THE DYNAMICS OF GEODESIC FLOWS*

GUSTAV A. HEDLUND

1. **Introduction.** Geodesic systems, particularly those on two-dimensional manifolds, have been a rich source in the determination and display of the possible types of macroscopic behavior of the motions of dynamical systems. In connection with the question of the existence of periodic motions, Poincaré [1]† investigated the geodesics on convex surfaces. Hadamard [1] has constructed open surfaces of negative curvature and proved the existence of interesting classes of geodesics on these surfaces. By an ingenious use of symbolism to characterize these geodesics, Morse [1] proved the existence of nonperiodic recurrent geodesics of discontinuous type. Birkhoff [1] has constructed closed surfaces of nonpositive curvature and has shown that, among many other types, there exist transitive geodesics on these surfaces.

There is another group of mathematicians who have made numerous contributions in connection with geodesic systems on surfaces of constant negative curvature. As will be seen, these surfaces have a close relationship with Fuchsian groups, and in addition to their work having other connections with these groups, Artin [1], Myrberg [1, 2, 3], Nielsen [1, 2], Koebe [1], and Löbell [1, 2, 3, 4] have derived many properties of the geodesics.

With the recent developments in ergodic theory, interest has been centered on those properties of geodesic flows associated with transitivity in some form, as for example, regional transitivity, metric transitivity, and mixture. The conditions under which regional transitivity holds have been greatly extended by Morse [3]. Geodesic systems have furnished some of the few known examples of metrically transitive dynamical systems (cf. Hedlund [1, 2], E. Hopf [1]). As will be indicated, a number of new results concerning transitivity can be added, both in the case of constant curvature and in the case of variable curvature.

An enormous body of results has been attained, and an hour is entirely inadequate to permit a description of all. For this reason I propose to restrict the discussion to transitivity properties of geodesic flows. It has been conjectured (Birkhoff [3], p. 370) that these are the important properties in that they are *general* in some

* An address delivered before the New York meeting of the Society on October 27, 1938, by invitation of the Program Committee.

† Numbers in brackets refer to the bibliography at the end of the paper.

sense, but the evidence gathered to date is not at all conclusive. Outstanding problems remain unsolved, a notable one being the problem of metric transitivity of the geodesic flow on a closed analytic surface of variable negative curvature.

It is of interest to note the influence of Poincaré in many phases of our considerations. It was Poincaré who first emphasized macro-analysis and probability considerations in the study of dynamical systems. We are indebted to him for the Poincaré fundamental group and much of our knowledge of Fuchsian groups which play an important role in the analysis of geodesic flows on two-dimensional manifolds of negative curvature.

2. **Flows and transitivity types.** The geodesics on a Riemannian manifold are the solutions of the Euler equations, a system of second order differential equations derived by imposing the condition that the first variation of the arc length vanish. If the coefficients of the positive definite quadratic forms

$$g_{\alpha\beta}du^{\alpha}du^{\beta}$$

which define the Riemannian manifold locally are of class C'' (somewhat less than this is sufficient), a geodesic is uniquely determined by an element, that is, by a point and a direction at that point. Let us assume that the manifold is complete in the sense that each geodesic can be continued to infinite length (cf. H. Hopf and Rinow [1]). Then if g is the directed geodesic determined by the element e, and s is the sensed arc length on g measured from the point P at which e is situated, corresponding to s there is an element e_s, namely, the element of g at distance s along g from P. Furthermore, e_s varies continuously with e and s. The transformation $e \rightarrow e_s$ is a transformation T_s of the space Ω of elements on the manifold into itself, and the properties which we consider can be most simply stated in terms of such a one-parameter set of transformations T_s.

The conditions which will be imposed on the space in which the transformations are defined and on the transformations themselves will be fulfilled by the element spaces and transformations in them which we subsequently consider.

Let Ω be a metric, separable, complete space in which an l-measure in the sense of von Neumann (cf. von Neumann [1], p. 575) is defined. The measure defined by use of this l-measure, which is analogous to a Lebesgue outer measure, will be denoted by m. Let T_t be, for each real t, a one-to-one transformation of Ω into itself satisfying the following conditions:

(a) $T_0(P)=P$; $T_t[T_s(P)]=T_{t+s}(P)$.

(b) $T_t(P)$ is a continuous function of t and P.

(c) If A is a measurable subset of Ω, then $T_t(A)$ is measurable and $m[T_t(A)]=mA$.

Such a continuous one-parameter group of transformations will be called a *flow* in Ω.

It will be convenient to denote the set $T_t(A)$, $(A\subset\Omega)$, simply by A_t. The set A will be said to be *invariant* if A_t coincides with A for all t. The set P_t, $(-\infty<t<+\infty)$, will be called a *motion* or *trajectory*.

The properties of the flow T_t in Ω with which we shall be concerned are the following, where the set of points common to the sets A and B of Ω is denoted by $A\cdot B$ and the empty set is denoted by 0.

REGIONAL TRANSITIVITY. *Given D and D^*, arbitrary open sets in Ω, there exists a t such that $D_t\cdot D^*\neq 0$.*

TOPOMETRIC TRANSITIVITY. *Given M, any measurable set of positive measure in Ω, and D, any open set in Ω, there exists a t such that $M_t\cdot D\neq 0$.*

METRIC TRANSITIVITY. *Given M and M^*, arbitrary measurable sets of positive measure in Ω, there exists a t such that $M_t\cdot M^*\neq 0$.*

PERMANENT REGIONAL TRANSITIVITY. *Given D and D^*, arbitrary open sets in Ω, there exists a $\bar{t}$ such that $D_t\cdot D^*\neq 0$, $(|t|\geqq\bar{t})$.*

PERMANENT TOPOMETRIC TRANSITIVITY. *Given M, any measurable set of positive measure in Ω, and D, any open set in Ω, there exists a $\bar{t}$ such that $M_t\cdot D\neq 0$, $(|t|\geqq\bar{t})$.*

