SURFACES WITHOUT CONJUGATE POINTS(1)

BY

L. W. GREEN

1. **Introduction.** The fact the variational problems of dynamics lead naturally to the investigation of shortest paths in abstract spaces has been a stimulus of much research into the behavior of such paths in the large. In particular, the case where such paths are unique, at least for the universal covering space of the domain in question, has been treated by, among others, Hadamard, Morse, Cartan, and in its abstract formulations by Busemann. On the other hand, the dynamical investigations of Morse, Hedlund, and E. Hopf were concerned with manifolds of constant negative curvature, or with spaces which exhibited many of the properties of such manifolds. Morse and Hedlund established a dynamical property (topological transitivity) for a class of surfaces on which the main restriction was that the shortest paths (geodesics) in the universal covering surface be unique. The surfaces they considered had to satisfy the additional dynamical condition of Poisson stability. To show that such surfaces form a large class necessitates appeal to the Poincaré recurrence theorem, which introduces the anomaly of using measure theory to obtain a topological result. The Poisson stability hypothesis, however, was not used directly to establish transitivity: it was needed for the proof of a purely geometric property, namely, that on a surface of the type considered, intersecting geodesics diverge.

It turns out that this hypothesis is superfluous. In the following we prove that, under mild analytical restrictions, the nonexistence of conjugate points on geodesics (equivalent to uniqueness of shortest paths) is sufficient to ensure the divergence of intersecting geodesics. Certain examples due to Hilbert show that this result cannot be extended to the more general spaces which Busemann considers. Using this property, topological transitivity may be established for a wider class of surfaces.

Moreover, the divergence property of intersecting geodesics is established for certain surfaces with poles, even though they may contain geodesics with mutually conjugate points. The conditions under which this happens give restrictions on the set of poles of a surface for which divergence fails. This set has previously been investigated by von Mangoldt and Cohn-Vossen.

The methods are also applicable to the problems of parallels on two-

Presented to the Society, February 28, 1953; received by the editors May 11, 1953.

(1) This paper is part of a thesis submitted by the author in partial fulfillment of the requirements for a Ph.D. from Yale University. The author wishes to express his thanks to Professor G. A. Hedlund for helpful advice and criticism. Part of this work was done while the author was an Atomic Energy Commission Pre-doctoral Fellow.

dimensional Riemannian manifolds. For this an additional hypothesis is used, the nonexistence of focal points (unique perpendicularity). For such surfaces we show that the Gaussian curvature between any two geodesics which remain a bounded distance apart is zero. Thus simply-connected analytic surfaces with any parallels in this sense must be Euclidean. Also, cylinders without focal points which do not "open out" like an hyperboloid of one sheet must have Gaussian curvature zero everywhere. A corollary is that surfaces of the topological type of a torus which have no focal points are flat, a result of Morse and Hedlund. However, E. Hopf showed that for this theorem the hypothesis of nonconjugacy is sufficient. Whether that is also the case for the more general statement about cylinders is not known.

2. **The Jacobi equation.** In this section we shall study the behavior in the large of solutions of the familiar Jacobi differential equation

$$y''(x) + K(x)y(x) = 0, \qquad -\infty < x < \infty. \tag{J}$$

The assumptions throughout will be that $K(x)$ is continuous on the whole real axis, and that $K(x) \geqq -M$ for some positive constant M. We begin with a discussion of the associated Riccati equation

$$u'(x) + u^2(x) + K(x) = 0. \tag{R}$$

The following lemma uses a Sturmian argument.

LEMMA 2.1. *If $u(x)$ is a solution of* (R) *defined for $x \geqq x_0$ $(x \leqq x_0)$, then $u(x)$ is bounded for $x \geqq x_0$ $(x \leqq x_0)$.*

Proof. Choose $b > 0$ such that $b^2 > M$. Suppose that there exists an $x_1 > x_0$ such that $u(x_1) > b$. Then there is a solution $v(x)$ of

$$v'(x) = b^2 - v^2(x) \tag{1}$$

for which $v(x_1) = u(x_1)$. In fact, $v(x) = b \coth (bx - d)$ for a suitably chosen constant d. Subtracting (R) from (1) we get

$$v'(x) - u'(x) = b^2 + K(x) - v^2(x) + u^2(x). \tag{2}$$

For $x = x_1$ the right side of (2) is positive, so

$$v(x) \geqq u(x), \tag{3}$$

at least in a one-sided neighborhood, $x_1 \leqq x \leqq x_1 + \epsilon$, of x_1. But now it is clear that (3) holds for all $x \geqq x_1$, since at any point of intersection of the curves $y = u(x)$, $y = v(x)$, the right side of (2) would again be positive. Since $v(x)$ is bounded for $x \geqq x_1$, we conclude that $u(x)$ is bounded above for $x \geqq x_1$.

If, on the other hand, $u(x_1) < -b$, we compare $u(x)$ with the solution of (1) $w(x) = b \coth (bx - c)$ for which $w(x_1) = u(x_1)$. $w(x)$ is defined for $x < c/b$, but $\lim_{x \to c/b} w(x) = -\infty$. However, exactly the same argument as that employed above shows that (3) holds for $x \geqq x_1$, so that the assumption that $u(x_1) < -b$

is incompatible with the hypothesis that $u(x)$ is defined for all $x \geqq x_0$. Thus $u(x)$ must be bounded from below, which completes the proof.

Contained in the above proof is the fact that a solution of (R) which is defined for all x is bounded by $M^{1/2}$. (Compare E. Hopf [1](2).)

Let us now consider the Jacobi differential equation (J). By the well known existence theorems, a solution of (J) is uniquely determined by two initial conditions, and is defined for all x. Recall, also, that if $y(x)$ is a solution of (J), $u(x)=y'(x)/y(x)$ is a solution of (R) in any interval where $y(x)\neq 0$.

THEOREM 2.1. *If there is a solution $w(x)$ of* (J) *for which $w(x)>0$ for $x\geqq 0$, then* $\lim_{x\to\infty} y(x)=\infty$ *for any solution $y(x)$ of* (J) *such that* $y(0)=0$, $y'(0)>0$.

