"Über den Zusammenhang der Theorie der absoluten optischen Instrumente with einem Satz der Variationsrechnung," Sitz. der math.-phys. der Abt. Bay. Akad. der Wiss. zu München **1** (1926), 1-18.

On the connection between the theory of absolute optical instruments and a theorem in the calculus of variations

By C. Carathéodory

Presented at the session on 16 January 1926

Translated by D. H. Delphenich

Contents

	Page
Introduction.	1
Historical overview.	1
The Maxwell fisheye.	2
The general mapping theorem.	4
The stigmatic map of surfaces.	8
The stigmatic map of isotropic spaces.	12
Application to the calculus of variations.	14
	Introduction. Historical overview. The Maxwell fisheye. The general mapping theorem. The stigmatic map of surfaces. The stigmatic map of isotropic spaces. Application to the calculus of variations.

1. Introduction. – In the following article, a property of absolute optical instruments that has been proved only for homogeneous isotropic object spaces and similar image spaces will be generalized. That generalization can also be inferred from the analogous theorem for arbitrary symmetric variational problems, and its proof is even much simpler and more concise than that of the original theorem itself.

2. Historical overview. – In the year 1858, **J. C. Maxwell** used a very elementary method to prove the theorem that for an "absolute" optical instrument – i.e., one for which every point in the object space creates a sharp (viz., stigmatic) image in the image space – the object and the image must be equally large (as measured in light time) (¹). In the proof, he had generally neglected second-order quantities, such that the result would initially seem to be valid for only small objects.

^{(&}lt;sup>1</sup>) "On the general laws of optical instruments," Quart. J. pure appl. math. **11** (1858), 233-244 or *Scientific Papers* 1, pp. 271-285, see esp., Props. VIII and IX.

Later, in his famous work on the eikonal (¹), **H. Bruns** proved rigorously and in general that under the absolute map, the image would be similar or symmetric to the object. By contrast, **Bruns** did not emphasize that the image and object would have to be be equally large when measured in light time, although that fact was almost a direct consequence of his formulas. **F. Klein** had briefly inferred that last consequence in the context of a surprisingly elegant proof that he gave for the theorem in question (²). **Klein** employed the imaginary "minimal ray" in his considerations, which would not be refracted at the separation surface under the transition from one medium to the other, as he pointed out. **H. Liebmann** ultimately found a geometric proof that was just as simple as that of **Klein**, but which involved only real rays (³). **Liebmann**'s beautiful proof not only had the latter advantage, but above all, the fact that the only rays that were employed in his construction were ones that actually ran completely through the opening of his instrument, no matter how narrow that opening might be. It is obvious that such a restriction on the construction must always be required (⁴).

3. The Maxwell fisheye. – All of those proofs assumed *in an essential way* that the object space, as well as the image space, were isotropic and homogeneous, such that from a theorem of **E. Abbe**, the absolute instrument in question would generate a *collinear* map between the two spaces. From the theorem of Maxwell-Bruns-Klein that was just mentioned, the image for an absolute instrument is always congruent or symmetric to the object, and the plane mirror is the single optical instrument that one knows that produces such a map.

Now, **Maxwell** had remarked (⁵) that in a medium with varying index of refraction, it might be very likely that all rays that go through an arbitrary point will once more meet at a single point such that in such a medium, any sufficiently-small object will actually possess a stigmatic image.

Maxwell made that discovery in the context of his study of the spherical lens in the eye of a fish, whose refraction might be determined by the following formula: If one lets r denote the distance from a point in the lens of the eye to its center and n denotes the index of refraction at the point considered then the following equation will be true:

(1)
$$n = \frac{2ab}{b^2 + r^2},$$

^{(&}lt;sup>1</sup>) "Das Eikonal," Abhandl. der Kgl. Schs. Ges. d. Wiss., math.-phys. Klasse **21** (1895), zee esp. pp. 370.

^{(&}lt;sup>2</sup>) "Räumliche Kollineation bei optischen Instrumenten," Zeit Math. Phys. **46** (1901), 376-382 or Ges. Abh., Bd. II, pp. 607-612.

^{(&}lt;sup>3</sup>) "Der allgemeine Malussche Satz und der Brunssche Abbildungsatz," these Sitz. (1916), 183-200.

