The Hawking-Penrose Singularity Theorems

Draft version

4th March 2017

Contents

List of Figures

Chapter 1

Proof of the Hawking-Penrose Singularity Theorems

1.1 Proof of the Hawking-Penrose Singularity Theorem

1.1.1 Introduction

The singularity theorems are based on very powerful indirect arguments which show that black hole and cosmological singularities are generic in classical general relativity. Gives us confidence in the big bang. The classical theory predicts its own breakdown.

In the singularity theorems one does not define a singularity as a place the curvature diverges, one characterises singularities as the "holes" left behind by the removal of their pesence. These "holes" should be detectable by the fact that their will be geodesics that cannot be extended any further in at least one direction but still only have a finite range of affine parameter. In the singularity theorems one does not directly establish that the curvature diverges, but merely that there is an obstruction of sould beome sort to timelike or null geodesics being extendable within the spacetime to infinte length. However, it is not always true that this obstruction indeed arises because of the presence of diverging curvature, and the theorem does not directly show this.

Of course one can just "artificially" remove points from spacetimes which would then be considered to be singular. However, we can avoid such possibilities by restricting consideration only to spacetimes which are not part of a larger spacetime.

Spacetime itself (the structure $(\mathcal{M}, g)$) consists entirely of regular points at which g is well behaved.

Definition A spacetime is singular if it is timelike of null geodesically incomplete but cannot be embeded in a larger spacetime.

[?], [?]

Energy condition implies a tendency for geodesics to converge, that is, it guarantees that gravity is attractive. Causality conditions prevents geodesics from converging through causality violations.

Figure 1.1: Sir Roger Penrose and Stephen Hawking. Initiated by Penrose, Penrose and Hawking, together with Robert Geroch, contributed much of the work on the existence of spacetime singularities with the use of point-set topological methods.

1.1.2 Some Basic Terminology

Terms we'll encounter

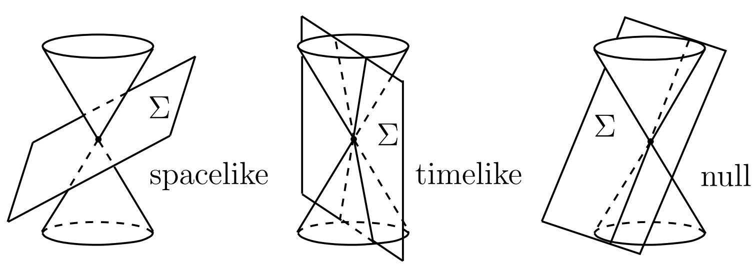


Figure 1.2: .

Properties of the spacetime:

Spacetime time orientable: At every point there are two cones of timelike vectors. The Lorentzian manifold is time orientable if a continuous choice of one of the cones,

termed future, can be made so it doesn't "turn upon itself" and be inconsistent.
Spacetime is Hausdorff: Events can be isolated.
Spacetime metric is non-degenerate: The metric is non-degenerate.

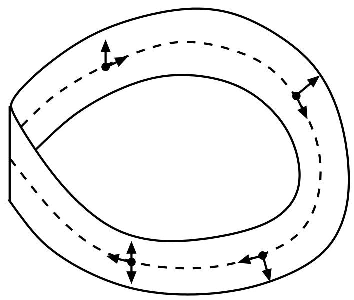


Figure 1.3: An example of a non-orientable space.

Paths:
Curves:
Trips:

Causally defined sets:

Chronalogical sets, I^+ and J^+: $I^+(x) := \{y \in \mathcal{M} : x \ll y\}$ is called the chronological future of x; $I^-(x) := \{y \in \mathcal{M} : y \ll x\}$ is called the chronological past of x; $J^+(x) := \{y \in \mathcal{M} : x \preceq y\}$ is called the causal future of x; $J^-(x) := \{y \in \mathcal{M} : y \preceq x\}$ is called the causal past of x. The chronological future of a set $S \subseteq \mathcal{M}$ is defined by the set $I^+(S) = \{y \in \mathcal{M} : x \ll y \text{ for some } x \in S\}$. Similar definitios hold for $I^-(S)$, $J^+(S)$ and $J^-(S)$.
Achronal sets: A set $S \subseteq \mathcal{M}$ is achronal if no two ponts of S are timelike related.
Achronal bounderies
Edge of a closed achronal set:

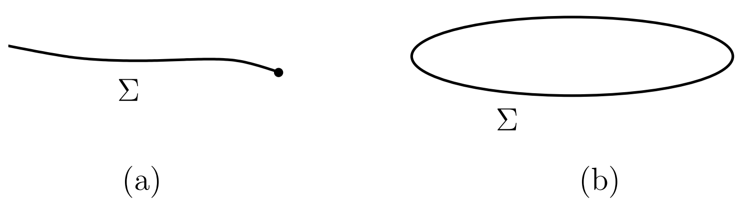


Figure 1.4: (a) A closed achronal set with edge. (b) A closed achronal set without edge.

Special regions:

convex normal neighbourhoods: An open set U is said to be a convex normal neighbourhood if for any $p, q \in U$ there is a unique geodesic lying in U joining p to q.
geodescially convex:
simple regions:

Causality conditions:

All spacetimes in general relativity locally have the same stucture qualitative causal structure as in special relativity, globally spacetimes may not be not be "causally well behaved". Here are some of the more imprtant conditions:

Chronology condition: There are no closed timelike curves.
Causality condition: Closed null curves can exist even when the chronology condition is satisfied, this motivates the definition of the causality conditin. The causality condition is satisfied when there are no closed causal curves.
Strong causality: there are no "almost closed" causal curves. This condition will be of particular importance for the singularity theorem.
Future (past) distinguishing: Any two points with the same chronological futurre (past) coincide, i.e., $(\mathcal{M}, g)$ is future distinguishing at $p \in \mathcal{M}$ if $I^+(p) \neq I^+(q)$ for $q \neq p$ and $q \in \mathcal{M}$.
stable causality: Stable causality is violated when a spacetime is "on the verge" of having closed timelike curves in the sense that an arbitrary small perturbation of the metric can produce a new spacetime that violates the chronology condition. See fig(1.1.2).
Global hyperbolicity: One definition of globally hyperbolic is that for $(\mathcal{M}, g)$ $\mathcal{M}$ is strongly causal and $J^+(p) \cap J^-(q)$ is compact for any $p, q \in \mathcal{M}$.

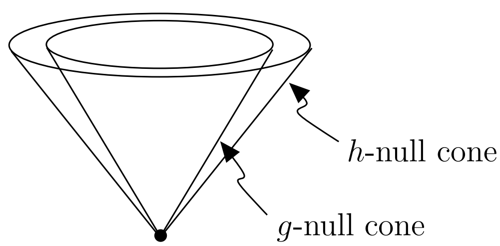


Figure 1.5: The $h-$null cone contains more timelike vectors than the $g-$null cone so there is more likelyhood to find closed timelike curves in $(\mathcal{M}, h)$ than in $(\mathcal{M}, g)$.

Topology of the manifold:

topological base:
compact: Compact basically mens the space is finite in size. It is easily shown that compact spacetimes necessarily have closed timelike curves and so are ruled out.
paracompactnes: A space may be non-compact, but if we have an open cover then we can "refine" it in a manner that still retains its finiteness properties locally on the space. If a manifold admits a Lorentz metric, it is paracompact. A manifold admits a positive definite metric if and only if it is paracompact. The availibility of a positive-definite metric is used in the singularity theorems.
Alexanderoff topology: The sets $\{I^+(a) \cap I^-(b) : a, b \in \mathcal{M}\}$ are open and form a basis for a topology on $\mathcal{M}$, called the Alexanderoff topology. An important result involved in the proof of the singularity theorems is that when a spacetime is strongly causal, the

Alexanderoff topology coincides with the topology of the spacetime.

Domains of dependence: Let S be an achronal subset of $\mathcal{M}$. Future and past domains of dependence os S and the total domain of dependence of S, respectively are defined as follows:
$D^+(S) = \{x \in \mathcal{M}$: every past endless causal curve from x intersects $S\}$.
$D^-(S) = \{x \in \mathcal{M}$: every future endless causal curve from x intersects $S\}$.
$D(S) = \{x \in \mathcal{M}$: every endless causal curve containing x intersects $S\}$.
Obviously, $D(S) = D^+(S) \cup D^-(S)$
Cauchy horizons: The future, past or total Cauchy horizon of an achronal closed set S is defined as (respectively):
$H^+(S) = \{x \in \mathcal{M} : x \in \overline{D^+(S)}$ but $I^+(x) \cap D^+(S) = \emptyset\}$,
$H^-(S) = \{x \in \mathcal{M} : x \in \overline{D^-(S)}$ but $I^-(x) \cap D^-(S) = \emptyset\}$,
$H(S) = H^+(S) \cup H^-(S)$
Cauchy surface: A Cauchy hypersurface for $\mathcal{M}$ is a non-empty set S for which $D(S) = \mathcal{M}$ **partial Cauchy surface:**
Global hyperbolicity:

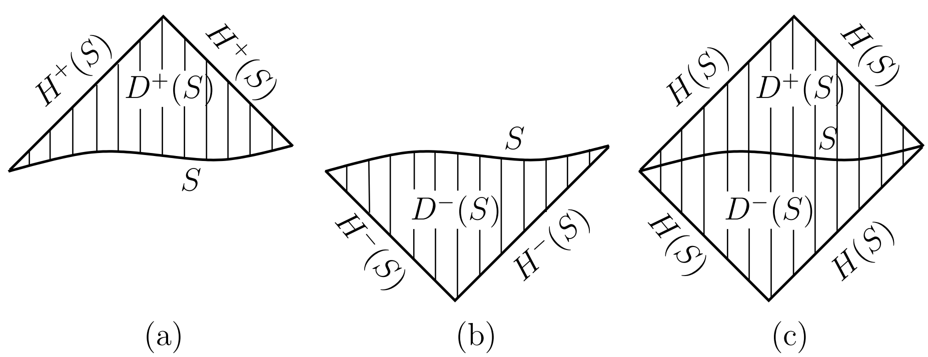


Figure 1.6: Domains of dependence. (a) The future domains of dependence of the achronal set S. (b) The past domain of dependence of the set S. (c) The total domain of dependence of S.

Energy conditions:

Weak energy condition: The weak energy condition states that $T_{ab}V^aV^b \geq 0$ for any timelike vector V. It is equivalent to saying that the energy density as measured by any observer is non-negative
Null energy condition: We say that an energy-momentum tensor T satifies the null energy condition if, for every null vector V, $T_{ab}V^aV^b \geq 0$. Combined with the field equations the null energy condition implies that for all timelike V, $R_{ab}V^aV^b \geq 0$.
Dominant energy condition: The dominant energy condition stipulates that, in addition to the weak energy condition holding true, for every future-pointing causal vector field (either timelike or null) V, the vector field $-T^a{}_bV^b$ must be a future-pointing causal vector. That is, mass-energy can never be observed to be flowing faster than light.

Strong energy condition: The strong energy condition stipulates that for every timelike vector field V, the trace of the tidal tensor measured by the corresponding observers is always non-negative: $\left(T_{ab} - \frac{1}{2}T g_{ab}\right) V^a V^b \geq 0$
Generic condition: The weak (resp. null energy) condition holds. If every timelike (resp. null) geodesic contains a point where there is some curvature that is not specially aligned with the geodesic. The generic condition is not satisfied by a number of known exact solutions, however, they are rather special. The generic condition specifically is that $t_{[a}R_{b]cd[e}t_{f]}t^c t^d \neq 0$ where t is the tangent vector. If the generic enery conditiond holds, each geodesic will encounter a region of gavitational focussing. As we will prove later, this will imply that there are pairs of conjugate points if one can extend the geodesic far enough in each direction.

Closed trapped surface: At a normal closed surface, the outgoing null rays from the surface diverge, while the ingoing rays converge. On a closed trapped surface, both the ingoing and outgoing null rays converge.
Future-trapped sets:

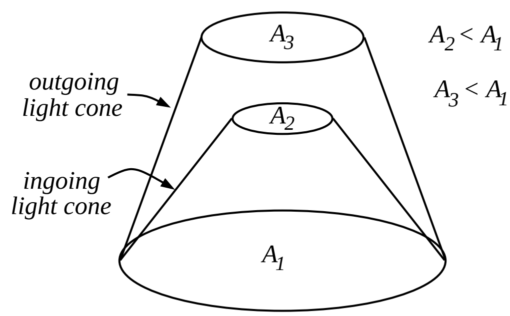


Figure 1.7: closedtrapArea. A closed trapped surface is when the outgoing light cone also converges.

Jacobi fields: We encounter Jacobi fields in the derivation of the Einstein's vacuum field equations. They are the connecting vectors of neighbouring geodesics.

Focal points: A focal point is where neighbouring geodesic intersect. Where neighbouring geodesics normal to a surface S intersect is also called a focal point.

Conjugate points: Focal points.

Space of causal curves: Space of causal curves.

1.1.3 The Singularity Theorem of Hawking and Penrose

It predicts that gravitational collapse , both at the Big Bang and inside black holes, brings about space-time singularities as at which the theory breaks down.

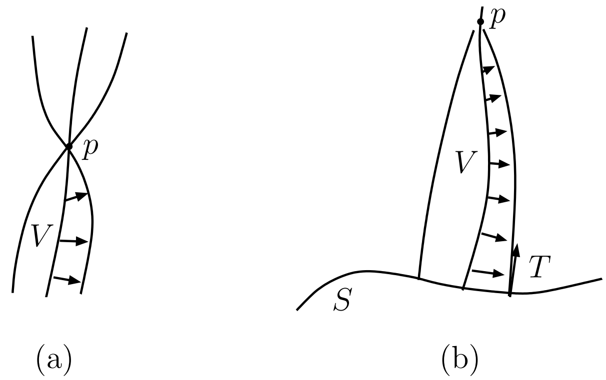


Figure 1.8: Focal points. (b) Focal point to a surface.

The Strong Energy Condition is the most important energy condition (Gravity is attractive)

The basic concepts needed to formulate and understand the singularity theorems will be developed as we work through the proof.

In the context of the singularity theorems, the working definition of a singularity is a kind of incompleteness of the space-time under consideration, more precisely, it is an obstruction of some sort to time-like or null geodesics from being indefinitely extendible. If the world line of a particle only exists for a finite proper time then clearly something has gone wrong but the nature of the singularity - whether it is physical reason or just a mathematical pathology is in general unclear.

Penrose Singularity Theorem Future max. development of an initial data set containing a (future) trapped surface is incomplete. Tools: Theory of geodesics in Lorentzian (Riemm.) geometry. Causal geometry. Raychadhouri equation plays the decisive role. The full force of the Einstein equations is not used.

The singularity theorem give very little information about the nature of the singularities, in effect they deny the existence of timelike or null geodesically complete spacetimes. The reason for the incompleteness is not predicted.

Many people believe that the resolution of the problem of singularities will come from modifications of the Einstein equations due to Quantum Gravity at the Planck scale. Explicitly demonstrated in loop quantum cosmology - a mini-superspace model which appropriately incorporates the discrete nature of quantum spacetime.

Theorem (Hawking and Penrose (1970))

Theorem 1.1.1 *1.* $R_{ab}K^aK^b \geq 0$ *for every non-spacelike vector* $\mathbf{K}$.
2. $\mathcal{M}$ *contains no closed timelike curves*

$$R_{ab}K^aK^b < 0 \tag{1.1}$$

(3) Every non-spacelike geodesic, with tangent vector K, contains a point at which

$$K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0. \tag{1.2}$$

(4) There exists at least one of the following:

(i) a compact achronal set without edge,
(ii) a closed trapped surface,
(iii) the null geodesics from some point are eventually focussed, or
the null geodesics from some closed 2-surface are all converging.

Condition (4) (i) is satisfied for any past-directed null-cone in our universe, because of the focussing affect of the black-body radiation; while (ii) is more or less the definition of a black hole, Condition (3) is a genericity condition is true for physical space-times except for certain highly symmetrical solutions.

An alternative version of the theorem is that the following three conditions cannot hold:

Theorem 1.1.2 *The following are mutually inconsistent in any space-time:*

(a) There are no closed trips

(b) Every endless causal geodesic contains a pair of conjugate points

(c) There is a future (past) - trapped set $\mathcal{S}$ in $\mathcal{M}$.

In fact it will be the alternative form of the theorem we will prove. After we have done that we will explain how the other theorem follows.

1.1.4 Basic Definitions

Definition A Lorentzian manifold is time-orientable if a continuous designation of future-directed and past=directed for non-spacelike vectors can be made over the manifold.

Definition A **path** is a smooth map from a connected set in $\mathbb{R}$ to $\mathcal{M}$.

Definition A **curve** is the point set image of a path. If $t \to \gamma(t)$ is a path, the curve determined by $\gamma(t)$ will be denoted γ.

Future timelike curves are identified with worldlines of material particles in $\mathcal{M}$.

A timelike curve is maximally extended in the past if it has no past endpoint (such a curve is also called past-inextendible). The idea behind this is that such a curve is fully extended in the past direction, and not merely a segment of some other curve. Similarly definitions hold for timlike curves extended into the future.

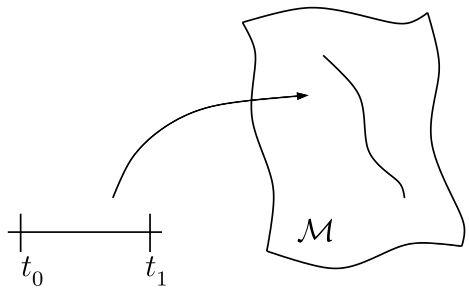


Figure 1.9: A path is a connected set in $\mathbb{R}$ to the space-time manifold $\mathcal{M}$.

The set of future-directed causal curves with a past endpoint a and a future endpoint b will be denoted $\mathcal{C}(a, b)$.

We will require that any smooth timelike curve to contain any end points it may have. This requirement eliminates curves that fail to be timelike at an end point, and which is null there instead.

Definition A causal curve is said to be **past (future)-inextendable** if it has no past (future) end point in $\mathcal{M}$.

The fig. (5.8 (a)) is a curve in Minkowski that can be prolonged while remaining timelike and hence is a timelike curve that is *not* inextendable. we have extended the curve α with a past directed, timelike curve which continues indefinitely. The new curve α' is past-inextendable. The new curve α'' is future-inextendable. In fig. (5.8(d)) a point has been removed we cannot prolong this curve any further into the future and so it is future-inextendable.

The null generators of the event horizon in fig.(**??**) in the sense that the continuation of the geodesic further into the past is no longer in the event horizon.

Let us model this by the artifical example given in fig(**??**) where we have removed a point from Minkowski spacetime. This provides an example of how the causal future of a point p will not neccessarily coincide with the closure of the chronological future of p.

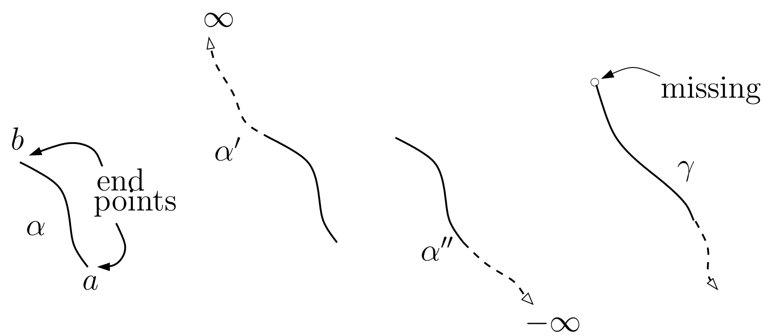


Figure 1.10: Examples in Minkowski spacetime. (a) The curve α is taken to contain its own past end point a and future end point b. (b) The curve α' is now future-inextendable . (c) The curve α'' is now past-inextendable (d) The curve γ is not future-inextendable because it cannot be prolonged any further.

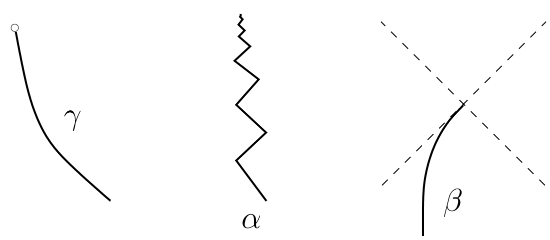


Figure 1.11: (a) A curve in Minkowski spacetime with point removed. (b) A curve is zig zag that it fails to have a well defined tangent vector at a point. (c) Here we have a time like curve which is null at its end point.

In fact this situatio is more general than the above example. Say infintesiamlly neighbouring null geodesics from p intersect at q. This means the point q will be conjugate to p along the null geodesic γ joing them. For points on γ beyond the conjugate point q there will be a variation of γ that gives a timelike curve from p. Thus γ cannot lie in the boundary of the future of p beyond the conjugate point q. So γ will have a future endpoint as a generator of the boundary of the future of p.

What we can say is the following.

Proposition 1.1.3 *For all subsets $S \subset \mathcal{M}$,*

(1) int $J^+(S) = I^+(S)$,

(2) $J^+(S) \subset \overline{I^+(S)}$.

Proof:

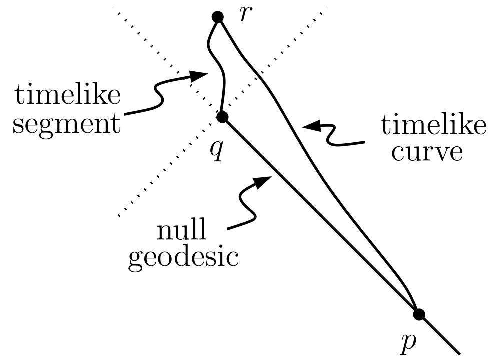


Figure 1.12: In flat spacetime, when a null geodesic curve joins onto a timlike curve, there exists a timelike curve between p and q.

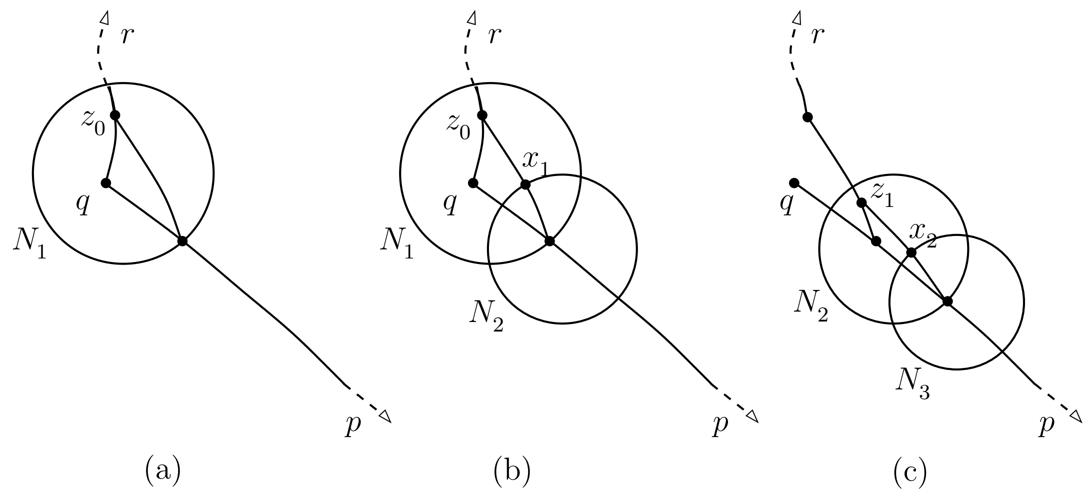


Figure 1.13: Say there are two points p and q connected by a null curve and a point r which is connected to q by a timelike curve. A timelike curve joining to the null geodesic. (b) Contiuing in this way, (c), we "peel" away a timelike curve that joins r and p.

Obvious.

□

1.1.5 Achronal Sets

Intuitively, a neighbourhood of a point is a set containing the point where you can move that point some amount without leaving the set.

Definition If X is a topological space and $p \in X$, a neighbourhood of p is a set V, which contains an open set U containing p,

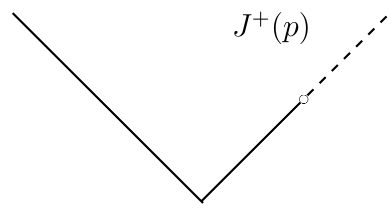


Figure 1.14: Artificial example of how the causal future of a point p will not neccessarily coincide with the closure of the chronological future of p.

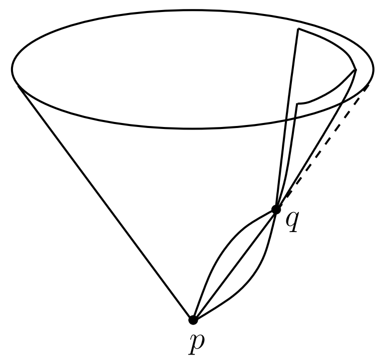


Figure 1.15: The point q is conjugate to p along null geodesics, so a null geodesic γ that joins p to q will leave the bounadry of the uture of p at q.

$$p \in U \subseteq V.$$

Definition A subset $S \subset \mathcal{M}$ is **achronal** provided no two of its points can be joined by a timelike curve.

Lemma 1.1.4 *An* **achronal boundary** *B is an achronal set.*

Proof: Let $B = \partial I^+[S]$, for some. If two points $x, y \in B$ satisify $x << y$, then $y \in I^+(x) \subseteq I^+[S]$ which is open. Hence y cannot be on the boundary of $I^+[S]$.

□

Lemma 1.1.5 *Let $\mathcal{C}(\lambda)$ be a causal curve that intersects $\dot{I}^+[S]$ at some point p. In the past the $\mathcal{C}(\lambda)$ reamins in $\dot{I}^+[S] \cup I^+[S]$.*

Proof: Consider an arbitrary point x on $\mathcal{C}(\lambda)$ such that $x \gg p$ and an arbitaryly small neighbourhood of x. A small deformation of the curve $\mathcal{C}(\lambda)$, between p and x, produces a timelike curve $\mathcal{D}$ from some point $q \in U(x)$. Since $p \in \dot{I}^+[S]$, a slight deformation of $\mathcal{D}$, keeping it timelike, produces a curve $\mathcal{E}$ from q to to some point in $I^+[S]$. But as q is in some arbitrary small neighbourhood of x, it must also lie in $I^+[S]$ or else in its boundary, $\dot{I}^+[S]$.

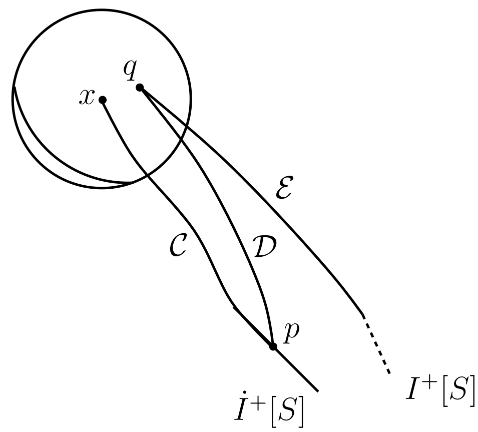


Figure 1.16: .

□

Lemma 1.1.6 *If $p \in \partial I^+(S)$ then $I^+(p) \subset I^+(S)$, and $I^-(p) \subset \mathcal{M}/\overline{I^+(S)}$.*

Note that if $q \in I^+(p)$ then $p \in I^-(q)$, and hence $I^-(q)$ is a neighbourhood of p. Since p is on the boundary of $I^+(S)$, it follows that $I^-(q) \cap I^+(S) \neq \emptyset$, and hence $q \in I^+(S)$. The second part of the lemma can be proven similarly.

□

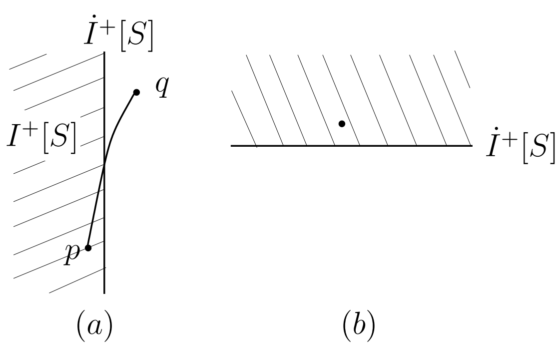


Figure 1.17: The future set cannot be bounded by timelike curves.

Proposition 1.1.7 *Let $B = \partial$, where F is a future set. Let $x \in B$ and suppose there is an open set Q containing x satisfying:*

(a) If $y \in Q \cap F$, there is a point z in F/Q, such that $z \ll y$.

or equivalently

(b) $I^-(y) \cap F/Q \neq \emptyset$, *for all* $y \in Q \cap F$

or equivalently

(c) $F = I^+[F/Q]$,

then x *is in the future endpoint of a null geodesic lying in* B.

The proof uses the method of taking a limit of causal curves. A technical difficulty arises however in that a limit of smooth causal curves need not be smooth. This leads to the need of the notion of a C^0 causal curve.

A C^0 causal curve is a continuous curve is a continuous curve that can be approximated with arbitrary precision by piecewise smooth causal curve.

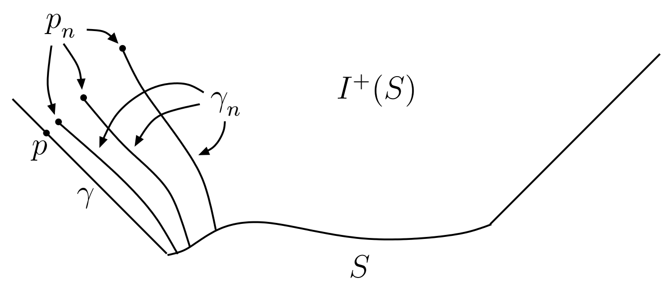


Figure 1.18: .

Theorem 1.1.8 $\dot{I}^+[S]$ *is generated by null geodesics which have no past endpoints.*

Proof: Take an arbitary point p in $\dot{I}^+[S]$. Construct an arbitrary neighbourhood $U(p)$. In $U(p) \cap I^+[S]$ construct a sequence of points $\{p_n\}$ which converges to to the point p. For each n, construct a causal curve γ_n extending from p_n to S. Let q_n be the intersection of γ_n with $\dot{U}(p)$, the boundary of $U[p]$. Since $\dot{U}(p)$ is a compact set, the sequence q_n must have a limit point, q (the proof of this fact of point set topology will be given in section 1.1.7). As there are causal curves from points p_n, arbitrarily close to p to points arbitrarily near q there must be a causal curve, $\mathcal{C}$, from p to q.

Since q is a limit point of a sequence of points in $I^+[S]$, q either lies in $I^+[S]$ or in its boundary $\dot{I}^+[S]$, or both. Suppose $q \notin \dot{I}^+[S]$. Then there is a small neighbourhood $U(q)$ is contained entirely in $I^+[S]$. Construct a causal curve from p to S by going from p to q along the causal curve $\mathcal{C}$, then from q along a timelike curve to some point $r \in U(q)$, and then from r to S along a causal curve. Since this curve from p to S has a timelike segment, it reaches any desired point s in some small neighbourhood $V(p)$. But this means that $V(p) \subset I^+[S]$, hence that $p \notin \dot{I}^+[S]$. This contradicts the definition of p, and so it must be that $q \in \dot{I}^+[S]$.

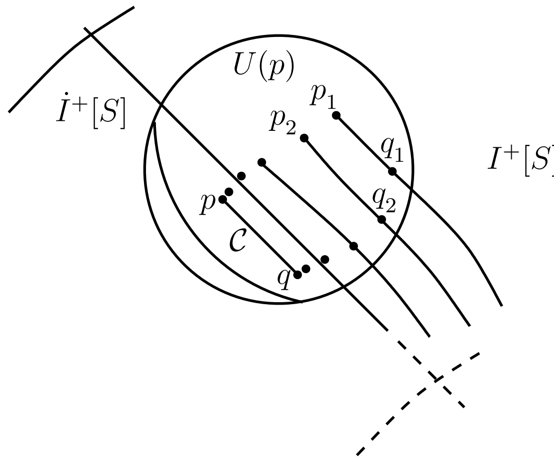


Figure 1.19: .

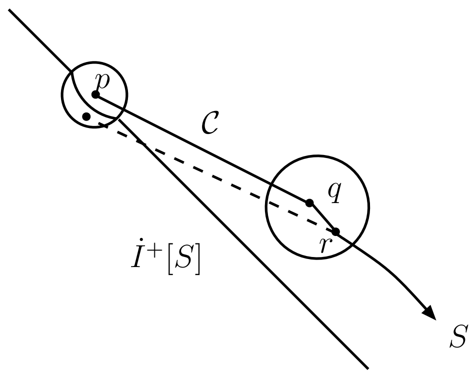


Figure 1.20: .

Proposition 1.1.9 *Once a a generator, being followed into the future, enters $\dot{I}^+[S]$, it can never leave $\dot{I}^+[S]$.*

□

Proof: Follow the generator $\mathcal{C}$ from p to p' and futher. This curve can never leave $\dot{I}^+[S]$.

□

1.1.6 Strong Causality

The **strong causality** condition holds on $\mathcal{M}$ if there are no closed or "almost-closed" timelike or null curves through any point of $\mathcal{M}$. If μ is any timelike curve in a spacetime that obeys the strong causality condition and x is any point not on μ, then there must be some neighborhood $\mathcal{N}$ of x that does not intersect μ. (Otherwise, μ would accumulate at x, and thereby give an almost-closed timelike curve.)

Globally hyperbolic $\rightarrow$ causally simple $\rightarrow$ stably causal $\rightarrow$ strongly $\rightarrow$ distinguishing $\rightarrow$ causal $\rightarrow$ chronological.

