by applying the above equations. Since the original ray was the optical axis, there is no distinction between the X and Y axes, and these equations apply equally well for a close ray in the YZ plane. For such a ray, replace dX by dY and dK by dL, for each part of the system. It should be noted that these equations hold for aspheric as well as spherical surfaces. Mathematically this is so because for the optical axis, $X = Y = 0$; hence by Equations (18) and (26), $W = 1 = G$, and by Equation (22), $E = c$. Physically the aspheric and the sphere are tangent at the optical axis and have the same curvature; hence a ray close to the axis intersects a surface of curvature c.

5.9.2.2 Paraxial ray calculations will be used so extensively to build up an understanding of optical systems, that a special notation will be used to refer to paraxial data. It is customary to use lower case letters for paraxial rays. Equations (54) and (55) will be written for a ray in the YZ plane and become

$$y = y_{-1} + \frac{t_{-1}}{n_{-1}} (n_{-1} u_{-1}) , \quad (56)$$

and

$$nu = n_{-1} u_{-1} + yc (n_{-1} - n) . \quad (57)$$

The differentials have been replaced by small letters indicating paraxial ray data. One can see that dY has been replaced by y, indicating a small displacement perpendicular to the optical axis. dL, which replaces dK for a paraxial ray in the YZ plane, is the change in the optical direction cosine of the originally traced ray. Since the original ray is the axial ray, and the original $L = 0$, $dL =$ new value of $L = n \cos \beta$, where β is the angle between the ray and the Y axis. Instead of $\cos \beta$, we can use $\sin U$, the angle between the ray and Z axis. Therefore $dL = n \sin U$. But U is a small angle, and we replace the $\sin U$ by U, the first order approximation. (See Section 5.11). Hence $dL = nU$, and using small letters, $dL = nu$. We see here why the term paraxial ray optics and first order optics are synonymous.

5.9.3 The use of finite angles and heights for paraxial rays.

5.9.3.1 Equations (56) and (57) were derived on the assumption that y and u are small, of the order of first order differentials. Physically, in order to form an image using paraxial rays, the actual rays must obey the condition that y and u are small. It is, however, both a remarkable and extremely useful fact that in ray tracing, we may use finite heights and angles, not necessarily small, for y and u. We will show this in the following paragraph.

5.9.3.2 Consider Figure 5.17 which indicates two rays from an axial object point O to the corresponding axial image point O'. Because u_{-1} and u in Equations (56) and (57) were assumed small, we can replace them by $\tan u_{-1}$ and $\tan u$ respectively. (The expansion of $\tan u$, in terms of u, shows that the first order approximation is $\tan u = u$, as in the case of $\sin u$. The third order approximation, how-

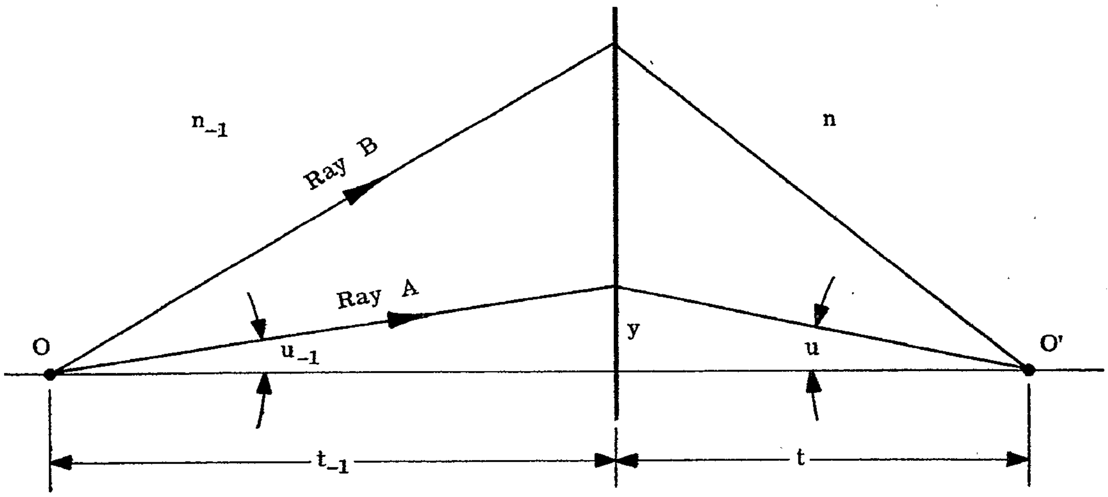

Figure 5.17. Paraxial rays through a single refracting surface.