^{(&}lt;sup>4</sup>) A very good survey of the results that were cited here was given by **H. Boegenhold**. One finds it in the new third edition of the book by **S. Czapski** and **O. Eppenstein**, *Grundzüge der optischen Instrumente nach Abbe*, Leipzig, J. A. Barth, 1924, pp. 213-216.

^{(&}lt;sup>5</sup>) "Solution of problems," Camb. and Dublin math. jour. **8** (1854), 188-193 or *Scient. Pap.* 1, pp. 74-79.

in which *a* and *b* mean positive constants. In the eightieth year of the present century, **L**. **Mathiessen** found that **Maxwell**'s formula (1) agreed quite well by measurements made with the lens in the eye of the cod and other fish (1).

Now, **Maxwell** found with the help of some rather elegant geometric considerations that when one fills up all of space with a medium whose index of refraction obeys the law (1), the light rays will all be circular or rectilinear, and that those of them that start from a point A in space that is different from the center O of the "fisheye" will all once more run through a second point A_1 in space. In that way, O always lies along the segment AA_1 and divides that segment into two intervals for which the relation:

$$AO \times OA_1 = b^2$$

is true. Those conditions suffice to completely characterize all light rays.

4. – Now, that result of Maxwell's follows at one stroke from the remark that in the equation:

(3)
$$\begin{cases} d\sigma = \frac{2ab}{b^2 + r^2} \sqrt{dx^2 + dy^2 + dz^2}, \\ r^2 = x^2 + y^2 + z^2, \end{cases}$$

the differential $d\sigma$, which defines the optical length of a line element inside of the Maxwell fisheye, can also be interpreted as the line element of the three-dimensional boundary of a four-dimensional sphere that is projected stereographically onto the space of *x*, *y*, *z*. The diameter of the sphere must be taken to be equal to 2a in that, and the distance from the space of *x*, *y*, *z* to the center of projection must be taken to be *b*. The extremals of the variational problem that corresponds to the line integral in (3) coincide with the images of great circles on our four-dimensional sphere. However, those images are the circles in the space of *x*, *y*, *z* that include two diametrically-opposite points of the sphere:

$$x^2 + y^2 + z^2 = b^2$$

They are then characterized by the facts that their planes include the origin O of the coordinates and that the power of the point O relative to each of those circles will always be equal to $-b^2$.

Each pair A, A_1 of conjugate points for which the relation (2) is true corresponds to a pair of diametrically-opposite points of our four-dimensional sphere, and since the distance between two points of the sphere is equal to the distance to their opposite points (while both distances are measured on the surface of the four-dimensional sphere), it will follow that for the variational problem (3), the extremal distance between two points A, B in the space of x, y, z must be equal to the extremal distance between the conjugate points A_1 and B_1 .

^{(&}lt;sup>1</sup>) L. Mathiessen, "Über ein merkwürdiges optisches Problem con Maxwell," F. Exners Repert. d. Phys. 24 (1888), 401-407.

It will then follow that for the stigmatic map that the "Maxwell fisheye" generates, every curve that is drawn on the object corresponds to a curve on the image *with precisely the same optical length*. Indeed, the map is no longer collinear, but it will be *true to scale*, as the aforementioned theorem in § 2 requires. We will see that this is an entirely general phenomenon and that the theorem of **Maxwell-Bruns-Klein** is not at all linked with collinear maps.

5. The general mapping theorem. – Let *J* be an arbitrary optical instrument that is intersected by a light ray ABA_1B_1 .

We shall not assume that the object space in which the segment AB of our light ray lies or the image space in which A_1B_1 lies are homogeneous, such that our light ray can be doubly-curved curve along its entire course.

The line element of the object space, which deviates neither in length nor direction very strongly from a line element of the part AB of our light ray, i.e., the ones that belong to a "narrow neighborhood" of AB, as one says in the calculus of variations, have the property that every light ray that include one of those line elements will go through both pupils of the instrument, just like ABA_1B_1 , and will arrive in the image space. We then say that the light ray lies *in the field of the instrument*.