Chronology: There are no closed timelike curves, collection of points $\pi\mathcal{M}$, such that $p_1 < p_2 \ldots p_n < p_1$.

Causality: There are no closed causal curves, collection of points $\{pi\}M, s.t.p_1 < p_2 < ... < p_n < p_1$.

Future / Past Distinguishing condition: Any two points with the same chronological future (past) coincide.

Strong Causality Condition: There are no almost closed timelike curves.

Stable Causality Condition: (M,g) is not "on the verge" of having a bad causal structure. (There is a neighbourhood of g in the Ck open topology ...) Definition: There exists a continuous nonzero tvf ta such that the metric $g'ab := gab$

the chronology and causality conditions a slightly stronger condition when the chronology condition is almost violated. Since an arbitrarily small perturbation of the metric will result closed timelike curves, such space-times also seems physically unreasonable. Techniquile reasons for considering strong causality. It will be of central importance in what follows.

Strong causality is violated at a point p if there are timelike curves starting at x which come arbitrarly close to p after leaving a given convex neighbourhood.

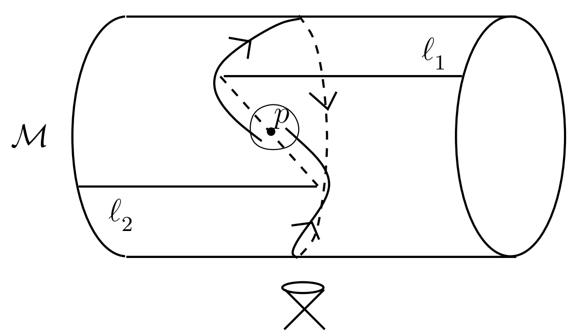
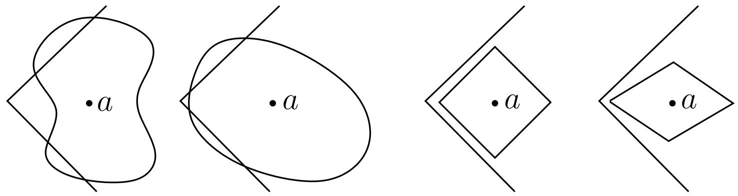


Figure 1.21: The lines ℓ_1 and ℓ_2 have been removed. This is a space-time which is causal but fails to be strongly causal.

We give a more precise definition of strong causality at a point.

A collection of open sets in $\mathcal{M}$ forms a neighbourhood basis of some point $x \in \mathcal{M}$ if every open set containing x has in it an open set from the collection with each open set in the collection containing x.

Definition (Strongly Causal) A space-time $\mathcal{M}$ is said to be *strongly causal* at $a \in \mathcal{M}$ if and only if a has a neighbourhood base $\{U_\alpha : \alpha \in \Lambda\}$ with the property that *no* U_α is intersected by a trip that then leaves U_α and then intersects U_α again. A space-time is strongly causal if it is strongly causal at each point.



(a) (b)

Figure 1.22: (a) . (b) Suitable causal basis.

In $\mathcal{M}^2$ (and elsewhere) one may construct various types of neighbourhoods at a point a. A neighbourdhood base at a constructed with sets similar to either of the first two in the figure is not a suitable **"causality base"**. $\mathcal{M}^2$ is strongly causal because there exists at each point a a neighbourhood base consisting of sets similar to either the third or the fourth.

Definition (Causally Convex) An open set $\mathcal{U}$ in a space-time $\mathcal{M}$ is said to be *causally convex* if no causal curve leaves $\mathcal{U}$ and then reenters $\mathcal{U}$.

Definition (Strongly Causal) A space-time is *strongly causal* at a point $p \in \mathcal{M}$ if every open neighbourhood of p contains a causally convex neighbourhood of p.

Definition Let N be a simple region containing a. The **local future** (*past*) of a point a, denoted $I_L^+(a)$ ($I_L^-(a)$), is

Definition Let N be a simple region containing $I_L^-(x) \cap I_L^+(y)$ neighbourhood base

$\{z : p \ll r \ll q$, where the trip lies in N.

In Minkowski space-time the points x and y

Proposition 1.1.10 *Let $a \in \mathcal{M}$. Strong causality fails at a if there is a point $b > a$, $b \neq a$, such that for all $x \in I^+(a)$ and all $y \in I^-(b)$, $y << x$.*

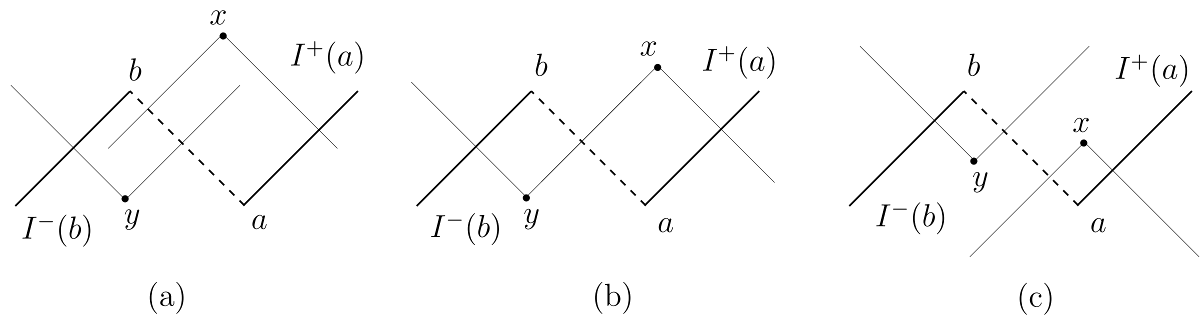


Figure 1.23: Minkowski space-time: b is joined to a by a null geodesic we consider $x \in I^+(a)$ and $y \in I^-(b)$. In the first in case (a) $y \ll x$. In case (b) $y < x$ but $y \not\ll x$. In case (c) x and y are not causally related, in particular $y \not\ll x$.

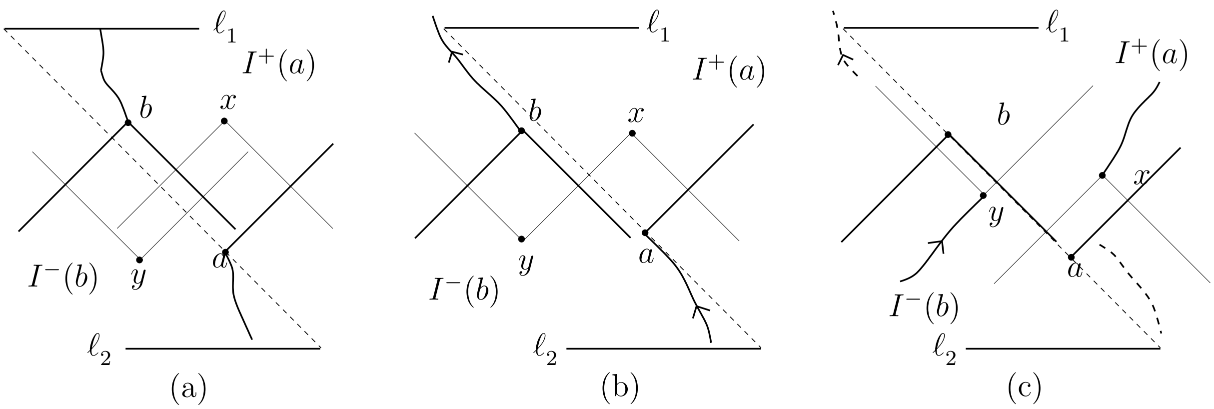


Figure 1.24: b is joined to a by a null geodesic we consider $x \in I^+(a)$ and $y \in I^-(b)$. In the first case (a) $y \ll x$ there are no closed timelike curves. In case (c) For some x and y.

Proof:

Suppose a and b exist satisfying the hypotheses. Separate a and b by disjoint neighbourhoods of a contained in $V(a)$ and $W(b)$. Let $U(a)$ be any neighbourhood of a contained in $V(a)$. Choose $x^1 \in I^-(a) \cap U(a)$ and $x \in I^+(a) \cap U(a)$ see fig.1.1.6 (a).

Since $x^1 << a$, and $a < b$, $x^1 << b$. Thus there is a trip from x^1 to b which must leave $U(a)$, see fig(1.1.6a). Choose y on this trip inside $W(b)$. Then $y << x$ by hypothesis, and the trip $[x^1yx]$ intersects $U(a)$ twice, fig.1.1.6 (b). Thus no nbd. of a contained in $V(a)$ can satisfy the strong causality condition and thus no nbd. base at a can satisfy it. So strong causality fails at a.

□

Now we consider the converse. We first need to introduce the following concept. Let N be a simple region containing a. Take any $x >> a$ such that x and a are joined by a unique timelike geodesic in N. x is said to be in the **local future** of a, denoted $x \in I_L^+(a)$.

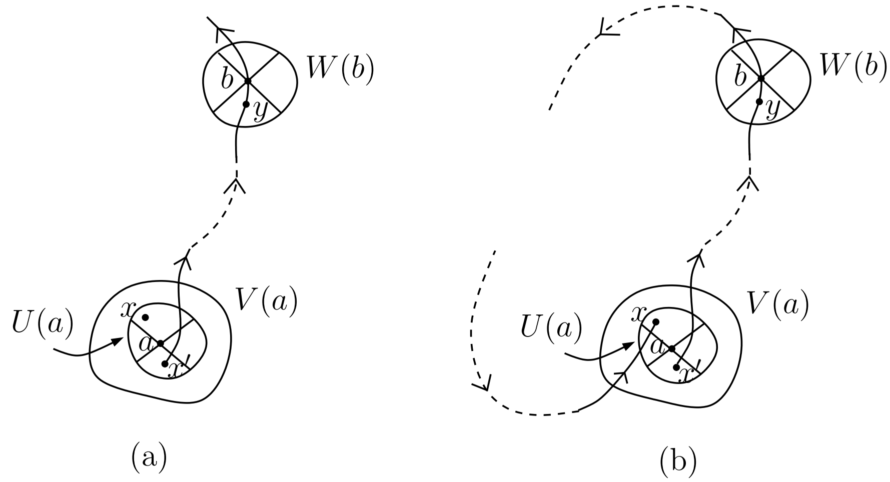


Figure 1.25: converse. (b) That $y \gg x$ to be true, there must be a timelike curve that renters $U(a)$.

Proposition 1.1.11 *Let $a \in \mathcal{M}$. Strong causality fails at a only if there is a point $b > a$, $b \neq a$, such that for all $x \in I^+(a)$ and all $y \in I^-(b)$, $y << x$.*

Proof:

If y is in the local past of a, then

$$I_L^+(x) \cap I_L^-(x) = \{z : y << z << x, \text{ where the trip lies in } N\}$$

is clearly an open neighbourhood of a. If a trip intersects this neighbourhood in a disconnected set, it clearly must leave and reenter N in order to do so.

If strong causality fails at a point a there must be a nested sequence of these neighbourhoods at a, $\{U_i : i = 1, 2, 3, \dots\}$ $U_i \supset U_{i+1}$; $\cap_i \{U_i\} = \{a\}$ with the property that each U_i contains the past end point of a trip γ_i which leaves N and then comes back to enter U_i.

$$U_i \ : i = 1, 2, 3, \dots; \quad U_I \supset U_{i+1} \ ; \ \cap_i \{U_i\} = a. \tag{1.3}$$

Being the limit of timelike curves $[c_i d_i]$ cannot by spacelike, but can be null or timelike??

$y << b_i$ as b_i is on γ_i $b_i << x$, hence $y << x$.

$$x \ll a_i \ll b_i \ll c_i \ll y \quad \text{for } i \text{ large enough} \tag{1.4}$$

□

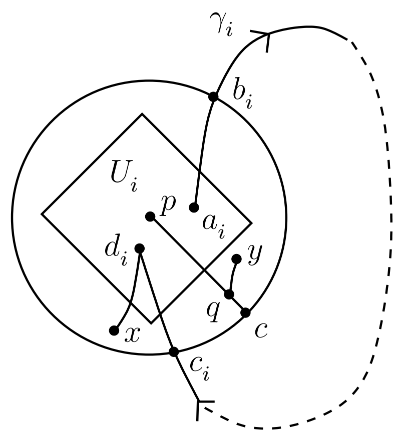


Figure 1.26: proof of .

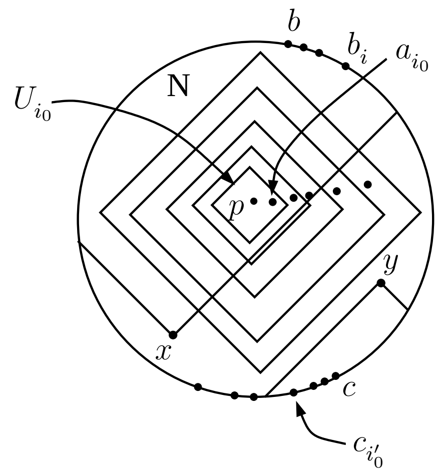


Figure 1.27: For $i \geq i_0$, $U_i \in I^+(x)$, from which we can conclude $x \ll a_i$. For $i \geq i'_0$, $c_i \in I^-(y)$. If we choose $i > i_0, i'_0$ then we have $x \ll a_i$ and $c_i \ll y$

Remark If, given the conditions of the 1.1.10, b is found to satisfy $a \ll b$, then for any $y \in I^-(b) \cap I^+(a)$, we have $y \ll y$; i.e. there are closed trips in $\mathcal{M}$.

Alexandroff Topology

A topological space is **Hausdorff** if any two given points can be isolated from each other by disjoint sets. Or more presicely, a topological space X is Hausdorff if any two given points $p_1, p_2 \in X$, there are two open sets U_1 and U_2 with $p_1 \in U_1$ and $p_2 \in U_1$ but with the intersection of U_1 and U_2 empty.

The finer the topology, the more open sets it has.

Theorem 1.1.12 *The following three conditions on a spacetime $\mathcal{M}$ are equivalent:*

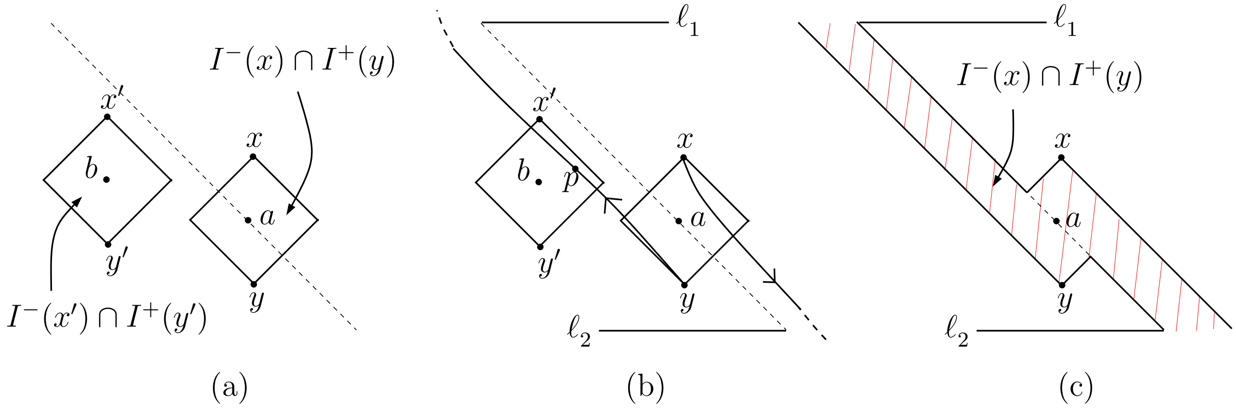


Figure 1.28: b a $x \in I^+(a)$ and $y \in I^-(b)$. In the first case (a) $y \ll x$ timelike curves. (b) Points in the open set of b are also in $I^+(x) \cap I^-(y)$ In case (c) .

(a) A spacetime $\mathcal{M}$ is strongly causal;

(b) the Alexandroff topology is Hausdorff;

(c) the Alexandroff topology coincides with the manifold topology.

Proof: If the alexanddorff topology is weaker than the manifold topology, there is an $\mathcal{M}$-open set $U(y)$ such that any A-nbd. of y gets outside of U, (for example in fig(1.1.6)). Since $\mathcal{M}$ is regular we may assume this holds for the $\overline{U}$ as well. Now let $V(y)$ be an arbitrary $\mathcal{M}$-open neighbourhodd of y. Let $b \in I^+(y)$, $a \in I^-(y)$. By hypoth, $\{I^-(b) \cap I^+(a)\}/\overline{U} \neq \emptyset$. Choose z in this set. Then $z \notin V(y)$, but $a \gg z \gg b$, and $V(y)$ disconnects the trip $[azb]$. Since $V(y)$ is arbitrary, strong causality fails at y.

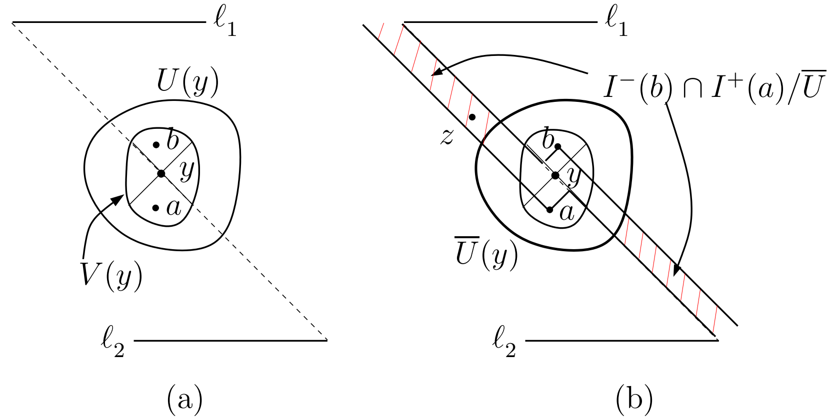
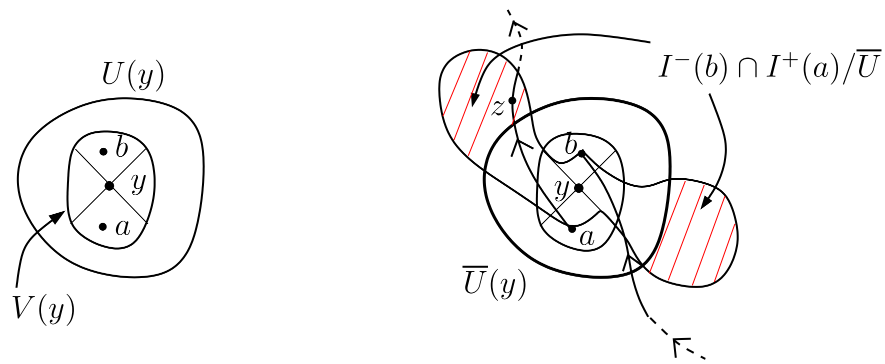


Figure 1.29: $a \in I^-(y) \cap V(y)$ and $b \in I^+(y) \cap V(y)$. In the first case (a) $y \ll x$ timelike curves. (b) Points in the open set of b are also in $I^+(x) \cap I^-(y)$.

Proposition 1.1.13 *For each $a \in \mathcal{M}$ define Q_a $\{x \in \mathcal{M} : x$ lies on a closed trip through $a\}$. Then Q_a is open, for every $a \in \mathcal{M}$; and either $Q_a \cap Q_b$, for all $a, b \in \mathcal{M}$.*



(a) (b)

Figure 1.30: $a \in I^-(y) \cap V(y)$ and $b \in I^+(y) \cap V(y)$. In the first case (a) $y \ll x$ timelike curves. (b) Points in the open set of b are also in $I^+(x) \cap I^-(y)$.

Proof:

Proposition 1.1.14 *If strong causality fails at $a \in \mathcal{M}$, then one (at least) of the following is true:*

(1) There are closed timelike curves through a.

(2) There is a future endless null geodesic through a along which strong causality fails; and there are closed tris near a.

(3) Time reverse of (2).

(4) There is an endless null geodesic through a along which strong causality fails.

In a spacetime $\mathcal{M}$, let G be defined as the set of points where strong causallity holds, i.e. $G \subset \mathcal{M}$ $G = \{x : \mathcal{M}$ is strongly causal at $x\}$.

Proposition 1.1.15 *Let $a \in \mathcal{M}$. Strong causality fails at $a \Leftrightarrow$ there is a point $b > a$, $b \neq a$, such that for all $x \in I^+(a)$ and all $y \in I^-(b)$, $y \ll x$.*

Definition The future domain of dependence, $D^+(\Sigma)$, is the set of points p in $\mathcal{M}$ for which every past-inextendable causal curve through p intersects Σ, fig(1.1.6).

We note that Penrose [**?**] and Geroch [**?**] use timelike curves to define the domain of depnedence, rather than non-spacelike curves used above, which agrees with Hawking and Ellis [**?**].

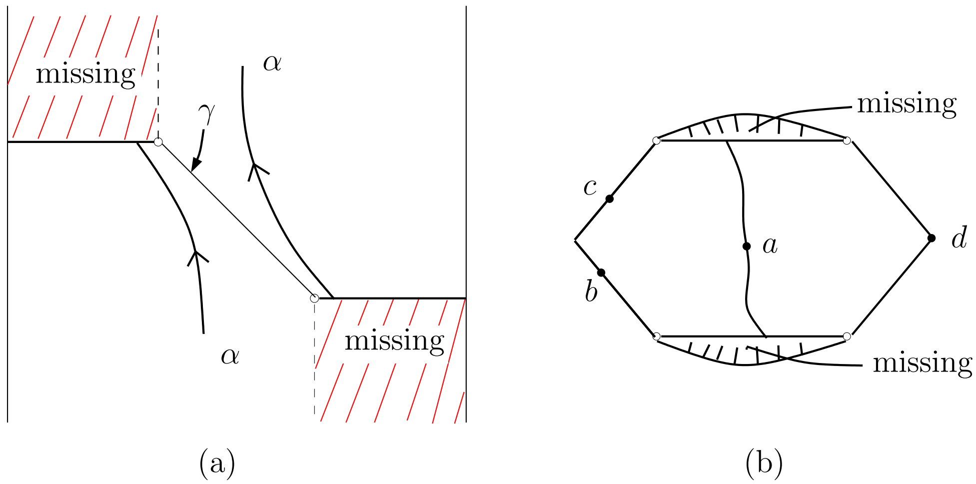


Figure 1.31: (a) There is an endless null geodesic along which strong causality is violated. (b) Strong causality is violated everywhere in R.

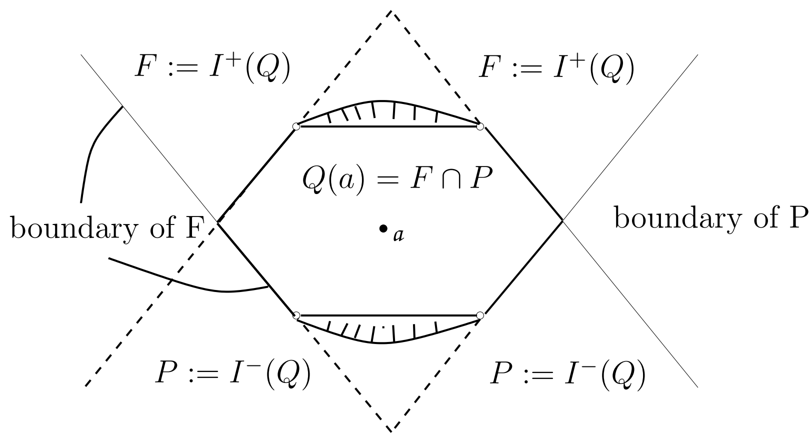


Figure 1.32: $F := I^+(Q)$ and $P := I^-(Q)$. $Q = F \cap P$, $\partial Q = F\partial + \partial P$.

The physical significance of the future (past) domain of development $D^+(\Sigma)$ is that it is the region in the future (past) of Σ that can be predicted from knowledge of the data on Σ.

Definition Let S be a closed achronal set. The edge of S is defined as a set of points $x \in S$ such that every neihbourhood $U(x)$ of x contains $y \in I^+(S)$ and $z \in I^-(S)$ and a timelike curve γ from z to y which does not meet S.

An example of a clsosed achronal surface without edge is given if fig ().

Proposition 1.1.16 *Let $S \subset \mathcal{M}$ be achronal and closed. Then:*

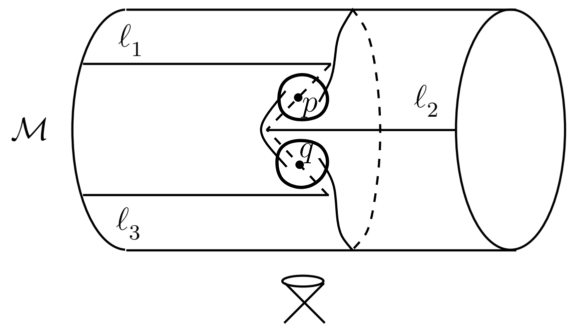


Figure 1.33: The lines ℓ_1, ℓ_2 and ℓ_3 have been removed.

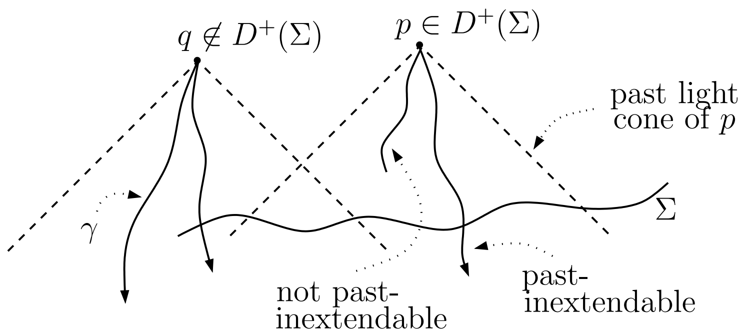


Figure 1.34: The future domain of dependence, $D^+(\Sigma)$, of Σ. p is in $D^+(\Sigma)$, q isn't because ther are past-inextendable causal curve through q that don't intersect Σ, e.g. the curve γ

(1) $H^+(S)$ *is achronal and closed.*

(2) $D^+(S)$ *is closed.*

(3) $x \in D^+(S) \Rightarrow \{I^-(x) \cap I^+[s] \subset D^+(s)$.

(4) $\partial D^+(S) = H^+(S) \cup S$.

(5) $I^+[H^+(S)] = I^+[s]/D^+(S)$.

Proof:

Definition Let $S \subset \mathcal{M}$ be achronal and closed. The **edge of** S is defined to be

$$\{x \in S : \text{for all} y, z, \text{ there is a trip}[zy] \text{ not meeting} S\}. \tag{1.5}$$

Proposition 1.1.17 *(a)* $x \in D^+(S)$ *implies* $I^-(x) \cap edge(S) = \emptyset$

(b) $edge(S) = edge(H^+(S))$.

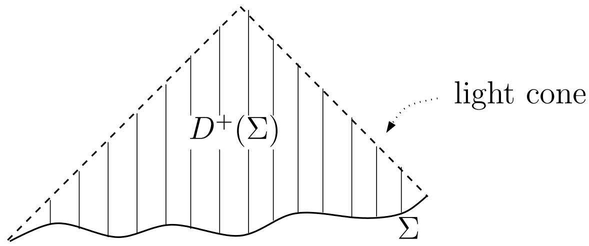


Figure 1.35: The future domain of dependence, $D^+(\Sigma)$, of a clsoed Σ in Minkowski spacetime.

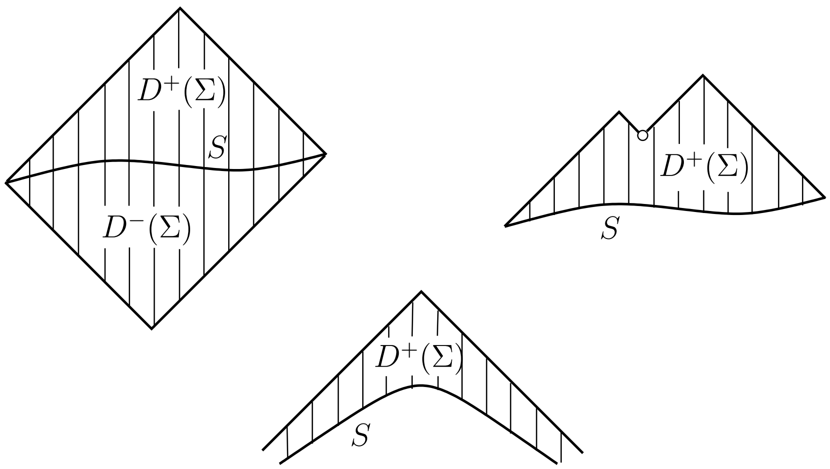


Figure 1.36: Domains of dependence.

Proposition 1.1.18 *Let $x \in H^+(S)/edge(S)$. Then there is a null geodesic on $H^+(S)$ with future endpoint x.*

Definition A non-empty, closed achronal set S is said to be *future-trapped* if $E^+[S]$ is compact. Past-trapped sets are defined analogously.

If we can associate with the topology some countable set of open sets and use the desirable properties of countability to learn something about the space. In conjuction with the requirement of strong causality, we establish the following result:

Proposition 1.1.19 *For $x \in \operatorname{int} D^+(S)$ the set $J^-(x) \cap J^+[S]$ is compact.*

Proof: Assume $K := J^-(x) \cap J^+[S]$ is not compact.

As we assume K to be non-compact, it has no cover with a finite subcover. There is an open cover $\{U_{i_1} : i = 1, 2, 3 \dots\}$ which is countable, locally finite and whose elements are simple regions.

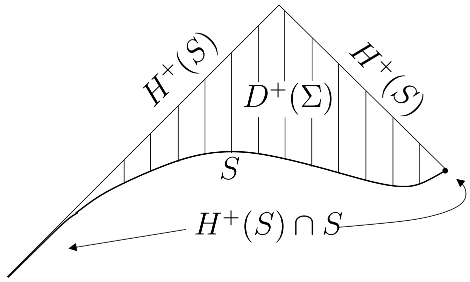


Figure 1.37: The future domain of dependence.

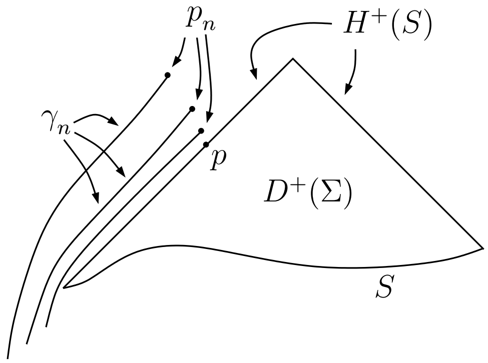


Figure 1.38: There exists a subsequence γ_m of past inextendble causal curves that do not meet S that converges to a past inextendble C^0 null geodesic γ starting at p.

1.1.7 The Space of Causal Curves

In a spacetime $\mathcal{M}$, let $G \subset \mathcal{M}$ be defined as the set of all points in $\mathcal{M}$ which are strong causality holds,

$$G := \{x : \mathcal{M} \text{ is strongly causal}\} \tag{1.6}$$

The set where srong causality fails being $\mathcal{M}/G$.

Proposition 1.1.20 *G is open.*

Proof We prove this by showing that $\mathcal{M}/G$ is closed. First note that a set is closed in $\mathcal{M}$ if and only if this set contains its own boundary. If we show that the limit point of every convergent series in $\mathcal{M}/G$ also belongs to $\mathcal{M}/G$, then by implication, every point in the boundary of $\mathcal{M}/G$ also belongs to $\mathcal{M}/G$. To this end, consider any series $\{x_i\} \in \mathcal{M}/G$ which converges to some point $x \in \mathcal{M}$. Take a simple region containing x_i, find the point b_i such that for all

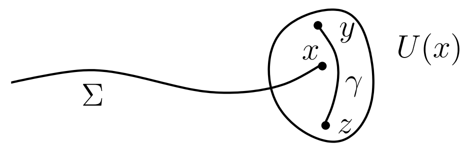


Figure 1.39: The edge of a closed achronal surface Σ.

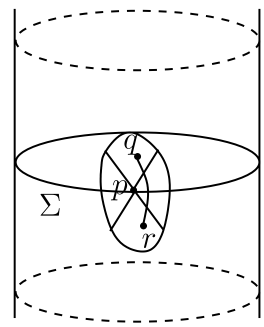


Figure 1.40: A simple example of a closed achronal surface without edge can be given by considering the spacetime $\mathbb{R} \times S$ with light cones locally at 45 degrees. For the open neighbourhood there is no $r \in I^-(p)$ and $q \in I^+(p)$ with a timelike curve between them that doesn't intersect S. The closed achronal set Σ has no edge.

$$a \in I^-(b_i) \text{ and } b \in I^+(x_i) \quad a \ll b. \tag{1.7}$$

Take the intersection of the geodesic $[x_i b_i]$ with the (compact) boundary of the simple region. These intersections have a cluster point y. Take two disjoint neighbourhoods of x and y. It is easy to construct a trip leaving this neighbourhood of x, entering the neighbourhood of y and then reentering the neighbourhood of x, showing that strong causality fails at x. Thus $x \in \mathcal{M}/G$.

□

The length of a smooth curve is

$$L(\alpha) = \int_p^q \left| g_{ab}(\alpha(t)) \frac{d\alpha^a(t)}{dt} \frac{d\alpha^b(t)}{dt} \right|^{1/2} dt \tag{1.8}$$

The length of a peicewise smooth curve is defined by adding the lengths of its smooth segments.

It is assumed that A and B are closed achronal sets contained in K, a compact subset of G. We denote by $\mathcal{C}_K(A, B)$ the set of causal curves from A to B lying in K, and by $\mathcal{T}_K(A, B)$ the set of causal trips from A to B lying in K.

