Math. Proc. Camb. Phil. Soc. (1979), **86**, 365
Printed in Great Britain

Cut points, conjugate points and Lorentzian comparison theorems

BY JOHN K. BEEM AND PAUL E. EHRLICH
University of Missouri, Columbia, Missouri

(*Received* 25 *September* 1978)

1. *Introduction.* The purpose of this paper is to study the global geometry of a space–time (M, g), which is related to the Lorentzian distance function induced on the manifold M by the Lorentzian structure. We will use the signature convention $(-, +, \ldots, +)$ for g and assume that (M, g) is time orientated. The first part of this paper deals with cut points and maximal geodesics, both of which were defined in (1) using the Lorentzian distance function in analogy to the standard concepts in Riemannian geometry. In (1), sections 2 and 3, some elementary properties of maximal geodesics were established. In particular, the principle that, for strongly causal space–times, a limit curve of a sequence of future-directed nonspacelike 'almost maximal' curves is a maximal geodesic was used to prove nonspacelike incompleteness ((1), theorem 6·3). Also null cut points were used to obtain results on null incompleteness ((1), section 5). In (2) we studied deeper properties of maximal geodesics and cut points using the technical tools developed in (1), sections 2 and 3. The first part of the present paper continues these investigations.

In the second part of this paper we first sketch the index theory of timelike geodesics in arbitrary space–times. We may then obtain related Lorentzian comparison theorems. There are striking differences from global Riemannian geometry in the application of these results, however. One example is furnished by the differences between the Lorentzian timelike diameter and the Riemannian diameter of a manifold. A complete Riemannian manifold of finite diameter is necessarily compact. Thus there is a minimal geodesic segment whose arc length realizes the diameter. There are also geodesics of arbitrary length, even though these geodesics are no longer minimal. In contrast, if a space–time has finite timelike diameter, then *all* timelike geodesics are incomplete and *all* have length less than or equal to the timelike diameter. In addition, no space–time of finite timelike diameter contains a geodesic *segment* whose arc length realizes the timelike diameter.

After recalling the definitions and basic properties of the Lorentzian distance function and the space–time cut locus in Section 2, we consider the continuity and differentiability properties of this function. In particular, we characterize strongly causal space–times in terms of local distance functions in Theorem 3·4. We then show in Section 4 that each point p of a globally hyperbolic space–time has a nonspacelike future (past) directed and future (past) inextendible maximal geodesic ray starting at p. Thus each point in a globally hyperbolic space–time has at least one future direction and one past direction which does not correspond to a cut point. If (M, g)

0305-0041/79/0000-6460 $03.50 © 1979 Cambridge Philosophical Society

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

is timelike geodesically complete as well as globally hyperbolic, then the causal future $J^+(p)$ less the future cut locus and future horismos of p is an open cell. This is the analogue of the well known result for compact Riemannian manifolds that the manifold is the union of an open cell and the cut locus ((13), pp. 100–101).

Even though in globally hyperbolic space–times we have constructed future and past directed rays at each point, this construction does not necessarily yield a nonspacelike line, i.e. an inextendible nonspacelike geodesic which is maximal between any two of its points. However, if a globally hyperbolic space–time is causally disconnected ((1), definition 6·1) then it must have a nonspacelike line ((1), proposition 6·4). The existence of this line is important because, if dim $M \geqslant 3$, the line must be an incomplete geodesic provided the space–time (M, g) satisfies the generic and strong energy conditions (1), (11), (12). Here we show in Section 5 that, for two-dimensional manifolds, all globally hyperbolic space–times are causally disconnected. Thus if (M, g) is a two-dimensional globally hyperbolic space–time, (M, g) admits a maximal inextendible nonspacelike geodesic. We also show that an arbitrary globally hyperbolic space–time possessing a null line is causally disconnected.

In Section 6 we consider the closure properties of the cut and conjugate loci. As in the Riemannian case, the cut point along a (nonspacelike) geodesic comes at or before the first conjugate point. We show that for globally hyperbolic space–times, the null cut locus, the nonspacelike cut locus and the nonspacelike conjugate locus are all closed subsets of M. The timelike cut locus need not be closed as may be seen from the Einstein static universe.

In Section 7 we sketch the index theory for timelike geodesics in an arbitrary space–time needed for Sections 8 and 9. Our treatment here parallels that given for Riemannian manifolds in (10), chapter 3. For completeness we also sketch G. Bölts' treatment ((4), unpublished, Bonn University Diplomarbeit) of the index form for null geodesics in an appendix.

In Section 8 we consider the cut locus and first conjugate locus in future 1-connected space–times. Here (M, g) is future 1-connected (8) if for each $q \in I^+(p)$, any two smooth future-directed timelike curves from p to q are homotopic through future-directed timelike curves with endpoints p and q. Suppose that (M, g) is globally hyperbolic and future 1-connected and that, for $p \in M$, the first future conjugate point of p along every timelike geodesic c with $c(0) = p$ is of order $\geqslant 2$. Then we show (theorem 8·3) that the future timelike cut locus of p and the locus of first future conjugate points of p coincide. Thus all geodesics from p maximize up to the first future conjugate point.

In the last section of this paper, we give the Lorentzian analogues of several important tools in global Riemannian geometry, namely, the Index Comparison Theorem ((10), p. 174), the Bonnet–Myers Diameter Theorem ((10), p. 212) and the Rauch Comparison Theorem ((10), p. 178).

2. *The space–time cut locus.* In this section we recall some notation from Beem and Ehrlich (2) needed in this paper.

Let (M, g) be an arbitrary connected space–time of dim $M \geqslant 2$. This means M is a

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

smooth manifold with a countable basis, a smooth Lorentz metric g of signature $(-, +, \ldots, +)$ and a time orientation. Let $\pi: TM \to M$ denote the tangent bundle of M. Recall that a nonzero tangent vector $v \in TM$ is said to be *timelike* (resp. *nonspacelike, null, spacelike*) according to whether $g(v,v) < 0$ (resp. $\leqslant 0, = 0, > 0$). We will use the standard notations $p \ll q$ if there is a future directed timelike curve from p to q and $p \leqslant q$ if there is a future directed nonspacelike curve from p to q. The chronological future of p is defined by $I^+(p) = \{q | p \ll q\}$ and the causal future of p is defined by $J^+(p) = \{q | p \leqslant q\}$. The sets $I^-(p)$ and $J^-(p)$ are defined dually.

The space–time is said to be *chronological* if $p \notin I^+(p)$ for all p. It is *strongly causal* if each point has arbitrarily small neighbourhoods such that no nonspacelike curve which leaves one of these neighbourhoods ever returns, see ((11), p. 192). A strongly causal space–time such that $J^+(p) \cap J^-(q)$ is compact for all $p, q \in M$ is called *globally hyperbolic*.

The *Lorentzian distance function* $d = d(g): M \times M \to \mathbb{R} \cup \{\infty\}$ of the space–time (M, g) is defined as follows. Given $p \in M$, set $d(p, q) = 0$ for $q \in M - J^+(p)$ and for $q \in J^+(p)$ let $d(p,q)$ be the supremum of lengths of all future directed nonspacelike curves from p to q. By construction this function is not symmetric in general. The Lorentzian distance function is always lower semi-continuous (11) where it is finite but it is not necessarily upper semi-continuous. For globally hyperbolic space–times, however, the distance function is always finite and continuous ((11), p. 215). This function has been used in establishing a number of singularity theorems (1), (11), (12), (15).

In (1) we defined a future directed curve γ from p to q with $p \leqslant q$, $p \neq q$, to be *maximal* if and only if its length $L(\gamma)$ is equal to $d(p,q)$. We then noted (Corollary 3·4) that any maximal future directed nonspacelike curve joining distinct points p and q is in fact a geodesic segment. A nonspacelike future directed curve $\gamma: (a, b) \to M$ defined on an open interval with $L(\gamma|[s,t]) = d(\gamma(s), \gamma(t))$ for all $a \leqslant s \leqslant t \leqslant b$ will also be called maximal. From the above result, each subsegment $\gamma|[s,t]$, $a < s \leqslant t < b$, is a geodesic. This implies γ itself is a geodesic and by maximality γ must be free of conjugate points. Also a maximal nonspacelike geodesic segment γ from p to q cannot contain a pair of conjugate points between p and q.

Let $T_{-1}M = \{v \in TM;\ g(v,v) = -1$ and v is future directed$\}$. Given $v \in T_{-1}M$, let c_v denote the unique timelike geodesic with $c_v'(0) = v$. Define a function $s: T_{-1}M \to [0, \infty]$ by $s(v) = \sup\{t \geqslant 0;\ d(c_v(0), c_v(t)) = t\}$. The *future timelike cut locus* $C_t^+(p)$ *of* p *in* M is then defined ((2), definition 4·12) to be the set of points $C_t^+(p) = \{\exp_p (s(v)\, v);\ v \in T_{-1}M|_p$ and $0 < s(v) < \infty\}$. The function $v \to s(v)$ thus measures the distance from p to the future cut point along c_v. Continuity properties of s were studied in section 4 of Beem and Ehrlich (2) where the continuity of $s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ was established for globally hyperbolic, timelike geodesically complete space-times.

