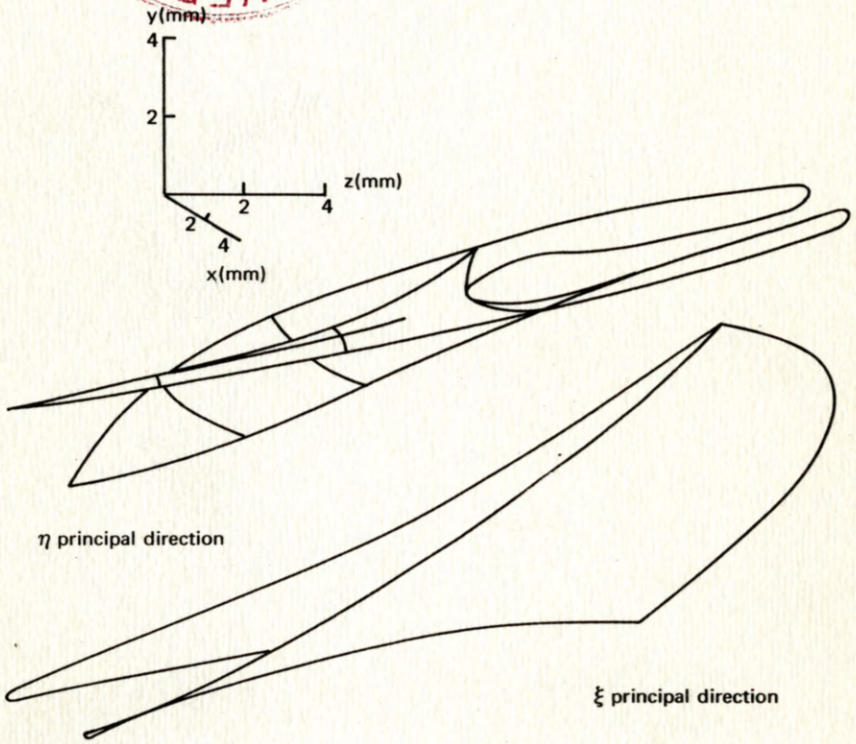
OPTICAL SCIENCES CENTER

University of Arizona



TECHNICAL REPORT NUMBER 71 NOV 1971

Properties and Applications of Generalized Ray Tracing

Steven C. Parker

PROPERTIES AND APPLICATIONS OF GENERALIZED RAY TRACING

Steven C. Parker

Optical Sciences Center, University of Arizona, Tucson, Arizona 85721
Technical Report 71, November 1971

FOREWORD

This technical report is adapted from a thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Optical Sciences at the University of Arizona.

The thesis was completed and approved on June 18, 1971.

ABSTRACT

Generalized ray tracing is a method of calculating the principal curvatures and directions of the wavefront associated with a ray as it is traced through an optical system. The results of such a ray trace provide important information about the structure of the image and have immediate application to lens design and image analysis.

The caustic surface formed by an optical system can be easily determined from the results of a generalized ray trace. An examination of several caustic surfaces formed by different optical systems provides valuable information about the relative quality of the images and indicates some of the advantages of the generalized ray tracing method.

CONTENTS

INTRODUCTION

Geometrical ray tracing has always been basic to any system of lens design and analysis. As useful as this approach is, however, it does not completely describe the characteristics of an optical system. For example, it would be extremely valuable to be able to describe completely the changes in shape a wavefront undergoes as it propagates through an optical system. Unfortunately this problem is still unsolved. However, if we can be satisfied to look at a very small region on the wavefront in the neighborhood of a ray, we can describe the changes in shape that it undergoes as it propagates through an optical system. Such a system of ray tracing, called *generalized ray tracing,* is the subject of this thesis.

This approach is not new. As early as 1906 Gullstrand derived equations to calculate the principal curvatures and directions of the surface associated with a ray. Despite the obvious value of the method, the complexity of the calculations greatly limited its applicability. More recently, Kneisly (1964) rederived the equations using modern vector notation. Perhaps the most comprehensive treatment of generalized ray tracing and its possible applications is to be found in a forthcoming book by O. N. Stavroudis (1972).

In this paper we will generally follow the derivation outlined by Kneisly and will use the notation of Stavroudis. Although this method can be applied to any media for which the ray tracing equations are known, we will assume for the sake of simplicity that the media are homogeneous and the refracting surfaces are spherical.

It should be pointed out that the term *wavefront* will be used strictly in the context of geometrical optics. In this context a wavefront is defined to be the surface orthogonal to a two-parameter family of rays. On this wavefront we can always choose an orthogonal coordinate system consisting of the unit normal to the surface, which is in the direction of the geometrical ray, and two unit vectors that are tangent to the surface in the principal directions.

Generalized ray tracing, then, is geometrical ray tracing generalized to include the calculation of the principal curvatures and directions of the wavefront in the neighborhood of the ray. As in geometrical ray tracing, the operations we will be concerned with are transfer and refraction. Both of these operations must now be extended to apply to the principal curvatures as well as to the ray.

Differential Geometry of Surfaces

Before we can properly treat wavefronts, we need to know some of the properties of surfaces in general. If we examine a small enough region around a point on a sufficiently smooth surface, we find that, no matter how complicated the surface is, it has some rather remarkable properties at that point. We shall discuss the theorems that describe those properties that are pertinent to generalized ray tracing.

The first theorem that we will need is that of Meusner (Struik, 1961, pp. 73-76). If we pass a plane through a point on the surface **W**, written as a vector function of the two essential parameters, u and v, then the curve **P** will be defined as the intersection of the plane and the surface. If we write u and v as functions of a third parameter t, then the expression for the curve can be written

$$\mathbf{P}(t) = \mathbf{W}\Big(u(t),v(t)\Big).$$

The normal to the curve, which we will call **n**, will then form an angle θ with the surface normal **N**. As the angle θ is varied, the curvature of **P** varies inversely with $\cos\theta$. Indeed, it can be shown that

$$\frac{\cos\theta}{\rho} = \frac{Lu_t^2 + 2Mu_tv_t + Nv_t^2}{Eu_t^2 + 2Fu_tv_t + Gv_t^2},$$

where $1/\rho$ is the curvature of **P**. The quantities E, F, and G are the first fundamental quantities, and L, M, and N are the second fundamental quantities. Unless otherwise indicated, subscripts represent partial derivatives. From the above expression and the fact that the fundamental quantities do not depend on θ, it can be seen that the curvature has a minimum value when θ is zero. In this case the curve **P** is known as a normal section and has a curvature

$$1/\rho = \frac{Lu_t^2 + 2Mu_tv_t + Nv_t^2}{Eu_t^2 + 2Fu_tv_t + Gv_t^2}. \tag{1}$$

Another important theorem that we will need is that of Gauss. When θ is zero and we rotate our plane about **N**, the curvature of the resultant normal section will in general pass through a maximum and a minimum value. The directions in which the curvature attains these extremal values are called principal directions. From Gauss' theorem it can be shown (Struik, 1961, p. 80) that no matter how complicated the shape of the surface is, as long as it is sufficiently smooth the principal directions will always be perpendicular to each other. The principal curvatures themselves can be calculated from the fundamental quantities (Struik, 1961, p. 81). If we let ξ and η represent the principal directions, then

$$1/\rho_\xi = L/E$$
$$1/\rho_\eta = N/G.$$

Once we know the curvature in the principal directions, then from Euler's theorem for normal curvatures we can calculate the curvatures of any perpendicular pair of normal sections (Struik, 1961, p. 81)

$$1/\rho_u = (1/\rho_\xi)\cos^2\phi + (1/\rho_\eta)\sin^2\phi \tag{2}$$

$$1/\rho_v = (1/\rho_\xi)\sin^2\phi + (1/\rho_\eta)\cos^2\phi \tag{3}$$

where ϕ is the angle of rotation between the normal section in the principal direction and the normal section in the u direction. For convenience later on we will introduce a new symbol

$$1/\sigma = M(EG)^{-1/2}.$$

This can be written in terms of the principal curvatures as

$$1/\sigma = \tfrac{1}{2}(1/\rho_\xi - 1/\rho_\eta)\sin 2\phi. \tag{4}$$

The reciprocal relations are

$$1/\rho_\xi = (1/\rho_u)\cos^2\phi + (1/\rho_v)\sin^2\phi + (2/\sigma)\sin\phi\cos\phi \tag{5}$$

$$1/\rho_\eta = (1/\rho_u)\sin^2\phi + (1/\rho_v)\cos^2\phi - (2/\sigma)\sin\phi\cos\phi \tag{6}$$

$$\tan 2\phi = (2/\sigma)/(1/\rho_v - 1/\rho_u). \tag{7}$$

Transfer

The transfer operations for the ray can be written quite simply in vector notation as

$$\mathbf{R}_{i+1} = \mathbf{R}_i - t\mathbf{A} + \bar{\lambda}\mathbf{N}$$

where $\mathbf{R}_i$ is the position of the ray on the ith surface, t is the distance between the vertices of the two surfaces, $\mathbf{A}$ is a unit vector in the direction of the optical axis, and $\mathbf{N}$ is a unit vector in the direction of the ray. The distance along the ray from one surface to the other is $\bar{\lambda}$, which can be written in terms of the other parameters (Stavroudis, 1972, Eq. VI-16) as

$$\bar{\lambda} = (1/u)[C_{i+1}(\mathbf{R}_i - t_i\mathbf{A})^2 - 2\mathbf{A}\cdot(\mathbf{R}_i - t_i\mathbf{A})]$$

where

$$u = (\mathbf{A}\cdot\mathbf{N}_i)(1 + C_{i+1}t_i) - C_{i+1}(\mathbf{N}_i\cdot\mathbf{R}_i) + B$$

$$B^2 = 1 - [(\mathbf{A}\times\mathbf{N}_i)(1 + C_{i+1}t_i) + C_{i+1}(\mathbf{N}_i\times\mathbf{R}_i)]^2.$$