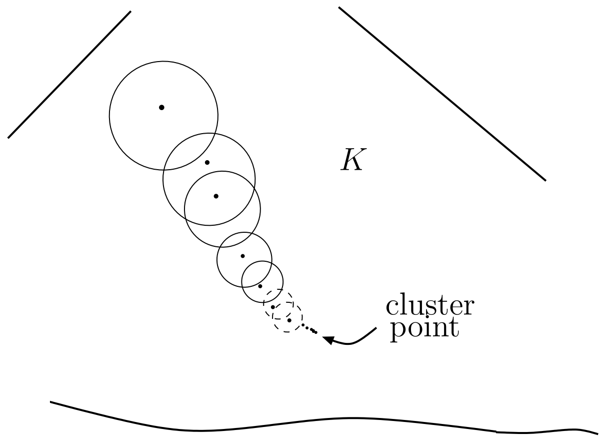


Figure 1.41: We can only get convergence to a cluster point in K if the space **wasn't** locally finite.

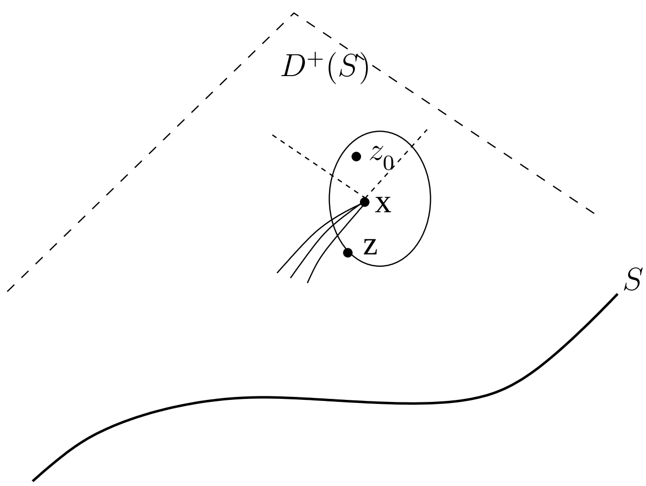


Figure 1.42: $K = J^+(S) \cap J^-(p)$ K is compact.

Lemma 1.1.21 *Any compact subspace C of a Hausdorff topological space X is closed in X.*

Proof: Let C be a compact subspace of a Hausdorff topological space X and let $a \in X - C$. We shall prove that there exist an open set U_a such that $a \in U_a \subset X - C$. Then $X - C = \cup_{a \in X-C} U_a$, so $X - C$ is open and hence C is closed.

Since X is Hausdorff, we can choose open subsets U_x and $V(x)$ of X such that $a \in U_x$, $x \in V(x)$, and U_x and $V(x)$ are disjoint. Choosing such subsets U_x and $V(x)$ for every point $x \in C$, we see that the sets $V(x)$ form an open cover of C. Since we assume C is a compact topological space, it follows that C is the union of the open sets $V_{x_1}, \ldots, V_{x_r}$ for some finite collection of points $x_1, \ldots, x_r \in C$. Let $U_a = \cap_{i=1}^r U_{x_i}$. Then U_a, as a finite intersection of open sets, is open. Also, $a \in U_a$ since $a \in U_{x_i}$ for every $i = 1, 2, \ldots, r$.

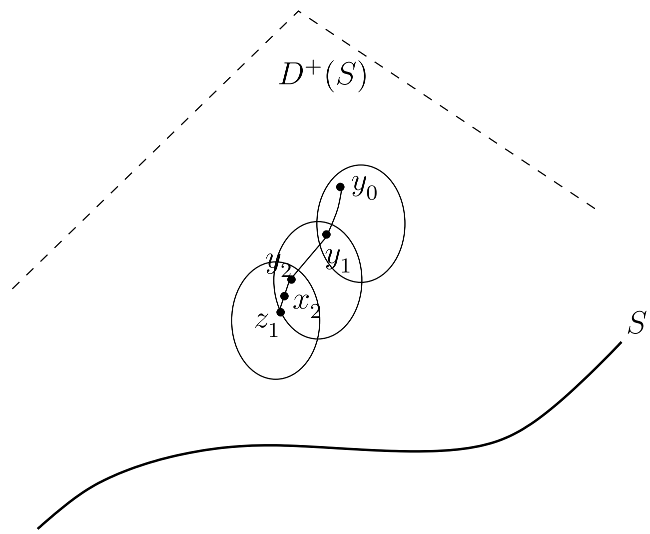


Figure 1.43: $K = J^+(S) \cap J^-(p)$ K is compact.

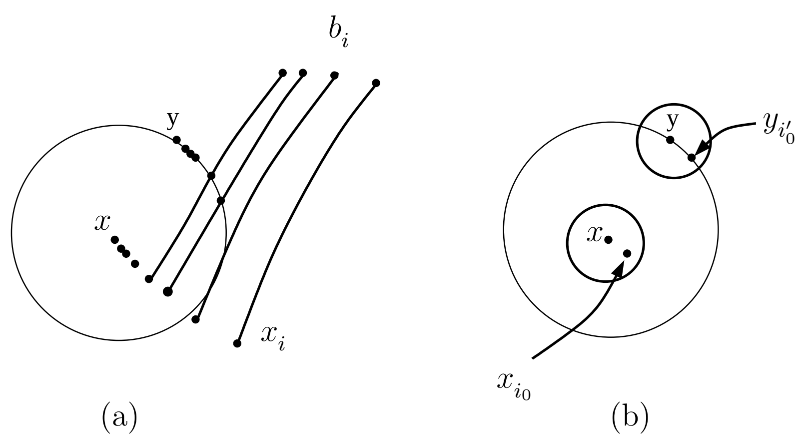


Figure 1.44: G is an open set.

Finally, if $b \in U_a$ then for any $i = 1, 2, \ldots, r$, $b \in U_{x_i}$ and hence $b \notin V(x_i)$, so $b \notin C$, since $C \subset \cup_{i=1}^r V(x_i)$. Thus $U_a \subset X - C$.

□

$\mathcal{M}$, as a topological space, is metrizable. Choose an arbitrary metric d compatible with the manifold topology. Then in G, d is compatible with the Alexandroff topology.

Recall that for a set E and a point p,

$$d(p, E) := \inf\{d(p, y) : y \in E\} \tag{1.9}$$

For a curve γ and some $\epsilon > 0$ we define the $\epsilon-$band $V_\epsilon(\gamma)$ about γ as

$$V_\epsilon(\gamma) := \{y \in \mathcal{M} : d(y, \gamma) < \epsilon\}. \tag{1.10}$$

For $\gamma \in \mathcal{C}_K(A, B)$ we define a metric on the space $\mathcal{C}_K(A, B)$ as the radius of the smallest band which when put around either curve encloses the other as well,

$$\rho(\gamma_1, \gamma_2) := \inf\{\epsilon : V_\epsilon(\gamma_1) \supset \gamma_2 \text{ and } V_\epsilon(\gamma_2) \supset \gamma_1\}. \tag{1.11}$$

ρ is the restriction to $\mathcal{C}_K(A, B)$ of the Hausdorff metric. (ρ is a pseudo-metric on the power set of $\mathcal{M}$ and a metric on the closed subsets of $\mathcal{M}$. The elements of $\mathcal{C}_K(A, B)$ are compact and thus closed).

The causal curves form the complete set of pointwise limits of the causal trips.

Proposition 1.1.22 *$\mathcal{T}(A, B)$ is dense in $\mathcal{C}(A, B)$.*

Proof: Let $\gamma \in \mathcal{C}(A, B)$.

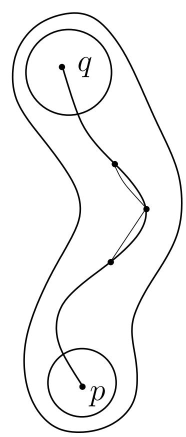


Figure 1.45: approximate a causal curve by a causal trip.

□

Mathematical preliminaries

Recall that a Cauchy sequence is not necessarily convergent. For example, consider the subspace $X = (0, 1]$ of the real line. the sequence defined by $x_n = 1/n$ is easily seen to be a Cauchy sequence in this space, but it is not convergent, since the point 0 is not a point of the space.

A metric space is sequentially compact if every sequence has a convergent subsequence.

Proposition 1.1.23 *Let (X, d) be a metric space. If a set $C \subseteq X$ is compact then it is sequentially compact.*

Proof: Let E denote the set of members of a sequence $\{x_n\}$ in C. First say E is finite. Then at least one point, say x, of E must be repeated infinitely often in the sequence $\{x_n\}$, and its occurrences form a subsequence converging to x, which is a point in C.

Now suppose E is infite. Suppose E has no limit point in C. Then for each $x \in C$, there exists $\epsilon(x)$ such that $S_{\epsilon(x)}(x)$ contains no point of E other than x. That is, $S_{\epsilon(x)}(x) \cap E = \emptyset$ or $\{x\}$. Now by compactness of C, the cover $\{S_{\epsilon(x)}(x) : x \in C\}$ has a finite subcover. But the union of any finite subcollection of the $S_{\epsilon(x)}(x)$ doesn't cover E, let alone C. This contradiction proves that E must have a limit point in C.

$\square$

A space is totally bounded if the cover

$$\{S_\epsilon(x) : x \in X\} \tag{1.12}$$

has a finite subcover for any $\epsilon > 0$.

Roughly, a metric space is complete if every series in it which tries to converge is successful, in the sense that it finds a point in the space to converge to.

Theorem 1.1.24 *Let (X, d) be a metric space. A set $C \subseteq X$ is sequentially compact if and only if it is both complete and totally bounded.*

Proof: Let C be sequentially compact. Every sequence has a convergent subsequence. Since the sequence is Cauchy, it follows it converges. We prove total boundedness by contracdiction. Suppose there is some $\epsilon > 0$ such that C is not totally bounded for ϵ. It is then possible to construct a sequence such that no term in the sequence is within ϵ of any other term in the sequence. This sequence has no Cauchy subsequences, for if it did there would be a subsequence who's terms would get close to each other. Hence there are no convergent subsequences, and hence C is not sequentially compact. We construct the desired sequence as follows: take any point p_1 in C. Choose p_2 to be any point not in $S_\epsilon(p_1)$. This is possible since C is not totally bounded. Choose p_3 to be any point not in $S_\epsilon(p_1) \cup S_\epsilon(p_2)$. Again this is possible since C is not totally bounded. Continuing in this way one constructs the desired sequence.

Now, let C be complete and totally bounded. Let E denote the set of members of a sequence $\{x_n\}$ in C. If E is finite it has a trivial convergent subsequence. Suppose, then, this is not the case. As C is totally bounded we can form a finite cover with a set of balls of radius 1. One of these balls, call it S_1 must contain an infinite subset of $\{x_i\}$. Take the first term in the subsequence, x_{i_1}, to be any term in the sequence $\{x_i\}$ that lies in

S_1. Now form a new cover of C using a finite set of balls of radius $1/2$. One of these balls, label it S_2, must contain an infinite subset of $S_1 \cap E$. Take the second term in the subsequence, x_{i_2}, to be any term in the sequence $\{x_i\}$ that lies in $S_1 \cap S_2$ and for $i_2 > i_1$. Continuing in this way, one constructs a Cauchy susequence $\{x_{i_k}\}$. Since C is complete, $\{x_{i_k}\}$ converges to a point in C. Thus C is sequentially compact.

□

Definition Given an open cover $\mathcal{U}$ of a metric space X, a fixed real number $\epsilon > 0$ is called a Lebesgue number for $\mathcal{U}$ if for any $x \in X$, there exists a set $U(x) \in \mathcal{U}$ such that $S_\epsilon(x) \subset U(x)$.

Theorem 1.1.25 *Any sequentially compact metric space (X, d) is compact.*

Proof: First we show there exists a Lebesgue number for any open cover of a sequentially compact metric space X. Suppose $\mathcal{U}$ be an open cover for which no Lebesgue number exists. Then for any integer n, there exists some point of X, say x_n, such that $S_{1/n}(x_n)$ is not contained in some U of $\mathcal{U}$. Consider a sequence of such points $\{x_n\}$. By sequentially compactness, $\{x_n\}$ has a subsequence $\{x_{n_k}\}$ converging to some point $x \in X$. Since $\mathcal{U}$ covers X, $x \in U_0$ for some $U_0 \in \mathcal{U}$. Since U_0 is open, $S_{2/m}(x) \subset U_0$ for some integer m. Now $S_{1/m}(x)$ contains x_{n_k} for all $k \geq M$. Choose $k \geq M$ so that $n_k \geq m$, and write $s = n_k$. Take y to be any point in $S_{1/s}(x_s)$, then by

$$d(x, y) \leq d(x, x_s) + d(x_s, y) \ < 1/m + 1/s \leq 2/m$$

we have $S_{1/s}(x_s) \subset S_{2/m}(x)$, and hence $S_{1/s}(x_s) \subset U_0$. This contradiction shows that there exists a Lebesgue number for $\mathcal{U}$.

Now, let $\mathcal{U}$ be an arbitrary open cover of a sequentially compact metric space X. By the preceding lemma there exists Lebesgue number ϵ for $\mathcal{U}$. As a sequentially compact metric space is totally bounded, there exists a finite cover for this Lebesgue number $\{S_\epsilon(x_i) : \{x_1, x_2, \ldots, x_n\}\}$ for X. Then each $S_\epsilon(x_i)$ is contained in some set, say U_i, in $\mathcal{U}$ by definition of Lebesgue number. Since $X \subset \cap_{i=1}^n S_\epsilon(x_i) \subset \cap_{i=1}^n U_i$, we have a finite subcover $\{U_1, U_2, \ldots, U_n\}$ of $\mathcal{U}$ for X.

□

Proposition 1.1.26 *Any closed subspace C of a compact metric space $\mathcal{T}$ is compact.*

Proof: Let $\mathcal{U}$ be any cover of C by sets open in $\mathcal{T}$. Since C is closed, $\mathcal{T} - C$ is open. $\mathcal{T} - C$ together with the collection $\mathcal{U}$ forms an open cover of $\mathcal{T}$. By the compactness of

$\mathcal{T}$, there is a finte subcover, say $\{U_1, U_2, \ldots, U_r\}$, one these being the set $\mathcal{T} - C$. The other U's provide a finite subcover of $\mathcal{U}$ for C.

$\square$

We now move onto proving $\mathcal{C}_K(A, B)$ is a compact metric space.

Proposition 1.1.27 *If K is compact and contained in G, $\mathcal{C}_K(A, B)$ is totally bounded.*

Proof: We wish to show the cover $\{S_\epsilon(\gamma_i) : \gamma_i \in \mathcal{C}_K(A, B)\}$ has a finite subcover for any $\epsilon > 0$.

Let $\epsilon < 0$ be given. Since K is compact it is totally bounded and may therefore be covered by finitely many Alexandroff neighbourhoods of diameter less than $\epsilon/2$, say $\{A_1, A_2, \ldots, A_N\}$. Consider the open sets $\{B_i\}$ formed by chains of A_i's, $A_{j_1} \cup \cdots \cup A_{j_n}$, such that they connect the sets A and B, i.e.,

$$A \cap A_{j_1} \neq \emptyset, \quad A_{j_n} \cap B \neq \emptyset, \quad A_{j_k} \cap A_{j_{k+1}} \neq \emptyset \tag{1.13}$$

and

$$I^+[A_j] \cap A_{j+2} \neq \emptyset. \tag{1.14}$$

The condition $A_{j_k} \cap A_{j_{k+1}} \neq \emptyset$ ensures there exists causal curves from one neighbourhood to the next, the condition $I^+[A_j] \cap A_{j+2} \neq \emptyset$ guarantees the chain B_i contains causal curves from A to B, by excluding cases such as in fig (1.1.7).

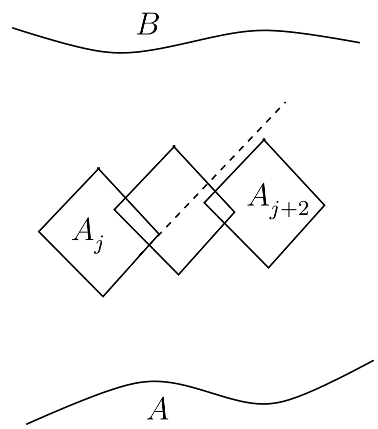


Figure 1.46: Imposition of the condition $I^+[A_j] \cap A_{j+2} \neq \emptyset$ avoids cases such as the above.

As there are a finite number of possible unions of the finite collection of sets A_i's, there are finitely many B_j's. Now, set

$$B_i' := \{\gamma \in \mathcal{C}_K(A,B) : \gamma \subset B_i\}. \tag{1.15}$$

By construction, we have $B_i' \neq \emptyset$ for each i; since every γ in $\mathcal{C}_K(A,B)$ is contained in one of the B_i's we have $\cup\{B_i'\} = \mathcal{C}_K(A,B)$; $\gamma \in B_i'$ implies $B_i' \subset S_\epsilon(\gamma)$.

....

Then the collection $\{S_\epsilon(\gamma_i)\}$ so determined is a finite cover of $\mathcal{C}_K(A,B)$.

□

Proposition 1.1.28 *$\mathcal{C}_K(A,B)$ is complete.*

Proof: Let $\{\gamma_i\} \subseteq \mathcal{C}_K(A,B)$ be a Cauchy sequence. Cover K by finitely many Alexandroff neighbourhoods each of which is contained in a simple region whose closure lies in G. Let $\{x_i\}$ be a sequence on A formed by intersections of the γ_i. As $\{x_i\}$ is Cauchy and A is compact (being a closed subset of the compact set K) x_i converges to a point $x_0 \in A$. Let A_1 be an Alexandroff neighbourhood in the cover which contains x_0. Introduce Minkowskian coordinates in $\overline{A_1}$ centered at x_0. Since $x_0 \in A_1^0$ there is some interval $[0, t_1)$ such that for each t satisfying $0 \leq t < t_1$, the compact 3-surface $\{t = \text{constant}\} \cap \overline{A_1}$ is intersected by infinitely many γ_i in points $x_i(t)$ in A_1. For each such t, $\{x_i(t)\}$ is a Cauchy sequence in the 3-surface and thus converges to a limit which we denote $x_0(t)$. It is clear that $0 < t' < t'' < t_1$ implies $x_0(t') < x_0(t'')$, and that $x_0(t)$ is continuous

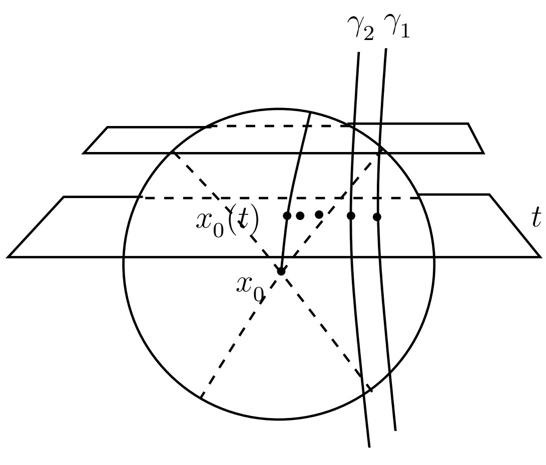


Figure 1.47: The geodesic λ.

....
....

Repeat the process, taking the Minkowskian coordinates in A_2 to have the value $(t_1, 0, 0, 0)$ at $x_0(t_1)$. The process is is finite since the curve $x_0(t)$ so constructed is a causal curve and cannot re-enter any Alexandroff neighbourhood it has left. Therefore $x_0(t)$ must strike B, since B is closed.

$\square$

Theorem 1.1.29 *$\mathcal{C}_K(A, B)$ is a compact metric space.*

Proof: $\mathcal{C}_K(A, B)$ is totally bounded and complete and hence compact.

$\square$

Corollary 1.1.30 *Let S be closed, achronal, and strongly causal. Let $x \in intD(S)$. Then $\mathcal{C}_K(A, B)$ is compact*

Proof: $\mathcal{C}_K(x, S) \subset K \equiv J^-(x) \cap J^+[S]$, which is compact and contained in a strongly causal region.

$\square$

Corollary 1.1.31 *Let S be closed, achronal, and strongly causal. Let $a, b \in intD(S)$. Then $\mathcal{C}_K(A, B)$ is compact*

$\square$

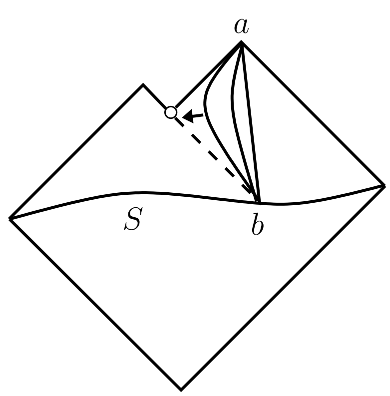


Figure 1.48: The sequence is a Cauchy sequence in $\mathcal{C}(a, b)$, but it is not convergent, since there is a point missing.

Definition A spacetime $\mathcal{M}$ is said to be globally hyperbolic if and only if $\mathcal{M}$ is strongly causal and $\mathcal{C}(a, b)$ is comapct, for all $a, b \in \mathcal{M}$.

Definition A Cauchy hypersurface for $\mathcal{M}$ is a closed achronal set S for which $D(S) = \mathcal{M}$.

Proposition 1.1.32 *$\mathcal{M}$ has a Cauchy hypersurface implies the spacetime $\mathcal{M}$ is globally hyperbolic.*

Proof: Corollary 1.1.31.

$\square$

Definition A partial Cauchy surface S is defined as an acausal set without edge.

A partial Cauchy surface is called a Cauchy surface or a global Cauchy surface if $D(S) = \mathcal{M}$.

Let S be a partial Cauchy suurface. Then let $\mathcal{N} = D^+(S) \cup D^-(S)$. Even though $\mathcal{M}$ may not be globally hyperbolic and S is not a Cauchy surface, the region int$D^+(S)$ or int$D^-(S)$ is globally hyperbolic in its own right as the surface S serves as a Cauchy surface for the manifold int$\mathcal{N}$.

An important property of globally hyperbolic spacetimes, or any globally hyperbolic subset of $\mathcal{M}$, which is relevant for the singularity theorems is the existence of maximum length non-spacelike geodesics between pairs of causally related points. This we will see in the next section.

If one removes a point from Minkowski spacetime, the resulting spacetime $\mathcal{M}$ admits no Cauchy surface and is therefore not gloabally hyperbolic.

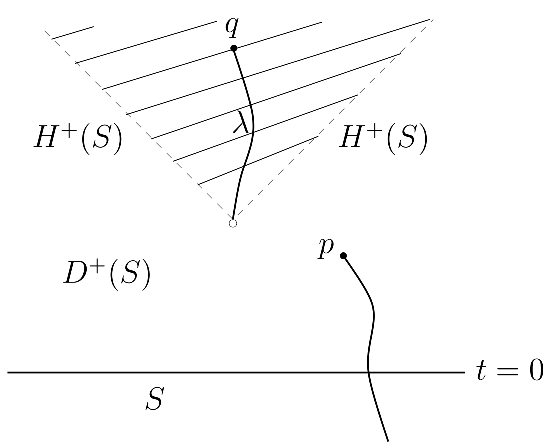


Figure 1.49: Minkowski space time with a point removed is not globally hyperbolic. The point q is not in $D^+(S)$ as there are non-spacelike curves like λ which do not meet S in the past.

The requirement that a spacetime posses a Cauchy hypersurface is very strong as we will see, such spacetimes have very regular global behaviour and all the pathological features in previous sections are excluded. It has been shown that a globally hyperblic spacetime is homeomorphic to $\mathbb{R} \times S$ where S is a three-dimensional submanifold and for each $t \in \mathbb{R}$, the spacelike hypersurface S_t of constant time is a Cauchy surface for the spacetime. A globally hyperblic spacetime has a unique topological structure and admits no topology change.

There are spacetimes homeomorphic to $\mathbb{R} \times S$ which do not have Cauchy hypersurfaces.

In **quantum gravity**, however, different kinds of topologies and, in particular, topology changes are conceivable. Perhaps it is possible to construct the quantum theory of the gravitational field based on the calssical assumption that $\mathcal{M} = \mathbb{R} \times \sigma$ and then lift this restriction in the quantum theory.

1.1.8 Conjugate Points

Mathematical preliminaries

Proposition 1.1.33 *If $f : \mathcal{T}_1 \to \mathcal{T}_2$ is a continuous map of topological spaces and $\mathcal{T}_1$ is compact then $f(\mathcal{T}_1)$ is compact.*

Proof: Suppose that $\mathcal{U}$ is an open cover of $f(\mathcal{T}_1)$ by sets open in $\mathcal{T}_2$. By continuity of f, $f^{-1}(U)$ is open in $\mathcal{T}_1$ for every U in $\mathcal{U}$. By compactness of $\mathcal{T}_1$, there is a finite subcover $\{f^{-1}(U_1), f^{-1}(U_2), \ldots, f^{-1}(U_r)\}$, and it is easily seen that $\{U_1, U_2, \ldots, U_r\}$ is then a finite subcover of $\mathcal{U}$ for $f(\mathcal{T}_1)$.

□

Corollary 1.1.34 *Compactness is a toplogical property: if $\mathcal{T}$ and $\mathcal{T}_2$ are homeomorphic then $\mathcal{T}_1$ is compact if and only if $\mathcal{T}_2$ is compact.*

Corollary 1.1.35 *Any continuous map from a compact space to a metric space M is bounded.*

Proof: We show that any compact subset C of a metric space is bounded, the corollary then follows from proposition 1.1.33. Let x_0 be any point of M. Then $C \subset \cup_{n=1}^{\infty} B_n(x_0)$, since for any $x \in C$, $d(x, x_0) < n$ for some integer n. Thus the collection $\{B_n(x_0) : n \in \mathbb{N}\}$ form an open cover for C, and since C is compact, there is a finite subcover, say $\{B_{n_1}, B_{n_2}, \ldots, B_{n_r}\}$. Let $n = \mathrm{Max}\{n_1, n_2, \ldots, n_r\}$. Then $C \subset \cup_{i=1}^{r} B_{n_i}(x_0) = B_n(x_0)$.

□

Definition If f is a real-valued function on a metric space E and $p_0 \in E$ we say that f attains a maximum at p_0 if $f(p_0) \geq f(p)$ for all $p \in E$, and that f attains a minimum at p_0 if $f(p_0) \leq f(p)$ for all $p \in E$.

Corollary 1.1.36 *Any continuous real-valued function on a nonempty compact metric space attains a maximum at some point, and also attains a minimum at some point.*

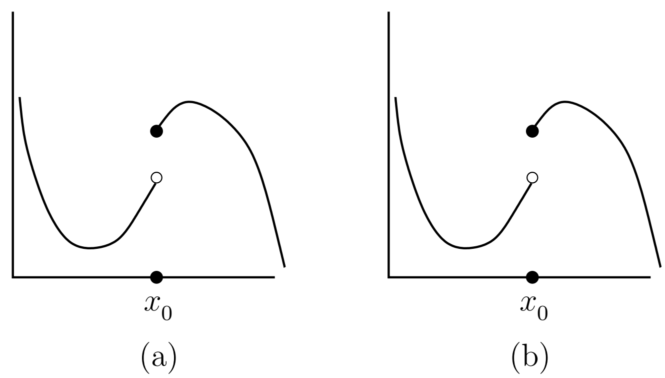


Figure 1.50: (a) An upper semi-continuous function. (b) A lower semi-continuous function.

Definition A real-valued function f defined on a topological space $\mathcal{T}$ is said to be **upper semicontinuous at a point** $x_0 \in \mathcal{T}$ if, given any $\epsilon > 0$, there exists a neighbourhood of x_0 in which $f(x) < f(x_0) + \epsilon$. Equivalently, this can be expressed as

$$\lim_{x \to x_0} \sup f(x) \leq f(x_0).$$

The function f is called **upper semicontinuous** if it is upper semicontinuous at every point of its domain. Similarly, f is said to be lower semicontinuous at x_0 if, given any $\epsilon > 0$, there exists a neighbourhood of x_0 in which $f(x) > f(x_0) - \epsilon$.

Lemma 1.1.37 *Any compact subspace C of a Hausdorff topological space X is closed in X.*

Proof: Let C be a compact subspace of a Hausdorff topological space X and let $a \in X - C$. We shall prove that there exist an open set U_a such that $a \in U_a \subset X - C$. Then $X - C = \cup_{a \in X - C} U_a$, so $X - C$ is open and hence C is closed.

Since X is Hausdorff, we can choose open subsets U_x and $V(x)$ of X such that $a \in U_x$, $x \in V(x)$, and U_x and $V(x)$ are disjoint. Choosing such subsets U_x and $V(x)$ for every point $x \in C$, we see that the sets $V(x)$ form an open cover of C. Since we assume C is a compact topological space, it follows that C is the union of the open sets $V_{x_1}, \dots, V_{x_r}$ for some finite collection of points $x_1, \dots, x_r \in C$. Let $U_a = \cap_{i=1}^r U_{x_i}$. Then U_a, as a finite intersection of open sets, is open. Also, $a \in U_a$ since $a \in U_{x_i}$ for every $i = 1, 2, \dots, r$. Finally, if $b \in U_a$ then for any $i = 1, 2, \dots, r$, $b \in U_{x_i}$ and hence $b \notin V(x_i)$, so $b \notin C$, since $C \subset \cup_{i=1}^r V(x_i)$. Thus $U_a \subset X - C$.

□

Theorem 1.1.38 *Let $f : M \to \mathbb{R} \cup \{+\infty\}$ be a lower semicontinuous function on the Hausdorff space M. Suppose that there is a real number r such that*

a) $[f \leq r] = \{x \in M : f(x) \leq r\} \neq \emptyset$ and

b) $[f \leq r]$ is sequentially compact.

Then there is a minimizing point x_0 for f on M:

$$f(x_0) = \inf f(x).$$

Proof: We begin by showing indirectly that f is lower bounded. If f is not bounded from below there is a sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $f(x_n) < -n$ for all $n \in \mathbb{N}$. For sufficiently large n the elements of the sequence belong to the set $[f \leq r]$, hence there is a subsequence $y_j = x_{n(j)}$ which converges to a point $y \in M$. Since f is lower semicontinuous we know $f(y) \leq lim\ inf_{j\to\infty} f(y_j)$, a contradition since $f(y_j) < -n(j) \to -\infty$. We conclude that f bounded from below and thus has a finite infimum:

$$-\infty < I = I(f, M) = \inf_{x\in M} f(x) \leq r.$$

Threfore there is a minimizing sequence $\{x_n\}_{n\in N}$ whose elements belong to $[f \leq r]$ for sufficiently large n. Since $[f \leq r]$ is sequentially compact there is again a subsequence $y_j = x_{n(j)}$ which converges to a point $x_0 \in [f \leq r]$. Since f is lower semmicontinuous we conclude

$$I \leq f(x_0) \leq \liminf_{j\to\infty} f(y_j) = I.$$

□

Theorem 1.1.39 *A finite upper semicontinuous function f defined on a compact topological space $\mathcal{T}$ is bounded from above.*

Proof:

□

Theorem 1.1.40 *A finite upper semicontinuous function f defined on a compact topological space $\mathcal{T}$ achieves its least upper bound on $\mathcal{T}$.*

Proof:

$\square$

Conjugate points

We consider a congruence of null geodesics with affine parameter v and tangent V^a. The expansion of the geodesics may be defined as $\theta = D_a V^a$, where D_a is the covariant derivative. The propagation equation for θ leads to this inequality:

$$\frac{d\theta}{dv} \leq -\frac{1}{2}\theta^2 - R_{ab}V^a V^b. \tag{1.16}$$

Suppose that (i) $R_{ab}V^aV^b \geq 0$ for all null vectors V^a (this is called the null convergence condition), (ii) the expansion, θ, is negative at some point $v = v_0$ on a geodesic γ, and (iii) γ is complete in the direction of increasing v (i.e., γ is defined for all $v \geq v_0$). Then $\theta \to -\infty$ along γ a finite affine parameter distance from v_0.

maximum and minimum of functions of one variable. If $f'(a) = 0$, then we know that $f(x)$ is stationary at $x = a$. We then go on to study the sign of $f''(a)$. If it is positive, then $f(x)$ has a minimum value at $x = a$, if it is negative, then $f(x)$ has a maximum value at $x = a$, and if $f''(a) = 0$.

$\tau(\alpha) := \int_a^b f(\alpha, t)dt$ where $f = (-t^a t_a)^{1/2}$. recall

$$\left.\frac{d\tau}{d\alpha}\right|_{\alpha=0} = 0. \tag{1.17}$$

for independent of the variation function η.

Interested in the sign of the **second variation** of the length function,

$$\left.\frac{d^2\tau}{d\alpha^2}\right|_{\alpha=0} \tag{1.18}$$

For it to be a maximum the sign of must be negative independently of the choice of the deviation function $\eta(x)$.

conjugate point if

$$\left.\frac{d^2\tau}{d\alpha^2}\right|_{\alpha=0} = 0. \tag{1.19}$$

Proposition 1.1.41 *Let A and B be achronal topological manifolds in a strongly causal spacetime $\mathcal{M}$. Then the length function is upper semi-continuous on $\mathcal{T}(A, B)$.*

Proof: Let $\gamma_k \to \gamma_0$ in $\mathcal{T}(A, B)$. We wish to show that $\limsup\{L(\gamma_k)\} \leq L(\gamma_0)$ Choose a subsequence of $\{\gamma_k\}$ and relabel so that $L(\gamma_k) \to \limsup\{L(\gamma_k)\}$. It is sufficient to prove the proposition in the case that γ_0 is a geodesic and $\mathcal{T}(A, B) \subset N$, a simple region.

