

A FUNDAMENTAL CLASS OF GEODESICS ON ANY CLOSED SURFACE OF GENUS GREATER THAN ONE*

BY

HAROLD MARSTON MORSE

INTRODUCTION

1. The study of geodesics on closed surfaces is influenced largely by the genus of the surface. Of closed surfaces of genus zero an important class has been studied by H. Poincaré,[†] namely, those closed surfaces which are everywhere convex. Of the surfaces with genus one, the torus has been studied by G. A. Bliss,[‡] and it seems probable that the types of geodesics found there will be found among the geodesics on any closed surfaces of genus one. It is the object of this paper to consider geodesics on any closed surface of genus greater than one.

In studying geodesics on a surface of negative curvature, the author[§] found that, of those geodesics which if extended indefinitely in either sense remained in a finite part of space, any particular one could be characterized in a manner which depended only upon a succession of fundamental contours of the surface. Now these surfaces of negative curvature are never closed. The question accordingly arose, is it possible also on closed surfaces, to characterize the geodesics in terms of the topographical elements of the surface, that is to characterize the geodesics of any particular closed surface in terms which would serve likewise for any other closed surface of the same genus? The answer to this question was readily seen to be no. However, on surfaces of genus $p > 1$, there appeared a fundamental class of geodesics which could be identified once and for all for all closed surfaces of the same genus. This paper is devoted to the definition and study of such a fundamental class of geodesics.

The surface is first mapped upon a hyperbolic non-euclidean plane. In the theory as developed each non-euclidean straight line represents a different

* Presented to the Society, September 8, 1921.

[†] *Sur les lignes géodésiques des surfaces convexes*, these Transactions, vol. 6 (1905), p. 237.

[‡] *The geodesic lines on the anchor ring*, Annals of Mathematics, ser. 2, vol. 4 (1903), p. 1.

[§] *Recurrent geodesics on a surface of negative curvature*, these Transactions, vol. 22 (1921), p. 84.

type of fundamental geodesic and it is with the aid of this representation that the fundamental class of geodesics is studied.

THE SURFACE

2. The surface defined. The surfaces to be considered are to be two-sided, closed surfaces of finite genus p , p greater than unity. They are to be without singularities. More specifically we will suppose that the points in the neighborhood of any point of the surface can be put into one to one continuous correspondence with the points in the neighborhood of some point in a plane, in such a manner, that for the neighborhoods considered, the cartesian coördinates, x , y , and z , of a point of the surface, be continuous functions of the cartesian coördinates, u , v , of the plane, provided with continuous partial derivatives up to the fourth order, while further

$$\left[\frac{D(xy)}{D(uv)} \right]^2 + \left[\frac{D(xz)}{D(uv)} \right]^2 + \left[\frac{D(yz)}{D(uv)} \right]^2 \neq 0.$$

The given surface will eventually be mapped upon the interior of a unit circle, in a manner that will be one to one and continuous as far as the neighborhoods of any two corresponding points are concerned, but by a correspondence that will be one to infinity as a whole. With that end in view there will now be defined a special group of linear transformations of a complex variable carrying the interior of the unit circle into itself. The method of definition will be similar to the method used by Poincaré in the article cited below.* From the "fundamental domain" of this group will be formed a canonical surface which can be mapped on the interior of the unit circle in the desired manner, and with the aid of which the given surface can also be so mapped.

3. The polygon S_0 . Let there be given a unit circle. In the plane of this circle let there be drawn a second circle concentric with the unit circle and with a radius r greater than one. Let p be a positive integer greater than one. On the circle of radius r let there be placed $4p$ equidistant points. With these points as centers let there be drawn $4p$ circles orthogonal to the unit circle. Now let r and the $4p$ circles vary, the $4p$ circles still remaining orthogonal to the unit circle and still having their centers equidistant on the circle of radius r . As r becomes arbitrarily large the $4p$ circles will approach straight lines passing through the center of the unit circle, and such that each makes an angle of $\pi - 2\pi/4p$ with its successor or predecessor. For very large values of r there will thus be formed a curvilinear polygon, lying

* *Théorie des groupes fuchsien*s, Acta Mathematica, vol. 1 (1882), p. 1.

within the unit circle, containing the center of the unit circle, and absolutely symmetrical with respect to the center of the unit circle (cf. Fig. 1). If r now decrease from very large values to small values, for some value of r the $4p$ circles will become tangent to each other. That is, the interior angles of the curvilinear polygon will diminish from $\pi - 2\pi/4p$ to zero. But for $p > 1$

$$\pi - \frac{2\pi}{4p} > \frac{2\pi}{4p} = \frac{\pi}{2p}$$

so that for a properly chosen value of r the curvilinear polygon will have interior angles of the magnitude of $\pi/2p$. The sides of this polygon will all be segments of circles orthogonal to the unit circle and will be equal in length. The polygon will contain the center of the unit circle and be absolutely symmetrical with respect to that center. The sum of the interior angles of the polygon will be 2π . This polygon will be denoted by S_0 and the corresponding value of r by r_0 . (Cf. Fig. 1, for the case $p = 2$.)

A property of S_0 to be used later is that *among the $4p$ circles bounding S_0 , alternating ones do not meet*. Suppose they did meet.

r and the common radius of the $4p$ circles should then be increased, the $4p$ circles still remaining orthogonal to the unit circle, the alternating circles would still continue to meet. Now the interior angles of S_0 have a magnitude $\pi/2p$ which is less than $\pi/2$ for $p > 1$. On the other hand if r should be increased beyond all limit the interior angles of the polygon thereby formed, say S' , would approach $\pi - 2\pi/4p$ which is greater than $\pi/2$ for $p > 1$. Thus for some value of $r > r_0$ the interior angles of S' would be right angles. Thus of any three successive circles bounding S' , the first and last would be orthogonal to the second and at the same time would be orthogonal to the unit circle. This is impossible. For two circles orthogonal to two non-tangent circles which meet, cannot themselves meet.

4. A group with principal circle. The polygon S_0 will be used to define a "group with principal circle."* The $4p$ sides of S_0 will be taken in one of their two circular orders and labelled as follows:

$$a_1 b_1 c_1 d_1 a_2 b_2 c_2 d_2, \dots, a_p b_p c_p d_p.$$

* H. Poincaré, *Théorie des groupes fuchsien*s, loc. cit. Also L. R. Ford, *An introduction to the theory of automorphic functions*, Edinburgh Mathematical Tracts.

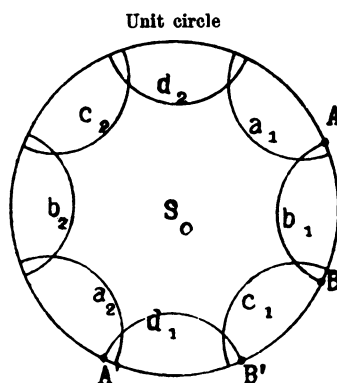


Fig. 1.

No confusion need arise if the circles bearing these sides be indicated by the same letters. We will term a_k conjugate to c_k , and b_k conjugate to d_k , where k ranges over the integers from 1 to p . Now a reflection of the plane in the radical axis of a_k and c_k will carry S_0 into itself and carry a_k into its conjugate side c_k . On the other hand a reflection of the plane in c_k will carry c_k into itself and S_0 into an adjacent polygon bordering S_0 along c_k . The product of these two reflections will be a transformation expressible as a linear transformation of a complex variable, and will carry the side a_k into the side c_k , and S_0 into a polygon bordering S_0 along c_k . Similarly there exist linear transformations of a complex variable which will carry any one of the sides of S_0 into its conjugate, and S_0 into a polygon bordering S_0 along the second of these conjugate sides. These transformations and all possible combinations of them will form a group G . All transformations of this group will carry the unit circle into itself. Points and regions which are the images of each other under this group will be termed *congruent*.

It is here necessary to make a convention, that just one side of each pair of conjugate sides of S_0 shall be considered as belonging to S_0 , and just one of S_0 's vertices. With this understood it may be stated that the set of all curvilinear polygons congruent to S_0 cover the interior of the unit circle once and only once. About each of the vertices of each of the polygons there will be grouped $4p$ polygons each with one corner at the given vertex. Each of these corner regions at such vertices will be congruent to just one corner in S_0 , and conversely each corner of S_0 will be congruent to just one corner at each such vertex. For the proofs of the results of this paragraph the reader is referred to the literature on the subject.

NATURE OF THE FUNDAMENTAL GROUP

5. Distances between congruent points. The circles orthogonal to the unit circle may be looked upon as representing the straight lines of a hyperbolic non-euclidean (written NE) geometry in which the points of the hyperbolic plane consist of the points interior to the unit circle. If we suppose the unit circle has its center at the origin of the euclidean plane, the NE length of any rectifiable curve interior to the unit circle is taken as the value of the integral

$$\int \frac{2 ds}{1 - x^2 - y^2},$$

taken along the given curve, and the NE distance between any two points within the unit circle as the NE length of the NE straight line joining the two given points. It is possible to prove directly that the NE length and

distance as defined by the above integral remain unchanged under any linear transformation of the complex variable that carries the interior of the unit circle into itself.

The NE distance between any two congruent points has a lower limit not zero.