PERMANENT METRIC TRANSITIVITY. *Given M and M^*, arbitrary measurable sets of positive measure in Ω, there exists a $\bar{t}$ such that $M_t\cdot M^*\neq 0$, $(|t|\geqq\bar{t})$.*

MIXTURE. *Given M, M^*, and $\overline{M}$, arbitrary measurable sets of Ω of finite positive measure,*

$$\lim_{t\to\pm\infty}\frac{m(M_t\cdot M^*)}{m(M_t\cdot\overline{M})}=\frac{mM^*}{m\overline{M}}\cdot$$

Regional transitivity is sometimes given a different but equivalent definition. A motion will be called *transitive* if the points on it form a set which is everywhere dense in Ω. *Regional transitivity is equivalent to the property that there exist a transitive motion* (cf., for example, Birkhoff [1], chap. 7).

The points on the transitive motions in Ω form a set which is the product of a denumerable set of open sets and is thus measurable. If this set coincides with Ω except for a set of measure zero, it will be said that *almost all motions are transitive. Topometric transitivity is a necessary and sufficient condition that almost all motions be transitive.*

Metric transitivity is a necessary and sufficient condition that in any division of Ω into two complementary, invariant, measurable sets, one of the sets is of measure zero. In this form the notion of metric transitivity was introduced by Birkhoff and Smith (cf. Birkhoff [**4**], p. 365). It plays a fundamental role in connection with ergodic theory. According to the ergodic theorem of Birkhoff [**2**], if $m\Omega$ is finite and M is any measurable set in Ω, the mean time of sojourn of a motion in M (that is, $\lim_{\beta-\alpha\to\infty} L_{\alpha,\beta}(P, M)/(\beta-\alpha)$, where $L_{\alpha,\beta}(P, M)$ is the linear measure of the part of the set P_t, $(\alpha \leqq t \leqq \beta)$, in M) exists except for a set of motions of measure zero. If metric transitivity holds, the mean time of sojourn in M is the same for almost all motions and is equal to $mM/m\Omega$.

If $m\Omega$ is finite, by replacing $\overline{M}$ by Ω the mixture property becomes

$$\lim_{t\to\pm\infty} m(M_t\cdot M^*) = \frac{mMmM^*}{m\Omega},$$

and, conversely, this implies the mixture property. Thus, any measurable set M of positive measure tends, with increasing or decreasing time, to occupy a definite fractional part of any other measurable set M^*, and the fraction is simply the fractional part of Ω which M occupies. Sets tend towards *homogeneous distribution* in Ω. (In this connection cf. E. Hopf [**2**], where references to the work of Koopman and von Neumann will be found.)

There are a number of evident relationships between the transitivity properties which have been defined. Metric transitivity implies topometric transitivity, which, in turn, implies regional transitivity. Any one of the permanent types of transitivity implies the corresponding non-permanent type. Permanent metric transitivity implies permanent topometric transitivity, which, in turn, implies regional transitivity. Mixture implies permanent metric transitivity and thus implies all the types of transitivity which have been defined here.

The flow defined by a suitably chosen family of parallel straight lines on a torus (rectangle with opposite sides identified), shown to be metrically transitive by Birkhoff and Smith (cf. Birkhoff [**4**], p. 368) yields an example which is not permanently regionally transitive and thus has none of the permanent transitivity properties. It seems to be difficult to give examples of flows which have one of the non-permanent types of transitivity without having the other non-permanent types, or which have one of the permanent types of transitivity without possessing the other permanent types. However, it will be possible to give an example of a geodesic flow which is regionally transitive (and even permanently regionally transitive) but not metrically transitive.

3. **Two-dimensional manifolds of constant negative curvature.** The simplest manifolds on which the geodesics display transitivity properties of the kind we are considering are two-dimensional manifolds of constant negative curvature. To define such manifolds, let Ψ be the interior of the unit circle U: $x^2+y^2=1$. To Ψ we assign the metric

$$ds^2 = \frac{4(dx^2 + dy^2)}{c(1 - x^2 - y^2)^2} = \frac{4\,|\,dz\,|}{c(1 - z\bar{z})^2}, \qquad c > 0. \tag{3.1}$$

The curvature of this simply connected Riemannian manifold is $-c$. The metric (3.1) assigns a length to curves in Ψ, and this length will be called *hyperbolic length*. Angle is euclidean angle, and the element of (hyperbolic) area is

$$\frac{4dxdy}{c(1 - x^2 - y^2)^2}. \tag{3.2}$$

The geodesics defined by (3.1) are arcs of circles orthogonal to U and will be called *hyperbolic lines*. Given two points P and Q of Ψ, there is a unique hyperbolic line segment joining P and Q, and the hyperbolic length of this hyperbolic line segment is the *hyperbolic distance* between P and Q.

The metric (3.1) is invariant under linear fractional transformarions which take Ψ into Ψ, so that under such transformations, hyperbolic distance, angle, and area are invariant. Such a transformation is either an elliptic transformation with fixed points inverse with respect to U, a parabolic transformation with fixed point on U, or a hyperbolic transformation with fixed points on U. These transformations are rigid motions of the well known hyperbolic geometry under consideration.

Now let F be a Fuchsian group with U as principal circle. That is, F is a group of linear fractional transformations, each of which transforms U into U and Ψ into Ψ, such that F is properly discontinuous in Ψ (cf. Ford [1], p. 35, chap. 3). Two sets of points in Ψ are *congruent* if there is a transformation of F taking one of these sets into the other. Either set will be said to be a *copy* of the other.

