EXPLORING SPACETIME AND SINGULARITIES

EXPLORING SPACETIME AND SINGULARITIES

By

AMI MAMOLO, B.Sc.

A Project

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2005)	McMaster University
(Mathematics)	Hamilton, Ontario

TITLE:	Exploring Spacetime and Singularities
AUTHOR:	Ami Mamolo, B. Sc. (McMaster University)
SUPERVISOR:	Professor M. Wang
NUMBER OF PAGES:	v, 64

Abstract

Hawking's Singularity Theorem establishes the existence of a cosmological singularity in a spacetime for which no global assumptions about causality are made. This theory has been useful for predicting the occurrence of singularities in a spacetime without solving Einstein's field equation. This paper is an exposition of the tools and some of the theory required to prove and apply Hawking's theorem. Emphasis is placed on practical methods for applying this result to the flat, dust-filled Robertson-Walker spacetime, and the black hole interior of the Kruskal extension of the Schwarzschild spacetime.

Acknowledgments

My sincerest thanks extends to Dr. McKenzie Wang. His guidance and patience throughout the completion of my project were a tremendous help and inspiration. I would also like to extend my gratitude to Drs. Hans Boden and Andrew Nicas for their invaluable insight and advice.

I owe a heartfelt thanks to my family for their unwavering love and support, without which this project would not have been written. I am especially grateful to Colin, who now knows more about geometry than he ever wanted to.

Contents

Chapter 0

Introduction

An *event* in space and time can be thought of as an occurrence at a particular instant in time and position in space. A 4-dimensional continuum of events is called a spacetime. Formally, a *spacetime* M is a connected 4-dimensional time-oriented manifold that is furnished with a non-degenerate metric tensor $\mathbf{g}$ of Lorentz signature. The unfamiliar concepts in this definition will be made precise in the following chapter.

In the early twentieth century, it was generally accepted that the laws of physics are independent of the choice of coordinates used to describe said laws. This *principle of general covariance* had a significant impact on the development of the Theory of General Relativity, wherein Einstein proposed that all physical laws can be described completely by a Lorentz metric and its covariant derivatives. Via Einstein's field equations, he postulated that the structure of a spacetime relates to its matter distribution. Provided solutions to Einstein's equation can be found and understood, mathematicians and physicists can link the curvature of $(M, \mathbf{g})$ to its stress-energy-momentum tensor. One of the first physicists to do so was Schwarzschild in 1916 when he derived a solution of the Ricci flat Einstein equations for a static, spherically symmetric body. Schwarzschild's solution and the Robertson-Walker solutions for a homogeneous and isotropic spacetime are the main examples of this exposition.

Example 0.0.1 The Robertson-Walker cosmological model is the spacetime with the metric

$$ds^2 = -dt^2 + a^2(t) \begin{cases} dr^2 + \sin^2 r \ (d\theta^2 + \sin^2\theta \ d\phi^2) \\ dr^2 + r^2(d\theta^2 + \sin^2\theta \ d\phi^2) \\ dr^2 + \sinh^2 r \ (d\theta^2 + \sin^2\theta \ d\phi^2). \end{cases} \tag{1}$$

The solutions for $a(t)$ (called the *Friedmann solutions* and listed in Table 1) correspond to the different spatial geometries of constant sectional curvature as well as the different types of matter that were commonly considered the dominant contributors to the stress-energy tensor. Up until recently, most of the energy density of the present universe was believed to be concentrated in galaxies. On the immense scales we are concerned with, each galaxy can be thought of as a "grain of dust." The detection of a cosmic microwave background suggests there are other forms of energy, such as dark energy. Analysis of the cosmic microwave background provides a glimpse of the universe as it was 300 000 years after the *big bang*, while research on a cosmic neutrino background may provide information about the universe just moments after its beginning.

The Robertson-Walker model seems to be a good approximation for the geometry of our universe on a large scale. The distribution of galaxies appears to be homogeneous and isotropic: the same in all directions when viewed from our galaxy. The three choices of (1) represent the possibilities that the universe is either "open" or "closed." Flat and hyperbolic spaces correspond to a universe that is expanding forever (open), whereas the spherical case corresponds to a universe that will eventually contract (closed).

Table 1
Friedmann solutions for Robertson-Walker
dust filled and radiation filled cosmological models

	Types of Matter	
Spatial geometry	Dust: models present universe	Radiation: models early universe
Flat	$a = \eta^2$ $t = \frac{1}{3}\,\eta^3$	$a = (4C')^{1/4}\, t^{1/2}$
Hyperbolic	$a = \frac{1}{2}\, C(\cosh\eta - 1)$ $t = \frac{1}{2}\, C(\sinh\eta - \eta)$	$a = \sqrt{C'}\, [(1 + \frac{t}{\sqrt{C'}})^2 - 1]^{1/2}$
Spherical	$a = \frac{1}{2}\, C(1 - \cos\eta)$ $t = \frac{1}{2}\, C(\eta - \sin\eta)$	$a = \sqrt{C'}\, [1 - (1 - \frac{t}{\sqrt{C'}})^2]^{1/2}$

$$C = \frac{8\pi\rho a^3}{3}, \quad C' = \frac{8\pi\rho a^4}{3}.$$

Example 0.0.2 The Schwarzschild metric

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2) \qquad (2)$$

describes the exterior ($r > 2m$) gravitational field of a static, spherically symmetric body with mass m in a spacetime for which the only source of gravitation is the star itself (hence the spacetime is Ricci flat). Relativistically significant, this model has accurately predicted the motion of planets, the "bending of light," and "time delay" effects observed within our solar system. When considering stars with large enough mass (at least twice that of our Sun), (2) depicts the exterior region of a black hole, that is, a star which has undergone complete gravitational collapse. One of the issues when interpreting (2) involves distinguishing between actual physical singularities and coordinate singularities. The coordinate singularity appearing at $r = 2m$ is known as the *Schwarzschild horizon.* A change of coordinates (cf ex. 1.5.3) will smoothly extend Schwarzschild's solution past the horizon and allow us to study the interior ($r < 2m$) region. The extension is known as the *Kruskal spacetime* and is integral to our appreciation of the geometry of a Schwarzschild black hole.

Both of these exact solutions to Einstein's equation produce models of spacetimes which are *singular*, that is, spacetimes for which there is a point where our current physical laws fail. The *big bang* of the Robertson-Walker solution and the black hole singularity of the Schwarzschild solution are popular topics of discussion, though the general notion of a singularity still eludes a universally accepted formulation.

Until the 1960s, the solutions to Einstein's equation that implied gravitational collapse assumed spherical symmetry. This assumption simplified calculations but left physicists wondering if the singularity found was merely a consequence of symmetry. Penrose was among those pondering whether the presence of perturbations that destroy spherical symmetry might alter the singular state of the spacetime. In his 1965 article "Gravitational collapse and space-time singularities," Penrose published the first result which showed the existence of a black hole singularity without any assumption of symmetry (see Theorem 5.2.3). This result led Hawking to establish similar theorems. In his 1967 article "The occurrence of singularities in cosmology. III. Causality and singularities," he addressed the following questions:

> "First, without any global assumption being made about causality, could it be proved that there are solutions which evolve from a non-singular state to an inevitable singularity and which are fully general in the sense that a small perturbation of the initial state could not prevent the occurrence of a singularity? Secondly, would there be a singularity in any solution which could

> represent the observed universe? Thirdly, what would be the nature of the singularity?" [H]

In Theorem 1 of [H] (cf Theorem 5.1.3), Hawking showed that a singularity is unavoidable in a spacetime containing a compact Riemannian surface whose normal geodesics are all converging.

This project is intended to be a comprehensive guide to understanding spacetimes and singularities for students who have taken introductory courses in manifold theory and Riemannian geometry. The main focus is to give an exposition of the tools and some of the theorems required to prove and apply Hawking's Singularity Theorem. In Chapters 1 and 2 the reader is introduced to some of the relevant aspects of Lorentzian geometry. Emphasis is placed on the particular properties that differ from Riemannian geometry. Chapters 3 and 4 delve into the causality relationships of a spacetime, i.e. the conditions under which an event may influence other events. The approach in Chapter 4 is more geometrical than that of Chapter 3. It sets up much of the theory required to prove the singularity theorems that are presented in Chapter 5. The two versions of Hawking's singularity theorems proved in this exposition (Theorems 5.1.2 and 5.1.4) appear as they are stated in [O]. The original version in [H] has a more relativistic approach, and the physics-inclined student is encouraged to read Chapters 6 and 8 of [HE].

The proofs of Hawking's theorems show that a spacetime M is geodesically incomplete provided some curvature conditions hold on M and that there is a timelike geodesic *(i)* which is orthogonal to a Riemannian hypersurface $S \subset M$, and *(ii)* for which there is a conjugate point of this surface. A *timelike* geodesic is one that travels along the "time dimension" of M. It maximizes the length (or "proper time") of curves joining two events (Proposition 1.2.2). In order to maximize the length of a timelike curve joining S to an event in the future of S, we must find a sequence of curves whose lengths approach this supremum and then take their limit (Theorem 4.4.1). Along this maximum geodesic, and in part because of the curvature conditions on M, there is a variation of timelike geodesics normal to S which focus to a point (Theorem 2.1.2). This conjugate point is where the known physical laws fail; it is the singularity of M. Conditions on the physical structure of M guarantee that no timelike curves reach an event in the future of the singularity. This is easily done if every event in M is predictable by S. Otherwise, further physical properties of M are needed to show the required result (Proposition 3.2.3).

Chapter 1

Geometric Preliminaries

The aim of this chapter is to discuss and present those concepts and computations in Lorentz geometry relevant to Hawking's singularity theorems. The emphasis will be on the geometrical properties of a Lorentz manifold that differ from those of a Riemannian one. Both Lorentz and Riemannian manifolds are special cases of *semi-Riemannian manifolds*, so many of the conventions of Riemannian geometry will hold for Lorentz manifolds as well. Unless otherwise indicated, the reader may assume that this is the case, though a change in sign may be necessary on occasion.

One key difference that arises in the study of Lorentz geometry is in the notions of distance and length for timelike curves. To appreciate these differences we begin by introducing Lorentz manifolds and discussing the causal character of lengths and curves.

1.1 Lorentz Metric and Causal Character

A smooth manifold M furnished with a nondegenerate metric tensor $\mathbf{g}$ of index $\nu \geq 0$ is called a *semi-Riemannian* (or pseudo-Riemannian) manifold. The *index* ν of $\mathbf{g}$ is the maximal dimension of a subspace in T_pM on which $\mathbf{g}$ is negative definite. For such a semi-Riemannian manifold M and point $p \in M$ there is an orthonormal basis $v_1, \ldots, v_n$ of T_pM such that $\mathbf{g}(\mathbf{v_j}, \mathbf{v_j}) = -1$ for $1 \leq j \leq \nu$, $\mathbf{g}(\mathbf{v_i}, \mathbf{v_i}) = 1$ for $\nu + 1 \leq i \leq n$, and $\mathbf{g}(\mathbf{v_i}, \mathbf{v_j}) = 0$ for $i \neq j$. If $x^1, \ldots, x^n$ is a local coordinate system on M, then the local components of $\mathbf{g}$ are $\mathbf{g_{ij}} = \mathbf{g}(\partial_i, \partial_j)$ $(\mathbf{1} \leq \mathbf{i}, \mathbf{j} \leq \mathbf{n})$, which is sometimes denoted as $\langle \partial_i, \partial_j \rangle$. For arbitrary vector fields $V = \sum V^i \partial_i$ and $W = \sum W^j \partial_j$,

$$\mathbf{g}(\mathbf{V}, \mathbf{W}) = \langle \mathbf{V}, \mathbf{W} \rangle = \sum \mathbf{g_{ij}} \mathbf{V^i} \mathbf{W^j}.$$

A *Lorentz manifold* M is a semi-Riemannian manifold of dimension ≥ 2 and index $\nu = 1$. The signature of a Lorentz metric is $-,+,\ldots,+$ whereas the signature of a Riemannian metric is $+,\ldots,+$. Spacetimes are defined as 4-dimensional Lorentz manifolds, so we will primarily be concerned with the case $\nu = 1$, though the theory presented for a Lorentz manifold can easily be generalized.

The analogue to Euclidean space for a Lorentz manifold is Minkowski space $\mathbb{R}^n_1$ with metric tensor $ds^2 = -dt^2 + (dx^1)^2 + \ldots + (dx^{n-1})^2$. Minkowski space with $n = 4$ is the relativistic model for a spacetime without any gravitational influence.

Similar to the Riemannian case, the metric tensor of M makes each tangent space T_pM isometric to $\mathbb{R}^n_1$. The isometry group of T_pM is the semi-orthogonal group called the *Lorentz group*, $O(1, n-1) = \{B \in Gl(n, \mathbb{R}) : B^t\ \eta\ B = \eta,\ \eta = \text{diag}(-1, 1, \ldots, 1)\}$. This is analogous to the Riemannian case where the orthogonal group $O(n) = \{A \in Gl(n, \mathbb{R}) : A^tA = I\}$. Notice that the number of negative signs in the diagonal matrix η corresponds to the value of the index ν. The Lorentz group can be used to describe the orientability of a Lorentz manifold, as we shall see in section 1.3.