It follows from the explicit nature (cf. § 4) of the transformations which carry S_0 into the polygons which are contiguous to S_0 , that in a sufficiently small neighborhood of any point of S_0 , or of its boundary, there are no pairs of mutually congruent points. Now if the lemma were false, it would be possible to pick out an enumerable infinity of pairs of congruent points whose NE distances would approach zero as the enumerating integer n became infinite. Each of these pairs could be replaced by a congruent pair of which one member at least might be taken so as to lie in S_0 or on its boundary. NE distances between the two points of any pair would be the same as before. These points in S_0 would have at least one limit point in S_0 or on its boundary. In every neighborhood of this limit point there could be found points which would be mutually congruent. This, as just stated, is impossible and the lemma is proved.

6. Nature of the group. *The group can contain no elliptic transformations.* An elliptic transformation carrying the unit circle into itself would have one fixed point within the unit circle and one without. Now there are in every neighborhood of a fixed point of an elliptic transformation points which are mutually congruent under the transformation. This is here impossible according to the result of the preceding section.

The group can contain no parabolic transformations.

To establish this fact it will be convenient to carry the interior of the unit circle into the upper half plane by means of a linear transformation of the complex variable. The integral

$$\int \frac{2 ds}{1-x^2-y^2}$$

will go over into the integral

$$\int \frac{ds}{y}.$$

If now the points above the axis of reals be considered as the points of a hyperbolic NE plane, the circles orthogonal to the axis of reals as the straight lines of this geometry, and the value of the integral $\int (1/y) ds$ taken along any curve segment in the upper half plane as the NE length of that curve segment, then any curve segment within the unit circle may be considered as

unchanged in N E length by the transformation from the unit circle to the upper half plane.

Corresponding to the fundamental group carrying the unit circle into itself there will exist a transformed group carrying the upper half plane into itself. If the first group had a parabolic transformation the transformed group would have one also. Now any parabolic transformation, say T , which carried the axis of reals into itself would have its fixed point on the axis of reals. Suppose now the upper half plane be carried into itself and thereby the group be still further transformed in such a manner that the fixed point of the parabolic transformation T goes into the point at infinity. N E distances would thereby be unchanged but the original parabolic transformation would now be transformed into a translation parallel to the axis of reals through a euclidean distance, say d . A point, no matter how remote from the axis of reals, would still be carried by this parabolic transformation through this euclidean distance d . If sufficiently remote, however, from the axis of reals, the N E distance between two such congruent points is seen with the aid of the integral

$$\int \frac{ds}{y}$$

to be arbitrarily small. This is contrary to the result of the preceding section. Thus the group can have no parabolic transformations.

Now the only linear transformations of a complex variable which carry the interior of a circle into itself are either elliptic, parabolic, or hyperbolic. *Hence all the transformations of the group are of the hyperbolic type.*

7. A lemma on polygons near the unit circle. The N E distance between any two points of S_0 has an upper limit. The same upper limit necessarily holds for distances between any two points on any one of the congruent polygons. Accordingly a polygon which has any point sufficiently near the unit circle will be arbitrarily small in its euclidean dimensions, for otherwise its N E dimensions as measured by the N E integral

$$\int \frac{2ds}{1-x^2-y^2}$$

would become arbitrarily large. *The polygons accordingly cluster about each point of the unit circle.*

The following lemma will now be proved.

LEMMA. *If a polygon lie sufficiently near the unit circle, all of its bounding circles, with the possible exception of one of them, will be arbitrarily small.*

Denote by O the center of the unit circle. Let A and B be two points on the unit circle at the extremities of a diameter. Let d be a small positive constant. Consider all the radii through O , except those that make an angle with OA numerically less than d . Now let O move along AB towards B . At the same time let each of the radii not excepted be replaced by an arc of a circle orthogonal to the unit circle, and drawn from O to the unit circle so as to make the same angle with OA as the radius to be replaced. Denote this pencil of circles through O by C . If O move along AB to a point sufficiently near B , it is obvious that the circles of the pencil C just described will become arbitrarily small.

To prove the lemma consider the $8p$ radii which pass from the center O to the points on the unit circle where the circles bounding the polygon S_0 (cf. § 3) meet the unit circle. Denote this set of radii by R . Let d , the constant of the preceding paragraph, here be a positive quantity less than half the magnitude of the angle between any two successive radii of the set R . From the result of the preceding paragraph it follows that if O and S_0 be carried by a transformation of the group respectively into a point O' and a new polygon sufficiently near the unit circle, then the set of circular segments, say C' , into which the radii of the set R are transformed will all be arbitrarily small *except* any of them which make an angle with $O'O$ of magnitude less than d . But from the choice of d there will be at most one circle of C' making an angle with $O'O$ of magnitude less than d . Now the circles bounding the transformed polygon and the circles of the set C' will intersect the unit circle in the same points, and accordingly, with the possible exception of one of them, each of the circles bounding the transformed polygon will meet the unit circle in two arbitrarily near points, and being orthogonal to the unit circle will consequently be arbitrarily small.

8. Fixed points of the transformations of the group. The distribution of the fixed points of the transformations of the group will now be considered.

(a). Let T be a transformation of the group. T is hyperbolic, and accordingly has two fixed points on the unit circle. Let A and B be two points on the unit circle not the fixed points of T , and let A' and B' be respectively the images of A and B under T . It follows from the nature of a hyperbolic transformation that in case the fixed points of T separate A and B , $ABA'B'$ will appear on the unit circle *in the circular order* $ABB'A'$, and that in the contrary case they will appear *in the order* $ABA'B'$ or else *in the order* $AA'BB'$.

(b). Consider any two conjugate circles of the circles bounding the polygon S_0 . Suppose the first of these two circles meets the unit circle in the points A and B and the second of the two circles meets the unit circle in the points A' and B'

where A is congruent to A' and B to B' (cf. Fig. 1). According to the result of § 3 no two conjugate circles intersect. A reference to the explicit description of the transformations which carry any circle bounding S_0 into its conjugate (cf. § 4) will now show that $ABB'A'$ appear on the unit circle in the order $ABB'A'$. After subjection to any other transformation of the group the circular order of these four points will still be the same.

With the aid of the results of the preceding paragraphs (a) and (b) the following lemma can be proved. This lemma will be used to show that among the geodesics considered the types that correspond to closed geodesics are everywhere dense in a sense to be explained later.

LEMMA. *There exists a transformation of the group which has fixed points arbitrarily near the end points of any preassigned arc of the unit circle.*

Let M_1 and M_2 be the end points of a preassigned arc of the unit circle. Let e be an arbitrarily small positive constant less than half the distance between M_1 and M_2 . According to the lemma of § 7, it is possible to choose a polygon S_1 so near M_1 that all of its bounding circles with the possible exception of one of them will lie within e of M_1 . Let S_2 be a second polygon similarly chosen relative to M_2 . Since $p > 1$ and the number of pairs of conjugate sides of a polygon is $2p$, S_1 and S_2 each have at least four pairs of conjugate sides, and on both S_1 and S_2 at least three of these pairs of conjugate sides lie entirely within e of M_1 and M_2 respectively. It is accordingly possible to choose a side of S_1 such that its intersections with the unit circle, say A_1 and B_1 , and the intersections of its conjugate side with the unit circle, say A'_1 and B'_1 , lie within e of M_1 , while at the same time the four points $A_2 B_2 A'_2 B'_2$ corresponding to $A_1 B_1 A'_1 B'_1$ under the transformation of the group that carries S_1 into S_2 also lie within e of M_2 .

According to the choice of e , the e neighborhood of M_1 is entirely distinct from the e neighborhood of M_2 . Accordingly $A_1 B_1 A_2 B_2$ have either the circular order $A_1 B_1 B_2 A_2$ or else the circular order $A_1 B_1 A_2 B_2$. In the first case it follows from the result of paragraph (a) that the transformation of the group that carries $A_1 B_1$ into $A_2 B_2$ will have one of its fixed points on the arbitrarily small arc between A_1 and B_1 and the other fixed point on the small arc between A_2 and B_2 , and the theorem is proved for that case.

Consider the second case. Here $A_1 B_1 A_2 B_2$ appear in the order $A_1 B_1 A_2 B_2$. According to the result of paragraph (b) $A_1 B_1 B'_1 A'_1$ always appear in the circular order $A_1 B_1 B'_1 A'_1$. It follows that in this second case $A_2 B_2 B'_1 A'_1$ appear in the order $A_2 B_2 B'_1 A'_1$. Using again the result of paragraph (a), the transformation of the group which carries $A_2 B_2$ into $A'_1 B'_1$ is seen to be one which has one fixed point on the small arc between A_2 and B_2 and the other fixed point on the small arc between A'_1 and B'_1 , and the lemma is proved for the second case as well.

THE GIVEN SURFACE MAPPED ON THE UNIT CIRCLE

9. Definition of the correspondence. Let a closed surface be formed from S_0 by continuously deforming S_0 so as to bring congruent points of conjugate sides into coincidence. S_0 has $2p$ pairs of conjugate sides. Hence the resulting closed surface, say T , will be of genus p . Each point P of T will arise from one point P_0 of S_0 . The correspondence between T and S_0 can be extended over the interior of the unit circle by requiring that P shall not only correspond to P_0 , but also to all points of the unit circle congruent to P_0 . The correspondence thereby established between the interior of the unit circle and T is everywhere continuous and one to one as far as the neighborhoods of corresponding points are concerned, but one to infinity as a whole.

Now by a fundamental theorem of analysis situs any two closed two-sided surfaces of the same genus can be put into one to one continuous correspondence. The original surface, say Σ (cf. § 2), can then be put into one to one continuous correspondence with that one of the above surfaces T which has the same genus as does Σ . The surface Σ will thereby be mapped on the interior of the unit circle in a manner essentially the same as the manner in which T has just been mapped upon the interior of the unit circle.

