BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 77, Number 6, November 1971

ERGODIC THEORY AND THE GEODESIC FLOW ON SURFACES OF CONSTANT NEGATIVE CURVATURE

BY EBERHARD HOPF[1]

1. **Introduction and abstract.** Famous investigations on the theory of surfaces of constant negative curvature have been carried out around the turn of the century by F. Klein and H. Poincaré in connection with complex function theory. The theory of the geodesics in the large on such surfaces was developed later in the famous memoirs by P. Koebe. This theory is purely topological. The measure-theoretical point of view became dominant in the late thirties after the advent of ergodic theory, and the papers of G. A. Hedlund and E. Hopf [2] on the ergodic character of the geodesic flow came into being. The present paper is an elaboration of the author's Gibbs lecture of this year and at the same time of the author's paper of 1939 on the subject, at least of its part concerning constant negative curvature. The author has had complaints about too much detail missing in the presentation of the material in the latter paper. This has been rectified in the present paper.

The very deep and very important recent papers by Sinai and Anosov on the subject are not touched upon in this paper. They are connected with the second phase of ergodic theory which came into being by the introduction of the notion of entropy into ergodic theory.

2. **Two-dimensional hyperbolic geometry.** The well-known Beltrami-Poincaré model of the hyperbolic plane is the interior of the unit circle endowed with the Riemannian metric

$$ds^2 = \frac{dx_1^2 + dx_2^2}{(1 - x_1^2 - x_2^2)^2} \tag{1}$$

which has curvature minus one. The isometries in this geometry are those Moebius transformations that map $x_1^2+x_2^2<1$ onto itself. An isometry is completely determined by the requirement that it carry a given orthogonal two-leg (ordered pair of directed line-elements with the same carrier point) again into such a thing. Line-elements

This is the text of the forty-fourth Josiah Willard Gibbs Lecture delivered before the Seventy-seventh Annual Meeting of the Society in Atlantic City, New Jersey, on January 21, 1971; received by the editors May 7, 1971.

[1] This paper was prepared under NSF grant 48-296-49.

are always understood to be directed. Of course, isometries leave all quantities unchanged which are determined by ds only, angles (= Euclidean angles), element of area

$$dA = \frac{4dx_1\,dx_2}{(1 - x_1^2 - x_2^2)^2}\,. \tag{2}$$

Geodesics are carried again into geodesics. They are the arcs in $x_1^2+x_2^2<1$ of the circles orthogonal to the unit circle. Hyperbolic distance between two points $x = (x_1, x_2)$ and $x' = (x_1', x_2')$ is denoted by $s(x, x')$.

Consider now the three-dimensional space of line-elements e in the hyperbolic plane. Each isometry of this plane induces, of course, a mapping of that space onto itself. We introduce as metric in the line-element space the expression

$$d\sigma^2 = ds^2 + d\chi^2 \tag{3}$$

where $d\chi$ is determined in the following way. Consider two nearby elements e, e' with carrier points x, x'. Move e from x to x' by parallel displacement, in other words, move the element to x' along the geodesic from x to x' in such a way that its direction always makes the same angle with the geodesic. $d\chi$ is then the angle between the element e in its final position and e'. Obviously $|d\chi|$ is independent of the order of the two elements. $d\sigma$ is therefore a Riemannian metric in e-space. $d\sigma$ is even invariant under the mappings induced by the isometries of the plane since the operation of parallel displacement is invariant under isometries. In other words, those mappings are themselves isometries in e-space relative to $d\sigma$. We denote by $\sigma(e, e')$ the invariant distance defined by $d\sigma$ in e-space. The invariant element of volume measure induced by $d\sigma$ in e-space is found to be

$$dm = dA\,d\phi \tag{4}$$

where $d\phi$ is the angle-differential in a point of the plane.

If we consider in e-space the motion along the directed geodesics with speed $ds/dt = 1$ we obtain a one-parameter group of mappings T^te, $T^{t+s} = T^tT^s$, of e-space onto itself: T^te is the position attained by e after it has moved t units along the geodesic determined by it. This is the geodesic flow in the e-space of the hyperbolic plane. It is a well-known fact of differential geometry that the geodesic flow on a surface leaves the element of measure (4) in e-space invariant. The same is, of course, true about the Lebesgue measure m determined by it in e-space. m is invariant not only under isometries but also under the geodesic flow.