Null cut points were also studied in Beem and Ehrlich (2) and in Beem and Ehrlich (1), section 5. Let $c: [0, a) \to M$ be a future directed null geodesic from $p = c(0)$. Set $t_0 = \sup\{t \in [0, a);\ d(p, c(t)) = 0\}$. If $0 < t_0 < a$, we call $c(t_0)$ the *future null cut point* of p on c. The *future null cut locus* $C_N^+(p)$ *of* p *in* M is then the set of all future null cut points of p. The *future nonspacelike cut locus* $C^+(p)$ is defined to be

$$C^+(p) = C_t^+(p) \cup C_N^+(p).$$

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

The past timelike cut locus $C_t^-(p)$, past null cut locus $C_N^-(p)$, and past nonspacelike cut locus $C^-(p)$ of p in M may all be defined dually.

Fix a point $p \in M$ and let $\exp_p\colon T_pM \to M$ be the exponential map. A tangent vector $v \in T_pM$ is called a *conjugate point* to p in T_pM if $(\exp_p)_*$ is singular at v. A point $q \in M$ is *conjugate to the point p along a geodesic c* if there is some conjugate point $v \in T_pM$ such that $\exp_p v = q$ and c is (up to reparametrization) the geodesic $\exp_p(tv)$. It is well known that $v \in T_pM$ is a conjugate point if and only if there is a non-trivial Jacobi field along $\exp_p(tv)$ that vanishes at p and $\exp_p v$.

3. *Local distance functions and strongly causal space-times.* We will use the same definition of convex normal neighbourhood as in Hawking and Ellis ((11), p. 34). Given any space–time (M, g)

Definition 3·1. A *local distance function* (D, U) *on* (M, g) is a convex normal neighbourhood U together with the distance function $D\colon U \times U \to \mathbb{R}$ induced on U by the space–time $(U, g|U)$.

More explicitly, if $p, q \in U$, then $D(p,q) = 0$ if there is no future directed timelike geodesic segment in U from p to q. Otherwise, $D(p, q)$ is the Lorentzian arc length of the unique future directed timelike geodesic segment in U from p to q.

We will let $I^+(p, U)$ (resp. $J^+(p, U)$) denote the chronological (resp. causal) future of p with respect to the space–time $(U, g|U)$.

LEMMA 3·2. *Let (M, g) be a space–time and let U be a convex normal neighbourhood of (M, g). Assume that $D\colon U \times U \to \mathbb{R}$ is the distance function for $(U, g|U)$. Then D is a continuous function on $U \times U$ and D^2 is differentiable on $U^+ = \{(p,q) \in U \times U;\ q \in I^+(p, U)\}$.*

Proof. Given $p, q \in U$ with $q \in J^+(p, U)$, let $c_{pq}\colon [0, 1] \to U$ denote the unique nonspacelike geodesic segment with $c_{pq}(0) = p$ and $c_{pq}(1) = q$. We then have

$$D(p,q) = (-g(c'_{pq}(0), c'_{pq}(0)))^{\frac{1}{2}} \quad \text{and} \quad (D(p,q))^2 = -g(c'_{pq}(0), c'_{pq}(0)).$$

From the differentiable dependence of geodesics on endpoints in convex neighbourhoods, it is immediate that D is continuous on $U \times U$ and D^2 is differentiable on U^+.

Remark 3·3. Minkowski space–time shows that D^2 fails to be differentiable across the null cones and thus fails to be smooth on all of $U \times U$.

It is not hard to see that the local distance function (D, U) uniquely determines the Lorentzian metric g on U. Consequently if $\{U_\alpha\}$ is a covering of M by convex normal neighbourhoods with associated local distance functions $\{(D_\alpha, U_\alpha)\}$, then $\{(D_\alpha, U_\alpha)\}$ uniquely determines g on M.

We now characterize strongly causal space–times in terms of local distance functions.

THEOREM 3·4. *A space–time (M, g) is strongly causal if and only if each point $r \in M$ has a convex normal neighbourhood U such that the local distance function (D, U) agrees on $U \times U$ with the distance function $d = d(g)\colon M \times M \to \mathbb{R} \cup \{\infty\}$.*

Proof. If (M, g) is strongly causal and $r \in M$, then there is some convex normal neighbourhood U of r such that no nonspacelike curve which leaves U ever returns. The local distance function for U then agrees with $d = d(g)|(U \times U)$.

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

Conversely, assume that strong causality breaks down at some point $r \in M$. Let U be a convex normal neighbourhood of r such that $D(p,q) = d(p,q)$ for all $p, q \in U$. There exists a neighbourhood $W \subset U$ of r such that any future directed nonspacelike curve $\gamma\colon (0,1] \to U$ with $\gamma(1) \in W$ and γ past inextendible in U contains some point not in $J^+(W,U)$. Since strong causality fails to hold at r, there is a future directed timelike curve $\gamma_1\colon [0,1] \to M$ with $r' = \gamma_1(0) \in W$, $\gamma_1(\frac{1}{2}) \notin U$ and $\gamma_1(1) \in W$. By construction of W, there is some point $p \in \gamma_1 \cap U$ with $p \notin J^+(r', U)$. Hence $D(r',p) = 0$. However, $d(r',p) > 0$ since $d(r',p)$ is at least as large as the length of γ_1 from r' to p. Thus $D(r',p) \neq d(r',p)$. Taking the contrapositive then establishes the theorem.

COROLLARY 3·5. *If (M, g) is strongly causal, then d is continuous on some neighbourhood of $\Delta(M) = \{(p,p); p \in M\}$ in $M \times M$.*

4. *Rays in globally hyperbolic space–times.* We first define the concept of a geodesic ray.

Definition 4·1. A future directed and future inextendible nonspacelike geodesic $\gamma\colon [0,a) \to M$, $0 < a \leqslant \infty$, is said to be a *future directed* nonspacelike *ray* if and only if $\gamma|[0,b]$ is maximal (i.e. $L(\gamma|[0,b]) = d(\gamma(0), \gamma(b))$) for any b with $0 < b < a$. A *past directed ray* may be defined dually.

We now show the existence of a future directed ray beginning at any point of a globally hyperbolic space–time. By duality there is a past directed ray beginning at any point.

THEOREM 4·2. *Let (M, g) be a globally hyperbolic space–time. Then given any point p of M, there is a future directed (resp. past directed) nonspacelike ray $\gamma\colon [0,a) \to M$ with $\gamma(0) = p$.*

Proof. Let $\{p_n\}$ be a sequence of distinct points of M with $p \ll p_n$ for all n and assume $\{p_n\}$ diverges (i.e. each compact subset of M contains only a finite number of elements of the sequence). Since (M, g) is globally hyperbolic, we may join p to p_n by a maximal future directed timelike geodesic segment γ_n with $\gamma_n(0) = p$. Extend each γ_n beyond p_n to a future inextendible nonspacelike curve, still denoted by γ_n. Let $\gamma\colon [0,a) \to M$ be a future directed and future inextendible nonspacelike limit curve of $\{\gamma_n\}$ with $\gamma(0) = p$, guaranteed by Lemma 6·2·1 of ((11), p. 185). We now show γ is maximal as a limit of maximal geodesic segments and hence is the required ray.

To this end, fix b with $0 < b < a$. Choose any $q \in M$ with $\gamma(b) \ll q$. Since $I^-(q)$ is open and γ is a limit curve of $\{\gamma_n\}$, we may find a subsequence $\{\gamma_m\}$ of $\{\gamma_n\}$ and a sequence $\{t_m\} \subset \mathbb{R}$ such that $\gamma_m(t_m) \to \gamma(b)$ and $\gamma_m(t_m) \in I^-(q) \cap I^+(p)$ for all m. Since $J^-(q) \cap J^+(p)$ is compact by the global hyperbolicity of (M, g), it follows that $\gamma_m|[0,t_m]$ is maximal for all m sufficiently large, as $\{p_n\}$ was divergent. Reparametrizing the geodesics γ_m, we may suppose that $\gamma_m\colon [0,b] \to M$ for all m with $\gamma_m(0) = p$, $\gamma_m(b) \to q$ and γ is a limit curve of $\{\gamma_m\}$. By Lemma 2·1 of Beem and Ehrlich (1), we may suppose, by passing to a subsequence of $\{\gamma_m\}$ if necessary, that $\gamma_m \to \gamma$ in the C^0 topology on curves (cf. ((14), pp. 49 ff.) for a discussion of this topology). Using upper

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

semi-continuity of length in the C^0 topology on nonspacelike curves in strongly causal space-times, we then have

$$L(\gamma|[0,b]) \geqslant \limsup L(\gamma_m|[0,b]) = \limsup d(p,\gamma_m(b))$$
$$= d(p,\gamma(b)).$$

The last equality holds since the distance function is continuous for globally hyperbolic space-times ((11), p. 215). As $d(p,\gamma(b)) \geqslant L(\gamma|[0,b])$ automatically from the definition of distance, we obtain $d(p,\gamma(b)) = L(\gamma|[0,b])$. Hence $\gamma|[0,b]$ is maximal for any $0 < b < a$. Therefore γ is a geodesic and thus the required ray.

Remark 4·3. It should be noted that the existence of a past and future inextendible ray at each point does *not* guarantee the existence of a line (Definition 5·1) through each point.

5. *Causally disconnecting sets in globally hyperbolic space-times.* We first recall some definitions and concepts from (1) and (2).