Let γ be an otherwise-arbitrary curved segment with continuously-varying tangents and nothing but light rays that lie in the field of our instrument will go through whose line elements. We would then like to say that the curve γ *lies tangentially in the field of J*. It is clear that every light polygon that is inscribed in γ will consist of nothing but light rays that lie in the field of *J* when its sides are chosen to be sufficiently small.

6. – With those preparatory considerations, which are true without restriction, we assume that we have an *absolute* instrument before us. That is, we assume that when all of the light rays that we consider start from a point A in the object space, they must cross at a point A_1 in the image space.

All of the rays that go through our instrument that connect points A and A_1 that correspond in that way will then have equal optical lengths between those points, as would emerge from a very elementary and well-known theorem in the calculus of variations. We would like to let $\varphi(A)$ denote that optical distance between an arbitrary point A in object space and its image point A_1 , which is therefore independent of the direction that the ray connecting A to A_1 might possess at the point A.

7. – If we let *h* denote the optical distance between *A* and *B* and let h_1 denote the optical distance between A_1 and B_1 then, from the figure below, we will have:

 $h + \varphi(B) = \varphi(A) + h_1$

or

(4)
$$h_1 = h + \varphi(B) - \varphi(A) .$$



We shall now consider a curve γ that lies tangentially to the optical instrument (§ 5) and connects the points *A* and *B*, and we denote the image of γ by γ_1 .

Let *APQB* be an arbitrary light polygon that is inscribed in γ whose sides lie in the field of the instrument, and let $A_1P_1Q_1B_1$ be its image that is inscribed in γ_1 . If we let u, v, w denote the optical lengths of the sides of the polygon that is inscribed in γ and let u_1 , v_1 , w_1 denote the optical lengths of the images of those sides then we will have the following equations, which are obtained in the same way as equation (4):

(5)
$$\begin{cases} u_1 = u + \varphi(P) - \varphi(A), \\ v_1 = v + \varphi(Q) - \varphi(P), \\ w_1 = w + \varphi(B) - \varphi(Q). \end{cases}$$

In that, one must consider that since the two light rays ABA_1 and APA_1 have the same optical length, $\varphi(A)$ has the same meaning in (4) and in the first of equations (5). One sees in exactly the same way that the values of $\varphi(P)$, $\varphi(Q)$, $\varphi(B)$ represent the same number in each of the two equations (4) or (5) in which they occur. Upon adding equations (4), one will then get:

(6)
$$u_1 + v_1 + w_1 = u + v + w + \varphi(B) - \varphi(A)$$
,

which is a relation that says that the difference between the optical lengths of the light polygon that is inscribed in γ and its image is equal to $\varphi(B) - \varphi(A)$. That property, which is independent of the number of sides to the inscribed polygon, can be carried over to the optical lengths of the curves γ and γ_1 by passing to the limit, and in that way we will get:

Theorem 1:

For any absolute optical instrument that maps a point in an object space \Re sharply to the point in an image space \Re_1 , the relation:

(7)
$$L_1 = L + \varphi(B) - \varphi(A)$$

will exist between the optical lengths L and L_1 of an arbitrary curve γ that lies tangentially in the field of the instrument and its image γ_1 , resp., in which $\varphi(A)$ and $\varphi(B)$ mean the optical distances from the endpoints A and B of γ to the endpoints A_1 and B_1 of γ_1 .

8. – The theorem that we have in mind will then be proved when we succeed in showing that $\varphi(A) = \varphi(B)$.

In regard to that, we remark that the optical lengths L and L_1 of $\gamma(\gamma_1, \text{resp.})$ can be represented by integrals along those curves. We can then write:

(8)
$$L = \int_{\gamma} F(x, y, z, \dot{x}, \dot{y}, \dot{z}) dt,$$

(9)
$$L_1 = \int_{\gamma_1} F_1(x, y, z, \dot{x}, \dot{y}, \dot{z}) dt.$$

In them, the two curves γ and γ_1 are represented with the help of a parameter *t*, and the functions *F* and *F*₁ are homogeneous of order one in \dot{x} , \dot{y} , \dot{z} (\dot{x}_1 , \dot{y}_1 , \dot{z}_1 , resp.). That last condition is known to have the consequence that the values of the integrals (8) and (9) are independent of the choice of parameter *t*. The two functions *F* and *F*₁ can then be completely different from each other. From our assumptions, e.g., the object space \Re can very well be crystalline, while the image space \Re_1 is isotropic.