10. Surface distances and NE distances. In contradistinction to NE lengths and distances, surface lengths and distances will now be defined.

DEFINITION. If a curve segment h within the unit circle corresponds on the original surface to a curve segment k that is rectifiable, the ordinary length of k will be termed the *surface length* of h . Before defining the surface distance between two points within the unit circle the following lemma is needed. *Of the curves within the unit circle joining any two non-coincident points, there is at least one which corresponds on the original surface to a geodesic segment whose surface length is at least as small as that of any other curve joining the two given points.* This lemma is essentially equivalent for the case of geodesics to Hilbert's *a priori* existence theorem of the calculus of variations.* A method which can be used without any essential change to prove this lemma is given by G. D. Birkhoff† for a similar problem in the article cited below.

DEFINITION. The lower limit of the surface lengths of curves joining two given points within the unit circle will be termed the *surface distance* between these two points. The surface distance between two given points will be denoted by D_g while the NE distance between the same two points will be denoted by D_n .

* Cf. Bolza, *Vorlesungen über Variationsrechnung*, 1909, p. 428.

† *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), § 9, p. 219.

LEMMA 1. *Corresponding to a positive constant a there exists a positive constant b such that any pair of points interior to the unit circle for which*

$$D_g \leq a$$

will be such that

$$D_n < b.$$

Any pair of points whose surface distance is at most a can be replaced by a congruent pair of which one member lies on S_0 and the other on the closed region consisting of all points whose surface distance from points of S_0 and its boundary is at most a . These new pairs will moreover have the same surface distances as the original pairs. But the NE distances between any two points lying on a closed region are all less than a properly chosen constant which we may take for b . The lemma thus is proved.

LEMMA 2. *There exist constants A and m such that for any pair of points interior to the unit circle, if either*

$$D_n > A,$$

or

$$D_g > A,$$

then

$$D_n < m D_g.$$

Let there be given two points whose surface distance is D_g . Starting from one end of a geodesic of surface length D_g joining the two given points, let there be marked off segments of surface length equal to the constant a of Lemma 1. The line will thereby be divided into not more than

$$\frac{D_g}{a} + 1$$

parts. According to the preceding lemma there is between the end points of each of these parts an NE distance less than the constant b . The NE distance between the end points of the geodesic is accordingly less than

$$b \left[\frac{D_g}{a} + 1 \right].$$

That is,

$$(1) \quad D_n < \frac{b}{a} \cdot D_g + b.$$

If then

$$(2) \quad D_g > b$$

(1) becomes

$$D_n < \frac{b}{a} \cdot D_g + D_g$$

or

$$(3) \quad D_n < \left(\frac{b}{a} + 1 \right) D_g.$$

Setting

$$\frac{b}{a} + 1 = m$$

we have

$$(4) \quad D_n < m \cdot D_g$$

subject to the condition (2). Now (1) holds unconditionally, and shows that D_g becomes infinite with D_n , so that the condition (2) will be fulfilled if for a sufficiently large constant c

$$(5) \quad D_n > c.$$

As seen now (4) holds if either (2) or (5) holds. Taking A as the larger of the constants b and c , and m as the constant m of (4) the inequalities of the lemma accordingly follow.

It may be observed that there is a large degree of reciprocity between D_n and D_g . In fact the two preceding lemmas and their proofs will remain true if the terms NE distance and surface distance be interchanged together with D_n and D_g . Lemma 2 and the lemma obtained by interchanging D_n and D_g in Lemma 2 may be combined into the following fundamental lemma:

LEMMA 3. *If either the NE distance or the surface distance between two points exceeds a properly chosen constant, then the NE distance lies between two positive constant multiples of the surface distance, and the surface distance lies between two positive constant multiples of the NE distance, where the constants involved are fixed once and for all for the whole unit circle and surface.*

11. Proof of Lemma 7. The question arises, do there exist geodesic segments which recede arbitrarily far from the NE straight lines joining their end points and which still have a surface length as small as that of any other geodesics joining their end points? This question is answered in Lemma 8. This section is devoted to proving Lemma 7 which is the principal aid in proving Lemma 8.

A slight modification of the proof of Lemma 2 will serve to establish the following lemma. The constants b and m of the following lemma are the constants b and m used in Lemmas 1 and 2.

LEMMA 4. *There exist constants b and m such that, if a geodesic segment has a surface length g for which $g > b$, then the end points of this segment can be joined by a curve all of whose points lie within an NE distance b of the given geodesic segment, and whose NE length, v , is such that*

$$v < mg.$$

LEMMA 5. *Corresponding to a positive constant μ there exists a positive constant R so large that if all the points of a curve segment h are at an NE distance R from some NE straight line, L , and the end points of h project through NE perpendiculars into the end points of a segment of L , say k , of more than unit length, then h has an NE length exceeding μ times the NE distance between its end points.*

The locus of points at a constant NE distance from L is a circle meeting the unit circle in the same points as does L . The segment h lies on this circle.

To simplify the proof let the interior of the unit circle be carried into the upper half plane by a linear transformation which carries the end points of L respectively into the origin, O , and the point at infinity. This can be done,

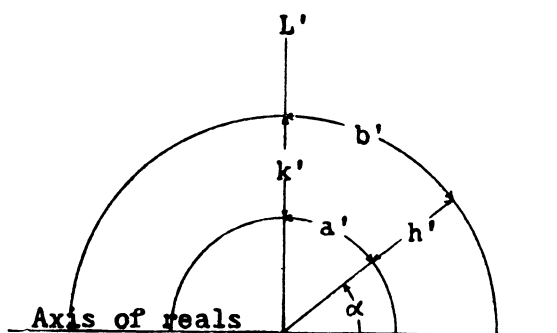


Fig. 2.

as in § 6, without altering any NE distances. L will go into an ordinary straight line, L' , perpendicular to the axis of reals at the origin (see Fig. 2). The transform of k , say k' , will lie on this line. The circle on which h lies will go into an ordinary straight line through the origin making an angle with the axis of reals which we

denote by α , measuring the angle from the axis of reals in the counterclockwise direction. We may suppose further that $\alpha < \pi/2$, since that could be brought about without altering any NE distances by a reflection of the upper half plane in L' . The transform of h , say h' , will lie on this line, making an angle α with the axis of reals. The NE perpendiculars from the end points of h to those of k will go respectively into two arcs, say a' and b' , of two circles both with center at the origin and leading from the end points of h' to those of k' . To compute the NE length of a' , b' , h' , and k' , suppose the

plane referred to polar coördinates ϱ and φ with pole at the origin. Denote by ϱ_1 the value of ϱ at the lower end point of k' , and by ϱ_2 its value at the upper end point. The polar coördinates of the end points of h' will be $(\varrho_1 \alpha)$ and $(\varrho_2 \alpha)$. The NE length of k' as evaluated through the integral $\int (ds/y)$ is

$$(1) \quad \int_{\varrho_1}^{\varrho_2} \frac{d\varrho}{\varrho} = \log \frac{\varrho_2}{\varrho_1}.$$

The NE length of h' , along which φ equals the constant α , is

$$(2) \quad \int_{\varrho_1}^{\varrho_2} \frac{d\varrho}{\varrho \sin \alpha} = \log \frac{\varrho_2}{\varrho_1} \cdot \frac{1}{\sin \alpha}.$$

The NE lengths of a' and b' are the same, namely

$$(3) \quad \int_{\alpha}^{\frac{\pi}{2}} \frac{\varrho_1 d\varphi}{\varrho_1 \sin \varphi} = \log \cot \frac{\alpha}{2}, \quad 0 < \alpha < \frac{\pi}{2}.$$

It is desired to show that if the end points of h' be joined by the broken line $a' k' b'$, there results a curve whose NE length has a ratio to the NE length of h' that approaches zero as R becomes infinite. The ratio of the NE length of the broken line $a' k' b'$ to that of h' is

$$\frac{2 \log \cot \frac{\alpha}{2} + \log \frac{\varrho_2}{\varrho_1}}{\log \frac{\varrho_2}{\varrho_1}} \sin \alpha = \left[\frac{2 \log \cot \frac{\alpha}{2}}{\log \frac{\varrho_2}{\varrho_1}} + 1 \right] \sin \alpha.$$

Now by hypothesis the NE length of k' , namely $\log (\varrho_2/\varrho_1)$, is greater than one, so that the above ratio is less than

$$(4) \quad \left(2 \log \cot \frac{\alpha}{2} + 1 \right) \cdot \sin \alpha.$$

R , the length of a' or b' , is given by (3) and becomes infinite only if α approaches zero. But as α approaches zero the expression (4) approaches

zero, so that for R sufficiently large the NE length of h' will certainly exceed μ times the NE length of the broken line $a'k'b'$ joining the end points of h' , and hence the NE length of h' will exceed μ times the shortest NE distance between the end points of h' , as was to be proved.

DEFINITION. Let k be any curve segment within the unit circle. The NE distance between the end points of k will be denoted by $D_n(k)$, and the surface distance between the end points of k will be denoted by $D_g(k)$.