It can be shown that corresponding to any such group F there exists a normal fundamental region R (cf. Ford [1], pp. 44, 69–70). This is a simply connected region bounded by arcs of hyperbolic lines which are congruent in pairs, such that no two interior points of R are congruent and any point of Ψ is congruent to some point within or on the boundary of R. If suitable conventions are made as to the inclusion of boundary points of R, no two copies of R have a common point and the totality of these copies fills Ψ.

If points which are congruent under F are considered identical, there is defined a two-dimensional manifold M_F^{-c} of constant negative curvature $-c$. These manifolds are non-euclidean space forms of hyperbolic type, and an extensive analysis of them can be found in the papers of Koebe [1] and Löbell [1]. By including in the group F transformations of the form

$$w = \frac{a\bar{z} + \bar{c}}{c\bar{z} + \bar{a}}, \qquad a\bar{a} - c\bar{c} = 1,$$

these authors consider non-orientable as well as orientable manifolds. For simplicity, the discussion will be restricted to manifolds defined by Fuchsian groups, though the results derived apply to the non-orientable cases.

The presence of elliptic transformations in F, such a transformation necessarily having one of its fixed points in Ψ, implies the existence of singular points on the manifold M_F^{-c}. The total angle at such a point is not 2π.

In the nonsingular case there are restrictions on the topological invariants of the manifolds. If the manifold is closed and orientable, its genus must be greater than one. (In the non-orientable case, the genus of a closed manifold must be greater than two.) Among the open manifolds are included manifolds of finite or infinite connectivity.

As a first classification of the manifolds M_F^{-c}, they are divided into first and second kind. The group F is of the first kind if it is not properly discontinuous on U. The corresponding manifold M_F^{-c} will be said to be of the *first kind* and denoted by M_{I}^{-c}. If F is not of the first kind, it is of the second kind, and the corresponding manifolds of the *second kind* will be denoted by M_{II}^{-c}. An essential difference between Fuchsian groups of the first and second kind lies in the behavior of the fundamental region. In the case of a group of the second kind, the fundamental region abuts on the circle U in an interval (cf. Ford [1], pp. 74–75). Thus manifolds of the second kind are necessarily open. In the case of groups of the first kind the boundary of the fundamental region cannot contain an interval of U and may or may not have points on U. Manifolds of the first kind include all closed orientable two-dimensional Riemannian manifolds of constant negative curvature. In addition there are included manifolds which are not closed and which may be of finite or infinite connectivity.

An element e in Ψ is a point of Ψ together with a direction at that point and can be specified by three coordinates (x, y, ϕ), where x and y are the coordinates of the point and ϕ, $(0 \leqq \phi < 2\pi)$, is an angular coordinate at the point measured positively in the counterclockwise sense from a direction parallel to the positive x-axis. The point $P(x, y)$ is the point *bearing* the element (x, y, ϕ). A *neighborhood* of the element $e_1(x_1, y_1, \phi_1)$ is the set (x, y, ϕ) such that

$$H(P, P_1) < \delta, \qquad \|\phi - \phi_1\| < \delta,$$

where P is the point (x, y), P_1 is the point (x_1, y_1), $H(P, P_1)$ denotes the hyperbolic distance between P and P_1 when $c=1$, $\|\phi-\phi_1\|$ denotes the least value of the set $|\phi-\phi_1+2n\pi|$, $(n=0, \pm 1, \pm 2, \cdots)$, and $\delta>0$. Let E denote the space of elements in Ψ with neighborhoods thus defined.

A transformation of F carries an element into a *congruent* element. The space Ω_F of elements on M_F^{-c} is the space obtained by identifying congruent elements of E. Neighborhoods are defined in Ω_F as the correspondents of the neighborhoods in E, and Ω_F is a Hausdorff space. By suitably defining a metric in Ω_F, Ω_F is made a metric, separable, complete space.* The points of Ω_F are in 1-1 correspondence with a subset E_R of E obtained by restricting the points bearing the elements to the fundamental region R (a properly chosen subset of the elements at boundary points of R included). Measurability is defined in E_R by considering this as a subset of the three-dimensional euclidean space (x, y, ϕ), and measure in E_R is defined by the integral

$$\iiint \frac{4dxdyd\phi}{c(1 - x^2 - y^2)^2}.$$

Measurability and measure in Ω_F are defined by the correspondence with E_R.

The geodesics on M_F^{-c} are *represented* in Ψ by sets of hyperbolic lines, congruent hyperbolic lines representing the same geodesic. The geodesics on M_F^{-c} define a flow in Ω_F which can be described simply as follows. Let p be any point of Ω_F, and let e be one of the

* The space Ω_F is metrizable. It may be necessary however to modify the metric thus obtained in order to assure completeness.

congruent elements in Ψ determining p. The element e determines a directed hyperbolic line h. Let s be the sensed hyperbolic length on h measured from the point Q bearing e. Let e_s be the element of h at the point with coordinate s, and let p_s be the point of Ω_F determined by e_s. The transformation $p \rightarrow p_s$ is a 1-1 continuous, measure preserving transformation T_s of Ω_F into itself. The flow thus defined is the flow in Ω_F which we consider and will be called the *geodesic flow* on M_F^{-c}.

A fundamental result and one which is useful in the derivation of transitivity properties is the following:

THEOREM 3.1. *There exist denumerably many periodic geodesics on any M_Γ^{-c}, and the elements on these geodesics form a set which is everywhere dense in Ω_F.*

A periodic geodesic on M_Γ^{-c} is represented in Ψ by an axis (the hyperbolic line joining the fixed points) of a hyperbolic transformation of F, and the statement of the theorem is equivalent to the statement that given arbitrary intervals I_1 and I_2 of U, there exists a hyperbolic transformation of F with one fixed point in I_1 and with the other fixed point in I_2.

This result was first proved in various special cases by Artin [1] and Herglotz, J. Nielsen [1], and Morse [2]. The general result is due to Koebe [1] and Löbell [2]. The proof of it is attained by simple geometrical arguments.

