General Relativity and Gravitation, Vol. 11, No. 2, 1979

The Space-Time Cut Locus

JOHN K. BEEM and PAUL E. EHRLICH

Department of Mathematics, University of Missouri, Columbia, Missouri 65211

Received February 27, 1978

Abstract

Let (M, g) be a space-time with Lorentzian distance function d. If (M, g) is distinguishing and d is continuous, then (M, g) is shown to be causally continuous. Furthermore, a strongly causal space-time (M, g) is globally hyperbolic iff the Lorentzian distance is always finite valued for all metrics g' conformal to g. Lorentzian distance may be used to define cut points for space-times and the analogs of a number of results holding for Riemannian cut loci may be established for space-time cut loci. For instance in a globally hyperbolic space-time, any timelike (or respectively, null) cut point q of p along the geodesic c must be either the first conjugate point of p or else there must be at least two maximal timelike (respectively, null) geodesics from p to q. If q is a closest cut point of p in a globally hyperbolic space-time, then either q is conjugate to p or else q is a null cut point. In globally hyperbolic space-times, no point has a farthest nonspacelike cut point.

§(1): *Introduction*

Let $c : [0, \infty) \to M$ be a geodesic in a complete Riemannian manifold starting at $p = c(0)$. Consider the set of all points q on c such that the portion of c from p to q is the unique shortest curve joining p to q. If this set of points has a farthest limit point, this limit point is called the cut point of p along the ray c. The cut locus $C(p)$ of p is then defined to be the set of cut points along all geodesic rays starting at p.

The cut locus has played a key role in modern global Riemannian geometry, notably in connection with Klingenberg's [1, 2] and Berger's [3] improvements of the sphere theorem first proved by Rauch [4] under nonoptimal curvature bounds–cf. Cheeger and Ebin [5], chapters 5-7. An earlier observation of Whitehead [6], crucial to [1, 2], and [3], was that if q is on the cut locus of p, then either q is conjugate to p or else there exist at least two geodesic segments of the same shortest length joining p to q. Klingenberg [1, p. 657] then showed

0001-7701/79/1000-0089$03.00/0

that if q is a closest cut point to p and q is not conjugate to p, there is a geodesic loop at p containing q. Klingenberg used this to obtain an upper bound for the injectivity radius of a positively curved complete Riemannian manifold in terms of a lower bound for the curvature and the length of the shortest smooth closed geodesic on M.

The importance of the cut locus in modern Riemannian geometry naturally suggests investigating the analogous concepts and results for timelike and null geodesics in a space-time. The central role that conjugate points, which are closely related to cut points, have played in singularity theory [7, 8, 9] in general relativity supports this idea. We have already utilized the null cut locus as a tool in [10], Corollary 5.3, to show that if (M, g) is a Friedmann cosmological model ($\Lambda = 0$), then there is a C^2-open neighborhood $U(g)$ of g in the space of Lorentzian metrics for M conformal to g such that *every* null geodesic in (M, g') is incomplete for *all* metrics $g' \in U(g)$. As usual the geometry of the space-time cut locus is more complicated than that of the Riemannian cut locus. First, physically realistic space-times are usually assumed to be noncompact. Second, there is no result for space-times as strong as the Hopf-Rinow theorem for Riemannian manifolds linking metric to geodesic completeness. Indeed, for space-times, null, timelike, and spacelike geodesic completeness are all known to be inequivalent—compare [11, 12, 13].

Since the causal structure of a space-time is so important in general relativity and since the timelike and null cut loci are defined using the Lorentzian distance function, it is natural to begin by considering how the standard causal structures for space-times, recalled in Section 2, are related to Lorentzian distance. This is done in Section 3, where we in particular provide proofs of two results we previously announced [14]. First if (M, g) is a distinguishing space-time with a continuous Lorentzian distance function, then (M, g) is causally continuous. Second, we have defined in [14] that a space-time (M, g) satisfies the *finite distance condition* iff $d(p, q) < \infty$ for all $p, q \in M$. We show that globally hyperbolic space-times may be characterized among strongly causal space-times as follows. A strongly causal space-time (M, g) is globally hyperbolic iff (M, g') satisfies the finite distance condition for all Lorentzian metrics g' for M conformal to g.

In Sections 4 and 5 we consider timelike and null cut points, respectively. By duality, it is enough to consider future cut points. Since it is possible to use unit timelike tangent vectors but not unit null vectors, it is convenient to treat the timelike and null cut loci separately. In addition, there are intrinsic differences between null and timelike cut points. For example, null cut points are invariant under conformal changes of metric, but timelike cut points are not.

Given a tangent vector $v \in TM$, let c_v denote the unique geodesic with initial condition $c_v'(0) = v$ and let $T_{-1}M$ denote the future observer bundle consisting of all future pointing tangent vectors $v \in TM$ with $g(v, v) = -1$. To handle future timelike cut points, we may define as in the Riemannian case the function

$s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ by $s(v) = \sup\{t \geq 0; d(c_v(0), c_v(t)) = t\}$. If $0 < s(v) < \infty$ and $c_v(s(v))$ exists, then we call $c_v(s(v))$ the *future timelike cut point of* $c_v(0)$ *along* c_v. For a complete Riemannian manifold, it is known that the analogous function is continuous on the unit sphere bundle. Here we show that s is upper semicontinuous on the future observer bundle if (M, g) is timelike geodesically complete and that s is lower semicontinuous if (M, g) is globally hyperbolic. Examples may be given of globally hyperbolic spaces-times for which s fails to be continuous. In a subsequent paper, the continuity of s will be used to show that if (M, g) is globally hyperbolic and timelike geodesically complete, then for any $p \in M$ the future set $J^+(p)$ is the disjoint union of an open cell and a closed subset of M.