LEMMA 6. *Corresponding to positive constants b and m there exists a positive constant r so large that if a curve segment, v , has its end points at an NE distance r from some NE straight line L , if further all of v 's points are at an NE distance from L exceeding $r - b$, and if v projects into a segment of L whose NE length exceeds unity, then the NE length of v exceeds m times the NE distance between v 's end points. (See Fig. 3.)*

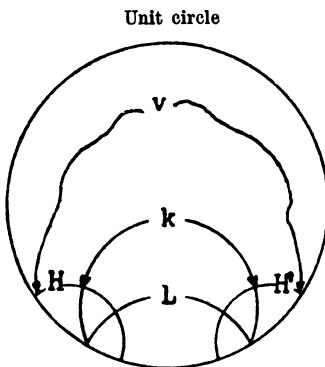


Fig. 3.

Denote by k a curve segment projecting through NE perpendiculars into the same segment on L as does v , and consisting further of points all at an NE distance $r - b$ from L . Let the NE length of v and k be denoted also by v and k respectively. We seek to prove that, for r sufficiently large,

$$(1) \quad v > m \cdot D_n(v).$$

To that end we shall obtain inequalities connecting v with k , k with $D_n(k)$, $D_n(k)$ with $D_n(v)$, and by combining these inequalities obtain the inequality (1).

By the aid of the construction of the preceding proof it is seen that

$$v > k.$$

Taking the constant μ of the preceding lemma as here equal to $m + 1$ it follows from that lemma that if $r - b$, and hence r , be sufficiently large,

$$k > D_n(k) \cdot (m + 1).$$

The two ends of v project into the two ends of k through NE perpendiculars to L , say H and H' respectively. Now the end points of v can be joined through the mediation of k by joining the first end point of v to the first end point of k , tracing out that segment of H that lies between these two

points, following this segment by tracing out the NE line joining the end points of k and proceeding finally to the second end point of v by tracing out that portion of H' which lies between the second end points of v and k . In accordance with one of the hypotheses of this lemma the portions of H and H' thereby traced out are each of NE length b , so that the whole broken curve thus traced out will have an NE length $2b + D_n(k)$. This length will be at the least $D_n(v)$, since $D_n(v)$ stands for the shortest NE distance between the end points of v . Thus

$$2b + D_n(k) \geq D_n(v) \text{ or } D_n(k) \geq D_n(v) - 2b.$$

Combining the preceding inequalities we have

$$(2) \quad v > (m+1)[D_n(v) - 2b] = (m+1)D_n(v) - (m+1)2b.$$

Now as r becomes infinite $D_n(v)$ will do likewise, and, for all cases in which the segment v projects on L into a segment of at least unit NE length, $D_n(v)$ will become infinite uniformly with r . Suppose r so large therefore that

$$(3) \quad D_n(v) > (m+1)2b.$$

Adding (2) and (3) we have

$$v > m \cdot D_n(v)$$

with the condition simply that r be sufficiently large. The proof is accordingly complete.

LEMMA 7. *There exists a constant r such that if all the points of a segment of a geodesic g are at an NE distance at least r from some NE straight line L , if further the end points of the geodesic segment g are exactly at an NE distance r from L , and if g projects in the NE sense into a segment of L of more than unit NE length, then the surface length g of the given geodesic exceeds the surface distance between its end points.*

It is required to prove that for r sufficiently large

$$(1) \quad g > D_g(g).$$

Now as r becomes infinite $D_n(g)$ will also become infinite. It follows from Lemma 3 that $D_g(g)$ will do likewise, so that according to Lemma 4 there exist positive constants b and \bar{m} such that if r exceed a properly chosen

positive constant the end points of g can be joined by a curve v whose points are each within an N E distance b of g and whose N E length v is such that

$$\bar{m}g > v.$$

We now have a relation between g and v . We shall obtain a relation between v and $D_n(v)$, and a relation between $D_n(v)$ and $D_g(v)$. Combining these relations (1) will be obtained.

If r exceed a second constant so large that $D_n(v)$ exceeds the constant required for the application of Lemma 3, we will have

$$D_n(v) > A D_g(v),$$

where A is a positive constant. If further r exceed the positive constant required for the application of Lemma 6 for the case where the m of Lemma 6 equals \bar{m}/A , then in addition to the preceding inequalities there will also result the inequality

$$v > \frac{\bar{m}}{A} D_n(v).$$

If the above inequalities be combined there results the inequality

$$g > D_g(v),$$

subject simply to the condition that r exceed a properly chosen constant.

A FUNDAMENTAL CLASS OF GEODESICS

12. Geodesics of Class A. DEFINITION. A geodesic segment h will be said to be of *Class A*, if, on the original surface, h is at least as short as any other rectifiable curve joining h 's end points and capable of being continuously deformed on the surface into h without moving its end points. An image of a geodesic segment of Class A on the hyperbolic plane for simplicity also will be termed a geodesic segment of Class A. The definition of geodesic segments of Class A reduces in the hyperbolic plane to the following. An image of a geodesic segment within the unit circle is of Class A if its surface length is at least as small as that of any curve joining its end points and possessing a surface length. An *unending geodesic* will be said to be of Class A if each of its finite segments is of Class A.

LEMMA 8. *There exists a positive constant R , so large, that within the unit circle, no segment of a geodesic of Class A can recede an NE distance greater than R from the NE straight line joining its end points.*

Denote the given geodesic segment by g , and the NE straight line joining its end points by L . Consider the constant r whose existence is affirmed in Lemma 7. If g has points at more than r NE units from L , let P' be such a point. Let P and Q be the first points following and preceding P' , respectively, that are at an NE distance r from L . If the segment PQ projects through NE perpendiculars into a segment on L of more than unit NE length, it follows from Lemma 7 that the surface length of the segment PQ exceeds the surface distance between P and Q contrary to the hypothesis that the given geodesic segment is of Class A.

If on the other hand the segment PQ projects into a segment on L of unit NE length or less than unit NE length, while P and Q are at an NE distance r from L , the NE distance between P and Q will be at most $2r + 1$, and hence the surface distance between P and Q will in all cases be inferior to some positive constant, say d . Since g is of Class A the surface distance between P' and P is at most that between P and Q , or at most d . It follows from Lemma 3 that the NE distance of P' from P is always less than some constant, say q . Hence P' is in all cases at most at NE distance $r + q$ from L , and the theorem is proved.

13. Unending curves of the same type. A definition will presently be given of what is meant by two unending curves being of the same type. This definition will be an extension of the definition of what is meant by two closed curves being mutually deformable. We begin accordingly by considering closed curves. A curve continuously deformable into a point either on the given surface or on the unit circle, will have an image deformable into a point on the other. A closed curve, say h , not deformable into a point on the given surface, if traced out just once will have as images on the unit circle an infinite number of curve segments whose end points are congruent, non-coincident points. Let A and B be respectively the end points of one of these images.

A is congruent to B under a certain transformation T . The set of all points within the unit circle congruent to A under T and its inverse all lie on a circle through the fixed points of T and cluster in the neighborhood of these fixed points. The set of all segments congruent to AB under T and its inverse join up these points congruent to A so as to form a single curve passing from one fixed point of T to the other. An unending curve such as this, consisting of an unending succession of segments all congruent under the same transformation of the group, will be called a *periodic curve*. Traced out in either sense such a curve corresponds on the original surface to a closed curve traced out an infinite number of times.

It follows from the fact that the correspondence between the surface and unit circle is one to one and continuous as far as the neighborhood of corresponding points is concerned, that if the given closed curve, h , be continuously deformed on the surface, then the image AB will vary continuously through corresponding points, A and B always remaining congruent under the same transformation. At the same time the variation of the complete set of points congruent to A under T and its inverse is one in which the circle on which these points lie turns about the fixed points of T while the points themselves vary continuously on the circle in such a manner as to remain always congruent under T . Thus two closed curves which on the original surface are not deformable into points, but which are mutually deformable into each other, will correspond within the unit circle, among other images, to pairs of periodic curves with end points at the same points on the unit circle. For the sake of completeness the following is added. Two curves which correspond to two periodic curves within the unit circle with the same end points on the unit circle and with successive portions congruent under the *same* transformation of the fundamental group, are mutually deformable on the original surface.

DEFINITION. Two unending curves within the unit circle will be said to be of the *same type* if there exists a positive constant C such that every point of either curve has an NE distance from some point of the other less than C . Two unending curves *on the original surface* will be said to be of the *same type* if among their respective images on the unit circle there are at least two curves of the same type.

Two circular arcs within the unit circle are of the same type if their end points on the unit circle are the same, and of different types if either of their end points are different. Every circular arc with two end points on the unit circle is thus of the type of the NE straight line with the same two end points on the unit circle. *Only those curves will be considered which are of the types of NE straight lines.* There is thereby included a type for every periodic curve, that is, curves which correspond on the original surface to closed curves. Future developments will still further show the inclusiveness of types of curves as restricted to NE straight lines. From Lemma 3, it follows that NE distance and surface distances become infinite together uniformly. Accordingly if two unending curves are of the same type there exists a constant D such that every point of either curve has a surface distance less than D from some point on the other curve. It is not difficult to show that if on the original surface two curves are closed and mutually deformable they have for their images on the unit circle an infinite number of pairs of periodic curves of the same type. It is in this sense that *the property of two curves being of the same type is an extension of the property of two curves being*

mutually deformable. The definition of what is meant by two curves being of the same type will be given in the following paragraph more in the spirit of analysis situs. The statements of the following paragraph will not be proved as they will not be used later.