Let x_k, y_k be intersections of γ_k with A and B respectively; and let μ_k be the unique causal geodesic $[x_k y_k]$ lying in N. Since $\gamma_k \to \gamma_0$, we have $x_k \to x_0$ and $y_k \to y_0$.

Also, since L is a continuous function on geodesics joinig two continuous surfacse, we have $L(\mu_k) \to L(\gamma_0)$. By local maximality of geodesics, for every k we have $L(\gamma_k) \leq L(\mu_k)$. Thus $\lim\{L(\gamma_k)\} = a \leq \lim\{L(\mu_k)\}$, and so the superior limit of the original sequence is less than or equal to $L(\gamma_0)$.

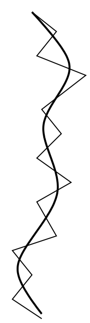


Figure 1.51: Two points joined by a timelike curve can be connected by a broken null geodesic.

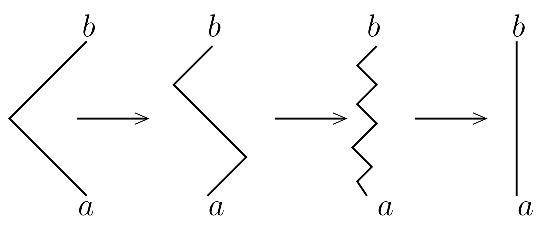


Figure 1.52: A sequence of null curves may converge to a timelike curve.

Two points joined by a timelike curve can be connected by a broken null geodesic, hence, the length function is not a lower continous function.

□

Theorem 1.1.42 *There is a curve $\gamma \in \mathcal{C}_K(A, B)$ with maximum length.*

An upper semi-continuous function on a compact set attains its maximum.

□

Note that γ needn't be unique and needn't be a geodesic.

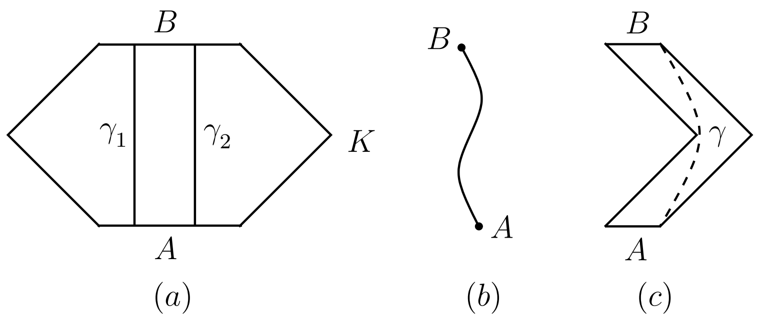


Figure 1.53: If A and B are parallel in Minkowskia spacetime then γ_1 and γ_2 are maximal. (b) Here there is only one element in $\mathcal{C}_K(A, B)$ which is necessarily maximal. It is not a geoesic. (c) Here the maximal eleement is a trip.

1.1.9 Timelike Congruences

We are considering a smooth congruence of timelike geodesics. The geodesics are parameterized by proper tiem τ...

....so that the vector fields V^a of tangents is normalised to unit length: $V^a V_a = -1$. This normalisation condition implies

$$V^a \nabla_c V_a = 0. \tag{1.20}$$

The geodesic equation is

$$V^c \nabla_c V_a = 0. \tag{1.21}$$

We define the transverse metric or "spatial metric" by

$$h_{ab} = g_{ab} + V_a V_b. \tag{1.22}$$

It is easily seen that the four tensor

$$h^a{}_b = g^{ac}h_{cb} = \delta^a{}_b + V^a V_b \tag{1.23}$$

is the projection operator onto the subspace of the tangent space perpendicular to V^a as

$$h^a{}_b V^b = (\delta^a{}_b + V^a V_b)V^b = 0 \quad \text{and if} \quad W^a V_a \quad \text{then} \quad h^a{}_b W^b = W^a.$$

We have

$$\begin{aligned} V^c \nabla_c h_{ab} &= V^c \nabla_c (g_{ab} + V_a V_b) = 0 & (1.24) \\ V^c \nabla_c h^{ab} &= V^c \nabla_c (g^{ab} + V^a V^b) = 0 & (1.25) \end{aligned}$$

where we have used $\nabla_c g_{ab} = 0$ nd the geodesic equation (1.21).

The expansion, twist and shear of the field

The expansion, shear and twist of the field are defined by

$$\theta_{ab} = h_a{}^c h_b{}^d \nabla_{(d} V_{c)} \tag{1.26}$$

$$\begin{aligned} \theta &= h^{ab} h_a{}^c h_b{}^d \nabla_{(d} V_{c)} \\ &= h^{cd} \nabla_{(d} \eta_{c)} & (1.27) \end{aligned}$$

$$\omega_{ab} = h_a{}^c h_b{}^d \nabla_{[d} V_{c]} \tag{1.28}$$

$$\sigma_{ab} = h_a{}^c h_b{}^d \nabla_{(d} V_{c)} - \frac{1}{3} \theta h_{ab} \tag{1.29}$$

$$\begin{aligned} \theta &= h^{ab} h_a{}^c h_b{}^d \nabla_{(d} V_{c)} \\ &= h^{cd} \nabla_d V_c & (1.30) \end{aligned}$$

The Raychaudhuri equation and evolution equations for shear and vorticity

Note that $\nabla_b V_a$ can be decomposed as

$$\nabla_b V_a = \frac{1}{3}\theta h_{ab} + \sigma_{ab} + \omega_{ab} \tag{1.31}$$

This decomposition of the gradient of the velocity vector V^a is directly analogous to the decomposition of the gradient of the fluid velocity vector in Newtonian hydrodynamics.

$$\begin{aligned} V^c\nabla_c(\nabla_b V_a) &= V^c\nabla_b\nabla_c V_d + V^c(\nabla_c\nabla_b - \nabla_b\nabla_c)V_d \\ &= V^c\nabla_b\nabla_c V_d + R_{cba}{}^d V^c V_d \\ &= \nabla_b(V^c\nabla_c V_d) - (\nabla_b V^c)(\nabla_c V_a) + R_{cba}{}^d V^c V_d \\ &= -(\nabla^c V_b)(\nabla_a V_c) + R_{cba}{}^d V^c V_d \end{aligned} \tag{1.32}$$

where we used the geodesic equation $V^c\nabla_c V_d = 0$.

Derivation of Raychaudhuri equation

We now derive the Raychaudhuri equation, i.e. the evolution eataion for θ. In the following it should be kept in mind that θ_{ab} and σ_{ab} are symmetric while ω_{ab} is anitsymmetric. Taking the trace over a and b and using (1.31) we obtain,

$$\begin{aligned} V^c\nabla_c\theta &= -\left(\frac{1}{3}\theta h^{ca} + \sigma^{ca} + \omega^{ca}\right)\left(\frac{1}{3}\theta h_{ac} + \sigma_{ac} + \omega_{ac}\right) - R_{cd}V^cV^d \\ &= -\frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} + \omega^{ab}\omega_{ab} - R_{cd}V^cV^d \end{aligned} \tag{1.33}$$

and so

$$\frac{d\theta}{ds} = -\frac{1}{3}\theta^2 - \sigma_{ab}\sigma^{ab} - \omega^{ab}\omega_{ab} - R_{cd}V^cV^d \tag{1.34}$$

where $\omega_{ab}\omega^{ab} \geq 0$ and $\sigma_{ab}\sigma^{ab} \geq 0$. This is the Raychaudhuri equation and is of great importance for the singularity theorem. Note that vorticity induces expansion which is in analogy with centrifugal force while shear induces cotraction.

We turn to the corresponding equation for the shear,

The trace-free, symmetric part of (1.32)

In general the trace-free part of a tensor X_{ab} is

$$X_{ab} - \frac{1}{3}h_{ab}h^{cd}X_{cd}$$

and the trace-free, symmetric part of a tensor X_{ab} is

$$X_{(ab)} - \frac{1}{3}h_{ab}h^{cd}X_{cd},$$

or

$$\left(\delta^c_{(a}\delta^d_{b)} - \frac{1}{3}h_{ab}h^{cd}\right)X_{cd}.$$

We wish to obtain the trace-free, symmetric part of the identity (1.32). First consider the trace-free, symmetric part of the LHS of (1.32),

$$V^c\nabla_c(\nabla_{(b}V_{a)}) - \frac{1}{3}h_{ab}h^{ef}V^c\nabla_c(\nabla_f V_e). \tag{1.35}$$

Recall we have $V^c\nabla_c h_{ab} = 0$ and $V^c\nabla_c h^{ab} = 0$. Therefore

$$h_{ab}h^{ef}V^c\nabla_c(\nabla_f V_e) = V^c\nabla_c(h_{ab}h^{ef}\nabla_f V_e).$$

Using this in (1.35) gives

$$\begin{aligned} V^c\nabla_c(\nabla_{(b}V_{a)}) - \frac{1}{3}h_{ab}h^{ef}V^c\nabla_c(\nabla_f V_e) &= V^c\nabla_c\left(\nabla_{(b}V_{a)} - \frac{1}{3}h_{ab}h^{ef}\nabla_f V_e\right) \\ &\equiv V^c\nabla_c\sigma_{ab}. \end{aligned} \tag{1.36}$$

We now consider the symmetric, trace-free part of the RHS of (1.32), i.e. the trace-free, symmetric part of $-(\nabla_b V^c)(\nabla_c V_a) + R_{cbad}V^cV^d$:

$$-\left[(\nabla_c V_{(a})(\nabla_{b)}V^c) - \frac{1}{3}h_{ab}h^{ef}(\nabla_c V_e)(\nabla_f V^c)\right] + R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d.$$

We substitute (1.31) into this and obtain

$$
\begin{aligned}
&= -\frac{1}{2}\left[\left(\frac{1}{3}\theta h_{ac} + \sigma_{ac} + \omega_{ac}\right)\left(\frac{1}{3}\theta h^c{}_b + \sigma^c{}_b + \omega^c{}_b\right) + a \leftrightarrow b\right] \\
&+\frac{1}{3}h_{ab}h^{ef}\left(\frac{1}{3}\theta h_{ec} + \sigma_{ec} + \omega_{ec}\right)\left(\frac{1}{3}\theta h^c{}_f + \sigma^c{}_f + \omega^c{}_f\right) \\
&+R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d \\
&= -\left(\frac{1}{9}\theta^2 h_{ab} + \frac{2}{3}\theta\sigma_{ab} + \sigma_{ac}\sigma^c{}_b + \omega_{ac}\omega^c{}_b\right) \\
&+\frac{1}{3}h_{ab}h^{ef}\left(\frac{1}{3}\theta^2 + \frac{2}{3}\theta h^{ef}\sigma_{ef} + \frac{2}{3}\theta h^{ef}\omega_{ef} + \sigma_{ec}\sigma^{ce} + 2\sigma_{ec}\omega^{ce} + \omega_{ec}\omega^{ce}\right) \\
&+R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d \\
&= -\left(\frac{1}{9}\theta^2 h_{ab} + \frac{2}{3}\theta\sigma_{ab} + \sigma_{ac}\sigma^c{}_b + \omega_{ac}\omega^c{}_b\right) + \frac{1}{3}h_{ab}\left(\frac{1}{3}\theta^2 + \sigma_{cd}\sigma^{cd} - \omega_{cd}\omega^{cd}\right) \\
&+R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d \\
&= -\frac{2}{3}\theta\sigma_{ab} - \sigma_{ac}\sigma^c{}_b - \omega_{ac}\omega^c{}_b + \frac{1}{3}h_{ab}\left(\sigma_{cd}\sigma^{cd} - \omega_{cd}\omega^{cd}\right) \\
&+R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d \qquad (1.37)
\end{aligned}
$$

where we have used the anti-symmetry of ω_{ab}. Using this together with (1.36), we have

$$
\begin{aligned}
V^c\nabla_c\sigma_{ab} = \frac{d\sigma_{ab}}{d\tau} &= -\frac{2}{3}\theta\sigma_{ab} - \sigma_{ac}\sigma^c{}_b - \omega_{ac}\omega^c{}_b + \frac{1}{3}h_{ab}\left(\sigma_{cd}\sigma^{cd} - \omega_{cd}\omega^{cd}\right) \\
&\quad +R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cfed}V^cV^d \qquad (1.38)
\end{aligned}
$$

We turn our attention to the second line on the RHS.

Symmetric trace-free part of $R_{cbad}V^cV^d$

The symmetric trace-free of $R_{cbad}V^cV^d$ is

$$
R_{c(ba)d}V^cV^d - \frac{1}{3}h_{ab}h^{ef}R_{cefd}V^cV^d = \left(\delta^e_{(a}\delta^f_{b)} - \frac{1}{3}h_{ab}h^{ef}\right)R_{cefd}V^cV^d. \qquad (1.39)
$$

In the following we will need

$$
\begin{aligned}
R_{cabd}V^cV^d &= R_{dabc}V^cV^d \\
&= R_{bcda}V^cV^d \\
&= R_{cbad}V^cV^d \qquad (1.40)
\end{aligned}
$$

where in the first line we have used that c and d are dummy indices, and after that we have used symmetries of the Riemann tensor. This implies $R_{c(ba)d}V^cV^d = R_{cbad}V^cV^d$ and $R_{c[ba]d}V^cV^d = 0$. So that the RHS of (1.39) can be written,

$$
\left(\delta_a^e\delta_b^f - \frac{1}{3}h_{ab}h^{ef}\right) R_{cefd}V^cV^d. \qquad (1.41)
$$

We use the decomposition of the Riemann tensor:

$$
R_{cbad} = C_{cbad} - g_{c[d}R_{a]b} - g_{b[a}R_{d]c} - \frac{1}{3}Rg_{c[a}g_{d]b}. \qquad (1.42)
$$

Note from this one can check that the Weyl tensor has the same symmetries as the Riemann tensor, but it also has the property that it vanishes upon contraction of any two of its indices.

We take (1.41) and decompose it as follows

$$
\left(\delta_a^e\delta_b^f - \frac{1}{3}h_{ab}h^{ef}\right)(C_{cefd} + [R_{cefd} - C_{cefd}])V^cV^d. \qquad (1.43)
$$

First note

$$
\begin{aligned}
\left(\delta_a^e\delta_b^f - \frac{1}{3}h_{ab}h^{ef}\right) C_{cefd}V^cV^d &= \left(\delta_a^e\delta_b^f - \frac{1}{3}h_{ab}(g^{ef} + V^eV^f)\right) C_{cefd}V^cV^d \\
&= C_{cbad}V^cV^d \qquad (1.44)
\end{aligned}
$$

where we have used $C_{cefd}V^cV^dV^eV^f \equiv 0$ and that the Weyl tensor is trace-free in any two indices.

Using $\delta_a^b = h_a{}^b - V_aV^b$ and the anti-symmetry in d and f, and the anti-symmetry in c and e of $R_{cefd} - C_{cefd}$, we obtain

$$\begin{aligned}
\left(\delta_a^f\delta_b^e-\frac{1}{3}h_{ab}h^{ef}\right)R_{cefd}V^cV^d &= C_{cbad}V^cV^d+\left(\delta_a^f\delta_b^e-\frac{1}{3}h_{ab}h^{ef}\right)(R_{cbad}-C_{cbad})V^cV^d \\
&= C_{cbad}V^cV^d+\left((h_a{}^f-V_aV^f)(h_b{}^e-V_bV^e)-\frac{1}{3}h_{ab}h^{ef}\right) \\
&\quad \left(R_{cefd}-C_{cefd}\right)V^cV^d \\
&= C_{cbad}V^cV^d+\left(h_a{}^fh_b{}^e-\frac{1}{3}h_{ab}h^{ef}\right)\left(R_{cefd}-C_{cefd}\right)V^cV^d
\end{aligned} \tag{1.45}$$

Now using (1.42) we have

$$(R_{cbad}-C_{cbad})V^cV^d=-g_{c[d}R_{a]b}V^cV^d-g_{b[a}R_{d]c}V^cV^d-\frac{1}{3}Rg_{c[a}g_{d]b}V^cV^d.$$

For the individual terms on the RHS we get

$$\begin{aligned}
-g_{c[d}R_{a]b}V^cV^d &= -\frac{1}{2}[g_{cd}R_{ab}-g_{ca}R_{db}]V^cV^d \\
&= \frac{1}{2}[R_{ab}+V_aR_{bd}V^d] \\
-g_{b[a}R_{d]c}V^cV^d &= -\frac{1}{2}[g_{ab}R_{cd}-g_{bd}R_{ac}]V^cV^d \\
&= \frac{1}{2}[-g_{ab}R_{cd}V^cV^d+V_bR_{ac}V^c] \\
-\frac{1}{3}Rg_{c[a}g_{d]b}V^cV^d &= -\frac{1}{2}\cdot\frac{1}{3}R[g_{ca}g_{db}-g_{cd}g_{ab}]V^cV^d \\
&= -\frac{1}{2}\cdot\frac{1}{3}R[V_aV_b+g_{ab}] \\
&= -\frac{1}{2}\cdot\frac{1}{3}Rh_{ab}.
\end{aligned}$$

Bringing these results together we have

$$(R_{cbad}-C_{cbad})V^cV^d=\frac{1}{2}R_{ab}-\frac{1}{2}g_{ab}R_{cd}V^cV^d+V_{(a}R_{b)d}V^d-\frac{1}{2}\cdot\frac{1}{3}Rh_{ab} \tag{1.46}$$

Substituting this into (1.45) gives

$$
\begin{aligned}
\left(\delta^f_a\delta^e_b - \frac{1}{3}h_{ab}h^{ef}\right)R_{cefd}V^cV^d &= C_{cbad}V^cV^d + \left(h_a{}^f h_b{}^e - \frac{1}{3}h_{ab}h^{ef}\right) \\
&\quad \left(\frac{1}{2}R_{fe} - \frac{1}{2}g_{fe}R_{cd}V^cV^d + V_{(f}R_{e)d}V^d - \frac{1}{2}\cdot\frac{1}{3}Rh_{fe}\right) \\
&= C_{cbad}V^cV^d + \frac{1}{2}h_a{}^e h_b{}^f R_{ef} - \frac{1}{2}\cdot\frac{1}{3}h_{ab}h^{ef}R_{ef} + \\
&- \frac{1}{2}\left(h_a{}^f h_b{}^e g_{fe} - \frac{1}{3}h_{ab}h^{ef}g_{fe}\right)R_{cd}V^cV^d \\
&+ \underbrace{h_a{}^f h_b{}^e V_{(f}R_{e)d}V^d}_{=0} - \frac{1}{3}\underbrace{h_{ab}h^{ef}V_{(f}R_{e)d}V^d}_{=0} \\
&- \frac{1}{2}\cdot\frac{1}{3}h_a{}^f h_b{}^e Rh_{fe} + \frac{1}{2}\cdot\frac{1}{3}\cdot\frac{1}{3}h_{ab}h^{ef}Rh_{fe}.
\end{aligned}
\tag{1.47}
$$

It is easy to check

$$
h_a{}^f h_{bf} = h_{ab} \qquad h_a{}^f h_b{}^e h_{fe} = h_{ab} \qquad h^{ef}h_{fe} = 3 \qquad h^{ef}g_{fe} = h^{ef}(h_{fe} - V_f V_e) = 3.
$$

Substituting these into (1.47) gives

$$
\left(\delta^f_{(a}\delta^e_{b)} - \frac{1}{3}h_{ab}h^{ef}\right)R_{cefd}V^cV^d = C_{cbad}V^cV^d + \frac{1}{2}\left(h_a{}^c h_b{}^d R_{cd} - \frac{1}{3}h_{ab}h^{cd}R_{cd}\right). \tag{1.48}
$$

Final equation for the evolution of the shear tensor

And so (1.38) becomes

$$
\begin{aligned}
\frac{d\sigma_{ab}}{ds} &= -\frac{2}{3}\theta\sigma_{ab} - \sigma_{ac}\sigma^c{}_b - \omega_{ac}\omega^c{}_b + \frac{1}{3}h_{ab}(\sigma_{cd}\sigma^{cd} - \omega^{cd}\omega_{cd}) \\
&\quad + C_{cbad}V^cV^d + \frac{1}{2}\left(h_{ac}h_{bd}R^{cd} - \frac{1}{3}h_{ab}h_{cd}R^{cd}\right).
\end{aligned}
\tag{1.49}
$$

Since the Wely tensor is trace-free is does not appear in the equation in the expansion equation(1.34). however as the terms $-2\sigma^2$ $(= -2\sigma^{ab}\sigma_{ab})$ occurs on the right-hand side of the expansion equation, the Weyl tensor induces convergence indirectly by inducing shear.

The Weyl tensor is the part of the Reiemann tensor not depending on the Ricci tensor. From the Einstein equations,

$$R_{ab} - \frac{1}{2}g_{ab}R + \Lambda g_{ab} = 8\pi T_{ab}.$$

we see that it is the Rici tensor, R_{ab}, is determined locally by the matter distributionis. Thus the Weyl tensor is the part of the curvature which is not determined locally by the matter distribution.

Derivative of vorticity

And from the anti-symmetric part of (1.32) yields

$$\begin{aligned}
V^c\nabla_c\omega_{ab} &= V^c\nabla_c\big(\nabla_{[b}V_{a]}\big) \\
&= -(\nabla_c V_{[a})(\nabla_{b]}V^c) + R_{c[ba]d}V^cV^d \\
&= -(\nabla_c V_{[a})(\nabla_{b]}V^c) \\
&= -\frac{1}{2}[(\nabla_c V_a)(\nabla_b V^c) - a \leftrightarrow b] \\
&= -\frac{1}{2}\left[\left(\frac{1}{9}\theta^2 h_{ab} + \frac{2}{3}\theta\sigma_{ab} + \frac{2}{3}\theta\omega_{ab} + \sigma_{ac}\sigma^c{}_b + 2\sigma^c{}_b\omega_{ac} + \omega_{ac}\omega^c{}_b\right) - a \leftrightarrow b\right] \\
&= -\frac{2}{3}\theta\omega_{ab} - 2\sigma^c{}_{[b}\omega_{a]c}
\end{aligned} \tag{1.50}$$

and so

$$\frac{d\omega_{ab}}{ds} = -\frac{2}{3}\theta\omega_{ab} - 2\sigma^c{}_{[b}\omega_{a]c}. \tag{1.51}$$

Conjugate points

A pair of points $p, q \in \gamma$ if there is a jacobi field η^a which is not identically zero but vanishes at both p and q. Or p and q are conjugate if an infinitesimally nearby geodesic intersects γ at both p and q.

Equation of geodesic deviation - the Jacobi equation

Let γ be a timelike geodesic with tangent V^a. Consider a congruence of timelike geodesics passing through p.

Write

$$x^a = x^a(\tau, \nu) \tag{1.52}$$

where τ is the proper time along the geodesic C_1 and ν paramaterises a curve connecdting the geodesic C_2.

$$V^a = \frac{dx^a}{d\tau} \tag{1.53}$$

and

$$\xi^a = \frac{dx^a}{d\nu}. \tag{1.54}$$

So V^a is the tangent vector to the timelike geodesic at each point along C_1 and ξ^a is a connecting vector connecting the neighbouring curves.

Now

$$\begin{aligned}
[V, \xi]^a &:= V^b \partial_b \xi^a - \xi^b \partial_b V^a \\
&= \frac{dx^a}{d\tau} \frac{\partial}{\partial x^b} \left(\frac{dx^a}{d\nu} \right) - \frac{dx^a}{d\nu} \frac{\partial}{\partial x^b} \left(\frac{dx^a}{d\tau} \right) \\
&= \frac{d^2 x^a}{d\tau d\nu} - \frac{d^2 x^a}{d\nu d\tau} \\
&= 0.
\end{aligned} \tag{1.55}$$

We can replace partial derivatives with covariant ones:

$$\begin{aligned}
0 &= V^b \partial_b \xi^a - \xi^b \partial_b V^a \\
&= V^b \partial_b (\xi^a + \Gamma^a_{bc} \xi^c) - \xi^b (\partial_b V^a + \Gamma^a_{bc} V^c) \\
&= V^b \nabla_b \xi^a - \xi^b \nabla_b V^c
\end{aligned} \tag{1.56}$$

where we have used $\Gamma^a_{bc} = \Gamma^a_{cb}$. Let use the directional derivative notation: $\nabla_X \equiv X^b \nabla_b$. So we have from (1.56)

$$\nabla_V \xi^a = \nabla_\xi V^a \tag{1.57}$$

Applying the directional derivative ∇_V to both sides gives

$$\nabla_V \nabla_V \xi^a = \nabla_V \nabla_\xi V^a. \tag{1.58}$$

Consider the identity

$$\nabla_X(\nabla_Y Z^a) - \nabla_Y(\nabla_X Z^a) - \nabla_{[X,Y]} Z^a = R^a{}_{bcd} Z^b X^c Y^d \tag{1.59}$$

If we set $X^a = Z^a = V^a$ and $Y^a = \xi^a$, then the second is $\nabla_\xi(\nabla_V V^a)$ vanishes because V^a is tangent to a geodesic,

$$\nabla_V V^a = V^b \nabla_b V^a = 0. \tag{1.60}$$

The third term, $[V, \xi]^b \nabla_b V^a$, vanishes by (1.55). Thus (1.59) becomes

$$\nabla_V \nabla_\xi V^a - R^a{}_{bcd} V^b V^c \xi^d = 0. \tag{1.61}$$

Substituting (1.58) into this we obtain

$$\nabla_V \nabla_V \xi^a - R^a{}_{bcd} V^b V^c \xi^d = 0. \tag{1.62}$$

By definition

$$\frac{d^2 \xi^a}{d\tau^2} \equiv \nabla_V \nabla_V \xi^a$$

and so, the geodesic deviation equation is

$$\frac{d^2 \xi^a}{d\tau^2} = R^a{}_{bcd} V^b V^c \xi^d. \tag{1.63}$$

We are only interested in the spatial part of ξ^a and its geodesic deviation equation. We thus define the orthogonal connecting vector η^a by

$$\eta^a := h^a{}_b \xi^b. \tag{1.64}$$

As

$$\xi^a = \eta^a - V^a V_b \xi^b$$

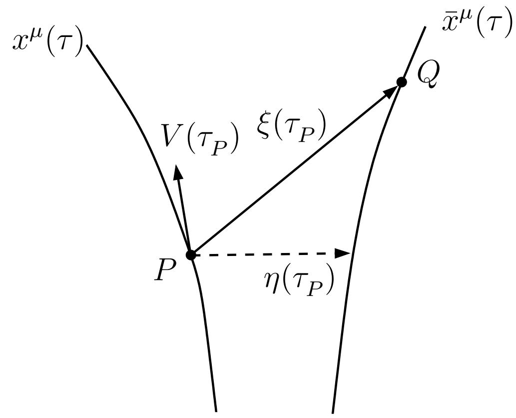


Figure 1.54: η is the orthogonal connecting vector.

we have

$$\begin{aligned}\frac{d\xi^a}{d\tau} &= V^c\nabla_c\xi^a \\ &= V^c\nabla_c(\eta^a - V^aV_b\xi^b) \\ &= V^c\nabla_c\eta^a - (V^c\nabla_cV^a)V_b\xi^b - V^a(V^c\nabla_cV_b)\xi^b - V^aV_b(V^c\nabla_c\xi^b) \\ &= V^c\nabla_c\eta^a \\ &= \frac{d\eta^a}{d\tau},\end{aligned} \tag{1.65}$$

where we used the geodesic equation $V^b\nabla_bV^a = 0$, $\nabla_V\xi^b = \nabla_\xi V^b$ and $V_b\nabla_cV^b = 0$. We also have

$$R^a{}_{bcd}V^bV^c\xi^d = R^a{}_{bcd}V^bV^c(\eta^d + V^dV_e\xi^e) = R^a{}_{bcd}V^bV^c\eta^d \tag{1.66}$$

since $R^a{}_{bcd}$ is anti-symmetric in c and d. So we have

$$\frac{d^2\eta^a}{d\tau^2} - R^a{}_{bcd}V^bV^c\eta^d = 0 \tag{1.67}$$

which is the same as (1.63) but with ξ^a replaced with η^a.

It is convenient to introduce an orthonormal basis of spatial vectors e_1^a, e_2^a, e_3^a orthogonal to V^a and parallely propagated along γ. The components, $\eta^\mu := \eta^a e_a^\mu$, of the vector η^a which represents the separation between a geodesic $\gamma(s)$ and a neighbouring geodesic then satisfies the Jacobi equation

$$\frac{d^2\eta^\mu}{d\tau^2} = -\sum_{\alpha,\beta,\nu} R_{\alpha\beta\nu}{}^{\mu} V^\alpha \eta^\beta V^\nu \tag{1.68}$$

The value of η at time s depends linearily on the initial data $\eta(0)$ and $(d\eta^\mu/ds)(0)$ at p. Since, by construction, $\eta^\mu(0) = 0$ for this congruence, we must have

$$\eta_\mu(s) = \sum_{\nu=1}^{3} A_{\mu\nu}(s) \frac{d\eta^\nu}{ds}(0) \tag{1.69}$$

The matrix $A_{\alpha\beta}$

By (1.69) we have

$$\begin{aligned} \frac{d^2\eta^\mu}{ds^2} &= \sum_{\nu=1}^{3} \frac{d^2}{ds^2} A^\mu{}_\nu(s) \frac{d\eta^\nu}{ds}(0) \\ &= -\sum_{\alpha,\beta,\sigma} R_{\alpha\beta\sigma}{}^{\mu} V^\alpha \eta^\beta V^\sigma \\ &= -\sum_{\alpha,\beta,\sigma} R_{\alpha\beta\sigma}{}^{\mu} V^\alpha V^\sigma \Big(\sum_{\nu=1}^{3} A^\beta{}_\nu(s) \frac{d\eta^\nu}{ds}(0) \Big) \end{aligned} \tag{1.70}$$

so that $A^\mu{}_\nu(s)$ satisfies the equation

$$\frac{d^2 A^\mu{}_\nu}{ds^2}(s) = -\sum_{\alpha,\beta,\nu} R_{\alpha\beta\sigma}{}^{\mu} V^\alpha V^\sigma A^\beta{}_\nu(\tau) \tag{1.71}$$

With

$$\frac{d\eta^\mu}{d\tau}(\tau) = \sum_{\nu=1}^{3} \frac{dA^\mu{}_\nu}{d\tau}(\tau) \frac{d\eta^\nu}{d\tau}(0) \tag{1.72}$$

clearly, we also have

$$A_{\mu\nu}(0) = 0, \qquad \frac{dA_{\mu\nu}}{d\tau}(0) = \delta_{\nu\mu}. \tag{1.73}$$

By definition of $\frac{d}{d\tau}$

$$\frac{d}{d\tau}\eta_\alpha = V_{\alpha;\beta}\eta_\beta \tag{1.74}$$

or in terms of $A_{\alpha\beta}$

$$\sum_{\nu=1}^{3}\frac{dA_{\mu\nu}}{d\tau}(\tau)\frac{d\eta^\nu}{d\tau}(0) = V_{\alpha;\beta}\sum_{\nu=1}^{3}A_{\beta\nu}(\tau)\frac{d\eta^\nu}{d\tau}(0) \tag{1.75}$$

or

$$\frac{dA_{\alpha\beta}}{d\tau}(\tau) = V_{\alpha;\gamma}\ A_{\gamma\beta}(\tau) \tag{1.76}$$

Assumming $A_{\alpha\beta}$ is invertible we can write

$$V_{\alpha;\beta}(\tau) = A^{-1}_{\gamma\beta}\frac{d}{ds}A_{\alpha\gamma} \tag{1.77}$$

This allows the vorticity and the expansion tensor as well as the expansion scalar to be expressed in terms of the matrix $A_{\alpha\beta}$ and it's inverse ${A^{-1}}_{\alpha\beta}$:

$$\omega_{\alpha\beta} = -A^{-1}_{\gamma[\alpha}\frac{d}{ds}A_{\beta]\gamma} \tag{1.78}$$

$$\theta_{\alpha\beta} = A^{-1}_{\gamma(\alpha}\frac{d}{ds}A_{\beta)\gamma} \tag{1.79}$$

$$\theta = trB = tr\left(\frac{dA}{d\tau}A^{-1}\right) \tag{1.80}$$

(so that $(\det A)^3$ measures a spacelike 3-volume element orthogonal to γ and transported along the curves of the congruence)? proof?

consider

$$\frac{d}{d\tau}\det A = \frac{dA_{ij}}{d\tau}\frac{\partial}{\partial A_{ij}}\det A \tag{1.81}$$

Recall from (**??**)

$$\frac{\partial}{\partial A_{ij}} \det A = \tilde{A}_{ij} = \det A (A^{-1})_{ji} \tag{1.82}$$

where $\tilde{A}_{ij}$ are the cofactors of $\mathbf{A}$. Together we get

$$\frac{d}{d\tau} \det A = \det A \left(\frac{dA_{ij}}{d\tau} (A^{-1})_{ji} \right) \tag{1.83}$$

So (1.112) can be written

$$\theta = \frac{d}{d\tau} (ln|detA|) \tag{1.84}$$

$$\sigma_{ab} = h_a{}^c h_b{}^d \nabla_{(d}\eta_{c)} - \frac{1}{3}\theta h_{ab} \tag{1.85}$$

Consider the quantity

$$A_{\alpha\delta}\omega_{\alpha\gamma}A_{\alpha\delta} = \frac{1}{2}\left((\frac{d}{dt}A_{\alpha\beta})A_{\alpha\delta} - A_{\alpha\beta}\frac{d}{dt}A_{\alpha\delta} \right) \tag{1.86}$$

It is easy to verify that this quantity is constant along $\gamma(t)$ by differating both sides and using (1.71),

$$\begin{aligned}
\frac{d}{ds}(A_{\alpha\delta}\omega_{\alpha\gamma}A_{\alpha\delta}) &= \frac{1}{2}\frac{d}{ds}\left(A_{\gamma\alpha}\frac{d}{ds}A_{\gamma\beta} - A_{\gamma\beta}\frac{d}{ds}A_{\gamma\alpha} \right) \\
&= \frac{1}{2}\left(A_{\gamma\alpha}\frac{d^2}{ds^2}A_{\gamma\beta} - A_{\gamma\beta}\frac{d^2}{ds^2}A_{\gamma\alpha} \right) \\
&= \frac{1}{2}(A_{\gamma\alpha}R_{\gamma 4\delta 4}A_{\delta\beta} - A_{\gamma\beta}R_{\gamma 4\delta 4}A_{\delta\alpha}) \\
&= \frac{1}{2}(A_{\gamma\alpha}R_{\gamma 4\delta 4}A_{\delta\beta} - A_{\gamma\beta}R_{\delta 4\gamma 4}A_{\delta\alpha}) \\
&= 0
\end{aligned} \tag{1.87}$$

where we used $R_{\gamma 4\delta 4} = R_{\delta 4\gamma 4}$. The RHS of (1.86) vanishes where $A_{\alpha\beta}$ is zero and so must be zero all along $\gamma(t)$,

$$A_{\alpha\delta}\omega_{\alpha\gamma}A_{\alpha\delta} = 0. \tag{1.88}$$

When $A_{\alpha\beta}$ is invertable, we can apply the inverse matrix of $\mathbf{A}$ to both sides and so conclude that $\omega_{\alpha\beta} = 0$ when $A_{\alpha\beta}$ is non-singular.

$$\frac{d^2}{ds^2} Z^\alpha = -R_{\alpha 4 \gamma 4} Z^\beta \tag{1.89}$$

$$\frac{d^2}{ds^2} A_{\alpha\beta} = -R_{\alpha 4 \gamma 4} A_{\gamma\beta} \tag{1.90}$$

The expansion θ is defined as $(detA)^{-1} d(detA)/ds$. Since $A_{\alpha\beta}$ obeys the ordinary differential equation (), $d(detA)/ds$ will be finite.