We conclude Section 4 by establishing the validity of Whitehead's result referred to above for future timelike cut points in globally hyperbolic space-times. That is, if q is the future timelike cut point of p along c, then either one or both of the following holds: (1) q is the first future conjugate point of p on c, or (2) there exist at least two maximal timelike geodesic segments from p to q.

Null cut points are defined in Section 5 as follows. Let c be a future-directed null geodesic with $p = c(0)$. Set $t_0 = \sup\{t \geq 0; d(p, c(t)) = 0\}$. If $0 < t_0 < \infty$ and $c(t_0)$ exists, then $c(t_0)$ is called the *future null cut point of p along c*. We show that if (M, g) is globally hyperbolic, then Whitehead's result applies to null cut points also. We prove a theorem in Section 5 corresponding to Klingenberg's result [1, p. 657] about the closest cut point mentioned above. The space-time analog has a new non-Riemannian flavor, however. We prove here that if (M, g) is a globally hyperbolic space-time, then the closest nonspacelike cut point to p is either conjugate to p or is a *null* cut point of p. We also establish that if (M, g) is globally hyperbolic, then there is no "farthest" future (or past) nonspacelike cut point to p for any $p \in M$.

§(2): *Preliminaries*

Let (M, g) be a connected space-time. This means M is a smooth n-dimensional manifold with a countable basis, a smooth Lorentz metric g of signature $(-, +, \ldots, +)$ and a time orientation. Let π: $TM \to M$ denote the tangent bundle of M. A nonzero tangent vector $v \in TM$ is said to be *timelike* (or, respectively, *nonspacelike*, *null*, *spacelike*) according to whether $g(v, v) < 0$ (respectively, ≤ 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 $p \leq q$ 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 $J^+(p) = \{q; p \leq q\}$. The sets $I^-(p)$ and $J^-(p)$ are defined dually.

If $I^+(p) \cap I^-(p) = M$ for all p, then (M, g) is said to be *totally vicious* [15, 16]. A space-time is said to be chronological if $p \notin I^+(p)$ for all p. It is *future* (respectively, *past*) *distinguishing* if $I^+(p) = I^+(q)$ [respectively, $I^-(p) = I^-(q)$] implies $p = q$. A *distinguishing* space-time is one that is both future and past distinguishing. A space-time is *strongly causal* if each point has arbitrarily small neighborhoods such that no nonspacelike curve that leaves one of these neighborhoods ever returns. If a strongly causal space-time has $J^+(p) \cap J^-(q)$ compact for all $p, q \in M$, then it is *globally hyperbolic*.

Consider a set valued function F on M which assigns an open subset of M to each $p \in M$. The function F is *inner continuous* if for each $p \in M$ and each compact subset $K \subset F(p)$ there is an open neighborhood U of p such that $K \subset F(q)$ for all $q \in U$. It is outer continuous if for each $p \in M$ and compact set $K \subset M - \overline{F(p)}$ there is some open neighborhood U of p such that $K \subset M - \overline{F(q)}$ for all $q \in U$.

The set valued functions $I^+(p)$ and $I^-(p)$ are inner continuous for all space-times—compare [17, p. 291]. A distinguishing space-time such that $I^+(p)$ and $I^-(p)$ are outer continuous is said to be *causally continuous* [17].

The collection of all smooth Lorentz metrics on M will be denoted by Lor (M). In this paper we will be interested in the C^r topologies on Lor (M) for $r = 0$ and $r = 2$. Let $\{B_i\}$ be a fixed countable collection of coordinate neighborhoods covering M such that any compact subset of M meets only finitely many of these neighborhoods. Let $\delta : M \to (0, \infty)$ be a continuous function. Then $|g_2 - g_1|_r < \delta$ on M if for each $p \in M$, all of the corresponding coefficients and derivatives up to order r of the two metric tensors are $\delta(p)$ close at p when calculated in the fixed coordinates of each of the corresponding coordinate charts B_i containing p. The sets $\{g_2 \in \text{Lor}(M); |g_2 - g_1|_r < \delta\}$ with $g_1 \in \text{Lor}(M)$ and $\delta : M \to (0, \infty)$ arbitrary form a basis for the C^r topology on Lor (M). If $r < s$, then the C^s topology on Lor (M) is finer than the C^r topology.

Stable causality may be defined in terms of the C^0 topology. The space-time (M, g) is *stably causal* if g has a C^0 neighborhood in Lor (M) such that all metrics in this neighborhood are chronological.

Two metrics g and g' in Lor (M) are *conformal* if there is a smooth function $\Omega : M \to (0, \infty)$ such that $g' = \Omega^2 g$. This is an equivalance relation on Lor (M) and the equivalence class $C(M, g)$ consists of all metrics on M conformal to g. Since $C(M, g) \subset \text{Lor}(M)$, the C^r topologies on Lor (M) induce C^r topologies on $C(M, g)$. If $g' \in C(M, g)$, then a curve γ is timelike (respectively, nonspacelike, null, spacelike) for (M, g') iff it is timelike (respectively, nonspacelike, null, spacelike) for (M, g). Thus conformal metrics induce the same causal structure of M.

Fix a point $p \in M$ and let $\exp_p : T_pM \to M$ 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—see [18, p. 146]. 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 known [18, p. 146] that $v \in T_pM$ is a conjugate point iff there is a nontrivial Jacobi field along $\exp_p(tv)$ that vanishes at p and $\exp_p v$.

A geodesic $c_v(t) = \exp_p(tv)$ is *complete* if $c_v(t)$ is defined for all positive and negative values of the affine parameter t. For timelike and spacelike geodesics this means the geodesic has infinite length in both directions. A space-time is *timelike* (respectively, *nonspacelike*, *null*, *spacelike*) *geodesically complete* if all timelike (respectively, nonspacelike, null, spacelike) geodesics can be extended to complete geodesics—compare [8, 12].