To determine how the principal curvatures are changed by a transfer operation, we will examine two successive positions of the wavefront described by the vector function, $\mathbf{W}(u,v,s)$. If the distance between these two positions is τ, we may write

$$\mathbf{W}(u,v,s_o + \tau) = \mathbf{W}(u,v,s_o) + \tau\mathbf{N}(u,v)$$

or

$$\mathbf{W}^* = \mathbf{W} + \tau\mathbf{N}. \tag{8}$$

If at some point on the surface we choose the parameters u and v so that they are in the principal directions, then the fundamental quantities F and M will both be zero at that point. Differentiating Eq. (8) gives us

$$\mathbf{W}_u^* = \mathbf{W}_u + \tau\mathbf{N}_u \tag{9}$$

$$\mathbf{W}_v^* = \mathbf{W}_v + \tau\mathbf{N}_v. \tag{10}$$

The equations of Weingarten (Struik, 1961, p. 108) give us

$$\mathbf{N}_u = [(FM-GL)\mathbf{W}_u + (FL-EM)\mathbf{W}_v] \,/\, (EG-F^2) \tag{11}$$

$$\mathbf{N}_v = [(FN-GM)\mathbf{W}_u + (FM-EN)\mathbf{W}_v] \,/\, (EG-F^2). \tag{12}$$

But since $F = M = 0$, these reduce to

$$\mathbf{N}_u = (L/E)\mathbf{W}_u$$

$$\mathbf{N}_v = (N/G)\mathbf{W}_v.$$

Substituting these into Eqs. (9) and (10) gives us

$$\mathbf{W}_u^* = [1 - \tau(L/E)]\,\mathbf{W}_u$$

$$\mathbf{W}_v^* = [1 - \tau(N/G)]\,\mathbf{W}_v.$$

From this we can calculate the new fundamental quantities:

$$E^* = \mathbf{W}_u^{*2} = [1 - \tau(L/E)]^2 E$$

$$F^* = \mathbf{W}_u^* \cdot \mathbf{W}_v^* = 0$$

$$G^* = \mathbf{W}_v^{*2} = [1 - \tau(N/G)]^2 G$$

$$L^* = \mathbf{N}_u \cdot \mathbf{W}_u^* = [1 - \tau(L/E)]\,L$$

$$M^* = -\mathbf{N}_v \cdot \mathbf{W}_u^* = 0$$

$$N^* = -\mathbf{N}_v \cdot \mathbf{W}_v^* = [1 - \tau(N/G)]\,N.$$

Since $F^* = M^* = 0$, it is clear that the principal directions remain unchanged during transfer.

After transfer, the principal curvature in the ξ direction will be

$$\begin{aligned} 1/\rho_\xi^* &= L^*/E^* \\ &= \frac{L}{E}\,\frac{1}{1-\tau(L/E)} \\ &= 1/(\rho_\xi - \tau). \end{aligned} \tag{13}$$

Similarly,

$$1/\rho_\eta^* = 1/(\rho_\eta - \tau). \tag{14}$$

We conclude that the centers of curvature in the principal direction remain fixed during transfer.

Refraction

Using the notation of Stavroudis (1972), the refraction equation for the ray can be written in vector form as

$$\mathbf{N}' = \mu\mathbf{N} + \alpha\mathbf{N}^{\#} \tag{15}$$

where μ is the ratio of successive refractive indices

$$\mu = n_i/n_{i+1}$$

and where

$$\alpha = \mu \cos i + \cos i'.$$

$\mathbf{N}$ and $\mathbf{N}'$ are unit vectors in the direction of the incident and refracted ray, respectively, and $\mathbf{N}^{\#}$ is the unit normal to the refracting surface. If the refracting surface is a sphere, then

$$\mathbf{N}_i^{\#} = -C_i\mathbf{R}_i + \mathbf{A}.$$

Unfortunately the refraction operation for the wavefront is much more difficult than it is for the ray. The operation can be somewhat simplified, however, by a judicious choice of coordinates such as those described by Stavroudis (1962, p. 188).

In this coordinate system the unit vector $\mathbf{P}$ is defined at the point of incidence to be

$$\begin{aligned} \mathbf{P} &= (\mathbf{N} \times \mathbf{N}^{\#})/[1-(\mathbf{N} \cdot \mathbf{N}^{\#})^2]^{1/2} \\ &= (\mathbf{N} \times \mathbf{N}^{\#})/\sin i. \end{aligned}$$

Clearly **P** is perpendicular to the plane of incidence and is therefore tangent to the refracting surface as well as the incident and refracted wavefront.

Next the unit vector **Q** is defined,

$$\mathbf{Q} = \mathbf{N} \times \mathbf{P}.$$

Clearly **Q** is perpendicular to **P** and is tangent to the incident wavefront.

The vectors **N**, **P**, and **Q** now form a convenient orthogonal coordinate system on the incident wavefront (Fig. 1).

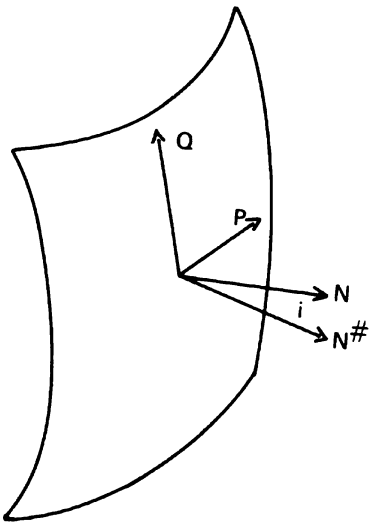


Fig. 1. The N, P, Q coordinate system.

A similar coordinate system can be introduced on the refracting surface and on the refracted wavefront by letting

$$\mathbf{Q}^{\#} = \mathbf{N}^{\#} \times \mathbf{P}$$

and

$$\mathbf{Q}' = \mathbf{N}' \times \mathbf{P}.$$

It should be noted that the three coordinate systems all have the vector **P** in common.

If we let u and v be in the directions of **P** and **Q**, respectively, then

$$\mathbf{P} = \mathbf{W}_u/(\mathbf{W}_u{}^2)^{1/2} = \mathbf{W}_u/E^{1/2}$$

and

$$\mathbf{Q} = \mathbf{W}_v/G^{1/2}.$$

Similarly for the refracting surface

$$\mathbf{P}^{\#} = \mathbf{R}_u/E^{\#1/2}$$

$$\mathbf{Q}^{\#} = \mathbf{R}_v/G^{\#1/2}$$

where we have let $\mathbf{R}(u,v)$ be the vector function describing the refracting surface.

For the refracted wavefront,

$$\mathbf{P}' = \mathbf{W}_u'/E'^{1/2}$$

$$\mathbf{Q}' = \mathbf{W}_v'/G'^{1/2}.$$

However, since

$$\mathbf{P} = \mathbf{P}^{\#} = \mathbf{P}',$$

we may write

$$\mathbf{P} = \mathbf{W}_u/E^{1/2} = \mathbf{R}_u/E^{\#1/2} = \mathbf{W}_u'/E'^{1/2}.$$

Assuming that we have chosen the parameter u correctly for the various surfaces,

$$E = E^{\#} = E'.$$

The importance of this relation will become apparent later.

We now need to differentiate the refraction equation, Eq. (15), with respect to u and v

$$\mathbf{N}_u' = \mu\mathbf{N}_u + \alpha\mathbf{N}_u^{\#} + \alpha_u\mathbf{N}^{\#}.$$

If we take the vector cross product of this with $\mathbf{N}^{\#}$, we have

$$\mathbf{N}_u' \times \mathbf{N}^{\#} = \mu(\mathbf{N}_u \times \mathbf{N}^{\#}) + \alpha(\mathbf{N}_u^{\#} \times \mathbf{N}^{\#}). \tag{16}$$

But from Weingarten's expression for $\mathbf{N}_u$, Eq. (11),

$$\begin{aligned}\mathbf{N}_u \times \mathbf{N}^{\#} &= [(L/E)\mathbf{W}_u + (M/G)\mathbf{W}_v] \times \mathbf{N}^{\#} \\ &= [(L/E^{1/2})\mathbf{P} + (M/G^{1/2})\mathbf{Q}] \times \mathbf{N}^{\#} \\ &= (L/E^{1/2})\mathbf{Q}^{\#} - (M/G^{1/2})(\mathbf{N} \times \mathbf{P}) \times \mathbf{N} \\ &= (L/E^{1/2})\mathbf{Q}^{\#} - (M/G^{1/2})\mathbf{P}\cos i.\end{aligned}$$