We need the following simple coordinate-representation of the geodesics and of the flow along them. We assume throughout that a geodesic is directed. A geodesic is characterized by its points of infinity θ^-, θ^+ on the unit circle, $\theta^- \neq \theta^+$, where θ is an angle (mod 2π). And a line-element e is characterized by the geodesic determined by it and by the position of its carrier point on this geodesic (orthogonal circular arc). This position is in turn characterized by the distance $s \gtrless 0$ of the point from the Euclidean midpoint of this directed arc,

$$e = (\theta^-, \theta^+, s). \tag{5}$$

The geodesic flow is then simply

$$T^t e = (\theta^-, \theta^+, s + t). \tag{6}$$

The invariant measure-differential dm in e-space has the simple form

$$dm = \rho(\theta^-, \theta^+)\, d\theta^-\, d\theta^+\, ds. \tag{7}$$

That ρ is independent of s follows from the invariance of dm under T^t. Positivity and continuity are the only properties of this function ρ which are needed in what follows.

We also need the following fundamental fact about hyperbolic geometry: To every geodesic and to every point x there exists precisely one geodesic through x that has the same point plus infinity on the unit circle. Two geodesics with the same point plus infinity we call positively asymptotic (to each other). Analogously is negative asymptoticity defined. The asymptotic character is expressed by the following:

PRINCIPAL LEMMA 1. *If two line-elements e, e' determine positively asymptotic geodesics then there exists a number a (depending on e, e') such that*

$$\sigma(T^{t+a}e,\ T^t e') \to 0$$

as $t \to \infty$. There is an analogous fact in the case of negative asymptoticity.

This is most easily proved if we map the interior of $x_1^2 + x_2^2 = 1$ by a Moebius transformation onto the upper half $y_2 > 0$ of a y-plane, $y = (y_1, y_2)$. We do this in such a way that the common point $+\infty$ of both geodesics goes into $y = \infty$. The metric becomes $ds^2 = (dy_1^2 + dy_2^2)/y_2^2$. The two geodesics become two straight lines l, l' orthogonal to the line $y_2 = 0$. Consider two tangent elements e on l, e' on l'. Evidently the desired value of a is the one for which $T^a e$ and e' have the same coordinate y_2. It is geometrically obvious that the hyperbolic distance $\int ds$ between the carrier points of the elements $T^{t+a}e$, $T^t e'$ tends to

zero as $t \to \infty$. At the same time the geodesic arc joining them becomes straighter and straighter in the Euclidean sense. Now, the distance

$$\sigma(T^{t+a}e,\ T^t e')$$

is not greater than the length of any path in e-space joining the two elements. Choose for this path the geodesic arc between the carrier points plus parallel displacement of the first element along it ($d\sigma = ds$), and then the remaining rotation of the element into the second element ($d\sigma - |d\chi|$). This makes the lemma obvious.

3. **Complete surfaces Σ of curvature minus one.** The classical construction of these surfaces is the following. It starts with a group $\mathfrak{G}$ of isometries S in the hyperbolic plane. We suppose that it is *discrete* or, in other words, that it does not contain isometries arbitrarily close to the identity. Well-known consequences of the discreteness are: (1) $\mathfrak{G}$ is countable, (2) the set of distinct points Sx, $S \in \mathfrak{G}$, congruent to a point x within the unit circle does not have a cluster point within this circle, (3) the fixed points of the isometries $S \in \mathfrak{G}$ different from the identity form a countable set of single points and orthogonal arcs (geodesic arcs). $\mathfrak{G}$ possesses a fundamental domain D in the unit disk. The classical construction of a simple fundamental domain is the following. Choose a point x^0 within the unit circle which is no fixed point of any $S \in \mathfrak{G}$ different from the identity. Then the set of all points x which satisfy the inequalities

$$s(x, x^0) < s(x, Sx^0)$$

for all $S \in \mathfrak{G}$ different from the identity is a fundamental domain or, rather, the interior of a fundamental domain $D = D(x^0)$ for $\mathfrak{G}$. D is geodesically convex. Its boundary is formed by countably many geodesic arcs and, perhaps, parts of the unit circle itself. This is so because the equation $s(x, x^1) = s(x, x^2)$ determines a geodesic in hyperbolic geometry. Fixed points of any $S \in \mathfrak{G}$ different from the identity can obviously occur only on the boundary of such a D.