With the aid of the preceding theorem it is easily shown that there exists a transitive geodesic, and the following theorem can be stated:

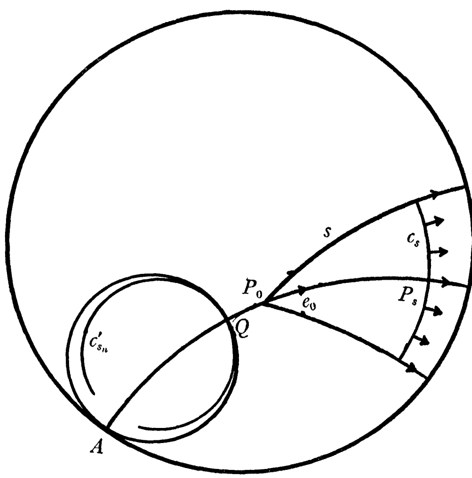
THEOREM 3.2. *The geodesic flow on any M_Γ^{-c} is regionally transitive.*

As in the case of the preceding theorem, this theorem was proved in various special cases by Artin [1] and Herglotz, by Myrberg [1, 2], and by J. Nielsen [1, 2], while the proof in the general case is due to Koebe [1] and Löbell [2].

The problem of permanent regional transitivity on any M_Γ^{-c} can be solved by an analysis of the transitivity properties of the horocycles. Since the method of solving this problem is characteristic of the methods of solving a number of other transitivity problems, it will be described briefly.

A *horocycle* is an ordinary euclidean circle which is internally tangent to U. Its importance here lies in the fact that it is a limiting curve of hyperbolic circles. A *hyperbolic circle* is the locus of points at constant hyperbolic distance from a fixed point (the center) in Ψ and is also a euclidean circle. If a sequence of hyperbolic circles which pass through a fixed point Q of Ψ and whose centers approach a point A of U is given, the limiting curve of this sequence is the horocycle which passes through Q and is internally tangent to U at A. This horocycle will be denoted by $C(Q, A)$, the point A being the *point at infinity* of $C(Q, A)$.

Permanent regional transitivity means that under the geodesic flow any open set D in Ω_F eventually (interpreting s as time) intersects any other open set D^* and after a sufficiently great length of time, positive or negative, the intersection is never empty. Assuming that F is of the first kind, let Ω_F be denoted by Ω_{I}. Since, according to Theorem 3.1, the elements on periodic geodesics are everywhere dense in Ω_{I}, D contains such a periodic element. This element is repre-



sented in Ψ by an element e_0 of a directed axis AB of a hyperbolic transformation T of the group F. Since D is an open set, it contains the set N_0 of Ω_{I} determined by a sector of elements E_0 at the point P_0 bearing e_0 and with e_0 as central element. The set D_s, $(s>0)$, then contains the set N_s of Ω_{I} determined by the elements E_s perpendicular to an arc c_s of a hyperbolic circle with center P_0 and hyperbolic radius s. The arc c_s is that determined by the hyperbolic rays with initial elements in E_0, and the elements E_s are directed outward.

Let Q be any point of the axis AB. The points $T^n(Q)$, $(n=0, \pm 1, \cdots)$, have A and B as limit points and the sequence $0<s_1<s_2<\cdots$, $\lim_{n\to\infty} s_n=+\infty$, can be so chosen that the point P_{s_n} where c_{s_n} cuts across AB is congruent to Q under a power T^{m_n} of T. As a matter of fact, this sequence can be taken in the form

$s_1+n\omega$, where ω is the hyperbolic distance through which T moves a point of the axis AB. The transformed arc $c'_{s_n}=T^{m_n}(c_{s_n})$ passes through Q, and as n becomes infinite, c'_{s_n} approaches the horocycle $C(Q, A)$. The transformed sets $E'_{s_n}=T^{m_n}(E_{s_n})$ have as limit elements the elements perpendicular to $C(Q, A)$ and directed outward. Congruent elements in Ψ determine the same points in Ω_{I}, so the set E'_{s_n} determines points in N_{s_n}, which is a subset of D_{s_n}. If the set of elements outwardly perpendicular to the horocycle $C(Q, A)$ determines an everywhere dense set in Ω_{I}, the set N_{s_n} will tend with increasing n towards a dense distribution in Ω_{I} and for sufficiently large n will contain points in D^*. If these properties hold for any choice of Q on the axis AB, it can be shown that the flow is permanently regionally transitive in the positive sense. Similar reasoning can be applied to the case $s<0$, where, however, outwardly directed normals must be replaced by inwardly directed normals.

The density properties of the sets of elements perpendicular to a horocycle can be determined by the density properties of the elements on the directed horocycle itself. The directed horocycles are called *right* or *left* according as the orientation is clockwise or counterclockwise. A right (left) horocycle is *transitive* if its elements determine an everywhere dense set in Ω_F.

The analysis of the transitivity properties of the horocycles has been carried through by the author (cf. Hedlund [**4**]). The methods are geometrically simple and depend on Theorem 3.1. With the aid of these results it is possible to establish the following theorem (essentially Theorem 3.1 of Hedlund [**4**]):

THEOREM 3.3. *The geodesic flow on any M_{I}^{-c} is permanently regionally transitive.*

It is of interest to note that if the fundamental region lies, together with its boundary, interior to U, all the right (left) horocycles are transitive and they define a flow in the three-dimensional space Ω_{I} in which all the motions are transitive.

As to metrical results with respect to manifolds M_F^{-c}, with an exception to be noted these have been derived only under the hypothesis that F has a finite set of generators. The corresponding manifolds will be denoted by $M_{\mathrm{I}f}^{-c}$. In the case of an $M_{\mathrm{I}f}^{-c}$, the (hyperbolic) area of the fundamental region is finite, which implies that the corresponding element space $\Omega_{\mathrm{I}f}$ is of finite measure. This fact seems to play an important role in the derivation of the results to be stated.

