# RECURRENT GEODESICS ON A SURFACE OF NEGATIVE CURVATURE*

BY

HAROLD MARSTON MORSE

## INTRODUCTION

The results necessary for the development of this paper are contained in a paper by G. D. Birkhoff,[†] in a paper by J. Hadamard,[‡] and in an earlier paper by the present writer.[§]

In this earlier paper, as in the present paper, only those geodesics on the given surfaces of negative curvature are considered which, if continued indefinitely in either sense, lie wholly in a finite portion of space. A class of curves is introduced, each of which consists of an unending succession of the curve segments by which the given surface, when rendered simply connected, is bounded. It is shown how a curve of this class can be chosen so as to uniquely characterize some geodesic lying wholly in a finite portion of space. Conversely, it is shown that every geodesic lying wholly in a finite portion of space, is uniquely characterized by some curve of the above class.

The results of the earlier paper on geodesics, and the representation obtained there, will be used in the present paper to establish various theorems concerning sets of geodesics and their limit geodesics. In particular, the existence of a class of geodesics called *recurrent geodesics of the discontinuous type*,[||] will be established. This class of geodesics offers the first proof that has been given in the general theory of dynamical systems, of the existence of recurrent motions of the discontinuous type.

For a more complete treatment of the questions of the existence of surfaces of negative curvature, the reader is referred to the paper by Hadamard, already cited.

---

* Presented to the Society, Dec. 28, 1920.

† *Quelques théorèmes sur le mouvement des systèmes dynamiques*, Bulletin de la Société Mathématique de France, vol. 40 (1912), p. 303.

‡ *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques pures et appliquées, (5), vol. 4 (1898), p. 27.

§ *A one to one representation of geodesics on a surface of negative curvature*, American Journal of Mathematics, vol. 42 (1920).

|| G. D. Birkhoff, loc. cit.

## THE SURFACE

§ 1. We will consider surfaces without singularities in finite space. We will suppose the surface divisible into overlapping regions, such that every point of the surface lying in a finite portion of space is contained as an interior point in some one of a finite number of these regions, and such that the Cartesian coördinates $x$, $y$, $z$ of the points of any one of these regions can be expressed in terms of two parameters, $u$ and $v$, by means of functions with continuous derivatives up to a convenient order, at least the third, and such that

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

By a *curve* on the surface we will understand any set of points on the surface in continuous correspondence with the points of an interval on a straight line, including one, both, or neither of its end points.

We will suppose the Gaussian curvature of the surface to be negative at every point, with the possible exception of a finite number of points, at which points the curvature will necessarily be zero. A first result, given by Hadamard in the paper already referred to, is that a surface of negative curvature cannot be contained in any finite portion of space.

§ 2. By a *funnel* of a surface will be meant a portion of a surface topographically equivalent to either one of the two surfaces obtained by cutting an unbounded circular cylinder by a plane perpendicular to its axis. We will consider surfaces of negative curvature whose points, outside of a sufficiently large sphere with center at the origin, consist of a finite number of funnels. Each of these funnels will be cut off from the rest of the surface along a simple closed curve. These curves will be taken sufficiently remote on the funnels to be entirely distinct from one another.

An unparted hyperboloid of revolution is an example of a surface of negative curvature with two funnels.

From the definition of a funnel it follows that, by a continuous deformation of the closed curve forming the boundary of the funnel, the funnel may be swept out in such a way that every point of the funnel is reached once and only once. Hadamard considers two classes of funnels: those which can be swept out by closed curves which remain less in length than some fixed quantity, and those which do not possess this property. Surfaces with funnels of the first sort are for several reasons of less general interest than those with funnels of the second sort. In the present paper surfaces with funnels only of the second sort will be considered. Hadamard showed that there exist surfaces of negative curvature possessing funnels all of the second sort, of any arbitrary number exceeding one, and such that the surface obtained by cutting off these funnels is of an arbitrary genus.

§ 3. We shall consider surfaces which possess at least two funnels of the second sort, and of the surfaces with just two funnels of the second sort, we will exclude those surfaces that are topographically equivalent to an unbounded circular cylinder. Hadamard proves that on such surfaces there exists one and only one closed geodesic that is deformable into the boundary of a given funnel, and that this geodesic possesses no multiple points, and no points in common with the other closed geodesics that are deformable into the boundaries of the other funnels.