We obtain a surface Σ of curvature minus one if we identify all points Sx, $S \in \mathfrak{G}$, congruent to a given point x or if we regard the set

$$p = \{Sx \mid S \in \mathfrak{G}\}$$

as a single point p. Distance between such a point and another such point

$$p' = \{Sx' \mid S \in \mathfrak{G}\}$$

is defined by

$$s(p, p') = \inf_{S\in\mathfrak{G};S'\in\mathfrak{G}} s(Sx, S'x') = \inf_{S\in\mathfrak{G}} s(Sx, x') = \inf_{S'\in\mathfrak{G}} s(x, S'x').$$

The last two equations hold by virtue of the invariance of s. Similarly, the directed line-elements P on Σ are defined by identification of congruent line-elements within the unit circle, and distance of two elements P is defined by

$$\sigma(P, P') = \inf_{S\in\mathfrak{G};\, S'\in\mathfrak{G}} \sigma(Se, S'e')$$

where S, S' mean the isometries induced in e-space and $\mathfrak{G}$ means the corresponding group. The three-dimensional space of line-elements P on Σ is denoted by Ω. All the other quantities which are defined within the unit circle and invariant under isometries define similar quantities on Σ. Plane measurability of a set of points p on Σ means measurability of the set of *all* representatives x in $x_1^2+x_2^2<1$ of all those p. A set in Σ is said to have measure zero if the set of all representative points x has this property. Of course, measure of a general measurable set on Σ has to be defined more carefully. It is defined as the measure $\int dA$ of the intersection of that set of all representatives x with a fundamental domain for $\mathfrak{G}$. It is a consequence of the invariance of $\int dA$ under isometries that the measure defined in this way is independent of the particular fundamental domain. Measure zero, according to this definition, agrees with measure zero as defined a moment ago because the set of all representatives x is the union of all copies under $\mathfrak{G}$ of those intersections and because these copies are countable in number. In a perfectly analoguous way m-measurability and measure m are defined in the space Ω of line-elements P on Σ. The measure m in Ω is invariant under T^t.

The geodesic flow T^tP is unambiguously defined on Σ, or rather, in Ω because

$$ST^te = T^tSe$$

holds for arbitrary isometries induced in Ω. This relation is most easily proved. In fact, S carries geodesic motion into geodesic motion. Hence, either side represents a geodesic motion starting from Se at $t=0$. But the geodesic motion with this initial value is uniquely determined.

A geodesic T^tP is said to be positively asymptotic to another geodesic T^tP' if some representative T^tSe of the first is positively asymptotic to some representative $T^tS'e'$ of the second in the unit disk. We also say in this case that the geodesic streamline is positively asymptotic to the second geodesic streamline.

PRINCIPAL LEMMA 2. *If the geodesic streamlines T^tP, T^tP' through two points P, P' of Ω are positively asymptotic to each other then there exists a number a (depending on P, P') such that*

$$\sigma(T^{t+a}P,\ T^tP') \to 0$$

holds as $t \to \infty$.

This is an easy consequence of Lemma 1 and of the fact that $\sigma(Q,\ Q') \leqq \sigma(e,\ e')$ holds for any respective representatives e, e' of the two points Q, Q' of Ω.

The following fact has to be used.