Proposition 1.1.43 (The focussing theorem) *Take any inextendable geodesic $\gamma(\lambda)$. If at some point $\gamma(\lambda_1)$ $(\lambda_1 > 0)$, the expansion has a negative value $\theta_1 < 0$ and if $R_{ab}V^aV^b$ everwhere positive then there will be a point conjugate to q along $\gamma(\lambda)$ between $\gamma(\lambda_1)$ and $\gamma(\lambda_1 + (3/-\theta_1))$.*

The Raychaudhuir equation for non-rotating flow implies

$$\dot{\theta} \leq -\frac{1}{3}\theta^2. \tag{1.91}$$

Writing it in the form $\theta^{-2}\dot{\theta} \leq -(1/3)$

$$\int_{\lambda_1}^{\lambda_2} d\lambda \theta^{-2}\dot{\theta} = \int_{\theta(\lambda_1)}^{\theta(\lambda_2)} d\theta \theta^{-2} = \big[-\theta^{-1}\big]_{\theta(\lambda_1)}^{\theta(\lambda_2)} \leq -\frac{1}{3}(\lambda_2 - \lambda_1) \tag{1.92}$$

rearranging gives:

$$-\theta(\lambda_2) \geq \left[-\theta(\lambda_1)^{-1} - \frac{(\lambda_2 - \lambda_1)}{3}\right]^{-1} \tag{1.93}$$

From this inequality we see that if the congruence is initially converging ($\theta(0) < 0$), then $\theta(\lambda) \to -\infty$ within an affine parameter $\lambda \leq 3/|\theta(0)|$. This demonstrates the irreversible tendency of neighbouring geodesics to converge. Of course, this represents merely a singularity in the congruence, not a singularity in the structure of spacetime itself. However, this result is key to the singularity theorems, as the above may not be possible if spacetime is geodesically incomplete, and as already stated, this incompleteness is interpreted as evidence of a singularity.

□

The time-reversed version of the argument holds ($s \to -s$, $V^a \to -V^a$, and $\omega \to -\omega = 0$) of this theorem where if $\theta(0) > 0$ at a point then there was must have been a point an affine parameter $\lambda \leq 3/|\theta(0)|$ distance away in the past where $\theta(-\lambda) \to -\infty$

The Weyl tensr expresses the tidal force that a body experiences when moving along a geodesic. The Weyl tensor differs from the Riemann curvature tensor in that it does not contains information on how the volume of the body changes, but rather only on how the shape of the body is distorted by the tidal force. The Ricci curvature contains precisely the information about how the volume changes in the presence of tidal forces. The Weyl curvature is the only part of the curvature that exists in free space. The effect of the Ricci tensor are due to a matter distribution and has a positive focussing effect..

attractive instability of gravity.

proposition (1.1.43), all that is required for the existence of conjugate points γ is that $R_{ab}\xi^a\xi^b \geq 0$ everywhere on γ and $R_{abcd}\xi^b\xi^d \neq 0$ at least one point of γ. We prove this in section 1.1.11 on the equivalence of the two versions of the singularity theorem.

$$\begin{aligned}
\frac{\partial \tau}{\partial \alpha}(\alpha) &= \int_a^b \frac{\partial}{\partial \alpha} f(\alpha) \\
&= \int_a^b X^a \nabla_a (-T^a T_a)^{1/2} \\
&= -\int_a^b \frac{1}{f} T_b X^a \nabla_a T^b dt \\
&= -\int_a^b \frac{1}{f} T_b T^a \nabla_a X^b dt \\
&= -\int_a^b T^a \nabla_a (X^b T_b / f) dt + \int_a^b X^b T^a \nabla_a (T_b / f) dt \\
&= \int_a^b X^b T^a \nabla_a (T_b / f) dt
\end{aligned} \tag{1.94}$$

where we have used that $T^a \nabla_a (X^b T_b / f) = \partial (f^{-1} T_b X^b) / \partial t$ and X^b vanishes at the end-points. Thus the condition $[d\tau / d\alpha]_{\alpha=0}$ is true for arbitrary X^b if and only if $T^a \nabla_a (T_b / f) = 0$ at $\alpha = 0$, which is the geodesic equation expressed in arbitrary parameterisation: $T^a \nabla_a T_b = (\dot{f}/f^2) T^a$.

The second variation of the arc length is

$$\frac{d^2 \tau}{d\alpha^2} = \int_a^b X^c \nabla_c [X^b T^a \nabla_a (T_b / f)] dt \tag{1.95}$$

Evaluating this at $\alpha = 0$, asuming λ_0 is a geodesic, we have

$$
\begin{aligned}
\left.\frac{d^2\tau}{d\alpha^2}\right|_{\alpha=0} &= \int_a^b X^c\nabla_c[X^bT^a\nabla_a(T_b/f)]dt \\
&= \int_a^b X^b(X^c\nabla_c T^a)\nabla_a(T_b/f)dt + \qquad \text{we used } T^c\nabla_c(T^b/f)=0, \\
&+ \quad \int_a^b X^bT^aX^c\nabla_c\nabla_a(T_b/f)dt \\
&= \int_a^b X^b(T^c\nabla_c X^a)\nabla_a(T_b/f)dt + \qquad \text{we used } T^a\nabla_a X^a = X^a\nabla_a T^a \\
&+ \quad \int_a^b X^bT^aX^c\nabla_a\nabla_c(T_b/f)dt + \int_a^b X^bT^aX^cR_{cab}{}^dT_d/f\,dt \\
&= \int_a^b X^bT^c\nabla_c[X^a\nabla_a(T_b/f)]dt + \int_a^b X^bT^aX^cR_{cab}{}^dT_d/f\,dt \qquad (1.96)
\end{aligned}
$$

where we used

$$
\begin{aligned}
\nabla_c\nabla_a(T_b/f) &= \nabla_c\left[\frac{1}{f}\nabla_a T_b - \frac{1}{f^2}T_b\nabla_a f\right] \\
&= \frac{1}{f}\nabla_c\nabla_a T_b - \frac{1}{f^2}\nabla_a T_b(\nabla_c f) - \frac{1}{f^2}\nabla_c T_b(\nabla_a f) - T_b\nabla_c\left[\frac{1}{f^2}\nabla_a f\right] \\
&= \frac{1}{f}\nabla_a\nabla_c T_b + R_{cab}{}^d(T_b/f) - \frac{1}{f^2}\nabla_a T_b(\nabla_c f) - \frac{1}{f^2}\nabla_c T_b(\nabla_a f) - T_b\nabla_a\left[\frac{1}{f^2}\nabla_c f\right] \\
&= \nabla_a\nabla_c(T_b/f) + R_{cab}{}^d(T_b/f) \qquad (1.97)
\end{aligned}
$$

The term in the square brackets of (1.96) can be rexpressed as

$$
\begin{aligned}
X^a\nabla_a(T_b/f) &= \frac{1}{f}X^a\nabla_a T_b - \frac{1}{f^2}T_bX^a\nabla_a f \\
&= \frac{1}{f}T^a\nabla_a X_b + \frac{1}{f^2}T_bX^a\nabla_a\left[\frac{1}{f}T^cT_c\right] \\
&= \frac{1}{f}T^a\nabla_a X_b + \frac{1}{f^2}T_bT^a\nabla_a\left[\frac{1}{f}T_cX^c\right] \qquad (1.98)
\end{aligned}
$$

choosing the curve parameterization so that $f=1$ along the geodesic γ_0 and also choosing the deviation vector to be orthogonal to T^a along γ_0.

$$\left.\frac{d^2\tau}{d\alpha^2}\right|_{\alpha=0} = \int_a^b X^b\{T^c\nabla_c(T^a\nabla_a X_b) + R_{cab}{}^d T^a T_d X^c\}dt$$
$$= \int_a^b X^b(\mathcal{O})_b dt \quad (1.99)$$

(1.99)

the matrix A is non-singular between p and q, $Y^\mu = \sum_\nu (A^{-1})^\mu{}_\nu X^\nu$.

Conjugate points for Timelike geodesics normal to spacelike three-surface

A point p on a geodesic γ of the geodesic congruence orthogonal to Σ is said to be conjugate to Σ along γ if there exists an orthogonal deviation vector η^a of the congruence which is nonzero on Σ but vanishes at p. Intuitively, p is conjugate to Σ if two "infintesimally nearby" geodesics orthogonal to Σ cross p as in fig ().

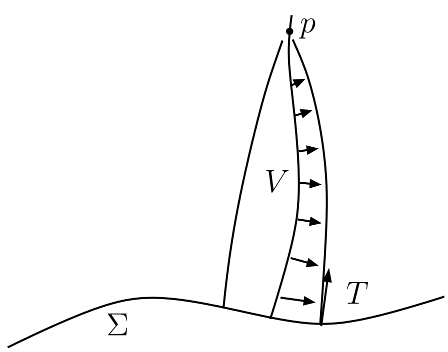


Figure 1.55: Displacement vectors for a hypersurface.

The three dimensional manifold Σ is defined locally by

$$f = 0$$

where f is second differentiable and $g^{ab}\partial_a f \partial_b f < 0$. The unit normal vector N to the surface Σ is

$$N^a = \frac{g^{ad}\nabla_d f}{(-g^{bc}\nabla_b f \nabla_c f)^{1/2}}$$

We define the second fundamental form χ of Σ as

$$\chi_{ab} = h_a{}^c h_b{}^d \nabla_d N_c \tag{1.100}$$

where as before

$$h_{ab} = g_{ab} + N_a N_b$$

is the induced metric tensor on Σ (also called the first fundamental form). It follows that χ_{ab} is symmetric,

$$\begin{aligned}
\chi_{ab} &= h_a{}^c h_b{}^d \nabla_d N_c \\
&= h_a{}^c h_b{}^d \left(\frac{\nabla_d \partial_c f}{\|f\|} - \partial_c f \nabla_d \frac{1}{\|f\|} \right) \\
&= \frac{h_a{}^c h_b{}^d (\partial_d \partial_c f - \Gamma^e_{dc} \partial_e f)}{\|f\|} - (h_a{}^c \partial_c f) h_b{}^d \nabla_d \frac{1}{\|f\|} \\
&= h_a{}^c h_b{}^d \nabla_c N_d \\
&= \chi_{ba}
\end{aligned} \tag{1.101}$$

where we used $\Gamma^e_{dc} = \Gamma^e_{cd}$ and $h_a{}^c \partial_c f = 0$.

The congruence of timelike geodesics orthonormal to Σ will consist of the timelike geodesics whose unit tangent vector V is the same as the unit normal N at Σ. Then

$$\nabla_b V_a = \chi_{ab} \ \text{ at } \Sigma. \tag{1.102}$$

When the congruence of geodesics are normal to the hypersurface the vorticity vanishes. Proof: The unit normal vector to the hypersurface is

$$V_a = \frac{\nabla_a f}{\|f\|}$$

We calculate the vorticity,

$$\begin{aligned}
\omega_{ab} &= h_a{}^c h_b{}^d \nabla_{[d} V_{c]} \\
&= \frac{h_a{}^c h_b{}^d}{2} \left[\nabla_d \left(\frac{\nabla_c f}{\|f\|} \right) - \nabla_c \left(\frac{\nabla_d f}{\|f\|} \right) \right] \\
&= \frac{h_a{}^c h_b{}^d}{2} \left[\frac{\nabla_d \nabla_c f - \nabla_c \nabla_d f}{\|f\|} + \nabla_c f \nabla_d \frac{1}{\|f\|} - \nabla_d f \nabla_c \frac{1}{\|f\|} \right] \\
&= 0
\end{aligned} \tag{1.103}$$

since $\nabla_c \nabla_d f - \nabla_d \nabla_c f = 0$ and $h_a{}^c \nabla_c f = 0$.

For the vector η which represents the separtion of a neighbouring geodesic normal to Σ from a geodesic $\gamma(s)$ normal to Σ, we again have the equation of geodesic deviation for the separating vector,

$$\frac{d^2}{ds^2}\eta_\alpha(s) = -R_{\alpha 4 \gamma 4}\eta_\gamma(s).$$

Also

$$\frac{d}{ds}\eta_\alpha(s) = \nabla_\beta V_\alpha \eta_\beta.$$

At a point on $\gamma(s)$ at Σ we have the initial condition

$$\frac{d}{ds}\eta_\alpha = \chi_{\alpha\beta}\eta_\beta \tag{1.104}$$

We express the Jacobi fields along $\gamma(s)$ which satisfy the above condition as

$$\eta^\mu(s) = \sum_{\nu=1}^{3} A^\mu{}_\nu(s)\eta^\nu(0) \tag{1.105}$$

This $A^\mu{}_\nu(s)$ satisfies the same differential equation

So that

$$\frac{d\eta^\mu}{d\tau}(\tau) = \sum_{\nu=1}^{3} \frac{dA^\mu{}_\nu}{d\tau}(\tau)\eta^\nu(0) \tag{1.106}$$

by

$$\frac{d\eta^\alpha}{d\tau}(\tau) = V_{\nu;\beta}\eta^\beta \tag{1.107}$$

or in terms of $A_{\alpha\beta}(\tau)$

$$\sum_{\nu=1}^{3} \frac{dA_{\alpha\beta}}{ds}(s)\eta^\beta(0) = V_{\alpha;\gamma}\sum_{\nu=1}^{3} A^\gamma{}_\beta(s)\eta^\beta(0) \tag{1.108}$$

or

$$\frac{dA_{\alpha\beta}}{ds}(s) = V_{\alpha;\gamma} A_{\gamma\beta}(s) \tag{1.109}$$

which is of the same form as (1.76). This allows the vorticity and the expansion tensor as well as the expansion scalar to be expressed in terms of the matrix $A_{\alpha\beta}$ and it's inverse ${A^{-1}}_{\alpha\beta}$ in the same way as before where there were different boundary conditions for $A_{\alpha\beta}$:

$$\omega_{\alpha\beta} = -A^{-1}_{\gamma[\alpha} \frac{d}{ds} A_{\beta]\gamma} \tag{1.110}$$

$$\theta_{\alpha\beta} = A^{-1}_{\gamma(\alpha} \frac{d}{ds} A_{\beta)\gamma} \tag{1.111}$$

$$\theta = trB = tr\left(\frac{dA}{d\tau} A^{-1}\right) \tag{1.112}$$

Again the quantity

$$A_{\alpha\delta}\omega_{\alpha\gamma}A_{\alpha\delta} = \frac{1}{2}\left((\frac{d}{dt} A_{\alpha\beta})A_{\alpha\delta} - A_{\alpha\beta}\frac{d}{dt}A_{\alpha\delta}\right) \tag{1.113}$$

is constant.

Thus on the hypersurface we have the boundary conditions

$$A_{\mu\nu}(0) = \delta_{\nu\mu}, \qquad \frac{dA_{\mu\nu}}{d\tau}(0) = \frac{dA_{\nu\mu}}{d\tau}(0). \tag{1.114}$$

Proposition 1.1.44 *Let $\mathcal{M}$ be a spacetime satisfying $R_{ab}\xi^a\xi^b \geq 0$ for all timelike ξ^a. Let Σ be a spacelike hypersurface with $K = \theta < 0$ at a point $q \in \Sigma$. Then within a proper time $\tau \leq 3/|K|$ there exists a point p conjugate to Σ along the geodesic γ orthogonal to Σ and passing through q, assuming that γ can be extended that far.*

This is proved using the Raychaudhuri equation (1.34) as in proposition 1.1.43. If the congruence is hypersurface orthogonal, we have $\omega_{ab} = 0$, so that the third term vanishes. The second term, $-\sigma_{ab}\sigma^{ab}$, is manifestly nonpositive. Thus we have

$$\frac{d\theta}{ds} + \frac{1}{3}\theta^3 \leq 0, \tag{1.115}$$

as before, which leads to

$$-\theta(s) \geq \left[-\theta^{-1}(0) - \frac{1}{3}s\right]^{-1}. \tag{1.116}$$

We are supposing that $\theta(0) < 0$, i.e. the congruence is initialy converging.

$\square$

Again the time reverse version holds, important in proof of simple singularity theorem on comological singulairty presented in the previous appendix.

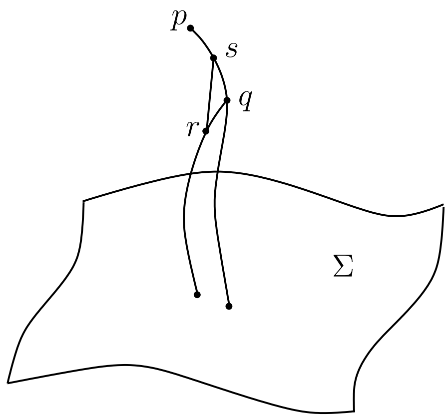


Figure 1.56: .

Conversely

$$\omega_{ab} = 0$$

implies

$$V_a = \nabla_a f$$

This is guaranteed by the Frobenius theorm.

The boundary conditions (1.114) will also be encountered in proposition 1.1.53.

1.1.10 The Singularity Theorems

Corollary 1: Let F be a future set, γ a null geodesic on ∂F. Then γ contains no proper segment (??a segment not γ itself or an empty set??) which has a pair of conjugate points.

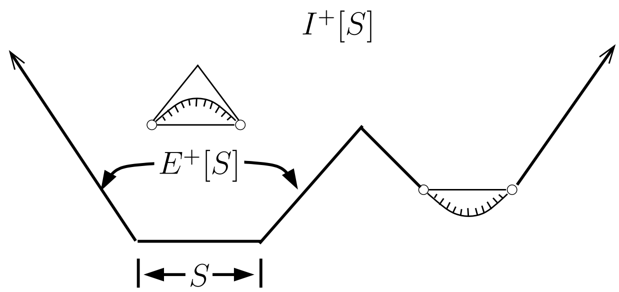


Figure 1.57: Future horozmos .

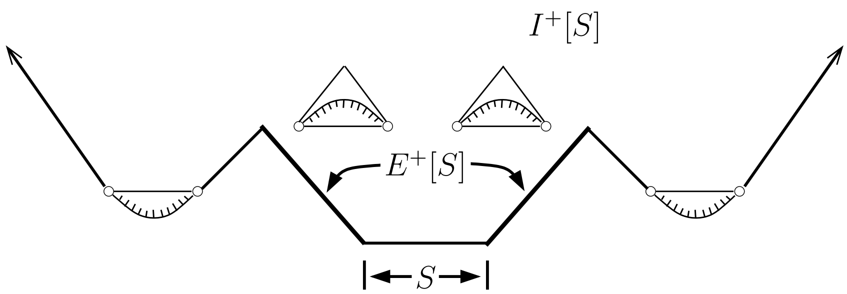


Figure 1.58: "doubling" future horozmos. $E^+[S]$ is compact

Proof: If $x < y < z$ along γ with x and y conjugate, then $z \in I^+(x)$ implies $z \in F$ or in other words $z \notin \partial F$, which is false.

□

Corollary 2: Let $\mathcal{M}$ be a spacetime satisfying:

(1) There are no closed trips

(2) *Every* endless null geodesic in $\mathcal{M}$ contains a pair of conjugate points.

Then is $\mathcal{M}$ strongly causal.

Proof:

If not, and there are no closed trips, strong causality fails on an endless geodesic γ. By hypothesis, there are conjugate points A and B on γ with $a < A$, $b > B$. Then strong causality fails at a and b, and by the theorem, $I^+(a) \cap I^-(b) \neq \emptyset$ and there are closed trips (see remark following 1.1.10), which is false.

□

Proposition 1.1.45 *Let S be future trapped and suppose strong causality holds in $\overline{I^+[S]}$. Then there is a future endless trip γ such that $\gamma \subset intD^+(E^+[S])$.*

Proof: First we nedd to show $H^+[E^+[S]]$, if it exists, must be non-compact. The basic idea of the argument is given in fig (1.1.10)

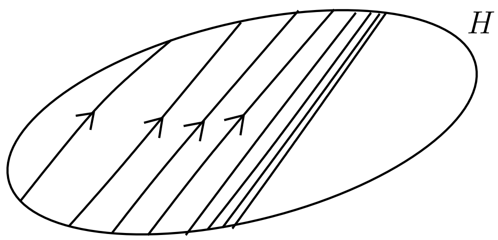


Figure 1.59: From the fact that H is a Cauchy horizon it follows that through every point of H there passes a maximally extened past-directed null geodesic that remains in H. Since H is compact, such a curve would have to come back arbitrarily close to itself - reentering some Alexandoff neighbourhood and so violating strong causality.

Let $H = H^+[E^+[S]]$. Any trip which leaves $\text{int}D^+(E^+[S])$ crosses H. If $H = \emptyset$, then the proposition is proven. If not, then H must be non-compact. For, if H is compact, there are finitely many Alexandoff neighbourhoods covering H, say $B_1, \ldots, B_k$. There exists a point $I^+[S]/D^+(E^+[S])$; suppose $p \in B_i$. Then there is a point $q_i \in I^+[S]/D^+(E^+[S])$ such that $q_{i_1} << p$ and $q_{i_1} \in B_{i_1} - B_i$ for some i_1. The process continues by induction to yeild $p >> q_{i_1} >> q_{i_2} >> \cdots >> q_{i_k} >> \cdots$. However, when the process has been repeated k times two (at least) of the q's must lie in the same B_{i_1}. By consrtuction, the trip between these two must leave and then reenter B_{i_1} contradicting the fact that $\overline{I^+[S]}$ is strongly causal.

So H is non-compact. Let ξ be any nowhere vanishing timelike vector field on $\mathcal{M}$. Given a non-zero vector field ξ defined over the manifold, this can be used to define a congruence of curves over the manifold (a congruence of curves is defined such that only one curve goes through each point of the manifold). We use ξ to construct the family of trajectories from $E^+[S]$. If one of these trajectories leaves $\text{int}D^+[E^+[S]]$, it intersects H. This mapping will have to be one-to-one as no point in this boundary can procede itself. If all these trajectories leave, then a homeomorphism is established between $E^+[S]$ and H, which is impossible since $E^+[S]$ is compact, and H is not. So one trajectory remains inside.

□

Theorem 1.1.46 *The following are mutually inconsistent in any spacetime:*

(a) There are no closed trips

(b) Every endless causal geodesic contains a pair of conjugate points

(c) There is a future (past) - trapped set $\mathcal{S}$ in $\mathcal{M}$.

Proof: Under the assumptions that all three conditions are satisfied, (a) and (b) imply that $\mathcal{M}$ is strongly causal. This implies the existence of a future endless geodesic in

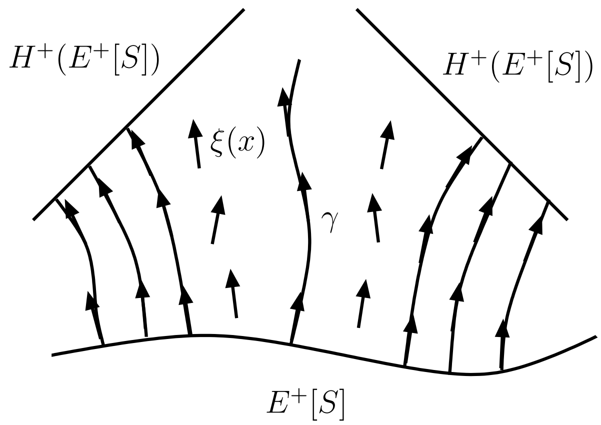


Figure 1.60: $E^+[S]$ is compact by defintion, however its Cauchy horizon is non-compact. As the two subsets can not homeomorphic there must be at least one trajectory γ which remains in $\text{int}D^+(E^+[S])$.

$\text{int}D^+(E^+(S))$ say γ. Define the compact set $T = \overline{I^-(\gamma)} \cap E^+(S)$. We argue that T is past-trapped. Since γ was contained in $\text{int}I^+(E^+(S)))$, $E^-(T)$ would consist of T and a portion of $\dot{J}^-(\gamma)$, see fig (1.1.10). Since γ was future inextendible, the null geodesic segments generating $\dot{J}^-(\gamma)$ could have no future endpoints. But by (b) every inextendible non-spacelike geodesic contains a pair of conjugate points. Thus each generating segment of $\dot{J}^-(\gamma)$ would enter $I^-(\gamma)$. However, as $\text{int}D^+(E^+[S])$ is globally hyperbolic, within this region on the null geodesic segments generating $\dot{J}^-(\gamma)$ there cannot occur a pair of conjugate points and so they intersect T. $E^-(T)$ is be compact being the intersection of the closed set $\dot{J}^-(\gamma)$ with a compact set generated by null geodesic segments from T of some bounded affine length. Then, by the time-reversed version of proposition 1.1.45 there exists α, a past endless causal geodesic in $\text{int}D^-(E^-(T))$.

Choose a sequence a_i receding into the past on α without limit point and a sequence c_i on γ going to the future. The sets $J^-(c_i) \cap J^+(a_i)$ are compact, for all i, and so there is a maximal geodesic μ_i in $\mathcal{C}(a_i, c_i)$ for each i. The intersection with T (as it is compact) have a limit point q and a limiting direction. Construct the causal geodesic μ which has this direction at q.

some nearby geodesic must meet γ more than once. But as γ is the limit curve of the sequence $\{\gamma_i\}$, so the nearby geodesic will meet at least one member γ_k of the sequence $\{\gamma_i\}$. This would mean we could lengthen that γ_k by rounding off corners, which contradicts that γ_k is the curve of maximum length from b_k to a_k.

□

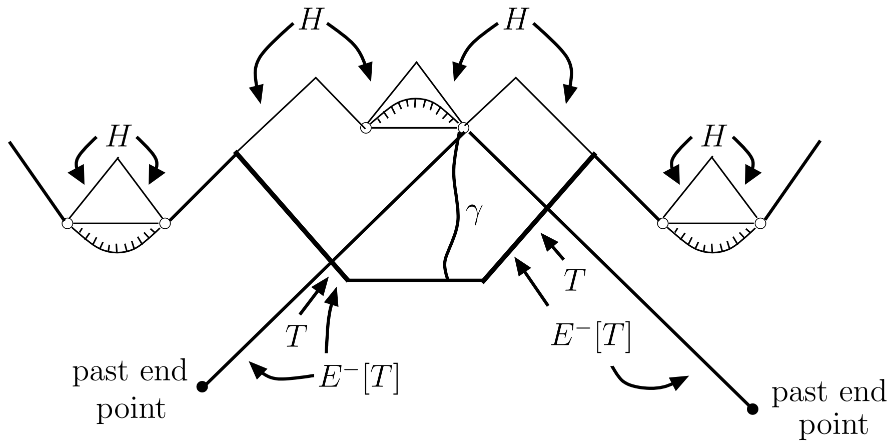


Figure 1.61: γ is a future endless causal geodesic in int$D^+(E^+(S))$. $H = H^+(E^+[S])$. As the intersection of the closed set $\dot{J}^-(\gamma)$ with a compact set generated by null geodesic segments from T of some bounded affine length, $E^-[T]$ is compact

1.1.11 Implication of the "Displayed" Singularity Theorem from the Established Version

We now establish theorem 1.1.1 from theorem 1.1.2.

move on to establish between

Theorem 1.1.47 *if $\mathcal{M}$ were timelike and null geodesically complete, (1) and (2) would imply (a) by 1.1.48. (3) is the same as (b). (1) and (4) would imply (c):*

in case (i) $\mathcal{L}$ would be the compact achronal set without edge and

$$E^+(\mathcal{L}) = E^-(\mathcal{L}) = \mathcal{L} \tag{1.117}$$

Refering to the two versions given in 1.1.1

Proof: To be given.

□

Theorem 1.1.48 *If for every non-spacelike vector K we have*

$$R_{ab}K^aK^b < 0 \tag{1.118}$$

and

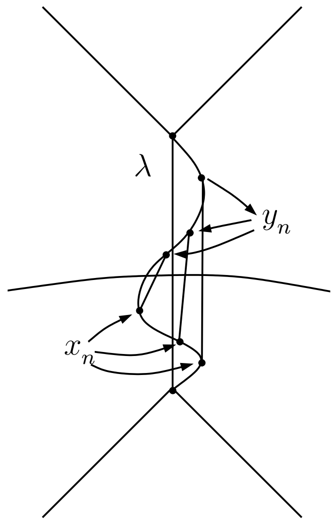


Figure 1.62: The geodesic λ.

(2) Every non-spacelike geodesic, with tangent vector K, contains a point at which

$$K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0. \tag{1.119}$$

then every inextendible non-spacelike geodesic will contain a pair of conjugate points.

Proof: To be given.

□

We say that a spacetime satisfying the condition that every timelike or null geodesic contains a point at which $K_{[a}R_{b]cd[e}K_{f]}K^cK^d$ is not zero satisfies the generic condition.

Before we give the proof of this theorem we discuss whether it is reasonable to assume the generic condition in physically realistic solutions. First note that for the timelike case the condition is $V_{[a}R_{b]cd[e}V_{f]}V^cV^d \neq 0$ equivalent to $R_{abcd}V^bV^c \neq 0$:

$$V_{[a}R_{b]cd[e}V_{f]}V^cV^d \neq 0 \tag{1.120}$$

Contracting this with V^aV^f, using $R_{abcd} = -R_{bacd} = -R_{abdc}$ and $V^2 = -1$ this becomes

$$R_{abcd}V^bV^c \neq 0. \tag{1.121}$$

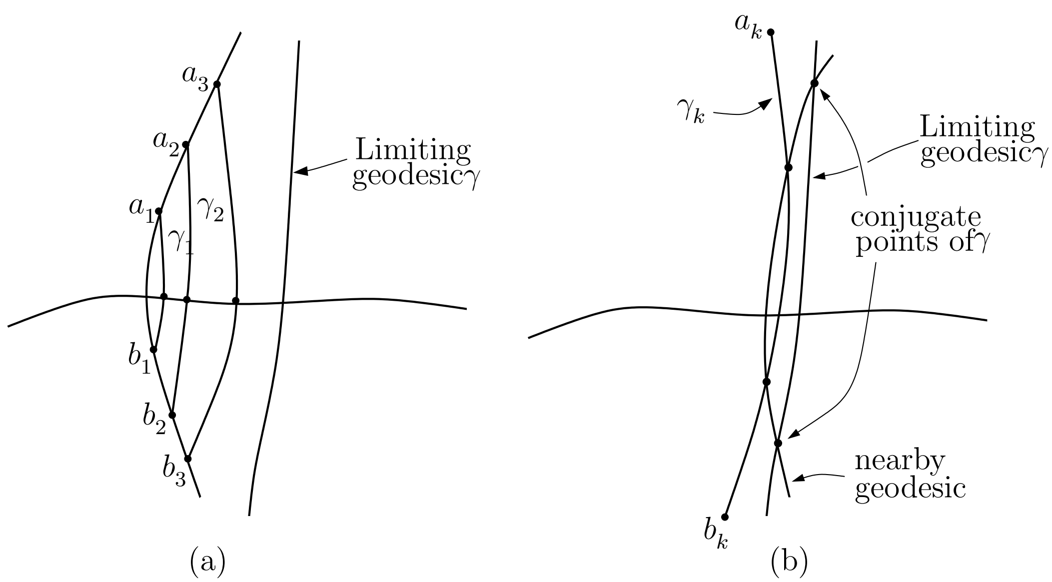


Figure 1.63: The limit geodesic γ contains conjugate points.