Similarly

$$\mathbf{N}_u' \times \mathbf{N}^{\#} = (L'/E'^{1/2})\mathbf{Q}^{\#} - (M'/G'^{1/2})\mathbf{P}\cos i'$$

and

$$\mathbf{N}_u^{\#} \times \mathbf{N}^{\#} = (L^{\#}/E^{\#1/2})\mathbf{Q}^{\#} - (M^{\#}/G^{\#1/2})\mathbf{P}.$$

Substituting these expressions into Eq. (14) and collecting terms gives us

$$\begin{aligned}&[(L'/E'^{1/2}) - \mu(L/E^{1/2}) - \alpha(L^{\#}/E^{\#1/2})]\mathbf{Q}^{\#} \\ &\quad - [(M'/G'^{1/2})\cos i' - \mu(M/G^{1/2})\cos i - \alpha(M^{\#}/G^{\#1/2})]\mathbf{P} = 0.\end{aligned}$$

If we differentiate the refraction equation with respect to v and follow the same steps as above, we will have

$$\begin{aligned}&[(M'/E'^{1/2}) - \mu(M/E^{1/2}) - \alpha(M^{\#}/E^{\#1/2})]\mathbf{Q}^{\#} \\ &\quad - [(N'/G'^{1/2})\cos i' - \mu(N/G^{1/2})\cos i - \alpha(N^{\#}/G^{\#1/2})]\mathbf{P} = 0.\end{aligned}$$

Since $\mathbf{P}$ and $\mathbf{Q}^{\#}$ are orthogonal, we can equate their coefficients to zero, giving us

$$(L'/E'^{1/2}) = \mu(L/E^{1/2})\cos i + \alpha(L^{\#}/E^{\#1/2}) \tag{17}$$

$$(M'/G'^{1/2})\cos i' = \mu(M/G^{1/2})\cos i + \alpha(M^{\#}/G^{\#1/2}) \tag{18}$$

$$(M'/E'^{1/2}) = \mu(M/E^{1/2}) = \alpha(M^{\#}/E^{\#1/2}) \tag{19}$$

$$(N'/G'^{1/2})\cos i' = \mu(N/G^{1/2})\cos i + \alpha(N^{\#}/G^{\#1/2}). \tag{20}$$

Now using the fact that $E = E^{\#} = E'$, Eqs. (17) and (18) become

$$L' = \mu L + \alpha L^{\#} \tag{21}$$

$$M' = \mu M + \alpha M^{\#}. \tag{22}$$

We now multiply Eq. (22) by $\cos i'/G'^{1/2}$ and subtract it from Eq. (18). After collecting terms, the result is

$$\mu M[(\cos i/G^{1/2}) - (\cos i'/G'^{1/2})] + \alpha M^{\#}[(1/G^{1/2}) - (\cos i'/G'^{1/2})] = 0.$$

Since M and $M^{\#}$ are properties of the incident wavefront and the refracting surface, respectively, they must be independent and their coefficients must vanish. This gives us the following relationships between the G's:

$$\cos i/G^{1/2} = \cos i'/G'^{1/2} = 1/G^{\#1/2}. \tag{23}$$

These expressions can be used to simplify Eq. (20), giving us

$$N' = \mu N + \alpha N^{\#}. \tag{24}$$

Using the relationship between the G's and the fact that $E = E^{\#} = E'$, we can rewrite Eqs. (21), (22), and (24) in terms of the principal curvatures associated with the ray. Equation (21) becomes

$$L'/E' = \mu(L/E) + \alpha(L^{\#}/E^{\#})$$

or

$$1/\rho_u' = \mu(1/\rho_u) + \alpha(1/\rho_u^{\#}). \tag{25}$$

Similarly, Eq. (24) becomes

$$1/\rho_v' = (\mu \cos^2 i)/(\rho_v \cos^2 i') + \alpha/(\rho_v^{\#} \cos^2 i'), \tag{26}$$

and finally, recalling that $1/\sigma = M/(EG)^{1/2}$, Eq. (22) becomes

$$1/\sigma' = (\mu \cos i)/(\sigma \cos i') + \alpha/(\sigma^{\#} \cos i').$$

However, in the case of a spherical refracting surface, $M^{\#}$ is zero, making $1/\sigma^{\#}$ also zero. Hence

$$1/\sigma' = (\mu \cos i)/(\sigma \cos i'). \tag{27}$$

The steps of the refraction operation are now obvious. We first use Eqs. (2), (3), and (4) to find the curvatures in the u and v directions. Next, we find the refracted quantities by applying Eqs. (25), (26), and (27), and finally we use Eqs. (5), (6), and (7) to find the new principal curvatures and directions. Reflection can be considered as a special case of refraction by letting $n_{i+1} = -n_i$.

APPLICATION OF GENERALIZED RAY TRACING TO A STUDY OF CAUSTIC SURFACES

The information gained from a generalized ray trace can be used in a number of different ways. For example, a necessary (although not sufficient) condition for a ray to be associated with a perfect image is that the two principal curvatures be equal. Hence, one application of generalized ray tracing would be to trace a number of rays through various locations in the entrance pupil and then adjust optical parameters to minimize the difference between the principal curvatures for each ray.

We shall use a somewhat different approach. As we shall see, generalized ray tracing gives us a convenient method to determine the caustic surface formed by an optical system. Our primary objective, then, is to write a computer program to perform the generalized ray tracing operations and to use this program to study the caustic surfaces formed by different lenses.

A Computer Program for Tracing Rays

To utilize the generalized ray tracing equations, it was necessary to write them as a computer program. The resultant program (see appendix) is written in BASIC to be run on the General Electric Time Sharing Service, MARK II.

Through a fixed object point an initial ray is chosen and the intersection of this ray with the first surface is found. The reciprocal of the distance between the object point and this point of interception is the curvature of the spherical wavefront centered at the object point and passing through the point of interception. This value of curvature is used as both initial principal curvatures in the subsequent calculation. As the ray is then traced through the optical system, the principal curvatures and directions of the wavefront are calculated at each refracting surface. After the ray has passed through the final surface, the positions of the principal centers of curvature relative to the vertex of the final surface are calculated and printed out. The computer then selects the next ray, which lies in the plane determined by the chief ray and the selected marginal ray, and the process is repeated until the specified number of rays in the fan has been traced. In practice the rays in the fan are selected beginning with the chief ray and then working out to the edge of the pupil until the specified number of rays has been traced or until one of the rays is internally reflected or exceeds the clear aperture of the surface.

Once the program was written and debugged, our main concern was with the accuracy of the calculations. One problem encountered with the use of vector notation was the roundoff errors inherent in any computer calculations. These errors caused the magnitude of our unit vectors to be slightly more or less than unity, which in turn led in some instances to taking the square root of a negative number. To avoid this, and at the same time to get an idea of the accuracy of the calculations, a subroutine was included to test the normalization of each unit vector and, if necessary, to renormalize it to within tolerable limits. It should be pointed out that, for all the data accumulated thus far, the vector normalization has been accurate to at least seven places.

As a final check on the accuracy of the program a number of rays were traced through a four-element aerial camera lens (Stavroudis and Sutton, 1965, p. 59). This generalized ray trace agreed to within 1% with data from a conventional meridional ray trace (Fig. 2).

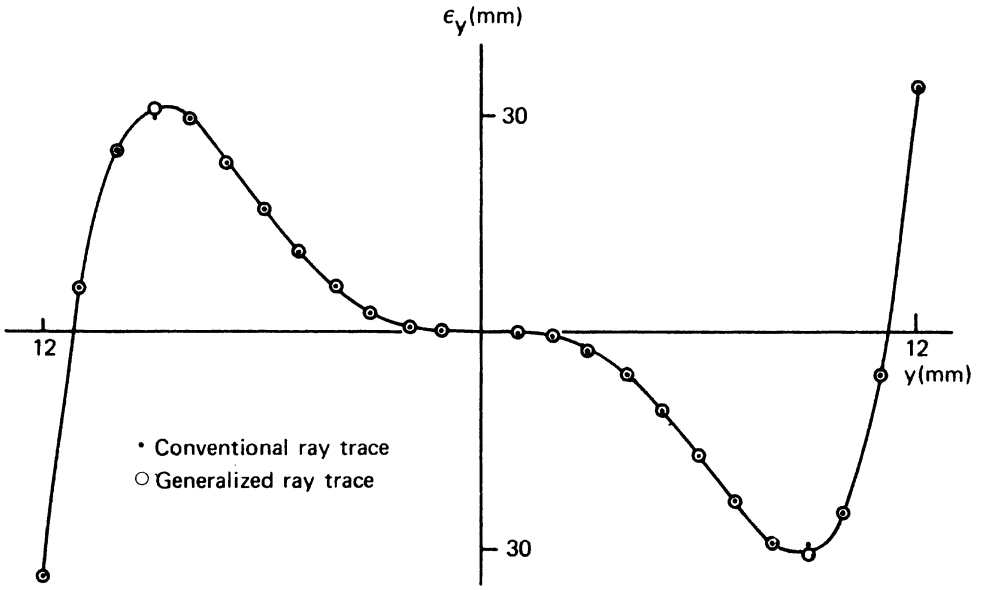


Fig. 2. Comparison of meridional ray fan plots from conventional and generalized ray tracing methods. Object is on axis at infinity.

