THE ELEMENTARY THEORY OF THE SYMMETRICAL OPTICAL INSTRUMENT.

I. APPROXIMATE FORMULAE FOR A SUCCESSION OF REFRACTIONS AT NEARLY NORMAL INCIDENCE.

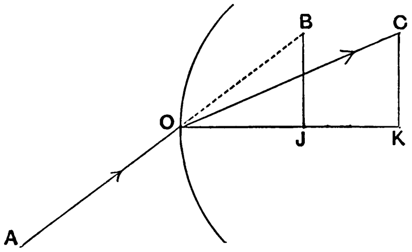
1. Analytical formulae expressing the laws of refraction.

These laws are: (i) The incident ray, the refracted ray, and the normal to the refracting surface at the point of incidence are in one plane. (ii) The ratio of the sines of the angles of incidence and refraction is a constant, depending only on the nature of the media in which the light is propagated.

It is important to express these laws in terms of the direction cosines of the lines involved. Let μ be the index of refraction of the medium in which is the incident ray, μ' the index of the medium in which is the refracted ray; let the cosines of the incident ray be (l, m, n), those of the refracted ray (l', m', n'), those of the normal to the refracting surface at the point of incidence, drawn towards the medium μ', (L, M, N); let the angles of incidence and refraction be ϕ and ϕ' respectively.

In the diagram AO is the incident ray, OC the refracted ray, and OK the normal. On the incident ray produced a point B is taken so that OB is μ units of length, and on the refracted ray a point C is taken so that OC is μ' units of length. BJ, CK are perpendiculars drawn to the normal. Then the angles BOJ, COK are ϕ, ϕ' respectively, and the lengths of JB and KC are $\mu \sin \phi$ units and $\mu' \sin \phi'$ units respectively.

The first law tells us that JB and KC are in the same plane through JK, and therefore parallel to one another. The second law tells us that JB and KC are of the same length. The two laws are expressed in the statement that JB and KC have equal projections on each of the coordinate axes. Now the projection of JB is the excess of the projection of OB over that of OJ; and OB is of length μ and



has cosines (l, m, n), while OJ is of length $\mu \cos \phi$ and has cosines (L, M, N). Likewise the projection of KC is the difference of the projections of OC and OK. So the equalities of the projections of JB and KC on the three axes of coordinates are expressed by equations of the type

$$\mu l - \mu \cos \phi L = \mu' l' - \mu' \cos \phi' L.$$

Rearranging, we get the following equations to express the laws of refraction:

$$\left.\begin{aligned} \mu' l' - \mu l &= (\mu' \cos \phi' - \mu \cos \phi) L \\ \mu' m' - \mu m &= (\mu' \cos \phi' - \mu \cos \phi) M \\ \mu' n' - \mu n &= (\mu' \cos \phi' - \mu \cos \phi) N \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots(1).$$

The three equations are not all independent, as is readily seen by multiplying them by L, M, N respectively and adding, it being remembered that $\Sigma L l' = \cos \phi'$, and $\Sigma L l = \cos \phi$.

2. Approximate formulae for nearly normal incidence.

When the incident ray is very nearly normal, it is readily seen that the refracted ray is also very nearly normal. It is therefore possible so to choose the axis of z that it shall be nearly parallel to the incident ray, the refracted ray, and the normal at the point of incidence. When the axis of z has been so chosen, l, m, l', m', L, M, are all small. We shall obtain approximate formulae on the hypothesis that these quantities are so small that their squares and products may be neglected; this is equivalent to the supposition that aberration is to be neglected.

On this hypothesis n, which is equal to

$$(1 - l^2 - m^2)^{\frac{1}{2}},$$

differs from unity by small quantities of the second order; so also

n' and N. Hence we may replace all three by unity. Also ϕ and ϕ' are small of the same order as l, etc.; so $\cos\phi$, $\cos\phi'$ differ from unity by small quantities of the second order, and may be replaced by unity.

Thus the third of equations (1) becomes, on the basis of the proposed approximation, an identity; and the other two take the forms

$$\left.\begin{aligned}\mu'l' - \mu l &= (\mu' - \mu)\,L \\ \mu'm' - \mu m &= (\mu' - \mu)\,M\end{aligned}\right\} \quad\ldots\ldots\ldots(2).$$

3. Expression of L, M in terms of the coordinates of the point of incidence.

The most usual application of the formulae (2) is to the case in which a pencil of rays pass nearly normally through a comparatively small portion of the refracting surface, so that the rays and the normals at all the points of incidence are very nearly parallel to one another. It is then advantageous to take as axis of z the normal at some one of the points of incidence, say at a point so centrally situated that it may be called the centre of the portion of the surface through which refraction takes place; the ray incident at this point may be called the central ray of the pencil; it need not be more precisely defined.