Let the given surface, $p > 1$, be cut so as to form a single simply-connected piece S . Let there be provided an enumerable infinity of copies of S each spread over the original surface in exactly the position in which they were before the surface was cut. Without altering the position of any of these copies of S , these copies of S can be successively joined in a manner analogous to the way in which the congruent polygons of the unit circle were joined, so as to "heal" up all of the cuts and to form an infinitely-sheeted, simply-connected, unbounded surface, say M , the neighborhood of each point of which will consist on M of an exact copy of the neighborhood of the point which is overhung on the original uncut surface. Two curves of M could now be defined as being of the same type if there exists a positive integer n , so large, that any point of either curve could be joined to some point of the other curve by a curve which has points in at most n of the copies of S making up M . From this definition it could be shown that *two curves of the same type on the given surface will correspond, on any surface which corresponds to the given surface in a one to one continuous manner, to two curves again of the same type*.

14. Existence of geodesics of Class A. An image on the unit circle of a geodesic on the original surface will be spoken of as a geodesic on the unit circle. By an *element* is meant a point and a direction through the point. By an element on a geodesic on the surface is meant a point on the geodesic together with a direction tangent to the given geodesic at the given point. By an *element on a geodesic on the unit circle* is meant a point P within the unit circle on such a geodesic, together with the direction of a tangent to the original geodesic at the point on the original surface corresponding to P . In either case an element may be represented by the coördinates x, y of its initial point in the unit circle, together with the direction cosines a, b, c of the direction tangent to the original geodesic on the surface. The element may be considered as represented by a point $(xyabc)$ in five dimensions. A set of elements will be said to have a *limit element* $(x_0 y_0 a_0 b_0 c_0)$ if the corresponding points in five dimensions have a limit point $(x_0 y_0 a_0 b_0 c_0)$. With these conventions the following definition is given.

DEFINITION. A geodesic G will be said to be a *limit geodesic* of a set of geodesics M not containing G , if every element on G is a limit element of elements on geodesics of the set M .

An element will be said to be on an NE straight line in the unit circle if its initial point P is on the NE straight line, and if its direction is the direction of a tangent to the NE straight line at P .

LEMMA 9. *If a geodesic is a limit geodesic of a set of geodesics of Class A, it is itself of Class A.*

Let H be an image of the limit geodesic within the unit circle. If H were not of Class A, it follows from the definition of geodesics of Class A, that H has some segment, h , whose end points can be joined by a curve k whose surface length k is less than the surface length h of the segment h . Set

$$k = h - e.$$

On the other hand, if the given geodesic is a limit geodesic of geodesics of Class A, it follows, from the property of continuous variation of a finite geodesic segment with its initial point and direction, that there exists a geodesic segment m , of Class A, with end points within $e/4$ of those of h , and with a surface length m , such that

$$m > h - \frac{e}{2}.$$

Now the end points of m can be joined through the mediation of k by first passing from one end point of m to the nearer end point of k , passing thereby along a curve which can be taken so as to have a surface length less than $e/4$; thence passing along k to k 's other end point, thereby passing over an additional surface length of $h - e$; finally passing from the second end point of k to the second end point of m along a curve taken so as to have a surface length less than $e/4$; in all passing over a curve of surface length less than $h - (e/2)$, that is, over a length less than that of m . This is impossible if m be of Class A. Thus the lemma is proved.

THEOREM 1. *Corresponding to any NE straight line, there exists at least one geodesic of Class A of the same type. Conversely, every geodesic of Class A is of the type of some one NE straight line.*

Let there be given an NE straight line L . Let

$$\cdots P_{-2} P_{-1} P_0 P_1 P_2 \cdots$$

be an unending succession of points on L which follow each other on L in the order of their subscripts, and which with increasing subscripts approach one end point of L on the unit circle, and with decreasing subscripts approach the other end point of L . Within the unit circle let P_{-n} be joined to P_n by a geodesic of Class A, say g_n . According to Lemma 8, g_n will have at least one point, say Q_n , within an NE distance R of P_0 . Let E_n be the geodesic

element on g_n at Q_n . The elements E_n will have at least one limit element, E . Let G be the geodesic within the unit circle bearing the element E . G will be shown to be of Class A, and to be a curve of the type of the given NE straight line.

If any finite segment of G be given, it follows from the definition of G , that there can be found among the segments g_n , one which possesses a sub-segment lying arbitrarily near the given segment of G , not only in position but also in direction and length. A repetition of the proof of the preceding lemma will suffice to show that the given segment of G is of Class A, as was to be proved.

Every point of the segments g_n lies within an NE distance R of some point of L (cf. Lemma 8). All the limit points of the points of the segments g_n , including the points of G , accordingly lie at most at an NE distance R from points of L . G thus lies in a region bounded by the two circles at an NE distance of R from L . Any two such circles and L meet the unit circle in the same two points. G passes from one of these points to the other. From this it is obvious that each point of L , also, lies at most at a distance R from some point of G , and G is thereby proved to be of the type of L .

To prove the converse proposition let G be a geodesic of Class A, within the unit circle, and let

$$\dots P_{-2} P_{-1} P_0 P_1 P_2 \dots$$

be a set of points appearing on G in the order of their subscripts, and so chosen that their surface distances from P_0 increase beyond all limit as n becomes infinite in absolute value. Denote by h_n an NE straight line joining P_n to P_{-n} . At least one point of h_n , say Q_n , will be at an NE distance from P_0 not greater than the constant R of Lemma 8. Denote by E_n , the element on h_n at the point Q_n . The elements E_n will have at least one limit element, say E . The NE straight line, L , through E , will be shown to be of the type of G .

Let P be any point on G . Let m be the smallest value of n such that all segments $P_{-n} P_n$ for which $n > m$ include the point P . For $n > m$ there is a point A_n on each h_n which is within an NE distance R (cf. Lemma 8) of the point P . But as n becomes infinite the points A_n will have at least one cluster point which, according to the definition of L , will lie on L . This cluster point will be at most at an NE distance R from the given point P of G . G thus lies within the band consisting of all points within an NE distance R of L . G and L will accordingly be of the same type, and the theorem is completely proved.

As just stated, a geodesic of Class A lies within a band consisting of all points within an NE distance R of the NE straight line of the same type. This constant R is the constant of Lemma 8, and is entirely independent of

the particular geodesic of Class A that is considered. Combining this fact with the fact given in Lemma 3, that N E distances and surface distances become infinite together uniformly for the whole unit circle, we have the following theorem:

THEOREM 2. *There exists a positive constant D , fixed for all types of geodesics of Class A, such that no geodesic of Class A, and N E straight line of the same type, can recede either a surface distance or an N E distance exceeding D from each other.*

15. Intersecting and asymptotic geodesics of Class A.

THEOREM 3. *Two geodesics of Class A can intersect at most once within the unit circle.*

If two such curves, G and H , intersected twice in two points, A and B , the portions, g and h , respectively, of G and H , between the points A and B , would have the same surface length. For if, for example, h were shorter than g , G could not be of Class A, as follows directly from the definition of curves of Class A. But g and h cannot have the same surface length. For consider a segment k of G large enough to contain the segment g as an interior segment. If the surface length of g equaled that of h , k would be unaltered in surface length if its subsegment g be replaced by the segment h . But the curve so obtained would have corners at A and B , and so could be shortened, contrary to the fact that k is of Class A, and hence could not be shortened. Thus the theorem is proved.

A geodesic of Class A within the unit circle passes from one end point of the N E straight line of the same type, to the other such end. Hence the theorem:

THEOREM 4. *If two N E straight lines have both end points distinct, the corresponding geodesics of Class A of the same type cross or do not cross each other within the unit circle, according to whether or not the given N E lines cross or do not cross each other.*

If two N E straight lines have one end point on the unit circle in common, the minimum N E distance from a point P on either to a point on the other approaches zero as the point P approaches the common end point. The N E straight lines are termed *asymptotic*. A direct result of Theorem 2 is accordingly

THEOREM 5. *If two geodesics of Class A are respectively of the types of two asymptotic N E straight lines, as they approach the common end point on the unit circle of the given N E straight lines, they come and remain within a finite surface distance of one another.*

THEOREM 6. *No two geodesics of Class A which are asymptotic can intersect within the unit circle.*

Let a and b be two geodesics of Class A which are asymptotic.

Choose as the positive senses of a and b the senses in which a and b are asymptotic. Suppose the theorem were false and that a and b did intersect in a point O (cf. Fig. 4). Now any broken geodesic which contained any one of the corners formed by a and b together with a segment of a or b on either side of O of at least unit surface length could be replaced by a curve joining its end points and with a surface length smaller by at least a positive constant e .

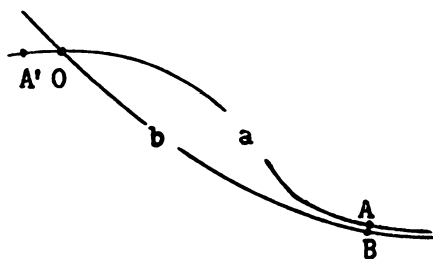


Fig. 4.

By hypothesis a and b are asymptotic. Accordingly it will be possible to find two points A and B , on a and b , so remote in the positive senses of a and b that they can be joined by a curve whose surface length AB is less than $e/4$:

$$AB < \frac{e}{4}.$$

Observe that the surface length of the segment AO of a cannot differ from that of BO on b by $e/4$ or more, for otherwise the larger of the two segments AO and BO , say AO , could be replaced by a broken curve of shorter surface length, namely ABO . Hence, in particular,

$$BO < AO + \frac{e}{4}.$$

Now let A' be any point on a , at least a unit surface distance beyond O in the negative sense. The segment AOA' of a can be replaced by a segment of shorter surface length. For adding to the two preceding inequalities the equality

$$OA' = OA',$$

we have

$$ABOA' < AOA' + \frac{e}{2}.$$

But the broken curve, $ABOA'$, contains a corner formed by a and b at O , so that it can be shortened by at least the constant e without altering its end points. The new curve will still join the end points of AOA' , and will now in addition have a surface length less than that of AOA' . This is impossible if a is of Class A. The theorem thus is proved.