We shall denote these closed geodesics, say $v$ in number, by

$$(1) \qquad\qquad g_1\, g_2\, \cdots\, g_v.$$

They will form the complete boundary of a part of the surface, contained in a finite part of space. We denote this bounded surface by $S$. As shown in § 18 and § 19 of my earlier paper on geodesics, $S$ may be rendered simply connected as follows: $S$ is first cut along a system of geodesics,

$$h_1\, h_2\, \cdots\, h_{v-1}.$$

each of which has one end point on an arbitrarily chosen point, $P$, on $g_v$, and the other, respectively, on the geodesic of the set

$$g_1\, g_2\, \cdots\, g_{v-1},$$

with the same subscript, and no two of which have a point other than $P$ in common, and no points other than their end points in common with the geodesics, of the set (1). There then results a surface with a single boundary. This surface can be rendered simply connected by $2p$ geodesics,

$$c_1\, c_2\, \cdots\, c_{2p},$$

which can be taken as beginning and ending at $P$, and which will have no other points than $P$ in common with any of the other geodesics or with each other.

We denote by $T$, the simply connected piece of surface obtained by cutting $S$ along the above geodesics. It may be proved as a consequence of the assumptions made concerning the representation of the given surface, that $T$ is topographically equivalent to a plane region consisting of the interior and boundary points of a circle.

## REPRESENTATION OF GEODESICS BY LINEAR SETS OR BY REDUCED CURVES

§ 4. We suppose that we have at our disposal an unlimited number of copies of the simply connected surface $T$, and that each of these copies of $T$ is entirely distinct from every other copy of $T$.

DEFINITION. Let $r$ be any integer, positive, negative or zero. Let $T_r$

denote a particular copy of $T$. By a *linear set* of copies of $T$ will be understood a surface consisting of a set of copies of $T$ of the form

$$(1) \qquad \cdots T_{-2} \, T_{-3} \, T_0 \, T_1 \, T_2 \cdots ,$$

or of the form of any subset of successive symbols of (1), in which no one copy of $T$ appears twice, and in which each copy of $T$ is joined along some one of its boundary pieces to that boundary piece of the succeeding copy of $T$ that arises from the opposite side of the same cut, while no copy of $T$ is joined to its predecessor and successor along the same boundary piece. A linear set which has no first or last copy of $T$ will be called an *unending* linear set.

Two linear sets will be considered the *same* if the two sets of their copies of $T$ can both be expressed by the same form (1), in such a manner that successive symbols represent successive copies of $T$ in the respective linear sets, joined along copies of the *same* cut.

A linear set in which the number of copies of $T$ is finite is seen to be a multiple-leaved, simply connected surface, bounded by a single closed curve.

Let the set of geodesic segments,

$$g_1 \, g_2 \cdots g_v; \; h_1 \, h_2 \cdots h_{v-1}; \; c_1 \, c_2 \cdots c_{2p} ,$$

described in § 3, be denoted by $H$.

DEFINITION. Let $r$ be any integer, positive, negative, or zero. Let $k_r$ be any member of the set $H$. By a *reduced curve* we shall understand any *continuous* curve that consists of a set of members of the set $H$, excluding $g_v$, of the form

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

or of the form of any subset of consecutive symbols of (1). In the special* case where a $k_r$ and a $k_{r+1}$ of (1) are copies of the same member of the set $H$, say $l$, we require that the end point of $k_r$ and the end point of $k_{r+1}$ which are joined, be points which on $l$ would be considered as opposite end points. A reduced curve without end points will be termed an *unending* reduced curve.

If a given reduced curve be traced out in an arbitrary sense, it follows from the last condition of the definition of a reduced curve that no two consecutive pieces of the given reduced curve will thereby appear as copies of the same piece of $H$ taken in opposite senses.

§ 5. The results of this section are established in § 12 and § 13, of my earlier paper on geodesics, already referred to.

A given unending reduced curve is contained in one and only one linear

---

*We admit the possibility of two symbols in (1) representing the same member of the set $H$, but as parts of the reduced curve we shall consider two such copies as distinct, in a manner analogous to the convention ordinarily made in the construction of a Riemann surface.

set, which set is an unending linear set. Conversely, every unending linear set contains one and only one unending reduced curve. Each copy of $T$ of an unending linear set that contains an unending reduced curve contains either a point or a single continuous segment of the given reduced curve, and no other points of the given reduced curve. The results necessary for the developments of this paper are summed up in the following:

THEOREM 1. *There is a one to one correspondence between the set of all unending reduced curves on $S$, and the set of all unending linear sets, in which each reduced curve corresponds to that linear set in which it is contained.*

§ 6. The results of this section follow from the results of §§ 21, 22, and 23, of the earlier paper on geodesics. For the purpose of representing geodesics that lie wholly on $S$, it will be convenient to suppose each closed geodesic replaced by that geodesic obtained by tracing out the given closed geodesic an infinite number of times in either sense.

Every geodesic lying wholly on $S$ is contained on one and only one linear set, which set must be an unending linear set. Conversely, every unending linear set contains one and only one of the geodesics lying wholly on $S$. Every copy of $T$ of a linear set that contains a geodesic lying wholly on $S$, contains either a point or a single continuous segment of this geodesic, and no other point on this geodesic.