A *time function* $h: M \to \mathbb{R}$ is a continuous function which is increasing along all future directed nonspacelike curves. If (M, g) is globally hyperbolic and h is a time function with $h^{-1}(t)$ a Cauchy surface for each t, then we will call h a *globally hyperbolic time function*.

§(3): *Lorentzian Distance and Causality*

In this section we will indicate how some of the standard concepts of causality theory in general relativity recalled in Section 2 may be formulated in terms of the Lorentzian distance function. Given $g \in \text{Lor}(M)$, the *Lorentzian distance function* $d = d(g): M \times M \to \mathbb{R} \cup \{\infty\}$ is defined as follows: Given $p \in M$, set $d(p, q) = 0$ if $q \in M - J^+(p)$ 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—compare [8]. It is known that if p and q can be joined by a nonspacelike curve that is not a geodesic, then they can be joined by a timelike curve [8, p. 112]. Furthermore, $d(p, q) > 0$ [respectively, $d(q, p) > 0$] iff $q \in I^+(p)$ [respectively, $q \in I^-(p)$]. This means that the distance function determines the chronology of (M, g). On the other hand, the chronology does not determine the distance function because conformal metrics give the same chronology but in general different Lorentzian distances.

Unlike the distance function of a Riemannian manifold, however, the Lorentzian distance function is by construction not symmetric in general. Indeed if $0 < d(p, q) = d(q, p) \leqslant \infty$, then $p \in I^+(q)$ and $q \in I^+(p)$, which yield $d(p, q) = d(q, p) = \infty$. Thus $d = d(g)$ is nonsymmetric where it is positive and finite valued. The distance function satisfies the reverse triangle inequality $d(p, q) \geqslant d(p, r) + d(r, q)$ whenever $p \leqslant r \leqslant q$. Also the distance function is lower semicontinuous where it is finite [8, p. 215]. In addition, if $p_n \to p$, $q_n \to q$, and $d(p, q) = \infty$, then $d(p_n, q_n) \to \infty$. For globally hyperbolic space-times, this function is continuous and finite valued [8, p. 215]. In more general space-times, however, the distance function may fail to be upper semicontinuous and thus fail to be continuous.

It is not difficult to establish the following proposition.

Proposition 3.1. (a) The space-time (M, g) is totally vicious iff $d(p, q) = \infty$ for all $p, q \in M$. (b) The space-time (M, g) is chronological iff d is identically zero on the diagonal $\Delta(M) = \{(p, p); p \in M\}$ of $M \times M$. (c) The space-time (M, g) is future (respectively, past) distinguishing iff for each pair of distinct p, $q \in M$, there is some $x \in M$ such that exactly one of $d(p, x)$ and $d(q, x)$ [respectively, $d(x, p)$ and $d(x, q)$] is zero. (d) The space-time (M, g) is stably causal iff there exists a neighborhood U of g in the C^0 topology on Lor (M) such that $d(g')(p, p) = 0$ for all $g' \in U$ and $p \in M$.

Recall that space-time distance in general fails to be upper semicontinuous. Thus the continuity of d should have implications for the causal structure of M. An example is the following result, which was stated but not proved in Beem and Ehrlich [14]. Here d is regarded as being continuous at $(p, q) \in M \times M$ if $d(p, q) = \infty$ because $d(p_n, q_n) \to \infty$ for all sequences $p_n \to p$ and $q_n \to q$.

Theorem 3.2. Let (M, g) be a distinguishing space-time. If $d: M \times M \to \mathbb{R} \cup \{\infty\}$ is continuous, then M is causally continuous.

Proof. We need only show that I^+ and I^- are outer continuous. Assume that I^+ is not outer continuous. There is then some compact set $K \subset M - \overline{I^+(p)}$ and some sequence $p_n \to p$ such that $K \cap I^+(p_n) \neq \phi$ for all n. Let $q_n \in K \cap I^+(p_n)$ and let $\{q_m\}$ be a subsequence of $\{q_n\}$ such that $\{q_m\}$ converges to some point q of the compact set K. Since $M - \overline{I^+(p)}$ is an open neighborhood of q, there is some $r \in M - I^+(p)$ with $q \ll r$. For sufficiently large m we then have $q_m \ll r$ and hence $p_m \ll q_m \ll r$. Thus $d(p_m, r) \geqslant d(p_m, q_m) + d(q_m, r)$. Using lower semicontinuity of distance and the causality relation $q \ll r$, we obtain $0 < d(q, r) \leqslant \liminf d(q_m, r)$. Consequently, $d(p_m, r) \geqslant d(q, r)/2 > 0$ for all sufficiently large m. However, since $r \notin I^+(p)$ we have $d(p, r) = 0$ and hence $d(p, r) \neq \lim d(p_m, r)$. Thus if d is continuous, then I^+ is outer continuous. A similar argument shows that I^- is outer continuous. Thus continuity of d implies that (M, g) is causally continuous. □

Remark 3.3. Examples may be constructed of causally continuous space-times that have discontinuous Lorentzian distance functions.

For the proof of Theorem 3.5 it is necessary to show that the usual definition of globally hyperbolic may be weakened.

Lemma 3.4. Let (M, g) be a strongly causal space-time. If $J^+(p) \cap J^-(q)$ has compact closure for all $p, q \in M$, then (M, g) is globally hyperbolic.