Topometric transitivity was first proved by Artin [**1**] and Herglotz, and by Myrberg [**1**] for a special group. Later Myrberg [**2, 3**] showed that topometric transitivity held for a large class of manifolds $M_{\text{I}f}^{-c}$. But these results are all included in the following theorem which is due to E. Hopf:

THEOREM 3.4. *The geodesic flow on any* $M_{\text{I}f}^{-c}$ *is metrically transitive.*

This important result was attained by ingenious methods involving harmonic functions, and Professor Hopf described this work in an address to this society in February, 1936 (cf. E. Hopf [**1**]). Previous to this the author had succeeded in establishing metric transitivity for a certain denumerable subclass of the manifolds $M_{\text{I}f}^{-c}$. The methods used involved symbolic characterizations of the geodesics due in one case to Artin [**1**] and Herglotz and in the remaining cases to J. Nielsen [**3**].

It is now possible to add a theorem which completes the solution of our transitivity problems with respect to manifolds $M_{\text{I}f}^{-c}$ (cf. Hedlund [**7**]).

THEOREM 3.5. *The geodesic flow on any* $M_{\text{I}f}^{-c}$ *is a mixture.*

The method of proof is an extension of that used to prove permanent regional transitivity on any M_{I}^{-c}. Since, in the case of an $M_{\text{I}f}^{-c}$, $m\Omega_{\text{I}f}$ is finite, the geodesic flow is a mixture if measurable sets tend with increasing or decreasing time towards homogeneous distribution in $\Omega_{\text{I}f}$ (cf. §2). As indicated in connection with permanent regional transitivity, sets in $\Omega_{\text{I}f}$ tend towards a distribution in sets determined by elements perpendicular to right horocycles. It would seem likely then that the mixture property holds if these sets of perpendicular elements are in some sense equidistributed in $\Omega_{\text{I}f}$, or, if the same is true of the elements on the right horocycles. The elements on a right horocycle determine a path in $\Omega_{\text{I}f}$. How is it possible to determine anything about the distribution of such paths in $\Omega_{\text{I}f}$? An obvious way is to show that these paths are the motions of a measure preserving flow in $\Omega_{\text{I}f}$ which is metrically transitive.

Similar to the way in which the directed hyperbolic lines define the geodesic flow T_s, the right horocycles define a flow ${}_RH_s$ which bears a simple relationship to the flow T_s. Due to this relationship it is possible to show that the flow ${}_RH_s$ is metrically transitive if the same is true of the geodesic flow. But, from Theorem 3.4, metrical transitivity of the geodesic flow holds on manifolds $M_{\text{I}f}^{-c}$; thus on such manifolds the right horocycle flow is metrically transitive. It follows from the ergodic theorem of Birkhoff that almost all the right horocycles determine paths which are equidistributed in the sense that the mean time of sojourn of a path in a measurable set $N \subset \Omega_{\text{I}f}$ exists and is equal to $mN/m\Omega_{\mathrm{I}f}$, except for a set of paths of measure zero. With the aid of this, Theorem 3.5 can be proved directly.

An example due to Seidel [1] shows that in the case of manifolds M_{I}^{-c}, permanent regional transitivity is not necessarily accompanied by metric transitivity when the number of generators of F is not finite. Seidel gives an example of a Fuchsian group F which is regionally transitive on U but not metrically transitive on U. It follows that F must be of the first kind and the geodesic flow on the manifold M_{I}^{-c} defined by F is permanently regionally transitive (Theorem 3.3). This flow cannot be metrically transitive, for this would imply the metric transitivity of F on U. The following theorem can be stated:

THEOREM 3.6. *There exist manifolds M_{I}^{-c} on which the geodesic flow is permanently regionally transitive but not metrically transitive.*

As to manifolds M_{II}^{-c}, the fundamental region has an interval of U on its boundary and $m\Omega_{\mathrm{II}}$ is therefore infinite. It is easily shown that regional transitivity cannot hold. It can be shown that on these manifolds almost all the geodesics are unstable in the sense that for both increasing and decreasing s, they eventually leave and remain outside of any finite domain of the manifold. This was first proved for manifolds $M_{\mathrm{I}f}^{-c}$ by E. Hopf [1]. The general result is a corollary of Theorem 4.6 of the following section.

THEOREM 3.7. *Almost all the geodesics on any M_{II}^{-c} are unstable.*

4. **Two-dimensional manifolds of variable negative curvature.** A number of the transitivity properties of the geodesic flows on manifolds of constant negative curvature can be shown to hold on manifolds which are not of constant curvature. The most general case in which transitivity properties have been derived is under some instability condition such as that of Morse (cf. Morse [**3**], Hedlund [**3**]).

To define the manifolds which we consider we again start with the unit circle U, but now assign to its interior the metric

$$ds^2 = \frac{\lambda^2(x, y)(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}, \tag{4.1}$$

where $\lambda(x, y)$ is a function of class C^7 in Ψ and is such that $0 < a \leq \lambda(x, y) \leq b$. The geodesics to be considered are those defined by (4.1).

Now let us assume that $\lambda(x, y)$ is invariant under a Fuchsian group F with principal circle U. Then the metric (4.1) is invariant under the transformations of the group, and if points in Ψ which are congruent under F are considered identical, there is defined a two-dimensional Riemannian manifold $M_F(\lambda)$. These manifolds form an extensive class. They include, in particular, all closed orientable surfaces of genus greater than one and defined by functions of at least class C^3.

The space of elements on $M_F(\lambda)$ can be taken as the space Ω_F defined in the case of constant negative curvature. The definition of measure in Ω_F is like that of the case of constant curvature except that the volume element now used is

$$\frac{\lambda^2(x, y)dxdyd\phi}{(1 - x^2 - y^2)^2}.$$

The flow to be considered in Ω_F is that defined by the geodesic determined by (4.1), and it is a continuous measure preserving flow.