The (stigmatic) map of the two spaces \Re and \Re_1 to each other can now be represented by the relations:

(10)
$$x_1 = \xi(x, y, z), \qquad y_1 = \eta(x, y, z), \qquad z_1 = \zeta(x, y, z).$$

If one sets:

(11)
$$\frac{d\xi}{dt} = \frac{\partial\xi}{\partial x}\dot{x} + \frac{\partial\xi}{\partial y}\dot{y} + \frac{\partial\xi}{\partial z}\dot{z},$$

with similar equations for $d\eta / dt$ and $d\zeta / dt$, and introduces the notation:

(12)
$$\Phi(x, y, z, \dot{x}, \dot{y}, \dot{z}) = F_1\left(\xi, \eta, \zeta, \frac{d\xi}{dt}, \frac{d\eta}{dt}, \frac{d\zeta}{dt}\right),$$

then one can replace the curve integral (9) over γ_1 with a curve integral over γ and write:

(13)
$$L_1 = \int_{\gamma} \Phi(x, y, z, \dot{x}, \dot{y}, \dot{z}) dt,$$

instead of (9).

With the help of (8) and (13), equation (7) will then assume the form:

$$\int_{\gamma} (\Phi - F) dt = \varphi(B) - \varphi(A)$$

However, the last equation means that the value of a curve integral over $(\Phi - F)$ depends upon only the endpoints *A*, *B*, but not upon the form of the curve γ . Indeed, the curve γ is not arbitrary: It must lie tangentially in the field of the instrument. However, that will in no way prevent us from concluding that the first variation of the curve integral over $(\Phi - F)$ must vanish identically and that the expression $(\Phi - F)$ will itself be the complete differential of a function $\varphi(x, y, z)$ then. We can then write:

(14)
$$\Phi - F = \psi_x \dot{x} + \psi_y \dot{y} + \psi_z \dot{z} .$$

9. – If the medium of the object space \Re is isotropic then the function *F* will possess the form:

(15)
$$F(x, y, z, \dot{x}, \dot{y}, \dot{z}) = f(x, y, z)\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}.$$

In this case, the equation F = 1 represents a sphere in the space of \dot{x} , \dot{y} , \dot{z} for fixed x, y, z. If \Re is crystalline then one must replace the function (15) with a more complicated one, in such a way that the Fresnel ray surface can be represented by the equation F = 1 in the space of \dot{x} , \dot{y} , \dot{z} (¹). However, in all cases, we have the relation:

(16)
$$F(x, y, z, -\dot{x}, -\dot{y}, -\dot{z}) = F(x, y, z, \dot{x}, \dot{y}, \dot{z})$$

[The relation (16) would no longer be true only when the object space \Re was found to be under the influence of an appreciable magnetic field.]

In exactly the same way, we can assume that the same identity also exists for the function F_1 . However, from equations (10), (11), and (12), we can then write:

(17)
$$\Phi(x, y, z, -\dot{x}, -\dot{y}, -\dot{z}) = \Phi(x, y, z, \dot{x}, \dot{y}, \dot{z})$$

If we then replace the quantities \dot{x} , \dot{y} , \dot{z} with $-\dot{x}$, $-\dot{y}$, $-\dot{z}$ in (14) then, due to (16) and (17), we will get:

$$\Phi - F = -(\psi_x \dot{x} + \psi_y \dot{y} + \psi_z \dot{z}),$$

^{(&}lt;sup>1</sup>) See, e.g., **P. Drude**, *Lehrbuch der Optik*, Leipzig, Hirzel, 1900, pp. 303.

and upon comparing the last equation with (14), we will get:

(18)
$$\Phi = F \,.$$

It follows from the latter equation that the optical lengths of the two curves are equal to each other for not only the curves γ that lie tangentially in the field, but also for any curve *C* at all that possesses an image C_1 :

Theorem 2:

For any absolute optical instrument, the optical length of a curve C whose points lie in the field of the instrument is equal to that of its image.