Unless the point at which the axis of z is normal, say $(0, 0, c)$, is a singular point on the refracting surface, the part of the surface in the neighbourhood of this point can be approximately represented by an equation of the type

$$2\,(z - c) + ax^2 + 2hxy + by^2 = 0 \quad\ldots\ldots\ldots(3),$$

the approximation neglecting terms of the third order in x and y. The constants a, h, b are of course known when the shape of the surface is known.

At the point (x, y, z) the cosines of the normal are approximately

$$(ax + hy),\ (hx + by),\ 1,$$

when the squares and products of x, y are neglected. The first two of these may be substituted for L, M in the formulae (2), and so we get the refraction of the ray incident at (x, y, z) determined by the equations

$$\left.\begin{aligned}\mu'l' - \mu l &= (\mu' - \mu)\,(ax + hy) \\ \mu'm' - \mu m &= (\mu' - \mu)\,(hx + by)\end{aligned}\right\} \quad\ldots\ldots\ldots(4).$$

It is, of course, possible to choose the coordinate planes of x and y

so that the h of formula (3) shall be zero. If this were done the approximate equation of the surface would take the form

$$2(z-c)+\frac{x^2}{\rho_1}+\frac{y^2}{\rho_2}=0 \quad \text{....................(5)},$$

and ρ_1, ρ_2 would be the principal radii of curvature of the surface at the point $(0, 0, c)$, reckoned positive when the corresponding convexities of the surface are towards the medium μ'. The formulae which would then take the place of (4) are

$$\left.\begin{aligned} \mu' l' - \mu l &= \frac{\mu'-\mu}{\rho_1}x \\ \mu' m' - \mu m &= \frac{\mu'-\mu}{\rho_2}y \end{aligned}\right\} \quad \text{..................(6).}$$

4. A series of refracting surfaces having a common normal.

When a ray traverses a succession of different media arranged in such a way that the refracting surfaces have a common normal with which the ray is always nearly coincident, it is interesting to see how the equations of the previous Article enable us to derive from a knowledge of the position of the ray before its first incidence a complete specification of the ray after its final emergence.

The incident ray (say in a medium μ_0) is known when we know its cosines l_0, m_0, and the coordinates (x_1, y_1, z_1) of the point where it meets the first surface,

$$2(z-c_1)+a_1x^2+2h_1xy+b_1y^2=0.$$

Clearly z_1 differs from c_1 only by quantities of the second order, so we may replace z_1 by c_1, and regard the incident ray as specified by the four quantities l_0, m_0, x_1, y_1.

After the first refraction (into a medium μ_1) the cosines of the ray are changed to l_1, m_1, given by the equations

$$\mu_1 l_1 - \mu_0 l_0 = (\mu_1-\mu_0)(a_1x_1+h_1y_1),$$
$$\mu_1 m_1 - \mu_0 m_0 = (\mu_1-\mu_0)(h_1x_1+b_1y_1).$$

Equations of this type, which determine the change of direction due to refraction, may be called "optical" equations.

The coordinates (x_2, y_2, c_2) of the point where the refracted ray meets the second refracting surface

$$2(z-c_2)+a_2x^2+2h_2xy+b_2y^2=0,$$

can be obtained by putting $z=c_2$ (a sufficient approximation) in the equations of the ray, i.e. of the line which proceeds from (x_1, y_1, c_1) in the direction $(l_1, m_1, 1)$. Now the equations of this line are

$$(x - x_1)/l_1 = (y - y_1)/m_1 = z - c_1,$$

and so x_2, y_2 are determined by the equations

$$\left.\begin{aligned} x_2 &= x_1 + l_1\ (c_2 - c_1) \\ y_2 &= y_1 + m_1(c_2 - c_1) \end{aligned}\right\} \qquad \text{(7)}.$$

Equations of this type, which determine the coordinates of a point of refraction in terms of those of the previous point of refraction, may be called "geometrical" equations.