In the terminology of Morse [3], two curves in Ψ are of the same *type* if there exists a constant d such that any point of either one of these curves is at a hyperbolic distance less than d from some point of the other curve. It can be shown that given an arbitrary hyperbolic line h in Ψ, there exists a geodesic g defined by (4.1) such that g and h are of the same type. The geodesics defined by (4.1) in Ψ satisfy the condition of *unicity* if there is just one of the type of a given hyperbolic line. The condition of unicity is satisfied if the curvature is negative, but, more generally, as shown by Morse [3], unicity is implied by a condition of uniform instability which is defined in terms of the equations of variation of the geodesics.

Let $M_F{}^u$ denote a manifold $M_F(\lambda)$ for which the condition of unicity is satisfied. A one-to-one correspondence between the geodesics on $M_F{}^{-c}$ and those on $M_F{}^u$ can be defined by means of the correspondence between the hyperbolic lines and geodesics of (4.1) in Ψ. This correspondence preserves many properties of the geodesics. To a periodic geodesic on $M_F{}^{-c}$ corresponds a periodic geodesic on $M_F{}^u$; to a transitive geodesic on $M_F{}^{-c}$ corresponds a transitive geodesic on $M_F{}^u$. Thus if $M_F{}^u$ is denoted by $M_{\mathrm{I}}{}^u$ when F is of the first kind, the following theorem is implied by Theorem 3.2:

THEOREM 4.1. *The geodesic flow on any $M_{\mathrm{I}}{}^u$ is regionally transitive.*

This theorem is essentially due to Morse [3]. It includes a result due to Birkhoff ([1], pp. 238–248), who had previously shown the existence of transitive geodesics on certain surfaces of nonpositive variable curvature. It is possible to establish regional transitivity under a condition of ray instability which differs from the instability condition of Morse (cf. Hedlund [3]).

Until recently, the preceding theorem was the only transitivity property known to hold without the restriction that the curvature be constant. It is now possible to add a number of results. These have been obtained only under the assumption that the curvature of the manifold lies between two negative constants, but it seems likely that they can be extended to hold under a condition approximating that of uniform instability. The manifolds will be denoted by M_{I}^n, M_{II}^n or $M_{\mathrm{I}f}^n$ according to the properties of the Fuchsian group under which (4.1) is invariant. The following theorem was proved by Grant [1]:

THEOREM 4.2. *The geodesic flow on any manifold M_{I}^n is permanently regionally transitive.*

This was derived by extending the notion of horocycles to the case under consideration. The hyperbolic circles are replaced by geodesic circles. If we consider a sequence of geodesic circles, all passing through a fixed point P of Ψ and with centers approaching a point A of U, the geodesic circles approach a limiting curve which can be shown to have many of the properties of the horocycles. These *generalized horocycles* will be referred to simply as horocycles.

By arguments similar to those given in the case of constant negative curvature, sectors of elements tend, under the geodesic flow on M_{I}^n, towards a distribution along elements perpendicular to horocycles. An analysis of the transitivity properties of the horocycles then yields the stated theorem.

These methods yield more than this, however, in the case of manifolds $M_{\mathrm{I}f}^n$. The following metrical result can be stated (cf. Hedlund [6]):

THEOREM 4.3. *The geodesic flow on any manifold $M_{\mathrm{I}f}^n$ is topometrically transitive.*

Topometric transitivity is the property that any set N of positive measure in $\Omega_{\mathrm{I}f}$ eventually intersects any given open set D in a non-empty set. Such a set N does not necessarily contain all the points of $\Omega_{\mathrm{I}f}$ determined by a sector of elements in Ψ, but this condition can be approximated. Some one of the sets of elements E_{P_0} at a point P_0 of Ψ must contain a linearly measurable subset $E_{P_0}(N)$ determining points of N and of positive linear measure (that is, linear measure in terms of the angular coordinate ϕ at P_0). There must be an element e_0 belonging to $E_{P_0}(N)$ at which the linear metric density of the set $E_{P_0}(N)$ is unity. A sector with e_0 as central element can be so chosen that the ratio of the linear measure of the subset of $E_{P_0}(N)$ in the sector to the measure of the whole sector is nearly 1.

Furthermore, since $m\Omega_{If}$ is finite, almost all points of Ω_{If} are on motions which are stable in the sense of Poisson for both positive and negative time (cf. Birkhoff [1], p. 190). This follows essentially from a well known recurrence theorem of Poincaré. Thus the element e_0 can be chosen as an element of a motion which is stable in the sense of Poisson. Let AB be the directed geodesic in Ψ determined by e_0. Let s be the sensed arc length (as measured by 4.1) on AB measured from P_0, and let e_s be the element of AB at the point with coordinate s. The Poisson stability implies the existence of a sequence $s_1 < s_2 < \cdots$, $\lim_{n\to\infty} s_n = +\infty$, such that $e_{s_n} = e_n$ has a congruent element e_n' with $\lim_{n\to\infty} e_n' = e_0$.

The arguments are now somewhat similar to those used in proving permanent regional transitivity on manifolds of constant negative curvature. Here the arc c_s is an arc of a geodesic circle, and the set N_{s_n} does not contain all the points determined by the elements outwardly perpendicular to c_{s_n}, but only those determined by elements $E_{P_0}(N)_{s_n}$ on geodesic rays with initial elements in $E_{P_0}(N)$. However, if T_n denotes the transformation of F such that $T_n(e_n) = e_n'$, it can be shown that the elements $T_n[E_{P_0}(N)_{s_n}]$ have as limit elements all the elements perpendicular to the (generalized) horocycle which passes through P_0 and has A as point at infinity. It can be shown that these perpendicular elements determine a dense set in Ω_{If}, and the stated theorem follows.