Caustic Surfaces

The caustic surface is defined to be the envelope of a two-parameter orthotomic system of rays. Associated with this system of rays is a family of wavefronts **W**, which we have written

$$\mathbf{W}(\xi,\eta,s) = \mathbf{W}(\xi,\eta) + s\mathbf{N}(\xi,\eta)$$

where ξ and η are assumed to be in the principal directions. Since each ray is normal to the wavefront and tangent to the caustic, the caustic surface is the evolute of the wavefront (Eisenhart, 1909, pp. 179-180).

If we let $\delta(\xi,\eta)$ be the distance along the ray from its starting point on the wavefront to its point of tangency on the caustic surface, then the caustic surface can be described by the vector function

$$\mathbf{C}(\xi,\eta) = \mathbf{W}(\xi,\eta) + \delta(\xi,\eta)\,\mathbf{N}(\xi,\eta).$$

The tangent vectors to the caustic surface will be

$$\mathbf{C}_\xi = \mathbf{W}_\xi + \delta\mathbf{N}_\xi + (\partial\delta/\partial\xi)\mathbf{N}$$

$$\mathbf{C}_\eta = \mathbf{W}_\eta + \delta\mathbf{N}_\eta + (\partial\delta/\partial\eta)\mathbf{N}.$$

Using Eqs. (11) and (12), these become

$$\mathbf{C}_\xi = (1 - \delta/\rho_\xi)\mathbf{W}_\xi + (\partial\delta/\partial\xi)\mathbf{N} \tag{28}$$

$$\mathbf{C}_\eta = (1 - \delta/\rho_\eta)\mathbf{W}_\eta + (\partial\delta/\partial\eta)\mathbf{N}. \tag{29}$$

The first fundamental quantities of the caustic surface can now be calculated

$$\hat{E} = (1 - \delta/\rho_\xi)^2 E + (\partial\delta/\partial\xi)^2$$

$$\hat{F} = (\partial\delta/\partial\xi)(\partial\delta/\partial\eta)$$

$$\hat{G} = (1 - \delta/\rho_\eta)^2 G + (\partial\delta/\partial\eta)^2.$$

The normal to the caustic surface is

$$\begin{aligned}\mathbf{C}_\xi \times \mathbf{C}_\eta = {} & (1 - \delta/\rho_\xi)(1 - \delta/\rho_\eta)\mathbf{W}_\xi \times \mathbf{W}_\eta \\ & + (\partial\delta/\partial\eta)(1 - \delta/\rho_\xi)\mathbf{W}_\xi \times \mathbf{N} \\ & - (\partial\delta/\partial\xi)(1 - \delta/\rho_\eta)\mathbf{W}_\eta \times \mathbf{N}.\end{aligned}$$

Using the fact that $\mathbf{N} = (\mathbf{W}_\xi \times \mathbf{W}_\eta)/(EG)^{1/2}$, the above expression may be written

$$\begin{aligned}\mathbf{C}_\xi \times \mathbf{C}_\eta = {} & (1 - \delta/\rho_\xi)(1 - \delta/\rho_\eta)(EG)^{1/2}\mathbf{N} \\ & - (\partial\delta/\partial\eta)(1 - \delta/\rho_\xi)[E/(EG)^{1/2}]\mathbf{W}_\eta \\ & - (\partial\delta/\partial\xi)(1 - \delta/\rho_\eta)[G/(EG)^{1/2}]\mathbf{W}_\xi. \end{aligned} \tag{30}$$

However, since this is normal to the caustic, it must be perpendicular to the ray, and therefore the coefficient of $\mathbf{N}$ must be zero. Hence

$$(1 - \delta/\rho_\xi)(1 - \delta/\rho_\eta) = 0.$$

The two solutions to this are

$$\delta_1 = \rho_\xi \quad \text{and} \quad \delta_2 = \rho_\eta.$$

Clearly the caustic surface is the locus of the principal centers of curvature. Since the principal curvatures are generally not equal, we can treat the caustic as a two-sheeted surface with each sheet corresponding to the locus of the centers of curvatures in one of the principal directions. Thus, the equations for the two sheets of the caustic surface may be written

$$\mathbf{C}_1 = \mathbf{W} + \rho_\xi\mathbf{N} \quad \text{and} \quad \mathbf{C}_2 = \mathbf{W} + \rho_\eta\mathbf{N}.$$

Another important property of the caustic is that the normal to the caustic surface at any point lies in one of the principal directions associated with the ray at that point. To see this we must first use Eqs. (28) and (29) to write the tangent vectors to each sheet of the caustic

$$\mathbf{C}_{1\xi} = (\partial\rho_\xi/\partial\xi)\mathbf{N}$$

$$\mathbf{C}_{1\eta} = (1 - \rho_\xi/\rho_\eta)\mathbf{W}_\eta + (\partial\rho_\xi/\partial\eta)\mathbf{N}$$

$$\mathbf{C}_{2\xi} = (1 - \rho_\eta/\rho_\xi)\mathbf{W}_\xi + (\partial\rho_\eta/\partial\xi)\mathbf{N}$$

$$\mathbf{C}_{2\eta} = (\partial\rho_\eta/\partial\eta)\mathbf{N}.$$

Next, the normal vector to each sheet is found by taking the appropriate cross product

$$\begin{aligned}\mathbf{C}_{1\xi} \times \mathbf{C}_{1\eta} &= (\partial\rho_\xi/\partial\xi)(1 - \rho_\xi/\rho_\eta)\mathbf{N} \times \mathbf{W}_\eta \\ &= (-\partial\rho_\xi/\partial\xi)(1 - \rho_\xi/\rho_\eta)\mathbf{W}_\eta \times [\mathbf{W}_\xi \times \mathbf{W}_\eta](EG)^{-1/2} \\ &= -G(\partial\rho_\xi/\partial\xi)(1 - \rho_\xi/\rho_\eta)\mathbf{W}_\xi/(EG)^{1/2}.\end{aligned}$$

This can be normalized by dividing by

$$(\tilde{E}\tilde{G}-\tilde{F}^2)^{1/2} = G^{1/2}(\partial\rho_\xi/\partial\xi)(1-\rho_\xi/\rho_\eta).$$

Hence, the unit normal to the first sheet of the caustic surface is

$$\mathbf{N}_1 = \mathbf{W}_\xi/E^{1/2}. \tag{31}$$

A similar expression is written for the normal to the second sheet of the caustic,

$$\mathbf{N}_2 = \mathbf{W}_\eta/G^{1/2}. \tag{32}$$

We conclude that the normals to the caustic surface do indeed lie in one of the principal directions associated with the ray.

Our primary objectives in studying caustic surfaces were to become familiar with some of their properties and, if possible, to find some correlations between the structure of the caustic surface and image quality. Obviously, if the image formed were perfect, the caustic surface would degenerate into a single point. Hence, one criterion of the image quality is the form of the caustic surface relative to that of a perfect image.

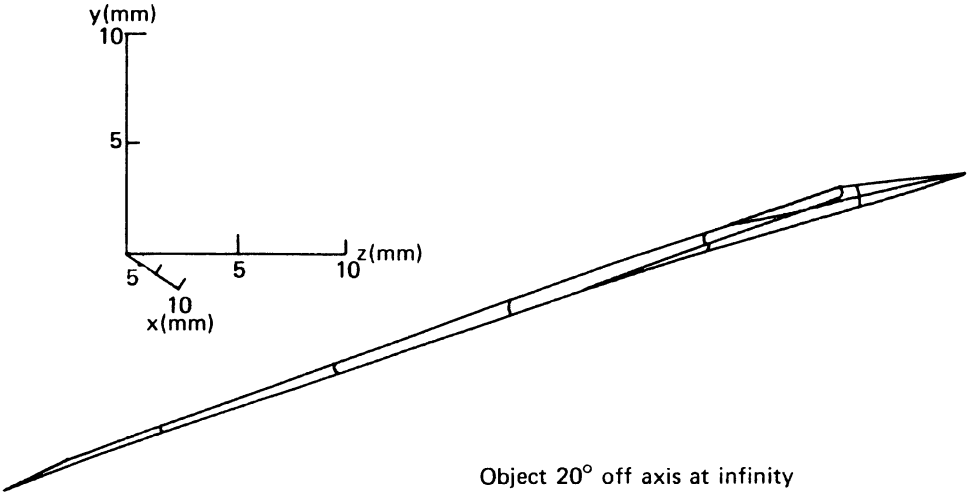
Because we have limited ourselves to considering rotationally symmetric optical systems, the exiting wavefront and also the caustic will be symmetric about the meridional plane. If, in addition, the wavefront is rotationally symmetric, such as when the object is on axis, then one sheet of the caustic will degenerate into a segment of the axis of revolution and the other will form a surface of revolution about the axis (Born and Wolf, 1965, p. 170).

A Plano-Convex Singlet

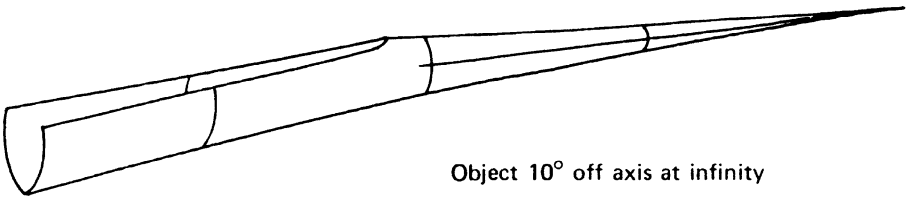
The first lens that we studied was a plano-convex singlet with a diameter of 50 mm and a thickness of 5 mm. The radius of curvature of the convex surface was 100 mm and the index of refraction was 1.5168. In all cases we will let the object be a point at infinity and will choose ξ and η so that, for a meridional ray, η will represent the curvature in the meridional plane and ξ will represent the curvature perpendicular to the meridional plane.