Proof. The existence of $w(x)$ and the initial condition $y'(0)>0$ imply, by the Sturmian separation theorem, that $y(x)>0$ for $x>0$. Choose $x_0>0$. Set $b=y(x_0)/w(x_0)$ and define $u(x)=bw(x)$. $u(x)$ is a solution of (J) which is linearly independent of $y(x)$. For $x\geqq 0$,

$$\frac{d}{dx}\left[\frac{y(x)}{u(x)}\right] = \frac{u(x)y'(x) - u'(x)y(x)}{u^2(x)} > 0,$$

because the Wronskian

$$u(x)y'(x) - y(x)u'(x) = bw(0)y'(0) > 0.$$

Therefore $y(x)/u(x)$ is an increasing function for $x\geqq 0$. Suppose that the conclusion of the theorem were false. Then there would exist a sequence $0<x_0<x_1<\cdots$ with no finite limit point for which $\lim_{n\to\infty} y(x_n)=c<\infty$. Choose numbers $\{a_n\}$ such that $u(x_n)=a_n y(x_n)$, $n=0, 1, 2, \cdots$. Because $u(x)/y(x)$ is decreasing, $\{a_n\}$ is a monotone decreasing sequence of positive numbers. Set $a=\lim_{n\to\infty} a_n$. Then $a\geqq 0$. Form a solution of (J) by

$$z(x) = u(x) - ay(x).$$

By its definition, $z(x)$ is linearly independent of $y(x)$ and $\lim_{n\to\infty} z(x_n)=0$.

$$z'(x)y(x) - y'(x)z(x) = d, \tag{4}$$

where d is a nonzero (in fact, negative) constant. Since neither $z(x)$ nor $y(x)$ vanishes for $x>0$, we may apply Lemma 2.1 to the functions $z'(x)/z(x)$ and $y'(x)/y(x)$ and conclude that they are bounded for $x\geqq x_0$. By assumption, the numbers $\{y(x_n)\}$ are bounded; it follows that $\{y'(x_n)\}$ is a bounded sequence. But dividing (4) by $z(x)$ we get

$$\frac{z'(x)}{z(x)}y(x) - y'(x) = \frac{d}{z(x)}\cdot$$

As x takes on the values x_n the left side of this equation remains bounded,

(2) References will be found in the bibliography at the end of the paper.

while the right side approaches $-\infty$. This contradiction proves the theorem.

A point x_1 is said to be *conjugate* to x_0 if there is a nontrivial solution $y(x)$ of (J) with $y(x_1)=y(x_0)=0$. If there is no solution with two zeros, we shall say that (J) has no conjugate points.

COROLLARY 2.1. *If* (J) *has no conjugate points and* $y(x)$ *is a solution with* $y(x_0)=0, y'(x_0)>0$ $(y'(x_0)<0)$, *then we have* $\lim_{x\to\infty} y(x)=\infty$, $\lim_{x\to-\infty} y(x)=-\infty$ $(\lim_{x\to\infty} y(x)=-\infty, \lim_{x\to-\infty} y(x)=\infty)$.

This corollary follows immediately from Theorem 2.1 if we recall the well known fact that, when (J) has no conjugate points, there exists a solution which never vanishes. More generally, it is easily seen that the conclusion of the corollary holds whenever x_0 is interior to an interval of points none of which have conjugates on the whole axis.

A point x_1 is said to be a *focal point* of x_0 if there is a solution $y(x)$ of (J) with $y(x_0)=1$, $y'(x_0)=0$, and $y(x_1)=0$. If, for each x_0, there is no such point x_1, then we shall say that (J) has no focal points. Clearly, the property of having no focal points is more restrictive than that of having no conjugate points, as the equation (J) with $K(x)=\cos x/(a+\cos x)$, $a>1$, shows. Since there is always a solution satisfying the initial conditions $y(\bar{x})=1$, $y'(\bar{x})=0$, we can state

COROLLARY 2.2. *If* 0 *has no focal point* $x_1>0$, *then the conclusion of Theorem* 2.1 *holds. If* x_0 *has no focal points, then the conclusion of Corollary* 2.1 *holds.*

The literature dealing with equations (J) and (R) is vast. Their mutual interdependence is investigated in Wintner [1], where further references may be found.

3. **Manifolds with poles.** Consider a two-dimensional open (i.e., noncompact) Riemannian manifold M of class C^3 which is simply connected. Since such a manifold is homeomorphic to the (x, y)-plane, we may consider its metric defined by a positive definite quadratic form

$$ds^2 = g_{11}dx^2 + 2g_{12}dxdy + g_{22}dy^2, \tag{1}$$

where the functions $g_{ij}(x, y)$ are of class C^3 for all x, y.

Functions $x(t)$, $y(t)$, $t_0 \leqq t \leqq t_1$, define an arc of class C^1 if they are of class C^1 and

$$(\dot{x})^2 + (\dot{y})^2 \neq 0, \qquad t_0 \leqq t \leqq t_1.$$

A continuous image of the unit interval is an arc of class D^1 if it consists of finitely many segments of class C^1. The length, $L(m)$, of an arc m: $x(t)$, $y(t)$, $t_0 \leqq t \leqq t_1$, of class D^1 is defined by

$$L(m) = \int_{t_0}^{t_1} \{g_{11}(\dot{x})^2 + g_{12}\dot{x}\dot{y} + g_{22}(\dot{y})^2\}^{1/2}dt. \tag{2}$$

Curves whose coordinates satisfy the Euler differential equations for the first variation of the integral (2) are called geodesics, and they are of class C^2 when arc-length is introduced as parameter.

A point Q: $x(s_1)$, $y(s_1)$ is said to be conjugate to the point P: $x(s_0)$, $y(s_0)$ on the geodesic g: $x(s)$, $y(s)$ if s_1 is conjugate to s_0 in the Jacobi differential equation

$$y''(s) + K(s)y(s) = 0, \tag{J}$$

where $K(s)$ is the Gaussian curvature of M along g, expressed in terms of the arc-length, s. Focal points are defined similarly. A geodesic segment minimizes the integral $L(m)$ relative to neighboring curves of class D^1 with the same end points if it contains no two mutually conjugate points. (See, for example, Bliss [1].)

Let $D(P, Q)$ be the greatest lower bound of the lengths of all curves of class D^1 with end points P and Q. M becomes a metric space with metric $D(P, Q)$, and in the following, unless otherwise noted, any topological statement about M will be understood to refer to the topology induced by $D(P, Q)$ (which, however, agrees with the original topology of M; see Seifert and Threlfall [1]).