Definition 5·1. A past and future inextendible nonspacelike future directed geodesic γ: $(a, b) \to M$ is said to be a *line* if $d(\gamma(s), \gamma(t)) = L(\gamma|[s, t])$ for all s, t with $a < s \leqslant t < b$.

Definition 5·2. The space-time (M, g) is said to be *causally disconnected* if there exist a compact set $K \subset M$ and divergent sequences $\{p_n\}, \{q_n\}$ with

$$0 < d(p_n, q_n) < \infty$$

for all n such that, for each n, every future directed nonspacelike curve from p_n to q_n meets K.

The following result was established in (1), proposition 6·4.

PROPOSITION 5·3. *If the strongly causal space-time (M, g) is causally disconnected by the compact set K, then (M, g) contains a nonspacelike line γ: $(a, b) \to M$ which intersects K.*

If (M, g) is nonspacelike complete, then this line γ is defined for all real values of an affine parameter. Otherwise γ may only be defined on a proper subset of $\mathbb{R}$, in which case the space-time is nonspacelike incomplete and hence singular.

In the next proposition we give a sufficient condition for a globally hyperbolic space-time to have a causally disconnecting set.

PROPOSITION 5·4. *Let (M, g) be globally hyperbolic. If (M, g) has a null line, then (M, g) is causally disconnected.*

Proof. Let $c: (a, b) \to M$ be an inextendible maximal future directed null geodesic of M. Let r be any point of c and choose K to be any compact subset of M with $r \in \mathrm{Int}(K)$. Choose sequences $t_n \to a^+$ and $t'_n \to b^-$ such that $c(t_n) \leqslant r \leqslant c(t'_n)$ for each n. Let $p_n = c(t_n)$. Because (M, g) is globally hyperbolic and hence all distances are finite, it suffices to show that, for each n, there is some q_n with $c(t'_n) \ll q_n$ such that all nonspacelike curves from p_n to q_n meet K.

If no such q_n existed for some fixed n, then there would be a sequence of points $\{x_k\}$ converging to $c(t'_n)$ with $c(t'_n) \ll x_k$ for each k and a sequence of future directed

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

nonspacelike curves γ_k from $c(t_n)$ to x_k such that $K \cap \gamma_k = \phi$ for each k. The sequence $\{\gamma_k\}$ would have a future directed nonspacelike limit curve γ starting at $c(t_n)$. Since (M, g) is globally hyperbolic, γ would join $c(t_n)$ to $c(t_n')$. On the other hand, because c is a maximal null geodesic, the only (up to parametrization) nonspacelike curve from $c(t_n)$ to $c(t_n')$ is the null geodesic $c|[t_n, t_n']$. Thus $\gamma \subset c$ and the sequence $\{\gamma_k\}$ must intersect K for some large k. This establishes the result.

Remark 5·5. The Einstein static universe ((11), p. 121) shows that there are globally hyperbolic, causally disconnected space–times which do not have inextendible maximal null geodesics. Thus the existence of a null line is *not* a necessary condition for a globally hyperbolic space–time to be causally disconnected.

Proposition 5·4 shows that Minkowski space–time, de Sitter space–time, the Friedmann cosmological models and the steady state universe are all causally disconnected.

The next proposition gives a second sufficient condition for a globally hyperbolic space–time to be causally disconnected.

PROPOSITION 5·6. *Let (M, g) be globally hyperbolic. If $p \in M$ is not the origin of a future (past) directed maximal null geodesic ray, then (M, g) is causally disconnected.*

Proof. Assume p is not the origin of any future directed maximal null geodesic ray. The future horismos ((11), p. 184) of the set $\{p\}$ is $E^+(p) = J^+(p) - I^+(p)$. It is sufficient to show $E^+(p)$ is compact since this implies $\{p\}$ is a future trapped set and the future horismos of a future trapped set is a causally disconnecting set, compare (1).

If $E^+(p)$ is not compact, there is a divergent sequence $\{q_n\} \subset E^+(p)$. However, $q_n \in J^+(p) - I^+(p)$ implies there is a maximal null geodesic γ_n from p to q_n. Extend each γ_n to a future inextendible nonspacelike curve and let γ be a limit curve of the sequence $\{\gamma_n\}$. Using Lemma 2·1 of (1) and the fact that the sequence $\{q_n\}$ diverges, it is not hard to show that γ is a maximal future inextendible null geodesic starting at p. This contradicts the hypothesis.

We now show that for two-dimensional globally hyperbolic space–times, causally disconnecting sets always exist. We first establish the following lemma.

LEMMA 5·7. *Let (M, g) be a two-dimensional globally hyperbolic space–time. If $E^+(p)$* (resp. *$E^-(p)$*) *is not compact, then both of the future (resp. past) directed and future (resp. past) inextendible null geodesics starting at p are maximal.*

Proof. Assume $E^+(p)$ is noncompact for some $p \in M$. Let

$$c_1 : [0, a) \to M \quad \text{and} \quad c_2 : [0, b) \to M$$

be the two future directed, future inextendible null geodesics with $c_1(0) = c_2(0) = p$. If c_1 is not maximal, there is some null cut point $q = c_1(t_0)$ of p along c_1 with $0 < t_0 < a$. Let $t_n \to t_0^+$ and $t_n < a$ for each n. Since $c_1(t_n) \in I^+(p)$ and (M, g) is globally hyperbolic, there is a maximal future directed timelike geodesic γ_n from p to $c_1(t_0)$ for each n. The sequence $\{\gamma_n\}$ has a limit curve γ which is a nonspacelike geodesic, and γ joins p to q since (M, g) is globally hyperbolic. Furthermore, $d(p, q) = 0$ implies that γ is a null geodesic. Since dim $M = 2$, γ is either a geodesic subsegment of c_1 or a subsegment

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

of c_2. If $\gamma \subset c_2$, then c_2 passes through q at some parameter value t_0'. In this case $E^+(p) = c_1|[0, t_0] \cup c_2|[0, t_0']$ is compact and we have a contradiction.

Assume $\gamma \subset c_1$ and let $U(q)$ be a convex normal neighbourhood about q. Let $U(q)$ be chosen so small that no nonspacelike curve which leaves $U(q)$ ever returns. The inextendible null geodesics of $U(q)$ may be divided into two disjoint families $\mathscr{F}_1$ and $\mathscr{F}_2$ each of which cover $U(q)$ simply, see (5), p. 245, and (3). Let $\mathscr{F}_1$ be the class which contains the null geodesic $c_1 \cap U(q)$. Let c_3 be the unique null geodesic of $\mathscr{F}_2$ which contains q. For some fixed large n, the timelike geodesic γ_n must intersect c_3 at some point r. We must have $q \leqslant r$ since if $r \leqslant q$ then $p \ll r$ would yield $p \ll q$ which is false. However, $q \leqslant r$, $q = c_1 \cap c_3$, $q \leqslant c_1(t_n)$ and $r \in c_3$ imply $r \notin c(t_n)$; compare (5), p. 245. This contradicts the fact that $r \ll c_1(t_n)$ because r comes before $c_1(t_n)$ on the timelike geodesic γ_n. This establishes the lemma.

THEOREM 5·8. *All two-dimensional globally hyperbolic space–times are causally disconnected.*

Proof. Let (M, g) be two-dimensional and globally hyperbolic. If $E^+(p_0)$ is compact for some $p_0 \in M$, then each future directed null geodesic starting at p_0 has a cut point. Thus Proposition 5·6 yields the existence of a causally disconnecting set.

Now suppose $E^+(p)$ is noncompact for each $p \in M$. Assume c: $(a, b) \to M$ is a future directed, inextendible null geodesic. Let s, t with $a < s \leqslant t < b$ be arbitrary. Applying Lemma 5·7 at the point $p = c(s)$, we find that $c|[s, b)$ is maximal and thus

$$d(c(s), c(t)) = L(c|[s, t]).$$

This implies that c is maximal and hence c is a null line. Using Proposition 5·4, it follows that (M, g) is causally disconnected.

We obtain as a corollary to Proposition 5·3 and Theorem 5·8 the following.

COROLLARY 5·9. *Let (M, g) be any globally hyperbolic space–time of dimension 2. Then (M, g) contains a nonspacelike line.*

In connexion with Theorem 5·8 it should be noted

Remark 5·10. Let (M, g) be globally hyperbolic and causally disconnected. Denote by $C(M, g)$ the space of metrics for M globally conformal to g. Then all space–times (M, g'), $g' \in C(M, g)$ are causally disconnected.

This result is immediate from the two facts that (1) all metrics in $C(M, g)$ are globally hyperbolic and (2) if (M, g') is globally hyperbolic, then all Lorentzian distances $d(g')(p, q)$ are finite for all $p, q \in M$. Thus if K, $\{p_n\}$ and $\{q_n\}$ causally disconnect one (M, g') with $g' \in C(M, g)$, then K, $\{p_n\}$ and $\{q_n\}$ causally disconnect all metrics in $C(M, g)$.

6. *Closure properties of the cut locus.* In this section we show that if (M, g) is globally hyperbolic, then the null cut locus, the nonspacelike cut locus and the nonspacelike conjugate locus are all closed subsets of M.