LEMMA 3. *Let Δ be a set of directions in a point p of a surface Σ. Consider a second point p' on Σ and a geodesic through p' which is positively asymptotic to some geodesic passing through p in a direction belonging to Δ. Denote by Δ' the set of directions in p' of all these geodesics through p'. Then, if Δ has angular measure zero so does Δ'.*

PROOF. Suppose that

$$P = \{Se \mid S \in \mathfrak{G}\}, \qquad P' = \{S'e' \mid S' \in \mathfrak{G}\}.$$

Then there holds

$$T^tP = \{T^tSe \mid S \in \mathfrak{G}\}, \qquad T^tP' = \{T^tS'e' \mid S' \in \mathfrak{G}\}.$$

To say that T^tP is positively asymptotic to T^tP' means exactly that, for some $S \in \mathfrak{G}$ and some $S' \in \mathfrak{G}$, T^tSe is positively asymptotic to $T^tS'e'$. This is equivalent to saying that ST^te is positively asymptotic to $S'T^te'$. In introducing the operator $T^\infty\bar{e}$ for the point $+\infty$ on the unit circle of the geodesic $T^t\bar{e}$ through $\bar{e}$, we can write the last asymptoticity relation (we use the fact that T^∞ as well as T^t commutes with any S) simply

$$S\pi = S'\pi' \tag{8}$$

where

$$\pi = T^\infty e, \qquad \pi' = T^\infty e'. \tag{8'}$$

We suppose that the carrier points x, x' of the elements e, e' are kept fixed,

$$e = (x, \vartheta), \qquad e' = (x', \vartheta')$$

where ϑ, ϑ' are the angles with two fixed directions in the points x, x', respectively. For fixed $S \in \mathfrak{G}$, $S' \in \mathfrak{G}$, (8) represents a one-to-one mapping of the unit circle into itself. As $\mathfrak{G}$ is countable there are only

countably many different mappings (8). Each of these maps a set of measure zero on the unit circle again into such a set because the Moebius transformations of $\mathfrak{G}$ are still one-to-one and bianalytic on this unit circle. Each of the two relations (8′) is bianalytic. Therefore, null sets correspond to null sets under them. Each composite mapping resulting from the relations (8) and (8′)—these are the mappings considered in Lemma 3—does therefore the same. Since there are only countably many such mappings involved, it is plain that Lemma 3 is right.

4. **The two classes of surfaces Σ.** Consider a geodesic motion γ on a surface Σ. We call it positively divergent on Σ if, for $t\to\infty$, $s(p_t, p^0)\to\infty$ holds where p_t is the p-coordinate of T^tP and where p^0 is a fixed point on Σ. Of course, if the statement holds true for one p^0 then it does so for any other fixed point p^0. Quite analogously, a geodesic streamline T^tP in the P-space Ω over Σ is called positively divergent if, for $t\to\infty$, $\sigma(T^tP, P^0)\to\infty$ holds where P^0 is an arbitrarily fixed point in Ω. It is plain that both concepts are equivalent because the general inequalities

$$s(p, p') \leqq \sigma(P, P') \leqq s(p, p') + \pi$$

hold where p, p' are the carrier points of P, P', respectively. The second of these inequalities follows by parallel displacement of P along a geodesic arc from p to p' and by a rotation of the line-element in p'.

An important consequence of the Principal Lemma 2 is: If a geodesic on Σ (streamline in Ω) is positively divergent in $\Sigma(\Omega)$ so is every other geodesic (streamline in Ω) which is positively asymptotic to the first. This statement remains, of course, valid if the word "positively" is replaced by the word "negatively" throughout $(t\to-\infty)$.

Lemma 3 permits the use of the abundance of positively-divergent geodesics for a subdivision of the class of surfaces Σ into two subclasses.

DEFINITION. *A surface Σ is of first class if the divergent geodesics issuing from a fixed point p of Σ form a set of directions in p of angular measure zero. Σ is called of second class if it is not of first class. If this is true for one fixed point $p\in\Sigma$ it is, by virtue of Lemma 3, true for any other fixed point p.*

The possible surfaces Σ of first class represent, topologically, a vast variety of surfaces of constant negative curvature. There are, as is well known, many types of closed surfaces. For instance, every closed orientable surface of genus >1 is topologically represented. A closed

surface Σ is of first class simply because no divergent geodesic exists on such a surface. More generally, the surface is of first class if the group $\mathfrak{G}$ of covering transformations S possesses a fundamental domain D, the closure of which is entirely in $x_1^2+x_2^2<1$. Many surfaces with boundary fall into this category. In this case the problem of the geodesics is either totally or at least partially a hyperbolic billiard problem with elastic reflection at the respective boundaries.