If $\mathfrak{S}$ is a set of points in M and P is a point of M, we define the distance from P to $\mathfrak{S}$ by

$$D(P, \mathfrak{S}) = \operatorname*{g.l.b.}_{Q \in \mathfrak{S}} D(P, Q).$$

If $\mathfrak{S}$ and $\mathfrak{R}$ are point sets of M, the *type distance* between these sets (see Hausdorff [1]) is defined by

$$D(\mathfrak{S}, \mathfrak{R}) = \max\left(\operatorname*{l.u.b.}_{P \in \mathfrak{S}} D(P, \mathfrak{R}), \operatorname*{l.u.b.}_{Q \in \mathfrak{R}} D(Q, \mathfrak{S})\right),$$

where $+\infty$ is an accepted value for $D(\mathfrak{S}, \mathfrak{R})$. Two sets are said to be of the same type if their type distance is finite.

A geodesic ray is a continuous image of a half-open interval, every closed segment of which is a geodesic. A complete geodesic, or for brevity, just a geodesic, is a similar image of an open interval such that both the geodesic rays determined by a point of the interval are infinitely long. A ray g is said to be continued to a ray g' if g is contained in g' and they have the same initial point. Continuations of a ray, then, are determined merely by their lengths, and in the following all rays will be understood to have been continued to infinite length. In order that this be possible, some additional assumption on the manifold is necessary.

We shall deal exclusively with *complete* manifolds, that is, with manifolds which are complete in the metric defined above. This concept was defined and investigated in Hopf and Rinow [1], where it was proved that on such a manifold every geodesic ray may be continued to infinite length. Moreover, between any two points, P, Q, of a complete, connected manifold there is a geodesic segment of length $D(P, Q)$. (See also Cartan [1], where a complete manifold is called normal.) Such a segment must contain the unique geodesic segment of shortest length connecting any two of its interior points, for otherwise a curve could be found between P and Q of length less than or equal to $D(P, Q)$ which contained corners, and such a path can be shortened.

P is called a pole of the space M if no geodesic ray with P as initial point contains a point conjugate to P. We shall call M a *manifold with pole* P if (i) M is a two-dimensional Riemannian manifold of class C^3, (ii) M is complete and simply-connected, (iii) the Gaussian curvature of M is bounded from below, and (iv) P is a pole of M. Examples of manifolds obeying all these conditions are: the universal covering surface of any compact, two-dimensional C^3 Riemannian manifold with nonpositive curvature; a paraboloid of revolution; the Euclidean plane.

In a manifold with pole P the geodesic rays with P as initial point cover the space simply, and polar coordinates (r, ϕ) may therefore be introduced in M. r is the distance from P along the geodesic making an angle of ϕ with a fixed direction at P. In these coordinates the form (1) becomes

$$ds^2 = dr^2 + G^2(r, \phi)d\phi^2, \tag{3}$$

where the function $G(r, \phi)$ is of class C^2 in $0 < r < \infty$, $0 \leqq \phi < 2\pi$, periodic in ϕ, and satisfies the equation

$$G_{rr}(r, \phi) + K(r, \phi)G(r, \phi) = 0, \tag{J'}$$

with boundary conditions $G(0, \phi) = 0$, $G_r(0, \phi) = 1$.

A geodesic ray with initial point P is the shortest path between any two of its points, and no two such rays intersect again. Moreover, a mapping by means of the polar coordinates onto the interior of the unit circle shows that, because M is complete, two different rays with P as initial point separate M into two simply-connected components.

Two geodesic rays are said to be of the same type if they are of the same type considered as point sets in M. This is equivalent to saying that, given an arbitrary point on either ray, its distance to the other ray lies below a uniform bound. In Euclidean and hyperbolic geometries two intersecting geodesic rays cannot be of the same type. The following theorem develops sufficient conditions for this phenomenon. Although (a) implies (b), it is convenient to include them both in the statement of the theorem.

THEOREM 3.1. *Let M be a manifold with pole P. Two geodesic rays with P as initial point cannot be of the same type if either of the following conditions is satisfied*:

(a) *P has no focal points on any ray with initial point P*;

(b) *P is in the interior of a set of points which are poles of M.*

Proof. Let g and q be the two geodesic rays in question, and assume that they are of the same type. For each integer $n>0$ let P_n be that point on g with $D(P, P_n)=n$. If R is a fixed number greater than the type distance of g and q, there must exist a point Q_n on q for which $D(P_n, Q_n)<R$. A geodesic segment exists with length $D(P_n, Q_n)$ and whose end points are P_n and Q_n; call this segment h_n. For $n>R$, h_n cannot be identical to the arc cut off on $g\cup q$ by P_n and Q_n, since the latter has length at least n. Thus for large enough n each arc h_n must lie entirely, except for its end points, in one of the two components into which $g\cup q$ separates M, for the property of being a shortest connection between P_n and Q_n would be destroyed should h_n have more than one point (but not all) in common with either g or q. At least one of these components must contain an infinite number of the curves $\{h_n\}$; choose such a component and set up polar coordinates with P as pole in such a way that g has the equation $\phi=0$, q has the equation $\phi=\phi_0$, and the points in the selected component have coordinates (r, ϕ), where $0<\phi<\phi_0$. Renumber the arcs h_n (and end points P_n, Q_n) which lie in this component so that, for them, n takes on the values $1, 2, 3, \cdots$.

Since h_n is a shortest arc from P_n to Q_n, it cannot intersect a ray from P in more than one point. For every ray from P is a shortest arc between any two of its points, so that its length between the first and last points of intersection with h_n is less than or equal to the length of the segment of h_n cut off by these points. Consequently the arc from P_n to the first such point of intersection, along the geodesic ray to the last such point, and along h_n to Q_n, has corners and may be shortened to a curve of length less than $L(h_n)$.

Because h_n cannot intersect a ray $\phi=\phi_1$, $0<\phi_1<\phi_0$, in more than one point, the equation for h_n in polar coordinates may be written $r_n=r_n(\phi)$, $0\leqq\phi\leqq\phi_0$. Moreover, for no fixed ϕ_1 can the sequence $\{r_n(\phi_1)\}$ have a finite limit point, since the distance from the points S_n: $(r_n(\phi_1), \phi_1)$ to P_n is bounded by R and $D(P, P_n)\leqq D(P, S_n)+D(S_n, P_n)\leqq r_n(\phi_1)+R$.