We have already defined what is meant by a limit geodesic of a class of geodesics (cf. § 14).

DEFINITION. We here define an NE straight line L to be a *limit line* of a set of NE straight lines M not including L , if every element of L is a limit element of elements on lines of the set M . A necessary and sufficient condition that L be a limit line of the set M is that there be in the set M lines whose end points on the unit circle lie, in the euclidean sense, arbitrarily near those of L .

THEOREM 7. *If a set M of NE straight lines have an NE straight line L for a limit line, the geodesics of Class A of the type of the lines of M will have at least one geodesic of the type of L as a limit geodesic.*

Let A be any element on L . It follows from the hypothesis of the theorem that there exists a set of elements,

$$A_1 A_2 A_3 \cdots,$$

all on lines of M , and such that

$$\lim_{n \rightarrow \infty} A_n = A.$$

Let L_n be the NE straight line of M bearing A_n . Let C_n be a geodesic within the unit circle of Class A and of the type of L_n . According to Lemma 8, there exists on C_n at least one point, say P_n , within an NE distance R of the initial point of the element A_n . Let E_n be the geodesic element on C_n at P_n . The elements E_n will have at least one limit element, E . It follows from the result of Lemma 9, § 14, that the geodesic G , defined by E , is of Class A. It will now be proved that G is of the type of the NE straight line L .

It follows, from the property of continuous variation of a finite segment of a geodesic with its initial element, that every point P of G is a limit point, as n becomes infinite, of a properly chosen set of points on the geodesics C_n . But the latter points are each at an NE distance less than the constant R of Lemma 8 from some point of the corresponding NE straight line L_n . These points on the lines L_n will, however, as n becomes infinite, approach some point of L , say H , and this point H will be at most at an NE distance R from the given point P on G . Thus G is of the type of L as was to be proved.

THEOREM 8. *If there is more than one geodesic of Class A of a given type within the unit circle, there always exist two particular non-intersecting geodesics of Class A of the given type bounding a region outside of which no geodesics of Class A of the given type can pass.*

Let L be an NE straight line of the given type. Let M be a set of NE straight lines which have L as a limit line, which have no points in common

with L either within or on the unit circle, and which all lie on one side of L . Let N be a similar set of NE straight lines all on the other side of L . Let M' and N' be two sets of geodesics consisting of one geodesic of Class A of the type of each NE straight line of the sets M and N respectively. According to Theorem 7 the geodesics of the sets M' and N' each have at least one limit geodesic, say M_0 and N_0 , respectively, that is of Class A, and of the type of L . The NE straight lines of the sets M and N do not cross L or have any end points in common with L . It follows from Theorem 4 that all geodesics of Class A of the type of L will lie in the region between any geodesic of the set M' , and any second geodesic of the set N' . It follows further from the property of continuous variation of any finite segment of a geodesic with its initial element that the geodesics M_0 and N_0 , by virtue of the fact that they are respectively limit geodesics of the sets M' and N' , will, in case they are non-coincident, be non-intersecting, and have between them all geodesics, other than M_0 and N_0 , of Class A and of the type of L .

CLOSED GEODESICS

16. **The variation of the distance of an arbitrary geodesic V of Class A from a given type of periodic curve h .** Let there be given an NE straight line, h , made up of successive segments all congruent under a transformation, T , of the fundamental group. We have previously termed h a periodic curve. Let k be a second NE straight line crossing h . Let K be the geodesic of Class A of the type of k . The transformation T carries h into itself. Accordingly the successive images of k ,

$$\cdots k_{-2} k_{-1} k_0 k_1 k_2 \cdots,$$

under T and its inverse, will consist of a set of non-intersecting circles, all meeting h and shrinking down about the two ends of h , the fixed points of T . The successive images of K under T and its inverse, say

$$\cdots K_{-2} K_{-1} K_0 K_1 K_2 \cdots,$$

will accordingly not meet each other (cf. Theorem 4), and will, like the k_n , shrink down about the two fixed points of T . Now let there be given any NE straight line, v , and let V be a geodesic of Class A of the type of v . If v crosses any one of the k_n , V will cross the geodesic K_n with the same subscript.

Three cases are distinguished here. The end points of v on the unit circle may be distinct from those of the given periodic curve h . In this case v

crosses at most a finite number of the k_n . Or v may be asymptotic to h , in which case v will cross all of the k_n with subscripts exceeding a certain integer, or else less than a certain integer. Finally v may coincide with h , in which case v will cross all of the k_n . A study will now be made of the positions of the points of intersection of V with the different K_n .

Let A_0 be any point on K_0 . Let A_n on K_n be the image of A_0 under T , or its inverse. NE distances will be measured along each K_n from the corresponding A_n . The sense that corresponds to positive distances on each of the K_n will be chosen so as to be unchanged through application of T . Now if V intersects K_n , denote this point of intersection by P_n and let s_n be the surface distance of P_n from A_n , measured along K_n . A proof will be given that the s_n behave in one of the following ways: (1) *remain constant*, (2) *increase with their subscripts*, (3) *decrease with their subscripts*, (4) *increase to a maximum and thereafter decrease*, or (5) *decrease to a minimum and thereafter increase*. In the latter two cases the maximum or minimum may be taken on by a single s_n or by two successive s_n . The proof will be given under the heads A, B, C, D, E.

A. *If three successive s_n are equal all of the s_n are equal.* An application of the transformation T to V will carry V into a new geodesic of Class A, say V' . If in particular $s_1 = s_2 = s_3$, P_1 and P_2 on V would go into P_2 and P_3 on V' so that V and V' would intersect in P_2 and P_3 . According to Theorem 3, V must then be identical with V' . Thus each P_n will be carried by T into P_{n+1} , so that s_n equals s_{n+1} .

B. *If just two successive s_n are equal, the next preceding and the next following s_n , if they exist, are either both greater or both less than the equal s_n .*

Of the two regions into which V divides the unit circle, that region which a point enters upon being projected from V along any of the curves K_n in the positive sense of that particular K_n will be termed the region above V , and the other region the region below V . As before let V' be the image of V under the transformation T . Suppose now that $s_2 = s_3$ and that $s_1 > s_2$. To prove $s_3 < s_4$: V' will meet K_2 above V since $s_1 > s_2$. V' will meet K_3 in the same point as does V since $s_2 = s_3$, and at this point will pass from above to below V . Thereafter V' will remain below V , since it cannot cross V a second time. In particular V' will meet K_4 below V , so that $s_3 < s_4$ as was to be proved. Similarly it could be shown that if $s_2 = s_3$ and $s_1 < s_2$, then $s_3 > s_4$.

The preceding proof includes a proof of the following.

C. *If the s_n are equal for $n = m$ and $n = m - 1$, V and V' meet on K_m .*

D. *If the s_n have a relative maximum or minimum for $n = m$, then between K_m and K_{m+1} , V and V' cross each other.* More explicitly suppose that

s_2 is greater than both s_1 and s_3 . Then V' meets K_2 below V since $s_1 < s_2$, and meets K_3 above V since $s_2 > s_3$. Thus V' must cross V between K_2 and K_3 . The case of a minimum among the s_n is similarly treated.

E. *The s_n cannot have more than one extremum of any sort.* For if the s_n had more than one extremum it follows from C and D that V and V' would have more than one point of intersection. V would then have to be identical with V' . All the s_n would be equal in that case.

A proof can now be given that the s_n behave only in the ways enumerated. Any sequence of numbers such as the s_n are either (1) all equal, or (2) increase or decrease monotonically with n , or (3) have at least one extremum. The first case is admittedly possible. The second case can be achieved, according to A and B, only when no two of the s_n are equal, and then is admittedly possible.

The third case, according to E, can occur only when there is a single extremum among the s_n . In this case the extremum will be taken on by not more than two successive s_n , as follows from A. Further, no two of the remaining s_n are equal, for if just two of them were equal, it would follow from B that they would constitute another extremum among the s_n , contrary to E; and if three or more of the remaining s_n were equal all of the s_n would be equal and we would have the first case.

Thus all three cases reduce to the cases originally stated to be possible.

17. Periodic geodesics of Class A. DEFINITION. Of the transformations of the fundamental group which have the same fixed points there are two which will be called the corresponding *primitive transformations*. They are inverses of each other. Each is such that every other transformation of the group with the same fixed points is either its multiple or the multiple of its inverse. Such transformation exists. For every transformation of the group which has the two given points for fixed points carries the NE straight line through these points into itself. With the aid of NE distances measured along this line the existence of the primitive transformations can be proved in exactly the same way as the existence of a primitive period is proved in the theory of simply periodic functions.

THEOREM 9. *A geodesic of Class A which is carried into itself by a transformation T' is also carried into itself by the corresponding primitive transformations.*

Denote the given geodesic by V . We shall make use of the notation of the preceding section, identifying the V and T of this section with the V and T of that section. Now according to the result of the preceding paragraph, there exists a positive integer m , such that

$$T' = T^m$$

where T is one of the two primitive transformations. Accordingly, upon applying T^m times to V , V will be carried into itself. Hence among the s_n used in the preceding section those whose subscripts differ by m must be equal. But if all of the s_n were not then equal, the s_n would have an infinite number of maxima and minima, which has been shown to be impossible. The s_n are then all equal. The result of applying T to V must carry V into itself. For the geodesic into which V is carried by T would otherwise intersect V in each of the points at which the s_n were equal; according to Theorem 3, this is impossible. Thus T carries V into itself, as was to be proved.