Take, for example, an equilateral hyperbolic triangle with the three interior angles equal to $2\pi/2n=\pi/n$ where n is an integer >3 (the sum of the three angles must be $<\pi$) and take the three hyperbolic reflections at the three sides as generators of a group $\mathfrak{G}$. The condition on the angle insures that the images under $\mathfrak{G}$ of the triangular domain cover the hyperbolic plane simply or, in other words, that this triangle is actually a fundamental domain for this group $\mathfrak{G}$. The surface Σ generated by $\mathfrak{G}$ is then this triangular area, and the geodesics problem is the hyperbolic billiard problem.

Quite generally, we may say that a surface Σ is of first class if its area is finite. The reason is this. If Σ has finite area its P-space Ω has finite volume $m(\Omega)=2\pi A$. By virtue of Poincaré's recurrence theorem, almost all streamlines of the geodesic flow in Ω (with invariant measure m) are recurrent. However, if Σ were of second class then, by virtue of the definition of the second class, the divergent streamlines from any fixed point of Σ would form a set of angular measure >0 and, therefore, the set in Ω of all divergent streamlines would have positive measure which would be a contradiction. There are many types of surfaces Σ of finite area but of infinite hyperbolic diameter. A well-known example is that of the modular group $\mathfrak{G}$ which has a geodesic triangle for fundamental domain with one corner or, rather, cusp at infinity (on $x_1^2+x_2^2=1$). The cusp does not prevent the domain from having a finite area.

A surface Σ of second class is realized if the fundamental domain of the group $\mathfrak{G}$ has on its boundary an open arc of the unit circle. Any geodesic ray that ends up on this arc must be divergent on Σ. In fact, such a ray stays ultimately in D, and the point running along it gets farther and farther removed, in the sense of the distance s, from the finite part of the boundary of D. Consequently, its minimal distance from all the points Sx^0, $S\neq$identity, congruent to a fixed point x^0 inside of D (they are outside of D) tends a fortiori to infinity. Obviously, the initial directions of the geodesic rays issuing from a fixed point and ending on that arc fill a whole angle. To the arc on the unit circle there corresponds an infinite funnel of the surface Σ. However, surfaces of the second class can be much more complicated in that they

can reach to infinity in much more intricate ways.

We mention that Koebe's two "kinds" of surfaces Σ are not the same as our two classes. "Kind" is a purely topological and, as such, very natural notion. From the standpoint of ergodic theory, however, the natural division is the one into the two "classes." A more detailed discussion between kinds and classes is found in our original memoir [2]. The first kind is actually larger than the first class.

5. **Ergodic theory and the two classes of surfaces Σ.** We are now ready to begin with the proof of the two principal theorems on the geodesic flow on surfaces Σ.

FIRST THEOREM. *For a surface of first class the geodesic flow is ergodic. In other words, if $f(P)$ and $g(P)>0$ are m-integrable in Ω then*

$$\lim_{\tau\to\infty} \frac{\int_0^\tau f(T^tP)\,dt}{\int_0^\tau g(T^tP)\,dt} = \frac{\int_\Omega f\,dm}{\int_\Omega g\,dm}$$

holds for almost every $P\in\Omega$ in the sense of the measure m. The same holds for the limit at $\tau\to-\infty$.

SECOND THEOREM. *For a surface of second class the geodesic flow is dissipative or, in other words, for almost every $P\in\Omega$ in the sense of the measure m the streamline T^tP is divergent in Ω, positively as well as negatively.*

Both these theorems have a common root, namely, the existence of asymptotic geodesics and the relations between them peculiar to hyperbolic geometry. Their effect is expressed in the following:

MAIN LEMMA 4. *Let B_+ and B_- be two m-measurable sets in the line-element space Ω of a surface Σ. Suppose that they satisfy the following conditions.* (a) *Each set is invariant under the geodesic flow.* (b) *With every streamline in B_+, every streamline positively asymptotic to it is in B_+; the same holds for B_- with respect to negative asymptoticity.* (c) *The set of all points P in one set but not in the other has measure $m=0$. Then, under these conditions, either B_+ and B_- both have measure $m=0$ or their complements in Ω both have measure $m=0$.*