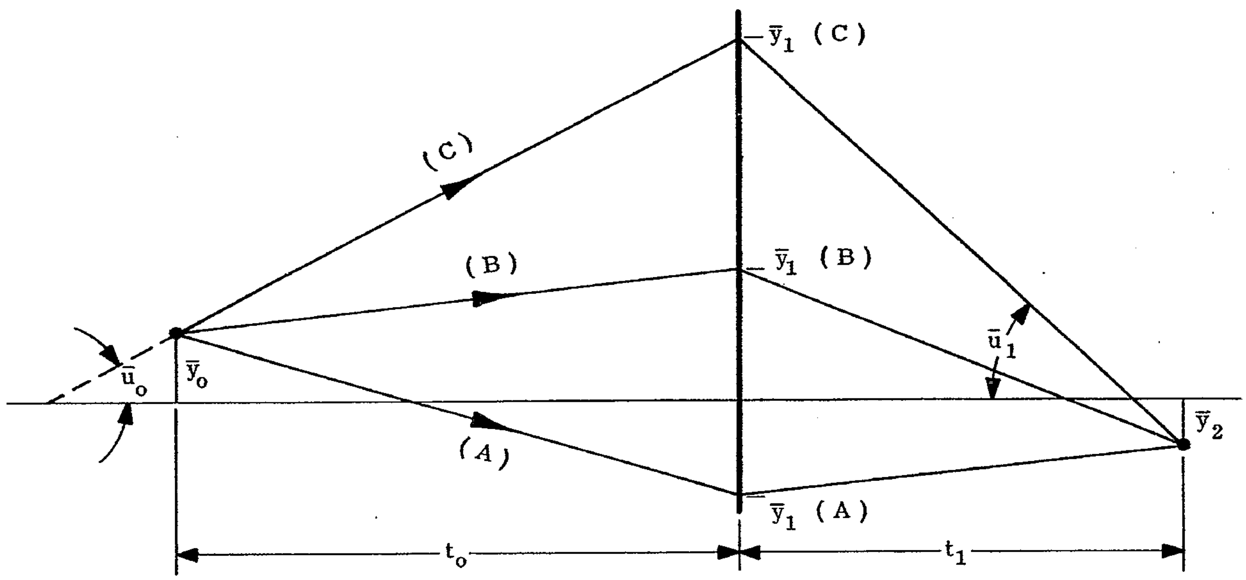

Figure 5.18 - Paraxial rays through a single refracting surface.

ever, differs from that of sin u). Remembering that u is negative, we substitute into Equation (57) and find

$$n \; -\left(\frac{y}{t}\right) = n_{-1}\,\frac{y}{t_{-1}} + \frac{y}{r}\,(n_{-1} - n)\,.$$

Upon rearrangement we get

$$\frac{n_{-1}}{t_{-1}} + \frac{n}{t} = \frac{n - n_{-1}}{r}\,,$$

which is the familiar form of the paraxial equation for a single surface. The important thing to note is that u_{-1} , u , and y no longer appear in this equation. This fact is interpreted as meaning that mathematically we may consider the image O' formed by any ray leaving the object O . Thus both rays (A) and (B) intersect the axis at the same image point.

5.9.3.3 Figure 5.17 and the above paragraph apply to axial object and image points. The same conclusions concerning finite heights and angles hold for rays through off-axis object and image points. Hence, in Figure 5.18, all rays (A), (B), and (C) intersect at one image point. Neither the angles $\bar{u}_o$ or $\bar{u}_1$, nor the heights $\bar{y}_o$, $\bar{y}_1$, or $\bar{y}_2$, need be small. *

5.10 GRAPHICAL RAY TRACE FOR PARAXIAL RAYS

5.10.1 Specialization of the general graphical method.

5.10.1.1 Paraxial rays may be traced graphically through a lens system by a construction very similar to the construction shown in Section 5.7. This is done by replacing the refractive index circles by tangent planes, and the curves of the lens surfaces by tangent planes through the vertices of the surfaces. The justification for these replacements will be given in Paragraph 5.10.1.3. For paraxial rays, the construction will appear as shown in Figure 5.19.