Proof: It is only necessary to show $J^+(p) \cap J^-(q)$ is always closed. Assume $r \in \mathrm{cl}\,(J^+(p) \cap J^-(q)) - J^+(p) \cap J^-(q)$. Choose a sequence $\{r_n\}$ of points in $J^+(p) \cap J^-(q)$ with $r_n \to r$. For each n let $\gamma_n: [0, 1) \to M$ be a future-directed, future-inextendible nonspacelike curve with $p = \gamma_n(0)$ and $q, r_n \in \gamma_n$. By

Proposition 6.2.1 of [8, p. 185], there is some future-directed, future-inextendible nonspacelike limit curve $\gamma:[0, 1) \to M$ of the sequence $\{\gamma_n\}$. Furthermore, $p = \gamma(0)$. The limit curve γ cannot be future imprisoned in any compact subset of M because (M, g) is strongly causal [8, p. 195]. Consequently, there is some point x on γ with $x \notin \mathrm{cl}\,(J^+(p) \cap J^-(q))$. The definition of limit curve yields a subsequence $\{\gamma_m\}$ of $\{\gamma_n\}$ and points $x_m \in \gamma_m$ with $x_m \to x$. Since $x \notin \mathrm{cl}\,(J^+(p) \cap J^-(q))$, we have $x_m \notin J^+(p) \cap J^-(q)$ for all large m. Using $\gamma_m \subset J^+(p)$, it follows that $x_m \notin J^-(q)$ for large m. Hence q lies between p and x_m on γ_m for large m. Let $\gamma[p, x]$ (respectively, $\gamma_m[p, x_m]$) denote the portion of γ (respectively, γ_m) from p to x (respectively, x_m). By Lemma 2.1 of [10] we may assume, by taking a subsequence of $\{\gamma_m[p, x_m]\}$ if necessary, that $\{\gamma_m[p, x_m]\}$ converges to $\gamma[p, x]$ in the C^0 topology on curves. Hence $q \in \gamma_m[p, x_m]$ for large m implies $q \in \gamma[p, x]$. Also $r_m \to r$ and $r_m \leqslant q$, which yield $r \in \gamma[p, q]$. Thus $r \in J^+(p) \cap J^-(q)$ in contradiction. □

The following metric condition on (M, g) was defined in [14] and utilized in [10] proving singularity theorems.

Finite Distance Condition. (M, g) satisfies the finite distance condition iff $d(g)(p, q) < \infty$ for all $p, q \in M$.

This condition may be used to characterize globally hyperbolic space-times among strongly causal space-times.

Theorem 3.5. The strongly causal space-time (M, g) is globally hyperbolic iff (M, g') satisfies the finite distance condition for all $g' \in C(M, g)$.

Proof. Assume (M, g) is globally hyperbolic. Since global hyperbolicity is a conformally invariant property, all metrics of $C(M, g)$ are globally hyperbolic. This implies [8, p. 215] that $d(g')$ is finite valued for all $g' \in C(M, g)$.

Conversely, assume that (M, g) is not globally hyperbolic. Lemma 3.4 implies that there exist $p, q \in M$ such that $J^+(p) \cap J^-(q)$ does not have compact closure. Let h be an auxiliary geodesically complete positive definite metric on M—compare [19]. Let $d_0: M \times M \to \mathbb{R}$ be the Riemannian distance function induced on M by h. The Hopf-Rinow theorem [18, p. 163] implies that all subsets of M that are bounded with respect to d_0 have compact closure. Thus $J^+(p) \cap J^-(q)$ is not bounded. Hence we may choose $p_n \in J^+(p) \cap J^-(q)$ such that $d_0(p, p_n) > n$ for each n. Choose p' and q' with $p' \ll p \ll q \ll q'$. We wish to show there exists a conformal factor Ω such that $d(\Omega^2 g)(p', q') = \infty$. For each $n > 1$ choose γ_n to be a future-directed timelike curve from p' to p_n such that $\gamma_n \mid [1/2, 3/4] \subset \{r \in M; n - 1 < d_0(p, r) < n\}$. For each $n > 1$ let $\Omega_n: M \to \mathbb{R}$ be a smooth conformal factor such that $\Omega_n(x) = 1$ if $x \notin \{r; n - 1 < d_0(p, r) < n\}$ and such that the length of $\gamma_n \mid [1/2, 3/4]$ is greater than n for the metric $\Omega_n^2 g$. Let $\Omega = \Pi\, \Omega_n$. This infinite product is well defined on M since for each $x \in M$ at most one of the factors Ω_n is not unity. Then $d(\Omega^2 g)(p', p_n) > n$ for each $n > 1$. Since $d(\Omega^2 g)(p', q') > d(\Omega^2 g)(p', p_n)$ for all n, it follows that $d(\Omega^2 g)(p', q') = \infty$. □

§(4): *The Timelike Cut Locus*

We first recall a definition from Beem and Ehrlich [10].

Definition 4.1. A future-directed nonspacelike curve c from p to q is said to be *maximal* if $d(p, q) = L(c)$, where $L(c)$ denotes the Lorentzian arc length of c defined as in [8, p. 105].

The importance of this definition stems from the following lemma.

Lemma 4.2. ([10], Corollary 3.4). If the future-directed causal curve c is maximal, then c is a geodesic.

We also recall the following analog of a classical result from Riemannian geometry. It may be shown along the lines of p. 99 of Kobayashi [20], using the fact that if $p \ll q$ and p and q are contained in a convex normal neighborhood, then p and q may be joined by a maximal timelike geodesic segment which lies in this neighborhood.

Lemma 4.3. Let $c: [0, a] \to M$ be a maximal timelike geodesic segment. Then for any s, t with $0 \leqslant s < t < a$, the curve $c \mid [s, t]$ is the *unique* maximal geodesic segment (up to parametrization) from $c(s)$ to $c(t)$.

