STUDIES IN OPTICS

II. Analysis of a Given System with the Help of the Characteristic Function, Using the Direct Method of Analysis*

BY

M. HERZBERGER

Communication No. 961 from the Kodak Research Laboratories

In previous papers[1,2,3] the author has developed a direct method of geometrical optics better adapted to the actual procedures of optical calculation than previous methods. In this paper we shall introduce in this direct theory Hamilton's characteristic function, and with it we shall find the image of an arbitrary surface in a given optical system with rotational symmetry. The author apologizes for changing his notations again. In the previous papers a ray was specified by the coordinates (x, y) of its intersection point with the object plane and by the optical direction cosines ξ, η, $\zeta = \sqrt{n^2 - \xi^2 - \eta^2}$, i.e., the direction cosines multiplied by the respective refractive indices. In this paper, x, y represent the optical lengths (i.e., lengths multiplied by the refractive index) of the coordinates of the intersection point with the coordinate plane, while ξ, η, $\zeta = \sqrt{1 - \xi^2 - \eta^2}$ are the direction cosines. The reader may prove for himself that the fundamental formulas given in the former papers remain unaffected by this change in coordinates.

1. **Hamilton's characteristic function and the direct method.** In the previous papers we introduced the vector $\mathbf{b}$ with coordinates x, y, the vector $\mathbf{t}$ with coordinates ξ, η, and analogously the vectors $\mathbf{b}'(x', y')$ and $\mathbf{t}'(\xi', \eta')$ for the image ray, x', y' being the optical lengths of the coordinates of the intersection point of the image ray with the coordinate plane $z' = 0$, and ξ', η' the direction cosines of this ray.

We found the equation

$$\mathbf{b}' = \alpha\mathbf{b} + \beta\mathbf{t}, \qquad \mathbf{t}' = \gamma\mathbf{b} + \delta\mathbf{t}, \tag{1}$$

where α, β, γ and δ depend only on the symmetric functions

$$\mathbf{b}^2 = u, \qquad \mathbf{bt} = v, \qquad \mathbf{t}^2 = w. \tag{2}$$

They are connected by one finite equation,

$$\alpha\delta - \beta\gamma = 1, \tag{3}$$

and three differential equations.

We can find a first integral of these three differential equations by using Hamilton's characteristic function. Hamilton has shown that there exists a function V which

is a characteristic of the optical path (sum of optical distances) between the object and image planes. If we know this function V as function of x, y, x', y', we find that

$$\xi = -\frac{\partial V}{\partial x}, \qquad \xi' = \frac{\partial V}{\partial x'}, \qquad \eta = -\frac{\partial V}{\partial y}, \qquad \eta' = \frac{\partial V}{\partial y'}. \tag{4}$$

These equations can be combined into the vector equation

$$dV = \mathbf{t}'d\mathbf{b}' - \mathbf{t}d\mathbf{b}, \tag{5}$$

where the left side is a total differential.

Since (5) does not depend upon the choice of variables, we can therefore assume x, y, ξ, η to be independent variables, and replace $\mathbf{b}$ and $\mathbf{t}$ by their expressions in (1). We find then

$$dV = (\gamma\mathbf{b} + \delta\mathbf{t})(\alpha d\mathbf{b} + \beta d\mathbf{t} + \mathbf{b}d\alpha + \mathbf{t}d\beta) - \mathbf{t}d\mathbf{b}$$
$$= \tfrac{1}{2}[\alpha\gamma du + 2\beta\gamma dv + \beta\delta dw] + \gamma(u d\alpha + v d\beta) + \delta(v d\alpha + w d\beta). \tag{6}$$

The reader may convince himself that the differential equations in the former paper are simply the three integrability conditions of (6), and that we obtain them by calculating E_{uv}, E_{uw}, E_{vw} from (6). Equation (6) is equivalent to

$$V_u = \tfrac{1}{2}\alpha\gamma + \gamma(u\alpha_u + v\beta_u) + \delta(v\alpha_u + w\beta_u),$$
$$V_v = \beta\gamma + \gamma(u\alpha_v + v\beta_v) + \delta(v\alpha_v + w\beta_v), \tag{7}$$
$$V_w = \tfrac{1}{2}\beta\delta + \gamma(u\alpha_w + v\beta_w) + \delta(v\alpha_w + w\beta_w).$$