We first establish:

LEMMA 6·1. *Let (M, g) be a strongly causal space–time. Assume $p \leqslant q_n$ and $q_n \to q \neq p$. If $\exp_p(v_n) = q_n$, $\exp_p(v) = q$ and the directions of the nonspacelike vectors v_n converge to the direction of v, then $v_n \to v$.*

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

Proof. Choose numbers $a_n > 0$ such that the vectors $w_n = a_n v_n$ converge to v. Then the sequence of points $r_n = \exp_p(w_n)$ must be defined for large n and must converge to q. Since (M, g) satisfies the causality condition, there is only one value t_n of t such that $\exp_p(t_n w_n) = q_n$, namely $t_n = a_n^{-1}$. Using $r_n \to q$, $q_n \to q$, and the strong causality of (M, g) near q, we find $t_n \to 1$. Since $v_n = t_n^{-1} w_n$, the sequence $\{v_n\}$ converges to v.

Fix a point $p \in M$. A tangent vector $v \in T_p M$ is a conjugate point to p in $T_p M$ if $(\exp_p)_*$ is singular at v. The conjugate points to p in $T_p M$ must form a closed subset of the domain of $\exp_p$ because $(\exp_p)_*$ is nonsingular on an open subset of $T_p M$. A point $q \in M$ is conjugate to p along a geodesic c if there is some conjugate point $v \in T_p M$ such that $\exp_p v = q$ and c is (up to reparametrization) the geodesic $\exp_p(tv)$. If $q \in M$, $p \leqslant q$ and q is conjugate to p along a nonspacelike geodesic, then we will say q is a *future nonspacelike conjugate* point of p. Past nonspacelike conjugate points are defined dually. The nonspacelike conjugate locus should be physically more interesting than the total conjugate locus since (presumably) no physical particle can traverse a spacelike curve. If (M, g) is a causal space–time then p cannot be in its own future (or past) nonspacelike conjugate locus because a nonspacelike geodesic passes through p at most once and $(\exp_p)_*$ is nonsingular at the origin of $T_p M$.

Remark 6·2. It is possible even in globally hyperbolic space–times for a point to be conjugate to itself along *spacelike* geodesics, however. This happens in any Einstein static universe of dim $\geqslant 3$.

LEMMA 6·3. *Let (M, g) be a globally hyperbolic space–time. Then the future (resp. past) nonspacelike conjugate locus in M is a closed subset of M.*

Proof. Assume that $\{q_n\}$ is a sequence of future nonspacelike conjugate points of p with $q_n \to q$. Then $q \neq p$, $p \leqslant q$ and $p \leqslant q_n$ for each n. Let $v_n \in T_p M$ be a nonspacelike vector such that $(\exp_p)_*$ is singular at v_n and $\exp_p(v_n) = q_n$. Then $c_n(t) = \exp_p(tv_n)$ is a future directed, future inextendible geodesic starting at p and containing q. The geodesics c_n must have a limit curve γ by lemma 6·2·1 of (**11**), p. 185. Since (M, g) is globally hyperbolic, γ must contain q. Using the strong causality of (M, g) and the fact that γ is a limit curve of nonspacelike geodesics we find that γ is itself a nonspacelike geodesic. Let v be the unique (nonspacelike) tangent vector to γ at $\gamma(0) = p$ such that $\exp_p(v) = q$. Since γ is a limit curve of $\{c_n\}$ there must be some subsequence $\{m\}$ of $\{n\}$ such that the directions of the vectors v_m converge to the direction of v. Lemma 6·1 then implies that $v_m \to v$. Since the vectors v_m belong to the total conjugate locus of p in $T_p M$ and this conjugate locus is a closed subset of the domain of $\exp_p$, v is a conjugate point of p in $T_p M$. Because v is a future directed nonspacelike vector, the point $q = \exp_p(v)$ is a future nonspacelike conjugate point of p in M. This establishes the closure in M of the future nonspacelike conjugate locus. The dual proof shows that the past nonspacelike conjugate locus is also closed.

In the rest of this section we will use the sets $C_t^+(p)$, $C_N^+(p)$ and $C^+(p)$ as well as the function $s\colon T_{-1} M \to \mathbb{R} \cup \{\infty\}$. These were defined in section 2.

Example 6·4. Let $M = S^1 \times \mathbb{R} = \{(\theta, t) | 0 \leqslant \theta \leqslant 2\pi, t \in \mathbb{R}\}$ with the flat metric

$$ds^2 = d\theta^2 - dt^2.$$

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

This is the two-dimensional Einstein static universe which is globally hyperbolic. There are no conjugate points in this space-time. If $p = (0, 0)$, the future null cut locus $C_N^+(p)$ consists of the single point (π, π). The future timelike cut locus $C_t^+(p)$ consists of $\{(\pi, t) | t > \pi\}$ and is not closed. On the other hand, $C^+(p) = C_t^+(p) \cup C_N^+(p)$ is closed.

If we delete the two points $(\pm \pi/4, 2\pi)$ from M, we obtain a space-time $(M', g|M')$ which is not globally hyperbolic. In $(M', g|M')$ the future nonspacelike cut locus $C^+(p)$ is *not* closed.

PROPOSITION 6·5. *Let (M, g) be a globally hyperbolic space-time. For any $p \in M$, the sets $C_N^+(p)$, $C_N^-(p)$, $C^+(p)$ and $C^-(p)$ are all closed subsets of M. In particular, the null cut locus and the nonspacelike cut locus of p are each closed.*

Proof. Since the four cases are similar we will only show $C_N^+(p)$ is closed in M.

To this end, let $q_n \in C_N^+(p)$ and $q_n \to q$. Since (M, g) is globally hyperbolic, $q \neq p$ and $q \in J^+(p)$. From the definition of $C_N^+(p)$ we see that $d(p, q_n) = 0$. Using the continuity of Lorentzian distance for globally hyperbolic space-times ((11), p. 215), it follows that $d(p, q) = \lim d(p, q_n) = 0$. Let γ be any nonspacelike geodesic from p to q. Then $d(p, q) = 0$ implies that γ is a maximal null geodesic and that the cut point to p along γ cannot come before q.

There are two cases to consider. Either infinitely many points q_n are future nonspacelike conjugate to p or else no q_n is future nonspacelike conjugate to p for large n. In the first case we apply Lemma 6·3 and find that q is a future nonspacelike conjugate point to p. If q is future nonspacelike conjugate to p along γ, then q is the cut point along γ because the cut point along γ comes before or at the first conjugate point along γ. If q is a future nonspacelike conjugate point to p along some other nonspacelike geodesic γ', then γ' must be null and lemma 5·1 of Beem and Ehrlich (2) shows that q is the cut point along both γ and γ'.

Assume now that q_n is not future nonspacelike conjugate to p for all sufficiently large n. Theorem 5·3 of Beem and Ehrlich (2) implies that for large n there exist at least two maximal null geodesics from p to q_n. Thus for large n there are two null vectors v_n and w_n with $\exp_p(v_n) = \exp_p(w_n) = q_n$. The future directed nonspacelike directions at p form a compact set of directions. Thus the directions determined by $\{v_n\}$ and $\{w_n\}$ have limit directions v and w respectively. If v and w determine different directions, we apply lemma 5·2 of (2) to obtain two maximal null geodesics from p to q. In this case q is a cut point of p. On the other hand, v and w may determine the same direction. If this is so, we first note that there are constants $a > 0$ and $b > 0$ such that $av = bw$ and $\exp_p(av) = \exp_p(bw) = q$. Applying Lemma 6·1 it follows that for some subsequence $\{m\}$ of $\{n\}$ we have $v_m \to av$ and $w_m \to av$. However $v_m \neq w_m$ and $\exp_p(v_m) = \exp_p(w_m)$ then yield that $(\exp_p)_*$ must be singular at av. Thus q is conjugate to p along $\gamma(t) = \exp_p(tav)$ and since $d(p, q) = 0$ we find that q is the cut point of p along $\gamma(t)$.

Surprisingly the above result holds without any assumptions about the timelike or null geodesic completeness of (M, g). In particular, the null cut locus and the nonspace-

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

like cut locus in globally hyperbolic space-times are closed by Proposition 6·5 even though the function s: $T_{-1}M \to \mathbb{R} \cup \{\infty\}$ may fail to be upper semi-continuous.

If (M, g) is timelike geodesically complete as well as globally hyperbolic, then the function s is continuous on $T_{-1}M$. This continuity will be needed in the next proposition.

Recall the future horismos is defined by $E^+(p) = J^+(p) - I^+(p)$.

PROPOSITION 6·6. *Let (M, g) be globally hyperbolic and timelike geodesically complete. For each $p \in M$ the set $J^+(p) - (C^+(p) \cup E^+(p))$ is an open cell.*

Proof. Let $B = J^+(p) - (C^+(p) \cup E^+(p))$. Then $q \in B$ if and only if there is a maximal future directed timelike geodesic which starts at p and extends beyond q. Thus $B = I^+(p) - C^+(p)$ which shows that B is open. Now $T_{-1}M|_p = \{v \in T_pM;\ v$ is future directed and $g(v, v) = -1\}$ is homeomorphic to $\mathbb{R}^{n-1}$. Let H: $T_{-1}M|_p \to \mathbb{R}^{n-1}$ be a homeomorphism. Then s: $T_{-1}M|_p \to \mathbb{R} \cup \{\infty\}$ induces a map $\bar{s}$: $\mathbb{R}^{n-1} \to \mathbb{R} \cup \{\infty\}$ by $\bar{s} = s \circ H^{-1}$. There is an induced homeomorphism of B with

$$\{(x, t) \in \mathbb{R}^{n-1} \times \mathbb{R};\ 0 < t < \bar{s}(x)\}$$

defined by $q \to (H(v), t)$, where $\exp_p(tv) = q$ and $v \in T_{-1}M$. Let f: $[0, \infty] \to [0, 1]$ be a homeomorphism with $f(0) = 0$ and $f(\infty) = 1$. Then the map $(x, t) \to (x, f(t)/f(\bar{s}(x)))$ shows B is homeomorphic to $\mathbb{R}^{n-1} \times (0, 1)$ which establishes the proposition.