Multiplying $R_{bcde}V^cV^d \neq 0$ by V_aV_f and antisymmetrizing between a and b, and between e and f gives us back Eq.(1.120), so the two conditions Eq.(1.120) and Eq.(1.121) are equivalent [3].

One would expect every timelike geodesic to encounter some mater or some gravitational radiation and so contain some point where the tidal force is non-zero, i.e. a point where $R_{abcd}V^bV^d$ is non-zero. Thus one expect every timelike geodesic to contain pairs of conjugate points, provided that it could be extended sufficiently far in both directions.

As in the timelike case, this condition will be satisfied for a null geodesic which passes through some matter provided that the matter is not pure radiation () and moving in the direction of the geodesic tangent vector K. It will be satisfied in empty space if the null geodesic contains some point where the Weyl tensor is non-zero and where K does not lie in one of the directions at the point for which $K_{[a}R_{b]cd[e}K_{f]}K^cK^d = 0$.

We break it down proof of 1.1.48 down into timelike then null case. First timelike case, in which it reads: If $R_{ab}V^aV^b \geq 0$ and if at some point $p = \gamma(s_1)$ the tidal force $R_{abcd}V^bV^d$ is non-zero, there will be values s_0 and s_2 such that $q = \gamma(s_0)$ and $r = \gamma(s_2)$ will be conjugate along γ, providing that $\gamma(s)$ can be extended to these values.

Mathematical preliminaries

Some mathematical preliminaries are needed in order to make lemma 1.1.52 more rigorous, although it is written so that it can be 'understood' without going into all such details.

[3]How do we know that the righthand side of Eq.(1.120), when contracted with V^aV^f is not zero? Well, if $R_{abcd}V^bV^c = 0$ this would mean that that $V_{[a}R_{b]cd[e}V_{f]}V^cV^d = 0$ which is false.

Definition A set B is bounded if there is a point p in S and a radius $r > 0$ such that B is a subset of the neigbourhood $N_r(p)$.

Definition A function f from S to T is said to be bounded if the image set $f(S) = \{f(p) : p \in S\}$ is a bounded subset of T.

Proposition 1.1.49 *Suppose that A is a compact subset of S and that f is a continuous function from A to T. Then the image $f(A) = \{f(p) : p \in A\}$ is a compact subset of T. In particular, f is bounded.*

Proof:

Suppose $\mathcal{V}$ is an open cover of $f(A)$. Let $\mathcal{U}$ be the collection of subsets of S consisting of the inverse images of the sets in the collection $\mathcal{V}$. As f is a continuous function the inverse image of an open set is open in A. Given a point p in A, $f(p)$ belongs to some $V \in \mathcal{V}$, so p belongs to the set $f^{-1}(V)$ from the collection $\mathcal{U}$. Thus $\mathcal{U}$ is an open cover of A. Let $\{U_1, \ldots, U_n\}$ be a finite subcover. Then each U_j is $f^{-1}(V_j)$ for some V_j in the collection $\mathcal{V}$, and the collection $\{V_j\}$ is a finite subcover of $f(A)$.

□

The properties of a norm are:

$$\begin{aligned} \|f\| &\geq 0, \quad \|f\| = 0 \Longleftrightarrow f \equiv 0 \\ \|cf\| &= |c| \cdot \|f\|, \quad c \in \mathbb{R} \\ \|f + g| &\leq \|f\| + \|g\|. \end{aligned} \tag{1.122}$$

Any norm defines a metric via

$$\begin{aligned} d(f, g) = \|f - g\| &\geq 0, \quad d(f, g) = \|f - g\| = 0 \Longleftrightarrow f \equiv g \\ d(f, g) &= \|f - g\| = \|g - f\| = d(g, f) \\ d(f, h) = \|f - h\| &= \|(f - g) + (g - h)| \\ &\leq \|(f - g)\| + \|(g - h)\| \\ &= d(f, g) + d(g, h). \end{aligned} \tag{1.123}$$

Thus everything proved for metric spaces carries over to the case of normed linear spaces.

Definition Supremum (sup) or least upper bound. If S is a nonempty set of real numbers, then a real number u is an upper bound of S if $x \leq u$ or every x in S. A number v is aleast upper bound of S if v is an upper bound and no number less than v is an upper bound of S. Thus, the least upper bound is the smallest real number that is greater than or equal to every number in S.

The result concerning uniform continuity pertains to metric spaces. the standard metric in the space of continuous functions is obtained by starting with the norm

$$\|f\| = sup\{|f(x)| : x \in X\} \tag{1.124}$$

We check the that the norm $\|f\|$ has the properties that define a norm:

$$\begin{aligned} sup\{|f(x)| : x \in X\} &\geq 0, \quad sup\{|f(x)| : x \in X\} = 0 \Longleftrightarrow f \equiv 0 \\ sup\{|cf(x)| : x \in X\} &= |c| \cdot sup\{|f(x)| : x \in X\}, \quad c \in \mathbb{R} \\ sup\{|f(x) + g(x)| : x \in X\} &\leq sup\{|f(x)| : x \in X\} + sup\{|g(x)| : x \in X\}. \end{aligned} \tag{1.125}$$

We can define a distance

$$d(f, g) = \|f - g\| = sup\{|f(x) - g(x)| : x \in X\} \tag{1.126}$$

for arbitrary $f, g \in N$. The distance between f and g is the maximum distance between their graphs. The fact that (1.126) it is a metric follows at once from the properties (1.122):

We can define different norms for matrices.

This is well defined and finite because the image of a compact space under a continuous function is compact, in particular bounded.

Lemma 1.1.50 *Let X and Y be metric spaces and $f : X \to Y$ a continuous function. If X is compact, then f is uniformally continuous.*

Proof:

To prove the lemma we start with a number $\epsilon > 0$ and find a number $\delta > 0$ such that if $p, q \in X$ are any points such that $d_X(p, q) < \delta$ then $d_Y(f(p), f(q)) < \epsilon$.

First for each $p \in X$ we find a number $\delta(p)$ such that if $q \in X$ and

$$d_X(p,q) < \delta(p) \text{ then } d_Y(f(p), f(q)) < \epsilon/2; \tag{1.127}$$

this is always possible since f is continuous at p. Let $B(p)$ be the open ball in X with center p and radius $\delta(p)/2$. The open balls $B(p)$ with p ranging over all points of X provides an open cover for X. Since X is compact, there is a finite sub-cover. Thus there exists a finite number of points of X, say $p_1, p_2, \ldots, p_n$, such that X is the union of their open balls. Now define

$$\delta = min\{\delta(p_1)/2, \delta(p_2)/2, \ldots, \delta(p_n)/2\}$$

This will be the δ that satisfies our original demand. For suppose that $p, q \in X$, with

$$d_X(p,q) < \delta. \tag{1.128}$$

For some i we have $p \in B(p_i)$, so that

$$d_X(p_i, p) < \delta(p_i)/2.$$

Also

$$\begin{aligned} d_X(p_i, q) &\le d_X(p_i, p) + d_X(p, q) \\ &< \delta(p_i)/2 + \delta \\ &\le \delta(p_i). \end{aligned}$$

Thus

$$d_X(p_i, p), d_X(p_i, q) \le \delta(p_i)$$

Now from (1.127) this means

$$d_Y\big(f(p_i), f(p)\big) < \epsilon/2 \;\text{ and }\; d_Y\big(f(p_i), f(q)\big) < \epsilon/2.$$

Therefore

$$\begin{aligned} d_Y(f(p), f(q)) &\le d_Y\big(f(p), f(p_i)\big) + f_Y\big(f(p_i), f(q)\big) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned} \tag{1.129}$$

whenever (1.128) is satisfied.

□

We prove some facts about the differentiability, smoothness and continuity of Taylor series. Functions that have derivatives of all orders are called smooth. If a function is differentiable at a point, then it must also be continuous there (the converse isn't always true, for example the function $y = |x|$ is continuous at $x = 0$ but not differentiable there).

Before we move on to this question, let us remind the ourselves of one method used t determined if a series expansion to be absolutely convergent, namely the ratio test for absolute convergence: let $\sum a_n$ be a positive-term series, and suppose

$$\lim_{n\to\infty} \frac{a_{n+1}}{a_n} = L < 1.$$

If $L < 1$ then the series is convergent. The proof of this is as follows: suppose $\lim_{n\to\infty}(a_{n+1}/a_n) = L < 1$. Let r be any number such that $0 \le L < r < 1$. Since a_{n+1}/a_n is close to L if n is large, there exists an integer N such that whenever $n \ge N$,

$$\frac{a_{n+1}}{a_n} < r \quad \text{or} \quad a_{n+1} < a_n r.$$

Substituting N, $N+1$, $N+2$, … for n, we obtain

$$\begin{aligned} a_{N+1} &< a_N r \\ a_{N+2} &< a_{N+1} r < a_N r^2 \\ a_{N+3} &< a_{N+2} r < a_N r^3 \end{aligned} \tag{1.130}$$

and in general

$$a_{N+m} < a_N r^m \quad \text{whenever} \quad m > 0$$

Recall the basic comparison test; given two positive-terms series $\sum a_n$ and $\sum b_n$, if $\sum b_n$ converges and $a_n < b_n$ for every positive integer n greater than some N, then obviously the series $\sum a_n$ converges. It follows from the comparison test that the seies

$$a_{N+1} + a_{N+2} + \cdots a_{N+m} + \ldots$$

converges, sinse its terms are less than the correspondig terms of the convergent geometric series

$$a_N r + a_N r^2 + \cdots a_N r^3 + \ldots.$$

Since the convergence or divergence is unaffected by discarding a finite number of terms, the series $\sum_{n=0}^{\infty} a_n$.

Lemma 1.1.51 *If the series*

$$\sum_{n=0}^{\infty} a_n x^n \tag{1.131}$$

is absolutely convergent and has radius of convergence $R > 0$, then the associated function $f(x) = \sum_{n=0}^{\infty} a_n x^n$ is differentiable at every point of the range of convergence and

$$f'(x) = \sum_{n=0}^{\infty} n a_n x^{n-1} = a_1 + 2a_2 x + 3a_3 x^2 + \ldots \tag{1.132}$$

Moreover, it will be smooth.

Proof:

First we prove that if the series (1.131) has a radius of convergence R then so does

$$\sum_{n=0}^{\infty} (n+1) a_{n+1} x^n \quad \text{and} \quad \sum_{n=0}^{\infty} (n+2)(n+1) a_{n+2} x^n \tag{1.133}$$

The second of these have the same relation to the first as the first does to (1.131), so we need only consider the first. Multiplicatin by x gives the series $\sum_{n=0}^{\infty} n a_n x^n$, which therefore converges for the same values of x and has the same radius of convergence $\sum_{n=0}^{\infty} (n+1) a_{n+1} x^n$ itself.

Note

$$(y-x)(y^{n-1} + y^{n-2}x + \cdots + x^{n-1}) = y^n - x^n$$

$$\begin{aligned}
a_n \left(\frac{y^n - x^n}{y-x} - n x^{n-1} \right) &= a_n (y^{n-1} + y^{n-2} x + \cdots + x^{n-1} - n x^{n-1}) \\
&= a_n [(y^{n-1} - x^{n-1}) + (y^{n-2} - x^{n-2}) x + \cdots + (y-x) x^{n-2}] \\
&= a_n (y-x) \left[\frac{y^{n-1} - x^{n-1}}{y-x} + \frac{y^{n-2} - x^{n-2}}{y-x} x + \cdots + \frac{y-x}{y-x} x^{n-2} \right]
\end{aligned} \tag{1.134}$$

Since $|x| < r$ and $|y| < r$,

$$\left| \frac{y^k - x^k}{y - x} \right| = |y^{k-1} + y^{k-2}x + \cdots + x^{k-1}| < kr^{k-1}. \tag{1.135}$$

Using (1.135) in (1.134) we get

$$\begin{aligned} \left| a_n \left(\frac{y^n - x^n}{y - x} - nx^{n-1} \right) \right| &< |x - y| \, |a_n| \, [(n-1)r^{n-2} + (n-2)r^{n-3}r + \cdots + 2rr^{n-3} + r^{n-2}] \\ &= |x - y| \, |a_n| \, r^{n-2} \, [(n-1) + (n-2) + \cdots + 1] \\ &= \frac{1}{2} n(n-1)|x - y| \, |a_n| r^{n-2} \end{aligned} \tag{1.136}$$

The power series with coefficients $|a_n|$ has the same radius of convergence as that with a_n.

$$\begin{aligned} \sum_{n=2}^{\infty} n(n-1)|a_n|r^{n-2} &= \sum_{n=0}^{\infty} (n+2)(n+1)|a_{n+2}|r^n \\ &= K < +\infty \end{aligned} \tag{1.137}$$

we get

$$\left| \frac{f(y) - f(x)}{y - x} - \sum_{n=1}^{\infty} n a_n x^{n-1} \right| < \frac{1}{2} K |y - x| \tag{1.138}$$

proving (1.132).

$\square$

Theorem for timelike geodesics from a point

Lemma 1.1.52 *For any $s^* > 0$, there exists $C > 0$ such that if $max|A'_{\alpha\beta}(0)| \geq C$, $A(0) = I_{3\times 3}$, $Tr(A'(0)) \leq 0$, $A'(0)$ symmetric,*

$$\frac{d^2 A_{\alpha\beta}}{ds^2} + A_{\alpha\gamma} R_{\gamma 4 \beta 4} = 0 \tag{1.139}$$

then $det(A(s_1)) = 0$ for some $s_1 \in [0, s^*]$.

Proof:

As (1.139) is a linear ODE, it suffices to show that:

"For any $s^* > 0$, there exists $\sigma > 0$ such that if $max|\tilde{A}'_{\alpha\beta}(0)| = 1$, $\tilde{A}(0) = \sigma_{\tilde{A}} I_{3\times 3}$ where $0 < \sigma_{\tilde{A}} < \sigma$, $tr(\tilde{A}'(0)) \leq 0$, $\tilde{A}'(0)$ symmetric,

$$\frac{d^2\tilde{A}_{\alpha\beta}}{ds^2} + \tilde{A}_{\alpha\gamma}R_{\gamma 4\beta 4} = 0,$$

then $det(\tilde{A}(s_1)) = 0$ for some $s_1 \in [0, s^*]$."

This sufficiency is established as follows: we assume the above statement in quotation marks is correct and then show that the conditions involved in lemma 1.1.52 can be converted into the conditions involved in this above statement.

Set $max|A'_{\alpha\beta}(0)| = \tilde{C}$ and note $C \leq \tilde{C}$. Define $\tilde{A} = A/\tilde{C}$ and set

$$\sigma_{\tilde{A}} = \frac{1}{\tilde{C}} \quad \text{and} \quad \sigma = \frac{1}{C}.$$

Then say for any $s^* > 0$, there exists $\sigma > 0$ (equivalent to $C = \frac{1}{\sigma} > 0$) we have $max|\tilde{A}'_{\alpha\beta}(0)| = 1$, $\tilde{A}(0) = \sigma_{\tilde{A}} I_{3\times 3}$ where $0 < \sigma_{\tilde{A}} < \sigma$ and obviously, $tr\tilde{A}'(0) \leq 0$ and $\tilde{A}'(0)$ is symmetric, and given that (1.139) is a linear ODE we have

$$\frac{d^2\tilde{A}_{\alpha\beta}}{ds^2} + \tilde{A}_{\alpha\gamma}R_{\gamma 4\beta 4} = 0$$

which was to be proved.

Now we turn to the proof of the second form of the lemma. Obviously, at $s = 0$ we have that the eigenvalues of $A_{\alpha\beta}$ are all equal to $\sigma_A > 0$. To prove the lemma we shall show that there is a negative eigenvalue of $A_{\alpha\beta}$ for some $s_0 \in [0, s^*]$, as this implies by the mean value theorem that an eigenvalue must become zero at some point $s_1 \in [0, s_0]$ and hence $det(A)(s_1) = 0$ for some $s_1 \in [0, s^*]$. Let us proceed.

Write $A_{\alpha\beta}(s)$ as a first order Taylor polynomial plus 'remainder',

$$A_{\alpha\beta}(s) = A_{\alpha\beta}(0) + A'_{\alpha\beta}(0)s + r_{\alpha\beta}(s)$$

We can find an expression for the remainder in terms of the value of $A''_{\alpha\beta}$ for some value of s, $s_{\alpha\beta}$, in the interval $[0, s]$. To this end, if t is any number in the interval $[0, s]$, let $H_{\alpha\beta}$ be the function defined by

$$H_{\alpha\beta}(t) := A_{\alpha\beta}(s) - \Big[A_{\alpha\beta}(t) + A'_{\alpha\beta}(t)(s-t)\Big] - r_{\alpha\beta}(s)\frac{(s-t)^2}{s^2}$$

If we differentiate both sides of this equation with respect to t, regarding s as a constant, we obtain

$$H'_{\alpha\beta}(t) = -A''_{\alpha\beta}(t)(s-t) + r_{\alpha\beta}(s)\cdot 2\frac{(s-t)}{s^2}$$

By the defintion of $H_{\alpha\beta}$ we easily see that $H_{\alpha\beta}(s) = 0$. We also note that

$$\begin{aligned} H_{\alpha\beta}(0) &= A_{\alpha\beta}(s) - [A_{\alpha\beta}(0) + A'_{\alpha\beta}(0)s] - r_{\alpha\beta}(s)\frac{s^2}{s^2} \\ &= A_{\alpha\beta}(s) - [A_{\alpha\beta}(0) + A'_{\alpha\beta}(0)s + r_{\alpha\beta}(s)] \\ &= 0. \end{aligned}$$

Hence there is a number $s_{\alpha\beta}$ between 0 and s such that $H'_{\alpha\beta}(s_{\alpha\beta}) = 0$; that is

$$-A''_{\alpha\beta}(s_{\alpha\beta})(s - s_{\alpha\beta}) + r_{\alpha\beta}(s)\cdot 2\frac{(s-s_{\alpha\beta})}{s^2} = 0.$$

giving

$$r_{\alpha\beta}(s) = \frac{A''_{\alpha\beta}(s_{\alpha\beta})}{2!}s^2.$$

where $s_{\alpha\beta}$ is some function of s and where $0 \leq s_{\alpha\beta} \leq s$. Using this we obtain

$$A_{\alpha\beta}(s) = \sigma_A\delta_{\alpha\beta} + sA'_{\alpha\beta}(0) + \frac{s^2}{2}A''_{\alpha\beta}(s_{\alpha\beta}(s)), \tag{1.140}$$

which is what we wished to prove. Next we want to show that $|A_{\alpha\beta}(s)|$ on $[0, s^*]$ is bounded by a constant which is independent of the boundary conditions specified in the lemma. To this end we note that the differential equation (1.139) is equivalent to a system of first order differential equations which can be written in matrix form as

So we need to show that $|A_{ij}(s)|$ on $[0, s^*]$ is bounded by a constant which is independent of the choice of $A(0)$ and $A'(0)$.

We first explain how to derive a formal solution. We note that the differential equation $A''_{\alpha\beta} + A_{\alpha\gamma}R_{\gamma 44\beta} = 0$ is equivalent to a system of first order differential equations which can be written in matrix form as

$$\frac{d}{ds}\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{da_{\alpha\beta}}{ds}(s) \end{pmatrix} = \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ -R_{\beta 44\gamma}(s) & 0 \end{pmatrix}\begin{pmatrix} A_{\alpha\gamma}(s) \\ \frac{dA_{\alpha\gamma}}{ds}(s) \end{pmatrix}$$

This can be written as an integral equation,

$$\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{dA_{\alpha\beta}}{ds}(s) \end{pmatrix} = \begin{pmatrix} A_{\alpha\beta}(0) \\ \frac{dA_{\alpha\beta}}{ds}(0) \end{pmatrix} + \int_0^s \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ -R_{\beta 44\gamma}(s_1) & 0 \end{pmatrix}\begin{pmatrix} A_{\alpha\gamma}(s_1) \\ \frac{dA_{\alpha\gamma}}{ds}(s_1) \end{pmatrix} ds_1$$

Which can be substituted into itself,

$$\begin{aligned}\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{dA_{\alpha\beta}}{ds}(s) \end{pmatrix} &= \begin{pmatrix} A_{\alpha\beta}(0) \\ \frac{dA_{\alpha\beta}}{ds}(0) \end{pmatrix} + \int_0^s \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ -R_{\beta 44\gamma}(s_1) & 0 \end{pmatrix}\begin{pmatrix} A_{\alpha\gamma}(0) \\ \frac{dA_{\alpha\gamma}}{ds}(0) \end{pmatrix} ds_1 \\ &+ \int_0^s \begin{pmatrix} 0 & \delta_{\beta\gamma_1} \\ -R_{\beta 44\gamma_1}(s_1) & 0 \end{pmatrix}\int_0^{s_1} \begin{pmatrix} 0 & \delta_{\gamma_1\gamma} \\ -R_{\gamma_1 44\gamma}(s_2) & 0 \end{pmatrix}\begin{pmatrix} A_{\alpha\gamma}(s_2) \\ \frac{dA_{\alpha\gamma}}{ds}(s_2) \end{pmatrix} ds_2 ds_1\end{aligned}$$

Continuing in this way we obtain:

$$\begin{aligned}\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{dA_{\alpha\beta}}{ds}(s) \end{pmatrix} &= \sum_{n=0}^{\infty}\int_0^s ds_1 \int_0^{s_1} ds_2 \cdots \int_0^{s_{n-1}} ds_n \\ &\begin{pmatrix} 0 & \delta_{\beta\gamma_1} \\ -R_{\beta 44\gamma_1}(s_1) & 0 \end{pmatrix}\cdots\begin{pmatrix} 0 & \delta_{\gamma_{n-1}\gamma} \\ -R_{\gamma_{n-1}44\gamma}(s_n) & 0 \end{pmatrix}\begin{pmatrix} A_{\alpha\gamma}(0) \\ \frac{dA_{\alpha\gamma}}{ds}(0) \end{pmatrix}\end{aligned}$$

where we have put $\gamma \equiv \gamma_n$. We define the time-ordered product of two matrices $M(s_1)$ and $M(s_2)$ as

$$\mathcal{T}\{M_{JK_1}(s_1)M_{K_1K}(s_2)\} := \begin{cases} M_{JK_1}(s_1)M_{K_1K}(s_2) & : s_1 > s_2 \\ M_{JK_1}(s_2)M_{K_1K}(s_1) & : s_2 > s_1 \end{cases}$$

This easily generalises. If $s_{\alpha(1)} > s_{\alpha(2)} > \cdots > s_{\alpha(n)}$ where α is the permutation of $\{1, 2, \ldots, n\}$, then the time-ordered product of n matrices is defined as

$$\mathcal{T}\{M_{JK_1}(s_1)M_{K_1K_2}(s_2)\cdots M_{K_{n-1}K_n}(s_n)\} := M_{JK_1}(s_{\alpha(1)})M_{K_1K_2}(s_{\alpha(2)})\cdots M_{K_{n-1}K_n}(s_{\alpha(n)}).$$

It can be shown that, we can then formally write the solution as:

$$\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{dA_{\alpha\beta}}{ds}(s) \end{pmatrix} = \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^s ds_1 \int_0^s ds_2 \cdots \int_0^s ds_n$$
$$\mathcal{T}\left\{ \begin{pmatrix} 0 & \delta_{\beta\gamma_1} \\ -R_{\beta 44\gamma_1}(s_1) & 0 \end{pmatrix} \cdots \begin{pmatrix} 0 & \delta_{\gamma_{n-1}\gamma} \\ -R_{\gamma_{n-1}44\gamma}(s_n) & 0 \end{pmatrix} \right\} \begin{pmatrix} A_{\alpha\gamma}(0) \\ \frac{dA_{\alpha\gamma}}{ds}(0) \end{pmatrix}$$

or

$$\begin{pmatrix} A_{\alpha\beta}(s) \\ \frac{dA_{\alpha\beta}}{ds}(s) \end{pmatrix} = \mathcal{T}\exp\left\{ \int_0^s \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ -R_{\beta 44\gamma}(t) & 0 \end{pmatrix} dt \right\} \begin{pmatrix} A_{\alpha\gamma}(0) \\ \frac{dA_{\alpha\gamma}}{ds}(0) \end{pmatrix}$$

We see from this that any solution for $a_{\alpha\beta}(s)$ will be of the general form

$$A_{\alpha\beta}(s) = A_{\alpha\gamma}(0) f_{\gamma\beta}(s) + \frac{d}{ds} A_{\alpha\gamma}(0) h_{\gamma\beta}(s)$$

where the functions $f_{\gamma\beta}(s)$ and $h_{\gamma\beta}(s)$ are solely determined by the matrix

$$\mathcal{T}\exp\left\{ \int_0^s \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ -R_{\beta 44\gamma}(t) & 0 \end{pmatrix} dt \right\}$$

and as such are independent of the initial conditions placed on a_{ij}. So we want to show that this matrix is bounded on $[0, s^*]$!

If we define

$$Q_{\beta\gamma}(s) := -R_{\beta 44\gamma}(s)$$

$$
\begin{aligned}
& \mathcal{T}\exp\left\{\int_0^s \begin{pmatrix} 0 & s\delta_{\beta\gamma} \\ sQ_{\beta\gamma} & 0 \end{pmatrix} dt\right\} \\
= & \begin{pmatrix} \delta_{\beta\gamma} & 0 \\ 0 & \delta_{\beta\gamma} \end{pmatrix} + \int_0^s ds_1 \begin{pmatrix} 0 & \delta_{\beta\gamma} \\ Q_{\beta\gamma}(s_1) & 0 \end{pmatrix} \\
+ & \frac{1}{2!}\int_0^s ds_1 \int_0^s ds_2 \mathcal{T}\left\{\begin{pmatrix} \delta_{\beta\gamma_1}Q_{\gamma_1\gamma}(s_2) & 0 \\ 0 & Q_{\beta\gamma_1}(s_1)\delta_{\gamma_1\gamma} \end{pmatrix}\right\} \\
+ & \frac{1}{3!}\int_0^s ds_1 \int_0^s ds_2 \int_0^s ds_3 \mathcal{T}\left\{\begin{pmatrix} 0 & \delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}()\delta_{\gamma_2\gamma} \\ Q_{\beta\gamma_1}()\delta_{\gamma_1\gamma_2}Q_{\gamma_2\gamma}() & 0 \end{pmatrix}\right\} \\
+ & \frac{1}{4!}\int_0^s ds_1 \int_0^s ds_2 \int_0^s ds_3 \int_0^s ds_4 \\
& \mathcal{T}\left\{\begin{pmatrix} \delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}(s_2)\delta_{\gamma_2\gamma_3}Q_{\gamma_3\gamma}(s_4) & 0 \\ 0 & Q_{\beta\gamma_1}(s_1)\delta_{\gamma_1\gamma_2}Q_{\gamma_2\gamma_3}(s_3)\delta_{\gamma_3\gamma} \end{pmatrix}\right\} \\
+ & \frac{1}{5!}\int_0^s ds_1 \int_0^s ds_2 \int_0^s ds_3 \int_0^s ds_4 \int_0^s ds_5 \\
& \mathcal{T}\left\{\begin{pmatrix} 0 & \delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}(s_2)\delta_{\gamma_2\gamma_3}Q_{\gamma_3\gamma}(s_4)\delta_{\gamma_4\gamma} \\ Q_{\beta\gamma_1}(s_1)\delta_{\gamma_1\gamma_2}Q_{\gamma_2\gamma_3}(s_3)\delta_{\gamma_3\gamma_4}Q_{\gamma_4\gamma}(s_5) & 0 \end{pmatrix}\right\} \\
+ & \frac{1}{6!}\int_0^s ds_1 \int_0^s ds_2 \int_0^s ds_3 \int_0^s ds_4 \int_0^s ds_5 \int_0^s ds_6 \\
& \mathcal{T}\left\{\begin{pmatrix} \delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}(s_2)\delta_{\gamma_2\gamma_3}Q_{\gamma_3\gamma}(s_4)\delta_{\gamma_4\gamma_4}Q_{\gamma_4\gamma}(s_6) & 0 \\ 0 & Q_{\beta\gamma_1}(s_1)\delta_{\gamma_1\gamma_2}Q_{\gamma_2\gamma_3}(s_3)\delta_{\gamma_3\gamma_4}Q_{\gamma_4\gamma}(s_5)\delta_{\gamma_4\gamma} \end{pmatrix}\right\} + \dots
\end{aligned}
$$

The function $f_{\gamma\beta}(s)$ corresponds to the top left hand quadrant:

$$
\begin{aligned}
f_{\gamma\beta}(s) &= \delta_{\beta\gamma} + \sum_{n=1}^{\infty} \frac{1}{(2n)!} \int_0^s ds_1 \cdots \int_0^s ds_{2n} \mathcal{T}\{\delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}(s_2) \cdots \delta_{\gamma_{2n}\gamma_{2n-1}}Q_{\gamma_{2n-1}\gamma}(s_{2n})\} \\
&= \delta_{\beta\gamma} + \sum_{n=1}^{\infty} \frac{s^n}{(2n)!} \int_0^s ds_1 \cdots \int_0^s ds_n \mathcal{T}\{Q_{\beta\gamma_1}(s_1) \cdots Q_{\gamma_{n-1}\gamma}(s_n)\}
\end{aligned}
$$

The function $h_{\gamma\beta}(s)$ corresponds to the top right hand quadrant:

$$
\begin{aligned}
h_{\gamma\beta}(s) &= s\delta_{\beta\gamma} + \sum_{n=1}^{\infty} \frac{1}{(2n+1)!} \int_0^s ds_1 \cdots \int_0^s ds_{2n+1} \mathcal{T}\{\delta_{\beta\gamma_1}Q_{\gamma_1\gamma_2}(s_2) \cdots \delta_{\gamma_{2n}\gamma_{2n-1}}Q_{\gamma_{2n-1}\gamma_{2n}}(s_{2n})\delta_{\gamma_{2n}\gamma}\} \\
&= s\delta_{\beta\gamma} + \sum_{n=1}^{\infty} \frac{s^{n+1}}{(2n+1)!} \int_0^s ds_1 \cdots \int_0^s ds_n \mathcal{T}\{Q_{\beta\gamma_1}(s_1) \cdots Q_{\gamma_{n-1}\gamma}(s_n)\}
\end{aligned}
$$

We have

$$|f_{\gamma\beta}(s)| \leq \delta_{\beta\gamma} + \sum_{n=1}^{\infty} \frac{s^n}{(2n)!} \int_0^s ds_1 \cdots \int_0^s ds_n T\{|Q_{jk_1}(s_1)| \cdots |Q_{k_{n-1}k}(s_n)|\} \tag{1.141}$$

Define q as

$$q := \{max|Q_{\beta\gamma}(s)| \text{ for all } \beta, \gamma \text{ and } s \in [0, s^*]\}$$

and define the matrix D to be the 3×3 matrix with all entries equal to 1. Then obviously

$$|Q_{\beta\gamma_1}(s_1)| \cdots |Q_{k_{n-1}k}(s_n)| \leq q^n (D^n)_{jk} = 3^{n-1} q^n D_{jk}$$

Using this in (1.141)

$$|f_{\gamma\beta}(s)| \leq \delta_{\beta\gamma} + D_{\beta\gamma} \sum_{n=1}^{\infty} \frac{s^{2n}}{(2n)!} 3^{n-1} q^n$$

which obviously converges. Similarly, for $h_{\gamma\beta}(s)$ we have

$$|h_{\gamma\beta}(s)| \leq s\delta_{\beta\gamma} + D_{\beta\gamma} \sum_{n=1}^{\infty} \frac{s^{2n+1}}{(2n+1)!} 3^{n-1} q^n$$

Therefore both the functions $|f_{\gamma\beta}(s)|$ and $|h_{\gamma\beta}(s)|$ are bounded on $[0, s^*]$. Now applying the conditions of the lemma, namely that $A_{ij}(0) = \epsilon_A \delta_{\alpha\beta}$ and $\max|A'_{\alpha\beta}(0)| = 1$, we deduce:

$$\begin{aligned} |A_{\alpha\beta}(s)| &\leq \epsilon_A |f_{\alpha\beta}(s)| + \sum_k \frac{d}{ds} |A_{\alpha\gamma}(0) h_{\gamma\beta}(s)| \\ &\leq \epsilon_A |f_{\alpha\beta}(s)| + \sum_k |\frac{d}{ds} A_{\alpha\gamma}(0)| |h_{\gamma\beta}(s)| \\ &\leq \epsilon |f_{\alpha\beta}(s)| + \sum_k |h_{\gamma\beta}(s)| \end{aligned}$$

where we have included the summation for clarity. Thus we see that $A_{\alpha\beta}(s)$ on $[0, s^*]$ is bounded by a constant which is independent of of the boundary conditions.

Now, as $R_{4\alpha 4\gamma}$ is bounded on $[0, s^*]$, by $A''_{ij} + A_{ik}R_{k44j} = 0$, $A''_{\alpha\beta}$ is also bounded on $[0, s^*]$. Hence,

$$|A''_{\alpha\beta}(s_{\alpha\beta})| \leq C_{s^*}, \tag{1.142}$$

where C_{s^*} is independent of the choice of $A(0)$ and $A'(0)$.