Let us compute the length of h_n. By (3)

$$L(h_n) = \int_0^{\phi_0} \left\{\left(\frac{dr_n}{d\phi}\right)^2 + G^2(r_n(\phi), \phi)\right\}^{1/2} d\phi \geqq \int_0^{\phi_0} G(r_n(\phi), \phi)d\phi.$$

By assumption, $L(h_n)\leqq R$ for all n; therefore we have

$$\liminf_{n\to\infty} \int_0^{\phi_0} G(r_n(\phi), \phi)d\phi \leqq \liminf_{n\to\infty} L(h_n) \leqq R.$$

The functions $G(r_n(\phi), \phi)$ are non-negative and integrable, so from Fatou's Lemma we conclude that $\lim\inf_{n\to\infty} G(r_n(\phi), \phi)$ is integrable over $(0, \phi_0)$, and is accordingly finite for almost all ϕ in the interval.

Up to this point we have not used either condition of the theorem. If P has no focal points, Corollary 2.2 may be applied to the function $G(r, \phi)$. Since $\lim_{n\to\infty} r_n(\phi) = \infty$ for any ϕ, by the conclusion of that corollary we see that lim $\inf_{n\to\infty} G(r_n(\phi), \phi) = \infty$ for any fixed ϕ. If condition (b) holds, the same result may be inferred by using the remark after Corollary 2.1. In either case, this contradicts the initial assumption which, as we have shown, implies that this limit inferior is finite almost everywhere. This completes the proof of the theorem.

As examples of manifolds with poles which satisfy conditions (a) or (b) we may call attention to the following: (a) No geodesic on the universal covering surface of a manifold with nonpositive curvature has focal points; for example, the plane considered as the covering surface of an hyperboloid (of revolution) of one sheet. (b) The obvious pole on one component of a two-sheeted hyperboloid of revolution is interior to a region of poles (von Mangoldt [1]).

I. NONCONJUGACY HYPOTHESIS. There is no pair of mutually conjugate points on any geodesic of M.

We denote by M(I) a manifold with pole which satisfies the nonconjugacy hypothesis. Every point of an M(I) is a pole, and every geodesic segment is the unique shortest path between its end-points. A direct application of Theorem 3.1 yields

COROLLARY 3.1. *On an M(I) no two geodesic rays with the same initial point can be of the same type.*

In Hedlund [1, Theorem 1.2] this result was obtained under the assumption that the geodesics of the manifold were uniformly unstable. Morse and Hedlund (Theorems 5.5 and 12.1 of Morse and Hedlund [1]—this paper will be referred to as M+H in the following) obtain the same result without that hypothesis, but with the restriction that the space was actually the universal covering surface of a particular class of manifolds (of the topological type of a torus or Klein bottle for Theorem 5.5, Poisson stable and of hyperbolic type for Theorem 12.1).

Busemann ([1], hereafter referred to as B) has studied general metric spaces in which geodesics, abstractly defined, exist and are unique. He calls such spaces S.L. (straight line) spaces; two-dimensional open (that is, containing no closed geodesics) S.L. spaces are homeomorphic to the plane, and are called S.L. planes. If an S.L. plane is a C^3 Riemannian manifold with Gaussian curvature uniformly bounded below it is an M(I) in our sense. Thus Corollary 3.1 is an extension of Theorem 6 [B, p. 103]. That the Riemannian character of the metric is essential for this stronger result may be seen by examining certain Hilbert geometries ([B, example 2, p. 108]—the author is indebted to Professor Busemann for pointing out that there are even *differentiable* Hilbert metrics for which the divergence property does not hold). For manifolds with nonpositive curvature the result is trivial (Hadamard [1], Cartan [1], and, for S.L. spaces, Busemann [2]). An M(I) admits regions of positive curvature, although how much can be tolerated without destroying the nonconjugacy property is not known. For some results on this aspect, see E. Hopf [2, p. 592]. However, his instability restriction is stronger even than Hedlund's (Hedlund [1]).

Let M be a space with pole P. The complement, $U(r)$, in M of every disk $S(P, r)$, $r>0$ ($S(P, r)$ is the set of points Q with $D(P, Q)<r$), is a surface homeomorphic to D', the closed unit disk of the Euclidean plane minus its center, 0. Every simple closed curve in $U(r)$ either is contractible to a point in $U(r)$ or has an image in D' which separates 0 from the unit circle. Following Cohn-Vossen [1] we call the latter type of curve a *girdle* of $U(r)$ if it is rectifiable. Let $g(r)$ be the greatest lower bound of all girdles of $U(r)$. Set $g(P)=\lim\inf_{r\to\infty} g(r)$.

Examining the proof of Theorem 3.1 we notice that the full force of the assumption that the rays were of the same type was not used, but only that $\lim\inf_{n\to\infty} L(h_n)$ was finite. A weaker formulation of this remark may be stated as

COROLLARY 3.2. *If M is a manifold with pole P and $g(P)$ is finite, then P is not an interior point of the set of poles of M, and P must have focal points.*

The paraboloid of revolution shows that the converse is not necessarily true. On the other hand, the surface of revolution of $f(x)=1/(x+1)+e^x \sin^2 x$, $x\geqq 0$, suitably smoothed to have a pole at $x=-1$, $y=0$, has $g(P)=0$, although the meridians are not of the same type.

4. **Nonfocality.** For a fixed point P we denote the set of points Q with $D(P, Q)<r$ by $S(P, r)$. Cl (A) will stand for the closure of the set A.

Let M be an M(I) and let g be a geodesic on M. Then any sequence of points of g which has no limit point on g has no limit point on M and (Hopf and Rinow [1]) M is finitely compact (that is, closed bounded sets are compact). It follows that, in an M(I), $g\cap \mathrm{Cl}\ (S(P, r))$ is compact (or empty) for any geodesic g, point P, and positive number r. Therefore, there always exists at least one point Q on g for which $D(P, Q)=D(P, g)$.

If g and h are two nonintersecting geodesics on an M(I) the complement of $g\cup h$ has three components. One of these components, call it A, has $g\cup h$ as boundary. We shall say that a point is between g and h if it is in A.