However, that is the generalization of the theorem of **Hamilton-Bruns-Klein** that we have in mind.

10. The stigmatic map of surfaces. – We would like to say of a two-dimensional piece S that it *lies tangentially in the field of an instrument J* when one can lay at least one light ray through every point P of S that first of all contacts the surface S and secondly goes through the instrument J.

We would now like to assume that *J* is not actually an absolute instrument, but that every point of the surface element *S* possesses a sharp point-like image. Now, let γ be an arbitrary curve segment that first of all lies on *S* and secondly lies tangentially in the field of our instrument. We let *A*, *B* denote the endpoints of γ , while *L* denotes the optical length of that curve segment, and *L*₁ denotes the optical length of its image. Precisely as in § 7, we can then prove that the following equation exists:

$$L_1 = L + \varphi(B) - \varphi(A) .$$

One can express the map between *S* and its image S_1 by saying that one can represent *S* and S_1 with the help of two parameters *u*, *v* in such a way that a point *P* of *S* and its image P_1 on S_1 will correspond to the same point in the parameter plane of *u*, *v*. The curve segments γ and γ_1 will then correspond to the same curve *C* in the *uv*-plane, and the optical lengths of those curve segments can be represented by curve integrals along *C*. We can then write:

$$L = \int_C \Phi(u, v, \dot{u}, \dot{v}) dt, \qquad L_1 = \int_C \Phi_1(u, v, \dot{u}, \dot{v}) dt$$

We can now conclude, as in § 8, that $(\Phi - \Phi_1)$ is a complete differential, so it has the form $(\chi_u \dot{u} + \chi_v \dot{v})$. It will then follow once more from:

that $\Phi = \Phi_1$. In other words, we have the:

Theorem 3:

If a surface patch S that lies tangentially in the field of an instrument J is mapped point-bypoint to a surface patch S_1 in the image space then any arbitrary curve on S will have the same optical length as its image on S_1 . The two surface patches S and S_1 can then be developed (optically) into each other.

11. – That last result, which seems to be new, is all the more remarkable because it is entirely linked with the condition that *S* lies tangentially in the field of *J*. Namely, it has been known for a long time (¹) that one can connect the rays of the object space with the rays in image space in such a way that first of all the **Malus** condition is satisfied and secondly that two given surfaces (that do not, however, lie tangentially in the field) will be mapped to each other in an entirely arbitrary, but stigmatic way. It is therefore necessary to investigate the basis for that apparent discrepancy. In that way, for greater clarity, we would then like to assume that the image and object spaces are homogeneous and isotropic.

We once more let S and S_1 denote the two surfaces that are to be mapped stigmatically to each other and define that map itself when we, in turn, establish that every point inside of a certain region in the *uv*-plane will be associated with two corresponding points S and S_1 .

Therefore, we assume that a light ray goes through the point (19) in the object space and defines direction cosines p, q, r with the positive axes, and that after it goes through the instrument, it will go to a ray in the image space that includes the point (20) and subtends the direction cosines p_1 , q_1 , r_1 with the positive axes of an axis-cross in image space. In that way, the quantities p_1 , q_1 , r_1 will be functions of p, q, r, u, v that can be calculated explicitly with the help of Malus's theorem, which has been known for a long time.

12. – To that end, we let $\varphi(u, v)$ denote the optical distance from the point (19) of *S* to its image point (20) and let *n* (*n*₁, resp.) denote the indices of refraction in the two spaces \Re and \Re_1 . We further consider the two points:

(21) $X = x + \lambda \cdot p$, $Y = y + \lambda \cdot q$, $Z = z + \lambda \cdot r$,

^{(&}lt;sup>1</sup>) See, e.g., Bruns, *loc. cit.*, pp. 371-375. Indeed, **E. Abbe** has occasionally asserted that a stigmatic map of two surface segments to each other is only approximately possible (Ges. Abh., Bd. 1, pp. 216), but already in 1890, **M. Thiessen** aptly remarked that Abbe's assertion will arise from switching the two different angles [Berl. Sitz. **2** (1890), pp. 812].