□

In the next step we will need the following fact:

$$\lambda_{min} = min_{\{x_1^2+x_2^2+x_3^2=1\}} A'_{\alpha\beta}(0)x_\alpha x_\beta.$$

where λ_{min} is the minimum eigenvalue of $A'_{\alpha\beta}(0)$, which we proceed to prove. Let $v^{(m)}_\alpha$ be a normalised eigenvector of $A'_{\alpha\beta}(0)$. Then any vector, x, can be expanded as $x_\alpha = \sum_m c_m v^{(m)}_\alpha$, where x can be chosen unit by requiring $\sum_m c_m^2 = 1$. The quantity $A'_{\alpha\beta}x_\alpha x_\beta$ is expressed as

$$x_\alpha A'_{\alpha\beta}(0)x_\beta = \sum_{m,m'} c_m v^{(m)}_\alpha A'_{\alpha\beta}(0) c_{m'} v^{(m')}_\beta = \sum_{m,m'} c_m c_{m'} \lambda_{m'} v^{(m)}_\alpha v^{(m')}_\alpha = \sum_m c_m^2 \lambda_m$$

From which we obviously have

$$\lambda_{min} = \sum_m c_m^2 \lambda_{min} \leq \sum_m c_m^2 \lambda_m = x_\alpha A'_{\alpha\beta}(0) x_\beta$$

□

Next, $A'_{\alpha\beta}(0)$ is symmetric. Let λ_1 λ_2, and λ_3 be the eigenvalues of $A'_{\alpha\beta}$ with $\lambda_1 \leq \lambda_2 \leq \lambda_3$. We claim that $\lambda_1 \leq -\frac{1}{2}$. We will be using

$$\lambda_1 = min_{\{x_1^2+x_2^2+x_3^2=1\}} A'_{\alpha\beta}(0)x_\alpha x_\beta.$$

Case 1. $|A'_{\alpha\alpha}(0)| = 1$ for some α (no summation over α). Without loss of generality, we can assume that $\alpha = 1$. For the case where $A'_{11}(0) = +1$, $A'_{11}(0) + A'_{22}(0) + A'_{33}(0) \leq 0$ implies $A'_{22}(0) + A'_{33}(0) \leq -1$. Which means that either $A'_{22}(0)$ or $A'_{33}(0)$ is less than or equal to $-\frac{1}{2}$. By taking $x = (1, 0, 0)$ we have $\lambda_1 \leq A'_{11}(0)$, by taking $x = (0, 1, 0)$ we have $\lambda_1 \leq A'_{22}(0)$, and by taking $x = (0, 0, 1)$ we have $\lambda_1 \leq A'_{33}(0)$. As a result we have $\lambda_1 \leq -\frac{1}{2}$. For $A'_{11}(0) = -1$, we have $A'_{22}(0) + A'_{33}(0) \leq +1$, implying $\lambda_1 \leq A'_{11}(0) = -1$.

Case 2. $|A'_{\alpha\beta}(0)| = 1$ for some $\alpha \neq \beta$. Without loss of generality, we assume $\alpha = 1$ and $\beta = 2$. For $A'_{12}(0) = +1$ we let $x = (1/\sqrt{2}, -1/\sqrt{2}, 0)$. We have

$$
\begin{aligned}
\lambda_1 &\le (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0) \begin{pmatrix} A'_{11}(0) & A'_{12}(0) & \cdot \\ A'_{12}(0) & A'_{22}(0) & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \\
&= \frac{1}{2}(1, -1, 0) \begin{pmatrix} A'_{11}(0) - A'_{12}(0) \\ A'_{12}(0) - A'_{22}(0) \\ \cdot \end{pmatrix} \\
&= \frac{1}{2}(A'_{11}(0) - 2A'_{12}(0) + A'_{22}(0)) \\
&= \frac{1}{2}(A'_{11}(0) - 2 + A'_{22}(0)). \qquad (1.143)
\end{aligned}
$$

If $A'_{11}(0) + A'_{22}(0) \le 1$, then $\lambda_1 \le -\frac{1}{2}$. If $A'_{11}(0) + A'_{22}(0) \ge 1$, then $A'_{33}(0) \le -1$ (by $Tr(A'(0)) \le 0$). So λ_1 is still less than or equal to -1. For the case $A'_{12}(0) = -1$, we take $x = (1/\sqrt{2}, 1/\sqrt{2}, 0)$ in which case

$$
\begin{aligned}
\lambda_1 &\le (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0) \begin{pmatrix} A'_{11}(0) & A'_{12}(0) & \cdot \\ A'_{12}(0) & A'_{22}(0) & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \\
&= \frac{1}{2}(1, 1, 0) \begin{pmatrix} A'_{11}(0) + A'_{12}(0) \\ A'_{12}(0) + A'_{22}(0) \\ \cdot \end{pmatrix} \\
&= \frac{1}{2}(A'_{11}(0) + 2A'_{12}(0) + A'_{22}(0)) \\
&= \frac{1}{2}(A'_{11}(0) - 2 + A'_{22}(0)). \qquad (1.144)
\end{aligned}
$$

Proceeding as before, we also find $\lambda_1 \le -1$. The most definitive statement we can conclude from all of the above is that

$$
\lambda_{min} \le -\frac{1}{2} \qquad (1.145)
$$

which is what we set out to prove.

By the above,

$$
|A'_{\alpha\beta}(0) + \frac{s}{2}A''_{\alpha\beta}(s_{\alpha\beta}) - A'_{\alpha\beta}(0)| \le \frac{s}{2}C_{s^*}. \qquad (1.146)
$$

This implies the deviation from $A'_{\alpha\beta}(0)$ can be made arbtrarily small by suitable choice of s. Now, the eigenvalues of a 3 by 3 matrix are roots of a cubic equation which has an explicit formula to solve, the minimum eigenvalue is a continuous function of its entries. Thus it is always possible, by choising s_0 small enough, so that the new λ_1 is less than or equal to say $-\frac{1}{4}$. In actual fact this choice can be made independent of $A'_{\alpha\beta}(0)$. We now turn to the technical part of the proof.

We denote by A the space given in the lemma, namely

$$A := \{A_{\alpha\beta}(0) = \sigma_A \delta_{\alpha\beta}; A'_{\alpha\beta}(0) = A'_{\beta\alpha}(0);\ max|A'_{\alpha\beta}(0)| = 1;\ tr(A'_{\alpha\beta}(0)) \leq 0)\}. \quad (1.147)$$

This is compact by the Heine-Borel theorem (a subset of $\mathbb{R}$ that is closed and bounded is compact). Then define the space $\tilde{A}$ by

$$\tilde{A} := [0, s] \times A \quad (1.148)$$

which is also compact. Define the function h by

$$h(s; A_{\alpha\beta}(0), A'_{\alpha\beta}(0)) := A'_{\alpha\beta}(0) + \frac{s}{2} A''_{\alpha\beta}(s_{\alpha\beta}(s)) \quad (1.149)$$

$$h : \tilde{A} \to B.$$

And by (??) we have

$$h(s; A_{\alpha\beta}(0), A'_{\alpha\beta}(0)) := \sigma_A h_{\alpha\beta}(s) + g_{\alpha\gamma}(s) A'_{\alpha\beta}(0) - \sigma_A \delta_{\alpha\beta} \quad (1.150)$$

from which we see that h is a continuous function over $\tilde{A}$ implying that B is compact and bounded. Now the minimum eigenvalue λ is a continuous function over B

$$\lambda : B \to C \subset \mathbb{R}.$$

implying it is bounded. Moreover, as it is a continuous function over a compact space it is also uniformally continuous, that is, given any $\epsilon > 0$ there exists a $\delta > 0$ such that if $p, q \in B$ and

$$d_B(p, q) < \delta$$

then

$$|\lambda(p) - \lambda(q)| < \epsilon.$$

We pick p to correspond to $h(s = 0; A'_{\alpha\beta}(0)) \in B$ and pick ϵ to be equal to $1/4$, then we can find a δ

$$\begin{aligned} & |h(s = 0; A'_{\alpha\beta}(0)) - h(s; A'_{\alpha\beta}(0))| \\ = & \ |A'_{\alpha\beta}(0) + \frac{s}{2} A''_{\alpha\beta}(s_{\alpha\beta}) - A'_{\alpha\beta}(0)| < \delta, \end{aligned} \tag{1.151}$$

such that

$$\left|\lambda\big(A'_{\alpha\beta}(0) + \frac{s}{2} A''_{\alpha\beta}(s_{\alpha\beta})\big) - \lambda\big(A'_{\alpha\beta}(0)\big)\right| < 1/4. \tag{1.152}$$

From reference to (1.146), we see this δ can be achieved by taking s to satisfy

$$\frac{sC_{s^*}}{2} < \delta$$

or

$$s < \frac{\delta}{2C_{s^*}}$$

Let s_0 denote the value of s that corresponds to $\epsilon = 1/4$. The choice s_0 is independent of $A_{\alpha\beta}(0)$ and $A'_{\alpha\beta}(0)$ and follows from the fact that the minimum eigenvalue is a continuous function over a compact space and as such is uniformally continuous.

Now we turn to the full solution with $s = s_0$,

$$A_{\alpha\beta}(s_0) = \sigma_A \delta_{\alpha\beta} + s_0 (A'_{\alpha\beta}(0) + \frac{s_0}{2} A_{\alpha\beta})). \tag{1.153}$$

Note that what we have established is that the minimum eigenvalue of the matrix $A'_{\alpha\beta}(0) + \frac{s_0}{2} A_{\alpha\beta}(s_{\alpha\beta(s)})$ is $\leq -(1/4)$. Actually what we really need is the corresponding reuslt for this matrix multiplied by $s_0 > 0$, namely the matrix $s_0(A'_{\alpha\beta}(0) + \frac{s_0}{2} a_{\alpha\beta}))$. It is easy to see then for this matrix the minnimum eigenvalue is $\leq -(s_0/4)$.

Now consider (1.153). Say we wanted to find the eigenvalues of a matrix M, we would solve the equation

$$det|M - \lambda| = 0. \quad (1.154)$$

What if we then added $\sigma_A I$ to M, what would be the new eigenvalues? We would have to solve

$$det|(M + \sigma_A I) - \lambda_{new} I| = 0$$

which is the same as solving

$$det|M - (\lambda_{new} - \sigma_A) I| = 0 \quad (1.155)$$

Comparison of (1.154) with (1.155) implies

$$\lambda_{new} = \lambda + \sigma_A.$$

Obviously, by adding the constant σ_A times the unit matrix all this does is shift the value of the eigenvalues of M by the positive number σ_A. By choosing σ small enough we can ensure that the minimum eigenvalue is less than or equal to say $-\frac{s_0}{8}$ (namely, with the choice $\sigma = s_0/8$).

We have now established that the full solution for $s = s_0$ has a minimum eigenvalue $\leq -(s_0/8)$ provided we take σ small enough. Now, all the einganvalues of $A_{\alpha\beta}(0)$ are $\sigma_A > 0$, but we have shown that there is a negative eigenvalue of $A_{\alpha\beta}(s_0)$. By the mean value theorem, there exists a zero eigenvalue of $A_{\alpha\beta}(s_1)$ for some $s_1 \in [0, s_0]$. So $det(A)(s_1) = 0$.

$\square$

Proposition 1.1.53 *If the strong energy condition holds and the generality condition holds at a point then there is a pair of conjugate points along any causal geodesic provided that the geodesic can be extended far enouigh.*

Proof: We break the proof into four parts

a) Prove there is a conjugate point to the future in the case of non-positive expansion, $\theta \leq 0$.

b) The map η that takes you from the inital condition, using the Jacobi equation, to the first focal point is a continuous function.

c) For $S = \{b \text{ is a 3 by 3 symmetric matrix with} tr(b) \leq 0\}$. We prove that $\eta(S)$ is bounded.

d) Take a future focal point and prove the existence of a past focal point.

Let γ be a time-like geodesic.

A solution for a Jacobi field $A_{\alpha\beta}(s)$ (satisfying (1.90)) along $\gamma(s)$ is uniquely determined by the value of $A_{\alpha\beta}$ and $dA_{\alpha\beta}/ds$ at p. Consider the space of solutions for which

$$A_{\alpha\beta}|_p = \delta_{\alpha\beta} \quad \text{and} \quad \frac{d}{ds}A_{\alpha\beta}|_p \quad \text{is a symmetric matrix with trace } \theta|_p \leq 0. \tag{1.156}$$

We specify

$$J_\alpha(s) = \sum_{al=1} A_{\alpha\beta}(s)e_\beta(0). \tag{1.157}$$

This implies

$$\frac{d^2}{ds^2}\sum_{\beta=1} A_{\alpha\beta}(s)e_\beta(0) = -R_{\alpha 4\gamma 4}\sum_{\beta=1} A_{\alpha\beta}(s)e_\beta(0) \tag{1.158}$$

or

$$\frac{d^2}{ds^2}A_{\alpha\beta} = -R_{\alpha 4\gamma 4}A_{\alpha\beta} \tag{1.159}$$

along γ.

We let $S = \{b \text{ is a 3 by 3 symmetric matrix with } tr(b) \leq 0\}$. we claim for ony $b \in S$, if $A_{\alpha\beta}(0) = I_{3\times 3}$, $A'_{\alpha\beta}(0) = b$

Proof a):

First we prove that for all such solutions there is a finite interval on $\gamma(s)$ from the point $\gamma(s_1)$ where the matrix $A_{\alpha\beta}(s)$ first becomes singular.

First, for each solution in P there will be some $s_3 > s_1$ for which $A_{\alpha\beta}(s_3)$ is singular, since either

$$\omega_{\alpha\beta}(s) = -\frac{1}{2}(A^{-1}_{\gamma\alpha}(s)\frac{d}{ds}A_{\beta\gamma}(s) - A^{-1}_{\gamma\beta}(s)\frac{d}{ds}A_{\alpha\gamma}(s))$$

So for $\omega_{\alpha\beta}(0)$ we find

$$
\begin{aligned}
\omega_{\alpha\beta}(0) &= -\frac{1}{2}(A^{-1}_{\gamma\alpha}(0)A'_{\beta\gamma}(0) - A^{-1}_{\gamma\beta}(0)A'_{\alpha\gamma}(0)) \\
&= -\frac{1}{2}(\delta_{\gamma\alpha}A'_{\beta\gamma}(0) - \delta_{\gamma\beta}A'_{\alpha\gamma}(0)) \\
&= -\frac{1}{2}(A'_{\beta\alpha}(0) - A'_{\alpha\beta}(0)) \\
&= 0.
\end{aligned}
$$

By

$$\frac{d}{ds}(A_{\gamma\alpha}(s)\omega_{\gamma\delta}(s)A_{\delta\beta}(s)) = 0$$

$\omega_{\alpha\beta}(s)$ will be zero everywhere along $\gamma(s)$ where $A_{\alpha\beta}(s)$ is non-singular.

(i) $\theta|_p < 0$, in which case it follows from the focussing theorem, or

For $tr(b) < 0$, we have at the origin (using (1.112))

$$\theta(0) = tr(A'_{\alpha\beta}(0)A^{-1}_{\alpha\beta}(0)) = tr(b) < 0.$$

So by the focussing theorem (proporsition 1.1.43) there will be a point where $det(A(s)) = 0$ for some $s > 0$. For $tr(b) = 0$ we have $\theta(0) = 0$.

(ii) $tr(b) = 0$ or $\theta|_p = 0$: First suppose that

$$R_{ab}\gamma'^a(0)\gamma'^b(0) + \sigma_{ab}\sigma^{ab} = 0$$

as $\sigma_{ab}\sigma^{ab}$ is nonnegative this implies,

$$R_{ab}\gamma'^a(0)\gamma'^b(0) = 0 \quad \text{and} \quad \sigma_{ab}\sigma^{ab} = 0$$

By the Raychaudhuri equation for shear (1.49) reduces to

$$\frac{d\sigma_{ab}}{ds} = R_{cbad}V^cV^d \tag{1.160}$$

at $\gamma(0)$. Since we have a non-zero tidal force at $\gamma(0)$ this implies

$$\frac{d\sigma_{ab}}{ds} \neq 0$$

at $\gamma(0)$, which implies

$$\sigma_{ab}\sigma^{ab} > 0$$

on an open segment of γ whose closure includes $\gamma(0)$. By the Raychaudhuri equation

$$\frac{d}{ds}\theta = -R_{ab}V^aV^b - \sigma_{cd}\sigma^{cd} - \frac{1}{3}\theta^2 \tag{1.161}$$

positive $\sigma_{cd}\sigma^{cd}$, ($R_{ab}V^aV^b \geq 0$ for $s > 0$), will cause θ to become negative for $s > 0$. Again we will have $det(A(s)) = 0$ for some $s > 0$. So the Weyl tensor produces convergence indirectly by inducing shear.

Now suppose that

$$R_{ab}\gamma'^a(0)\gamma'^b(0) + \sigma_{ab}\sigma^{ab} \neq 0$$

then by the Raychaudhuri equation we have $det(A(s)) = 0$ for some $s > 0$.

Proof b):

Let $S = \{b \text{ is a 3 by 3 symmetric matrix with } tr(b) \leq 0\}$. Let $\eta : S \to [0, +\infty)$ such that $\eta(b) = min\{s \in [0, +\infty] : \ det(A(s)) = 0 \text{ with } A_{\alpha\beta}(0) = I_{3\times 3},\ A'_{\alpha\beta}(0) = b \text{ and } \frac{d^2}{ds^2}A_{\alpha\beta} + A_{\alpha\gamma}R_{\gamma 44\beta} = 0\}$. We claim η is continuous.

Suppose the map is not continuous at some point $b \in S$. Then there exists a $\epsilon > 0$ such that for all $n \in \mathbb{N}$, there exists some $b_n \in S$ with $max|(b_n)_{\alpha\beta} - b_{\alpha\beta}| < \frac{1}{n}$ but

$$|\eta(b_n) - \eta(b)| \not< \epsilon. \tag{1.162}$$

Note condition (1.162) is equivalent to

$$\eta(b_n) \notin (\eta(b) - \epsilon, \eta(b) + \epsilon). \tag{1.163}$$

Case i) $tr(b) < 0$ at $\gamma(0)$.

We have that b_n is assumed to converge to b. Now even though we may have $tr(b_n) = 0$ for some n, the series $\{tr(b_n)\}$ cannot converge to zero as $tr(b)$ is assumed to be less than zero. Then for some $n > N \in \mathbb{N}$ we will have that

$$tr(b_n) < \frac{tr(b)}{2}.$$

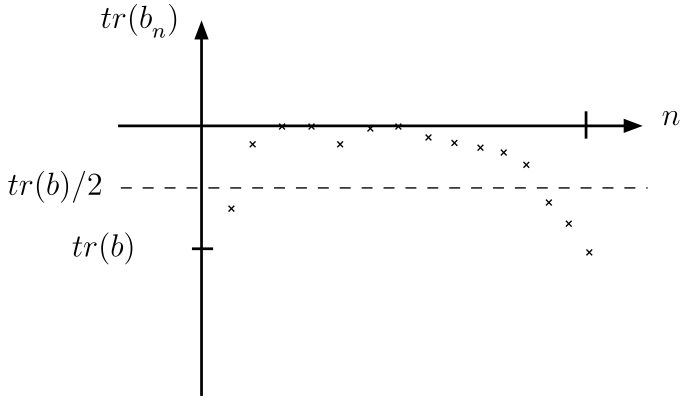


Figure 1.64: Since $b_n \to b$, $tr(b_n) < tr(b)/2$ for all $n > N$.

(see fig 5.63). Recall that the focussing theorem maintains if $\theta(s) < 0$ then θ will tend to $-\infty$ within $\Delta s \leq 3/|\theta(s)|$. Hence $\eta(b_n) \in [0, -6/tr(b)]$. As $[0, -6/tr(b)]$ is a compact set, $\eta(b_n)$ converges to $\xi \in [0, -6/tr(b)]$, where, by assumption of discontinuity, $\xi \neq \eta(b)$.

As

$$(A_{b_n})_{\alpha\beta}(s) = h_{\alpha\beta}(s) + g_{\alpha\gamma}(s)(b_n)_{\gamma\beta}$$

is obviously continuous with respect to the components of b_n, we have

$$A_{b_n}(s) \to A_b(s). \tag{1.164}$$

as $b_n \to b$. Indeed, $\eta(b_n)$ converging to ξ means that

$$\begin{aligned} |A_{b_n}(\eta(b_n)) - A_b(\xi)| &= |A_{b_n}(\eta(b_n)) - A_{b_n}(\xi) + A_{b_n}(\xi) - A_b(\xi)| \\ &\leq |A_{b_n}(\eta(b_n)) - A_{b_n}(\xi)| + |A_{b_n}(\xi) - A_b(\xi)| \end{aligned} \tag{1.165}$$

By the continuity of $A_{b_n}(s)$ with respect to s and (1.164)

$$A_{b_n}(\eta(b_n)) \to A_b(\xi)$$

As the determinant is a continuous function of the matrix elements and

$$det(A_{b_n}(\eta(b_n)) = 0 \quad \text{for all } n,$$

we have

$$det(A_b(\xi)) = 0. \tag{1.166}$$

If $\xi < \eta(b)$, then by (1.166) there is a contradiction since $\eta(b)$ is the first point where $det(A_b(s)) = 0$. If $\xi > \eta(b)$, then there is a sequence of expansions $\theta(h_n)$ associated to b_n such that $\theta(h_n)$ tends to $-\infty$ when h_n tends to $\eta(b)$ from the 'left'; $\{h_n\}$ is a sequence of values of s tending to $\eta(b)$ and $\theta(h_n)$ is the expansion corresponding to the solution with intial condition b_n with the sequence $\{h_n\}$ chosen such that $\theta(h_n)$ is finite for finite n. Since h_n converges to $\eta(b)$, for $n > N \in \mathbb{N}$ we have

$$h_n \in (\eta(b) - \epsilon, \eta(b)).$$

Again by the focussing theorem if $\theta(s) < 0$ then θ will tend to $-\infty$ within $\Delta s \leq 3/|\theta(s)|$; for large n (and with $n > N$) the expansion θ corresponding to b_n is so negative that the first point where $det(A_{b_n}) = 0$, i.e. $\eta(b_n)$, is within the distance

$$\Delta s \leq \epsilon.$$

This is achieved for $|\theta(h_n)|$ satisfying

$$|\theta(h_n)| \geq 3/\epsilon.$$

Therefore $\eta(b_n)$ lies inside the open interval $(\eta(b) - \epsilon, \eta(b) + \epsilon)$ in contradiction to the assumption (1.163).

Case ii) $tr(b) = 0$ at $\gamma(0)$. By the above argument, we have $tr(b) < 0$ on an open interval whose closure contains $\gamma(0)$. We apply the same arguent as in case i) and find a contradiction.

Proof c):

We prove that $\eta(S)$ is bounded, i.e. $\eta(S) \subseteq [0, s_5]$. First let $C > 0$ be a constant. Define

$$S_C := \{b \in S : max|b_{\alpha\beta}| > C\}.$$

By lemma 1.1.52, $\eta(S_C)$ is bounded. The set $S - S_C$ corresponds to the subspace of S consisting of all matrices whose component's absolute values are less than or equal to C, i.e.,

$$S - S_C = \{b \in S : max|b_{\alpha\beta}| \leq C\}.$$

Being closed and bounded it is compact. As η is continuous $\eta(S - S_C)$ will be bouded by proposition 1.1.49.

This means that there is some $s_5 > s_1$ such that $\eta(S)$ is contained in the segment from $\gamma(s_1)$ to $\gamma(s_5)$.

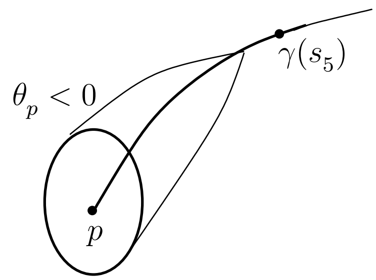


Figure 1.65: $\eta(S)$ is contained in the bounded segment from $\gamma(s_1)$ to $\gamma(s_5)$.

Proof d):

Then we consider a point r beyond the this interval. Suppose that there is no conjugate to $\gamma(s_5)$ along $\gamma[0, s_5]$. Otherwise, the propositin is done. We will see that it follows that there must be a point say q before $\gamma(s_1)$ conjugate to r. First we need to prove there is a Jacobi field J along γ such that $J_\alpha(0) = e_\alpha$ and $J_\alpha(s_5) = 0$. Let us write the general solution of the equation (1.139) in terms of the boundary conditions given at the point $r = \gamma(s_5)$,

$$A_{\alpha\beta}(s) = h_{\alpha\gamma}(s - s_5)A_{\gamma\beta}(s_5) + g_{\alpha\gamma}(s - s_5)\frac{d}{ds}A_{\gamma\beta}(s_5) \tag{1.167}$$

At $s = 0$, this is

$$A_{\alpha\beta}(0) = h_{\alpha\gamma}(-s_5)A_{\gamma\beta}(s_5) + g_{\alpha\gamma}(-s_5)\frac{d}{ds}A_{\gamma\beta}(s_5) \tag{1.168}$$

We choose

$$A_{\gamma\beta}(s_5) = 0, \tag{1.169}$$

to make $J_\alpha(s_5) = 0$. Then we have

$$A_{\alpha\beta}(0) = g_{\alpha\gamma}(-s_5)\frac{d}{ds}A_{\gamma\beta}(s_5). \tag{1.170}$$

We consider the solutions for which $\frac{d}{ds}A_{\gamma\beta}(s_5)$ is non-singular. As we are taking $A_{\alpha\beta}(0)$ to be non-singular, by $detA = detg\ det\frac{d}{ds}A_{\gamma\beta}(s_5)$, $g_{\alpha\gamma}(-s_5)$ is also non-singular. Let us choose

$$\frac{d}{ds}A_{\gamma\beta}(s_5) = (g^{-1}(-s_5))_{\gamma\beta} \tag{1.171}$$

so that

$$A_{\alpha\beta}(0) = \delta_{\alpha\beta}. \tag{1.172}$$

By this chioce $\frac{d}{ds}A_{\gamma\beta}(s_5)$ is symmetric.

The corresponding Jacobi field will have zero vorticity: in general we have

$$A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta} = \frac{1}{2}\left(A_{\gamma\alpha}\frac{d}{ds}A_{\gamma\beta} - A_{\gamma\beta}\frac{d}{ds}A_{\gamma\alpha}\right)$$

the Jacobi field vanishes at $\gamma(s_5)$ so

$$A_{\gamma\alpha}\omega_{\gamma\delta}A_{\delta\beta} = 0$$

at $s = s_5$. As $A_{\alpha\beta}$ has an inverse on $\gamma[0, s_5)$. Therefore $\omega_{\alpha\beta} = 0$ on $\gamma[0, s_5)$. From this we see that

$$0 = \omega_{\alpha\beta}(0) = -\frac{1}{2}(A^{-1}_{\gamma\alpha}(0)A'_{\beta\gamma}(0) - A^{-1}_{\gamma\beta}(0)A'_{\alpha\gamma}(0)) \tag{1.173}$$

this implies

$$\begin{aligned} 0 &= A^{-1}_{\gamma\alpha}(0)A'_{\beta\gamma}(0) - A^{-1}_{\gamma\beta}(0)A'_{\alpha\gamma}(0) \\ &= \delta_{\gamma\alpha}A'_{\beta\gamma}(0) - \delta_{\gamma\beta}A'_{\alpha\gamma}(0) \\ &= A'_{\beta\alpha}(0) - A'_{\alpha\beta}(0) \end{aligned} \tag{1.174}$$

so $A'_{\alpha\beta}(0)$ is symmetric. We have now established a family of Jacobi fields that zanish at $\gamma(s_5)$ and correspond to $A'_{\alpha\beta}(0)$ symmetric, $A_{\alpha\beta}(0) = \delta_{\alpha\beta}$ and zero vorticity.

From this we see that if there is no point conjugate to $\gamma(s_5)$ between $\gamma(s_5)$ and $\gamma(0)$ the Jacobi fields which are zero at $\gamma(s_5)$ must have expansion which is positive at $\gamma(0)$ otherwise they would correspond to families of Jacobi fields with zero vorticity which have

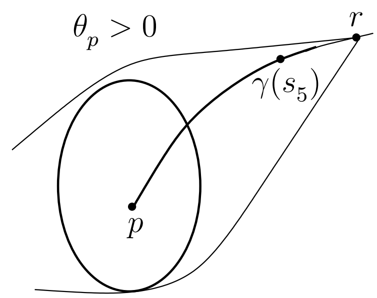


Figure 1.66: The Jacobi fields which are zero at r must have expansion θ which is positive at p otherwise r would lie in the bounded interval from $\gamma(s_1)$ to $\gamma(s_5)$.

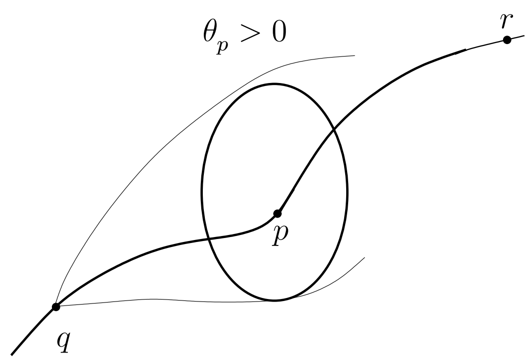


Figure 1.67: With $R_{abcd}V^bV^c \neq 0$. Non-positive expansion at p ($\theta > 0$) implies there is a point q conjugate to r, in the past of p. This is just the *time reversed* version of the focussing theorem.

non-posiitve expasion at $\gamma(0)$. It follows from the (time-reversed) focussing theorem that there must be a point say q before $\gamma(s_1)$ conjugate to r.

$\square$

Theorem for null geodesics from a point

In the appendix J on horizons we introduced adapted basis to describe the geometry of congruences of null geodesics.

The equation

$$\frac{d^2}{dv^2}Z^m = -R_{m4n4}Z^n \quad (m, n = 1, 2) \tag{1.175}$$

along a null geodesic $\gamma(v)$, a Jacobi field along $\gamma(v)$. The components Z^m are taken to be the components with respect to the basis $\mathbf{E_1}$ and $\mathbf{E_2}$, of a vector in the space S_q at each point q. The point p is conjugate to q along the null geodesic $\gamma(v)$ if there is a Jocobi field along $\gamma(v)$, not identically zero, which vanishes at q and p.

$$Z^m(v) = \hat{A}_{mn} \frac{d}{dv} Z^n \big|_q. \tag{1.176}$$

The null expansion scalar of S with respect to K is the scalar field $\hat{\theta}$ on S has a natural geometric interpretation. Let Σ be the intersection of S with hypersurface in $\mathcal{M}$ which is tranverse to K near $p \in S$; Σ will be a co-dimensional two spacelike submanifold of $\mathcal{M}$, along which K is orthogonal.

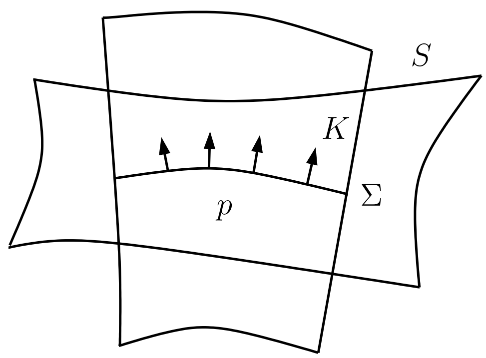


Figure 1.68: The null expansion scalar $\hat{\theta}$.

One has as before: $\hat{A}_{lm} \hat{\omega}_{lk} \hat{A}_{kn} = 0$, so the vorticity of the Jacobi fields which are zero at p vanishes so long as $\hat{A}_{mn}$ is invertible. Also p will be conjugate to q along $\gamma(v)$ if and only if

$$\hat{\theta} = (\det \hat{A})^{-1} \frac{d}{dv} (\det \hat{A}) \tag{1.177}$$

becomes infinite at p. We have the null version of the focussing theorem:

Proposition 1.1.54 *If $R_{ab} K^a K^b \geq 0$ and if at some point $\gamma(v_1)$ the expansion $\hat{\theta}$ has negative value $\hat{\theta}_1 < 0$, then there will be a point conjugate to q along $\gamma(v)$ between $\gamma(v_1)$ and $\gamma(v_1 + (2/-\hat{\theta}_1))$ provided that $\gamma(v)$ can be extended that far.*

The expansion $\hat{\theta}$ of the matrix $\hat{A}_{mn}$ obeys (??):

$$\frac{d}{dv} \hat{\theta} = -R_{ab} K^a K^b - 2\hat{\sigma}^2 - \frac{1}{2} \hat{\theta}^2, \tag{1.178}$$

and so the proof proceeds as before.

□

Proposition 1.1.55 *If $R_{ab}V^aV^b \geq 0$ and if at some point $p = \gamma(v_1)$ $K^cK^dK_{[a}R_{b]cd[e}V_{f]}$ is non-zero, there will be values v_0 and v_2 such that $q = \gamma(v_0)$ and $r = \gamma(v_2)$ will be conjugate along γ, providing that $\gamma(v)$ can be extended to these values.*

If $K_{[a}R_{b]cd[e}K_{f]}K^cK^d \neq 0$ is non-zero then so is R_{m4n4}. The proof is then similar to that of proposition **??**.