LEMMA 4.1. *Let g and h be two geodesics of the same type on an M(I). If every point of a geodesic ray r is between g and h, r is of the type of a ray contained in g.*

Proof. By Corollary 3.1, g and h cannot intersect, so it makes sense to speak of points between g and h. Let P_0 be a point on g. Then there is a point Q_0 on h such that $D(P_0, Q_0)=D(P_0, h)$. Denote by s_0 the geodesic segment with end points P_0 and Q_0; s_0 must separate the set of points between g and h into two components, A_1, A_2. Since r is a geodesic ray, it can intersect s_0 at most once, so all but a compact subset of r lies entirely in either A_1 or A_2;

suppose it is A_2. Choose P_n on $g \cap \mathrm{Cl}\,(A_2)$ with $D(P, P_n) = n$, and let Q_n on h be such that $D(P_n, Q_n) = D(P_n, h)$. All of the arcs s_n connecting P_n and Q_n, except for end points and s_0, lie in A_2. Moreover, $L(s_n) \leqq R$, where $R = D(g, h) < \infty$, by hypothesis. s_{n-1}, s_n, and the parts of g and h between P_{n-1} and P_n, Q_{n-1} and Q_n, respectively, form a simple closed curve which is the boundary of a closed set B_n in $\mathrm{Cl}\,(A_2)$. Every point of r, except, perhaps, for a compact segment, is in some B_n, and every B_n for large enough n contains points of r. It is therefore sufficient to prove that diam (B_n) is uniformly bounded (diam $(A) = \max_{P,Q \in A} D(P, Q)$), since in that case the subray of g with initial point P_0 which contains P_n will be of the same type as r. Because of the way it was constructed, B_n is geodesically convex (the geodesic arc joining any two points of B_n lies in B_n). Hence $\max_{P,Q \in B_n} D(P, Q)$ is assumed for P, Q on the boundary of B_n, so it is easy to see that diam $(B_n) \leqq 2R + 1$. This completes the proof.

LEMMA 4.2. *Let $y(x)$ be a solution of*

$$\text{(J)} \qquad y''(x) + K(x)y(x) = 0, \qquad K(x) \geqq -M,$$

which never vanishes. If lim inf $|y(x)| < \infty$ *for $x \to \infty$ and $x \to -\infty$, then every solution of* (J) *which never vanishes is of the form $cy(x)$, where c is a constant.*

Proof. The hypothesis that $y(x)$ never vanishes implies that (J) has no conjugate points or, what is equivalent, that no two integral curves have more than one point in common. We may assume that $y(x) > 0$ for all x. Suppose $w(x)$ is a solution of (J) which never vanishes. Then

$$u(x) = y(x) - \frac{y(0)}{w(0)}\, w(x)$$

is a solution of (J) for which $u(0) = 0$ and $y(x) > u(x) \geqq 0$ for either $x > 0$ or $x < 0$; assume the former case. Then $\liminf_{x \to \infty} u(x) < \infty$, so by Corollary 2.1, $u(x) = 0$ for all x, completing the proof of the lemma.

II. NONFOCALITY HYPOTHESIS. No geodesic of M contains two points such that one is a focal point of the other.

We designate an M(I) which satisfies the nonfocality hypothesis by M(II). Every simply-connected complete surface with bounded nonpositive curvature is an M(II). Unless otherwise stated, in the remainder of this section we deal with an M(II). That II is a powerful assumption is shown in part by

LEMMA 4.3. *If P is a point not contained in the geodesic g, there is precisely one geodesic h containing P with the following properties:*

(i) *h is perpendicular to g at their point of intersection Q;*

(ii) *$D(P, Q) < D(P, R)$ for any point R on g, $R \neq Q$.*

Proof. Let $Q(s)$ denote a point on g, where s is arc-length on g. At any point $Q(s_0)$ where the function $J(s)=D(P, Q(s))$ assumes a stationary value, the geodesic connecting P to $Q(s_0)$ is perpendicular to g, by the well known formula (see, for instance, Seifert and Threlfall [1, p. 100])

$$\frac{dJ}{ds} = \cos \alpha,$$

where α is the angle this connecting geodesic makes with g. Since there exists a point Q on g for which $D(P, Q)=D(P, g)$, it is sufficient to prove that this point is unique. Suppose R were another such point. Then, as $Q(s)$ varies on the closed segment of g between Q and R, $J(s)$ is either constant or attains a maximum value at some interior point of this segment. In either case, there is a point Q' interior to this arc for which $\alpha=\pi/2$ and at which $J(s)$ does not have a strong minimum relative to neighboring geodesics from P to g. But this is contrary to the assumption concerning the nonexistence of focal points (Bliss [1, p. 151]). This completes the proof(3).

It is easy to see that the function $J(s)$ introduced above can take on but one extreme value, so either of the conditions (i) or (ii) is sufficient to characterize the geodesic with these properties. We call such a geodesic the perpendicular from P to g.

THEOREM 4.1. *If g and h are two geodesics of the same type on an M(II) and P is a point between g and h, then $K(P)=0$. ($K(P)$ is the Gaussian curvature at P.)*

Proof. Let k' be the unique perpendicular from P to g. Using P as origin from which to measure arc-length t on k', choose $\epsilon>0$ so that the points $P(t)$ on k', $-\epsilon\leqq t\leqq\epsilon$, are all between g and h, and call this segment of k', k. Set up geodesic normal coordinates with k' as base line. A point will have coordinates (t, u) if its distance to k' is u and the perpendicular from this point to k' intersects k' at $P(t)$. By Lemma 4.3 this coordinate system is one-to-one in the large. The line element becomes

$$ds^2 = du^2 + F^2(u, t)dt^2, \tag{1}$$

where $F(0, t)=1$, $F_u(0, t)=0$, and F satisfies the Jacobi equation

$$F_{uu}(u, t) + K(u, t)F(u, t) = 0. \tag{J''}$$

Consider a geodesic $h(t)$ which is orthogonal to the segment k at $P(t)$, $-\epsilon\leqq t\leqq\epsilon$. $h(t)$ cannot intersect g, for then there would exist two geodesics perpendicular to k' from the same point. On the other hand, $h(t)$ cannot intersect h twice, since they are geodesics. Therefore some subray of $h(t)$ lies entirely between g and h, and by Lemma 4.1 this ray must be of the

(3) This result can also be obtained without appeal to the calculus of variations by taking the second derivative of J with respect to s and using the fact that $d\alpha/d\phi=-\partial G^{1/2}/\partial r$.

same type as a subray of h. Consequently, by Corollary 3.1, $h(t)$ cannot intersect h at all, must lie entirely between g and h, and is a geodesic of the same type as h or g. Since the relation of being of the same type is transitive, we have, in particular, that $h(-\epsilon)$ and $h(\epsilon)$ are of the same type.