Remark 6·7. Since $I^+(p) \subset J^+(p)$, Proposition 6·6 yields some indirect topological information about $I^+(p)$. The Einstein static universe shows that $I^+(p)$ need not be an open cell even if (M, g) is globally hyperbolic. However, $(I^+(p), g|I^+(p))$ is globally hyperbolic whenever (M, g) is globally hyperbolic. Thus $I^+(p)$ may be expressed topologically as a product $I^+(p) = S \times \mathbb{R}$, where S is an $(n-1)$-dimensional manifold, compare Geroch (9).

7. *The index theory of timelike geodesics.* While several studies of the Morse theory of the space of causal curves joining two points of a globally hyperbolic space-time have been made ((7), (17), (18)), no complete development of the index theory for timelike geodesic segments in an arbitrary space-time has been published. In this section we give such a treatment for use in the sequel. Our treatment parallels that for Riemannian manifolds given by Gromoll, Klingenberg and Meyer in (10), sections 4·5 and 4·6.

Let (M, g) be an *arbitrary* space-time of dimension $\geqslant 2$ throughout this section. We will denote by $\langle , \rangle$ the Lorentzian metric g in this section. Also, all timelike geodesics c: $[a, b] \to M$ will be assumed to have unit speed, i.e. $\langle c'(t), c'(t) \rangle = -1$ for all $t \in [a, b]$.

Definition 7·1. A *piecewise smooth vector field Y along* (the timelike geodesic) c: $[a, b] \to M$ is a map Y: $[a, b] \to TM$ with $Y(t) \in T_{c(t)}M$ for all $t \in [a, b]$, such that there exists some finite partition $t_0 = a < t_1 < \ldots < t_k = b$ of $[a, b]$ so that $Y|(t_i, t_{i+1})$ is smooth for each i with $0 \leqslant i \leqslant k-1$. Let $V_{\text{per}}(c)$ denote the $\mathbb{R}$-vector space of all piecewise smooth vector fields Y along c with $\langle Y(t), c'(t) \rangle = 0$ for all $0 \leqslant t \leqslant b$. Also let $V_{\text{per},0}(c) = \{Y \in V_{\text{per}}(c);\ Y(a) = Y(b) = 0\}$ and $N(c(t)) = \{v \in TM;\ \langle v, c'(t) \rangle = 0\}$.

Remark 7·2. Since $c'(t)$ is timelike, $N(c(t))$ is a vector space of dimension $(n-1)$ consisting of spacelike tangent vectors and thus $\{v \in N(c(t));\ \langle v, v \rangle = 1\}$ is compact.

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

Definition 7·3. Let $c: [a,b] \to M$ be a timelike geodesic. Then $t_1, t_2 \in [a,b]$ with $t_1 \neq t_2$ are *conjugate with respect to* c if and only if there is a nontrivial vector field J along c with $J(t_1) = J(t_2) = 0$ satisfying the Jacobi equation $J'' + R(J,c')c' = 0$. Here J' denotes the covariant derivative operator of vector fields along c induced by the Lorentzian connexion of $\langle , \rangle$ on M (cf. (10), p. 47). A vector field J along c satisfying $J'' + R(J,c')c' = 0$ is called a *Jacobi field along* c.

We now define the Lorentzian index form $\mathrm{I}: V_{\mathrm{per}}(c) \times V_{\mathrm{per}}(c) \to \mathbb{R}$.

Definition 7·4. The *index form* $\mathrm{I}: V_{\mathrm{per}}(c) \times V_{\mathrm{per}}(c) \to \mathbb{R}$ is the symmetric bilinear form given by

$$I(X, Y) = -\int_a^b (\langle X', Y' \rangle - \langle R(X,c')c', Y \rangle)\, dt$$

for any $X, Y \in V_{\mathrm{per}}(c)$.

If $X \in V_{\mathrm{per}}(c)$ is smooth, we also have

$$I(X, Y) = -\langle X', Y \rangle \Big|_a^b + \int_a^b \langle X'' + R(X,c')c', Y \rangle\, dt.$$

Thus if $Y \in V_{\mathrm{per},0}(c)$ and X is smooth, one obtains the formula

$$I(X, Y) = \int_a^b \langle X'' + R(X,c')c', Y \rangle\, dt$$

linking the index form to Jacobi fields.

Remark 7·5. With our definition, the index form has the following geometric interpretation (cf. (11), lemma 4·5·6). Given $Y \in V_{\mathrm{per},0}(c)$, define a variation of c by setting $V(t,s) := \exp_{c(t)}(sY(t))$. If we put $L(s) := L(t \mapsto V(t,s))$, then $L'(0) = 0$ since c is a smooth timelike geodesic and $L''(0) = I(Y, Y)$.

We need the following

LEMMA 7·6. *For* $Y \in V_{\mathrm{per},0}(c)$, *the following are equivalent:*
(*a*) Y *is a Jacobi field along* c.
(*b*) $I(Y, Z) = 0$ *for all* $Z \in V_{\mathrm{per},0}(c)$.

Proof. First $(a) \Rightarrow (b)$ is obvious from the formula for the index form restricted to $V_{\mathrm{per},0}(c)$. On the other hand, $(b) \Rightarrow (a)$ by the standard calculus of variations proof in Riemannian spaces in (10), p. 144, lemma 2, because

$$Y',\ Y'',\ Y'' + R(Y,c')c' \in V_{\mathrm{per}}(c) \quad \text{if} \quad Y \in V_{\mathrm{per}}(c)$$

(just differentiate $\langle Y, c' \rangle = 0$).

Using the negative semi-definiteness of the index form established in Hawking and Ellis (11), lemma 4·5·8, one may now prove negative definiteness. This has been noted in Bölts (4), Satz 4·4·5.

PROPOSITION 7·7. *For a timelike geodesic* $c: [a,b] \to M$, *the following are equivalent:*
(*a*) c *has no conjugate points in* $(a,b]$.
(*b*) $I: V_{\mathrm{per},0}(c) \times V_{\mathrm{per},0}(c) \to \mathbb{R}$ *is negative definite.*

Proof. $(a) \Rightarrow (b)$. Lemma 4·5·8 of (11) implies $I(Y, Y) \leqslant 0$ if $Y \in V_{\mathrm{per},0}(c)$. Thus we need only show that, if $Y \in V_{\mathrm{per}.0}(c)$ and $I(Y, Y) = 0$, then $Y = 0$. To this end, let

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

$Z \in V_{\text{per},0}(c)$ be arbitrary. By Remark 7·2, we have $Y - tZ \in V_{\text{per},0}(c)$ for any $t \in \mathbb{R}$ and thus $I(Y - tZ, Y - tZ) \leqslant 0$. Since $I(Y - tZ, Y - tZ) = -2tI(Y, Z) + t^2 I(Z, Z)$, it follows that $I(Y, Z) = 0$. Hence $Y = 0$ by Lemma 7·6.

$(b) \Rightarrow (a)$: First note that if Y is a Jacobi field with $Y(a) = Y(t_1) = 0$, $a < t_1 \leqslant b$, then $Y \in V_{\text{per}}(c)$ since $\langle Y, c' \rangle'' = \langle Y'', c' \rangle = -\langle R(Y, c')c', c' \rangle = 0$. Thus if t_1 is a conjugate point, a nontrivial Jacobi field Y with $Y(a) = Y(t_1) = 0$ may be extended to a nontrivial $Z \in V_{\text{per},0}(c)$ by setting $Z(t) = Y(t)$ for $a \leqslant t \leqslant t_1$ and $Z(t) = 0$ for $t \geqslant t_1$. A calculation then shows that $I(Z, Z) = 0$.

A consequence of Proposition 7·7 is the following maximality property of Jacobi fields with respect to the index form.

THEOREM 7·8. *Let $c: [a, b] \to M$ be a timelike geodesic with no conjugate points and let $J \in V_{\text{per}}(c)$ be any Jacobi field. Then, for any $Y \in V_{\text{per}}(c)$ with $Y \neq J$ and $Y(a) = J(a)$, $Y(b) = J(b)$, we have $I(J, J) > I(Y, Y)$.*

Proof. Calculate $0 > I(J - Y, J - Y) = I(Y, Y) - I(J, J)$ as in (10), p. 145, reversing the inequality signs.

Now that Theorem 7·8 is established, a Morse Index Theorem may be formulated and proved along the lines of (10), section 4·6. First we must define the index of any timelike geodesic $c: [a, b] \to M$. The definition given makes sense because $V_{\text{per},0}(c)$ is a vector space.