(22)
$$X_1 = x_1 + \lambda_1 p_1$$
, $Y_1 = y_1 + \lambda_1 q_1$, $Z_1 = z_1 + \lambda_1 r_1$

on a light ray that goes through the instrument, in which λ and λ_1 mean two parameters. The optical distance ρ between the two points (21) and (22), the first of which lies in object space and the second of which lies in image space, is now given by the equation:

(23)
$$\rho = \varphi(u, v) + n_1 \lambda_1 - n \lambda.$$

We shall now replace p, q, r, λ with arbitrary functions of u and v and determine λ_1 by the condition that the quantity ρ in (23) should be a constant. The coordinates of the points (21) and (22) will then be certain functions of u, v, and those points themselves will describe certain surfaces \mathfrak{F} and \mathfrak{F}_1 . Malus's theorem now says that whenever the functions of u, v that we have replaced p, q, r, λ with have the property that the normals to \mathfrak{F} at each point possess the components p, q, r, at the same time, the normals to Φ_1 must have the components p_1 , q_1 , r_1 . In other words, the relation that $\sum p_1 dX_1 = 0$ must follow from $\sum p dX = 0$.

Now, when one considers the relations:

(24)
$$p^2 + q^2 + r^2 = 1$$
 and $p \, dp + q \, dq + r \, dr = 0$,

one will get:

$$\sum p \, dX = \sum p \, dx + \lambda \sum p \, dp + d\lambda \sum p^2 = \sum p \, dx + d\lambda.$$

The condition that $\sum p \, dX = 0$ is then equivalent to the relation:

$$d\lambda = -(p \, dx + q \, dy + r \, dz) \,,$$

and one will likewise find that the condition $\sum p_1 dX_1 = 0$ is equivalent to the relation:

$$d\lambda_1 = -(p_1 \, dx_1 + q_1 \, dy_1 + r_1 \, dz_1) \, .$$

Finally, when one sets $\rho = \text{const.}$, it will follow from (23) that:

$$d\varphi + n_1 \, d\lambda_1 - n \, d\lambda = 0 \, .$$

Malus's theorem is then equivalent to the following relation:

$$n_1 (p_1 dx_1 + q_1 dy_1 + r_1 dz_1) = n (p dx + q dy + r dz) + d\varphi.$$

However, that equation is only an abbreviation for the following two:

(25)
$$\begin{cases} n_1 \left(p_1 \frac{\partial x_1}{\partial u} + q_1 \frac{\partial y_1}{\partial u} + r_1 \frac{\partial z_1}{\partial u} \right) = n \left(p \frac{\partial x}{\partial u} + q \frac{\partial y}{\partial u} + r \frac{\partial z}{\partial u} \right) + \frac{\partial \varphi}{\partial u}, \\ n_1 \left(p_1 \frac{\partial x_1}{\partial v} + q_1 \frac{\partial y_1}{\partial v} + r_1 \frac{\partial z_1}{\partial v} \right) = n \left(p \frac{\partial x}{\partial v} + q \frac{\partial y}{\partial v} + r \frac{\partial z}{\partial v} \right) + \frac{\partial \varphi}{\partial v}, \end{cases}$$

and together with:

(26)
$$p_1^2 + q_1^2 + r_1^2 = 1$$

they will allow us to calculate the quantities p_1, q_1, r_1 as functions of p, q, r, u, v.

13. – In order to grasp the geometric consequences of equations (25), we would like to choose the parameters u, v and the two axis-crosses x, y, z and x_1 , y_1 , z_1 in such a way that those equations will take on the simplest-possible form for a certain pair of corresponding points. In regard to that, we remark that it is known that two mutually-perpendicular line elements can be found at each point A of S that will be mapped to mutually-orthogonal line elements of S_1 . We can then assume from the outset, with no loss of generality, that the parameter curves u = const. and v = const.intersect perpendicularly on *both* the surfaces S and S_1 . With that, we can choose the x and y-axes to be parallel to the directions of the two parameter curves at a point A of S and assume that the axis-cross of x_1 , y_1 , z_1 has a corresponding position with respect to the parameters of S_1 at the image point A_1 of A. The eight quantities:

$$\frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}, \frac{\partial x}{\partial v}, \frac{\partial z}{\partial v}, \frac{\partial z}{\partial v}, \frac{\partial y_1}{\partial u}, \frac{\partial z_1}{\partial u}, \frac{\partial x_1}{\partial v}, \frac{\partial z_1}{\partial v}$$

will then vanish in (25), and those equations will assume the simple form:

(27)
$$\alpha p_1 = p + a, \qquad \beta q_1 = q + b.$$

One easily convinces oneself that the parameters α , β mean the magnification ratios of the two surfaces (as measured in light time) in the directions of the curves v = const. (u = const., resp.) and that a and b are proportional to the first derivatives of $\varphi(u, v)$.