THEOREM 10. *A geodesic, V , of Class A, which if traced out in one of its senses, say the positive sense, comes and remains at an NE distance less than an assignable positive constant from a periodic NE straight line L , is either a periodic geodesic of the type of L , or else is asymptotic in its positive sense to a geodesic that is itself periodic, and is of the type of L , and of Class A.*

L is made up of successive portions congruent under a primitive transformation, say T . We will again make use of the notation of the preceding section identifying the V , T and L of the preceding section with the V , T and L of this theorem. We suppose that the subscripts n of the s_n of the preceding section are here so chosen that they increase as V is traced out in its positive sense. If the s_n are all equal, V must be periodic. The case where the s_n are not all equal will be considered now.

According to the result of the preceding section, if the s_n are not all equal, then for values of n exceeding a properly chosen integer, say m , the s_n must either all increase with n , or else all decrease with n . It follows from the hypothesis that V traced out in its positive sense comes and remains within an assignable positive NE distance of L , that the s_n are bounded, for $n > m$. Hence as n becomes infinite the quantities s_n approach a limit, say b .

Let k be any positive integer greater than m . The transformation which carries the geodesics K_k of the preceding section into K_0 , carries L into itself, and V into a curve of Class A, say V_k , for which the corresponding s_n will be obtained from the s_n for V by advancing the subscripts by k . Thus the point of intersection of V_k with K_0 , say Q_k , will be at an NE distance s_{0+k} from the point A_0 of K_0 , from which NE distances along K_0 are being measured. As k becomes infinite the points Q_k will approach a point on K_0 , say Q , whose surface distance from A_0 will equal $\lim_{k \rightarrow \infty} s_k = b$. Let E_k be the element on V_k , at the point Q_k . The elements E_k will have one or more limit elements. Let E be one such limit element. The initial point of E will be Q . The geodesic, say G , defined by E will be of Class A.

The point of intersection of G with any particular K_n , say K_s , for example, will be the limit of the points of intersection of the geodesics V_k with that

particular geodesic K_s , as k becomes infinite; so that G will intersect K_s at a surface distance from A_s on K_s equal to $\lim_{k \rightarrow \infty} s_{s+k} = b$. Similarly G will meet all other K_n at a surface distance from the corresponding A_n equal to b . It follows that the transformation T will carry G into itself, so that G is a periodic geodesic.

It remains to be proved that V is asymptotic to G . Now E was defined as a limit element of the elements E_k . But the elements E_k can have no other limit element than E . For if the elements had another limit element E' , then E' , like E , would serve to define a second periodic geodesic, say G' , passing through the point Q . G' and G would intersect not only in Q but also in all the points congruent to Q under the transformation T . This is impossible for two geodesics of Class A. Hence $G' = G$, and $E' = E$, and the set E_k has no other limit element than E . This may be written

$$(1) \qquad \lim_{k \rightarrow \infty} E_k = E.$$

Now E_k is the initial element of the portion of V_k , say h_k , between K_0 and K_1 , while E is the initial element of the portion of G between K_0 and K_1 . Hence for k sufficiently large, h_k , according to (1), will lie arbitrarily near the portion of G between K_0 and K_1 . But the portion of V between K_k and K_{k+1} is, by definition of V_k , congruent to h_k , and accordingly for k sufficiently large lies arbitrarily near the portion of G between K_k and K_{k+1} . Thus V is asymptotic to G in its positive sense. The proof of the theorem is complete.

DEFINITION. The two non-intersecting geodesics of Class A of a given type which form the two boundary curves of the region outside of which no geodesics of Class A of the given type can pass, will be termed the *boundary geodesics* of that type (cf. Theorem 8). The name will be applied even in the special case where the two bounding geodesics are identical.

THEOREM 11. *The boundary geodesics of the type of a periodic NE straight line L are themselves periodic.*

Let T be a transformation of the group that carries L into itself. T carries L into itself, and hence every curve of the type of L into another such curve. Further, T , as well as every other transformation of the group, preserves surface distances, and hence carries every geodesic of Class A into a geodesic of Class A. Thus T carries a geodesic of Class A of the type of L into another such geodesic.

To prove the theorem, observe first that the two given boundary geodesics, say G_1 and G_2 , divide the unit circle into three non-overlapping regions, say S_1 , S_2 , and S , of which S lies between G_1 and G_2 , S_1 lies between G_1 and the unit circle, and S_2 lies between G_2 and the unit circle. S , in case G_1 is identical

with G_2 , will reduce simply to the geodesic $G_1 = G_2$. Now S_1 contains in its interior no geodesic of Class A of the type of L , but is, however, bounded within the unit circle by one such geodesic. These two properties of S_1 must be preserved in the region, say S' , into which S_1 is carried by the transformation T . For otherwise, upon applying the inverse of T to S' , we could conclude that the two properties attributed to S_1 did not really belong to S_1 . Now the only regions of the unit circle which have these properties of S_1 , and like S_1 are bounded in part by the unit circle, are the regions S_1 and S_2 . Hence the transformation T must carry S_1 either into S_1 , or else into S_2 . But T cannot carry S_1 into S_2 ; for that portion of the unit circle that bounds S_1 lies between the two fixed points of T , and under a hyperbolic transformation, such as T , is carried into the same portion of the unit circle. Thus S_1 goes into S_1 . Hence G_1 , as the boundary of S_1 within the unit circle, goes into G_1 . It follows similarly that G_2 is carried into G_2 by T , and the theorem is proved.

THEOREM 12. *If two closed geodesics of Class A are of the same type they are also of the same length.*

Since the two given closed geodesics are of the same type they will be represented on the unit circle by an infinite number of pairs of periodic curves, the members of each pair of the same type. Let C and D be the members of one such pair. Let T be a primitive transformation which carries C and D into themselves. Let c be a segment of C of which one end point A is carried into the other end point by T . Let d be a segment of D of which one end point B is carried into the other end point by T . Denote the surface lengths of c and d also by c and d , respectively.

Suppose now that the theorem were false, and that

$$d - c = e > 0.$$

Let h be the surface distance between A and B . Let n be a positive integer such that

$$n > \frac{2h}{e}.$$

Consider now a continuous portion of C , say c' , consisting of all of the images of c obtained by applying T^n times to c . Let d' be a similar piece of D obtained by applying T^n times to d . The end points of d' can be joined by a curve whose surface length is shorter than that of d' , by passing from the end point B of d' to the end point A of c' through a surface distance h , passing along c through a surface distance nc , thence returning to the second end point of d' through a surface distance h , in all over a path whose surface

length would be $2h + nc$ which according to the preceding inequalities would be less than nd , or the surface length of d' . This is impossible if D is of Class A. The theorem thus is proved.

18. Geodesics of Class A asymptotic to periodic geodesics of Class A.

THEOREM 13. *A geodesic a of Class A which is asymptotic to a periodic geodesic b of Class A within the unit circle cannot cross any periodic geodesic of Class A of the type of b .*

Suppose that a did cross a periodic geodesic c of Class A of the type of b , at a point P . (See Fig. 5.)

Let there be assigned to a , b , and c , positive senses that lead to their common end point on the unit circle. Let m be the length of the closed geodesic which on the original surface corresponds to b .

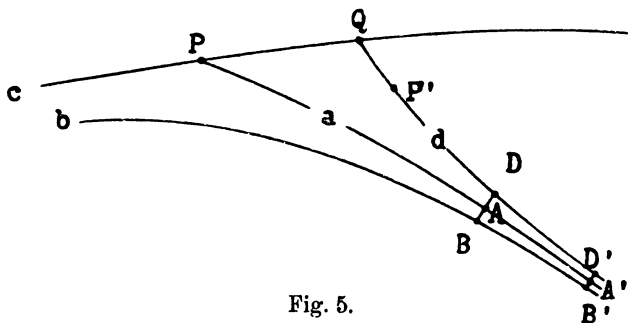


Fig. 5.

According to Theorem 12, the length of the closed geodesic which corresponds to c is also m . On c let there be laid off from P in c 's positive sense a segment whose surface length is m ; denote the second end point of this segment by Q . The transformation T which carries P into Q , and b and c into themselves, carries a into a geodesic d of Class A that is again asymptotic to b .

Let P' be any point on d , following Q , and neighboring Q . Any curve which contains the corner PQP' , formed at Q by c and d , can be shortened in surface length, without disturbing P and P' , by a positive constant, say e . Let B and B' be two points on b which are the two end points of a segment of b of surface length m , and which are so remote on b that two geodesics perpendicular to b respectively at B and B' meet the asymptotes a and d in points within a surface distance $e/4$ of B and B' respectively, and cut out of a and d respectively two segments whose surface lengths differ from m by less than $e/4$. Denote the point in which the geodesic perpendicular to b at B meets a by A , and the point at which it meets d by D , and denote the point in which the congruent geodesic perpendicular to b at B' meets a by A' , and d by D' .