Definition 7·9. The *extended index* $\text{Ind}_0(c)$ and the *index* $\text{Ind}(c)$ of the timelike geodesic $c: [a, b] \to M$ are defined by $\text{Ind}_0(c) = \text{lub}\{\dim A; A \text{ is a vector subspace of } V_{\text{per},0}(c) \text{ and } I|A \times A \text{ is positive semi-definite}\}$ and $\text{Ind}(c) = \text{lub}\{\dim A; A \text{ is a vector subspace of } V_{\text{per},0}(c) \text{ and } I|A \times A \text{ is positive definite}\}$. Also let $J_t(c)$ denote the $\mathbb{R}$-vector space of Jacobi fields Y along c with $Y(a) = Y(t) = 0$, $a < t \leqslant b$.

Using Theorem 7·8 and the standard finite dimensional approximation of $V_{\text{per},0}(c)$ by piecewise smooth Jacobi fields, it may be established as in the Riemannian case ((10), p. 149) that $\text{Ind}(c)$, $\text{Ind}_0(c)$ are finite and $\text{Ind}_0(c) = \text{Ind}(c) + \dim J_t(c)$. One may then establish the

MORSE INDEX THEOREM. *Let $c: [a, b] \to M$ be a timelike geodesic and, for $t \in (a, b]$, let $J_t(c)$ denote the vector space of smooth Jacobi fields along c with $Y(a) = Y(t) = 0$. Then*

$$\text{Ind}(c) = \sum_{a<t<b} \dim J_t(c)$$

and

$$\text{Ind}_0(c) = \sum_{a<t\leqslant b} \dim J_t(c).$$

If $t_1, t_2, \ldots, t_r$ are the conjugate points of c in $(a, b]$ with multiplicities $m_1, \ldots, m_r$, then

$$\text{Ind}(c) = \sum_{i=1}^{r} m_i.$$

Proof. Recall that $J_t(c) \subset V_{\text{per}}(c)$ as was shown in the proof of $(b) \Rightarrow (a)$ of Proposition 7·7. The proof given in ((10), pp. 150–152) of the Morse Index Theorem works in the Lorentzian case if the E on page 151 is changed to $E = N(c(t_1)) \times \ldots \times N(c(t_{k-1}))$ rather than using the full tangent spaces. Even though $(M, \langle , \rangle)$ is Lorentzian, the set S defined in (10) from the modified E is compact by Remark 7·2.

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

8. *Morse Theory and cut points.* Let (M, g) be globally hyperbolic in this section. Our aim in this section is to prove a Lorentzian analogue of Theorem 5·11 given in Cheeger and Ebin ((6), p. 101). Namely, in a complete, simply connected Riemannian space, if for some point p, the first conjugate point along every geodesic starting at p is of order $\geqslant 2$, then the cut locus of p coincides with the first conjugate locus of p.

To this end recall (8).

Definition 8·1. M is *future 1-connected* if and only if, for each q in the chronological future of p, any two (smooth) future directed timelike curves from p to q are homotopic through future directed timelike curves with endpoints p and q.

Note that, if M is future 1-connected, then the path space of (smooth) timelike curves from p to q is connected. Thus using lemma 4·11–(2) of (6), p. 85, and standard path space Morse Theory (cf. (7)), it follows that

PROPOSITION 8·2. *Let (M, g) be future 1-connected and globally hyperbolic. Fix $p \in M$ and suppose that every future directed timelike geodesic starting at p has index 0, or index $\geqslant 2$. Given $q \in I^+(p)$ such that q is not conjugate to p along any future directed timelike geodesic from p to q, there is exactly one future directed timelike geodesic from p to q of index 0, namely, the unique maximal geodesic from p to q.*

Now we are ready to prove

THEOREM 8·3. *Let (M, g) be future 1-connected and globally hyperbolic. Suppose that for $p \in M$, the first future conjugate point of p along every timelike geodesic γ with $\gamma(0) = p$ is of order $\geqslant 2$. Then the future timelike cut locus of p and the locus of first future conjugate points of p coincide. Thus all timelike geodesics from p maximize up to the first future conjugate point.*

Proof. All geodesics will be unit speed and future timelike during the course of this proof. It suffices to show that, if $\gamma: [0, t] \to M$ with $\gamma(0) = p$ has index 0 and $\gamma(t)$ is not conjugate to p along $\gamma|[0, t]$, then γ is maximal. Since the set of the singular points of $\exp_p$ is closed, it follows from the Morse Index Theorem that there exist $\epsilon_1, \epsilon_2 > 0$ such that if $\measuredangle(v, \gamma'(0)) < \epsilon_1$ and $\epsilon < \epsilon_2$, then the future timelike geodesic

$$\sigma: [0, t+\epsilon] \to M$$

with $\sigma'(0) = v$ is of index 0.

By Sard's Theorem, we may find a sequence of points $\{p_i\} \subset I^+(p)$ with $p_i \to \gamma(t)$ such that every timelike geodesic segment from p to p_i has nonconjugate endpoints. Thus, by hypothesis, every such segment has index 0 or index $\geqslant 2$. By Proposition 8·2 there is a unique timelike maximal geodesic segment γ_i from p to p_i of index 0.

Since $(\exp_p)_*$ is nonsingular at $t\gamma'(0)$, for i sufficiently large, there are geodesic segments $\bar{\gamma}_i$ from p to p_i with $\bar{\gamma}_i \to \gamma$. Since $\measuredangle(\bar{\gamma}_i'(0), \gamma'(0)) \to 0$ these segments have index 0 for large i. It follows that, for i sufficiently large, $\bar{\gamma}_i = \gamma_i$ and hence $\bar{\gamma}_i$ is maximal. Thus γ is maximal as a limit of maximal geodesics.

9. *Lorentzian comparison theorems and timelike diameter.* The purpose of this section is to give for arbitrary space–times the Lorentzian analogues of two notable tools in Riemannian geometry, namely the Index Comparison Theorem ((10), p. 174) and the Rauch Comparison Theorem ((10), p. 178). We also establish the analogue of the Bonnet-Myers Diameter Theorem ((10), p. 212) for globally hyperbolic space–times.

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

Let (M_1, g_1) and (M_2, g_2) be arbitrary space–times of the same dimension $n \geqslant 2$. Let $c_1\colon [0,\beta] \to M_1$ and $c_2\colon [0,\beta] \to M_2$ be unit speed future directed timelike geodesic segments. Throughout this section, we will denote both g_1 and g_2 by $\langle , \rangle$ and both the index form on $V_{\mathrm{per}}(c_1) \times V_{\mathrm{per}}(c_1)$ and that on $V_{\mathrm{per}}(c_2) \times V_{\mathrm{per}}(c_2)$ by I.

We first need to define an isomorphism $\phi\colon V_{\mathrm{per}}(c_1) \to V_{\mathrm{per}}(c_2)$ so that

$$\langle \phi X, \phi X\rangle|_t = \langle X, X\rangle|_t \quad \text{for all} \quad t \in [0,\beta].$$

This may be done following the classical parallel translation construction in Riemannian geometry ((10), p. 175). We first define an isometry $\phi_t\colon T_{c_1(t)}M_1 \to T_{c_2(t)}M_2$ as follows. Let $P_t\colon T_{c_1(t)}M_1 \to T_{c_1(0)}M_1$ (resp. $Q_t\colon T_{c_2(t)}M_2 \to T_{c_2(0)}M_2$) denote the vector space isomorphism of parallel translation along c_1 (resp. c_2) from $T_{c_1(t)}M_1$ to $T_{c_1(0)}M_1$ (resp. from $T_{c_2(t)}M_2$ to $T_{c_2(0)}M_2$). Let $i\colon (T_{c_1(0)}M_1, \langle , \rangle) \to (T_{c_2(0)}M_2, \langle , \rangle)$ be an isometry with $i(c_1'(0)) = c_2'(0)$. Then the map $\phi_t\colon (T_{c_1(t)}M_1, \langle , \rangle) \to (T_{c_2(0)}M_2, \langle , \rangle)$ given by $\phi_t = Q_t^{-1} \circ i \circ P_t$ is an isometry since parallel translation preserves the Lorentzian structures. Given $X \in V_{\mathrm{per}}(c_1)$, we may then define $\phi X \in V_{\mathrm{per}}(c_2)$ by $(\phi X)|_t = \phi_t(X|_t)$. It also follows, as in ((10), p. 175), that $(\phi X)' = \phi(X')$.

Let $G_{2,t}(c_i)$ denote the set of all timelike plans σ (i.e. all 2-dimensional subspaces of $T_{c_i(t)}M_i$ spanned by a spacelike and a timelike vector) containing $c_i(t)$ for $i = 1, 2$. There is then an induced map $\phi_t\colon G_{2,t}(c_1) \to G_{2,t}(c_2)$ defined as follows. If $\sigma \in G_{2,t}(c_1)$, we may write $\sigma = \mathrm{span}\{v, c_1'(t)\}$, where v is spacelike. Put $\phi_t(\sigma) = \mathrm{span}\{\phi_t(v), c_2'(t)\} \in G_{2,t}(c_2)$.

We are now ready to state the

INDEX COMPARISON THEOREM. *Let (M_1, g_1) and (M_2, g_2) be space–times of the same dimension and let $c_1\colon [0,\beta] \to M_1$, and $c_2\colon [0,\beta] \to M_2$ be unit speed future directed timelike geodesics. Suppose that, for all t with $0 \leqslant t \leqslant \beta$ and all timelike planes $\sigma \in G_{2,t}(c_1)$, the sectional curvature $K_{M_1}(\sigma) \geqslant K_{M_2}(\phi_t \sigma)$. Then for any $X \in V_{\mathrm{per}}(c_1)$ we have*

(1) $I(X, X) \leqslant I(\phi X, \phi X)$.