5.10.1.2 In the above paragraph we have indicated that Figure 5.19 is correct for paraxial rays.

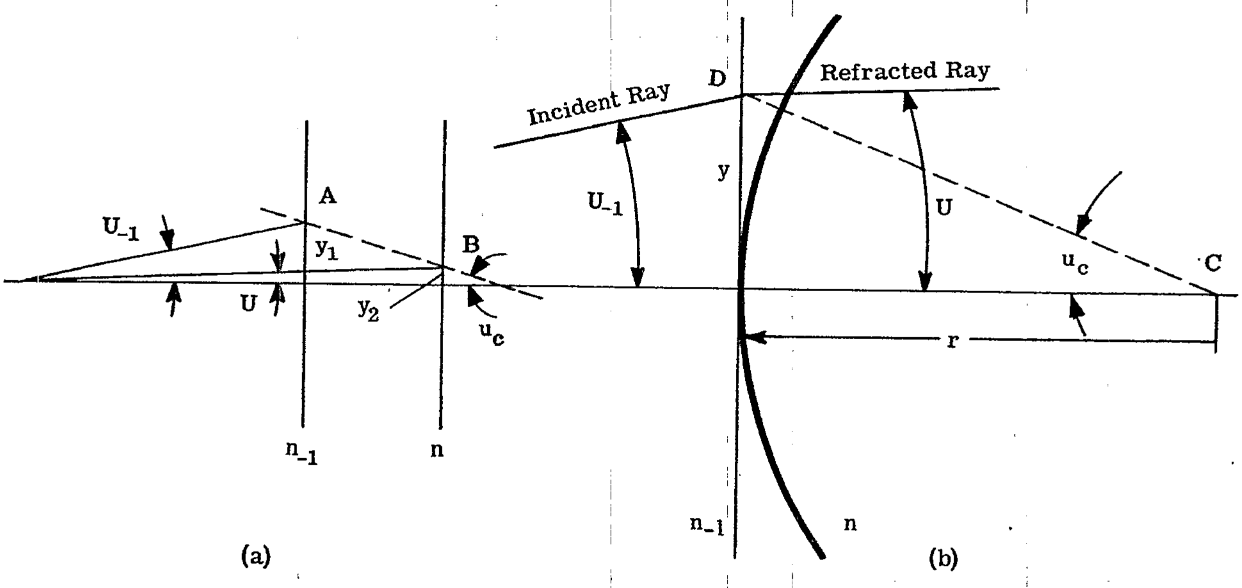

Figure 5.19 - The method for tracing paraxial rays graphically.

* The angles and heights corresponding to rays through off-axis object and image points are written with a line or bar over the symbol, as $\bar{u}$ and $\bar{y}$.

Assuming this let us use the drawing to reexamine and extend the ideas discussed in Section 5.9.3. From Figure 5.19 it can be seen that

$$y_1 = n_{-1} \tan U_{-1}$$

and

$$y_2 = n \tan U .$$

Since the line connecting A B (a) is parallel to line DC (b) it is clear that, from similar triangles,

$$\frac{y_1 - y_2}{y} = \frac{n - n_{-1}}{r} .$$

By inserting the expressions for y_1 and y_2 into the last equation, we find on rearranging

$$n \tan U = n_{-1} \tan U_{-1} + yc(n_{-1} - n) .$$

5.10.1.3 This last equation, derived from Figure 5.19, is correct for small angles. This is easily seen, because when the $\tan U_{-1}$ and $\tan U$ are replaced by the angles u_{-1} and u respectively, Equation (57) results. However let us assume for the moment that both Figure 5.19 and the last equation are correct for any angles U and U_{-1} . In particular, these may be finite angles and do not have to be small. Now if we compare this equation to Equation (57), which is true for small angles u and u_{-1} , and therefore for paraxial rays, we see that $\tan U$ and $\tan U_{-1}$ correspond to u and u_{-1} , respectively. This indicates that the equation derived from Figure 5.19 can be used in connection with paraxial rays, provided the angles u and u_{-1} are replaced by $\tan U$ and $\tan U_{-1}$ respectively. Since U can have any value, $\tan U$ and therefore u can have any value. Equation (57) and Figure 5.19 can therefore be used for paraxial rays incident at any finite height and making any finite angle with the optical axis. Equation (56) and Figure 5.19 can also be used to accurately transfer the value of y from one surface to another for paraxial rays. This equation and figure can also be used with non-paraxial meridional rays to transfer between plane surfaces; in this case u_{-1} is replaced by $\tan U_{-1}$.