Having found x_2, y_2 by the geometrical equations, we are in a position to use the optical equations corresponding to the next refraction, viz.

$$\mu_2 l_2 - \mu_1 l_1 = (\mu_2 - \mu_1)(a_2 x_2 + h_2 y_2),$$
$$\mu_2 m_2 - \mu_1 m_1 = (\mu_2 - \mu_1)(h_2 x_2 + b_2 y_2),$$

equations which we could not use till we knew x_2, y_2, but which now give us the values of l_2, m_2.

Eliminating l_1, m_1 from the six equations we are left with four equations which give us explicit formulae for x_2, y_2, l_2, m_2 in terms of l_0, m_0, x_1, y_1; that is, the quantities specifying the ray in the medium μ_2 in terms of the quantities that specify the incident ray.

Thus the problem of refraction is solved for the case of two surfaces. If there are n surfaces and $n+1$ media, we get in a similar manner $2n$ optical equations and $2n-2$ geometrical equations. From these we can eliminate successively the $4n-6$ quantities

$$l_1,\ m_1,\ x_2,\ y_2,\ l_2,\ m_2,\ x_3,\ y_3,\ \dots\ x_{n-1},\ y_{n-1},\ l_{n-1},\ m_{n-1},$$

and obtain finally four equations expressing x_n, y_n, l_n, m_n, the quantities specifying the emergent ray, in terms of l_0, m_0, x_1, y_1, the quantities that specify the incident ray*.

5. Case in which all the refracting surfaces are symmetrical about the same two planes through the axis.

By suitable choice of the planes of x and y it is always possible to make the coefficient of xy in the equation of one of the refracting surfaces vanish; but in general this choice would leave the corresponding coefficients for all the other surfaces different from zero. But if the surfaces are such that their indicatrices at the points where they are met by the common normal (the axis of z) all have their principal

* Cf. Prof. R. A. Sampson on Gauss's *Dioptrische Untersuchungen*, *Proc. London Math. Soc.* xxxix. 1898, p. 33.

axes in the same two directions, then the same choice of coordinate planes will make all the h's vanish simultaneously.

In this case the equations of the preceding Article are greatly simplified, for when all the h's are zero, the equations divide themselves into two sets, one set involving only x's and l's, the other set involving only y's and m's. To solve the problem of n surfaces we now have only to eliminate the $2n-3$ quantities $l_1, x_2, l_2, \ldots x_{n-1}, l_{n-1}$, from the n optical and the $n-1$ geometrical equations of the first set. The result, with suitable change of symbols, serves also for the corresponding equations of the second set; so that in this particular case the labour of elimination is much less than half that required in the general case.

6. Form of the results in the general case.

Taking the x's, y's, l's, m's in the order which presents itself naturally as one follows the course of a ray, we see that each of these quantities is a homogeneous linear function of those that precede it. Consequently when the elimination has been performed we get x_n, y_n, l_n, m_n as homogeneous linear functions of l_0, m_0, x_1, y_1. A more useful result is arrived at if we specify the positions of the entering and the emergent rays, not by the coordinates (x_1, y_1), (x_n, y_n) of the points where they meet the planes $z=c_1$, $z=c_n$, but by the coordinates (ξ_0, η_0), (ξ, η) of the points where they respectively meet two other arbitrarily selected planes $z=c_1-p$, $z=c_n+q$, which we call planes of reference. This implies the introduction of four more geometrical equations, namely two of the type

$$x_1 = \xi_0 + pl_0,$$

and two of the type

$$\xi = x_n + ql,$$

and the addition of x_1, y_1, x_n, y_n to the quantities to be eliminated.

If we denote by a's with double suffix the coefficients in the expressions for l, m, x_n, y_n in terms of x_1, y_1, l_0, m_0, so that, for example, $l=a_{11}x_1+a_{12}y_1+a_{13}l_0+a_{14}m_0$, it is readily verified that the suggested elimination leads to

$$\left.\begin{aligned} \xi &= (a_{31}+qa_{11})\,\xi_0 + (a_{32}+qa_{12})\,\eta_0 \\ &\quad + (a_{33}+pa_{31}+qa_{13}+pqa_{11})\,l_0 + (a_{34}+pa_{32}+qa_{14}+pqa_{12})\,m_0 \\ \eta &= (a_{41}+qa_{21})\,\xi_0 + (a_{42}+qa_{22})\,\eta_0 \\ &\quad + (a_{43}+pa_{41}+qa_{23}+pqa_{21})\,l_0 + (a_{44}+pa_{42}+qa_{24}+pqa_{22})\,m_0 \end{aligned}\right\} \ldots(8),$$

and two others which we may regard as contained in these two, since they may be derived by differentiating with respect to q and remembering that

$$\frac{\partial \xi}{\partial q} = l, \quad \frac{\partial \eta}{\partial q} = m.$$

Thus the whole theory of the set of n surfaces is contained (to the degree of approximation postulated at the outset) in two linear equations involving a rather large array of constants*.