We can eliminate one of the four coefficients, α, β, γ, or δ, using (3). Eliminating γ, we are led to

$$\beta V_u = \delta[\tfrac{1}{2}\alpha^2 + (\alpha u + \beta v)\alpha_u + (\alpha v + \beta w)\beta_u] - (\tfrac{1}{2}\alpha + \alpha_u u + \beta_u v),$$
$$\beta V_v = \delta[\alpha\beta + (\alpha u + \beta v)\alpha_v + (\alpha v + \beta w)\beta_v] - (\beta + \alpha_v u + \beta_v v), \tag{8}$$
$$\beta V_w = \delta[\tfrac{1}{2}\beta^2 + (\alpha u + \beta v)\alpha_w + (\alpha v + \beta w)\beta_w] - (\alpha_w u + \beta_w v).$$

We can eliminate δ from (8), and thus get two differential equations connecting α and β. If V is a given function of u, v, w, we can thus (theoretically) determine α and β, and therefore γ and δ, by means of this single function V. The differential equations in question are

$$\frac{\beta V_u + \tfrac{1}{2}\alpha + \alpha_u u + \beta_u v}{\tfrac{1}{2}\alpha^2 + (\alpha u + \beta v)\alpha_u + (\alpha v + \beta w)\beta_u} = \frac{\beta V_v + \beta + \alpha_v u + \beta_v v}{\alpha\beta + (\alpha u + \beta v)\alpha_v + (\alpha v + \beta w)\beta_v}$$
$$= \frac{\beta V_w + \alpha_w u + \beta_w v}{\tfrac{1}{2}\beta^2 + (\alpha u + \beta v)\alpha_w + (\alpha v + \beta w)\beta_w}, \tag{9}$$

or

$$V_u[\alpha\beta + (\alpha u + \beta v)\alpha_v + (\alpha v + \beta w)\beta_v] - V_v[\tfrac{1}{2}\alpha^2 + (\alpha u + \beta v)\alpha_u + (\alpha v + \beta w)\beta_u]$$
$$- \beta\alpha_u v - \beta\beta_u w + \tfrac{1}{2}\alpha\alpha_v v + \tfrac{1}{2}\alpha\beta_v w + (\alpha_u\beta_v - \alpha_v\beta_u)(uw - v^2) = 0,$$
$$V_v[\tfrac{1}{2}\beta^2 + (\alpha u + \beta v)\alpha_w + (\alpha v + \beta w)\beta_w] - V_w[\alpha\beta + (\alpha u + \beta v)\alpha_v + (\alpha v + \beta w)\beta_v]$$
$$+ \tfrac{1}{2}\beta^2 + \tfrac{1}{2}\beta u\alpha_v + \beta v\alpha_w + \tfrac{1}{2}\beta v\beta_v + \beta w\beta_w + (\alpha_v\beta_w - \alpha_w\beta_v)(uw - v^2) = 0. \tag{10}$$

Let us now assume that for a given optical system we have calculated V, and therefore α, β, γ, δ; and let us proceed to investigate the image of an arbitrary surface.

2. Point and diapoint, the diapoint function. In a former paper[4] the author introduced the definition of a *diapoint*. Let us consider an object point P and a ray originating from it. The plane through the object point and the symmetry axis of the system is called the meridian plane. We define the diapoint as the point where the image ray intersects the meridian plane, and the diamagnification v as the ratio of the distances of a point and diapoint from the axis. The function F describing the optical path from the object point to its diapoint shall be called the diacharacteristic. To each point of the object space belongs a diacharacteristic, which we shall compute later.

To a given arbitrary point belongs a two-dimensional manifold of diapoints, one for each ray, and all these diapoints lie in the meridian plane. The diapoint characteristic for a given object point thus assumes a two-dimensional manifold of values. These may, however, be singular points; for instance, points having a sharp image. In this case, all the diapoints fall together, and, according to Fermat's law, the optical path from point to diapoint is constant along all rays. The diapoint characteristic is a constant.

An intermediate case is the one in which all rays intersect a curve in the image space. It can be proved that, in this case, the image rays can be split up into a one-dimensional manifold of pencils, each pencil converging to a point of this curve in such a manner that they form a series of circular cones with the curve tangent to the axis of each cone at its apex.[5] For each ray in such a cone, the diacharacteristic must again have a constant value, and thus the diacharacteristic can assume only a one-dimensional manifold of values. Such an image has been called half-symmetric by M. Boegehold[6] and the author.