THEOREM 2. *There is a one to one correspondence between the set of all geodesics lying wholly on $S$, and the set of all unending linear sets, in which each geodesic corresponds to that linear set in which it is contained.*

The results of Theorems 1 and 2 can be combined in the following:

THEOREM 3. *There is a one to one correspondence between the set of all geodesics lying wholly on $S$, and the set of all unending reduced curves, in which each geodesic corresponds to that unending reduced curve that is contained in the same linear set.*

THEOREM 4. *If an unending reduced curve consists wholly of repetitions of a closed curve, the geodesic that passes through the same linear set consists wholly of successive repetitions of a closed geodesic. Conversely, if a geodesic consists wholly of successive repetitions of a closed curve, the unending reduced curve that passes through the same linear set, consists wholly of successive repetitions of a closed curve.*

## VARIATION OF GEODESICS WITH INITIAL ELEMENTS

§ 7. On a surface representable in the manner in which the given surface is representable, there is one and only one geodesic through a given point, and tangent to a given direction.

DEFINITION. A point on the surface and a direction tangent to the surface will be called an *element,* and will be said to define that sensed geodesic that

passes through the initial point of the given element, and is such that its positive tangent direction at that point agrees with the direction of the given element.

If $u$ and $v$ are parameters in any representation of a part of the surface, and if $\theta'$ is the angle which a given tangent direction makes at the point $(u'\,v')$ with the positive tangent to the curve, $u = u'$, then $(u'\,v'\,\theta')$ will represent an element of the given surface. We shall understand by each statement of metric relations between elements, the same statement of metric relations between the points in space of three dimensions obtained by considering the complex $(u'\,v'\,\theta')$ as the Cartesian coördinates of a point.

Let $G$ be any geodesic segment lying on the original uncut surface. $G$ is an extremal in the Calculus of Variations problem of minimizing the arc length, from which theory we can readily obtain the following theorem that describes the nature of the variation of $G$ with variation of its initial element.*

THEOREM 5. *Corresponding to any positive constants e and h, there exists a positive constant d so small, that if any two elements, with initial points on the bounded surface S, lie within d of each other, and if a second pair of elements lie respectively on the two geodesics defined by the first two elements, and if further the initial points of this second pair of elements lie respectively at a distance, measured along the given geodesics from the geodesics' initial points, that is the same in both cases and that does not exceed h, the second pair of elements will lie within e of each other.*

The following theorem describes the manner in which a geodesic varies with the reduced curve contained in the same linear set. It is given in § 24 of the earlier paper on geodesics.

THEOREM 6. *Corresponding to any positive constant e, there exists a positive constant k, so large, that if two unending reduced curves possess in common a continuous segment of length exceeding k, the two corresponding geodesics each have at least one element within e of some element on the other, and with initial point in the same copy of T, in the geodesic's linear set, as the mid point of the common reduced curve segment.*

*Conversely, corresponding to any positive constant k, there exists a positive constant e, so small, that if on each of two geodesics there exists some element within e of some element on the other, the two corresponding reduced curves possess in common a segment of length k, with mid point in the same copy of T in the reduced curve's linear set, as the initial point of either of the two elements.*

## REPRESENTATION OF GEODESICS BY SETS OF NORMAL CURVES

§ 8. The previous representation of geodesics by means of linear sets and reduced curves can now be replaced by another representation which will be

---

* Cf. Bolza, *Vorlesungen über Variationsrechnung* (1909), p. 219.

fundamental in the work of this paper.   This representation will be in terms
of the geodesic segments,

$$(1) \qquad c_1\, c_2\, \cdots\, c_{2p}\,,$$

$$(2) \qquad g_1\, g_2\, \cdots\, g_{v-1}\,,$$

which form a subset of the boundary pieces of each copy of $T$.

DEFINITION.   I. Each one of the geodesic segments of (1) and (2) will be
called a *normal segment*.

II. Let $m$ be any integer, positive, negative, or zero.   Let $C_m$ represent
any sensed normal segment.   By a *normal set $C$*, will be understood an un-
ending ordered set of sensed normal segments, in the form

$$(3) \qquad \cdots\, C_{-2}\, C_{-1}\, C_0\, C_1\, C_2\, \cdots\,,$$

in which no two successive members are the same normal segment taken in
opposite senses.

III. Two normal sets $C$ will be considered the *same* if they contain the
same normal segments in the same order with the same senses.

A normal set $C$ will not in general constitute a reduced curve.   For a re-
duced curve may include any normal segment, and in addition any geodesic
segment of the set,

$$(4) \qquad h_1\, h_2\, \cdots\, h_{v-1}\,.$$

However, it is readily seen that, with the aid of the members of the set
(4), there can be formed from a given normal set $C$ one and only one sensed
reduced curve whose normal segments taken in the order and with the senses
in which they appear on the given reduced curve constitute the given normal
set $C$.   Conversely, if there be given any unending sensed reduced curve,
its normal segments taken in the order and with the senses in which they
appear on the given unending sensed reduced curve, constitute a normal set $C$.