Definition 1.1.1 *The* causal character *of $v \in T_pM$ is:*

spacelike if $\langle v, v \rangle > 0$ *or* $v = 0$,

null if $\langle v, v \rangle = 0$ *and* $v \neq 0$,

timelike if $\langle v, v \rangle < 0$.

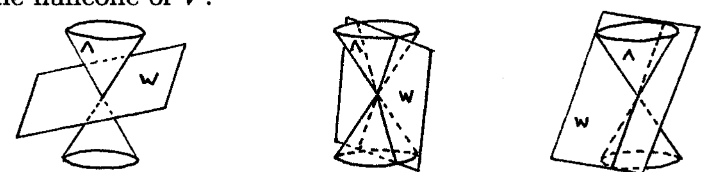
The norm *(or speed)* $|v|$ *of* v *is defined as* $|\langle v, v \rangle|^{1/2}$.

The set of null vectors in T_pM is called the *nullcone* Λ_p (or lightcone) at $p \in M$. Timelike vectors lie in the interior of Λ_p, whereas spacelike vectors lie outside.

It is possible to study T_pM abstractly as a scalar product space of index 1 and dimension ≥ 2. In such a case, T_pM is called a *Lorentz vector space.* The causality of a subspace W of a Lorentz vector space V is defined as follows:

1. W is *spacelike* if $\mathbf{g}|_\mathbf{W}$ is positive definite.
2. W is *timelike* if $\mathbf{g}|_\mathbf{W}$ is nondegenerate of index 1.
3. W is *lightlike* if $\mathbf{g}|_\mathbf{W}$ is degenerate.

Furthermore, W is timelike only if $W^{\perp} = \{v \in V : \langle u, v \rangle = 0 \ \forall \ u \in W\}$ is spacelike, and vise versa. W is lightlike (that is, degenerate) if and only if $W^{\perp}$ is lightlike. The following diagram helps illustrate these points. Define Λ as the nullcone of V.



(i) W spacelike (ii) W timelike (iii) W lightlike

Proposition 1.1.2 *Let $v, w \in T_pM$ be timelike vectors such that $\langle v, w \rangle < 0$.*

(i) The reverse Schwartz inequality holds:

$$| \langle v, w \rangle | \geq |v||w| \text{ with equality iff } v \text{ and } w \text{ are collinear.}$$

(ii) The reverse triangle inequality holds:

$$|v| + |w| \leq |v + w|, \text{ with equality iff } v \text{ and } w \text{ are collinear.}$$

Proof. (i) Since $\langle v, w \rangle < 0$ there is an $a > 0$ such that $w = av + u$, where u is orthogonal to v (and hence spacelike). w is timelike so

$$\langle w, w \rangle = \langle av + u, av + u \rangle = a^2 \langle v, v \rangle + \langle u, u \rangle < 0.$$

Then,

$$\langle v, w \rangle^2 = \langle v, av + u \rangle^2 = a^2 \langle v, v \rangle^2 = (\langle w, w \rangle - \langle u, u \rangle) \langle v, v \rangle \geq \langle w, w \rangle \langle v, v \rangle .$$

If equality holds, then $u = 0$ and $w = av$.

$$\begin{aligned} \text{(ii) } |v + w|^2 &= | \langle v + w, v + w \rangle | = - \langle v + w, v + w \rangle \\ &= - \langle v, v \rangle - 2 \langle v, w \rangle - \langle w, w \rangle \\ &\geq - \langle v, v \rangle + 2|v||w| - \langle w, w \rangle \text{ by part (i)} \\ &= |v|^2 + 2|v||w| + |w|^2 \\ &= (|v| + |w|)^2. \end{aligned}$$

Equality will hold iff $|v||w| = - \langle v, w \rangle$ and part (i) shows collinearity. □

Smooth curves can also have a causal character. A curve $\alpha : I \longrightarrow M$ is called spacelike if all of its velocity vectors $\alpha'(s)$ are spacelike; likewise for null and timelike. Intuitively, a timelike curve describes the motion of an object through time; its movement into the future or into the past. An arbitrary curve may not necessarily have a causal character. However, a geodesic $\gamma : I \longrightarrow M$ always does since γ' is parallel and hence the causal character of γ' remains unchanged along γ.

1.2 Reparametrizations and Timelike Geodesics

In Riemannian geometry a piecewise smooth curve segment $\alpha : I \longrightarrow M$ with $\alpha'(s) \neq 0$ for $s \in I$ can be reparametrized by arc length to a unit speed piecewise smooth curve segment. In Lorentz geometry the unit speed parameter depends on the causality of the curve.

Definition 1.2.1 *Let $\alpha : [a,b] \longrightarrow M$ be a piecewise smooth curve segment in a semi-Riemannian manifold M. The arc length of α is*

$$L(\alpha) = \int_a^b |\alpha'(s)| \, ds.$$

For a spacelike curve, this definition of arc length is the same as Riemannian arc length. However for a timelike curve, the sign is different and we use the term "proper time" synonymously with length,

$$\tau = \int_a^b \Big(- \langle \alpha'(s), \alpha'(s) \rangle \Big)^{1/2} ds.$$

Proper time is an important concept for our discussion of singularities and will be dealt with in more detail in Chapter 4 where we will reparametrize timelike geodesics by τ.

As in the Riemannian case, a *geodesic* $\gamma : I \longrightarrow M$ in a Lorentz manifold is a curve whose acceleration is zero: $\gamma'' = 0$. The standard existence theorems for ordinary differential equations hold, as do the properties of the exponential map $\exp_p : T_pM \longrightarrow M$, which takes lines through the origin of T_pM to geodesics of M through p. Consequently, $\exp_p$ maps a small enough star-shaped neighbourhood of $0 \in T_pM$ diffeomorphically onto a *normal neighbourhood* $\mathcal{U}$ of p in M. Unlike the Riemannian case however, timelike geodesics do not have the property of being locally length minimizing.

Proposition 1.2.2 *Let M be a Lorentz manifold and $\mathcal{U}$ a normal neighbourhood of $p \in M$. If there exists a timelike curve in $\mathcal{U}$ from p to q, then the longest such curve is the unique (up to reparametrization) timelike geodesic in $\mathcal{U}$ from p to q.*

Sketch of proof. This is a straightforward analogue of the Riemannian case. One uses the Gauss Lemma [L] (p. 102) to decompose σ' as

$$\sigma' = - \langle \sigma', U \rangle \, U + N,$$

where U is a timelike unit vector field along σ and N is spacelike. Then,

$$|\sigma'|^2 = (-\langle\sigma',\sigma'\rangle) = |\langle\sigma',U\rangle U|^2 - |N|^2 \leq |\langle\sigma',U\rangle|^2.$$

However, $|\langle\sigma',U\rangle| = \frac{d(r\circ\sigma)}{ds}$, where r is the radius function of M defined by $r(q) = |\exp_p^{-1}(q)|$. The length of σ is

$$\begin{aligned} L(\sigma) &= \int_0^1 |\sigma'|\, ds \leq \int_0^1 |\langle\sigma',U\rangle|\, ds \\ &= \int_0^1 \frac{d(r\circ\sigma)}{ds}\, ds \\ &= r(q). \end{aligned}$$

Equality will hold if and only if $|\sigma'|$ is constant.

A curve $\alpha : I \longrightarrow M$ is called *inextendible* (or maximal) if I is the largest possible domain for α; if $\alpha : J \longrightarrow M$ with $J \supset I$ then $J = I$. An inextendible timelike geodesic with finite length is called a *timelike incomplete* geodesic. Timelike geodesic incompleteness means that particles could move freely along a timelike geodesic whose future or past histories do not exist outside of a finite interval of proper time.

A null curve has zero length, so a null geodesic is called incomplete if it cannot be extended to arbitrary values of its affine parameter.

Timelike and null (geodesic) incompleteness are frequently used to characterize a singular spacetime. On a manifold with Lorentz metric any two points that are connected by a timelike curve can be connected by one of arbitrarily small proper time, so there is no generalization of Riemannian distance. Note moreover, that unlike a Riemannian metric, a Lorentz metric does not define a topological metric. Consequently, the Hopf-Rinow theorem does not hold and we are left with only geodesic incompleteness to describe a singular spacetime.

1.3 Time-Orientability

Time orientability allows a continuous choice of the future nullcone in T_pM to each point as the point varies over M. It is the Lorentz counterpart to the orientability associated to a Riemannian manifold.

An orientable Riemannian manifold N can be defined by considering how the orthogonal group $O(n)$ acts on orthonormal bases of T_qN, ($q \in N$). Let $e_1,\ldots,e_n$ and $f_1,\ldots,f_n$ be orthonormal bases of T_qN. Then $f_j =$

$\sum b_{ij} e_i$ $(1 \leq j \leq n)$ defines a matrix $(b_{ij}) \in O(n)$. If (b_{ij}) is an element of the special orthogonal group $SO(n) \subset O(n)$, then the two bases have the same orientation. The index of $SO(n)$ $[O(n), SO(n)]$ is equal to 2, so this equivalence relation describes two equivalence classes of *orientations* of $T_q N$. An *orientation* of N is a function λ that smoothly assigns an orientation of $T_q N$ to each $q \in N$. What is meant by "smooth" is that for each $q \in N$ there is a coordinate system whose induced local orientation agrees with λ on some neighbourhood of q. N is *orientable* provided it admits an orientation.

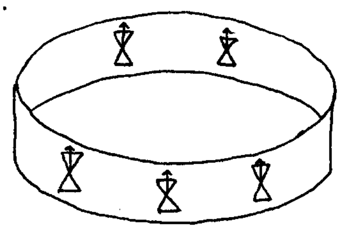
The orientation of a Lorentz manifold M is described in much the same way. However the orthogonal group $O(1, n-1) = \{B \in Gl(n, \mathbb{R}) : B^t\, \eta\, B = \eta,\ \eta = \text{diag}(-1, 1, \ldots, 1)\}$ now has four connected components (unlike $O(n)$ which has 2). Let $(h_{ij}) \in O(1, n-1)$ be the Lorentz counterpart of (b_{ij}) as defined above. The components of the Lorentz group are designated by:

$$O^{++}(1, n-1) = \{(h_{ij}) \in O(1, n-1) : h_{11} > 0, \text{ and } \det(h_{ij}) > 0\ i, j \neq 1\}$$
$$O^{+-}(1, n-1) = \{(h_{ij}) \in O(1, n-1) : h_{11} > 0, \text{ and } \det(h_{ij}) < 0\ i, j \neq 1\}$$
$$O^{-+}(1, n-1) = \{(h_{ij}) \in O(1, n-1) : h_{11} < 0, \text{ and } \det(h_{ij}) > 0\ i, j \neq 1\}$$
$$O^{--}(1, n-1) = \{(h_{ij}) \in O(1, n-1) : h_{11} < 0, \text{ and } \det(h_{ij}) < 0\ i, j \neq 1\}$$

Corollary 1.3.1 *(see [O] 9.7(3)) The sets $O^{++} \cup O^{--}$, $O^{++} \cup O^{+-}$, and $O^{++} \cup O^{-+}$ are subgroups of $O(1, n-1)$.*

Each of these subgroups has index 2 (in the group theoretic sense), and consequently each describes two equivalence classes, which allow three possible ways to orient M. If $(h_{ij}) \in SO(1, n-1) = O^{++} \cup O^{--}$ then $T_p M$ and M are orientable in the Riemannian sense. If $(h_{ij}) \in O^{++} \cup O^{+-}$ then $h_{11} > 0$, and $T_p M$ admits a *time-orientation.* Similarly, if $(h_{ij}) \in O^{++} \cup O^{-+}$ then $\det(h_{ij}) > 0$ $(i, j \neq 1)$ and $T_p M$ admits a *space-orientation.*

Intuitively, a manifold for which future and past directions can be consistently chosen is called time-orientable, and a particular choice of orientation will *time-orient* M.



M is time-orientable if a future and past nullcone can be consistently chosen for each $p \in M$.

Lemma 1.3.2 *Let M be time-orientable. There exists a smooth timelike vector field $X \in \mathfrak{X}(M)$.*

Sketch of proof. Let $\widetilde{\tau}$ be a time-orientation of M. $\widetilde{\tau}$ is smooth, so for each $p \in M$ there is a smooth vector field $X_{\mathcal{U}}$ on a neighbourhood $\mathcal{U}$ of p such that $X_{\mathcal{U}} \in \widetilde{\tau}_q$ for each $q \in \mathcal{U}$. Let Φ be a partition of unity subordinate to the covering of M by neighbourhoods $\mathcal{U}_\alpha$. $\varphi_\alpha \in \Phi$ is non-negative and timecones are convex, so the vector field $X = \sum \varphi_\alpha X_{\mathcal{U}_\alpha}$ is timelike.

Exercise: Prove the converse of Lemma 1.3.2.

If M is time-oriented, a differentiable timelike curve is future-directed if the tangent at each of its points lies in the interior of the future null-cone of those points. Similarly, a *future-directed causal* (or non-spacelike) curve is one whose tangent vectors to each of its points is either future-pointing timelike, or null. Corresponding definitions hold for past-directed curves. Relativistically, a future-pointing timelike curve $\alpha : I \to M$ such that $|\alpha'(\tau)| = 1$ for the proper time parameter $\tau \in I$ is called a *material particle*. A future-pointing null geodesic $\gamma : I \to M$ is called a *lightlike particle*. A photon, for instance, is a lightlike particle. The tangent directions of light determine a lightcone and the tangent lines of a material particle lie inside of this cone. Thus any material particle must travel at a speed less than the speed of light.