Before commencing our study of the timelike cut locus, we need to define the unit future observer bundle $T_{-1}M$—cf. Thorpe [21]; also [22].

Definition 4.4. Let $T_{-1}M = \{v \in TM; g(v, v) = -1 \text{ and } v \text{ is future directed}\}$. Given $p \in M$, let $T_{-1}M_{|p}$ denote the fiber of $T_{-1}M$ at p. Also given $v \in T_{-1}M$, let c_v denote the unique timelike geodesic with $c_v'(0) = v$.

We may now define a function that measures the distance along a timelike geodesic to the future cut point.

Definition 4.5. Define the function $s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ by $s(v) = \sup \{t \geqslant 0; d(\pi(v), c_v(t)) = t\}$.

We may first note that if $d(p, p) = \infty$, then $s(v) = 0$ for all $v \in T_{-1}M$ with $\pi(v) = p$. Also $s(v) > 0$ for all $v \in T_{-1}M$ if (M, g) is strongly causal. The number $s(v)$ may be interpreted as the "largest" parameter value t such that c_v is a maximal geodesic between $c_v(0)$ and $c_v(t)$. Indeed from Lemma 4.3 we know the following.

Corollary 4.6. For $0 < t < s(v)$, the geodesic $c_v : [0, t] \to M$ is the *unique* maximal timelike curve (up to reparametrization) from $c_v(0)$ to $c_v(t)$.

The function s fails to be upper semicontinuous for general space-times. For timelike geodesically complete space-times we have the following proposition.

Proposition 4.7. If (M, g) is timelike geodesically complete, then $s : T_{-1}M \to \mathbb{R} \cup \{\infty\}$ is upper semicontinuous.

Proof. It suffices to show the following. Let $v_n \to v$ in $T_{-1}M$ with $\{s(v_n)\}$ converging in $\mathbb{R} \cup \{\infty\}$. Then $s(v) \geqslant \lim s(v_n)$. If $s(v) = \infty$, there is nothing to prove. Hence we assume that $s(v) < \lim s(v_n) = A \leqslant \infty$ and derive a contradiction.

We may choose $\delta > 0$ such that $s(v) + \delta < A$ and assume that $s(v_n) \geqslant s(v) + \delta = b$ for all n. Let $c_n = c_{v_n}$. Since $b \leqslant s(v_n)$, we have $d\,(\pi(v_n), c_n(b)) = b$ for all n. Since $v_n \to v$, we have by lower semicontinuity of distance $d\,(\pi(v), c_v(b)) \leqslant \liminf d(\pi(v_n), c_n(b)) = b$. Thus $d(\pi(v), c_v(b)) \leqslant b = L(c_v \mid [0, b])$—this last equality follows from the definition of arc length. On the other hand, $d\,(\pi(v), c_v(b)) \geqslant L(c_v \mid [0, b])$ so that $d\,(\pi(v), c_v(b)) = L(c_v \mid [0, b]) = b$. Hence $s(v) \geqslant b = s(v) + \delta$, in contradiction. □

In order to prove the lower semicontinuity of s for globally hyperbolic space-times, it will first be useful to establish the following lemma.

Lemma 4.8. Let (M, g) be a globally hyperbolic space-time and let $\{p_n\}$ and $\{q_n\}$ be two infinite sequences of points with $p_n \to p$ and $q_n \to q$, where $p \ll q$. Assume $c_n : [0, d(p_n, q_n)] \to M$ is a unit speed maximal geodesic segment from p_n to q_n and set $v_n = c_n'(0) \in T_{-1}M$. Then the sequence $\{v_n\}$ has a timelike limit vector $w \in T_{-1}M$. Moreover, $c_w : [0, d(p, q)] \to M$ is a maximal geodesic segment from p to q.

Proof. Choose $p_0 \in I^-(p)$ and let $h : M \to \mathbb{R}$ be a globally hyperbolic time function for M. If $t = h(q)$, then $K = h^{-1}(-\infty, t + 1] \cap J^+(p_0)$ is compact [8, p. 210] and $p, q \in \mathrm{Int}\,(K)$. Thus since $p_n \to p$ and $q_n \to q$, we may assume $\{p_n\}$, $\{q_n\} \subset \mathrm{Int}\,(K)$. We then have $c_n\,[0, d\,(p_n, q_n)] \subset \mathrm{Int}\,(K)$ for all n. There is then a nonspacelike future-directed limit curve λ of $\{c_n\}$ with $\lambda(0) = p$—compare [8, p. 185]. Since K is a compact subset of a globally hyperbolic space-time and $q_n \to q$, λ must pass through q. Let α be the first parameter value, $\alpha > 0$, for which $\lambda(\alpha) = q$ and set $c = \lambda \mid [0, \alpha]$. By Lemma 2.3 of [10], we have

$L(c) \geq \limsup L(c_n) = \lim d(p_n, q_n) = d(p, q) > 0$. Thus as $d(p, q) \geq L(c)$, it follows that $L(c) = d(p, q) > 0$. Hence, by Corollary 3.4 of [10], the curve c is a maximal timelike geodesic segment from p to q. Finally, $w = c'(0)/(-g(c'(0), c'(0))^{1/2}$ is the required tangent vector. □

We are now ready to prove the following proposition.

Proposition 4.9. If (M, g) is globally hyperbolic, then $s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ is lower semicontinuous.

Proof. It suffices to prove that if $v_n \to v$ in $T_{-1}M$ and $s(v_n) \to A$ in $\mathbb{R} \cup \{\infty\}$, then $s(v) \leq A$. If $A = \infty$, there is nothing to prove. Thus suppose $A < \infty$. Assuming $s(v) > A$, we will derive a contradiction.