Just as in Theorem 3.1 we construct geodesic segments u_n connecting points $(n, -\epsilon)$ on $h(-\epsilon)$ with $h(\epsilon)$, whose equations in normal coordinates are $u_n = u_n(t)$, $-\epsilon \leqq t \leqq \epsilon$, $n = \pm 1, \pm 2, \pm 3, \cdots$. By the same argument we find that $\lim_{n\to\infty} u_n(t) = \infty$, $\lim_{n\to-\infty} u_n(t) = -\infty$, for each t. Since $h(-\epsilon)$ and $h(\epsilon)$ are of the same type, $\liminf L(u_n) < \infty$ for $n \to \pm\infty$, where the lengths $L(u_n)$ are now given by the integrals

$$L(u_n) = \int_{-\epsilon}^{\epsilon} \{(\dot{u}_n)^2 + F^2(u_n(t), t)\}^{1/2} dt.$$

Applying Fatou's lemma exactly as is done in Theorem 3.1, we conclude that $\liminf_{n\to\infty} F(u_n(t), t)$ and $\liminf_{n\to-\infty} F(u_n(t), t)$ are both integrable functions. Thus there exist two sets, M^+ and M^-, each of measure 2ϵ, such that

$$\liminf_{n\to\infty} F(u_n(t), t) < \infty, \qquad t \in M^+, \tag{2}$$

and

$$\liminf_{n\to-\infty} F(u_n(t), t) < \infty, \qquad t \in M^-. \tag{2'}$$

$M = M^+ \cap M^-$ is again a set of measure 2ϵ, and for each $t \in M$ the inequalities (2), (2′) hold simultaneously.

Let $t_0 \in M$. Then $F(u, t_0)$ is a solution of (J″) satisfying the hypotheses of Lemma 4.2. Hence any two solutions of

$$F_{uu}(u, t_0) + K(u, t_0)F(u, t_0) = 0 \tag{3}$$

which are never zero are linearly dependent. But the hypothesis of nonfocality is precisely the assumption that at each point u_0 there is a never zero solution of (3), say $H(u)$, with $H'(u_0) = 0$. Hence $F_u(u, t_0) = 0$ for all u, so that, because of the initial conditions, $F(u, t_0) = 1$ for all u. Then (3) reduces to

$$K(u, t_0) = 0, \qquad -\infty < u < \infty. \tag{4}$$

Since t_0 was an arbitrary point of a set of measure 2ϵ, and K is continuous, (4) holds for all t in $[-\epsilon, \epsilon]$. In particular, (4) is true for $u = 0$, $t_0 = 0$; that is, $K(P) = 0$, which was to be proved.

COROLLARY 4.1. *If a (real) analytic* M(II) *contains two geodesics of the same type, it is flat; that is,* $K = 0$ *and the surface is the Euclidean plane.*

Proof. K may be developed in a power series about some point between the two geodesics, and by Theorem 4.1 all the coefficients of this series are zero. The proof is completed by continuing K analytically to every point of the surface.

COROLLARY 4.2. *If every point P of an M(II) is between some pair of geodesics of the same type (depending on P), the manifold is flat.*

Busemann [3, p. 283] has obtained a similar result for spaces which are not necessarily differentiable. However, in the Riemannian case the surfaces he considers must have nonpositive curvature(4). Morse and Hedlund [1, Theorem 5.7] have shown that the hypotheses of Corollary 4.2 are satisfied on the covering surface of a doubly periodic Riemannian manifold of Euclidean type without focal points (for definitions, see the paper cited). Thus their Theorem 8.2, that every such surface, that is, every torus or Klein bottle without focal points, is flat, follows from Corollary 4.2. (See, however, E. Hopf [1], where the same result is obtained for any surface of these topological types whose universal covering surface is merely an M(I).)

Let C be a surface of the topological type of a cylinder which is a complete Riemannian manifold. It is easy to see (compare, for example, Cohn-Vossen [1, p. 114]) that on such a surface there always exists an infinite geodesic g without conjugate points. Let $P(t)$ designate a point on g, where t is arc-length, and let $c(t)$ be a simple closed curve through $P(t)$ which is not contractible to a point on C. We may assume that $c(t)$ intersects g only in $P(t)$. For $t \neq 0$ let $U(t)$ be the component of $C-c(t)$ which does not contain $P(0)$. $U(t)$ is a region of the kind considered in Corollary 3.2, and we may accordingly define girdles of $U(t)$ and their greatest lower bound, $g(t)$. Call C a *tube* if $\liminf g(t) < \infty$ for $t \to \infty$ and $t \to -\infty$. The following is an analogue for cylinders of the result of Morse and Hedlund for tori.

COROLLARY 4.3. *If C is a tube whose covering surface is an M(II), C is flat.*

Proof. Every point of the covering surface of C lies between two copies of the geodesic g. Such lines may not be of the same type under our hypotheses, but an examination of the proof of Theorem 4.1 shows that $\liminf g(t) < \infty$ for $t \to \pm\infty$ is sufficient to conclude that the curvature is zero.

5. **Manifolds of hyperbolic type.** An M(I) as defined in §3 is homeomorphic to the plane, and hence is also homeomorphic to the interior D of the unit circle C. Each such homeomorphism will be called a *representation* if it is of class C^3 when considered as a map between the differentiable manifolds M(I) and D. A representation is conformal if the Riemannian metric

(4) Professor Busemann has pointed out that Theorem 4.1 cannot be extended to Finsler spaces. In particular, Hilbert geometries (see the remark after Corollary 3.1) may be constructed satisfying the hypotheses of Corollary 4.2. However, these are not Minkowskian, the Finsler space analogue of flatness.

it induces in D has the form

$$(1) \qquad ds^2 = \frac{4f^2(u, v)(du^2 + dv^2)}{(1 - u^2 - v^2)^2}, \qquad u^2 + v^2 < 1.$$

Every conformal representation of an $M(\mathrm{I})$ induces a geometry in D in which the angles are the same as the Euclidean angles. The case in which $f(u, v) = 1$ for all u, v yields the familiar Poincaré metric for hyperbolic geometry. The geodesics in this metric are arcs of (Euclidean) circles orthogonal to C. We shall call them hyperbolic lines or H-lines, and define similarly H-rays and hyperbolic distance $H(P, Q)$ between two points in D. Euclidean distance between P, Q in $C \cup D$ will be denoted by $E(P, Q)$.