For the remainder of this exposition, consider M an n-dimensional time-oriented Lorentz manifold unless otherwise specified.

1.4 Submanifolds and Warped Products

Definition 1.4.1 *Let $i : P \to M$ be the inclusion map of a regular submanifold $P \subset M$, where M is a semi-Riemannian manifold. P is a* semi-Riemannian submanifold *of M if the pull back $i^*(\mathrm{g})$ is a metric tensor on P.*

A vector field $X \in \mathfrak{X}(M)$ can be orthogonally decomposed as $X = X^\top + X^\perp$ where the tangential component of X is $X^\top \in \mathfrak{X}(P)$ and the normal component $X^\perp \in \mathfrak{X}(P)^\perp$. The Levi-Civita connection $\overline{\nabla}$ of M can also be orthogonally decomposed into tangential and normal components:

$$\overline{\nabla}_V W = \nabla_V W + II(V, W) \qquad V, W \in \mathfrak{X}(M).$$

∇ is the Levi-Civita connection on P and $\nabla_V W = (\overline{\nabla}_V W)^{\top}$. The normal component of $\overline{\nabla}_V W$ defines the Second Fundamental Form tensor, $II(V, W)$ of $P \subset M$.

Example 1.4.2 *The Levi-Civita Connection*
Let M be the Lorentz manifold with metric

$$ds^2 = 2F(u,v)dudv = F(u,v)(du \otimes dv + dv \otimes du),$$

for $F(u,v) > 0$. Consider the spacelike curve $\Delta = \{u = v\}$ with orthonormal frame $X = \frac{1}{\sqrt{2F}}(\partial_v + \partial_u)$, and let $U = \frac{1}{\sqrt{2F}}(\partial_v - \partial_u)$ be the future-poiting timelike unit normal to Δ.

The covariant derivative of X can be determined by considering

$$\langle \overline{\nabla}_X X, U \rangle = \left\langle \overline{\nabla}_{\frac{1}{\sqrt{2F}}(\partial_v + \partial_u)} \left(\frac{1}{\sqrt{2F}}(\partial_v + \partial_u) \right), \frac{1}{\sqrt{2F}}(\partial_v - \partial_u) \right\rangle.$$

The calculation can be reduced to

$$\begin{aligned} \langle \overline{\nabla}_{(\partial_v + \partial_u)}(\partial_v + \partial_u), \partial_v - \partial_u \rangle &= -\langle \partial_v + \partial_u, \overline{\nabla}_{\partial_v + \partial_u}(\partial_v - \partial_u) \rangle \\ &= -\langle \partial_v + \partial_u, \overline{\nabla}_{\partial_v - \partial_u}(\partial_v + \partial_u) \rangle \\ &= -\frac{1}{2} \overline{\nabla}_{\partial_v - \partial_u} \langle \partial_u + \partial_v, \partial_u + \partial_v \rangle \\ &= -\frac{1}{2} \overline{\nabla}_{\partial_v - \partial_u}(2F) \\ &= -\left.\frac{d}{dt}\right|_{t=0} F((u,u) + t(-1,1)) \\ &= \left.\left(\frac{\partial F}{\partial u} - \frac{\partial F}{\partial v} \right)\right|_{(u,u)} \end{aligned}$$

In the next chapter we will make use of the shape operator $\widetilde{II}$ of $P \subset M$ in order to simplify calculations. For $V \in \mathfrak{X}(P)$ and $Z \in \mathfrak{X}(P)^{\perp}$, define $\widetilde{II}(V, Z) = (\overline{\nabla}_V Z)^{\top}$. $\widetilde{II}$ and II provide the same information, in fact $\langle \widetilde{II}(V, Z), W \rangle = -\langle II(V, W), Z \rangle$ for all $V, W \in \mathfrak{X}(P)$, $Z \in \mathfrak{X}(P)^{\perp}$. To see this, consider $\overline{\nabla}_V \langle Z, W \rangle = \langle \overline{\nabla}_V Z, W \rangle + \langle Z, \overline{\nabla}_V W \rangle = 0$. Then

$$\langle \overline{\nabla}_V Z, W \rangle = -\langle Z, \overline{\nabla}_V W \rangle \Leftrightarrow \langle \widetilde{II}(V, Z), W \rangle = -\langle II(V, W), Z \rangle.$$

In the special situation of warped products (defined below) the second fundamental form may be computed easily, as developed by O'Neill (1983).

Definition 1.4.3 *Suppose B and F are semi-Riemannian manifolds and let $f > 0$ be a smooth function on B. The* warped product *$M = B \times_f F$ is the product manifold $B \times F$ furnished with metric tensor*

$$\mathbf{g} = \pi^*(\mathbf{g}_B) + (f \circ \pi)^2 \sigma^*(\mathbf{g}_F)$$

where π and σ are the projections of $B \times F$ to B and F respectively.

Lemma 1.4.4 *(see [O] 7.35) If V, W are vector fields in the lift of F, then the Second Fundamental Form of the fiber can be expressed as*

$$II(V,W) = -(\langle V,W\rangle / f)\ \mathrm{grad}\ f$$

with grad $f = \sum_{i,j} \mathbf{g}^{\mathrm{ij}} \frac{\partial \mathrm{f}}{\partial \mathrm{x}^{\mathrm{i}}} \partial_{\mathrm{j}}$.

Proof. Let X be in the lift of B. Then

$$\langle II(V,W), X\rangle = -\left\langle \widetilde{II}(V,X), W\right\rangle = -\langle \overline{\nabla}_V X, W\rangle = -\langle \overline{\nabla}_X V, W\rangle . \quad (1.1)$$

Furthermore, $X\langle V,W\rangle = X(f^2 \langle V,W\rangle|_F \circ \sigma)$, writing f for $f \circ \pi$.
Claim: $\overline{\nabla}_X V = (Xf/f)V$.

$$\begin{aligned} X\langle V,W\rangle &= 2fXf(\langle V,W\rangle|_F \circ \sigma) \\ &= 2(Xf/f)\langle V,W\rangle \end{aligned}$$

and by the Koszul formula:

$$\begin{aligned} 2\left\langle \overline{\nabla}_X V, W\right\rangle &= X\langle V,W\rangle + V\langle W,X\rangle - W\langle X,V\rangle \\ &\quad - \langle X,[V,W]\rangle + \langle V,[W,X]\rangle + \langle W,[X,V]\rangle \end{aligned}$$

we get

$$2\left\langle \overline{\nabla}_X V, W\right\rangle = X\langle V,W\rangle = 2(Xf/f)\langle V,W\rangle \ \text{ and } \ \overline{\nabla}_X V = (Xf/f)V.$$

Equation (1.1) becomes

$$-\langle (Xf/f)V, W\rangle = -(\langle V,W\rangle/f)Xf = -(\langle V,W\rangle/f)\langle \mathrm{grad} f, X\rangle$$

and the result follows. □

1.5 Examples and Computations in Lorentz Geometry

Example 1.5.1 *Ricci Tensor, $R_{\alpha\beta}$*

Consider the Robertson-Walker spacetime M with metric

$$ds^2 = -dt^2 + a^2(t)\Big(dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta\ d\phi^2)\Big),$$

where $a(t)$ is as in example 0.0.1.

To compute the Ricci tensor $R_{\alpha\beta} = \sum R_{\alpha\gamma\beta}{}^{\gamma}$ we begin by computing the nonzero Christoffel symbols

$$\Gamma^k_{ij} = \frac{1}{2}\sum_m \mathbf{g}^{\mathbf{km}}\Big(\frac{\partial \mathbf{g_{jm}}}{\partial \mathbf{x^i}} + \frac{\partial \mathbf{g_{im}}}{\partial \mathbf{x^j}} - \frac{\partial \mathbf{g_{ij}}}{\partial \mathbf{x^m}}\Big).$$

Taking $a' = a'(t)$ we get:

$$\begin{array}{ll}
\Gamma^t_{rr} = aa' & \Gamma^r_{\phi\phi} = -\sin r\cos r\sin^2\theta \\
\Gamma^t_{\theta\theta} = aa'\sin^2 r & \Gamma^\theta_{r\theta} = \Gamma^\phi_{r\phi} = \frac{\cos r}{\sin r} \\
\Gamma^t_{\phi\phi} = aa'\sin^2 r\sin^2\theta & \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta \\
\Gamma^r_{\theta\theta} = -\sin r\cos r & \Gamma^\phi_{\theta\phi} = \frac{\cos\theta}{\sin\theta} \\
\Gamma^r_{tr} = \Gamma^\theta_{t\theta} = \Gamma^\phi_{t\phi} = \frac{a'}{a} &
\end{array}$$

The Riemannian tensor can now be calculated using the equation

$$R^i_{jkm} = \frac{\partial \Gamma^i_{mj}}{\partial x^k} - \frac{\partial \Gamma^i_{kj}}{\partial x^m} + \sum_l (\Gamma^l_{mj}\Gamma^i_{kl} - \Gamma^l_{kj}\Gamma^i_{ml}).$$

$$\begin{array}{ll}
R^r_{trt} = R^\theta_{t\theta t} = R^\phi_{t\phi t} = -\frac{a'}{a} & \\
R^t_{rtr} = aa'' & R^\theta_{r\theta r} = R^\phi_{r\phi r} = (a')^2 + 1 \\
R^t_{\theta t\theta} = aa''\sin r & R^r_{\theta r\theta} = R^\phi_{\theta\phi\theta} = ((a')^2+1)\sin^2 r \\
R^t_{\phi t\phi} = aa''\sin^2 r\sin^2\theta & R^r_{\phi r\phi} = R^\theta_{\phi\theta\phi} = ((a')^2+1)\sin^2 r\sin^2\theta
\end{array}$$

Thus the components of the Ricci tensor are:

$$\begin{array}{ll}
R_{tt} = -3\frac{a''}{a} & R_{\theta\theta} = (aa'' + 2(a')^2 + 2)\sin^2 r \\
R_{rr} = aa'' + 2(a')^2 + 2 & R_{\phi\phi} = (aa'' + 2(a')^2 + 2)\sin^2 r\sin^2\theta.
\end{array}$$

The scalar curvature is equal to $R = \sum \mathbf{g^{ii}R_{ii}} = \mathbf{6}\Big(\frac{\mathbf{a''}}{\mathbf{a}} + \frac{\mathbf{(a')^2}}{\mathbf{a^2}} + \frac{\mathbf{1}}{\mathbf{a^2}}\Big).$

Example 1.5.2 *Einstein's Equation*

Using the equations from the above example, we can verify Einstein's field equation for the Robertson-Walker spherical spacetime. In component form,

$$G_{ab} = R_{ab} - \frac{1}{2}R\mathbf{g_{ab}} = 8\pi\mathbf{T_{ab}}$$

where G_{ab} is Einstein's gravitational tensor, R is the scalar curvature, and T_{ab} is the stress-energy tensor. The most general form that the stress-energy tensor can take for a homogeneous and isotropic spacetime is that of a perfect fluid:

$$T = (\rho + \mathfrak{p})U^*U^* + \mathfrak{p}\mathbf{g}.$$

where U^* is the one-form metrically equivalent to the future-pointing time-like unit vector field U on M, ρ is the energy density function, and $\mathfrak{p} = (\rho/3)$ is the pressure function. The stress-energy tensor components are

$$T_{tt} = \rho \quad T_{\cdot\cdot} = \mathfrak{p}\mathbf{g}_{\cdot\cdot} \quad \text{for } \bullet \neq \mathbf{t}.$$

Thus the components of Einstein's gravitational tensor are

$$\begin{aligned} G_{tt} &= 3\,\frac{(a')^2}{a^2} + \frac{3}{a^2} &= 8\pi\rho \\ G_{\cdot\cdot} &= -2\,\frac{a''}{a} - \frac{(a')^2}{a^2} - \frac{1}{a^2} &= 8\pi\mathfrak{p}. \end{aligned}$$

Rewriting these equations gives one of the general evolution equations for homogeneous, isotropic spacetimes:

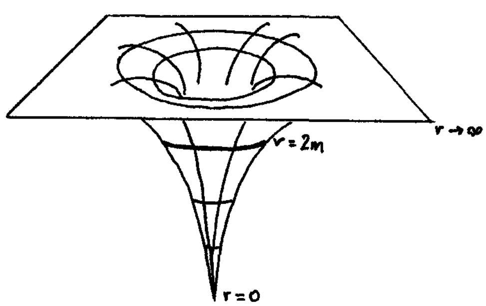
$$3\frac{a''}{a} = -4\pi(3\mathfrak{p} + \rho).$$

Exercise: Verify that Schwarzschild's model of a spacetime (cf ex. 0.0.2) is Ricci flat.

Example 1.5.3 *Change of Coordinates*

Consider the Schwarzschild spacetime M as the warped product $M = \mathbb{R} \times \mathbb{R}_+ \times S^2$ with metric

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2). \qquad (1.2)$$



This cosmological model of a black hole appears to have two singularities; one at $r = 0$ (the actual singularity) and the other at $r = 2m$ (the horizon). With a change of coordinates we will show that the singularity at the horizon is merely due to a breakdown of coordinates in (1.2).