Thus there is a one to one correspondence between the set of all unending
sensed reduced curves and the set of all normal sets $C$, in which each normal
set $C$ corresponds to that unending sensed reduced curve whose normal seg-
ments, taken in the order and with the senses in which they appear on the
given unending sensed reduced curve, constitute the given normal set $C$.

DEFINITION.   If an unending sensed reduced curve and a normal set $C$
correspond in the sense of the preceding statement, the normal set $C$ will be
said to *represent* the given unending sensed reduced curve, and also that
sensed geodesic that passes through the same linear set of copies of $T$ in the
same sense as does the given unending sensed reduced curve.

By virtue of Theorem 3, § 6, every sensed geodesic lying wholly on the
surface $S$, is represented by one and only one normal set $C$, while every
normal set $C$ represents one and only one sensed geodesic.

CLOSED GEODESICS

§ 9. If a normal set $C$ of the from (3) of the preceding section represents a closed geodesic it follows from Theorem 4, § 6, that there exists a positive integer $p$, such that in the set $C$

$$C_m = C_{m+p},$$

where $m$ is any integer, positive, negative, or zero. The given normal set will then be said to be *periodic*, and to have the period $p$. Theorem 4, § 6 now becomes the following:

THEOREM 7. *A necessary and sufficient condition that a geodesic be closed, is that the normal set $C$ representing that geodesic be periodic.*

Let $q$ be the smallest period of a periodic set $C$. Then any other period $p$ must either equal $q$, or else be a multiple of $q$. For if $p$ were not equal to $q$ or a multiple of $q$, it follows from Euclid's Algorithm that there exist three integers, $A$, $B$, and $r$, of which $r$ is less than $q$, and is greater than zero, and which are such that

$$Aq + Bp = r.$$

It follows from this equation that $r$ is also a period of the given periodic set $C$, contrary to the assumption that $q$ was the smallest period of the given periodic set $C$.

DEFINITION. If $q$ is the smallest period of a periodic normal set $C$, then any $q$ successive sensed normal segments of the given set $C$ will be called a *generating set* of the given set $C$, and also of the closed geodesic represented by the given set $C$.

If $B$ is a generating set of a normal set $C$, this set $C$ consists merely of an unending succession of sets $B$, which we will write in the form,

$$\cdots B\ B\ B\ B\ B\ B\ B\ B \cdots$$

All generating sets of a periodic set $C$, can evidently be obtained from any one such generating set by a circular permutation of the sensed normal segments composing the given generating set.

§ 10. We consider now the question of the arbitrary formation of sets that may serve as generating sets of some geodesic. To that end we form a finite ordered set of sensed normal segments, in which neither the first and last members, nor any two successive members are the same normal segment taken in opposite senses, and which cannot be obtained through repetitions of a similar set containing fewer sensed normal segments. Denote the set so obtained by $D$. The set

(1)                    $$\cdots D\ D\ D\ D\ D \cdots$$

is, in the first place, a normal set $C$. For $D$ is made up of sensed normal

segments in which neither the first and last members, nor any two successive members are the same normal segment taken in opposite senses. Further $D$ is a generating set of the set (1), for otherwise (1) would have a period smaller than the number of successive segments in $D$, and hence a period that is a divisor of the number of successive segments in $D$. $D$ could then be obtained by a finite number of repetitions of a similar set containing fewer sensed normal segments, contrary to the last hypothesis made concerning $D$.

The number of different periodic normal sets $C$ is seen to equal the number of generating sets not obtainable one from the other by a circular permutation of their normal segments. The number of such generating sets is readily seen to be an enumerable infinity. From this result, together with the theorem of the preceding section, we have the result given by Hadamard:

*There are an enumerable infinity of distinct closed geodesics on the surface $S$.*

### LIMIT GEODESICS OF SETS OF GEODESICS

§ 11. DEFINITION. A geodesic $G$ will be said to be a *limit geodesic of a set of geodesics* if a set of elements, $M$, lying on the given set of geodesics, have as a limit an element $E$, on $G$, while all the initial points of those elements of the set $M$ that lie on $G$, are at distances, measured along $G$ from the initial point of $E$, exceeding a fixed positive quantity.

From the property of continuous variation of a geodesic with its initial element, as given in Theorem 5, § 7, it follows that if one element on $G$ is a limit element of elements on a given set of geodesics, then every element on $G$ is a limit element of elements on the given set of geodesics.

If a closed geodesic should be considered as replaced by an unclosed geodesic that traces out the given closed geodesic an infinite number of times in either sense, the latter geodesic would be a limit geodesic of itself. In this sense any closed geodesic will be considered a limit geodesic of itself.