Thus

(2) $\mathrm{Ind}\,(c_1) \leqslant \mathrm{Ind}\,(c_2)$,

and

(3) $\mathrm{Ind}_0(c_1) \leqslant \mathrm{Ind}_0(c_2)$.

Proof. Recalling that $\langle c_1', c_1'\rangle = -1$ and $\langle X, c_1'\rangle = 0$, we obtain

$$\langle R(X, c_1')\, c_1', X\rangle = -\langle X, X\rangle\, K(X, c_1').$$

Similar formulas hold for ϕX and c_2'. Thus

$$\begin{aligned} I(\phi X, \phi X) &= \int_0^\beta (-\langle (\phi X)', (\phi X)'\rangle + \langle R(\phi X, c_2')\, c_2', \phi X\rangle)\, dt \\ &= \int_0^\beta (-\langle \phi(X'), \phi(X')\rangle + \langle R(\phi X, c_2')\, c_2', \phi X\rangle)\, dt \\ &\geqslant \int_0^\beta (-\langle X', X'\rangle + \langle R(X, c_1')\, c_1', X\rangle)\, dt = I(X, X). \end{aligned}$$

COROLLARY 9·1. *If (M, g) satisfies the curvature condition that all timelike planes have sectional curvature $\leqslant -k < 0$ and $c\colon [0,\beta] \to M$ is any timelike geodesic with $L(c) > \pi/\sqrt{k}$,*

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

then c has a pair of conjugate points on $(0, \beta)$, i.e. there is a nontrivial Jacobi field $Y \in V_{\text{per}}(c)$ with $Y(t_1) = Y(t_2) = 0$, $0 < t_1 < t_2 < \beta$.

Using Corollary 9·1, we may obtain the analogue of Bonnet's Theorem (6), p. 27. First we need (cf. Tipler (15), p. 17)

Definition 9·2. The *timelike diameter* $\operatorname{diam}(M, g)$ of (M, g) is defined to be

$$\operatorname{diam}(M, g) = \sup\{d(p, q);\ p, q \in M\}.$$

Physically the timelike diameter represents the supremum of possible proper times any particle could possibly experience in the given space–times. A space–time of finite timelike diameter is singular in a striking way.

Remark 9·3. If $\operatorname{diam}(M, g) < \infty$, then *all* timelike geodesics are incomplete and, in fact, must have length $\leqslant \operatorname{diam}(M, g)$.

From a physical point of view, the most interesting space–times of finite timelike diameter are the Wheeler Universes (cf. Tipler (16), p. 500). In particular, the 'closed' Friedmann cosmological models are examples of Wheeler Universes.

For a complete Riemannian manifold, the diameter is finite if and only if the manifold is compact. In this case we may always find two points of M whose distance realizes the diameter. On the other hand, for space–times

PROPOSITION 9·4. *Let (M, g) be an arbitrary space–time and suppose that there exist points $p, q \in M$ such that $d(p, q) = \operatorname{diam}(M, g)$. Then $d(p, q) = \infty$.*

Proof. Suppose $d(p, q) = \operatorname{diam}(M, g) < \infty$. Let $q' \in I^+(q)$ be arbitrary. Then

$$d(p, q') \geqslant d(p, q) + d(q, q') > d(p, q) = \operatorname{diam}(M, g),$$

in contradiction.

As a consequence of this proposition, we see that (1) the timelike diameter is never realized by any pair of points in a space–time of finite timelike diameter and (2) the timelike diameter is never achieved in a globally hyperbolic space–time.

We now prove the Lorentzian form of Bonnet's Theorem.

THEOREM 9·5. *Let (M, g) be a globally hyperbolic space–time with all timelike sectional curvatures $\leqslant -k < 0$. Then $\operatorname{diam}(M, g) \leqslant \pi/\sqrt{k}$.*

Proof. Suppose that $\operatorname{diam}(M, g) > \pi/\sqrt{k}$. We may then find $p, q \in M$ with

$$d(p, q) > \pi/\sqrt{k}$$

by definition of $\operatorname{diam}(M, g)$. Since (M, g) is globally hyperbolic, there is a maximal timelike geodesic segment $c: [0, \beta] \to M$ with $c(0) = p$, $c(\beta) = q$. On the other hand, since the timelike sectional curvatures satisfy $K(\sigma) \leqslant -k < 0$, the geodesic

$$c: [0, \beta] \to M$$

contains a pair of conjugate points in $(0, \beta)$ by Corollary 9·1. Hence c is not maximal, in contradiction.

Similar arguments using Theorem 7·8 may be used to establish the Lorentzian analogue of Myers' Theorem (cf. (6), pp. 27–28)).

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

THEOREM 9·6. *Let (M, g) be a globally hyperbolic space–time with*

$$\mathrm{Ric}(v, v) \geqslant (n-1)k > 0$$

for all unit timelike vectors $v \in TM$. Then $\mathrm{diam}(M, g) \leqslant \pi/\sqrt{k}$.

We conclude with a Lorentzian analogue of the Rauch Comparison Theorem ((10), p. 178).

TIMELIKE RAUCH COMPARISON THEOREM. *Let (M_1, g_1) and (M_2, g_2) be two space–times of the same dimension. Let $c_1: [0, \beta] \to M_1$ and $c_2: [0, \beta] \to M_2$ be future directed timelike unit speed geodesic segments. Suppose, for all t with $0 \leqslant t \leqslant \beta$ and any $\sigma \in G_{2,t}(c_1)$, that $K_{M_1}(\sigma) \geqslant K_{M_2}(\phi_t \sigma)$. Let $Y_1 \in V_{\mathrm{per}}(c_1)$ and $Y_2 \in V_{\mathrm{per}}(c_2)$ be Jacobi fields satisfying the initial conditions*

(1) $Y_1(0) = Y_2(0) = 0$,

(2) $\|Y_1'(0)\| = \|Y_2'(0)\|$.

If c_2 has no conjugate points in $(0, \beta)$, then

(3) $\|Y_1(t)\| \geqslant \|Y_2(t)\| \quad (0 \leqslant t \leqslant \beta)$.

In particular, c_1 has no conjugate points in $(0, \beta)$.

Proof. This may be given along the lines of ((10), pp. 180–1), except the proof of the inequality (7), p. 181, must be modified as follows. Let $Z \in V_{\mathrm{per}}(c_2)$ be the unique Jacobi field along c_2 with $Z(0) = 0$ and $Z(t_0) = (\phi Y_1)(t_0)$. We must show that

$$\langle Y_1, Y_1' \rangle|_{t_0} \geqslant \langle Z, Z' \rangle|_{t_0}.$$

To this end let $c_3 = c_1|[0, t_0]$ and $c_4 = c_2|[0, t_0]$. Then

$$\begin{aligned} \langle Y_1, Y_1' \rangle|_{t_0} &= -I(Y_1|c_3, Y_1|c_3) \\ &\geqslant -I((\phi Y_1)|c_4, (\phi Y_1)|c_4) \end{aligned}$$

by the Index Comparison Theorem (a) proved above

$$\geqslant -I(Z|c_4, Z|c_4)$$

by the maximality of Jacobi fields (Theorem 7·8)

$$= \langle Z, Z' \rangle|_{t_0}.$$

Appendix. The index form for a null geodesic. The purpose of this appendix is to sketch Gerhard Bölts' treatment of the index theory for a null geodesic in an arbitrary space–time given in chapter 4 of Bölts' unpublished Bonn University Diplomarbeit (4). While Bölts' discussion of the null index theory was motivated by the treatment of the second variation of arc length for null geodesics in Hawking and Ellis (11), pp. 113–116), no explicit treatment of the null index theory is given in (11). Bölts' Diplomarbeit also contains a complete treatment of the index theory for null geodesics orthogonal to a spacelike 2-surface in a 4-dimensional space–time, although we will not discuss this below.

Let (M, g) be an arbitrary space–time of dimension $\geqslant 2$ and let $\beta: [a, b] \to M$ be an arbitrary null geodesic. As in Definition 7·1, let $V_{\mathrm{per}}(\beta)$ denote the $\mathbb{R}$-vector space of all piecewise smooth vector fields Y along β with $\langle Y(t), \beta'(t) \rangle = 0$ for all $a \leqslant t \leqslant b$. Also

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

let $N(\beta(t)) = \{v \in T_{\beta(t)}M;\ \langle v, \beta'(t)\rangle = 0\}$. The difficulty in developing the index theory of null (as opposed to timelike) geodesics is that β' and $t\beta'$ are both Jacobi fields in $V_{\text{per}}(\beta)$, but neither of these Jacobi fields give rise to conjugate points. Also Jacobi fields Y along the null geodesic β that are *not* orthogonal to β' may *not* be split in the form $Y = f\beta' + N$ with f an affine function and $\langle N, \beta'\rangle = 0$. For if such a decomposition existed, we would have $\langle Y, \beta'\rangle = 0$ (Bölts(4), Bemerkung 1·4).