First we prove this lemma and then the two main theorems. It is interesting to observe that the lemma is true if in the hypothesis (c) and in the conclusion the statement "set of measure $m=0$" is replaced by the statement "empty set." In this altered form the lemma is not

only true but almost trivial. In fact, (c) means then that $B_+ = B_-$, and the conclusion says that either $B_+ = B_- = 0$ or $B_+ = B_- = \Omega$ holds. This modified lemma is true simply because to two arbitrary geodesics on Σ there always exists at least one geodesic on Σ which is negatively asymptotic to the first and positively asymptotic to the second (on the universal covering surface there is precisely one such geodesic). If $B_+ = B_-$ is nonvoid then this argument can be applied to an entirely arbitrary streamline and to a streamline in B^+. However, we need the lemma in the form stated because, in both our applications, the exceptional null sets in the hypothesis (c) are, in general, nonvoid.

PROOF OF THE MAIN LEMMA. This proof rests upon the same simple argument but then the exceptional sets of measure zero require careful handling. A set of entire streamlines in the element-space Ω of a surface Σ can be represented in two ways, as a flow-invariant point set in Ω and as a set of pairs of points θ^-, θ^+ on the unit circle which is invariant under the simultaneous transformations $S \in \mathfrak{G}$. Remember that m is the flow-invariant measure in Ω and that $\int\int d\theta^- \, d\theta^+$ is a measure in the space of those pairs θ^-, θ^+. We prove first: A flow-invariant point set in Ω has measure $m = 0$ if and only if the corresponding set of pairs θ^-, θ^+ has measure $\int\int d\theta^- \, d\theta^+ = 0$. To see this denote by C the invariant set in Ω. Let C' be the set of all elements e in the universal covering surface $x_1^2 + x_2^2 < 1$ which represents the elements contained in C. We know that $m(C) = 0$ implies $m(C') = 0$ and vice versa. C' is invariant under the geodesic flow in the covering surface. On using the coordinates θ^-, θ^+, s in (6) for the elements of C' we see that C' is a cylindrical set, with its base in (θ^-, θ^+)-space equal to that set of pairs of points. By virtue of (7), $m(C') = 0$ means that that set has measure $\int\int \rho \, d\theta^- \, d\theta^+ = 0$. As ρ is > 0 and continuous this means that the set has measure $\int\int d\theta^- \, d\theta^+ = 0$. The statement is thereby proved.

The main lemma is now proved as follows. We assume that

$$m(B_-) > 0 \tag{9}$$

and show that $\overline{B}_- = \Omega - B_-$ and $\overline{B}_+ = \Omega - B_+$ have measure $m = 0$. Retaining the letters B_-, B_+ for the sets corresponding to these sets in (θ^-, θ^+)-space we infer from the hypothesis (b) of the lemma that both these sets are product sets in that space,

$$B_- = b_- \times L, \qquad B_+ = L \times b_+, \tag{10}$$

where L is the line of θ and where b_-, b_+ are certain measurable subsets of it. Strictly, we would have to exempt the diagonal-line $\theta^- = \theta^+$

from $L \times L$ since no geodesics correspond to the points of this line. Since, however, it is of measure $\int\int d\theta^- \, d\theta^+ = 0$ it may safely be disregarded in this proof. Hypothesis (c) says that

$$(11) \qquad B_- \cap \overline{B}_+ = b_- \times \bar{b}_+, \qquad \overline{B}_- \cap B_+ = \bar{b}_- \times b_+$$

have measure zero. This may be understood in the sense of the product measure $\int\int d\theta^- \, d\theta^+$. By virtue of (9), b_- has nonzero measure on the θ-line. Consequently, since the first set has product measure zero, $\bar{b}_+$ has θ-measure zero. Since the second set has product measure zero it now follows that $\bar{b}_-$ has θ-measure zero. In other words, the complements of both sets (10) have measure zero. These complements in product space correspond exactly to the sets $\overline{B}_-$, $\overline{B}_+$ in the space Ω. Consequently, these latter sets have measure $m = 0$. The lemma is herewith completely proved.