There is a difficulty in the present case which is not encountered in the case of constant curvature. This lies in the relationship between the sets of elements $E_{P_0}(N)$ and $E_{P_0}(N)_{s_n}$; more exactly, in the relationship between the linear measure of the set $E_{P_0}(N)$ and the measure (linear on c_{s_n} measured in terms of hyperbolic length) of the set bearing $E_{P_0}(N)_{s_n}$. The proof of the desired relationship is not simple and involves an analysis of the dependence of the function $G(r, \theta)$ on θ, where $G(r, \theta)$ is defined by the quadratic form

$$ds^2 = dr^2 + G^2(r, \theta)d\theta^2$$

obtained by setting up geodesic polar coordinates with P_0 as center.

The following theorem is essentially a corollary of the preceding theorem, the statement with regard to non-orientable surfaces being derivable by the same methods.

THEOREM 4.4. *Almost all the geodesics on any closed surface of class C^3 and of negative curvature are transitive.*

We state without proof the following theorem which can be derived by these methods:

THEOREM 4.5. *The geodesic flow on any closed surface of class* C^3 *and of negative curvature is permanently regionally transitive.*

These theorems complete the known results concerning transitivity properties of the geodesic flows on two-dimensional manifolds. Many problems remain unsolved. The manifolds M_I^u, on which, according to Theorem 4.1, there exist transitive geodesics, form a class which is restricted in a topological as well as in a dynamical sense. For example, there is no known analytic two-dimensional manifold which is homeomorphic to a sphere or to a torus and has on it a transitive geodesic. The interrelation of the transitivity problem with the difficult problem of stability has been pointed out by Birkhoff [**3**].

By considering Fuchsian groups of the second kind we define manifolds M_{II}^n. As in the case of constant curvature, these are all open and the geodesic flow on any such manifold is not regionally transitive. Let a geodesic on M_{II}^n be *unstable* if, given any finite (compact) region on M_{II}^n, the points of the geodesic which lie in this region lie on a finite segment of the geodesic. Then the following theorem can be stated (cf. Hedlund [**6**]):

THEOREM 4.6. *Almost all the geodesics on any* M_{II}^n *are unstable.*

The method of proof is again based on horocycles.

Manifolds M_{II}^n have properties similar to the surfaces of negative curvature which Hadamard constructed (Hadamard [**1**]). A preliminary survey indicates that the methods used here can be applied to these Hadamard surfaces and that the perfect sets of geodesics constructed by Hadamard form sets of measure zero.

5. **Three-dimensional manifolds of constant negative curvature.** By assigning the metric

$$ds^2 = \frac{4(dx^2 + dy^2 + dz^2)}{c(1 - x^2 - y^2 - z^2)^2}$$

to the interior of the unit sphere $x^2+y^2+z^2=1$, and by identifying points which are congruent under the transformations of a properly discontinuous group which leaves this metric invariant, it is possible to define three-dimensional manifolds of constant negative curvature (cf. Tuller [**1**], where other references will be found). In contrast to the two-dimensional case, little seems to be known about the possible topological types of such manifolds. It is only recently that Löbell

[6] has constructed examples of this kind such that the manifolds are closed.

However, several of the methods which have been used in solving transitivity problems in connection with two-dimensional manifolds can be applied successfully to the case of three dimensions. Löbell [4] has derived theorems analogous to Theorems 3.1 and 3.2. Tuller [1] has shown that if almost all the geodesics are stable (where the definition of stable is analogous to that given in §4) almost all of them are transitive. Thus topometric transitivity holds on such manifolds which, in particular, include all those which are closed. Moreover, the work of Tuller indicates that under conditions similar to those in which the results hold in two dimensions, permanent regional and permanent topometric transitivity hold. E. Hopf (Zentralblatt für Mathematik, vol. 18, p. 273) states that his methods can be extended so as to prove metric transitivity in a large number of cases. The problem of mixture remains unsolved.

There appear to be no results concerning transitivity properties of the geodesics on three-dimensional manifolds which are not of constant curvature.

6. **Symbolic dynamics.** Symbolic methods have been used frequently in the derivation of transitivity properties of geodesics. These involve a characterization of the geodesic by an unending sequence of symbols called a symbolic trajectory. To a transitive geodesic corresponds a transitive symbolic trajectory; that is, one which contains every possible finite block, subject to certain rules of admissibility determined by the manifold under consideration. Conversely, to a transitive symbolic trajectory corresponds a transitive geodesic. With such a characterization available it is often a simple matter to construct a transitive symbolic trajectory, and thus prove the existence of a transitive geodesic.

In the case of the modular group (the interior of the unit circle being replaced by the upper half-plane with the Poincaré metric $(dx^2+dy^2)/y^2$) Artin [1] and Herglotz devised a symbolic characterization of the geodesics and with the aid of this characterization proved not only regional but topometric transitivity. Myrberg [1] independently derived similar results. Further analysis of this symbolic characterization enabled the author [2] to prove that metric transitivity holds in this case.

In the case of certain Fuchsian groups with symmetric fundamental region, Nielsen [2] employed symbolic methods to prove regional transitivity. The symbolic characterization of Nielsen (as given in [3]) was used by the author [1] in proving metric transitivity.

Koebe ([1], IV) has developed symbolisms in connection with general two-dimensional manifolds of constant negative curvature and used them to prove regional transitivity.

A symbolic analysis of the geodesics on certain surfaces of non-positive variable curvature and a proof of regional transitivity by means of this symbolism has been given by Birkhoff ([1], pp. 238–248). As shown by Birkhoff, much more than the existence of transitive geodesics can be inferred from the symbolic characterization. The symbolism enables one to dominate the problems concerning the qualitative behavior in the large of the geodesics.