Choose $\delta > 0$ such that $A + \delta < s(v)$. Define $b_n = s(v_n) + \delta$ and let N_0 be such that $b_n < s(v)$ for all $n \geq N_0$. Put $c_n = c_{v_n}, p_n = c_n(0)$ and $q = c_v(A + \delta)$. Since $v_n \to v$ and c_v is defined for some parameter values beyond $A + \delta$, the geodesics c_n must be defined for some parameter values past b_n whenever n is larger than some $N \geq N_0$. Let $q_n = c_n(b_n)$ for $n \geq N$. Now $c_n \mid [0, b_n]$ cannot be maximal since $b_n > s(v_n)$. Because M is globally hyperbolic and $c_n(0) \ll c_n(b_n)$, we may find maximal unit speed timelike geodesic segments $\gamma_n : [0, d(p_n, q_n)] \to M$ from p_n to q_n. Set $w_n = \gamma_n'(0)$. Since $c_v \mid [0, s(v))$ is a maximal geodesic and thus has no conjugate points, it is impossible for v to be a limit direction of $\{w_n; n \geq N\}$. Thus the maximal geodesic c_w joining p to q given by Lemma 4.8 applied to $\{w_n\}$ is different from c_v. This then implies that $s(v) \leq A + \delta$, which contradicts $A + \delta < s(v)$.

Remark 4.10. Examples of strongly causal but not globally hyperbolic space-times for which s fails to be lower semicontinuous may be constructed. Also globally hyperbolic examples may be found for which s is not upper semicontinuous.

Combining Propositions 4.7 and 4.9 we obtain the following theorem:

Theorem 4.11. If (M, g) is globally hyperbolic and timelike geodesically complete, then $s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ is continuous.

We are now ready for the following definition:

Definition 4.12. The *future timelike cut locus* $\Gamma^+(p)$ in T_pM is defined to be $\Gamma^+(p) = \{s(v)\,v; v \in T_{-1}M_{|p}$ and $0 < s(v) < \infty\}$. The *future timelike cut locus* $C_t^+(p)$ *of* p *in* M is defined to be $C_t^+ = \exp_p [\Gamma^+(p)]$. If $0 < s(v) < \infty$ and $c_v(s(v))$ exists, then the point $c_v(s(v))$ is called the *future cut point* of $p = c_v(0)$ *along* c_v. The *past timelike cut locus* $C_t^-(p)$ and *past cut points* may be defined dually.

We may then interpret $s(v)$ as measuring the distance from p up to the future cut point along c_v. Thus Theorem 4.11 implies that for globally hyperbolic, timelike geodesically complete space-times, the distance from a fixed $p \in M$ to its future cut point in the direction $v \in T_{-1}M|_p$ is a continuous function of v.

We now give Lorentzian analogs of two well-known results relating cut and conjugate points on complete Riemannian manifolds. The following theorem is well known {cf. Hawking and Ellis [8, pp. 111-116] or Lerner [23], Theorem 4(b)}:

Theorem 4.13. A timelike geodesic is not maximal beyond the first conjugate point.

In the language of Definition 4.12 this may be restated as the following corollary.

Corollary 4.14. The future cut point of $p = c_v(0)$ along c_v comes no later than the first future conjugate point of p along c_v.

Utilizing this fact, we may now prove the second basic result on cut and conjugate points.

Theorem 4.15. Let (M, g) be globally hyperbolic. If $q = c(t)$ is the future cut point of $p = c(0)$ along the timelike geodesic c from p to q, then either one or possibly both of the following holds: (1) q is the first future conjugate point of p along c; (2) there exist at least two maximal timelike geodesic segments from p to q.

Proof. Without loss of generality we may suppose that $c = c_v$ for some $v \in T_{-1}M$ and thus that $t = d(p, q) = s(v)$. Let $\{t_n\}$ be a monotone decreasing sequence of real numbers converging to t. Since $c(t) \in M$, the points $c(t_n)$ exist for n sufficiently large. By global hyperbolicity, we may join $c(0)$ to $c(t_n)$ by a maximal timelike geodesic $c_n = c_{v_n}$ with $v_n \in T_{-1}M|_p$. Since $t_n > t = s(v)$, we have $v \neq v_n$ for all n. Let $w \in T_{-1}M$ be the timelike limiting vector for $\{v_n\}$ given by Lemma 4.8. If $v \neq w$, then c and c_w are two maximal timelike geodesic segments from p to q.

It remains to show that if $v = w$, then q is the first future conjugate point of p along c. If $v = w$, then there is a subsequence $\{v_m\}$ of $\{v_n\}$ with $v_m \to v$. If v were not a conjugate point, there would be a neighborhood U of v in $T_{-1}M|_p$ such that $\exp_p : U \to M$ is injective. On the other hand, since c_n and $c \mid [0, t_n]$ join $c(0)$ to $c(t_n)$ and $v_m \to v$, no such neighborhood U can exist. Thus q is a future conjugate point of p along c. □

Theorem 4.15 has the immediate implication that for globally hyperbolic space-times, $q \in C_t^+(p)$ iff $p \in C_t^-(q)$.

The timelike cut locus of a timelike geodesically complete, globally hyperbolic space-time has the following structural property which refines Theorem 4.15. We know from this theorem that if $q \in C_t^+(p)$ and q is not conjugate to p, then there exist at least two maximal geodesic segments from p to q. Accordingly, it makes sense to consider the set

$$\mathrm{Seg}\,(p) = \{q \in C_t^+(p);\ \text{there exist at least two future directed maximal geodesic segments from } p \text{ to } q\}$$

Since $s: T_{-1}M \to \mathbb{R} \cup \{\infty\}$ is continuous by Theorem 4.11 and maximal geodesics joining any pair of causally related points exist in globally hyperbolic space-times, it may be shown that Seg (p) is dense in $C_t^+(p)$ for all $p \in M$. The proof may be given along the lines of Wolter's proof of Lemma 2 (cf. [26, p. 93]) for complete Riemannian manifolds. The dual result also holds for the past timelike cut locus $C_t^-(p)$.