PROOF OF THE SECOND THEOREM. Denote by B_- (B_+) the set of all elements $P \in \Omega$ on negatively (positively) divergent streamlines. As Σ is of second class, the angular measure $\int d\theta$ of the positively divergent directions in a point $p \in \Sigma$ is positive. Hence there holds

$$m(B^+) = \int\int_{B_+} d\phi \, dA = \int_\Sigma \left\{ \int d\phi \right\} dA > 0.$$

That the two sets B_-, B_+ satisfy hypothesis (c) of the main lemma follows from a general theorem ([1] or [2]): If $T^t P$, $T^0 P = P$, $T^{t+s} = T^t T^s$ is a continuous flow in a complete metric space Ω with invariant measure m (m σ-finite) then the set of all streamlines that are divergent in one direction but not in the other has measure $m = 0$. That hypothesis (c) is satisfied was remarked before. Therefore the conclusion of the main lemma holds, and the second theorem is thereby proved. By virtue of Lemma 3 we obtain a somewhat sharper result: If Σ is of second class then, in any point $p \in \Sigma$, the geodesic rays issuing from p are divergent for almost all initial directions.

PROOF OF THE FIRST THEOREM. Just as in the beginning of the preceding proof we infer that, for a surface Σ of first class, the set of all positively or negatively divergent streamlines in Ω is a set of measure $m = 0$. We have to mention now that the general theorem referred to in the preceding proof is part of the following general theorem [1]:

Under the same hypothesis as in that theorem, Ω splits into two T^t-invariant parts, the conservative and the dissipative part. The first contains almost no positively nor negatively divergent streamlines. The second consists almost exclusively of positively as well as negatively divergent streamlines. In the conservative part,

$$\int_0^\infty g(T^t P)\,dt = \infty \tag{12}$$

holds almost everywhere for any $g(P) > 0$. In our present case, geodesic flow on a surface Σ of first class, the flow is purely conservative. For a conservative flow with invariant measure m the ergodic theorem states this [1]: If $f(P)$, $g(P) > 0$ are m-integrable in Ω, then

$$\lim_{\tau\to\infty} q_\tau(P) = f^*(P), \qquad q_\tau(P) = \frac{\int_0^\tau f(T^t P)\,dt}{\int_0^\tau g(T^t P)\,dt}, \tag{13}$$

exists almost everywhere (m) in Ω; gf^* is m-integrable, f^* is T^t-invariant and satisfies the relation

$$\int_\Omega gf^* h\,dm = \int_\Omega fh\,dm \tag{14}$$

for every bounded and measurable $h(P)$ that is invariant under T^t. The average in the past

$$\lim_{t\to-\infty} q_\tau(P) = f_*(P) \tag{15}$$

exists almost everywhere too (apply theorem to the flow $\hat{T}^t = T^{-t}$ which has the same invariant functions as T^t), and $f_*(P)$ satisfies the same relation (14) as $f^*(P)$. Consequently,

$$\int_\Omega g(f^* - f_*)h\,dm = 0$$

must hold for every bounded invariant h, so for instance, for $h = \operatorname{sign}(f^* - f_*)$. Hence,

$$f_*(P) = f^*(P) \tag{16}$$

must hold almost everywhere.

Our aim is to show that $f^*(P)$ is constant almost everywhere. From (14), $h = 1$, it would then follow that this constant has the value $\int f\,dm / \int g\,dm$. To prove constancy of f^* for any m-integrable f it suffices to prove this for every f in a set of f's that is dense in L_m, the space of all functions f m-integrable over Ω. Reason: The linear operator $f^* = T^* f$ (g is kept fixed) satisfies

$$\int_\Omega g\,|f^*|\,dm \leqq \int_\Omega |f|\,dm$$

(apply (14) with $h = \operatorname{sign} f^*$). We use this fact as follows in our present case (Σ, Ω). Choose the fixed function $g > 0$ such that

$$\frac{g(P') - g(P)}{g(P')} \to 0 \qquad \text{as } \sigma(P, P') \to 0 \tag{17}$$

holds *uniformly* with respect to P, P' in Ω. In the case where Σ has finite area we can simply choose $g = 1$. For functions f we take only those which have the similar property that

$$\frac{f(P') - f(P)}{g(P')} \to 0 \qquad \text{as } \sigma(P, P') \to 0 \tag{18}$$

holds uniformly in Ω. In the case where Σ has finite area and where $g = 1$ this simply means that f is uniformly continuous in Ω. It can be shown by means of classical arguments that the set of these f is dense in L_m.