5.10.2 Two approaches to the treatment of paraxial rays.

5.10.2.1 We have shown that paraxial rays can be considered from either of two points of view:

(1) We use small angles and finite curvatures for surfaces. This led to Equation (57).

(2) We use finite angles and zero curvatures for surfaces. This led to Figure 5.19. It must be emphasized that we do not have to combine these and use small angles with plane surfaces.

5.10.2.2 It is convenient then to think of paraxial rays as passing through the optical system at finite heights, striking the surfaces on the tangent planes instead of the actual curved surfaces. Since the two Equations (56) and (57) are linear equations, and since the location of images are found for values of $y_k = 0$, it makes no difference what value of u is used. It is instructive to trace paraxial rays through a lens at heights equal to the actual ray heights, and note the difference in path for a paraxial ray and an actual ray. This is demonstrated for a single surface refraction in Figure 5.20. The ray traced through the curved surface crosses the axis at M , closer to the surface than the point P . The paraxial ray crosses at P , further away from the surface. This defect of focus is called spherical aberration.

5.11 THE DIFFERENT "ORDERS" OF OPTICS

5.11.1 Expansion of the sine function.

5.11.1.1 The fundamental equations which have been discussed and used in tracing rays are: (1) the transfer equations, and (2) the refraction equations. Both have been put into a form explicitly using the cosine function of various angles, such as the angles of incidence and refraction, and the angles which the ray makes with the coordinate axes. Both equations could have been written in terms of the sine function; so as to explain the meaning of the phrase orders of optics we will deal with the sine function.

5.11.1.2 The optical axis is a special ray for which both angles of incidence and refraction are zero. In addition the angles which this ray makes with the X , Y , and Z axes are 90°, 90°, and 0° respectively. For a meridional ray near the axis, the angles of incidence and refraction, and the angle with the Z axis, are small. The ray trace equations, therefore, involve the sines of small angles. As the meridional ray