In a warped product, the leafs $\mathbb{R} \times \mathbb{R}_+ \times \{q\}$ are totally geodesic and isometric to each other. Thus, in order to derive the extension of M it is reasonable to assume it will likewise be a warped product with fiber S^2, and we will focus on the plane

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2$$

in order to remove the coordinate singularity.

We begin by looking at outgoing and ingoing radial null geodesics: null geodesics that travel away from and toward the horizon and which satisfy

$$0 = -(1 - \frac{2m}{r})(\frac{dt}{d\tau})^2 + (1 - \frac{2m}{r})^{-1}(\frac{dr}{d\tau})^2.$$

To solve for t in terms of r, rewrite $(\frac{dt}{dr})^2 = (\frac{r}{r-2m})^2$ as $dt = \frac{r}{r-2m}dr$. Then

$$t = \pm\Big(r + 2m\ln(\frac{r}{2m} - 1)\Big) + C, \quad C \in \mathbb{R}.$$

Let $\widetilde{r} = r + 2m\ln(\frac{r}{2m} - 1)$ and define null coordinates w and x by

$$w = t - \widetilde{r} \qquad x = t + \widetilde{r}.$$

Therefore,

$$ds^2 = -(1 - \frac{2m}{r})dwdx. \tag{1.3}$$

r is now a function of w and x described by $r + 2m\ln(\frac{r}{2m} - 1) = \frac{w-x}{2}$. Solving for $(\frac{r}{2m} - 1)$, (1.3) can be written as

$$ds^2 = -(\frac{2m\ e^{-r/2m}}{r}\ e^{\frac{w-x}{4m}})dwdx. \tag{1.4}$$

Let $U = e^{w/4m}$ and $V = -e^{-x/4m}$; (1.4) becomes

$$ds^2 = -\frac{32m^3}{r}e^{-r/2m}dUdV, \tag{1.5}$$

where $UV = -(\frac{r}{2m} - 1)e^{r/2m}$. This form of the metric allows us to infer easily that there is no singularity at $r = 2m$ (see the discussion in (i) below). In order to simplify future calculations we will make two more coordinate changes, rewriting (1.5) as a metric with (i) null coordinates and (ii) a timelike and spacelike coordinate.

(i) Let $u = \sqrt{\frac{2m}{e}}\, U$ and $v = -\sqrt{\frac{2m}{e}} V$, with $uv = (r - 2m)\ e^{r/2m-1}$ and $t = 2m \ln|\frac{v}{u}|$. Then

$$ds^2 = \frac{16m^2}{r}\ e^{1-r/2m} du dv.$$

Establish the notation $f(r) = (r - 2m)e^{r/2m-1}$ for $r > 0$. Since $f' > 0$, f is a diffeomorphism onto the open set $(\frac{-2m}{e}, \infty)$. The equation $r = f^{-1}(uv)$ defines a smooth positive function on the region $uv > \frac{-2m}{e}$ in the uv-plane. Thus, r is defined implicitly by the relation $f(r) = uv$, and $r = 2m$ is shown to be a break down of coordinates rather than a physical singularity. For further discussion, please refer to example 5.1.6. The extended Schwarzschild metric becomes

$$ds^2 = \frac{16m^2}{r}\ e^{1-r/2m} du dv + r^2(d\theta^2 + \sin^2\theta\ d\phi^2). \tag{1.6}$$

(ii) Let $T = \frac{U+V}{2}$ and $X = \frac{V-U}{2}$, with $X^2 - T^2 = (\frac{r}{2m} - 1)e^{r/2m}$ and $t = 2m \ln|\frac{T+X}{X-T}| = 4m \tanh^{-1}(\frac{T}{X})$. Then (1.2) can be rewritten as

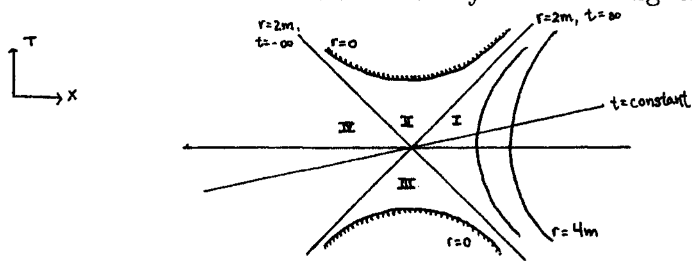
$$ds^2 = \frac{32m^3}{r} e^{-r/2m}(-dT^2 + dX^2) + r^2(d\theta^2 + \sin^2\theta\ d\phi^2). \tag{1.7}$$

This smooth extension of the Schwarzschild solution was first described by Kruskal in the article "Maximal Extension of Schwarzschild Metric" (1960). A manifold with metric (1.6) or (1.7) is called a *Kruskal spacetime* and will be denoted by the warped product $K = B \times_r S^2$. From the form of the metric (1.7), one sees that the metric on $B \times \{q\}$ is conformally equivalent to an open set in Minkowski space $\mathbb{R}^2_1$.

Physical interpretation. In the (X, T) coordinates a spacelike curve of B takes the form

$$X^2 - T^2 = (\frac{r}{2m} - 1)e^{r/2m} \quad \text{for } r > 0.$$

The physical singularity at $r = 0$ is represented by the curve $X^2 - T^2 = -1$. The Kruskal extension can be illustrated by the following diagrams.



Region I is the asymptotically flat region that corresponds to the exterior ($r > 2m$) of the Schwarzschild spacetime. An in-falling material (or light-like) particle will cross the horizon $X = T$ ($r = 2m$) and enter region II. In region II the particle experiences such a strong gravitational pull that within finite proper time it will fall into the singularity at $X = (T^2 - 1)^{1/2}$, never able to escape back to region I. Region III is known as a "white hole," it has the time reversed properties of II. A particle in region II must have originated at the singularity $X = -(T^2-1)^{1/2}$ and will leave III within finite time, entering region IV. Region IV has properties identical to region I. For additional physical interpretations of the Kruskal extension, please see the detailed discussion in [W] section 6.4.

Example 1.5.4 *Second Fundamental Form*
If we consider the Robertson-Walker spacetime $M = \mathbb{R} \times_a S^3$ and the Kruskal spacetime $K = B \times_r S^2$ as warped products we can use Lemma 1.4.4 to compute II of their fibers.

For M, $II(V, W) = \langle V, W \rangle (a'/a)U$, where $U = \partial_t$. This follows immediately since grad $a = -a'U$.

For K, we can compute II of the orthogonal frame of S^2.

$$\begin{aligned} II(\partial_\theta, \partial_\theta) &= -\frac{\langle \partial_\theta, \partial_\theta \rangle}{r} \text{ grad } r = -r \text{ grad } r \\ II(\partial_\phi, \partial_\phi) &= -\frac{\langle \partial_\phi, \partial_\phi \rangle}{r} \text{ grad } r = -r \sin^2\theta \text{ grad } r \end{aligned}$$

Solving for grad r in terms of ∂_u and ∂_v will allow us to write II in the Kruskal null coordinates u and v.

$$\begin{aligned} \text{grad } r &= \sum_{i,j} \mathrm{g}^{ij} \frac{\partial r}{\partial x^i} \, \partial_j \\ &= -(1 - \frac{2m}{r})^{-1} \frac{\partial r}{\partial t} \, \partial_t + (1 - \frac{2m}{r}) \frac{\partial r}{\partial r} \, \partial_r \end{aligned}$$

$$= (1 - \frac{2m}{r})\ \partial_r.$$

One may compute dr as a linear combination of du and dv from the relation
$uv = (r-2m)e^{\frac{r}{2m}-1}$ (cf example 1.5.3(i)):

$$dr = 2m\ (1 - \frac{2m}{r})\ \left(\frac{du}{u} + \frac{dv}{v}\right), \quad uv \neq 0.$$

From this, the dual relationship follows easily. We get,

$$\partial_r = \frac{1}{4m}\ (1 - \frac{2m}{r})^{-1}(u\partial_u + v\partial_v) \text{ and grad } r = \frac{1}{4m}(u\partial_u + v\partial_v).$$

Thus,

$$\begin{aligned} II(\partial_\theta, \partial_\theta) &= \frac{-r}{4m}\ (u\partial_u + v\partial_v), \\ II(\partial_\phi, \partial_\phi) &= \frac{-r\sin^2\theta}{4m}\ (u\partial_u + v\partial_v). \end{aligned}$$

Also for K, we may consider the curve $\{u = v\}$ as a subspace of B and compute its second fundamental form. By example 1.4.2,

$$\langle \nabla_X X, U\rangle = \left\langle \nabla_{\frac{1}{2\sqrt{F}}(\partial_v + \partial_u)}\left(\frac{1}{2\sqrt{F}}(\partial_v + \partial_u)\right), \frac{1}{2\sqrt{F}}(\partial_v - \partial_u)\right\rangle = 2\left(\frac{\partial F}{\partial u} - \frac{\partial F}{\partial v}\right).$$

Define $F(u,v) = h(r)$ and $uv = f(r)$. Then

$$\frac{\partial F}{\partial u} = h'(r)\ \frac{\partial r}{\partial u} = \frac{h'(r)}{f'(r)}\ v,$$

and,

$$\frac{\partial F}{\partial v} = h'(r)\ \frac{\partial r}{\partial v} = \frac{h'(r)}{f'(r)}\ u.$$

Thus at $u = v$ the covariant derivative, and hence, the second fundamental form vanishes.

Chapter 2

Focal Points and Convergence

With some background in Lorentz geometry, we are now able to begin exploring the circumstances under which a spacetime may be singular.

Define a *variation* of a curve segment $\alpha : [a, b] \to M$ as a two-parameter map $x : [a, b] \times (-\delta, \delta) \to M$ such that $\alpha(s) = x(s, 0)$ for all $a \leq s \leq b$. x is called a *fixed endpoint variation* if its first and last transverse curves (t-parameter curves) are constant. In this case, all of the longitudinal curves (s-parameter curves) go from $p = \alpha(a)$ to $q = \alpha(b)$, and the variation vector field V vanishes at a and b. If every longitudinal curve of x is a geodesic, then x is a one-parameter family of geodesics.

By studying variations through geodesics we can examine the effect of curvature on nearby geodesics. In order to study geodesics with respect to a variation, it is helpful to consider the set of all piecewise smooth curve segments $\alpha : [0, b] \to M$ from p to q as a manifold. We can then think of (the nonnull) geodesics as the critical points of the length function L. The first variation of arc length, L', vanishes at its critical points, at which the second derivatives of L become interesting.

Definition 2.0.5 *Let $\sigma : [0, b] \to M$ be a nonnull geodesic such that $\sigma(0) = p$ and $\sigma(b) = q$. The* index form *I_σ of σ is*

$$I_\sigma(V, V) = \frac{\epsilon}{|\sigma'|} \int_0^b \left\langle V'^{\perp}, V'^{\perp} \right\rangle - \langle R_{V\sigma'} V, \sigma' \rangle \, du - \frac{\epsilon}{|\sigma'|} \langle \sigma', A \rangle \Big|_0^b,$$

where V is the variation vector field $V(s) = \frac{d}{dt}\big|_{t=0} x(s, t)$, A is the transverse acceleration vector field of x, and $\epsilon = sgn\langle \sigma', \sigma' \rangle = \pm 1$ ($+1$ if σ is spacelike, -1 if σ is timelike).

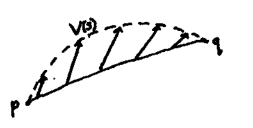
By definition $I_\sigma(V,V) = L''(0)$, so the index form provides the same information as the second variation of arc length. If the variation vector field V is tangent to σ, then $V^\perp = 0$ and $I_\sigma(V,V) = 0$. Thus, without loss of generality, we can restrict the index form to $I_\sigma^\perp$, the index form of variation vector fields V that are everywhere orthogonal to σ.

The index form provides a quantitative way to measure the effect of curvature on a variation through geodesics. In particular $I_\sigma^\perp$ offers information about the existence of conjugate points, and consequently, whether geodesics fail to be minimizing or maximizing (depending on their causal character).

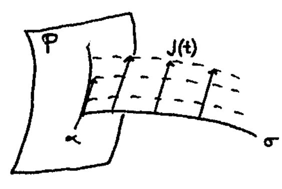
Points $p, q \in M$ are *conjugate* along a geodesic segment σ, if there is a Jacobi field $J(t)$ along σ that vanishes at p and q, but is not everywhere zero (see figure (i) below). An important property of points conjugate to p is that they are precisely the points where the exponential map $\exp_p$ fails to be a local diffeomorphism. For instance, every geodesic starting at a point p in the sphere $S^n(R)$ of radius R will meet at the antipodal point. The exponential map is a diffeomorphism on the ball in $T_pS^n(R)$ centered at 0 with radius πR. However, there are points on the boundary of this ball for which $\exp_p$ is not a local diffeomorphism. Moreover, each Jacobi field on $S^n(R)$ that vanishes at p has its first zero at the antipodal point.

Similarly, the existence of conjugate points can relate the curvature of a spacetime with its topological structure. The main area of interest in this chapter is the existence of focal points, points which are conjugate to a submanifold rather than another point. As we shall see in Chapters 4 and 5, points that are conjugate to submanifolds describe the situation when a timelike geodesic fails to maximize the proper time between two points. Furthermore, we shall see that focal points are the points at which a spacetime may fail to be timelike (or null) complete.

To describe a focal point, we must first define a P-Jacobi field. Essentially a P-Jacobi field, as depicted in figure (ii), is the variation vector field of a variation x through normal geodesics such that $J(0)$ is tangent to the submanifold $P \subset M$.