Fixing our attention on a single conformal representation Φ of $M(\mathrm{I})$ (if there is one), we say that an H-line h is of the same type as a geodesic g of $M(\mathrm{I})$ if h and $\Phi(g)$ are of the same type in D with respect to the metric (1). Type distance is defined similarly.

An $M(\mathrm{I})$ will be called a *manifold of hyperbolic type*, denoted by $M(\mathrm{I}, H)$, if there exists a conformal representation Φ of $M(\mathrm{I})$ onto D for which the following conditions hold: H-T *Hyperbolic type conditions*: (i) There exists a finite constant R such that, if g is the geodesic segment joining P and Q, and h is the H-segment between $\Phi(P)$ and $\Phi(Q)$, the type distance between g and h is less than R. (ii) There is a constant $c > 0$ such that $cH(\Phi(P), \Phi(Q)) \leqq D(P, Q)$ for all P and Q.

In Morse [1] and M+H (p. 379) it is proved that a manifold defined in D by the form (1), where $f(u, v)$ is of class C^3 and lies between fixed positive bounds, satisfies conditions H-T. In addition, if its Gaussian curvature is bounded below and there are no conjugate points, such a manifold will be an $M(\mathrm{I}, H)$. The following facts are consequences of conditions H-T (Morse [1, p. 44]):

(A) Every geodesic is of the type of some H-line.

(B) Every geodesic ray is of the type of some H-ray with the same initial point, and for every H-ray there is a geodesic ray of its type with the same initial point. ("Same initial point" refers to the image in D.)

We now fix our attention on one $M(\mathrm{I}, H)$ with a specific conformal representation Φ. Its geometry is then equivalent to that of D with the metric (1), so we shall from now on talk only about the point set D under the metrics $D(P, Q)$, induced by (1), $H(P, Q)$, and $E(P, Q)$. (The latter will also apply to $D \cup C$.) We call geodesics of D, or just geodesics, the geodesics defined by (1).

An easy consequence of H-T (ii) and the behavior of the hyperbolic metric is

Lemma 5.1. *If $\{P_n\}$, $\{Q_n\}$ are points in D such that $D(P_n, Q_n)$ are uniformly bounded and $E(P_n, P) \to 0$ for some P on C, then $E(Q_n, P) \to 0$.*

Since an H-line is completely determined by its "points at infinity" (the intersections of the Euclidean circle defining it with C), property A associates two such points with each geodesic of D. Lemma 5.1 implies that these points are unique (and hence the H-line of the same type is unique), so we may speak of the points at infinity of geodesics of D. Similarly, a geodesic ray has a unique point at infinity, and the H-ray which (B) says exists is unique. Therefore two geodesic rays with the same initial point and the same point at infinity are of the same type as a single H-ray and, consequently, are themselves of the same type. An application of Corollary 3.1 proves

LEMMA 5.2. *Given $P \in D$ and $Q \in C$, there is at most one geodesic ray with initial point P and point at infinity Q.*

The corresponding statement may not be true for complete geodesics; that is, there may exist many geodesics with the same pair of points at infinity. We remark further that, since the points at infinity of geodesic rays coincide with the points at infinity of the H-rays of the same type (Lemma 5.1), the existence of the ray from P to Q is assured by (B) and the fact that there is such an H-ray.

A pair (P, ϕ), where P is a point of D and ϕ is an angle between 0 and 2π, will be called an *element*. To each element (P, ϕ) there corresponds precisely one geodesic ray with P as initial point and ϕ as tangent direction at P, for the solutions of the differential equations defining geodesics are uniquely determined by these boundary conditions. Let E denote the set of elements of D, and topologize E as a product space, the topology for the angle being that of the real numbers modulo 2π. $P(t, \phi)$ will denote the point on the geodesic rays with initial element (P, ϕ) and distance t from P. Then, since the solutions of the differential equations depend continuously on the initial conditions, $P_n(t, \phi_n) \to P(t, \phi)$ uniformly for $0 \leqq t \leqq t_0$ if $P_n \to P$ and $\phi_n \to \phi$ (t_0 is any fixed positive number). Under these circumstances the geodesic rays with initial elements (P_n, ϕ_n) will be said to converge to the ray with initial element (P, ϕ).

LEMMA 5.3. *Let $E(P_n, P) \to 0$, $P_n \in D$, $P \in C$. Then the geodesic rays r_n from P_1 through P_n, $n = 2, 3, 4, \cdots$, converge to a ray with P as point at infinity.*

Proof. Let (P_1, ϕ) be the initial element of the ray through P_1 with point at infinity P, and suppose that there is a subsequence, which we renumber (P_1, ϕ_n), $n = 2, 3, 4, \cdots$, of initial elements of r_n for which $\phi_n \to \theta \neq \phi$. There is an $\epsilon > 0$ such that $\theta \notin [\phi - \epsilon, \phi + \epsilon]$. Let g_1 and g_2 be the geodesic rays with initial elements $(P_1, \phi - \epsilon)$, $(P_1, \phi + \epsilon)$, and Q_1, Q_2 their points at infinity. P, Q_1, and Q_2 are distinct, by Lemma 5.2, so there is a Euclidean neighborhood of P whose intersection with D lies entirely in the region bounded by

g_1, g_2, and the arc of C between Q_1 and Q_2 which contains P. But there are an infinite number of points P_n outside this neighborhood, and this contradiction of the hypothesis proves the lemma.

Let G be a group of transformations of the unit disk onto itself which leave the metric (1) invariant. If T is in G, $T(P)$ is called the point *congruent* to P under T, and the collection $\{T(P) \mid T \in G\}$ is the set of all points congruent to P under G. G is called *properly discontinuous* in a set A if, for any point P in A, P is not a limit point of points congruent to P under G. If G is properly discontinuous in D but ceases to be so in every subset of C, G will be called a group of the *first kind*. Because the metric (1) is left invariant, T preserves angle measurement and must be a conformal or inversely conformal transformation. Thus T may also be regarded as an isometry in the space of elements E. Moreover, any such transformation is an isometry of the hyperbolic geometry.