14. – Equations (27) make it very easy for one to exhibit the conditions for the rays in the object and image spaces that correspond to each other to be real. Namely, in order for the ray with the direction components p_1 , q_1 , r_1 to be real, equation (26) must be fulfilled, from which it will follow that $p_1^2 + q_1^2 < 1$, or due to (27):

(28)
$$\frac{(p+a)^2}{\alpha^2} + \frac{(q+b)^2}{\beta^2} < 1.$$

One likewise finds that one must have:

(29)
$$p^2 + q^2 < 1$$
.

The instrument in question will then allow light rays to pass through at best when the ellipse in the pq-plane whose surface is defined by (28) has interior points in common with the circle (29).

The light rays that go through the instrument and simultaneously contact both surfaces S and S₁ at the mutually-corresponding points A and A₁ are associated with points in the *pq*-plane that simultaneously lie on the boundaries of the surface patches (28) and (29). Thus, whenever one does have a = b = 0 and $\alpha = \beta = 1$ simultaneously, there will be only at most *four* such rays. However, it can happen that no rays of that sort exist.

For a pair of conjugate aplanatic points on the axis of a rotationally-symmetric instrument, e.g., due to symmetry, the two ovals (28) and (29) must be concentric circles that therefore possess no real point of intersection. In that case, one must have: a = b = 0 and $\alpha = \beta \neq 1$. In place of equations (27), one will then have:

$$(30) \qquad \qquad \alpha p_1 = p , \qquad \alpha q_1 = q ,$$

i.e., equations from which the famous sine law of E. Abbe will follow immediately.

15. – We are now in a position to completely comprehend the connection between our Theorem 3 and the known results on the stigmatic maps between two surfaces *S* and *S*₁. Namely, if *S* lies tangentially in the field of the instrument (§ 9) then there will be *infinitely-many* rays through any point *A* of *S* that simultaneously contact *S* and *S*₁. From the previous section, the ellipse (28) must be identical to the unit circle (29) then, from which it will follow that a = b = 0 and $\alpha = \beta = 1$.

Consistent with the result in § 10, one infers that the derivatives φ_u and φ_v vanish and that $\varphi(u, v) = \varphi(A)$ is constant. However, the second condition $\alpha = \beta = 1$ says that the magnification ratio, as measured in light time, is equal to one for two mutually-perpendicular directions, and therefore for any possible direction. However, with that we have once more proved Theorem 3 for isotropic and homogeneous object and image spaces with the help of the theory of the eikonal.

16. The stigmatic map of isotropic spaces. – Under the assumption that the two media in object space \Re and image space \Re_1 are isotropic, but not necessarily homogeneous, Theorem 2 of § 9 will lead to some remarkable consequences. Namely, if one lets f(x, y, z) and $f_1(x_1, y_1, z_1)$ denote the indices of refraction of the two spaces at two points that correspond to each other by

means of the stigmatic map, as in § 9, and lets ds and ds_1 denote two corresponding line elements of \Re and \Re_1 at those same points then it will follow from our mapping theorem that:

$$f_1(x_1, y_1, z_1) ds_1 = f(x, y, z) ds$$
.

The ratio ds_1 : ds of the line elements that get mapped to each other is then *independent of their directions* at each point, from which it will follow with no further analysis that the stigmatic map of two spaces to each other must be *conformal*.

Now, there is a known theorem of differential geometry that was first found and proved by **Liouville** (¹), according to which every conformal map between three-dimensional regions will be identical to either a collineation that transforms every figure into a similar one or a transformation by reciprocal radii or a transformation that is composed of the two. We then have the:

Theorem 4:

Any stigmatic map of two isotropic spaces to each other that is produced by an absolute optical instrument is either a similarity transformation or a transformation by reciprocal radii or a transformation that can be represented by a transformation by reciprocal radii, followed by a similarity transformation.