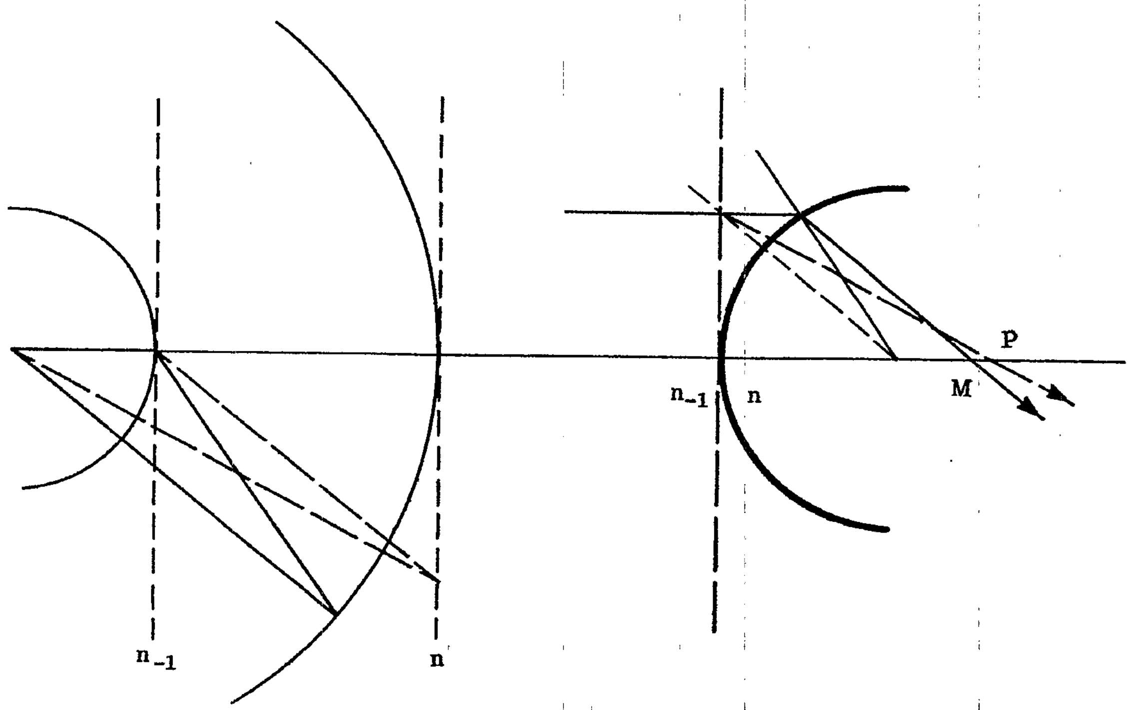

Figure 5.20 - Comparison between a paraxial and an actual ray showing spherical aberration.

makes larger and larger angles with the Z axis, we have to be concerned with the sines of larger and larger angles.

5.11.1.3 One reason the ray trace equations are complicated is that they involve the trigonometric functions of angles, instead of just the angles. (We have seen in Section 5.9 how the equations are greatly simplified when they can be expressed in terms of angles, instead of trigonometric functions). To relate the $\sin \alpha$ to the angle α, we expand the sine function in a series, thus

$$\sin \alpha = 0 + \alpha - 0 - \frac{\alpha^3}{3!} + 0 + \frac{\alpha^5}{5!} - 0 - \frac{\alpha^7}{7!} + \ldots$$

5.11.1.4 The terms given explicitly as zero have been written down to clarify the situation. Whenever a function is expanded in a series the "first" term is called the <u>zeroth approximation</u> or the <u>zeroth order</u>, and successive terms are called the <u>first</u>, <u>second</u>, <u>third</u>, etc., <u>orders</u>. In the case of the expansion of the sine function, the zeroth, second, fourth, and all even order terms are identically zero; only the odd orders remain.

5.11.2 <u>First order optics.</u> If in ray trace equations the sine is replaced by the angle, we are using the zeroth and the first order terms in the above expansion. The resulting equations and design procedures are called <u>first order optics</u>, and the rays concerned are paraxial rays. One of the fascinating parts of geometrical optics is the extensive understanding of lens systems one can obtain by tracing two paraxial rays. With two paraxial rays one can predict the location and size of any image formed with paraxial rays, and by making further calculations based on these paraxial ray data it is possible to predict the approximate magnitude of image errors. The following sections, 6 and 7, will be devoted to the development and use of the equations of first order optics. This development will be based on the two simple equations, (56) and (57).

5.11.3 <u>Third order optics.</u>

5.11.3.1 If the first and third order terms in the expansion of the sine are retained, the resulting equations are part of <u>third order optics.</u> But this term has an added meaning, pertaining to aberrations, and it is usually in this latter sense that the term is used.

5.11.3.2 The intersection of a ray with the image surface locates the image. If the intersection has been computed using the skew ray trace equations, the intersection is the true image. If paraxial ray trace equations have been used, the resulting paraxial image will generally be displaced from the true image. The difference between the true image and the first order approximation (paraxial approximation) is known by the general term aberration. (We are considering here monochromatic light only. Aberrations due to non-monochromatic light are considered in Paragraph 5.11.3.4.) In the same way that the sine was expanded in a series, the aberrations can be expanded. The first term in the expansion is known as the third order aberration. The reason for this is that it represents the first approximation to the total aberration, and hence can be considered as the difference between the paraxial image and the image using the third order approximation for the sine. Third order optics then has come to mean the equations and procedures dealing with the first approximation to the aberrations. It is fortunate, as will be evident in a later section (Section 8) that these third order aberrations can be calculated from first order (paraxial) ray trace data.

5.11.3.3 The next term in the expansion of the aberration, after the third order aberration, is called the fifth order aberration. Fifth order optics deals with the aberrations through the fifth order aberration term.* Hence fifth order optics deals with fifth order aberration, or the second approximation to the aberration.

5.11.3.4 Aberrations due to non-monochromatic light can also be expanded in a series. The first term gives the aberration appearing in paraxial images, hence is referred to as first order aberration. This is treated in Section 6, dealing with first order optics.

* In some countries other than the United States, for example England, the first, second, third, etc., terms in the aberration expansion are referred to as primary, secondary, tertiary, etc. aberration.