(i) conjugate points



(ii)P-Jacobi field

Definition 2.0.6 *A Jacobi field J along a geodesic σ normal to $P \subset M$ is called a P-*Jacobi field *if it satisfies the following conditions:*

(i) $J(0)$ *is tangent to* P,

(ii) $(\overline{\nabla}_{\sigma'(0)}J)^{\top} = \widetilde{II}(J(0), \sigma'(0))$.

A P-Jacobi field of a fixed endpoint variation x is tangent to P at $J(0)$ and vanishes at $J(b)$. For such an x, the term $\langle \sigma', A\rangle |_0^b$ of the index form reduces to $-\langle \sigma'(0), A(0)\rangle$ since the last transverse curve of x is constant. If α is the first transverse curve of x, then

$$\langle \sigma'(0), A(0)\rangle = \left\langle \sigma'(0), (\alpha'(0))^{\perp}\right\rangle = \langle \sigma'(0), II(\alpha'(0), \alpha'(0)\rangle = \langle \sigma'(0), II(J(0), J(0))\rangle .$$

Thus the index form can be written as

$$I_\sigma(V, V) = \frac{\epsilon}{|\sigma'|}\int_0^b \langle V'^{\perp}, V'^{\perp}\rangle - \langle R_{V\sigma'}V, \sigma'\rangle\, du - \frac{\epsilon}{|\sigma'|}\langle \sigma', II(J(0), J(0))\rangle .$$

Since the Jacobi equation is linear, the set of normal Jacobi fields forms a $(2n-2)$-dimensional vector space. The set of P-Jacobi fields forms an $(n-1)$-dimensional vector space, so the space of P-Jacobi fields on σ that vanish at a focal point has dimension at most $n-1$. It is left as an exercise to the reader to verify these statements.

Proposition 2.0.7 *(see [O] 10.28) A Jacobi field* J *along a geodesic* σ *is a* P*-Jacobi field if and only if it is the variation vector field of a variation* x *of* σ *through normal geodesics.*

Proof. Given such a variation x, it is not difficult to see that the variation vector field must satisfy the conditions of a P-Jacobi field. Consider $J(s) = \frac{\partial}{\partial t}|_{t=0}\ x(s,t)$. The first transverse curve α of x lies in P with $\alpha'(0) = J(0)$, see figure (ii) above. Thus $J(0)$ is tangent to P. What is left to show is $\widetilde{II}(J(0), \sigma'(0)) = \left(\overline{\nabla}_{\sigma'(0)}J\right)^{\top}$. Let $Z(t) = \frac{\partial}{\partial s}|_{s=0}\ x(s,t)$ be the vector field on α that is normal to P, so that $Z(0) = \sigma'(0)$. Then

$$\left(\overline{\nabla}_{\alpha'(0)}Z\right)^{\top} = \widetilde{II}(\alpha'(0), Z(0)) = \widetilde{II}(J(0), \sigma'(0)).$$

However

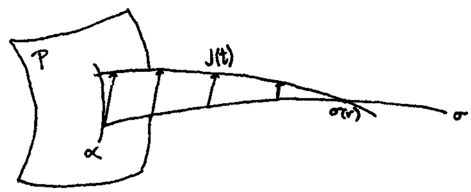
$$\overline{\nabla}_{\alpha'(0)}Z = \frac{\partial}{\partial t}|_{t=0}\frac{\partial}{\partial s}|_{s=0}\ x(s,t) = \frac{\partial}{\partial s}|_{s=0}\frac{\partial}{\partial t}|_{t=0}\ x(s,t) = \overline{\nabla}_{\sigma'(0)}J,$$

and the second condition follows.
For the reverse direction, see [O]. □

If x of Proposition 2.0.7 is a fixed endpoint variation of $\sigma : [0, b] \to M$, then $\sigma(b)$ is a focal point of P.

Definition 2.0.8 *Let σ be a geodesic of M that is normal to $P \subset M$. Then $\sigma(r)$, $r \neq 0$ is a* focal point *of P along σ provided there is a nonzero P-Jacobi field $J(t)$ on σ with $J(r) = 0$.*

One may think of a focal point as the intersection point of two geodesics that are orthogonal to P and infinitesimally close together, as in the diagram below (diagram not to scale).



In order to determine the existence of focal points one can examine the initial rate of convergence of geodesics passing orthogonally through P. This rate of convergence (i.e. the convergence of P as defined below) measures the "bending" of P in the spacetime M. Intuitively, one might think of how light bends and focuses depending on the concavity of a lens. Similarly, the convergence of P establishes whether geodesics spread out or focus. This in turn predicts how P will change as it flows along these geodesics. If the convergence is nonzero, the resulting surfaces will either expand or contract depending on the direction of the outward normal. Analogous concepts are discussed in [HE] and [W]. The expansion and extrinsic curvature of a Riemannian hypersurface in the spacetime M provide a relativistic explanation of this geometrical property. The description used here is the one stated in [O] and is defined explicitly as follows.

Definition 2.0.9 *Let P be a semi-Riemannian submanifold of M with mean curvature vector field H. The* convergence *of P is the real-valued function $\boldsymbol{k}$ on the normal bundle NP such that*

$$\boldsymbol{k}(z) = \langle z, H_p \rangle = \frac{1}{dimP} \; trace \; II_z \; for \; z \in T_pP^{\perp}$$

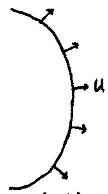
For a spacelike hypersurface in M,

$$H_p = \frac{1}{n-1} \sum_{i=1}^{n-1} II(e_i, e_i),$$

where $e_1, \ldots, e_{n-1}$ is any orthonormal basis for T_pP.

If P is a spacelike hypersurface in M there are two choices for an outward normal vector. Either z is future-pointing or z is past-pointing. In the former case, $k > 0$ is a necessary condition for the focusing of orthogonal geodesics. Conversely, if z is past-pointing, then k should be negative for geodesics to focus.

(i) U is past-pointing, $k < 0$



(ii) U is future-pointing, $k < 0$

2.1 Focal Points and Timelike Curves

The existence of focal points along timelike geodesics reveals much about the intrinsic geometry of a semi-Riemannian manifold. For instance, as we will discuss later, the expansion (or contraction) of our universe can be predicted by its curvature and the convergence of the spacelike hypersurface $S = \{t = constant\}$. If the initial rate of convergence of S is positive along a future-pointing timelike geodesic, S will contract as $t \to \infty$. In particular, the timelike geodesics will be focused toward a point conjugate to S. Initially, we look at how the existence of focal points affects the index form of P-Jacobi fields.

Theorem 2.1.1 *(see [O] 10.34)*

1. *Let $\sigma : [0, b] \to M$ be a geodesic segment from P to a point q (with sign ϵ) that is nonnull, and such that $\sigma'(s)^{\perp} \subset T_{\sigma(s)}M$ is spacelike. If there are no focal points of $\sigma(0) \in P$ along σ, then the index form $I_{\sigma}^{\perp}$ is definite (positive if $\epsilon = 1$, negative if $\epsilon = -1$).*

2. *If there is a focal point $\sigma(s)$, $0 < s < b$, along σ, then I_{σ} is not semi-definite.*

Proof of 1. Choose a basis $Y_1, \ldots, Y_{n-1}$ for the P-Jacobi fields on σ. Let V be the variation vector field of a fixed endpoint variation x along σ. It is possible to show that there exist unique piecewise smooth functions f_i defined on $[0, b]$ such that $V = \sum f_i Y_i$. This is left as an exercise for the reader. We would like to evaluate $\epsilon I_{\sigma}(V, V)$. To simplify the calculation, we first establish two lemmas.

Lemma 1: *If V, W are P-Jacobi fields on σ then $\langle V', W\rangle = \langle V, W'\rangle$.*
We show this in two parts:

i) $\langle V', W\rangle - \langle V, W'\rangle$ is constant.

$$\begin{aligned}\langle V', W\rangle' &= \langle V'', W\rangle + \langle V', W'\rangle \\ &= \langle R_{V\sigma'}\sigma', W\rangle + \langle V', W'\rangle \text{ by Jacobi equation} \\ &= -\langle R_{V\sigma'}W, \sigma'\rangle + \langle V', W'\rangle\end{aligned}$$

By symmetry of the Riemannian curvature tensor, $\langle V', W\rangle' = \langle V, W'\rangle'$. Integrate both sides to get $\langle V', W\rangle - \langle V, W'\rangle =$ constant.

ii) $\langle V'(0), W(0)\rangle = \langle V(0), W'(0)\rangle$.
Since W is a P-Jacobi field, $W(0)$ is tangent to P. Therefore,

$$\begin{aligned}\langle V'(0), W(0)\rangle &= \langle V'(0)^{\top}, W(0)\rangle \\ &= \left\langle \widetilde{II}(V(0), \sigma'(0)), W(0)\right\rangle \qquad (2.1) \\ &= -\langle II(V(0), W(0)), \sigma'(0)\rangle, \qquad (2.2)\end{aligned}$$

where (2.1) follows from part (*ii*) of Definition 2.0.6, and (2.2) from the fact that V and W are perpendicular to σ. The result follows from the symmetry of II.

Lemma 2: *Let $Y_1, \ldots, Y_r$ be Jacobi fields along σ such that $\langle Y_i', Y_j\rangle = \langle Y_i, Y_j'\rangle$ for all i, j. If $V = \sum f_i Y_i$ then,*

$$\langle V', V'\rangle - \langle R_{V\sigma'}V, \sigma'\rangle = \langle A, A\rangle + \langle V, B\rangle',$$

where $A = \sum f_i' Y_i$, $B = \sum f_i Y_i'$.
Write $V' = A + B$. Then $\langle V', V'\rangle = \langle A, A\rangle + 2\langle A, B\rangle + \langle B, B\rangle$. Now compute $\langle V, B\rangle'$:

$$\begin{aligned}\langle V, B\rangle' &= \langle V', B\rangle + \langle V, B'\rangle \\ &= \langle A + B, B\rangle + \left\langle V, \sum f_i' Y_i'\right\rangle + \left\langle V, \sum f_i Y_i''\right\rangle \\ &= \langle A, B\rangle + \langle B, B\rangle + \left\langle \sum f_j Y_j, \sum f_i' Y_i'\right\rangle + \left\langle V, \sum f_i R_{Y_i\sigma'}\sigma'\right\rangle \\ &= \langle A, B\rangle + \langle B, B\rangle + \sum f_j f_i' \langle Y_j, Y_i'\rangle + \langle V, R_{\sum f_i Y_i \sigma'}\sigma'\rangle \\ &= \langle A, B\rangle + \langle B, B\rangle + \left\langle \sum f_j Y_j', \sum f_i' Y_i\right\rangle + \langle V, R_{V\sigma'}\sigma'\rangle \\ &= \langle A, B\rangle + \langle B, B\rangle + \langle A, B\rangle - \langle \sigma', R_{V\sigma'}V\rangle.\end{aligned}$$

$$\begin{aligned}\text{Thus, } \langle A, A\rangle + \langle V, B\rangle' &= \langle A, A\rangle + 2\langle A, B\rangle + \langle B, B\rangle - \langle R_{V\sigma'}V, \sigma'\rangle \\ &= \langle V', V'\rangle - \langle R_{V\sigma'}V, \sigma'\rangle, \text{ which finishes the proof of this lemma.}\end{aligned}$$

Assuming $|\sigma'| = 1$, then by Lemma 2 the index form becomes:

$$\begin{aligned}\epsilon I_\sigma(V,V) &= \int_0^b \langle V',V'\rangle - \langle R_{V\sigma'}V,\sigma'\rangle\, du - \langle \sigma'(0), II(V(0),V(0))\rangle \\ &= \int_0^b \langle A,A\rangle\, du + \langle V,B\rangle\,|_0^b - \langle \sigma'(0), II(V(0),V(0))\rangle\,. \quad (2.3)\end{aligned}$$

However, x is a fixed endpoint variation so $V(b) = 0$. We can rewrite the term $\langle V,B\rangle\,|_0^b$ as:

$$\begin{aligned}\langle V,B\rangle\,|_0^b &= -\langle V(0),B(0)\rangle \\ &= -\left\langle V(0), \sum f_i(0)Y_i'(0)\right\rangle \\ &= -\sum f_i(0)\,\langle V(0),Y_i'(0)\rangle \\ &= -\sum f_i(0)\,\langle V(0),Y_i'(0)^{\top}\rangle \\ &= -\sum f_i(0)\left\langle V(0),\widetilde{II}(Y_i(0),\sigma'(0))\right\rangle \\ &= -\left\langle V(0),\widetilde{II}(V(0),\sigma'(0))\right\rangle \\ &= \langle \sigma'(0), II(V(0),V(0))\rangle\,.\end{aligned}$$

Hence $\langle V,B\rangle\,|_0^b$ cancels out with the last term in (2.3). Furthermore, $\langle A,A\rangle \geq 0$ since A is a linear combination of spacelike Jacobi fields. Therefore, $\epsilon I_\sigma(V,V) \geq 0$. If $\epsilon I_\sigma(V,V) = 0$ then $\int_0^b \langle A,A\rangle = 0 \Rightarrow \langle A,A\rangle = 0 \Rightarrow A = 0$, which implies that each f_i is constant and $V = 0$. □

The rigorous proof of part 2 of Theorem 2.1.1 is left as an exercise. Intuitively, the argument goes as follows. Suppose $\sigma(s)$ is a focal point of σ. There is a timelike geodesic β that is infinitely close to $\sigma|_{[0,s]}$ with the same endpoints and length. The broken timelike curve $\beta + \sigma|_{(s,b]}$ has length equal to $L(\sigma)$. By smoothing out this corner we produce a timelike curve whose length is greater than that of σ. Now, if V is a smooth tangent vector field along σ pointing toward the longer curve and such that $V(0) = V(b) = 0$, then $I_\sigma(V,V)$ should be positive.