From Theorem 6, § 7, it follows that a necessary and sufficient condition that a geodesic $G$ be a limit geodesic of a set of geodesics $J$, not including $G$, is that every finite segment of the unending reduced curve corresponding to $G$ be contained in the unending reduced curve corresponding to some geodesic of the set $J$. In terms of normal sets $C$, this result becomes the following:

THEOREM 8. *A necessary and sufficient condition that a geodesic $G$ be a limit geodesic of a set of geodesics $J$, not including $G$, is that every subset of consecutive normal segments of the normal set representing $G$, be a subset of consecutive normal segments of some normal set representing a geodesic of the set $J$.*

The following theorem is given by Hadamard, with a proof, however, that is different from the following.

THEOREM 9. *Every geodesic lying wholly on a surface of negative curvature*

*for which* $2p + v - 1 \geqq 2$ (cf. section 3), *is a limit geodesic of the set of all closed geodesics on that surface.*

The number of different normal segments equals $2p + v - 1$. Hence on any of the surfaces considered, there are at least two different normal segments. Since any closed geodesic is a limit geodesic of itself, we need only consider the case of a geodesic not a closed geodesic. Let $G$ be any geodesic lying wholly on $S$, and not a closed geodesic. Let there be given an arbitrary finite subset of consecutive normal segments of the normal set $C$ representing $G$. If this subset does not begin and end with the same normal segment taken in opposite senses, we denote the subset by $D$; in the other case we add to the given subset a normal segment different from the first and last normal segment, and denote this set also by $D$.

In either case,

$$\cdots D\ D\ D\ D\ D\ D\ D \cdots$$

will be a normal set $C$. This normal set is periodic; according to Theorem 7, § 9, it then represents a closed geodesic. Further this normal set contains as a subset of successive normal segments the given arbitrary subset of the normal set representing $G$. From Theorem 8, it accordingly follows that $G$ is a limit geodesic of the set of all closed geodesic on $S$, and the theorem is proved.

THEOREM 10. *On a surface of negative curvature for which* $2p + v - 1 \geqq 2$, *there exists at least one geodesic which has for a limit geodesic every geodesic lying wholly on* $S$.

The set of all possible finite subsets of consecutive normal segments of normal sets $C$, form an enumerable set which may accordingly be put into one to one correspondence with the set of all integers, positive, negative, or zero. In this correspondence that one of these subsets that corresponds to the integer $n$, we denote by $B_n$. The set,

$$\cdots B_{-2}\ B_{-1}\ B_0\ B_1\ B_2 \cdots$$

will be a normal set $C$, unless for some integer $n$, the last sensed normal segment of $B_{n-1}$ and the first sensed normal segment of $B_n$ are the same normal segments taken in opposite senses. In every such case we insert between $B_{n-1}$ and $B_n$ a normal segment different from the normal segment in question. The resulting set will be a normal set, which we denote by $C'$.

$C'$ contains each subset $B_n$ as a subset of consecutive normal segments. It follows from Theorem 8, that every geodesic lying wholly on $S$, with the possible exception of the geodesic represented by $C'$, is a limit geodesic of the geodesic represented by $C'$. That the geodesic represented by $C'$ is a limit geodesic of itself, follows from the fact that every *closed* geodesic is a

limit geodesic of the geodesic represented by $C'$, while *every* geodesic lying wholly on $S$ is a limit geodesic of the set of all closed geodesics on $S$.

<div align="center">RECURRENT GEODESICS</div>

§ 12. The following definition, and Theorems 11, 12, and 13, are restatements for the case of geodesics of what is given by Professor Birkhoff for a dynamical system, in the paper referred to in the introduction.

DEFINITION. By a *minimal set* of geodesics we shall understand any set of geodesics lying wholly on $S$, each of which has every other geodesic of the set, and no other geodesic, as a limit geodesic. Any geodesic of a minimal set will be called a *recurrent* geodesic.

A closed geodesic constitutes a minimal set in which it is the only geodesic.

The following theorem serves as an existence proof for recurrent geodesics.

THEOREM 11. *Every geodesic lying wholly on $S$ contains among its limit geodesics at least one minimal set of geodesics.*

Concerning the number of recurrent geodesics in a minimal set, we have

THEOREM 12. *The power of any minimal set not simply a closed geodesic, is that of the continuum.*

The characteristic property of a recurrent geodesic is given by the following:

THEOREM 13. *A necessary and sufficient condition that a geodesic lying wholly on $S$ be a recurrent geodesic is that, corresponding to any arbitrary positive constant $e$, there exist a positive constant $h$, so large, that if $L$ be any segment of the given geodesic of length at least equal to $h$, any element of the given geodesic lies within $e$ of some element of $L$.*

§ 13. Let there be given a set of symbols of the form,

$$(1) \qquad \cdots R_{-2}\, R_{-1}\, R_0\, R_1\, R_2 \cdots.$$

Let $m$ and $n$ be any integers, positive, negative, or zero.