Thus, the characteristic describes the image of any object point. Let us investigate an object point with the coordinates x_0, y_0, z_0 (optical lengths) and an arbitrary ray through it with the direction cosines ξ, η, $\zeta = \sqrt{1-(\xi^2+\eta^2)} = \sqrt{1-w}$. Let λ be the optical distance along the ray from the coordinate plane to the object point. We find that

$$\lambda = z_0 \sqrt{1-w}. \tag{11}$$

Let $\mathbf{b_0}$ be the vector (x_0, y_0). We then have

$$\mathbf{b_0} + \lambda\mathbf{t} = \mathbf{b}. \tag{12}$$

Inserting this into (1), we find that

$$\mathbf{b'} = \alpha\mathbf{b} + \beta\mathbf{t} = \alpha\mathbf{b_0} + (\beta + \alpha\lambda)\mathbf{t}, \qquad \mathbf{t'} = \gamma\mathbf{b_0} + (\delta + \gamma\lambda)\mathbf{t}. \tag{13}$$

An arbitrary point on the image ray is given by the vector $\mathbf{b_0'}$ with

$$\mathbf{b_0} = \mathbf{b'} + \lambda'\mathbf{t'} = (\alpha + \lambda'\gamma)\mathbf{b_0} + (\beta + \alpha\lambda + \lambda'(\delta + \gamma\lambda))\mathbf{t}. \tag{14}$$

We find the diapoint by choosing λ' such that $\mathbf{b_0'}$ and $\mathbf{b_0}$ are parallel, or that

[4] M. Herzberger, *A new theory of optical image formation*, J. Opt. Soc. Amer. **26**, 197–204 (1936).

[5] M. Herzberger, *Die Gesetze erster Ordnung in optischen Systemen*, Zeit. für Physik, **45**, 86–96 (1927).

[6] H. Boegehold, *Raumsymmetrische Abbildung*, Zeit. für Inst. **56**, 98–109 (1938).

$$\lambda' = (\alpha\lambda - \beta)/(\delta - \gamma\lambda). \tag{15}$$

This gives for the diamagnification

$$v = \alpha + \lambda'\gamma = 1/(\delta - \gamma\lambda). \tag{16}$$

The diacharacteristic is now given by

$$D = V - \lambda + \lambda' = V - \lambda + (\alpha\lambda - \beta)/(\delta - \gamma\lambda), \tag{17}$$

where $V(u, v, w)$ is the optical path between the object and the image coordinate planes. D is given in (17) as a function of u, v, w, and z, but for a given object point, these are not independent variables. Equation (12) shows that

$$x_0 - \lambda\xi = x, \qquad y_0 - \lambda\eta = y. \tag{18}$$

If we set

$$x_0^2 + y_0^2 = u_0, \tag{19}$$

then u_0 is constant for the given object point, and we find that u, v, w are connected by the equation

$$u_0 = u + 2\lambda v + \lambda^2 w. \tag{20}$$

Equation (20) allows us to eliminate u. Inserting this into (17), we obtain D as a function of v and w alone, with u_0 and z_0 as parameters. This function $D(u_0, z_0, v, w)$ is the diacharacteristic. We obtain the points that are imaged sharply if we find the value of u_0 and z for which the following identities are true:

$$D_v \equiv 0, \qquad D_w \equiv 0, \tag{21}$$

for any value of v and w. If a point has a half-symmetric image, we must have

$$D_{vv}D_{ww} - D_{vw}^2 \equiv 0. \tag{22}$$

Points u_0, z_0 for which (22) is fulfilled identically have half-symmetric images.

3. **Refraction at a sphere.** As an example let us investigate the image formation of a sphere of unit radius. The object and image planes will be at the center. If $\mathbf{s}$, $\mathbf{s}'$ are unit vectors along the ray (Fig. 1), we then have

FIG. 1.

$$\mathbf{a}' = \alpha\mathbf{a}, \qquad \mathbf{s}' = \gamma\mathbf{a} + \delta\mathbf{s}, \tag{23}$$

with

$$\alpha\delta = 1. \tag{24}$$

We denote the optical path from the object point to the refracting sphere (vector $\mathbf{a}_r$) by λ, and from the refracting sphere to the image reference plane by λ'. We then have

$$\mathbf{a} + \lambda\mathbf{s} = n\mathbf{a}_r, \qquad n'\mathbf{a}_r + \lambda'\mathbf{s}' = \mathbf{a}'. \tag{25}$$
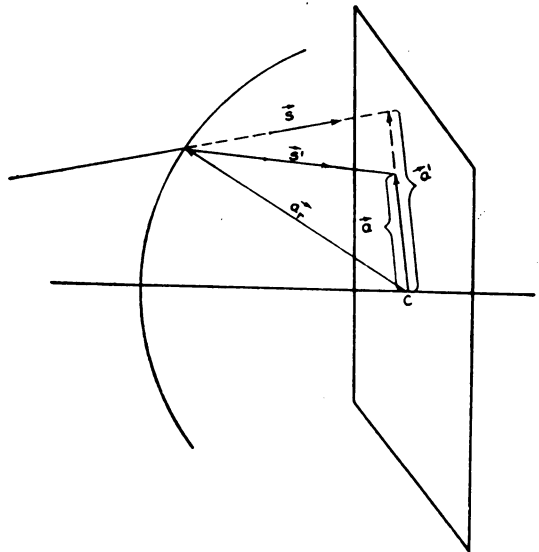
Inserting (25) into (23), we find that