After these preparations we now turn to the main point of the proof. Consider f, g as indicated above. We show that

$$\frac{\int_0^\tau f(T^t P)\, dt}{\int_0^\tau g(T^t P)\, dt}, \qquad \frac{\int_0^\tau f(T^t P')\, dt}{\int_0^\tau g(T^t P')\, dt} \tag{19}$$

have the same limit as $t \to \infty$ if the two streamlines occurring here are positively asymptotic to each other. By means of (12), it is easy to see that the limit behaviour of the left fraction in (19) is not affected if the variable t is replaced by $t + a$ where a is an arbitrary constant. Therefore it is to be shown that

$$\frac{\int_0^\tau f(T^{t+a} P)\, dt}{\int_0^\tau g(T^{t+a} P)\, dt}, \qquad \frac{\int_0^\tau f(T^t P')\, dt}{\int_0^\tau g(T^t P')\, dt} \tag{20}$$

have the same limit as $\tau \to \infty$. We choose for a the a of the Principal Lemma 2. From this lemma and from (17), (18) it is clear that

$$\frac{f(T^{t+a} P) - f(T^t P')}{g(T^t P')}, \qquad \frac{g(T^{t+a} P) - g(T^t P')}{g(T^t P')}$$

go to zero as $t \to \infty$. On the other hand, a calculation shows that the difference right minus left in (20) equals the difference of

$$\frac{\int_0^\tau ((f_t' - f_t)/g_t')g_t'\,dt}{\int_0^\tau g_t'\,dt} \tag{21}$$

and

$$\frac{\int_0^\tau f_t\,dt}{\int_0^\tau g_t\,dt} \cdot \frac{\int_0^\tau ((g_t' - g_t)/g_t')g_t'\,dt}{\int_0^\tau g_t'\,dt} \tag{22}$$

where there is

$$f_t = f(T^{t+a}P), \qquad f_t' = f(T^tP')$$

and, similarly, for g_t, g_t'. From (12) it follows that

$$\int_0^\infty g_t'\,dt = \infty,$$

and hence it follows that (21) and the right-hand factor in (22) go to zero as $\tau \to \infty$. The left-hand factor in (22) remains bounded provided that

$$f(P)/g(P) < C \qquad \text{in } \Omega. \tag{23}$$

This additional restriction does, however, not invalidate the previous conclusion that, for $g(P) > 0$ fixed, the functions $f(P)$ admitted are dense in L_m. It has therefore been shown that the difference of the two fractions in (19) goes to zero as $\tau \to \infty$ provided that the functions f, g admitted satisfy all the conditions stated. The same is true as $\tau \to -\infty$ if the two streamlines are negatively asymptotic. Consequently, the two invariant sets

$$B_- = \{P \mid f_*(P) \geqq C\}, \qquad B_+ = \{P \mid f^*(P) \geqq c\}$$

satisfy hypothesis (b) of the Main Lemma 4 no matter what the value of the constant c is. (16) implies the validity of hypothesis (c). From that lemma we can therefore infer that either B_+ or its complement has measure $m = 0$. This holds for any value of c or, in other words, f^* is constant almost everywhere. To be sure, constancy of f^* has been proved for a set of functions f dense in L_m but, as mentioned before, from that there follows this constancy for every $f \in L_m$. From the

general ergodic theorem as mentioned before there now follows the truth of the first theorem.

References

1. Eberhard Hopf, *Ergodentheorie*, Ergebnisse der Mathematik, Band 5, no. 2, Springer-Verlag, Berlin, 1937.

2. ———, *Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung*, Ber. Verh. Sächs. Akad. Wiss. Leipzig **91** (1939), 261–304. MR **1**, 243.

Indiana University, Bloomington, Indiana 47401