§(5): *The Null Cut Locus*

We recall the definition of null cut point from [10], Section 5, where the concept was used to prove null geodesic incompleteness for certain classes of space-times. Let $\gamma: [0, a) \to M$ be a future-directed null geodesic with endpoint $p = \gamma(0)$. Set $t_0 = \sup\ \{t \in [0, a); d\,(p, \gamma(t)) = 0\}$. If $0 < t_0 < a$, we will say $\gamma(t_0)$ is the *future null cut point* of p on γ. Past null cut points are defined dually. Let $C_N^+(p)$ [respectively, $C_N^+(p)$] denote the *future* (respectively, past) *null cut locus* of p consisting of all future (respectively, past) null cut points of p. The definition of $C_N^+(p)$ together with the lower semicontinuity of distance yields $d\,(p, q) = 0$ for all $q \in C_N^+(p)$. We then define the future nonspacelike cut locus to be $C^+(p) = C_t^+(p) \cup C_N^+(p)$. The past nonspacelike cut locus is defined dually.

The geometric significance of null cut points is similar to that of timelike cut points. The geodesic γ is maximizing from p up to and including the cut point $\gamma(t_0)$, i.e., $L\,(\gamma \mid [0, t]) = d\,(p, \gamma(t)) = 0$ for all $t \leqslant t_0$. Thus there is no timelike curve joining p to $\gamma(t)$ for any t with $t \leqslant t_0$. In contrast, the geodesic γ is no longer maximizing beyond the cut point $\gamma(t_0)$. In fact each point $\gamma(t)$ for $t_0 < t < a$ may be joined to p by a timelike curve.

Utilizing Proposition 2.19 of Penrose [24, p. 15] and the definition of maximality, the following can easily be seen.

Lemma 5.1. Let (M, g) be a causal space-time. If there are two null geodesic segments from p to q, then q comes on or after the null cut point of p on each of the two segments.

The cylinder $S^1 \times \mathbb{R}$ with the metric $ds^2 = d\theta dt$ shows that causality is needed in Lemma 5.1.

We now prove the null analog of Lemma 4.8.

Lemma 5.2. Let (M, g) be globally hyperbolic and let $c:[0, t] \to M$ be a future-directed null geodesic from $p = c(0)$ to $q = c(t)$ with $d(p, q) = 0$. Assume $p_n \to p, q_n \to q$ and $p_n \leqslant q_n$. Let c_n be a maximal geodesic joining p_n to q_n with initial direction v_n. Then the set of directions $\{v_n\}$ has a limit direction w and c_w is a maximal null geodesic from p to q.

Proof. Choose p' and q' with $p' \ll p \leqslant q \ll q'$. Then for n sufficiently large, the segments c_n all lie in the compact set $J^+(p') \cap J^-(q')$. Using Lemma 6.2.1 of Hawking and Ellis [8, p. 185], one may show there is a future-directed limit curve λ from p to q. Since $d(p, q) = 0$, λ must satisfy $L(\lambda) = 0$. It follows that λ is a maximal null geodesic. □

We may now obtain the null analog of Theorem 4.15.

Theorem 5.3. Let (M, g) be globally hyperbolic and let $q = c(t)$ be the future null cut point of $p = c(0)$ along the null geodesic c. Then either one or possibly both of the following holds: (1) q is the first future conjugate point of p along c; (2) there exist at least two maximal null geodesic segments from p to q.

Proof. Let $v = c'(0)$ and let t_n be a monotone decreasing sequence of real numbers with $t_n \to t$. Since $q \in M$, we know that $c(t_n)$ exists for all sufficiently large n. Since (M, g) is globally hyperbolic, we may find maximal nonspacelike geodesics c_n with initial directions v_n joining p to $c(t_n)$. By Lemma 5.2, the set of directions $\{v_n\}$ has a limit direction w. If $v \neq w$, then the geodesic c_w is a second maximal null geodesic joining p to q as $d(p, q) = 0$. If $v = w$, then q is conjugate to p along c. Since $d(p, q) = 0$, q must be the first conjugate point to p along c (cf. Lerner [23, p. 39]). □

We now give the Lorentzian analog of an important result of Klingenberg, compare [1, p. 657], [20, p. 118], [25, p. 244].

Theorem 5.4. Let (M, g) be a globally hyperbolic space-time and assume that $p \in M$ has a closest future (or past) cut point q. Then q is either conjugate to p or else q is a *null* cut point of p.

Proof. Let q be a future cut point of p which is a closest cut point of p with respect to the Lorentzian distance d. Assume q is neither conjugate to p nor a null cut point of p. Then $p \ll q$ and by Theorem 4.15, there exist at least two future-directed maximal timelike geodesics c_1 and c_2 from p to q. Let $\gamma:[0, a) \to M$ be a past-directed timelike curve starting at q. By choosing $a > 0$ sufficiently small we may assume the image of γ lies in the chronological future of p. Then $p \ll \gamma(t) \ll q$ for $0 < t < a$ implies $d(p, q) \geqslant d(p, \gamma(t)) + d(\gamma(t), q) > d(p, \gamma(t))$ using the reverse triangle inequality. Since q is a closest cut point, the point $\gamma(t)$ comes before a cut point of p on any timelike geodesic from p to $\gamma(t)$. Thus any timelike geodesic from p to $\gamma(t)$ is maximal. Since q is not conjugate to p along c_1 there is a timelike geodesic from p to $\gamma(t)$ near c_1

for all sufficiently small t. Similarly there exists a timelike geodesic from p to $\gamma(t)$ near c_2 for all small t. The existence of two maximal timelike geodesics from p to $\gamma(t)$ implies $\gamma(t)$ is a cut point of p and yields a contradiction since $d(p, \gamma(t)) < d(p, q)$. □

We now prove a nonexistence theorem for farthest cut points in a globally hyperbolic space-time.