$$\lambda' = - n'^2\lambda/(n^2 + nn'\gamma\lambda), \qquad \alpha = n'/n + \gamma\lambda', \qquad \delta = n/n' + \gamma\lambda, \qquad (26)$$

whence

$$V = \lambda \frac{n^2 - n'^2 + nn'\gamma\lambda}{n^2 + nn'\gamma\lambda} = \lambda + \lambda'. \qquad (27)$$

Finally, we find from (25) that

$$\mathbf{a} \times \mathbf{s} = n(\mathbf{a}_r \times \mathbf{s}) = n'(\mathbf{a}_r \times \mathbf{s}') = (\mathbf{a}' \times \mathbf{s}'). \qquad (28)$$

Thus, $\mathbf{a} \times \mathbf{s}$ is an invariant vector. Since $\mathbf{a} \cdot \mathbf{s} = v a^2 = u$ (cf. (2)), we then have

$$n\mathbf{a}_r \cdot \mathbf{s} = \sqrt{n^2 - u + v^2} = Q, \qquad n'\mathbf{a}_r \cdot \mathbf{s}' = \sqrt{n'^2 - u + v^2} = Q'. \qquad (29)$$

Equations (25) and (29) now give

$$\alpha = \frac{nn'}{n^2 + (Q - v)(Q' - Q)}, \qquad \beta = 0,$$
$$\gamma = (Q' - Q)/nn', \qquad \delta = n/n' + (Q - v)(Q' - Q)/nn'; \qquad (30)$$

and for the characteristic function we have

$$V = \frac{(Q - v)[n^2 - n'^2 + (Q - v)(Q' - Q)]}{n^2 + (Q - v)(Q' - Q)}. \qquad (31)$$

Since

$$n'^2 - n^2 = Q'^2 - Q^2, \qquad (31a)$$

Eq. (31) can be written in the form

$$V = \frac{(Q - v)(Q - Q')(Q' + v)}{u + (Q - v)(Q' + v)}. \qquad (32)$$

This is the characteristic function of the sphere. We now form the diapoint characteristic. Let μ be the distance of an object point x_0, y_0, z_0 from the reference plane measured along the object ray. Equation (17) gives

$$F = V - \mu + \frac{\alpha\mu}{\delta - \gamma\mu} = (\lambda - \mu)\left[1 - \frac{1}{(\delta - \gamma\mu)(\delta - \gamma\lambda)}\right]. \qquad (33)$$

In these equations we substitute for u from the relation

$$u = u_0 + \mu v + \mu^2 w. \qquad (34)$$

If we write

$$u_0 + z^2 = \rho^2, \qquad (35)$$

where ρ is the distance of the object point from the center, and

$$\tau = v + \mu = v + z(1 - w)^{-1/2}, \qquad (36)$$

we then find that all our functions depend only on ρ^2 and τ. We have in particular

$$Q^2 = n^2 - \rho^2 + \tau^2, \qquad Q'^2 = n'^2 - \rho^2 + \tau^2, \qquad (37)$$

and finally

$$F = \frac{(Q' - Q)(Q - \tau)(Q' + \tau)}{\rho^2 + (Q - \tau)(Q' + \tau)} . \tag{38}$$

F as function of v and w assumes only a one-dimensional manifold of values. Every point has a symmetrical image.

Equation (38) becomes independent of τ only if

$$Q = \tau \quad \text{or} \quad Q = -\tau. \tag{39}$$

The first case leads to $\rho = n$, i.e., the refracting surface itself has a sharp image. The second case leads to

$$\rho = -n', \tag{40}$$

i.e., to the so-called aplanatic surfaces. In both cases the optical path between the object and the image points vanishes.

A simpler method leads to the diapoint characteristic of a plane refracting surface. If z is the distance of the object point from the plane, we find that

$$F = z(1 - w)^{-1/2}(n'^2/n^2 - 1). \tag{41}$$

For a given object point, F is a function of w alone. Every object point is therefore half-symmetric. Only for $z = 0$ can F become independent of w. The plane therefore gives a sharp image only of itself (and of the infinite plane, whose points are given by $z/\zeta = z/\sqrt{1-w} = $ constant).