7. Optical interpretation of the equations for the general case.

Without going into detail, it may be useful to indicate how these equations may be interpreted optically, in other words to shew how they give us information as to the image formed by the rays of light that proceeded originally from a given bright point. We regard $(\xi_0, \eta_0, c_1 - p)$ as the coordinates of the bright point, and suppose a white screen to be placed in the plane $z = c_n + q$. If we fix attention on the ray which sets out from the source in a given direction, specified by given values of l_0 and m_0, the equations tell us the co-ordinates of the point in which the ray after refraction strikes the screen. If the pencil of refracted rays form a point image, it must be possible, namely by placing the screen where it will receive the image, to make all the refracted rays strike the screen in the same point; in fact, it must be possible to give such a value to q that ξ and η shall be independent of l_0 and m_0. This means that the giving of a suitable value to q makes four coefficients vanish. Though these four conditions are not all independent, they are in general equivalent to three independent conditions, and so cannot be simultaneously satisfied except in special cases.

Failing to obtain a point image, we next try to find the positions of the focal lines of the pencil of emergent rays. For the points in which the rays strike the screen to lie in a straight line, the necessary and sufficient condition is that there should be a linear relation between ξ and η with coefficients independent of l_0 and m_0. This is the case if the same value of λ makes the coefficients both of l_0 and of m_0 vanish in the expression for $\xi + \lambda\eta$. The condition is therefore the vanishing of

$$\begin{vmatrix} a_{33} + pa_{31} + qa_{13} + pqa_{11}, & a_{34} + pa_{32} + qa_{14} + pqa_{12} \\ a_{43} + pa_{41} + qa_{23} + pqa_{21}, & a_{44} + pa_{42} + qa_{24} + pqa_{22} \end{vmatrix}.$$

There are two values of q which satisfy this condition, and these define the positions of the two focal lines of the emergent pencil.

* Amongst the 16 constants there are 6 relations, which are the conditions for the existence of a Characteristic Function. For the form of the relations see Sampson, l.c. pp. 38, 39; for the connexion with the Characteristic Function see Bromwich, *Proc. London Math. Soc.* XXXI. 1899, p. 8.

8. Case of two planes of symmetry.

When the surfaces are as described in Article 5, the mathematical solution of the refraction problem is much simpler. The equations reduce to

$$\left.\begin{aligned} \xi &= (a_{31} + qa_{11})\,\xi_0 + (a_{33} + pa_{31} + qa_{13} + pqa_{11})\,l_0 \\ \eta &= (a_{42} + qa_{22})\,\eta_0 + (a_{44} + pa_{42} + qa_{24} + pqa_{22})\,m_0 \end{aligned}\right\} \quad \ldots\ldots(9),$$

and the optical interpretation is easier than in the general case.

II. THE MATHEMATICAL SOLUTION OF THE REFRACTION PROBLEM FOR A SYMMETRICAL INSTRUMENT.

9. The symmetrical optical instrument.

We proceed now to the detailed discussion of a very particular case of the general arrangement discussed in Article 4, namely the case in which all the refracting surfaces are symmetrical with respect, not merely to two planes, but to every plane through the axis of z. The refracting surfaces are then necessarily surfaces of revolution having the axis of z as common axis of revolution; they are generally spheres, but are sufficiently well represented by approximate equations of the type

$$2(z-c) + \frac{x^2+y^2}{\rho} = 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots(10).$$

Such an arrangement is called a symmetrical optical instrument, and is the kind of instrument most frequently employed in every-day life.

The approximate theory here developed applies to any symmetrical instrument which is used in such a way that the rays which it transmits are very nearly parallel to the axis, for example the telescope and opera-glass; it is practically useless in the case of wide-angle instruments such as the microscope or portrait-camera.