17. – The Maxwell fisheye (§§ 3 and 4) is an example of a stigmatic map, as would follow from the last theorem. One can easily show that the map of space to itself that is accomplished by the fisheye is the *only* map for which every point in the space at infinity \Re (with the exception of the center *O*) possesses a single sharp image. That is because among the transformations that were enumerated in Theorem 4, there are no other ones that are involutory (i.e., under which the image of A_1 is, in turn, A), and possess no double points.

However, it would be a mistake for one to conclude from that alone that the law for the index of refraction that produces such a stigmatic map must necessarily satisfy equation (1) of § **3**. Namely, from the form of the map of space to itself, one can only conclude that the light rays must be closed curves that are transformed into themselves by the stated map, but not, as one might imagine, be able to conclude that the index of refraction (1) of § **3** is the only one for which all of space will be mapped stigmatically to itself. The "fisheye problem" is the adaptation to three-dimensional space of a question that **W. Blaschke** had posed for closed surfaces, but which has still not been answered (²).

18. – When one observes that the image of a light ray that lies in the field of the instrument coincides with an elongation of the light ray itself, one will see that due to Theorem 4, the light

^{(&}lt;sup>1</sup>) Note VI in the 5th edition of **Monge** *Feuilles d'Analyse appliquées à la géomtérie*, Paris, 1850, that **Liouville** edited. See also **F. Klein**, *Einleitung in die höhere Geometrie*, autogr. lectures, Göttingen 1892-93, pp. 378, *et seq*.

^{(&}lt;sup>2</sup>) **W. Blaschke**, *Vorlesungen über Differentialgeometrie*, I, Berlin, Springer, 1921, 1st ed., § 86, pp. 155-158.

rays on image space must be circles of straight lines when the light rays in object space have that property. An application of that argument is the following one:

One assumes that object space is homogeneous and isotropic, as usual. However, one can try to enforce a stigmatic map in such a way that one assumes that the image space is isotropic but has a varying index of refraction. The following theorem shows how little there is to be gained by that, and it follows immediately from the foregoing arguments and the properties of the transformation by reciprocal radii that:

Theorem 5:

If the object space is homogeneous and isotropic then in order for a stigmatic map to be possible at all, it is necessary that the image space either has the same property or it exhibits a distribution of refractive power such that all light rays that go through it have the form of circles that all go through one and the same point in space.

19. Application to the calculus of variations. – It is almost self-evident that the proofs in §§ **7-10** can be adapted directly to arbitrary symmetric variational problems in spaces of arbitrarilymany dimensions. In that, we are calling a variational problem *symmetric* when the value of the curve integral:

$$\int F(x_i, \dot{x}_i) dt$$

is independent of the sense in which one performs the integration over the given curve, which will be the case if and only if the relation:

$$F(x_i, -\dot{x}_i) = F(x_i, \dot{x}_i)$$

exists identically.

In order to adapt our theorems, we must assume that we have two "mutually-coupled" variational problems, i.e., that we know a canonical transformation between the canonical variables of the two variational problems under which one of those variational problems will go to the other one $(^1)$.

It is known that the extremals of the two variational problems will be in one-to-one correspondence with each other under that coupling. Now, if the extremals in the first problem that go through a point A in space \Re go to extremals in the second problem that all cross at one and the

same point A_1 in the space \Re_1 under that association, and if that will always be the case as long as

A is found on a two-dimensional surface S that lies "tangentially to the field of the coupling" then all of the assumptions will be fulfilled that it takes to prove a theorem that is so completely analogous to our Theorem 3 in § 10 that we do not need to even state it.

^{(&}lt;sup>1</sup>) See, e.g., **Riemann-Weber**, *Differential- und Integralgleichungen der Mechanik und Physik*, 7th ed., Braunschweig, Vieweg, 1925, pp. 198.

Similar theorems also seem to be true when one considers coupled symmetric variational problems with differential equations as auxiliary conditions. However, the relationships in that case are more complicated, and for that reason, I shall be content to merely suggest that possibility.