□

Theorem for null geodesics orthogonal to a spacelike two-surface

Proposition 1.1.56

□

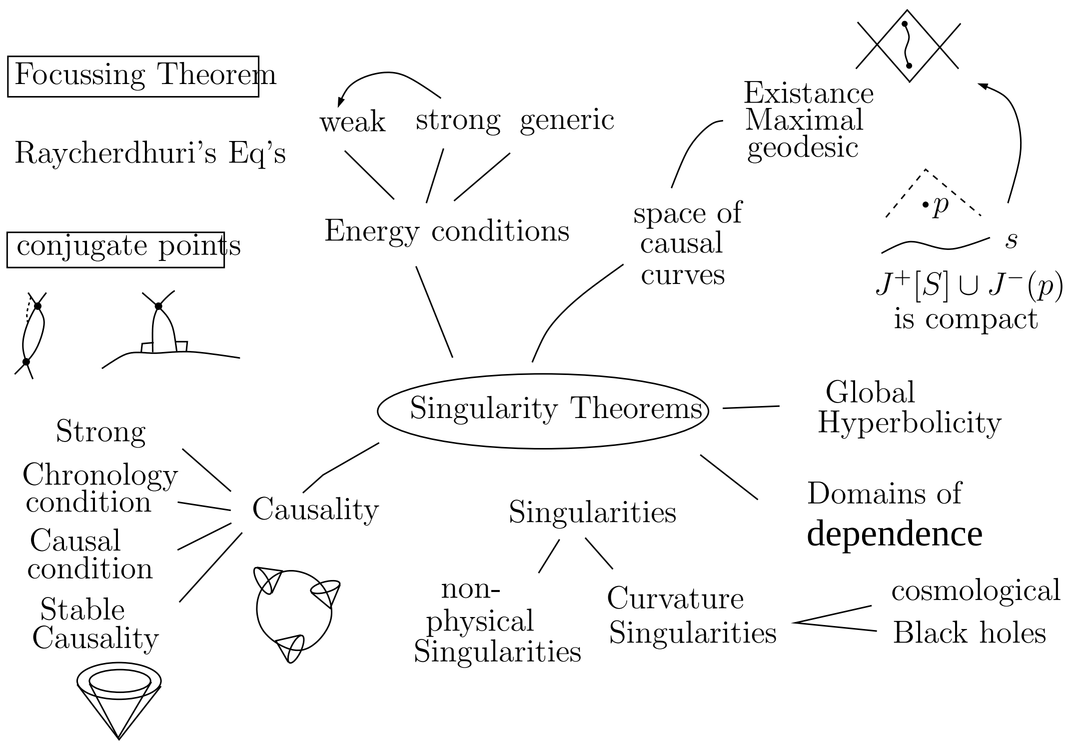


Figure 1.69: Singularity theorems.

1.2 Black Holes

It predicts that gravitational collapse , both at the Big Bang and inside black holes, brings about spacetime singularities as at which the theory breaks down.

Even if one perturbs around the Schwarzschild solution, specific values of expansions of null congruences may change, but for sufficiently small perturbations they will remain negative if they are negative for the unperturbed solution.

The Strong Energy Condition

$$T_{\hat{0}\hat{0}} + \sum_i T_{\hat{0}\hat{0}} \geq 0$$

is the most important energy condition (Gravity is attractive)

1.2.1 Collapse of a Star

In Newtonian gravity the force between particles goes as $1/r^2$. While the matter of the star is widely spaced the gravitational force between particles is weak. The gravitaional force increases as the particles come closer together. This increases the accelaration of the collapse and if nothing intervenes matter will collapse to a point. A star which has a high temperature has high pressure which balances the gravitational force and stops the collapse.

The gravity compresses the mater against the degeneracy pressure. If the neutron star is too large, the gravitational forces overwhelm the pressure gradients and collapse cannot be halted. The neutron star continues to shrink until it finally becomes a black hole.

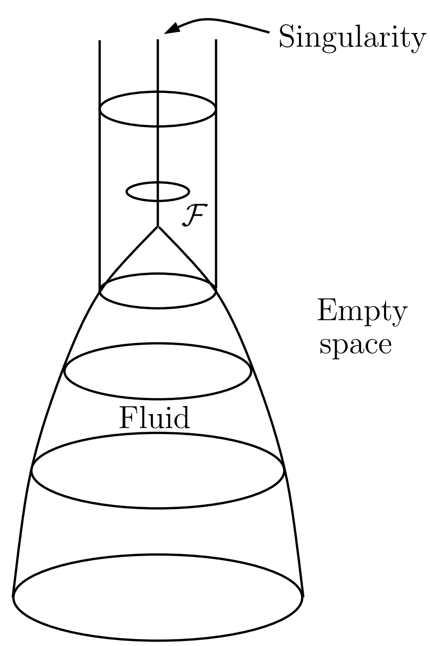


Figure 1.70: Diagram of collapse of a star.

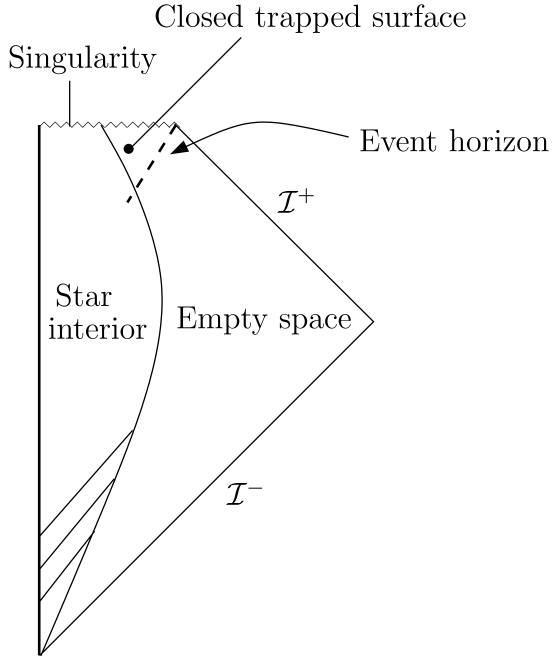


Figure 1.71: Penrose diagram of collapse of a star.

Using the Newtonian order of magnitude argument, we note that for a star with mass M and radius r_0, the gravitational unit volume is of the order $M/r_0^2 n m_n$, where $nm_n \simeq M/r_0^3$ is the mass density.

The gravitational force is balanced by a pressure gradient of order P/r_0, P being the average pressure in the star. Thus the pressure P can be expressed as:

$$P = \frac{M^2}{r_0^4} \approx M^{2/3} n^{4/3} m_N^{4/3} \tag{1.179}$$

When th density is sufficienlt low the main contribution to the pressure is from the degneracy of non-relativistic electrons, hence using (1.24) we have

$$P = \hbar^2 n^{5/3} m_e^{-1} \tag{1.180}$$

Equationg the two expressions for the pressure we obtain

$$M^{2/3} n^{4/3} m_N^{4/3} = \hbar^2 n^{5/3} m_e^{-1} \tag{1.181}$$

which gives the value of the number density n as

$$n = M^2 m_n^4 m_e^3 \hbar^{-6}. \tag{1.182}$$

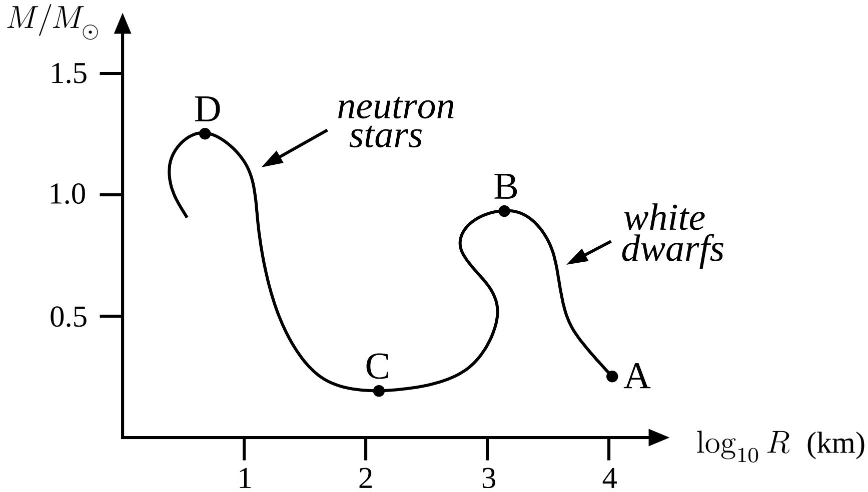


Figure 1.72: starCollaps.

The above value of n is based on the assumption that the self-gravity of the star is coming into play, this will be valid as long as this n is greater than the value of n given by() where self-gravity has no influence on the star as it is too small; and also this n must be less than $m_e^3\hbar^{-3}$ for the correctness of () In terms of pressures the relationship between small and large stasrs can be stated as

$$e^2 n^{4/3} < M^{2/3} n^{4/3} m_n^{4/3} \tag{1.183}$$

or equivalently as

$$e^3 m_n^{-2} < M. \tag{1.184}$$

On the other hand since

$$e^3 m_n^{-2} < M \tag{1.185}$$

other stuff

$$\frac{dP}{dr} = -\rho M(r) r^{-2} \tag{1.186}$$

where

$$M(r) = 4\pi \int_0^r \rho r^2 dr.$$

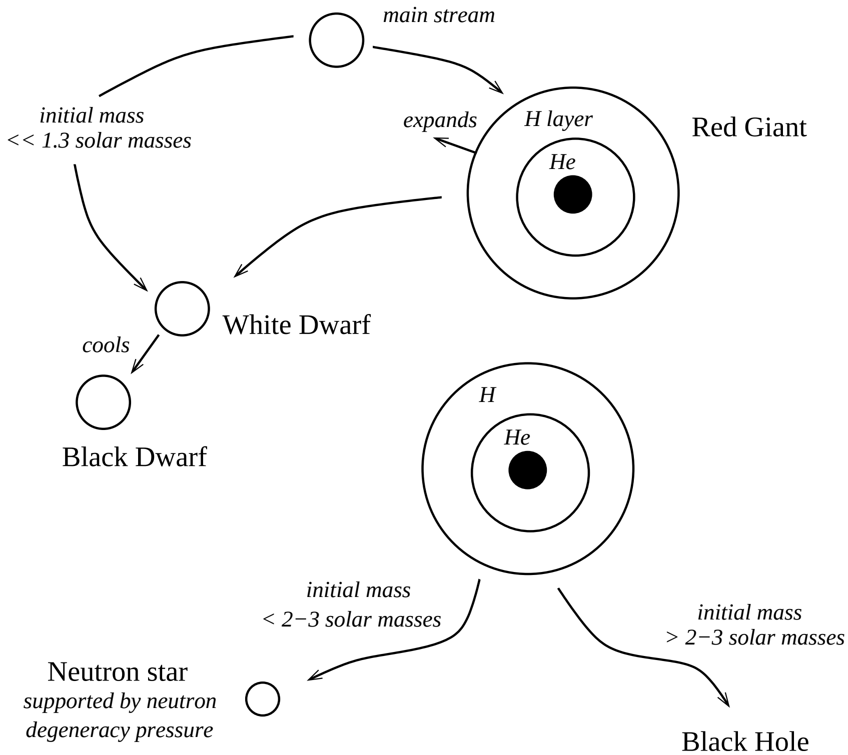


Figure 1.73: .

Multiply (1.186) by r^4 and integrate the LHS by parts from 0 to r, and since $p = 0$ at r_0 we obtain:

$$4\int_0^{r_0} pr^3 dr = \frac{(M(r_0))^2}{8\pi}. \tag{1.187}$$

On the other hand, since $\frac{dP}{dr}$ is never positive,

$$\frac{d}{dr}\left(\int_0^r pr'^3\right)^{3/4} < \frac{3\sqrt{2}}{4}p^{3/4}r^2. \tag{1.188}$$

Also p is never greater than $\hbar n^{3/4}$, hence,

$$\int_0^{r_0} pr'^3 dr' < \hbar\left(nr^2 dr\right)^{4/3} = \hbar(M(r_0))^{4/3}(4\pi m_n)^{-4/3}. \tag{1.189}$$

From (1.187), after simplification,

$$M(r_0) < (8\pi)^{3/2}(4\pi)^{-1/2}m_n^{-2} < 8\hbar^{3/2}\ m_n^{-2} \tag{1.190}$$

When the body is static, spherically symmetric and is composed of a perfect fluid, the Einstein field equations can be reduced to

$$\frac{dp}{dr} = -\frac{(\mu + p)(\hat{M}(r) + 4\pi r^3 p)}{r\big(r - 2\hat{M}(r)\big)} \tag{1.191}$$

where the radial coordinate is such that the area of the 2-surface (r =constant, t =constant.) is $4\pi r^2$. Similar to the Newtonian case the function $\hat{M}(r)$, represents the mass defined by the integral:

$$\hat{M}(r) = \int_0^r 4\pi r^2 \mu dr \tag{1.192}$$

where $\mu = \rho(1+\epsilon)$ is the total energy density, $\rho = nm_m$ (n times the mass of the nucleon) and ϵ is the relativistic increase of mass associated with the momentum of the fermions. ?? The function $\hat{M}(r_0)$ equals the Schwarzschild mass of the exterior Schwarzschild solution for $r > r_0$. For a bounded star $\hat{M}(r_0)$ will be less than the conserved mass:

$$\tilde{M} = \int_0^{r_0} \frac{4\pi\rho r^2 dr}{(1 - 2M/r)^{1/2}} \tag{1.193}$$

where N is the total number of nucleons in the star. The difference $\tilde{M} - \hat{M}$ represents the amount of energy (binding energy) radiated off to infinity since the formation of the star from dispered matter initially at rest.

1.2.2 Stability of Closed Trapped Surfaces

1.2.3 Apparent Horizons

The singularities that arise from localized gravitational collapse are associated with black holes. Intuitively, the Cosmic Censorship Hypothesis postulates that all singularities are hidden inside the event horizon, i.e. inside $\dot{I}^-(\mathcal{I}^+)$, the boundary of the past of future null infinity, $\mathcal{I}^+$, the latter is usually assume to be complete.

1.3 The Big Bang

Cosmic background radiation

1.4 Biblioliographical notes

In this chapter I have relied on the following refferences: [?], [?], [?]

The details of th proof of the existence of pair of conjugate points on every causal geodesic follows the proof given in “Causality, Conjugate Points and Singularity Theorems in Space-time” by TONG, Pun Wai.

1.5 Worked Exercies and Details

Details Hamiltonian.

$$
\begin{aligned}
R_{00} &= \frac{N''}{2a} - \frac{N'}{4a}\left(\frac{N'}{N} + \frac{a'}{a}\right) + \frac{N'}{ra'} && (1.194)\\
R_{11} &= -\frac{N''}{2N} - \frac{N'}{4N}\left(\frac{N'}{N} + \frac{a'}{a}\right) + \frac{a'}{ra'} && (1.195)\\
R_{22} &= 1 - \frac{r}{2a}\left(\frac{N'}{N} + \frac{a'}{a}\right) - \frac{1}{a} && (1.196)\\
R_{33} &= R_{22}\sin^2\theta. && (1.197)
\end{aligned}
$$

Thus the Ricci scalar

$$
R = g^{\mu\nu}R_{\mu\nu} = -6\frac{k + (a/N)(a\ddot{/}N)/c^2 + {(a\dot{/}N)}^2/c^2}{(a/N)^2} \tag{1.198}
$$

$$
\begin{aligned}
G_{00} &= 3\dot{R}/R^2c^2 + 3k/R^2 && (1.199)\\
G_{11} &= -\frac{k + 2R\ddot{R}/c^2 + \dot{R}^2/c^2}{1 - k\sigma^2} && (1.200)
\end{aligned}
$$

Details Hamiltonian.

(a) Find the action for isotropic spacetimes with scalar matter field $\phi(x)$.

(b) Check that the equations of motion of this action gives the Friedmann equation

$$
\left(\frac{\dot{a}}{a}\right)^2 = \frac{16\pi G}{3}\rho(a) \tag{1.201}
$$

$$\rho(a) = a^{-3}\mathcal{H}(a) = a^{-3}\left(\frac{1}{2}\frac{p_\phi^2}{a^3} + a^3 V(\phi)\right) \tag{1.202}$$

(a)

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-g}R + \int d^4x\sqrt{-g}\left[\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi)\right], \tag{1.203}$$

where R is the Ricci scalar curvature and V is a matter potential.

$$ds^2 = N^2(t)\,dt^2 - a^2(t)\,d\sigma^2 \tag{1.204}$$

$$d\sigma^2 = \frac{1}{1-Kr^2}\,dr^2 + r^2(\,d\theta^2 + \sin^2\theta + \,d\phi^2) \tag{1.205}$$

$$(g_{\mu\nu}) = \begin{pmatrix} N^2 & 0 & 0 & 0 \\ 0 & \frac{-a^2}{1-Kr^2} & 0 & 0 \\ 0 & 0 & -a^2r^2 & 0 \\ 0 & 0 & 0 & -a^2r^2\sin^2\theta \end{pmatrix}, \quad (g^{\mu\nu}) = \begin{pmatrix} \frac{1}{N^2} & 0 & 0 & 0 \\ 0 & \frac{-1+Kr^2}{a^2} & 0 & 0 \\ 0 & 0 & \frac{-1}{a^2r^2} & 0 \\ 0 & 0 & 0 & \frac{-1}{a^2r^2\sin^2\theta} \end{pmatrix}. \tag{1.206}$$

$$\begin{aligned} g_{00} &= g^{00} = 1 \quad g_{11} = \frac{1}{g^{11}} = \frac{-R^2}{1-k\sigma^2} \\ g_{22} &= \frac{1}{g^{22}} = -R^2\sigma^2 \quad g_{33} = \frac{1}{g^{33}} = -R^2\sigma^2\sin^2\theta. \end{aligned} \tag{1.207}$$

$$-\det(g_{\mu\nu}) = \frac{N^2a^6r^2\sin^2\theta}{1-Kr^2} \tag{1.208}$$

$$\begin{aligned} \int d^4x\sqrt{-g}\left[\frac{1}{2}(\partial_\mu\phi)^2 - V(\phi)\right] &= \int dt \int dr\,d\theta\sin\theta\,d\phi\left(\frac{r^2}{1-Kr^2}\right)^{1/2} Na^3\left[\frac{1}{2}\frac{\dot\phi^2}{N^2} - V(\phi)\right] \\ &= \mathcal{K}\int dtN(t)\left\{\frac{1}{2}a^3\frac{\dot\phi^2}{N^2(t)} - a^3V(\phi)\right\} \end{aligned} \tag{1.209}$$

where $\mathcal{K}$

$$\mathcal{K} = \int dr \frac{r}{(1-Kr^2)^{1/2}} \tag{1.210}$$

$\phi = \phi(r)$

$$\partial_\mu \partial^\mu \phi = \frac{1}{N^2}\dot{\phi}^2 + 0 \tag{1.211}$$

$$S = \int dt N(t) \left\{ \frac{3}{8\pi G} a \left(K - \frac{\dot{a}^2}{N^2(t)} \right) + \frac{1}{2} a^3 \frac{\dot{\phi}^2}{N^2(t)} - a^3 V(\phi) \right\}, \tag{1.212}$$

(b)

$$\frac{\partial \mathcal{L}}{\partial a} - \frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{a}} = 0 \tag{1.213}$$

$$\frac{\partial \mathcal{L}}{\partial a} = N(t) \left\{ \frac{3}{8\pi G} \left(K - \frac{\dot{a}^2}{N^2(t)} \right) + \frac{3}{2} a^2 \frac{\dot{\phi}^2}{N^2(t)} - 3a^2 V(\phi) \right\} \tag{1.214}$$

$$\frac{d}{dt}\frac{\partial \mathcal{L}}{\partial \dot{a}} = -\frac{6}{8\pi G}\frac{d}{dt}\left(\frac{\dot{\phi}}{N(t)} \right) = \frac{6}{8\pi G}\dot{N}(t)\frac{\dot{\phi}}{N^2(t)} - \frac{\ddot{\phi}}{N(t)} \tag{1.215}$$

$$N(t)\frac{3}{8\pi G}\left(K - \frac{\dot{a}^2}{N^2(t)} \right) = \frac{6}{8\pi G}\left(\frac{\ddot{\phi}}{N(t)} - \dot{N}(t)\frac{\dot{\phi}}{N^2(t)} \right) + N(t) 3a^2 V(\phi) - \frac{3}{2}a^2\frac{\dot{\phi}^2}{N(t)} \tag{1.216}$$

Singularity Theorems

Details Hamiltonian.

there would not be any null geodesic γ which is directed along a princpal null direction at six or more of its points.

Details Hamiltonian.

$\mathcal{M}$ may be topologically embedded in a finite dimensional Euclidean space, so $\mathcal{M}$ is Lindelof. If K is non-compact, there is an open cover with nonfinite subcover; by paracompactness, there is a locally finite open refinement of this; since the space is Lindelof, there is a countable subcover of this. Since simple regions form a basis, this final cover may be taken to consist of simple regions.

Not sure if I want this included?

Details Hamiltonian.

the matrix A is non-singular between p and q, $Y^\mu = \sum_\nu (A^{-1})^\mu{}_\nu X^\nu$.

$$2\omega_{\alpha\beta} = (A^{-1})_{\gamma\beta}\frac{d}{dt}A_{\alpha\gamma} - (A^{-1})_{\gamma\alpha}\frac{d}{dt}A_{\beta\gamma} \tag{1.217}$$

(1.99)

$$L = \int_a^b A_{\mu\nu}Y^\nu\left(\frac{d^2}{dt^2}(A_{\mu\gamma}Y^\gamma) + R_{\mu ab}{}^d T^a T_d A_{\mu\gamma}Y^\gamma\right)dt \tag{1.218}$$

(1.83)

$$L = \lim \epsilon \to 0 - \int_{a+\epsilon}^b A_{\alpha\beta}Y^\beta\left(2\frac{d}{dt}A_{\mu\gamma}\frac{d}{dt}Y^\gamma + A_{\mu\gamma}\frac{d^2}{dt^2}Y^\beta\right) \tag{1.219}$$

(We take the limit because the second derivative of Y^β may not be well defined at q.) Integrating the second term by parts

$$\begin{aligned}\int_a^b A_{\alpha\beta}Y^\beta A_{\mu\gamma}\frac{d^2}{dt^2}Y^\gamma &= -\int_a^b \frac{d}{dt}(A_{\alpha\beta}Y^\beta A_{\mu\gamma})\frac{d}{dt}Y^\gamma dt \\ &= -\int_a^b \Big[A_{\alpha\beta}Y^\beta\frac{d}{dt}A_{\mu\gamma}\frac{d}{dt}Y^\gamma + (A_{\alpha\beta}\frac{d}{dt}Y^\beta)(A_{\mu\gamma}\frac{d}{dt}Y^\gamma) \\ &\qquad + Y^\beta A_{\mu\gamma}\frac{d}{dt}A_{\alpha\beta}\frac{d}{dt}Y^\beta\Big]\,dt \end{aligned} \tag{1.220}$$

substituting this into (1.219)

$$L = -\Sigma\int_0^{s_p}\left\{(A_{\alpha\beta}\frac{d}{dt}Y^\beta)(A_{\alpha\delta}\frac{d}{dt}Y^\delta) + Y^\beta\left((\frac{d}{dt}A_{\alpha\beta})A_{\alpha\delta} - A_{\alpha\beta}\frac{d}{dt}A_{\alpha\delta}\right)\right\} \tag{1.221}$$

$$\left((\frac{d}{dt} A_{\alpha\beta}) A_{\alpha\delta} - A_{\alpha\beta} \frac{d}{dt} A_{\alpha\delta} \right) = -2 A_{\alpha\delta} \omega_{\alpha\gamma} A_{\alpha\delta} = 0. \tag{1.222}$$

$$L \leq 0. \tag{1.223}$$

empty

satisfy $E_1'(\phi_{\mathbf{1}}'') = 0$, $E_2'(\phi_{\mathbf{2}}'') = 0$ where E_1' is the Einstein operator with the coefficients $\mathbf{A_1'}, \mathbf{B_1'}$ and $\mathbf{C_1'}$ determined by $\phi_{\mathbf{1}}'$.

Since the coeffiecients

Appendix A

Space-time Diffeomorphism Invariance of General Relativity

Just as the action principle for electrodynamics should be invariant under gauge transformations, an action principle for general relativty should be invariant under its gauge transformations are infintesimal active diffeomorphisms, not to be confused with coordinate transformations.

It turns out, however, that actions that are invaraint under coordinate transformations are automatically invariant under active diffeomorphisms! This we prove in the following subsection.

A.0.1 Invariance of Integral Scalars Under Active Diffeomorphisms

With a coordinate transformation the volume element

$$d^4x = \frac{1}{4!}\epsilon_{\mu\nu\sigma\rho}dx^\mu dx^\nu dx^\sigma dx^\rho$$

transforms as

$$d^4x' = d^4xJ \tag{A.1}$$

where

$$J = \det\left(\frac{\partial x^{\mu'}}{\partial x^\alpha}\right).$$

Now

$$g'_{\mu\nu}(x') = \frac{\partial x^\alpha}{\partial x^{\mu'}} g_{\alpha\beta}(x) \frac{\partial x^\beta}{\partial x^{\nu'}}$$

We look on the right-hand side as the product of three matrices, and take the determinant of both sides giving

$$g' = J^{-2} g.$$

Thus

$$\sqrt{-g} = J\sqrt{-g'}. \tag{A.2}$$

Therefore

$$\sqrt{-g} d^4 x = \sqrt{-g'} d^4 x'$$

and so the quantity

$$\sqrt{-g} d^4 x$$

is invariant under coordinate transformations. Suppose F is a scalar field, $F = F'$. Then

$$\int F\sqrt{-g} d^4 x = \int F\sqrt{-g'} J d^4 x = \int F'\sqrt{-g'} d^4 x'$$

if the region of integration for x' corresponds to that for x. Therefore

$$\int F\sqrt{-g} d^4 x$$

is invariant under coordinate transformations. We refer to as an integral scalar.

Now let us investigate how an integral scalar transforms under an active diffeomorphism. We start with seeing how $\sqrt{-g} d^4 x$ transforms under active diffeomorphisms.

Consider the "small" volume depictured in fig (A.1). Under an active diffeomorphism the "corner" points of the infintesimal volume are changed but then evaluated at the original values of the coordinates, therefore

$$d^4x = d^4y.$$

and the quantity

$$d^4x$$

is invariant under active diffeomorphisms.

We see that in integrals over spacetime volume, the volume element d^4x does not change under an active diffeomorphism, while it does change under a coordinate transformation. However, the volume element $\sqrt{-g}\ d^4x$ is invariant under a coordinate transformation but not under an active diffeomorphism.

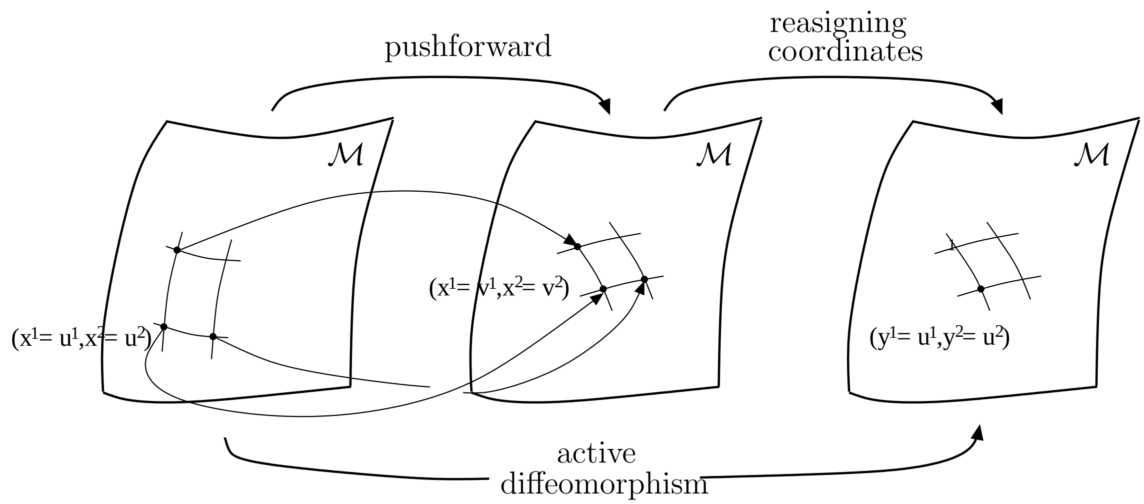


Figure A.1: We first do a pushforward mapping the "corners" of the infintesimal volume to new points. We then do a coordinate transformation to assign the new points the original coordinate values.

We will now show that any scalar integral is invariant under an active diffeomorphism that vanishes at the end points of integration. Let us find the effect of an active diffeomoprhism generated by the vector field ξ_ν on the metric $g_{\mu\nu}$.

$$y^\mu = x^\mu + \epsilon\xi^\mu(x)$$

Differentiating, we get

$$\frac{\partial y^\mu}{\partial x^\nu} = \delta^\mu_\nu + \epsilon\partial_\nu\xi^\mu. \tag{A.3}$$

$$
\begin{aligned}
g^{\mu\nu}(y) &= \frac{\partial y^\mu}{\partial x^\rho}\frac{\partial y^\nu}{\partial x^\sigma} g^{\rho\sigma}(x) \\
&= (\delta^\mu_\rho + \epsilon\partial_\rho \xi^\mu)(\delta^\nu_\sigma + \epsilon\partial_\sigma \xi^\nu) g^{\rho\sigma}(x) \\
&= g^{\mu\nu}(x) + \epsilon g^{\rho\nu}(x)\partial_\rho \xi^\mu + \epsilon g^{\mu\rho}(x)\partial_\rho \xi^\nu
\end{aligned}
\tag{A.4}
$$

$$
g^{\mu\nu}(y) = g^{\mu\nu}(x^\rho + \epsilon\xi^\rho) = g^{\mu\nu}(x) + \epsilon\xi^\rho\partial_\rho g^{\mu\nu}(x) \tag{A.5}
$$

$$
\frac{\delta g^{\mu\nu}}{\epsilon} = \frac{g^{\mu\nu}(y) - g^{'\mu\nu}(y)}{\epsilon} \tag{A.6}
$$

$$
\begin{aligned}
\delta g^{\mu\nu}/\epsilon &= \xi^\rho \partial_\rho g^{\mu\nu} - g^{\mu\rho}\partial_\rho \xi^\nu - g^{\rho\nu}\partial_\rho \xi^\mu \\
&= \xi^\rho \partial_\rho g^{\mu\nu} - g^{\mu\rho}(\nabla_\rho \xi^\nu - \Gamma^\nu_{\sigma\rho}\xi^\sigma) - g^{\rho\nu}(\nabla_\rho \xi^\mu - \Gamma^\mu_{\sigma\rho}\xi^\sigma) \\
&= \xi^\rho(\partial_\rho g^{\mu\nu} + g^{\mu\sigma}\Gamma^\nu_{\rho\sigma} + g^{\sigma\nu}\Gamma^\mu_{\rho\sigma}) - g^{\mu\rho}(\nabla_\rho \xi^\nu) - g^{\rho\nu}(\nabla_\rho \xi^\mu) \\
&= \xi^\rho \nabla_\rho g^{\mu\nu} - \nabla^\mu \xi^\nu - \nabla^\nu \xi^\mu \\
&= -\nabla^\mu \xi^\nu - \nabla^\nu \xi^\mu
\end{aligned}
\tag{A.7}
$$

We have under a diffeomorphism

$$
\delta g^{\mu\nu} = -\epsilon\nabla^\mu \xi^\nu - \epsilon\nabla^\nu \xi^\mu .
$$

Then from

$$
\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}
$$

we have

$$
\delta\sqrt{-g} = \epsilon(\nabla_\alpha \xi^\alpha)\sqrt{-g}. \tag{A.8}
$$

Putting this together and as the Lie derivative of a scalar is the directional derivative

$$
\begin{aligned}
\delta S_{\epsilon\xi} &= \int \mathcal{L}_{\epsilon\xi}(F\sqrt{-g})d^4x \\
&= \epsilon\int (\xi^\mu \nabla_\mu F + F\nabla_\mu \xi^\mu)\sqrt{-g}d^4x \\
&= \epsilon\int \nabla_\mu (F\xi^\mu)\sqrt{-g}d^4x \\
&= \epsilon\int F\xi^\mu d^3\Sigma_\mu
\end{aligned}
\tag{A.9}
$$

where we have used Gauuss' law. For variations with $\xi^\mu = 0$ at the bounadries, $\delta S_{\epsilon\xi} = 0$.

What is the physical implication of this? Note if the metric was a nondynamical object it could not be varied without changing the original action principle. Since we are assuming the metric is not a nondynamical object that would otherwise break the background independence of the theory - diffeomorphism invariance is formally equivalent to general covariance, namely the invariance of the field equations under arbitrary changes of the spacetime coordinates $\vec{x}$ and t. Therefore if the action principle is based on an integral scalar **it will have active diffeomorphisms as a gauge transformation**.

Bibliography

[1] S. W, Hawking and G. F. R, Ellis, “The Large Scale Structure of Space-Time” (Cambridge Monographs on Mathematical Physics).

[2] R. Penrose, (1972) “Techniques of Differential Topology in Relativity”, (the Society for Industrial and Applied Mathematics).

[3] D. Lerner, (1972), “Methods of Local and Global Differential Geometry in General Relativity”. Chapter 1, “Techniques of Topology and Differential Geometry in General Relativity”, (Proceedings of the Regional Conference on Relativity held at the University of Pittsburgh, Pittsburhg, Pennsylvania, July 13-17, 1970).

[4] Robert. M. Wald, (1984), “General Relativity”, (The University of Chicago Press).

[5] G. L. Naber, (1988), “Spacetime and Singularities An Introduction”, (Cambridge University Press).

[6] “Symmetry Transformations, the Einstein-Hilbert Action, and Gauge Invariance”, Edmund Bertschinger, (web.mit.edu/edbert/GR/gr5.pdf).

[7] Sobolev, S. L. (1963), “Applications of Functional Analysis in Mathematical Physics”, Vol 7, Translations of Mathematical Monographs (Am. Math. Soc., Providence).