We will now show that the segment PA on a can be replaced by a curve joining its end points, which has a smaller surface length. To that end we compare the surface length of PA with that of the broken curve $PQDA$. We shall use the letters PA to denote the surface length of PA measured along the geodesic a , and will extend this convention to all other segments to be

considered. Now since the segment PA is congruent to the segment QD' we have

$$(1) \quad PA = QD',$$

or

$$(2) \quad PA = QD + DD'.$$

Now, from the choice of B and B' ,

$$(3) \quad DD' > BB' - \frac{e}{4}.$$

Adding (2) and (3),

$$(4) \quad PA > QD + BB' - \frac{e}{4}.$$

Now

$$BB' = m = PQ.$$

Hence (4) becomes

$$(5) \quad PA > QD + PQ - \frac{e}{4}.$$

Further, from the choice of B ,

$$(6) \quad \frac{e}{4} > DA.$$

Adding (5) and (6)

$$PA > PQ + QD + DA - \frac{e}{2},$$

or

$$PA > PQDA - \frac{e}{2}.$$

Now the broken curve $PQDA$ contains the above mentioned corner at Q , and can be replaced by a curve joining its end points whose surface length will, according to the choice of e , be e less than that of $PQDA$. This new curve will accordingly join the end points of PA , and according to the last inequality be less in surface length than PA contrary to the fact that PA is a segment of Class A. Thus our assumption that a crosses a periodic geodesic of Class A of the type of b is impossible, and the theorem is proved.

One of the hypotheses of Theorem 10 requires that the geodesic V of that theorem be of Class A. However, what is essential in the proof of that

theorem can be carried over to the case where every segment on V from a certain point on is of Class A. In this way we get the following lemma:

LEMMA 10. *A portion of a geodesic which in one sense, say its positive sense, is unending, every segment of which portion is of Class A, and which in its positive sense comes and remains at an NE distance less than an assignable positive constant from some periodic geodesic b of Class A is either asymptotic in its positive sense to a periodic geodesic of the type of b , and of Class A, or is itself periodic of the type of b .*

There is a similar difference between Theorem 13 and the following lemma:

LEMMA 11. *A portion of a geodesic which in one sense from a certain point on is of Class A, and which is asymptotic in the given sense to a periodic geodesic b of Class A, cannot cross any periodic geodesic of Class A of the type of b .*

THEOREM 14. *If there exist two different periodic geodesics of Class A, namely, b and c , of the same type and between which there are no other periodic geodesics of Class A, then if to b and c there be assigned positive senses that lead on b and c to a common end point of b and c on the unit circle, there exist at least two geodesics of Class A lying in the region between b and c , of which one geodesic is asymptotic to b in b 's positive sense and to c in c 's negative sense, while the other geodesic is asymptotic to b in b 's negative sense and to c in c 's positive sense.*

Let $B_1 B_2 B_3 \dots$ be a set of points which follow one another on b in b 's positive sense, and whose surface distances from B_1 become infinite with the subscript n . Let $C_1 C_2 C_3 \dots$ be a set of points on c which follow each other on c in c 's negative sense and whose surface distances from C_1 become infinite with the subscript n . Let h_n be a segment of a geodesic of Class A joining B_n to C_n . Let P_n be a point on h_n whose surface distances from b and c are equal.

Let k be a segment of a geodesic of Class A joining some point of b to some point of c . Let T be a transformation of the fundamental group that carries b and c into themselves. (Cf. Theorem 9.) Under T , k will be carried into a second geodesic segment, say k' , also of Class A, and also joining a point of b to one of c . Between k and k' , b and c , there will be a region, say S , contained in the region, say S' , between b and c , and such that every point of S' is congruent under some multiple of T or its inverse to some point of S .

A proper multiple of T or its inverse will carry P_n into a point P'_n of S . The same transformation will carry h_n into a second geodesic segment of Class A, say h'_n , again joining a point of b to a point of c . Let E_n be the element on h'_n at the point P'_n . The elements E_n will have at least one limit element E . Denote the initial point of E by P . Let d be the geodesic defined by E . We

will show that d is asymptotic to b in b 's positive sense, and to c in c 's negative sense.

The h'_n all become infinite in surface length with n . (Cf. Lemma 3.) Accordingly d will become infinite in surface length, in at least one sense, without passing out of the region between b and c . That portion of d that lies between b and c , and contains the point P , will certainly be of Class A. For any of its finite segments between b and c permit an arbitrarily close approximation, both in position and in surface length, by a properly chosen segment of some one of the segments h'_n , which are themselves of Class A. (Cf. proof of Lemma 9, § 14.) According to Lemma 10, d traced out in that sense in which it is already proved to remain between b and c , becomes asymptotic to some periodic curve of Class A of the type of b and c . According to Lemma 11, d cannot then cross any periodic curve of Class A, of the type of b or c . In particular d cannot cross b or c , and hence remains between b and c . As stated before every segment of d which contains the point P , and remains between b and c , is of Class A. We can conclude therefore that every segment of d is of Class A.

Finally it would be impossible for d , as defined, to be asymptotic to b in b 's negative sense, and to c in c 's positive sense. For the methods of § 16, applied to any geodesic segment h_n , will show that if h_n be traced from its end point B_n to its end point C_n , its successive intersections with the geodesic segments congruent to k under T will recede from b and progress toward c , in the sense that their surface distances measured from b along the segments congruent to k will increase. Now according to the definition of d , there exists a geodesic segment h'_n , which, starting near the point P on d , will follow along d arbitrarily near d , for an arbitrarily long surface distance, measured on d from P , in either sense. Hence d can only be asymptotic to b in b 's positive sense, and to c in c 's negative sense.

The existence of a geodesic of Class A asymptotic to b in b 's negative sense and to c in c 's positive sense is similarly proved.

THEOREM 15. *It is impossible within the unit circle for a geodesic of Class A to be asymptotic in both of its senses to the same periodic geodesic of Class A.*

Suppose the theorem were false and that a geodesic b of Class A were asymptotic in both of its senses to a periodic geodesic c of Class A. Upon applying any transformation of the fundamental group which carries c into itself b would be carried into a second geodesic, say b' , of Class A also asymptotic in both of its senses to c . Now b and b' would intersect. For if b , for example, lay entirely within the region between b' and c the maximum NE distance of any of b 's points from c would be less than that of b' . This maximum, however, should be the same for both b and b' since they are congruent. Thus b would intersect b' contrary to the fact, as stated in Theorem 6,

that no two geodesics which are asymptotic can intersect. Thus the initial assumption of this proof is false, and the theorem is true.

19. Completion of results on periodic geodesics. Let L be any periodic NE straight line. Denote the set of all geodesics of Class A and of the type of L by M . The properties of the geodesics of the set M may be summed up and completed as follows:

If there is but one geodesic in the set M it is periodic. If there is more than one geodesic in M there are two non-intersecting geodesics of M , say b_1 and b_2 , called boundary geodesics of M , delimiting a region outside of which no geodesics of M can pass. The boundary geodesics are periodic. No periodic geodesics of M intersect. Between b_1 and b_2 there may be a finite or infinite number of periodic geodesics of M or none. If there exists a single point in the region between b_1 and b_2 through which there does not pass a periodic geodesic of M , there is at least one region bounded by two periodic geodesics of M , say p_1 and p_2 , between which there are no other periodic geodesics of M . Denote one end point of L on the unit circle by A , and the other end point by B . In the region between p_1 and p_2 there is at least one geodesic of M asymptotic to p_1 as A is approached, and asymptotic to p_2 as B is approached, and at least one geodesic of M asymptotic to p_1 as B is approached and to p_2 as A is approached. Any geodesic of M which is not periodic lies in a region between two periodic geodesics of M , between which there are no periodic geodesics of M . Any such non-periodic geodesic of M is always asymptotic in one sense to one of the periodic geodesics bounding the region in which it lies, and to the other such periodic geodesic in the other sense.

Geodesics of Class A, of the type of NE straight lines that cross L , cross all geodesics of the set M and recede indefinitely from all points of geodesics of M . Any geodesic, say c , of Class A, of the type of any one of the infinite number of NE straight lines asymptotic to L , lies wholly outside the region between b_1 and b_2 , and as A is approached becomes asymptotic to that one of the two geodesics, b_1 or b_2 , which separates c from the region between b_1 and b_2 . A similar statement may be made replacing the point A by the point B .

We conclude with a proof of the following theorem.

THEOREM 16. *The set of all periodic geodesics of Class A include among their limit geodesics all of the "boundary" geodesics. (Cf. § 17.)*

Let L be a given NE straight line. According to the lemma of § 8 there exists a transformation of the fundamental group, having its fixed points arbitrarily near the end points of any preassigned arc of the unit circle. The NE straight line passing through the fixed points of a transformation of the group is a periodic straight line. Accordingly it will be possible to choose a set of periodic NE straight lines, say M , which have L as a limit line,

which have no points in common with L either within or on the unit circle, and which all lie on one side of L . Let N be a similar set of periodic NE straight lines all on the other side of L . According to Theorem 11, § 17, the boundary geodesics of the type of any periodic NE straight line are themselves periodic. Accordingly, corresponding to each periodic NE straight line of the set M there exists at least one geodesic of Class A of the same type that is itself periodic; denote the set of such periodic geodesics by M' . Let N' be a similar set of periodic geodesics consisting of one periodic geodesic of Class A of the type of each NE straight line of the set N . Now exactly as was shown in the proof of Theorem 8, § 15, the geodesics of the sets M' and N' will have, the one, a limit geodesic M_0 , and the other, a limit geodesic N_0 , that are the two boundary geodesics of the geodesics of Class A of the type of L . Thus the theorem is proved.

CORNELL UNIVERSITY,
ITHACA, N. Y.