DEFINITION. I. A set of symbols of the form (1) will be said to be *recurrent*, if corresponding to any positive integer $r$, there exists a positive integer $s$, so large that any subset of (1) of the form,

$$(2) \qquad R_m\, R_{m+1} \cdots R_{m+r}$$

is contained in every subset of (1) of the form

$$(3) \qquad R_n\, R_{n+1} \cdots R_{n+s}.$$

II. The set (1) will be said to be *periodic*, if there exists a positive integer $p$, such that
$$R_n = R_{n+p},$$
whatever integer $n$ may be, and $p$ will be said to be a *period* of the set (1).

It appears at once that a set (1) that is periodic, is also recurrent.

Theorem 13, § 12, interpreted in terms of normal sets $C$ by means of Theorem 6, § 7, becomes the following

THEOREM 14. *A necessary and sufficient condition that a geodesic lying wholly on S be recurrent, is that the set $C$ representing the given geodesic be recurrent.*

<div align="center">EXISTENCE OF RECURRENT GEODESICS, NOT PERIODIC.</div>

§ 14. We come now to the question of the existence of recurrent geodesics that are not closed geodesics.

On a surface of negative curvature topographically equivalent to an unbounded circular cylinder, the only possible recurrent geodesic is a single closed geodesic. On a surface of negative curvature topographically equivalent to an unbounded plane, there are no recurrent geodesics whatever. The surfaces of negative curvature which we have been considering include neither of these two types of surfaces (cf. § 3).

We have seen in Theorem 7, § 9, that a geodesic that is periodic is represented by a normal set $C$ that is periodic; while Theorem 14, § 13, states that a normal set $C$ that is recurrent represents a geodesic that is recurrent. Hence, to prove the existence of a geodesic that is recurrent without being periodic, it is sufficient to prove the existence of a normal set $C$ that is recurrent without being periodic.

Now there are just $2p + v - 1$ normal segments (cf. § 8). We are considering surfaces for which $2p + v - 1 \geqq 2$. Hence any of the surfaces of negative curvature considered will possess at least two normal segments. We will seek a normal set $C$ that is composed solely of two normal segments. For that purpose the following lemma is introduced.

LEMMA. *There exists an unending set of symbols each of which is either 1 or 2, which forms a set that is recurrent without being periodic.*

By the juxtaposition of two or more symbols representing ordered sets of symbols, we shall mean here, as elsewhere, the ordered set obtained by taking the symbols of the given sets in the order in which the sets are written.

Let $n$ be any positive integer. We introduce the following definitions:

(1)
$$a_0 = 1,$$
$$b_0 = 2,$$
$$a_1 = a_0 b_0,$$
$$b_1 = b_0 a_0,$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot$$
$$a_{n+1} = a_n b_n,$$
$$b_{n+1} = b_n a_n.$$

We introduce the set of symbols,

(2)                    $\cdots d_{-2} \, d_{-1} \, d_0 \, d_1 \, d_2 \cdots ,$

of which

$$d_0 \, d_1 \cdots d_{2^n}$$

are defined respectively as the $2^n$ integers of $a_n$; further, if $m$ is any positive integer, $d_{-m}$ is defined as equal to $d_{m-1}$. The set (2), so defined, will be proved to be recurrent without being periodic.

For definiteness we write out (2) in part, beginning with $d_0$:

(3)                    1221 2112   2112 1221   2112 1221   $\cdots$

It follows from the definitions (1), that if the integers of (2) be grouped in groups of $2^n$ integers, then the set (2) can be expressed, beginning with $d_0$, by a succession of the sets $a_n$ and $b_n$, obtained by replacing the integers 1 and 2 in the set (2), respectively by $a_n$ and $b_n$. Thus beginning with $d_0$, (2) is given in part as

(4)                    $a_n \, b_n \, b_n \, a_n$   $b_n \, a_n \, a_n \, b_n$   $b_n \, a_n \, a_n \, b_n$   $\cdots$

The symbols of the set (2) that have negative subscripts, can be obtained, according to their definition, by taking the symbols of (2) with positive or zero subscripts in reverse order. It follows from the definitions (1) that the integers of $a_n$ and $b_n$, taken respectively in their reverse orders, give $a_n$ and $b_n$ when $n$ is even, and $b_n$ and $a_n$ when $n$ is odd. We have the result:

*Whatever integer $n$ may be, the set (2) can be expressed by a properly chosen succession of the sets, $a_n$ and $b_n$. Thus, if $r$ be any integer such that*

$$r \equiv 0 \; modulo \; 2^n ,$$

*then any subset of (2) of the form*

$$d_r \, d_{r+1} \cdots d_{r+2^n}$$

*is either a set $a_n$ or a set $b_n$.*

We will now prove that the set (2) is recurrent.