Theorem 5.5. Let (M, g) be a globally hyperbolic space-time and let $p \in M$ be any point of M. Then p has no farthest nonspacelike cut point.

Proof. Assume $q \in M$ is a farthest cut point of p. Then q is a cut point along a maximal geodesic segment γ from p to q. Choose a sequence of points $\{q_n\}$ such that $q \ll q_n$ for each n and $q_n \to q$. Since (M, g) is globally hyperbolic, there exist maximal timelike geodesic segments $c_n: [0, d(p, q_n)] \to M$ from p to q_n for each n. Extend each c_n to a future-inextendible geodesic ray. Since q is a farthest cut point of p and $d(p, q_n) \geqslant d(p, q) + d(q, q_n) > d(p, q)$, for each n the geodesic ray c_n contains no cut point of p. The sequence $\{c_n\}$ has a limit curve c that is a future directed and future inextendible nonspacelike curve starting at p—compare [8, p. 185]. By passing to a subsequence if necessary we may assume that $\{c_n\}$ converges to c in the C^0 topology on curves—compare Lemma 2.1 of [10]. Using the global hyperbolicity of (M, g) and $q_n \to q$, we find that $q \in c$. If $r \in c$ and $r_n \in c_n$ with $r_n \to r$, then $d(p, r_n)$ is the length of the curve c_n from p to r_n and $d(p, r_n) \to d(p, r)$. Using the upper semicontinuity of arc length [24, p. 54] for strongly causal space-times, we find that the length of c from p to r is at least as great as $\limsup d(p, r_n) = \lim d(p, r_n) = d(p, r)$. Thus c is a maximal geodesic ray. Since q is a cut point of p on γ, the geodesics c and γ are distinct maximal nonspacelike geodesics containing p and q. Either Lemma 4.3 or Lemma 5.1 now yields a contradiction to the maximality of c beyond q. □

We mention without proof a related result holding for globally hyperbolic space-times. Namely, given any $p \in M$, there is a future-directed maximal nonspacelike ray $c: [0, a) \to M$ with $c(0) = p$ which is future inextendible. Thus $d(c(0), c(t)) = L(c \mid [0, t])$ for all $0 \leqslant t < a$. Dually, there are past-directed maximal rays that are past inextendible—compare [27].

References

1. Klingenberg, W. (1959). *Ann. Math.*, **69**, 654.
2. Klingenberg, W. (1961). *Comment. Math. Helv.*, **35**, 47.
3. Berger, M. (1960). *Ann. Scuola Norm. Sup. Pisa, Ser. III*, **14**, 161.
4. Rauch, H. (1951). *Ann. Math.*, **54**, 38.
5. Cheeger, J., and Ebin, D. (1975). *Comparison Theorems in Riemannian Geometry* (North Holland, Amsterdam).
6. Whitehead, J. H. C. (1935). *Ann. Math.*, **36**, 679.

7. Hawking, S. W. (1967). *Proc. R. Soc. London*, **A300**, 187.
8. Hawking, S. W., and Ellis, G. F. R. (1973). *The Large Scale Structure of Space-Time* (Cambridge University Press, London).
9. Hawking, S. W., and Penrose, R. (1970). *Proc. R. Soc. London*, **A314**, 529.
10. Beem, J. K., and Ehrlich, P. E. (1979). *Math. Proc. Camb. Phil. Soc.*, **85**, 161.
11. Beem, J. K. (1976). *Gen. Rel. Grav.*, **7**, 501.
12. Geroch, R. P. (1968). *Ann. Phys. (N. Y.)*, **48**, 526.
13. Kundt, W. (1963). *Z. Phys.*, 172, 488.
14. Beem, J. K., and Ehrlich, P. E. (1977). *C. R. Acad. Sci. Paris Ser. A*, **285**, 1129.
15. Tipler F. (1977). *J. Math. Phys.*, **18**, 1568.
16. Tipler, F. (1977). *Ann. Phys. (N. Y.)*, **108**, 1.
17. Hawking, S. W., and Sachs, R. K. (1974). *Commun. Math. Phys.*, **35**, 287.
18. Hicks, N. J. (1965). *Notes on Differential Geometry* (D. Van Nostrand, Princeton, New Jersey).
19. Nomizu, K., and Ozeki, H. (1961). *Proc. Am. Math. Soc.*, **12**, 889.
20. Kobayashi, S. (1967). *Studies in Global Geometry and Analysis* (MAA Studies in Mathematics, no. 4, p. 96).
21. Thorpe, J. A. (1977). Abstracts of Contributed Papers to 8th Int. Gen. Rel. Conf., p. 334.
22. Thorpe, J. A. (1977). *J. Math. Phys.*, **18**, 960.
23. Lerner, D. (1972). *Springer Lecture Notes in Phys.*, **14**, 1.
24. Penrose, R. (1972). Techniques of Differential Topology in Relativity (Regional Conference Series in applied Math. 7, SIAM, Philadelphia).
25. Bishop, R. L., and Crittenden, R. J. (1964). *Geometry of Manifolds* (Academic Press, New York).
26. Wolter, F. E. (1979). *Arch. Math.*, **32**, 92.
27. Beem, J. K., and Ehrlich, P. E., *Math. Proc. Camb. Phil. Soc.*, (in press).