The following theorem establishes a relationship between the curvature of a spacetime and the existence of focal points along timelike geodesics. It is a key ingredient in the proof of Hawking's Singularity Theorem (cf Theorem 5.1.2).

Theorem 2.1.2 *([O] 10.37) Let P be a spacelike hypersurface in an n-dimensional Lorentz manifold M, and let σ be a geodesic normal to P at the point $p = \sigma(0)$. Suppose*

1. $\boldsymbol{k}(\sigma'(0)) = \langle \sigma'(0), H_p \rangle > 0$.

2. $Ric(\sigma', \sigma') \geq 0$.

Then there is a focal point $\sigma(r)$ *of* P *along* σ *with* $0 < r \leq \frac{1}{\boldsymbol{k}(\sigma'(0))}$, *provided* σ *is defined on this interval.*

Proof. To prove the theorem, it suffices to show that ϵI_σ is not positive definite. Suppose $|\sigma'| = 1$ and let $k = \boldsymbol{k}(\sigma'(0))$. Choose an orthonormal basis $\{e_i\}$ for $T_p(P)$ and extend it to a parallel orthonormal frame of spacelike vector fields $\{E_i\}$ along σ. Define a function $f : [0, \frac{1}{k}] \to [0,1]$ by $f(u) = 1 - ku$ such that fE_i is a piecewise smooth vector field with $(fE_i)(0) \in T_p(P)$ and $(fE_i)(\frac{1}{k}) = 0$. Now consider $\epsilon I_\sigma(fE_i, fE_i)$. Since fE_i is orthogonal to σ, $I_\sigma^\perp = I_\sigma$. Thus,

$$\begin{aligned}
\epsilon I_\sigma(fE_i, fE_i) &= \int_0^{\frac{1}{k}} \langle (fE_i)'^\perp, (fE_i)'^\perp \rangle - \langle R_{fE_i\sigma'} fE_i, \sigma' \rangle \, du \\
&\quad - \langle \sigma'(0), II((fE_i)(0), (fE_i)(0)) \rangle \\
&= \int_0^{\frac{1}{k}} \langle -kE_i^\perp, -kE_i^\perp \rangle - f^2 \langle R_{E_i\sigma'} E_i, \sigma' \rangle \, du \\
&\quad - \langle \sigma'(0), II(E_i(0), E_i(0)) \rangle \\
&= k - \int_0^{\frac{1}{k}} f^2 \langle R_{E_i\sigma'} E_i, \sigma' \rangle \, du - \langle \sigma'(0), II(e_i, e_i) \rangle .
\end{aligned}$$

The E_i's are spacelike, so summing over i gives:

$$\begin{aligned}
\sum \epsilon I_\sigma(fE_i, fE_i) &= (n-1)k - \int_0^{\frac{1}{k}} f^2 Ric(\sigma', \sigma') du - \langle \sigma'(0), (n-1)H_p \rangle \\
&= (n-1)k - \int_0^{\frac{1}{k}} f^2 Ric(\sigma', \sigma') du - (n-1)k \\
&= - \int_0^{\frac{1}{k}} f^2 Ric(\sigma', \sigma') du \qquad (2.4)
\end{aligned}$$

By assumption $Ric(\sigma', \sigma') \geq 0$. Therefore $\epsilon I_\sigma(fE_i, fE_i) \leq 0$ for at least one value of i, and by Theorem 2.1.1(1) there must be a focal point of P along σ. □

2.2 Focal Points and Null Curves

There is an analogue to Theorem 2.1.2 that examines the connection between the convergence of a submanifold with codimension 2 and the existence of focal points along a null geodesic. Whereas Theorem 2.1.2 pertains to the existence of maximal timelike geodesics as well as incomplete timelike curves, the subsequent proposition will provide insight on the existence of incomplete null curves.

Proposition 2.2.1 *([O] 10.43) Let $P \subset M$ be a spacelike submanifold with codimension 2 of the Lorentz manifold M, and let H be the mean normal curvature vector field of P. If σ is a null geodesic normal to P at $\sigma'(0) = p$ such that:*

i) $k = \boldsymbol{k}(\sigma'(0)) = \langle \sigma'(0), H_p \rangle > 0$

ii) $Ric(\sigma', \sigma') \geq 0$,

then there is a focal point $\sigma(r)$ of P along σ with $0 \leq r \leq \frac{1}{k}$ provided σ is defined on this interval.

A consequence of focal points along a null geodesic is stated below. This theorem is also useful for describing aspects of the causal structure of a singular spacetime.

Theorem 2.2.2 *([O] 10.51) Let P be a spacelike submanifold of M. If $\alpha : [0, b] \longrightarrow M$ is a piecewise smooth causal curve from P to a point $q \in M$, then there is a piecewise smooth timelike curve arbitrarily near α that runs from P to q. However, this does not hold if α is a null geodesic normal to P along which there are no focal points of P before q.*

Exercises:

1. Discuss the differences between the proofs of Theorem 2.1.2 and Proposition 2.2.1.
2. Prove Theorem 2.2.2.

Chapter 3

Causal Structure of Spacetimes

The causal character of curves and vectors was briefly discussed in Chapter 1, in this chapter we shall look at the conditions that determine the causal structure of M as whole. The focus will be mainly on the future definitions of sets which comprise the causal structure of M, however definitions and theory for past versions of these sets are described simply by reversing time-orientation.

Let $p, q \in M$. We say that p *chronologically precedes* q if there is a future-pointing timelike curve segment from p to q, written as $p \ll q$. The set of all such $q \in M$ is called the *chronological future* of p and is denoted

$$I^+(p) = \{q \in M : p \ll q\}.$$

The point p *causally precedes* q ($p < q$) if there is a future-pointing causal curve segment from p to q. The *causal future* of p is

$$J^+(p) = \{q \in M : p \leq q, \text{ i.e. } p < q \text{ or } p = q\}.$$

In Minkowski spacetime, $I^+(p)$ is an open set and $J^+(p)$ is closed; the boundary of both sets is formed by the future nullcone of p. This is not necessarily the case in an arbitrary spacetime (cf example 3.0.6).

Theorem 3.0.3 *Let M be a spacetime and let $p \in M$. There exists a convex normal neighbourhood $\mathcal{C}$ of p. Furthermore the chronological future of p in the manifold $\mathcal{C}$, $I^+(p, \mathcal{C})$, is open in $\mathcal{C}$ (and in M).*

For the proof of this theorem please refer to [HE]. Theorem 3.0.3 links the causality of a spacetime M to its topology. In particular, it implies that the relation $\ll$, and hence the set $I^+(p)$, is open. It follows that the chronological future of a subset $A \subset M$, defined as $I^+(A) = \bigcup_{p\in A} I^+(p)$, is also an open set in M. The causal future of A is defined similarly as $J^+(A) = \bigcup_{p\in A} J^+(p)$. Notice the relations $\ll$ and $\leq$ are transitive and that $I^+(A) = I^+(I^+(A)) = I^+(J^+(A)) \subset J^+(A)$.

Exercise: If $p, q, r \in \mathcal{C}$ then either $q \in J^+(p,\mathcal{C})$, $r \in I^+(q,\mathcal{C})$ or $q \in I^+(p,\mathcal{C})$, $r \in J^+(p,\mathcal{C})$ imply $r \in I^+(p,\mathcal{C})$.

Let $q \in J^+(p)$ and $\alpha : [0,b] \to M$ is a causal (nonspacelike) curve from p to q. The image of $[0,b]$ is compact, so we can cover α by a finite subcover of convex neighbourhoods. If α is not a null geodesic in some convex neighbourhood $\mathcal{C}'$, then using Theorem 3.0.3 it is possible to deform α into a timelike curve in $\mathcal{C}'$. We can then extend this deformation to the other convex neighbourhoods to obtain a timelike curve from p to q.

Corollary 3.0.4 *([W] p. 191) If $q \in J^+(p) \setminus I^+(p)$, then any causal curve connecting p to q must be a null geodesics.*

The following lemma is also a consequence of Theorem 3.0.3.

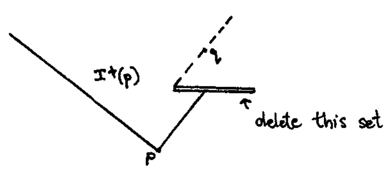
Lemma 3.0.5 *Let $A \subset M$.*

i) $intJ^+(A) = I^+(A)$,

ii) $J^+(A) \subset \overline{I^+(A)}$ with equality iff $J^+(A)$ is closed,

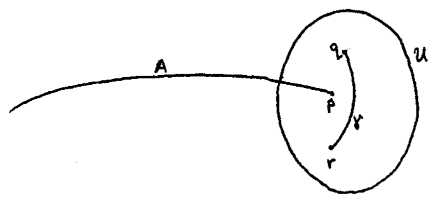
iii) $\partial J^+(A) = \partial I^+(A)$.

Example 3.0.6 Consider the spacetime $M = \mathbb{R}^n_1 \setminus \{p_1, \dots p_k\}$ and let $A \subset M$ be the single point $A = \{p\}$ as in the diagram below. While in Minkowski space $\overline{I^+(A)} = J^+(A)$, in this case $q \in \overline{I^+(A)}$ but $q \notin J^+(A)$. $I^+(A)$ is still an open subset of $J^+(A)$, but $J^+(A)$ is not closed.



A subset $A \subset M$ is called *achronal* if $I^+(A) \cap A = \emptyset$, i.e. for any two events $p, q \in A$, q cannot be in $I^+(p)$. Similarly, A is called *acausal* if $q \notin J^+(p)$, for $p \neq q$ in A. $I^+(A) \subset J^+(A)$, so an acausal set is an achronal one.

Definition 3.0.7 *Let A be a closed achronal set. The edge of A is the set of events $p \in \overline{A}$ such that for every open neighbourhood $\mathcal{U}$ of p there are points $q \in I^+(p,\mathcal{U})$ and $r \in I^-(p,\mathcal{U})$ and a timelike curve from r to q that misses A.*



A is a closed hypersurface if and only if edge$A \cap A = \emptyset$. This fact is not obvious (see [O] pages 413-414 for proof) but is useful in the following proposition. Define a *future set* F as a subset of M such that $I^+(F) \subset F$.

Proposition 3.0.8 *([O] 14.27) The (nonempty) boundary of a future set F (similarly a past set) is a closed achronal hypersurface.*

Proof. Let $p \in \partial F$. Pick $q \in I^+(p)$ such that $I^-(q)$ is a neighbourhood of p containing a point p' of F. By construction $q \in I^+(p')$ and consequently $q \in I^+(F) \subset F$ as well. Thus $I^+(p)$ must be contained in F. The complements of future sets are past sets, so $I^-(p) \subset M \setminus F$ also holds. Ergo, $I^+(\partial F) \cap I^-(\partial F) = \emptyset$ and ∂F is achronal. Clearly ∂F has no edge points since $I^+(p) \subset \text{int}F$ and $I^-(p)$ is contained in the exterior of F. □

3.1 Cauchy Hypersurfaces

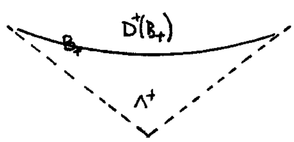
A curve is called *past-inextendible* if it is defined on the domain (a, b) but not on $(a - \delta, b)$ for $\delta > 0$. Inextendible curves are important in determining the causality and possible singularities of a spacetime M. For example they relay information about the predictability of the future of M. If A is an achronal set in M, the events predictable by A are the ones that can be reached by past-inextendible causal curves that cross A. This set of events corresponds to the *future Cauchy development* (or future domain of dependence) of A:

$$D^+(A) = \{p \in M : \text{ every past-inextendible causal curve through } p \text{ meets } A\}$$

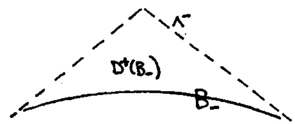
In particular, $A \subset D^+(A)$ and $D^+(A) \subset J^+(A)$. The *Cauchy development* (or full domain of dependence) of A is defined as $D(A) = D^+(A) \cup D^-(A)$.