Let there be given any subset of (2) of the form

(5)                    $d_s \, d_{s+1} \cdots d_{s+m} ,$

where $s$ is any integer, positive, negative, or zero, and $m$ is any positive integer. Let $r'$ be the largest integer less than $s$ such that

$$r' \equiv 0 \; modulo \; 2^m .$$

From the choice of $r'$, we have,

$$r' < s < s + m < r' + 2^{m+1} .$$

Hence the set (5) is a subset of the set,

(6) $$d_{r'} \, d_{r'+1} \cdots d_{r'+2^{m+1}}.$$

From the result of the preceding paragraph, it appears that (6) must be one of the four possible ordered combinations of $a_m$ and $b_m$, that is, one of the four sets,

$$a_m \, b_m, \quad b_m \, b_m, \quad b_m \, a_m, \quad a_m \, a_m.$$

Each of these four sets is a subset of $a_{m+3}$ and $b_{m+3}$; for from the equations (1), we have,

$$a_{m+3} = a_{m+2} \, b_{m+2} = a_{m+1} \, b_{m+1} \, b_{m+1} \, a_{m+1} = a_m \, b_m \, b_m \, a_m \, b_m \, a_m \, a_m \, b_m,$$

$$b_{m+3} = b_{m+2} \, a_{m+2} = b_{m+1} \, a_{m+1} \, a_{m+1} \, b_{m+1} = b_m \, a_m \, a_m \, b_m \, a_m \, b_m \, b_m \, a_m.$$

Since the set (2) can be expressed as a succession of the sets $a_{m+3}$ and $b_{m+3}$, each of which contains $2^{m+3}$ integers of the set (2), it appears that any subset of at least $2^{m+4}$ successive integers of (2), say $R$, contains at least one of the sets $a_{m+3}$ and $b_{m+3}$. Retracing the steps it is seen that $R$ contains a subset identical with the given set (5). The set (2) is thus recurrent.

We will now show that the set (2) is not periodic. Suppose that the set (2) had a period prime to 2. Since $p$ is prime to 2, there exists an integer $m$, greater than one, such that

(7) $$2 \equiv 2^m, \qquad \text{modulo } p.$$

Since the set (2) has the period $p$, it follows from (7) that the set

(8) $$d_2 \, d_3 \, d_4 \cdots$$

must be identical with the set

(9) $$d_{2^m} \, d_{2^m+1} \, d_{2^m+2} \cdots.$$

The set (8) commences with the integers

(10) $$2 \, 1 \, 2 \, 1 \, 1 \, 2 \cdots,$$

while the set (9) commences as does $b_m$, which is seen from the equations (1) to commence with the integers

(11) $$2 \, 1 \, 1 \, 2 \, 1 \, 2 \cdots.$$

The sets (10) and (11) are not identical. The set (2) can thus have no period prime to $p$.

Finally suppose that the set (2) had a period $2^r \, p$, where $r$ is any positive integer, and $p$ is prime to 2. Let the set (2), commencing with $d_0$, be written in terms of $a_r$ and $b_r$:

(12) $$a_r \, b_r \, b_r \, a_r \, b_r \, a_r \cdots.$$

Considered as a succession of symbols $a_r$ and $b_r$, (12) has the period $p$. But the original expression for (2) in terms of its integers, and commencing with $d_0$, is obtained from (12) by replacing the symbols $a_r$ and $b_r$ respectively by 1 and 2. Thus the expression for the set (2), in terms of its integers, and commencing with $d_0$, would have a period prime to 2. We have seen this to be impossible.

Thus the set (2) is recurrent without being periodic, and the lemma is proved.

§ 15. THEOREM 15. *On a surface of negative curvature for which*

$$2p + v - 1 \geqq 2,$$

*there exists a set of geodesics that are recurrent without being periodic, and this set has the power of the continuum.*

The number of different normal segments equals $2p + v - 1$. We are considering surfaces of negative curvature for which $2p + v - 1 \geqq 2$. Hence on any of the surfaces considered, there are at least two different normal segments. Let $N_1$ and $N_2$ be two different normal segments, each taken in an arbitrary sense.

In the preceding lemma we have established the existence of a set,

$$\cdots d_{-2}\, d_{-1}\, d_0\, d_1\, d_2 \cdots ,$$

that is composed entirely of the integers one and two, and which is a set that is recurrent without being periodic. The set

(1)                     $\cdots N_{d_{-2}}\, N_{d_{-1}}\, N_{d_0}\, N_{d_1}\, N_{d_2} \cdots$

is accordingly recurrent; from Theorem 14, § 13, it follows that the geodesic represented by the set (1) is recurrent. The set (1) is not periodic; it follows from Theorem 7, § 9, that the geodesic represented by (1) is not periodic. We have thus established the existence of a geodesic that is recurrent without being periodic.