We will first examine the caustic surfaces formed when the plane side of the lens is toward the object (Fig. 3). When the object is on axis, the first, or ξ, sheet of the caustic surface forms a spike along the optical axis while the second or η sheet forms a surface of revolution about the optical axis. The cusp is the paraxial image point and is located 193.5 mm from the vertex of the second surface.

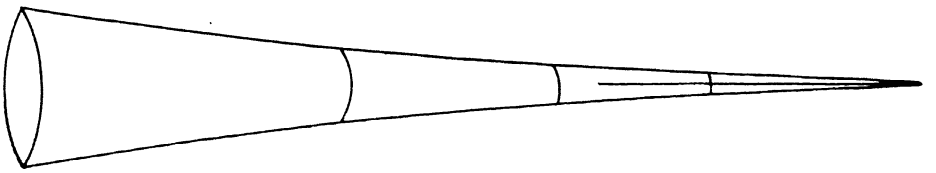
Although we shall not attempt an extensive correlation between the form of the caustic surface and conventional aberration analysis, it is significant to note that the length of the central spike corresponds to the sum of the different orders of longitudinal spherical aberration. For this lens, the length of the central spike is 14.3 mm, which is the same value we obtained from meridional ray trace data.



Object 20° off axis at infinity



Object 10° off axis at infinity



Object on axis at infinity

Fig. 3. Caustic surface formed by plano-convex singlet with plane side toward object.

When the object is 10° off axis, the most significant feature of the caustic is that the first sheet is still a spike and, although the second sheet is partially vignetted, it is obviously a surface of revolution with the spike as the axis of symmetry. The reason for this symmetry is clear. Because the object is at infinity, the incident wavefront is a plane surface. After refraction by the first surface, the wavefront is still a plane surface. At some point on the second surface the normal to the wavefront will coincide with the normal to the spherical refracting surface, and the refracted wavefront will be rotationally symmetric about the ray passing through this point.

If we choose the ray that is perpendicular to the second surface to be the chief ray, then, except for vignetting, the caustic will be symmetric about the chief ray. Choosing this new chief ray corresponds to shifting the stop to a position 62 mm in front of the first surface. Interestingly, this is precisely the position of the stop necessary for the elimination of third order coma as calculated by the stop shift equations associated with third order aberration theory (Hopkins and Hanau, 1962, pp. 8, 16, 18). An additional advantage of choosing this ray to be the chief ray is that the principal centers of curvature coincide, and hence for this ray there will be no astigmatism. If we choose any other ray to be the chief ray, the principal curvatures will not be equal and the principal center of curvature in the ξ direction will correspond to the sagittal focus while the principal center of curvature in the η direction will correspond to the tangential focus.

When the object is 20° off axis, the form of the caustic is similar to the two previous cases, except that now the vignetting of the η sheet is much greater.

When the lens is turned around so that the convex side is toward the object (Fig. 4), the shape of the caustic formed with the object on axis is similar to that in

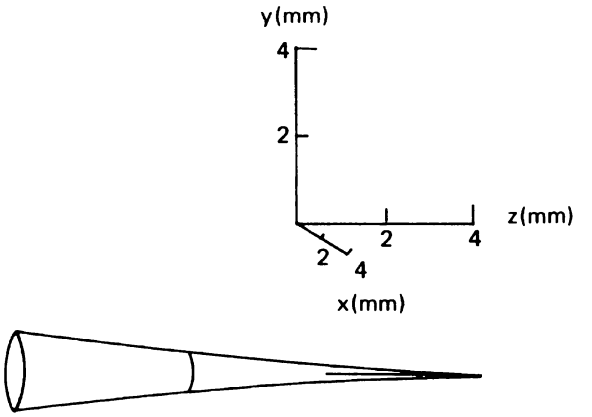


Fig. 4. Caustic surface formed by plano-convex singlet with convex side toward object.

Fig. 3. Again, it is a central spike surrounded by a surface of revolution, with a cusp at the paraxial image point located 190 mm from the second surface. The major difference is that the caustic now formed is much more compact. Its central spike is only 3.6 mm long compared to 14.3 mm in the previous case.

When the object is moved off axis in this case, all semblance of rotational symmetry is lost and the caustic surface becomes so highly deformed that it is difficult to describe and almost impossible to draw. Because of this we have resorted to representing the off-axis caustic surface by a series of meridional sections (Fig. 5) and sagittal projections (Fig. 6). The meridional sections are formed by tracing a meridional fan of rays through the entrance pupil. The section of the caustic surface thus formed will lie entirely in the meridional plane. The sagittal projections are made by tracing a skew fan of rays through the entrance pupil and projecting the principal centers of curvatures onto the x,z plane. This projection does not represent a plane section of the caustic surface because the principal centers of curvature formed in this manner do not generally lie in the same plane.

Examining first the meridional sections of the caustic surface in Fig. 5, we note that the ξ sheet separates from the η sheet and becomes more rounded as the object is moved off axis. The η sheet becomes somewhat stretched and deformed but still has a cusp.

In the sagittal projections (Fig. 6), the situation is reversed. The ξ sheet forms a cusp and the tip of the η sheet becomes indented and almost parabolic as the object is moved off axis.

Once the object is moved off axis, it is obvious that no ray has equal principal curvatures. This means that the astigmatism cannot be eliminated by merely shifting the stop to define a new chief ray as was done in the previous case. This does not mean, however, that a stop shift would not be advantageous. The most likely choice for the chief ray would be the ray that passes through the cusp of the η sheets in Fig. 5. This corresponds to locating the stop 7 mm behind the second surface of the lens. This is in contrast to the position of 20 mm predicted by stop shift equations for the elimination of third order coma.

Figure 7 is a drawing of the caustic surface formed when the object is 20° off axis and illustrates the complexity of the surface. For additional clarity the η and ξ sheets are presented separately in Fig. 8. The η sheet forms a surface that intersects itself near the center of the figure. Hence, if we were to locate the image plane near this point, the image formed would be essentially a straight line perpendicular to the meridional plane and could therefore be thought of as the tangential focus of an astigmatic image. Similarly, if we were to locate the image plane near the far end of the ξ sheet, the image formed would be essentially a straight line in the meridional plane and could be thought of as the sagittal focus of an astigmatic image. This is, of course, an oversimplification, but it does give an idea of the predominance of astigmatism in the caustic surface at this rather large field angle.