Example 3.1.1 *1.* Consider the warped product $M = \mathbb{R}^1_1 \times F$ with metric $\mathbf{g} = -\mathbf{dt^2} + \mathbf{f(t)^2 g_0}$. The achronal spacelike subset $A = \{t = c :$

c is constant$\}$ has future Cauchy development $D^+(A) = J^+(A) = \{(t, x, y, z) : t \geq c\}$. Minkowski space, $\mathbb{R}^1_1 \times \mathbb{R}^3$, is one example of a spacetime with a metric of cohomogeneity one, and so has sets A and $D^+(A)$ such as these. Also in Minkowski space, the future nullcone $\Lambda^+ = \{(t, x, y, z) : t > (x^2+y^2+z^2)^{1/2}\}$ contains the subset $B_+ = \{(t, x, y, z) : -t^2+x^2+y^2+z^2 = -1,\ t > 0\}$ with future Cauchy development as depicted below (i). Similarly, the set $B_- = \{(t, x, y, z) : -t^2 + x^2 + y^2 + z^2 = -1,\ t < 0\}$ in the past nullcone has future Cauchy development as in (ii):

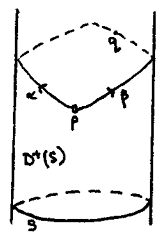


(i) B_+ $(t > 0)$



(ii) B_- $(t < 0)$

2. Let $M = \left(\mathbb{R}^1_1 \times S^1\right) \setminus \{p\}$, and S a spacelike circle as in the figure below. In this case there is a point $q \in \partial J^+(S)$ that does not have a past-endpoint in M. Thus $D^+(S) \neq J^+(S)$, and in fact, $D^+(S)$ is the union of S and the open region between S and the null geodesics α and β.

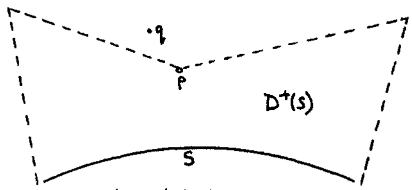


An achronal set $S \subset M$ for which $D(S) = M$ is called a *Cauchy hypersurface.* If $D(S) = M$ the entire future (resp. past) of M can be predicted by the events in S. An alternative definition for a Cauchy hypersurface is that S must be met exactly once by every inextendible timelike curve in M.

Exercise: Verify that these definitions are equivalent. *Hint* for one direction, derive a contradiction by constructing a past-inextendible timelike curve that misses S. Then construct an inextendible one.

3.2 Cauchy Horizons

When a subset is not a Cauchy hypersurface there may be points in the future (or past) of S which are not in $D^+(S)$ (or $D^-(S)$). Consider the manifold $M = \mathbb{R}^4_1 - \{p\}$ and let S be any subset of M that precedes p:



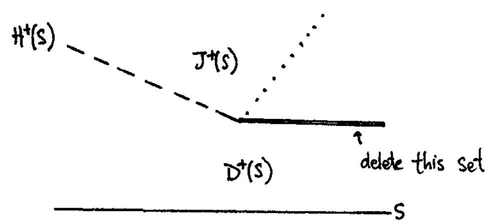
Then $q \in J^+(p)$ but $q \notin D^+(S)$. The future boundary of $D^+(S)$ is called the *future Cauchy horizon* of S.

Definition 3.2.1 *If S is an achronal set, its* future Cauchy horizon *is:*

$$H^+(S) = \overline{D^+}(S) \setminus I^-(D^+(S)) = \{p \in \overline{D^+}(S) : I^+(p) \text{ does not meet } D^+(S)\}$$

The *Cauchy horizon* of S is defined as $H(S) = H^+(S) \cup H^-(S)$.

Intuitively one can think of $H^+(S)$ as separating $D^+(S)$ from the rest of $J^+(S)$.



Example 3.2.2 Let $B_+ = \{(t,x,y,z) : -t^2 + x^2 + y^2 + z^2 = -1,\ t > 0\} \subset \Lambda^+$ and $B_- = \{(t,x,y,z) : -t^2 + x^2 + y^2 + z^2 = -1,\ t < 0\} \subset \Lambda^-$. Referring to diagram (i) in example 3.1.1(1), it is clear that $H^+(B_+) = \emptyset$ and $H^-(B_+) = \Lambda^+$. From diagram (ii) we infer that $H^+(B_-) = \Lambda^-$ and $H^-(B_-)$ is empty.

Lemma 3.2.3 *([O] 14.52) If A is a closed achronal set $\partial D^+(A) = A \cup H^+(A)$.*

Sketch of proof. Assume $p \in \partial D^+(A) \setminus A \setminus H^+(A)$. This implies p must also be in $I^+(A)$ (cf Lemma 1 below). It is now possible to find a point $q \in I^+(p) \cap D^+(A)$ such that $I^+(A) \cap I^-(q)$ is a neighbourhood of p contained in $D^+(A)$. This contradicts our choice of p. The reverse inclusion follows from the definition of $H^+(A)$.

Proposition 3.2.4 *([O] 14.53) Let S be a closed acausal hypersurface. Then*

1. $H^+(S) = I^+(S) \cap \partial D^+(S) = \overline{D^+}(S) \setminus D^+(S)$. *In particular,* $H^+(S) \cap S = \emptyset$.

2. *If $H^+(S) \neq \emptyset$ then $H^+(S)$ is a closed achronal hypersurface.*

3. *Starting at each point of $H^+(S)$ there is a past-inextendible null geodesic without conjugate points that is entirely contained in $H^+(S)$. (Future extended as far as possible in $H^+(S)$, such null geodesics are called* generators *of $H^+(S)$.)*

Proof. Part 1 is left as an exercise.
Part 2: *If $H^+(S) \neq \emptyset$ then $H^+(S)$ is a closed achronal hypersurface.*
By definition $H^+(S) = \overline{D^+(S)} \setminus I^-(\overline{D^+(S)}) = \overline{D^+(S)} \cup \{M \setminus I^-(D^+(S))\}$, so $H^+(S)$ is closed. For achronality, it suffices to show $I^-(H^+(S)) \cap H^+(S) = \emptyset$. This is clear as $I^-(H^+(S)) \subset I^-(\overline{D^+(S)}) = I^-(D^+(S)) \subset M \setminus H^+(S)$. Lastly, we need to show $H^+(S)$ is a hypersurface.
Claim: $P = D^+(S) \cup I^-(S)$ is a past set.
Consider $I^-(P) = I^-(D^+(S) \cup I^-(S)) = I^-(D^+(S)) \cup I^-(I^-(S))$. However, $I^-(S) = I^-(I^-(S))$ and $I^-(D^+(S)) = D^+(S) \setminus \partial D^+(S)$, so

$$I^-(P) = I^-(D^+(S)) \cup I^-(S) = [D^+(S) \setminus \partial D^+(S)] \cup I^+(S) \subset P.$$

By Proposition 3.0.8, ∂P is a closed achronal hypersurface. Since S is acausal $I^+(S) \cap I^-(S) = \emptyset$, and by part 1 $H^+(S) = I^+(S) \cap \partial D^+(S) = I^+(S) \cap \partial[D^+(S) \cup I^-(S)] = I^+(S) \cap \partial P$. This equality implies $H^+(S)$ is also a hypersurface.

Part 3: *Starting at each point of $H^+(S)$ there is a past-inextendible null geodesic without conjugate points that is entirely contained in $H^+(S)$.*
We begin by stating two lemmas, the first of which will be proved following this proposition.

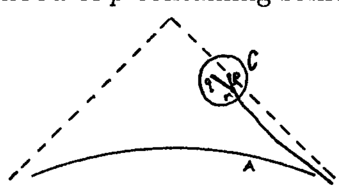
Lemma 1: *([O] 14.51) Let A be a closed achronal set. Then $\overline{D^+(A)} = \{p \in M$: every past-inextendible timelike curve through p meets $A\}$.*

Lemma 2: *([O] 14.30(2)) Let α be a past-inextendible causal curve starting at p that does not meet a closed set S. If α is not a conjugate-free null geodesic, there is a past-inextendible timelike curve starting at $\alpha(0)$ that does not meet S.*

By Part 1, if $p \in H^+(S)$ there exists a past-inextendible causal curve γ starting at p that does not meet S. Lemma 1 implies γ cannot be timelike and Lemma 2 asserts γ must be a conjugate-free null geodesic. What is left to show is that γ does not leave $H^+(S)$. Take $H^+(S) = \overline{D^+(S)} \setminus D^+(S)$ as in Part 1. If $\gamma(r) \in D^+(S)$ for some $r > 0$ then γ would have to meet S by definition of $D^+(S)$. On the other hand if $\gamma(s) \notin D^+(S)$ for some $s > 0$, there is a past-pointing past-inextendible timelike curve β starting at $\gamma(s)$

that avoids S. Applying Lemma 2 to the causal curve $\gamma|_{[0,s]} + \beta$ contradicts Lemma 1 since $\gamma(0) = p \in \overline{D^+}(S)$ by assumption. □

Proof of Lemma 1. $T \subset \overline{D^+(A)}$: Let $p \notin \overline{D^+(A)}$ and pick a point $r \in I^-(p, M \setminus \overline{D^+(A)})$. There is a past-inextendible causal curve that starts at r and misses A. By Lemma 2 of Proposition 3.2.4 there is a past-inextendible timelike curve through p that does not cross A, which implies $p \notin T$.
$\overline{D^+(A)} \subset T$: Assume there exists $p \in \overline{D^+}(A) \setminus T$. Let α be a past-inextendible timelike curve that starts at p and avoids A. Since $p \notin A$, p has a convex neighbourhood C which is disjoint from A. Let $\alpha(s) = r \in C$ be a point that chronologically precedes p ($r \ll p$). This implies $I^+(r, C)$ is an open neighbourhood of p containing some point $q \in D^+(A)$.



Denote the geodesic segment from q to r in C by σ_{qr}. The past-pointing timelike curve $\sigma_{qr} + \alpha|_{[s,\infty)}$ constitutes a past-inextendible timelike curve that does not meet A. This contradicts $q \in D^+(A)$. □

Problems:

1. Prove part 1 of Proposition 3.2.4. *Hint* use Lemma 3.2.3
2. Discuss the relevance of Theorem 2.2.2 to Lemma 2 above.
3. Prove Lemma 2 above.

3.3 Cauchy Hypersurfaces Revisited

The following proposition relates the predictability of future events of a submanifold with its future Cauchy horizon. This is relevant to the discussion of singularities since one of the conditions for a spacetime to be singular is that it possesses a future Cauchy hypersurface (as defined below).

Proposition 3.3.1 *Let S be a closed hypersurface in M. Then $H^+(S) = \emptyset$ if and only if $J^+(S) \subset D^+(S)$. In this case, S is called a* future Cauchy hypersurface.

Proof. Assume there is an event $p \in J^+(S) \setminus D^+(S)$. This is possible $\Leftrightarrow$ $\exists\, \gamma$ a causal curve segment and $r > 0$ s.t. $\gamma(r) \in \partial D^+(S) \setminus S \subset \partial D^+(S) \cap I^+(S)$. Thus, $p \in J^+(S) \setminus D^+(S) \Leftrightarrow \gamma(r) \in \partial D^+(S) \cap I^+(S) \Leftrightarrow H^+(S) \neq \emptyset$. □

Corollary 3.3.2 *A subset $S \subset M$ is a Cauchy hypersurface if and only if $H(S)$ is empty.*

Example 3.3.3 De Sitter spacetime is the warped product $M = \mathbb{R}^1_1 \times S^3$ with metric

$$ds^2 = -dt^2 + \alpha^2 \cosh^2(\alpha^{-1}t)\{dr^2 + \sin^2(d\theta^2 + \sin^2\theta\; d\phi^2)\}.$$

Referring to example 3.1.1(1), the hypersurface $S = \{t = c : c \text{ is constant}\}$ in M is Cauchy since $D^+(S) = J^+(S)$ (similarly $D^-(S) = J^-(S)$), and thus $D(S) = M$. Consequently,

$$H^+(S) = I^+(S) \cap \partial D^+(S) = I^+(S) \cap \partial J^+(S) = \emptyset.$$

The analogous property holds for $H^-(S)$, and therefore the Cauchy horizon $H(S) = H^+(S) \cup H^-(S) = \emptyset$.

Warped products have an additional property that relates the existence of a Cauchy hypersurface in the base to one in the spacetime.

Theorem 3.3.4 *([O] 14.33) Let $M = B \times_f F$ be a warped product with F complete. M has a Cauchy hypersurface if and only if B does.*

In Chapter 4 we will see that the two-dimensional base of the Schwarzschild spacetime M contains a Cauchy hypersurface (the curve $\{t = c\}$). By the above theorem, so will the exterior region of the black hole. We will in fact show more than this. By examining the Kruskal extension of M we can apply Theorem 3.3.4 to the entire spacetime, interior as well as exterior.

Chapter 4

Existence of Maximal Geodesic

In a Riemannian manifold two points contained in a convex set can be joined by a unique minimizing geodesic. In a Lorentz manifold timelike radial geodesics have the opposite feature of being the unique longest curves joining two points in a convex neighbourhood. A key ingredient in the proof of Hawking's singularity theorems is the existence of a maximal (or inextendible) timelike geodesic, which will be established in this chapter. Some interesting connections between the causality and geometry of a Lorentz manifold will be developed. Albeit the proofs are a bit technical, the examples and diagrams will help illustrate some of the important ideas. We begin by describing a tool that is used repeatedly throughout the section, most notably in Lemmas 4.1.2 and 4.4.2, and Theorem 4.2.6.