The development of a symbolic theory apart from its dynamical significance has recently been begun by Morse and the author (cf. Morse [4]). This initial work includes an extensive analysis of transitive symbolic trajectories. The full scope of these symbolic methods in dynamics is yet to be determined.

Bibliography

E. Artin

1. *Ein mechanisches System mit quasiergodischen Bahnen*, Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität, vol. 3 (1924), pp. 170–175.

G. D. Birkhoff

1. *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9, New York, 1927.
2. *Proof of a recurrence theorem for strongly transitive systems and proof of the ergodic theorem*, Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 650–660.
3. *Probability and physical systems*, this Bulletin, vol. 38 (1932), pp. 361–379.
4. (With P. A. Smith.) *Structure analysis of surface transformations*, Journal de Mathématiques Pures et Appliquées, (9), vol. 7 (1928), pp. 345–379.

L, R. Ford

1. *Automorphic Functions*, New York, 1929.

A. Grant

1. *Asymptotic transitivity on surfaces of variable negative curvature*, Bryn Mawr thesis, 1937

J. Hadamard

1. *Les surfaces à courbures opposées et leur lignes géodésiques*, Journal de Mathématiques Pures et Appliquées, (5), vol. 4 (1898), pp. 27–74.

G. A. Hedlund

1. *On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature*, Annals of Mathematics, (2), vol. 35 (1934), pp. 787–808.
2. *A metrically transitive group defined by the modular group*, American Journal of Mathematics, vol. 57 (1935), pp. 668–678.
3. *Two-dimensional manifolds and transitivity*, Annals of Mathematics, (2), vol. 37 (1936), pp. 534–542.
4. *Fuchsian groups and transitive horocycles*, Duke Mathematical Journal, vol. 2 (1936), pp. 530–542.
5. See M. Morse [4].

6. *On the measure of geodesic types on surfaces of negative curvature*, Duke Mathematical Journal, vol. 5 (1939).
7. *Fuchsian groups and mixtures*, Annals of Mathematics, (2), vol. 40 (1939).

E. Hopf
1. *Fuchsian groups and ergodic theory*, Transactions of this Society, vol. 39 (1936), pp. 299–314.
2. *Ergodentheorie*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 5, 1937.

H. Hopf and W. Rinow
1. *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Commentarii Mathematici Helvetici, vol. 3 (1931), pp. 209–225.

P. Koebe
1. *Riemannsche Mannigfaltigkeiten und nicht euklidische Raumformen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, I, 1927, pp. 164–196; II and III, 1928, pp. 345–442; IV, 1929, pp. 414–457; V and VI, 1930, pp. 304–364, 504–541; VII, 1931, pp. 506–534.

F. Löbell
1. *Die überall regulären unbegrenzten Flächen fester Krümmung*, Dissertation, Tübungen, 1927.
2. *Über die geodätischen Linien der Clifford-Kleinschen Flächen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 572–607.
3. *Zur Frage der Struktur der geschlossenen geodätischen Linien in den offenen Clifford-Kleinschen Flächen mit positiver Charakteristik*, Journal für die reine und angewandte Mathematik, vol. 162 (1930), pp. 125–131.
4. *Einige Eigenschaften der Geraden in gewissen Clifford-Kleinschen Räumen*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1930, pp. 556–569.
5. *Beispiele geschlossener dreidimensionaler Clifford-Kleinscher Räume negativer Krümmung*, Berichte über die Verhandlungen der sächsischen Akademie der Wissenschaften zu Leipzig, vol. 83 (1931), pp. 167–174.

Marston Morse
1. *Recurrent geodesics on a surface of negative curvature*, Transactions of this Society, vol. 22 (1921), pp. 84–100.
2. *A fundamental class of geodesics on any closed surface of genus greater than one*, Transactions of this Society, vol. 26 (1924), pp. 25–61.
3. *Instability and transitivity*, Journal de Mathématiques Pures et Appliquées, (9), vol. 14 (1935), pp. 49–71.
4. (With G. A. Hedlund) *Symbolic dynamics*, American Journal of Mathematics, vol. 60 (1938), pp. 815–866.

P. J. Myrberg
1. *Einige Anwendungen der Kettenbrüche in der Theorie der binären quadratischen Formen und der elliptischen Modulfunktionen*, Annales Academiae Scientiarum Fennicae, (A), vol. 23 (1925).
2. *Ein Satz über die fuchsschen Gruppen und seine Anwendung in der Funktionentheorie*, Annales Academiae Scientiarum Fennicae, (A), vol. 32 (1929).
3. *Ein Approximationssatz für die fuchsschen Gruppen*, Acta Mathematica, vol. 57 (1931), pp. 389–409.

J. von Neumann
1. *Einige Sätze über messbare Abbildungen*, Annals of Mathematics, (2), vol. 33 (1932), pp. 574–586.

J. Nielsen
1. *Om geodaetiske Linier i lukkede mangfoldigheder med konstant negativ Krumning*, Matematisk Tidsskrift, B, 1925, pp. 37–44.
2. *Zur Topologie der geschlossenen Flächen*, Mathematikerkongresses i Kopenhavn (1925), pp. 263–276.
3. *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Mathematica, vol. 50 (1927), pp. 189–358.

H. Poincaré
1. *Sur les lignes géodésiques des surfaces convexes*, Transactions of this Society, vol. 6 (1905), pp. 237–274.

W. Rinow
See H. Hopf and W. Rinow.

W. Seidel
1. *On a metric property of Fuchsian groups*, Proceedings of the National Academy of Sciences, vol. 21 (1935), pp. 475–478.

P. A. Smith
See G. D. Birkhoff [**4**].

A. Tuller
1. *The measure of transitive geodesics on certain three-dimensional manifolds*, Duke Mathematical Journal, vol. 4 (1938), pp. 78–94.

The Institute for Advanced Study