If an $M(\mathrm{I}, H)$, considered as the point set D with metric (1), has a group of isometries G of the first kind, it will be denoted by $M(G)$. Since $T \in G$ preserves distances and arc-length the images of geodesic segments under T are again geodesic segments. A geodesic ray r will be called *transitive* if the set of elements on r and all its congruent copies under G is dense in E.

We now know enough about manifolds of hyperbolic type to state the following theorem. Its proof is the same as that of Theorem 13.1 (M+H), and will therefore be omitted (compare also Theorem 3.1 of Hedlund [1] and Hauptsatz of Salenius [1]).

THEOREM 5.1. *On any $M(G)$ there exist transitive rays.*

Since G is properly discontinuous we may identify points of D which are congruent under G and obtain a manifold with fundamental group G and universal covering space D (Seifert and Threlfall [2, Chap. 8]). Because (1) is invariant under G, we may define a Riemannian metric in this space and call the resulting surface M. The set of elements of M, topologized in the usual fashion, is called the phase space, Ω, of M; it is precisely the tangent circle bundle.

Let e': (P, ϕ) be a point of E, e its image in Ω. e' determines a unique geodesic g of D; let s be its arc-length measured from P. We set e'_s equal to the tangent element of g whose base point is a (directed) distance s from P, and let $T_s(e)$ be the corresponding point in Ω. The group of transformations $\{T_s \mid -\infty < s < \infty\}$ defines the *geodesic flow* in Ω (see Hedlund [2]). If there exists a transitive ray in D there is an element e of Ω such that the set $\{T_s(e) \mid 0 < s < \infty\}$ (the positive semi-orbit of e) is dense in Ω, and conversely. In this case the flow is said to be *topologically transitive*. We restate Theorem 5.1 to conform to this order of ideas.

COROLLARY 5.1. *Let M be a Riemannian manifold whose universal covering* *space is an $M(G)$, where G is its fundamental group. Then the geodesic flow in Ω (the phase space of M) is topologically transitive.*

Theorem 5.1 includes Theorem 13.1 of M+H in the sense that, if the manifold they consider has Gaussian curvature bounded below, it is an $M(G)$ in our terminology. Since the curvature of the metric (1) as given by the Theorema Egregium is

$$K(u, v) = -\frac{1}{f^2} + \frac{1}{4f^4}(1 - u^2 - v^2)^2(f_u^2 + f_v^2 - ff_{uu} - ff_{vv})$$

we may state

COROLLARY 5.2. *Let the function f in (1) satisfy, for all u, v in D, the conditions*

(i) $$0 < a \leqq f(u, v) \leqq b, \qquad a, b \text{ constant};$$

(ii) $$\Delta f = f_{uu} + f_{vv} \leqq N, \qquad N \text{ constant}.$$

If M is a C^3 Riemannian manifold whose universal covering surface is D with metric (1) and fundamental group of the first kind, then the geodesic flow in the phase space of M is topologically transitive.

Thus we have replaced the requirement of Morse and Hedlund that the flow be Poisson stable (see M+H, p. 380) by condition (ii) of a purely analytical character. This proof, accordingly, makes no use of the Poincaré recurrence theorem, with its measure-theoretic considerations. In particular, if M is closed and of sufficiently high genus, it always has a conformal representation and conditions (i) and (ii) are automatically satisfied. We therefore obtain Corollary 13.1 of M+H:

COROLLARY 5.3. *If M is a closed orientable two-dimensional Riemannian manifold of class C^3 and of genus greater than one, and if no geodesic on M has conjugate points, then the geodesic flow in the phase space of M is topologically transitive. The same result is true if M satisfies all of the above conditions, except is nonorientable and of genus greater than two.*

BIBLIOGRAPHY

G. A. BLISS

1. *Lectures on the calculus of variations*, Chicago, 1946.

H. BUSEMANN

1. *Metric methods in Finsler spaces and in the foundations of geometry*, Princeton, 1942.
2. *Spaces with non-positive curvature*, Acta Math. vol. 80 (1948) pp. 259–310.
3. *Angular measure and integral curvature*, Canadian Journal of Mathematics vol. 1 (1949) pp. 279–296.

E. CARTAN

1. *Leçons sur la géométrie des espaces de Riemann*, Paris, 1946.

S. COHN-VOSSEN

1. *Kürzeste Wege und Totalkrümmung*, Compositio Math. vol. 2 (1935) pp. 69–133.

J. HADAMARD
1. *Les surfaces à courbures opposées*, J. Math. Pures Appl. (5) vol. 4 (1898) pp. 27–73.
F. HAUSDORFF
1. *Mengenlehre*, Leipzig, 1927.
G. A. HEDLUND
1. *Two-dimensional manifolds and transitivity*, Ann. of Math. (2) vol. 37 (1936) pp. 534–542.
2. *The dynamics of geodesic flows*, Bull. Amer. Math. Soc. vol. 45 (1939) pp. 241–260.
E. HOPF
1. *Closed surfaces without conjugate points*, Proc. Nat. Acad. Sci. U.S.A. vol. 34 (1948) pp. 47–51.
2. *Statistik der Lösungen geodätischer Probleme vom unstabilen Typus*, II, Math. Ann. vol. 117 (1940) pp. 590–608.
H. HOPF and W. RINOW
1. *Ueber den Begriff der vollständigen differentialgeometrischen Fläche*, Comment. Math. Helv. vol. 3 (1931) pp. 209–225.
H. VON MANGOLDT
1. *Geodätische Linien auf positiv gekrümmten Flächen*, J. Reine Angew. Math. vol. 91 (1881) pp. 23–53.
M. MORSE
1. *A fundamental class of geodesics on any closed surface of genus greater than one*, Trans. Amer. Math. Soc. vol. 26 (1924) pp. 25–60.
M. MORSE and G. A. HEDLUND
1. *Manifolds without conjugate points*, Trans. Amer. Math. Soc. vol. 51 (1942) pp. 362–386.
T. SALENIUS
1. *Ueber geodätische Linien in gewissen dreidimensionalen Mannigfaltigkeiten*, Dissertation, University of Helsinki, 1948.
H. SEIFERT and W. THRELFALL
1. *Variationsrechnung im Grossen*, Berlin, 1938.
2. *Lehrbuch der Topologie*, Leipzig, 1934.
A. WINTNER
1. *On the non-existence of conjugate points*, Amer. J. Math. vol. 73 (1951) pp. 368–380.

YALE UNIVERSITY,
NEW HAVEN, CONN.