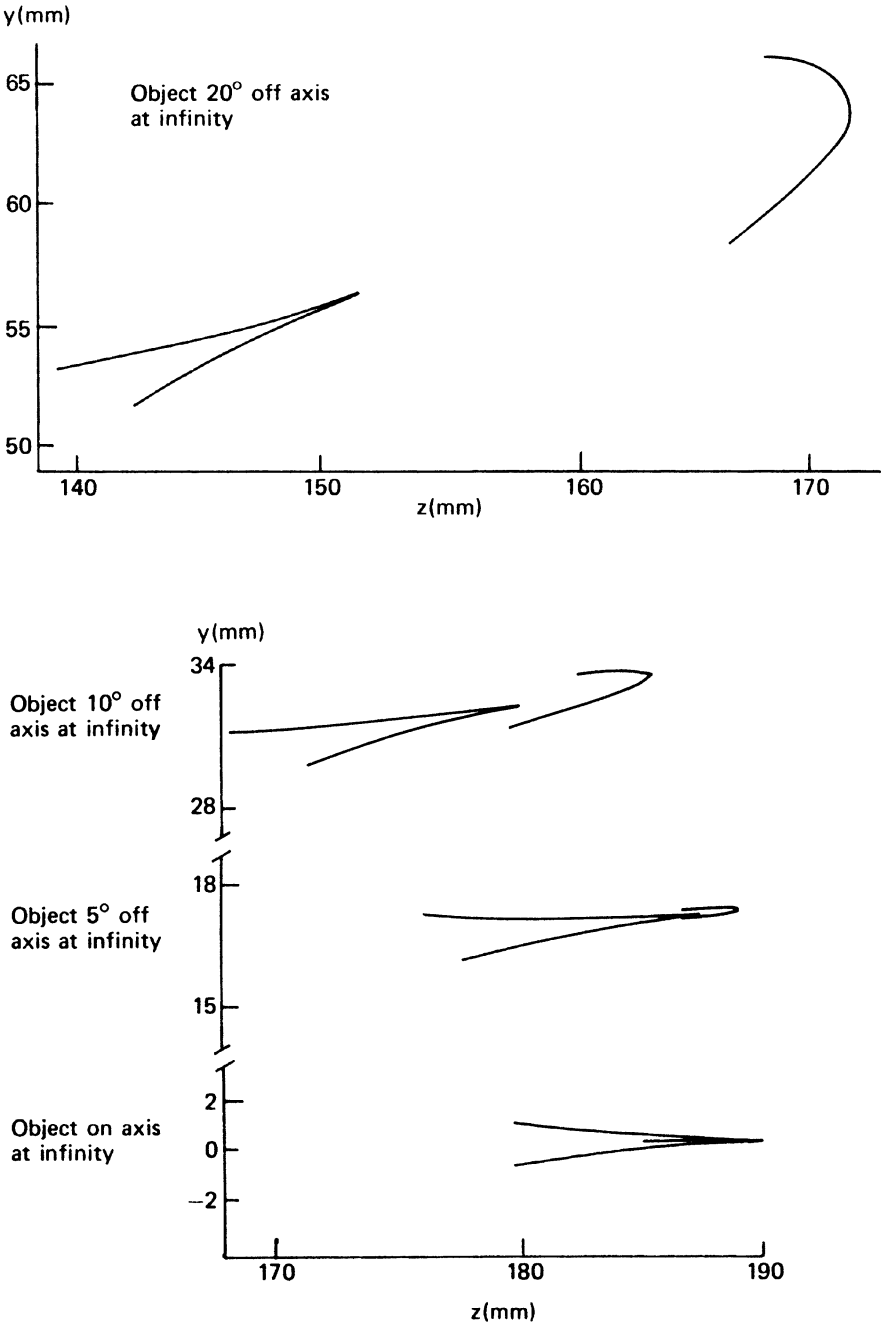


Fig. 5. Meridional sections of caustic surface formed by plano-convex singlet with convex side toward object.

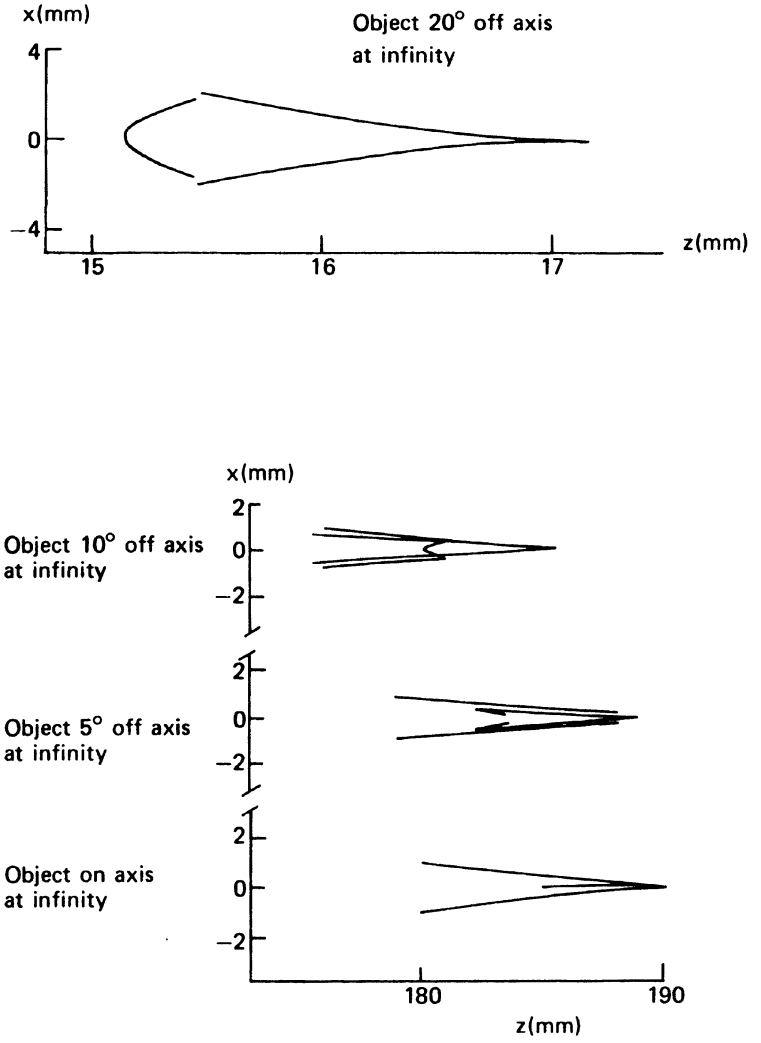


Fig. 6. Sagittal projections of caustic surface formed by plano-convex singlet with convex side toward object.

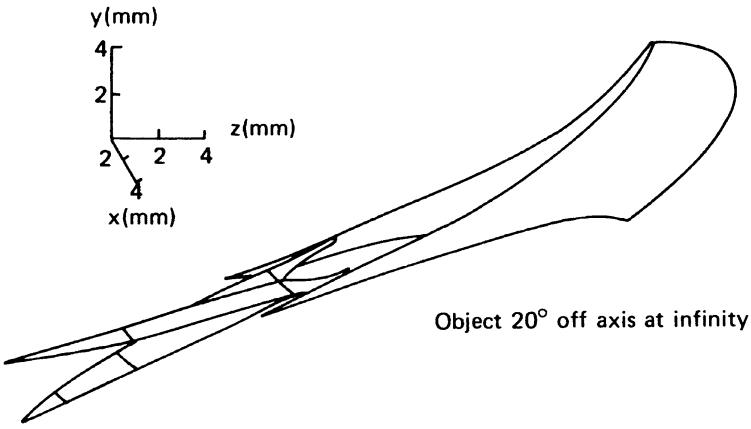


Fig. 7. Caustic surface formed by plano-convex singlet with convex side toward object.

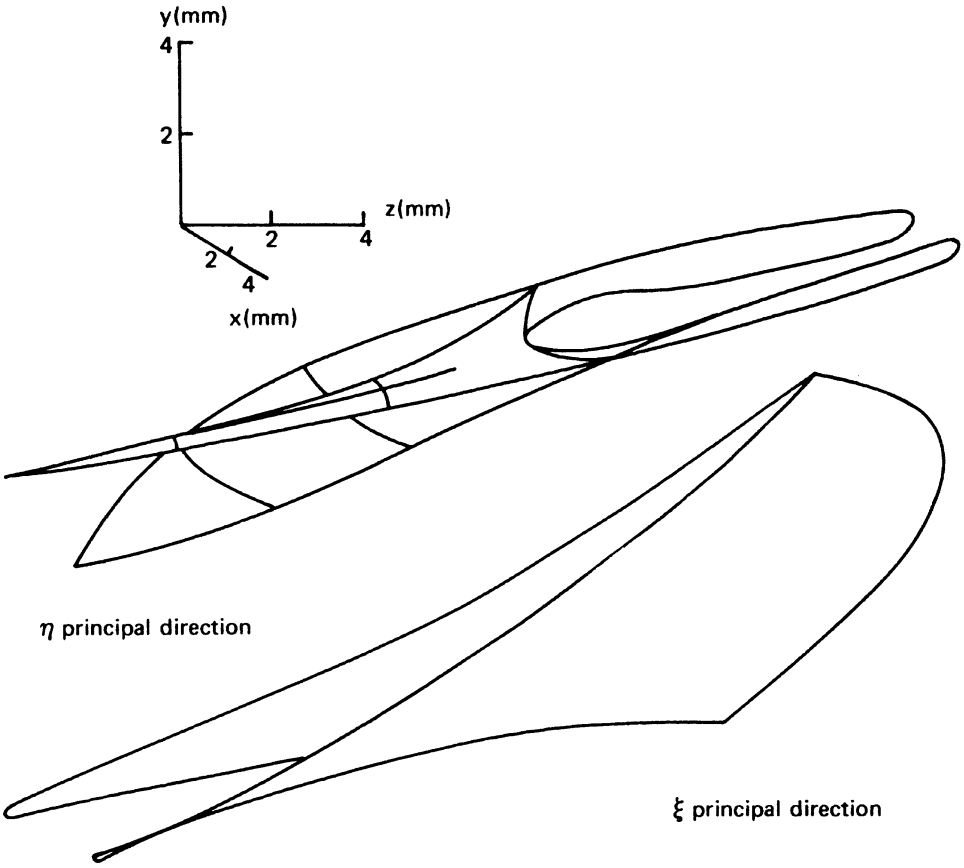


Fig. 8. Sheets associated with the η and ξ principal directions from the caustic surface of Fig. 6, top (20° off axis at infinity).

A Well Corrected Aerial Camera Lens

To get an idea of the shape of the caustic formed by a reasonably well corrected lens, we studied a four-element aerial camera lens designated by Stavroudis and Sutton (1965, p. 59) as No. ME-10. This is an f/6.3 lens with a focal length of 15 cm.

Figure 9 is the caustic surface formed by this lens with the object on axis at infinity. As with the previously discussed caustic surfaces formed by an object on axis, the ξ sheet has degenerated into a spike along the optical axis while the η sheet forms a surface of revolution about the optical axis. However, as Fig. 10 indicates, in this case the η sheet intersects itself.

It is somewhat surprising to note that the length of the η sheet is a rather long 10 mm while the central spike is only 0.87 mm long. It is interesting to compare the behavior of the central spike with the meridional ray trace of Fig. 2. The spike reaches its maximum negative value for the ray with an entrance pupil height between 8 and 9 mm. The spike then retraces itself, passing through the paraxial image point at $y = 11.1$ mm, and reaches its maximum positive value at $y = 12$ mm.

Another indication of the quality of the image formed with this lens is the relatively small diameter of the η sheet of the caustic surface. The diameter is greatest (1.2 mm) at the far end of the surface and smallest (0.012 mm) where it intersects itself.