According to Theorem 12, § 12, the existence of one geodesic that is recurrent without being periodic, is sufficient to establish that the power of the complete set of geodesics that are recurrent without being periodic is that of the continuum.

§ 16. THEOREM 16. *On a surface of negative curvature for which*

$$2p + v - 1 \geqq 2,$$

*the set of all geodesics that are recurrent without being periodic, has as a limit geodesic every geodesic lying wholly on $S$.*

Let there be given an arbitrary closed geodesic lying on $S$. Let $B$ be any finite subset of successive normal segments of the normal set $C$ representing the given closed geodesic. Let $N_1$ and $N_2$ be the two sensed normal segments

used in the proof of the preceding theorem, and $- N_1$ and $- N_2$ be, respectively, the same normal segments taken in opposite senses.

If now the set $B$ does not begin or end with $- N_1$, or $- N_2$, we denote the set $B$, by $D$. If the set $B$ begins with $- N_1$ or $- N_2$, we prefix $N_2$ or $N_1$, respectively, to the set $B$, while if the set $B$ ends with $- N_1$ or $- N_2$, we add $N_2$ or $N_1$, respectively to the set $B$, and in either case denote the resulting set by $D$. We interpose this set $D$ between each two successive sensed normal segments of the normal set $C$, given by (1) in the proof of the preceding theorem, and denote the resulting set by $C'$.

It is a consequence of the nature of the construction of the set $C'$, that no two of its successive sensed normal segments are the same normal segment taken in opposite senses. The set $C'$ is thus a normal set. The normal set (1) of the proof of the preceding theorem, is recurrent without being periodic; it follows that the set $C'$ is recurrent without being periodic. The geodesic represented by $C'$ is accordingly recurrent without being periodic. The set $C'$ contains $B$ as a subset of successive normal segments. It follows from Theorem 8, § 11, that the given closed geodesic is a limit geodesic of the set of all recurrent geodesics that are not periodic.

That every geodesic lying wholly on $S$ is a limit geodesic of the set of all geodesic that are recurrent without being periodic, follows now from the fact that *every* geodesic lying wholly on $S$ is a limit geodesic of the set of all closed geodesics on $S$.

### DISTRIBUTION OF ELEMENTS ON RECURRENT GEODESICS

§ 17. DEFINITION. Two elements $E'$ and $E''$ of a set of elements $M$ on a region $R$ of $S$, will be said to be *mutually accessible in $M$ and on $R$*, if corresponding to any positive constant $e$, there exists in the set $M$, a finite ordered subset of elements of which the first is $E'$, and the last $E''$, while each element of the subset, excepting the last, lies within a geodesic distance $e$, measured on $R$, of the following element.

The following theorem is established in § 25 of the earlier paper on geodesics.

THEOREM 17. *On any simply connected region $R$ of $S$, and in the set of all elements on $R$, and on geodesics lying wholly on $S$, no two elements on different geodesics are mutually accessible.*

A particular consequence of the preceding theorem is that, on any simply connected region $R$ of $S$, and in the set of all elements on $R$, and on geodesics that are *recurrent*, no two elements on different geodesics are mutually accessible. A set of recurrent geodesics with this property are of a type called *discontinuous* recurrent motions by Professor Birkhoff. Thus:

THEOREM 18. *The set of all recurrent geodesics on $S$ constitutes a set of recurrent motions of the discontinuous type.*

The proof, given in this paper, of the existence of a set of this type, is the first proof of the existence of a discontinuous set of recurrent motions.

§ 18. A recurrent geodesic was defined as a member of a minimal set,—a set in which every geodesic has every geodesic of the set, and no other geodesic, as a limit geodesic. In case a given recurrent geodesic is a closed geodesic, the minimal set containing the given geodesic consists merely of the given closed geodesic. In case a recurrent geodesic is not a closed geodesic, the power of the minimal set that contains the given recurrent geodesic, is, according to Theorem 12, § 12, that of the continuum.

From the definition of a minimal set, it appears that no two minimal sets that are not identical, have any geodesic in common. Each recurrent geodesic thus belongs to one and only one minimal set. The question arises as to how many different minimal sets there are on the given surface. That there are at least an enumerable infinity, follows at once from the fact that there are an enumerable infinity of closed geodesics. The number of minimal sets that do not consist simply of one closed geodesic still remains to be determined.

It has been seen that any geodesic lying wholly on $S$ can be completely characterized by means of normal sets of sensed normal segments. It may be inquired whether or not any minimal set may not be characterized in terms of sensed normal segments, and if so, what is the explicit nature of the characterization. These questions seem to indicate the opening to an interesting field of inquiry.

HARVARD UNIVERSITY,
      *June, 1917*