The method Bölts uses to circumvent these difficulties is to work with the quotient bundle $N(\beta(t))/[\beta'(t)]$, where $[\beta'(t)] = \{\lambda\beta'(t);\ \lambda \in \mathbb{R}\}$, rather than working directly with vector fields along β. Let $N(\beta)/[\beta']$ denote the quotient bundle $\bigcup_{a\leqslant t\leqslant b} N(\beta(t))/[\beta'(t)]$ along β, and let $\pi: N(\beta)\to N(\beta)/[\beta']$ denote the natural projection map. One may extend the Lorentzian metric g for M to a positive definite bilinear form

$$\bar{g}: (N(\beta)/[\beta']) \times (N(\beta)/[\beta'])\to\mathbb{R}$$

by setting $\bar{g}(\bar{X}, \bar{Y}) := g(X, Y)$ for any $X, Y \in N(\beta)$ with $\pi(X) = \bar{X}$ and $\pi(Y) = \bar{Y}$ (Bölts(4), p. 42).

Let $\bar{V}_{\text{per}}(\beta) = \pi(V_{\text{per}}(\beta))$ denote the vector space of piecewise smooth sections $\bar{Y}: [a, b]\to N(\beta)/[\beta']$, (i.e. $\bar{Y}(t) \in N(\beta(t))/[\beta'(t)]$ for each $a \leqslant t \leqslant b$), and given $Y \in V_{\text{per}}(\beta)$, let $\bar{Y}: = \pi(Y) \in \bar{V}_{\text{per}}(\beta)$. The covariant derivative of vector fields in $V_{\text{per}}(\beta)$ then extends (Bölts(4), p. 42) to sections in $\bar{V}_{\text{per}}(\beta)$ as follows. Given $\bar{X} \in V_{\text{per}}(\beta)$, choose any $X \in V_{\text{per}}(\beta)$ with $\pi(X) = \bar{X}$ and set $\bar{X}'(t) := \pi(X'(t))$. Also one may define

$$\bar{R}(\bar{X}(t), \beta'(t))\beta'(t) := \pi(R(X(t), \beta'(t))\beta'(t))$$

to obtain the curvature operator

$$\bar{R}(-, \beta'(t))\beta'(t): N(\beta(t))/[\beta'(t)]\to N(\beta(t))/[\beta'(t)]$$

with the usual properties. A section $\bar{W} \in \bar{V}_{\text{per}}(\beta)$ can then be defined to be a *Jacobi field* if and only if $\bar{W}$ satisfies the equation

$$\bar{W}'' + \bar{R}(\bar{W}, \beta')\beta' = \bar{0}.$$

It may then be shown (Bölts(4), p. 43), Satz 2·1) that if $\bar{W} = \pi(W) \in \bar{V}_{\text{per}}(\beta)$, where $W \in V_{\text{per}}(\beta)$ is a Jacobi field and $\bar{W}(a) = \bar{W}(b) = \bar{0}$, then there exists a Jacobi field $Z \in V_{\text{per}}(\beta)$ with $Z(a) = Z(b) = 0$ and $\pi(Z) = \bar{W}$.

Thus one may define (Bölts(4), p. 44, definition 2·1)

Definition A 1. For $\beta: [a, b]\to M$ a null geodesic, $t_0 < t_1$ in $[a, b]$ are said to be *conjugate* along β if and only if there exists a Jacobi field $\bar{W} \neq \bar{0}$ in $\bar{V}_{\text{per}}(\beta)$ with $\bar{W}(t_0) = \bar{W}(t_1) = \bar{0}$.

As in Section 7 the index form $I: V_{\text{per}}(\beta) \times V_{\text{per}}(\beta)\to\mathbb{R}$ is given by

$$I(X, Y) = -\int_a^b (\langle X', Y'\rangle - \langle R(X, \beta')\beta', Y\rangle)\, dt.$$

The index form may then be extended to

$$I: \bar{V}_{\text{per}}(\beta) \times \bar{V}_{\text{per}}(\beta)\to\mathbb{R}$$

by setting

$$I(\bar{Z}, \bar{W}) = -\int_a^b (\bar{g}(\bar{Z}', \bar{W}') - \bar{g}(\bar{R}(\bar{Z}, \beta')\beta', \bar{W}))\, dt.$$

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

One then has for any Z, $W \in V_{\text{per}}(\beta)$ with $\pi(Z) = \bar{Z}$ and $\pi(W) = \overline{W}$ that

$$I(\bar{Z}, \overline{W}) = I(Z, W),$$

((4), p. 111). Let $V_{\text{per},0}(\beta) = \{W \in V_{\text{per}}(\beta);\ W(a) = W(b) = 0\}$ and set

$$\overline{V}_{\text{per},0}(\beta) := \pi(V_{\text{per},0}(\beta)).$$

Bölts' checks ((4), pp. 111–112) that to study the definiteness or semi-definiteness of $I: V_{\text{per},0}(\beta) \times V_{\text{per},0}(\beta) \to \mathbb{R}$, it suffices to consider $I: \overline{V}_{\text{per},0}(\beta) \times \overline{V}_{\text{per},0}(\beta) \to \mathbb{R}$. For instance, $I(\overline{W}, \overline{W}) < 0$ for all $\overline{W} \in \overline{V}_{\text{per},0}(\beta)$ if and only if $I(Z, Z) < 0$ for all $Z \in V_{\text{per},0}(\beta)$, $Z \notin [\beta']$.

Bölts then obtains ((4), p. 123; Satz 4·5·5) the following theorem relating null conjugate points to the negative definiteness of the index form

$$I: \overline{V}_{\text{per},0}(\beta) \times \overline{V}_{\text{per},0}(\beta) \to \mathbb{R}.$$

THEOREM A 2. Let $\beta: [a, b] \to M$ be a null geodesic. Then β has no conjugate points in $[a, b]$ if and only if $I(\overline{W}, \overline{W}) < 0$ for all $\overline{W} \in \overline{V}_{\text{per},0}(\beta)$, $\overline{W} \neq \bar{0}$.

Bölts also obtains an analogous theorem ((4), p. 127; Satz 4·5) for null geodesics orthogonal to a spacelike 2-surface in a 4-dimensional space–time.

Since this paper was submitted, the authors have obtained a Morse index theorem for null geodesic segments in arbitrary space–times using the index form given in this appendix. Details will be published elsewhere.

In this work J.K.B. was partially supported by a grant from the Research Council of the Graduate School, University of Missouri-Columbia. P.E.E. was partially supported by a grant from the Research Council of the Graduate School, University of Missouri-Columbia and by the Sonderforschungsbereich Theoretische Mathematik at Bonn University.

REFERENCES

(1) BEEM, J. K. and EHRLICH, P. E. Singularities, incompleteness and the Lorentzian distance function. *Math. Proc. Cambridge Philos. Soc.* **85** (1979), 161–178.

(2) BEEM, J. K. and EHRLICH, P. E. The space–time cut locus. *General Relativity and Gravitation*, to appear.

(3) BEEM, J. K. and WOO, P. Y. Doubly timelike surfaces. *Mem. Amer. Math. Soc.* no. 92 (1969).

(4) BÖLTS, G. *Existenz und Bedeutung von konjugierten Werten in der Raum-Zeit* (Bonner Universität Diplomarbeit, January 1977).

(5) BUSEMANN, H. and BEEM, J. K. Axioms for indefinite metrics. *Rend. Circ. Mat. Palermo* (2) **15** (1966), 223–246.

(6) CHEEGER, J. and EBIN, D. *Comparison theorems in riemannian geometry* (North Holland, Amsterdam, 1975).

(7) EVERSON, J. and TALBOT, C. J. Morse theory on timelike and causal curves. *General Relativity and Gravitation* **7** (1976), 609–622.

(8) FLAHERTY, F. J. Lorentzian manifolds of nonpositive curvature. *A.M.S. Proc. Symp. Pure Math.* **27**, part 2 (Stanford, 1973), 395–399.

(9) GEROCH, R. P. The domain of dependence, *J. Math. Phys.* **11** (1970), 437–449.

(10) GROMOLL, D., KLINGENBERG, W. and MEYER, W. *Riemannsche Geometrie im Grossen. Lect. Notes Math.* no. 55, 1975.

(11) HAWKING, S. W. and ELLIS, G. F. R. *The large scale structure of space–time* (Cambridge University Press, 1973).

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press

(12) HAWKING, S. W. and PENROSE, R. The singularities of gravitational collapse and cosmology. *Proc. Roy. Soc. Lond. Ser. A* **314** (1970), 529–549.
(13) KOBAYASHI, S. On conjugate and cut loci. Studies in Global Geometry and Analysis (*MAA Studies in Math.* **4** (1967)), 96–122.
(14) PENROSE, R. Techniques of differential topology in relativity. *Regional Conference Series in Applied Math.* **7** (SIAM, Philadelphia, 1972).
(15) TIPLER, F. Singularities and causality violation. *Ann. of Phys.* **108** (1977), 1–36.
(16) TIPLER, F. Black holes in closed universes. *Nature* **270** (1977), 500–501.
(17) UHLENBECK, K. A Morse theory for geodesics on a Lorentz manifold, *Topology* **14** (1975), 69–90.
(18) WOODHOUSE, N. M. J. An application of Morse theory to space–time geometry. *Commun. Math. Phys.* **46** (1976), 135–152.

https://doi.org/10.1017/S0305004100056188 Published online by Cambridge University Press