10. Power of a single spherical refracting surface.

When refraction takes place from a medium of index μ to a medium of index μ' through a surface represented approximately by equation (10), the optical equations are equations (6) of Article 3, simplified by the equality of both ρ_1 and ρ_2 to ρ. They are, in fact,

$$\left.\begin{aligned} \mu' l' - \mu l &= \left(\frac{\mu'-\mu}{\rho}\right) x \\ \mu' m' - \mu m &= \left(\frac{\mu'-\mu}{\rho}\right) y \end{aligned}\right\} \quad \ldots\ldots\ldots\ldots\ldots\ldots(11).$$

Thus it appears that, to the approximation contemplated, the optical properties of the spherical refracting surface are all contained in the single constant

$$\frac{\mu' - \mu}{\rho} \qquad \text{............................(12).}$$

This constant is called the "diverging power" or simply the "power" of the surface, and is usually denoted by κ.

It is most important that the definition of the power of a surface should be understood precisely, without confusion on account of the conventions of sign which are implicitly involved in the above formula. These conventions are ultimately two, namely (i) that the light passes from the medium μ into the medium μ', and (ii) that ρ is positive when the convexity of the refracting surface is towards the second medium.

Suppose the direction of the ray of light to be reversed. Then μ' becomes the index of the first medium, and μ that of the second, so that $\mu - \mu'$ must take the place of $\mu' - \mu$ in formula (12); but convexity towards the medium μ' is concavity towards the medium μ, which is now the second medium, so that the case required by definition is that in which ρ is negative, and $-\rho$ must take the place of ρ in the denominator. Thus the power of the surface for a ray going from μ' to μ is

$$\frac{\mu - \mu'}{-\rho},$$

that is, the same as before. Thus it appears that the power of a surface is not altered by reversing the ray, and its definition has nothing to do with the sense in which the ray passes.

The second of the conventions is thus the only one that affects the definition of the power, and its bearing on the definition is clearly summed up in the following rule:

The power of a refracting surface is to be regarded as positive when the medium on the convex side of the surface has a greater refractive index than the medium on the concave side; the power is negative when the medium of greater index is on the concave side.

This is only another way of saying that κ is positive if $\mu' > \mu$ and $\rho > 0$, or if $\mu' < \mu$ and $\rho < 0$.

For a plane refracting surface ρ is infinite, and κ is zero.

The optical equations for a surface of power κ are

$$\mu' l' - \mu l = \kappa x, \quad \mu' m' - \mu m = \kappa y \qquad \text{..................(13).}$$

11. Reduced projected inclinations of a ray.

The equations of the ray in the medium μ being of the form

$$\frac{x-a}{l}=\frac{y-\beta}{m}=z-\gamma,$$

it follows that the equation of the projection of the ray on the plane of xz is of the form

$$x-a=l\,(z-\gamma).$$

Hence l is the tangent of the angle that this projection of the ray makes with the axis of z, or, to our degree of approximation, the angle itself. In fact l and m are the inclinations to the axis of z of the projections of the ray on the coordinate planes through that axis. We call l and m the "projected inclinations" of the ray.

Putting $\mu l=\delta$, $\mu m=\epsilon$, we may call δ and ϵ the "reduced projected inclinations" of the ray, the word "reduced" in this connexion meaning that the projected inclinations are multiplied by the index of the medium in which the ray is passing.

In terms of this notation, the optical equations for a single surface take the form

$$\delta'-\delta=\kappa x,\quad \epsilon'-\epsilon=\kappa y \qquad\text{.....................(14).}$$

12. Divergence produced by a refracting surface.

If the effect of refraction were an increase in the projected inclinations of each ray, a pencil of rays would, after passing through the surface, be more divergent or less convergent than before; the surface would literally produce divergence. But it is to be noticed that the optical equations give information, not as to the changes produced by the refraction in the projected inclinations, but as to the changes produced in the *reduced* projected inclinations; and the power of the surface measures the degree to which, for a given point of incidence, it is capable of increasing the reduced projected inclinations. It is therefore convenient to abandon the literal meaning of the word "divergence," and to apply the term to what is really quite different, namely increase of the reduced projected inclinations. With this understanding, a diverging surface is one whose power is positive, a converging surface is one whose power is negative. A plane surface produces neither convergence nor divergence.

In the case of an instrument consisting of several surfaces, if the first and last media are the same there is no difference between the literal and the special meanings of divergence. Thus a double convex