It is interesting to note that the over-all length of the η sheet can be reduced to 1.3 mm by stopping the lens down to about f/8.4 as in Fig. 10. In this case the length of the central spike is reduced to 0.44 mm.

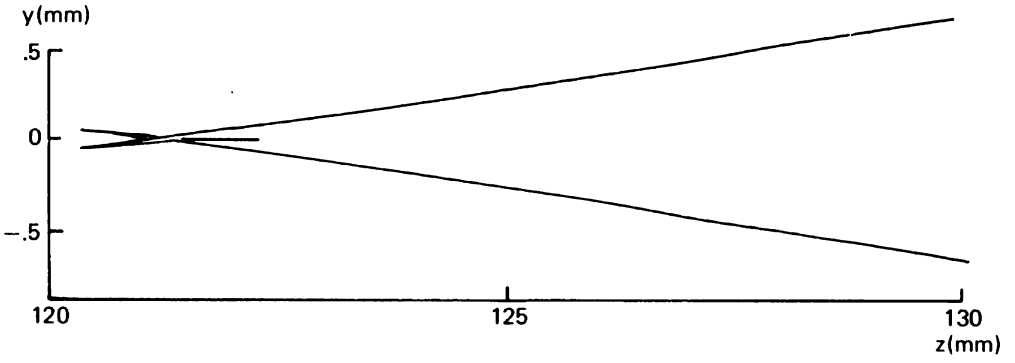


Fig. 9. Caustic surface formed by aerial camera lens with object on axis at infinity.

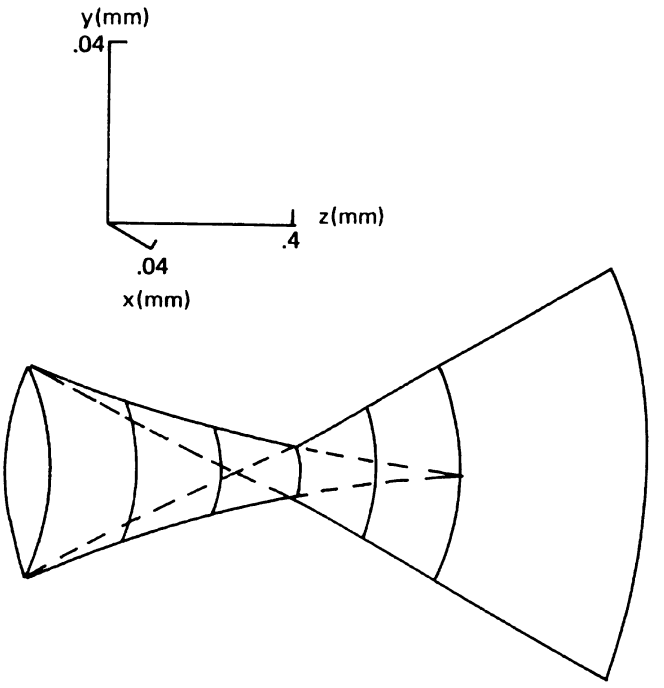


Fig. 10. Expanded view of caustic surface formed by aerial camera lens with object on axis at infinity.

When the object is moved 20° off axis, the caustic surface loses its rotational symmetry and becomes extremely complicated. A meridional section of this caustic surface (Fig. 11) shows that, although one segment of the η sheet (formed by the upper half of the entrance pupil) becomes very long, the rest of the caustic remains fairly compact, with a maximum cross section of about 0.1 mm.

A sagittal projection (Fig. 12) shows how the surface folds back and forth upon itself, with a maximum width of about 2 mm. The principal centers of curvature of the chief ray are separated by about 0.15 mm, and the two sheets of the caustic surface intersect each other about midway between these two centers of curvature and about 0.001 mm off axis.

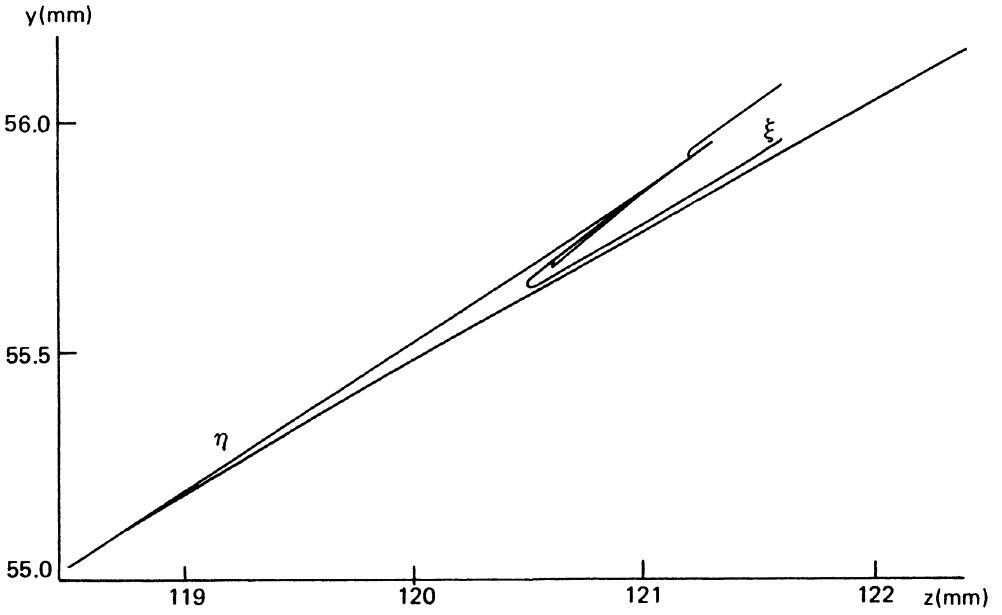


Fig. 11. Meridional section of caustic surface formed by aerial camera lens with object 20° off axis at infinity.

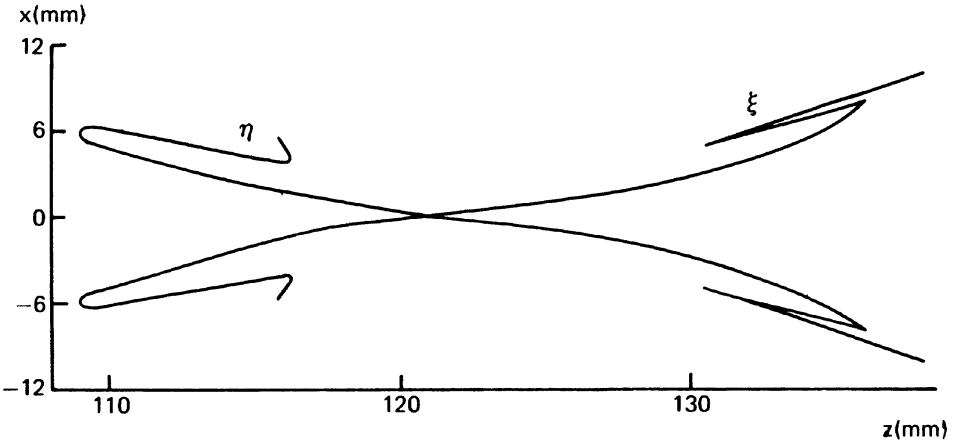


Fig. 12. Sagittal projection of caustic surface formed by aerial camera lens with object 20° off axis at infinity.

CONCLUSIONS

It is readily apparent that the method of generalized ray tracing as outlined in this paper does provide us with additional information about the image that cannot be obtained directly in any other manner. For example, a conventional ray trace gives us the position and direction cosines of each ray traced through the system. A generalized ray trace gives us, in addition to this information, the principal curvatures and directions of the wavefront associated with each ray.

Although we have restricted the treatment here to spherical refracting surfaces and homogeneous media, the method can be extended to any medium or surface where the ray tracing equations exist.

The question that remains to be answered is whether or not generalized ray tracing can be incorporated into an effective method of lens design. Before this question can be answered we must first decide what qualities of a caustic surface make a good image.

Clearly, one criterion of a good image is the relative size of the caustic surface. The magnitude of the central spike on axis appears to be especially important. Also important is the diameter of the outer sheet. Off axis, the size of the caustic is still important, but also important is the relation between the two sheets of the caustic surface.

Once criteria for a good image are decided upon, they must be programmed into some kind of minimization routine in a lens design program. As mentioned before, one approach to the lens design problem would be to select a few rays from various positions in the entrance pupil and then adjust parameters so that the difference between the principal curvatures of each ray is minimized.

It is interesting to compare generalized ray tracing to the procedure known as differential ray tracing (Hopkins and Hanau, 1962, pp. 5, 27, 31). Differential ray tracing involves tracing a ray through an optical system and then using differential transfer and refraction equations to trace additional rays that differ from the first by small first-order changes in the ray position and the direction cosines. In practice, several neighboring rays are traced to determine their intersection points.

One of the main difficulties in this approach is that there is no guarantee that two neighboring rays will intersect unless they are located in a principal section of the wavefront. Since there is no way to determine the principal directions, it is necessary to trace several rays to find the intersection points that correspond to the previously discussed principal centers of curvature.

Clearly, generalized ray tracing is a more direct approach since the principal curvatures and directions are determined uniquely for each ray traced through the system.

For a meridional ray, one of the principal directions is in the meridional plane and the other is perpendicular to it. In this case, we can use the Coddington equations to trace a neighboring ray in the meridional plane and another neighboring ray in a plane perpendicular to the meridional plane; then the neighboring rays are displaced along the principal directions and their intersections with the original ray will correspond to the principal centers of curvatures calculated from a general ray trace.

Appendix. GENERALIZED RAY TRACING PROGRAM

```
1  DIM H(3)
2  DIM A(3),B(3),E(3),F(3),G(3),M(3),N(3),P(3),R(3),T(3)
3  PRINT "PUPIL PØSITIØN",,"CAUSTIC PØSITIØN"
4  LET A(3)=1
5  READ J2,J4
6  PRINT,"XI",,"ETA"
7  LET J5=1
10 LET N1=1
15 FØR J3=J6 TØ J4-1 STEP J5
20 MAT READ B,R
22 READ D,C,N2,H1
25 IF J4 = 1 GØTØ 40
30 MAT R=(J5*J3/(J4-1))*R
37 MAT H=R
40 MAT F=(D)*A
45 MAT F =F-B
47 IF D<1E10 GØTØ 53
48 IF B(1)*B(1)+B(2)*B(2)<>0 GØTØ 55
49 MAT N=ZER
50 LET N(3)=1
51 LET L=D
52 GØTØ 75
53 MAT F=F+R
55 MAT G=F
60 GØSUB 1880
65 LET L=SQR(I)
66 MAT N = (1/L)*F
67 LET S$="N"
68 MAT G=N
69 GØSUB 1600
70 MAT N=G
75 IF L<1E10 GØTØ 95
80 LET K1=0
85 LET K2=0
90 GØTØ 105
95 LET K1=-1/L
100 LET K2=K1
105 FØR J1=1 TØ J2
```

```
106 IF J1=1 GØTØ 380
112 READ D,C,N2,H1
115 MAT F=R
120 MAT G=N
125 GØSUB 1850
130 MAT E = (-C)*B
140 MAT F = A
150 GØSUB 1850
160 MATB = (1+C*D)*B
170 MAT F = B+E
180 MAT G = F
190 GØSUB 1880
195 IF I>1 GØTØ 1050
200 LET R9 = SQR(1-I)
210 MAT F = R
220 MAT G = N
230 GØSUB 1880
240 LET U = C*I
250 MAT F = A
260 GØSUB 1880
270 LET U = I*(1+C*D)-U+R9
280 MAT E = (D)*A
290 MAT G = R-E
300 GØSUB 1880
310 LET L = 2*I
320 MAT F = G
325 GØSUB 1880
330 LET L = (C*I-L)/U
340 MAT B = (L)*N
350 MAT R = F+B
351 IF SQR(R(1)*R(1)+R(2)*R(2))>H1 GØTØ 1050
360 LET K1=K1/(1-L*K1)
370 LET K2=K2/(1-L*K2)
380 MAT M=(C)*R
390 MAT M=A-M
410 MAT G = M
415 MAT F=N
420 GØSUB 1880
430 LET C3 = I
434 IF I<0 GØTØ 1050
440 GØSUB 1850
447 LET S3=SQR(1-I*I)*SGN(N(2)-M(2)+1E-10*M(1)-1E-10*N(1))
448 IF S3<>0 GØTØ452
449 MAT P=ZER
450 LET P(1)=1
451 GØTØ460
452 MAT P=(1/S3)*B
453 MAT F=P
454 MAT G=P
455 GØSUB 1880
456 MAT P=(1/SQR(I))*P
460 MAT G=P
461 LET S$="P"
```

```
462 GØSUB 1600
463 MAT P=G
470 MAT F=T
480 GØSUB 1880
490 LET C4 = I
500 LET S4=SQR(1-I*I)*SGN(T(2)-P(2)+1E-10*P(1)-1E-10*T(1))
510 LET C1=I*I*K1+S4*S4*K2
520 LET C2=S4*S4*K1+I*I*K2
530 LET S=(K1-K2)*S4*I
600 LET U9 = N1/N2
605 IF U9=1 GØTØ 830
610 LET S3 = U9*S3
613 IF S3<-1 GØTØ 1070
614 IF S3>1 GØTØ 1070
620 LET C5 = SQR(1-S3*S3)
630 LET Y9 = C5-U9*C3
640 MAT M = (Y9)*M
650 MAT N = (U9)*N
660 MAT N = N+M
670 MAT G=N
671 LET S$="N'"
672 GØSUB 1600
673 MAT N=G
674 MAT F=N
680 MAT G = P
685 GØSUB 1850
686 LET S$="Q'"
687 MAT G=B
688 GØSUB 1600
690 MAT B=G
700 LET C1 = U9*C1+Y9*C
710 LET C2 = (U9*C3*C3*C2+Y9*C)/(C5*C5)
720 LET S = U9*C3*S/C5
727 IF C2-C1<>0 GØTØ 730
728 LET S4 = 0
729 GØTØ 740
730 LET S4 = ATN(2*S/(C2-C1))/2
740 LET S9 = SIN(S4*2)*S
750 LET C4 = CØS(S4)
760 LET S4 = SIN(S4)
770 LET K1=C1*C4*C4+C2*S4*S4+S9
780 LET K2=C1*S4*S4+C2*C4*C4-S9
790 MAT F = (C4)*P
800 MAT G = (S4)*B
805 MAT T=F+G
806 LET S$="T'"
807 MAT G=T
808 GØSUB 1600
810 MAT T=G
830 LET N1=N2
850 NEXT J1
860 MAT F=(1/K1)*N
870 MAT F=R+F
```

```
880 MAT G=(1/K2)*N
890 MAT G=R+G
894 PRINT
896 MAT PRINT H;
910 PRINT TAB(5);F(1);F(2);F(3);TAB(40);G(1);G(2);G(3)
950 RESTØRE
955 READ X,Y
960 NEXT J3
1000 IF J5=1 GØTØ1110
1010 GØTØ 1840
1050 PRINT "EXCEEDED CLEAR APERTURE ØN SURFACE #";J1
1060 GØTØ 1080
1070 PRINT "INTERNAL REFLECTIØN"
1080 IF J5=-1 GØTØ 1840
1090 RESTØRE
1100 READ X,Y
1110 MAT READ F,G
1120 GØSUB 1880
1130 IF I=0 GØTØ 1840
1140 RESTØRE
1150 READ X,Y
1160 LET J4=2-J4
1170 LET J5=-1
1180 LET J6=-1
1190 GØTØ 10
1590 STØP
1595 REM   TEST FØR ERRØR IN VECTØR NØRMALIZATIØN
1600 MAT F=G
1610 GØSUB 1880
1620 IF I<=1 GØTØ 1710
1625 PRINT
1630 PRINT "ERRØR IN VECTØR ";S$;" =";I-1
1640 FØR K5=1TØ3
1650 LET G(K5)=INT(G(K5)*1E7+5)*1E-7
1660 NEXT K5
1670 MAT F=G
1680 GØSUB 1880
1690 PRINT "RESIDUAL =";I-1
1700 GØTØ 1800
1710 IF 1-I<1E-6 GØTØ 1800
1715 PRINT
1720 PRINT "ERRØR IN VECTØR ";S$;" =";I-1
1725 PRINT
1800 RETURN
1840 STØP
1850 LET B(1) = F(2)*G(3)-F(3)*G(2)
1855 LET B(2) = -F(1)*G(3)+F(3)*G(1)
1860 LET B(3) = F(1)*G(2)-F(2)*G(1)
1865 RETURN
1880 LET I = F(1)*G(1)+F(2)*G(2)+F(3)*G(3)
1885 RETURN
2000 END
```

ACKNOWLEDGMENTS

The author wishes to acknowledge the encouragement and generous assistance of his thesis advisor, Dr. Orestes N. Stavroudis, during the preparation of this thesis.

This work was supported in part by Project THEMIS, administered by the Air Force Office of Scientific Research.

REFERENCES

Born, Max, and Emil Wolf, 1965, "Geometrical theory of optical imaging," Chap.4 in *Principles of Optics* (ed. 3, revised), London, Pergamon Press.

Eisenhart, L. P., 1909, "Fundamental equations," Chap. 5 in *A Treatise on the Differential Geometry of Curves and Surfaces,* New York, Dover, 474 pp.

Gullstrand, 1906, *cited in* Kneisly, 1964, p. 229.

Hopkins, R. E., and Richard Hanau, 1962, secs. 5 and 8, *MIL-HDBK-141,* Washington, D.C., Defense Supply Agency.

Kneisly, J. A. II, 1964, "Local curvature of wavefronts in an optical system," J. Opt. Soc. Am. 54(2):229-235.

Stavroudis, O. N., 1962, "Ray-tracing formulas for uniaxial crystals," J. Opt. Soc. Am. 52(2):187-191.

Stavroudis, O. N., 1972 (to be published), *Wavefronts, Rays, and Caustics,* New York, Academic Press.

Stavroudis, O. N., and L. E. Sutton, 1965, "Spot diagrams for the prediction of lens performance from design data," National Bureau of Standards Monograph 93, Washington, D.C., U.S. Government Printing Office, 96 pp.

Struik, D. J., 1961, *Lectures on Classical Differential Geometry,* ed. 2, Reading, Mass., Addison-Wesley, 232 pp.