Definition 4.0.5 *Let* $\{\alpha_n\}$ *be an infinite sequence of future pointing causal curves in* M *and let* $\mathfrak{R}$ *be a convex covering of* M*, i.e. a covering by convex sets for which* $\mathcal{U}, \mathcal{V} \subset \mathfrak{R}$ *implies* $\mathcal{U} \cap \mathcal{V} \subset \mathfrak{R}$*. Define a* limit sequence *for* $\{\alpha_n\}$ *relative to* $\mathfrak{R}$ *as a (finite or infinite) sequence* $p = p_0 < p_1 < \ldots$ *in* M *satisfying:*
(L1) for each p_i *there exists a subsequence* $\{\alpha_m\}$ *and for all* m*, numbers* $s_{m_0} < s_{m_1} < \ldots < s_{m_i}$ *such that*

1. $\lim_{m\to\infty}(\alpha_m(s_{m_j})) = p_j$ *for each* $j \leq i$
2. *for each* $j < i$ *the points* p_j, p_{j+1} *and the curve segments* $\alpha_m|_{[s_{m_j}, s_{m_{j+1}}]}$ *for all* m *are contained in a single set* $C_j \in \mathfrak{R}$*.*

(L2) if $\{p_i\}$ *is infinite then it is nonconvergent. If* $\{p_i\}$ *is finite then it has more than one point, and no strictly longer sequence satisfies (L1).*

Proposition 4.0.6 *([O] 14.8) Let $\{\alpha_n\}$ be a sequence of future-pointing causal curves that satisfy the conditions:*

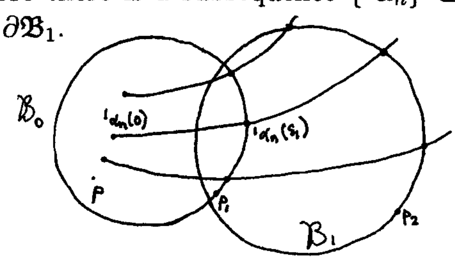
(i) $\{\alpha_n(0)\} \to p$

(ii) there is a neighbourhood of p that contains only finitely many of the curves α_n (use the notation $\{\alpha_n\} \nrightarrow p$).

Then with respect to any convex covering $\mathfrak{R}$, $\{\alpha_n\}$ has a limit sequence starting at p.

Proof. There are three parts to this proof:
1. Constructing $\{p_i\}$: Let $\mathfrak{R}'$ be a locally finite covering by open sets $\mathfrak{B}$ such that each $\bar{\mathfrak{B}}$ is compact and contained in a convex set in $\mathfrak{R}$. Pick $\mathfrak{B}_0 \subset \mathfrak{R}'$ such that infinitely many α_n's start in $\mathfrak{B}_0$ and leave $\bar{\mathfrak{B}}_0$. We can find a subsequence with the properties $\{^1\alpha_n(0)\} \to p$ and $\{^1\alpha_n(s_n)\} \to p_1$, for $p_1 \in \partial\mathfrak{B}_0$.

Now pick a set $\mathfrak{B}_1 \subset \mathfrak{R}'$ containing p_1. If infinitely many $^1\alpha_n' s$ leave $\mathfrak{B}_1$ then as before there is a subsequence $\{^2\alpha_n\} \subset \{^1\alpha_n\}$ that converges to a point p_2 on $\partial\mathfrak{B}_1$.



Repeat this process until we have a limit sequence of points $\{p_i\}$. Note that if there is more than one $\mathfrak{B}_i$ containing p_i we will pick the one used the fewest times before.

The condition (L1) of Definition 4.0.5 holds with C_i any element of $\mathfrak{R}$ that contains $\bar{\mathfrak{B}}_i$. It follows that $p_{i+1} > p_i$ since the relation $\leq$ is closed on C_i and by construction $p_{i+1} \neq p_i$.

2. If $\{p_i\}$ is infinite, we need to show it is nonconvergent. Assume $\{p_i\} \to q$ for some $q \in M$. Pick $\mathfrak{B} \in \mathfrak{R}'$ containing q, so that $p_i \in \mathfrak{B}$ for all but finitely many i. Since $\bar{\mathfrak{B}}$ is compact and $\mathfrak{R}'$ is locally finite, only finitely many elements of $\mathfrak{R}'$ meet $\mathfrak{B}$. Thus there must be some $\mathfrak{B}_i$ that was chosen with infinitely many p_i in its boundary by the construction in part 1. However $\mathfrak{B}$ contains almost all of the p_i's and with only finitely many in $\partial\mathfrak{B}$. This contradicts the way we chose p_i.

3. If $\{p_i\}$ is finite ($p = p_0 < p_1 < \ldots < p_k$), only finitely many ${}^k\alpha_n$ can leave $\mathfrak{B}_k$. Let $\{\alpha_m\} \subset \{{}^k\alpha_n\}$ be the subsequence of extendible curves trapped in $\mathfrak{B}_k$ (see exercise below). Now, assume

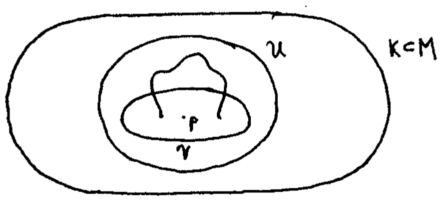
$$\alpha_m : [0, b_m] \longrightarrow M \text{ such that } \alpha_m(b_m) \to q \text{ for some } q \in \bar{\mathfrak{B}}_k$$

If $q = p_k$ then $p_0 < \ldots < p_k = q$ cannot be extended and still satisfy (L1). On the other hand, if $q \neq p_k$ then both (L1) and (L2) hold for $p_0 < \ldots < p_k < p_{k+1} = q$. $\square$

Exercise: Let α be a causal curve contained in a compact subset of a convex open set in M. Show α is extendible. *Hint* $\alpha : [0, B) \to M$ is extendible if for every $\{s_i\} \subset [0, B)$ with $s_i \to B$, there is a $q \in M$ such that $\alpha(s_i) \to q$.

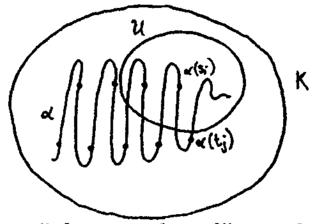
4.1 Back to the Future?

There are some properties of the causal structure of our universe (and any reasonable relativistic model for our universe) that are generally accepted. For instance traveling back and forth in time is something that perhaps only Marty McFly can do well. So, despite our dreams, it makes sense to consider noncompact spacetime manifolds which do not contain any closed timelike curves (such spacetimes are said to satisfy the *chronology condition*). However, we can demand even more, and consider spacetimes that satisfy not only the *causality condition* (no closed causal curves), but also the strong causality condition, i.e. manifolds for which there are no "almost closed" curves. The *strong causality condition* on a compact set $K \subset M$ says that for all $p \in K$ and any neighbourhood $\mathcal{U}$ of p, there exists a neighbourhood $\mathcal{V} \subset \mathcal{U}$ of p such that every causal curve segment with endpoints in $\mathcal{V}$ lies entirely in $\mathcal{U}$. So if a curve starts arbitrarily close to a point $p \in M$ and leaves a fixed neighbourhood of p it cannot return arbitrarily close to p.



Lemma 4.1.1 *([O] 14.13) Suppose the strong causality condition holds on a compact subset K of M. If α is a future-inextendible causal curve starting in K, then there is an $s > 0$ such that $\alpha(t) \notin K$ for all $t \geq s$. Namely, α will eventually leave K and never return.*

Proof. Assume otherwise, then if $\alpha : [0, B) \to M$ for $B \leq \infty$, there is a sequence $\{s_i\}$ in $[0, B)$ with $\{s_i\} \to B$ and $\{\alpha(s_i)\} \subset K$. K is compact, so we can find a subsequence $\{\alpha(s_j)\}$ that converges to a point $p \in K$. As α has no future endpoint it is possible to find another sequence $\{t_j\} \to B$ with $\{\alpha(t_j)\} \not\to p$. By taking a further subsequence we can suppose there is a neighbourhood $\mathcal{U}$ of p that contains none of the $\alpha(t_i)$. Since both $\{s_j\}$ and $\{t_j\}$ converge to B they have subsequences which alternate: $s_1 < t_1 < s_2 < t_2 \dots$



The curves $\alpha|_{[s_k, s_{k+1}]}$ are "almost closed", and this contradicts the strong causality of M at p. □

The following lemma will be used, among other things, to determine the length of a maximal geodesic. The proof relies on the notions of limit sequences and convergence of causal curves. If $\{p_j\}$ is a limit sequence for $\{\alpha_n\}$ let λ_j be the future pointing causal geodesics from p_j to p_{j+1} as in (L1). The future-pointing piecewise smooth causal curve $\lambda = \sum \lambda_j$ is called the *quasi-limit* of $\{\alpha_n\}$ with vertices p_j. A piecewise smooth curve whose segments are geodesics will be referred to as a *broken geodesic.*

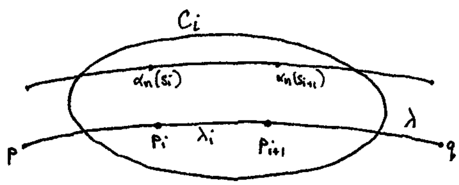
Lemma 4.1.2 *([O] 14.14) Suppose the strong causality condition holds on a compact subset K of M. Let $\{\alpha_n\}$ be a sequence of future pointing causal curve segments in K such that $\{\alpha_n(0)\} \to p$ and $\{\alpha_n(1)\} \to q \neq p$ for points $p, q \in K$. Then there is a future-pointing causal broken geodesic λ from p to q and a subsequence $\{\alpha_{n_j}\} \subset \{\alpha_n\}$ such that*

$$\lim_{n_j \to \infty} L(\alpha_{n_j}) \leq L(\lambda)$$

Proof. By Proposition 4.0.6, we can choose a convex covering so that $\{\alpha_n\}$ has a limit sequence $\{p_i\}$ starting at p. In the case $\{p_i\}$ is infinite, the corresponding quasi-limit λ is a future-inextendible causal curve starting at p. By Lemma 4.1.1 λ must leave K and never return. In particular, there is a vertex $p_i \notin K$, implying all of the $\{\alpha_n\}$'s must leave K and contradicting the hypotheses.

Therefore the limit sequence $\{p_i\}$ must be finite. It starts at p, ends at $\lim \alpha_{n_j}(1) = q$, and has quasi-limit λ. To simplify notation, write the index

n_j as m.



Consider a convex set C_i. The length of the i^{th} segment of α_m is at most the separation of its points in C_i:

$$L(\alpha|_{[s_{m_i},\ s_{m_{i+1}}]}) \leq |p_{m_i}p_{m_{i+1}}|, \quad \text{where } p_{m_i} = \alpha_m(s_{m_i}).$$

Hence,

$$L(\alpha_m) \leq L_m = \sum_{i=0}^{k} |p_{m_i}p_{m_{i+1}}|.$$

The vector pq and its norm $|pq|$ depend continuously on (p, q). Thus the sequence $\{L_m\}$ converges to $\sum |p_i p_{i+1}| = L(\lambda)$. Furthermore,

$$\lim_{m\to\infty} L(\alpha_m) \leq \lim_{m\to\infty} L_m = L(\lambda). \quad \square$$

4.2 Global Hyperbolicity

In addition to being strongly causal many reasonable models of our universe also appear to be globally hyperbolic. For every pair of events in the spacetime M that can be joined by a causal curve segment, there is a (maximal) causal geodesic joining them as well.

Definition 4.2.1 *M is* globally hyperbolic *provided*

(i) the strong causality condition holds

(ii) for each $p < q$, $J(p,q) = J^+(p) \cap J^-(q)$ is compact.

It can be proved that a spacetime is globally hyperbolic if and only if it contains a Cauchy hypersurface. Thus, the entire future and past history of a globally hyperbolic manifold is predictable. Definition 4.2.1 is motivated by the following proposition.

Proposition 4.2.2 *(see [O] 14.19) For $p < q$, if the set $J(p,q)$ is compact, and the strong causality condition holds on it, then there is a causal geodesic from p to q whose length is the supremum of the lengths of future-pointing causal curve segments from p to q.*

Example 4.2.3

1. Minkowski spacetime $\mathbb{R}^n_1$: is globally hyperbolic. This follows immediately from example 3.1.1(1), where we saw that $S = \{t = c\}$ is a Cauchy hypersurface.

2. Robertson-Walker spacetime $M = \mathbb{R} \times_a S$: the hypersurface $S(t_0) = \{(t_0, p) : p \in S\}$ with t_0 constant is a Cauchy hypersurface. Since this metric is of cohomogeneity one, $U = \frac{\partial}{\partial t}$ is the velocity vector for a timelike curve γ. U cannot equal zero, so γ will never reach an extreme point and change direction. Therefore, $S(t_0)$ will be met exactly once by each inextendible timelike curve. Global hyperbolicity also follows from the warped product structure of M.

3. Kruskal spacetime $K = B \times_r S^2$: Choosing the coordinates $\{T, X, \theta, \phi\}$ (cf ex. 1.5.3) the metric of K takes the form

$$ds^2 = \frac{32m^3}{r} \exp^{-r/2m}(-dT^2 + dX^2) + r^2(d\theta^2 + \sin^2\theta \, d\phi^2).$$

Claim: K is globally hyperbolic.

Consider the Kruskal plane B with metric $ds^2 = \frac{32m^3}{r} \exp^{-r/2m}(-dT^2 + dX^2)$. For $r > 0$, B is conformally equivalent to Minkowski space ($ds^2 = -dT^2 + dX^2$) and conformal metrics determine the same causal structure (since they have the same nullcone). Thus $\{T = c\}$ is Cauchy and B is globally hyperbolic. In $K = B \times_r S^2$, the spacelike hypersurface $\{T = c\} \times S^2$ is also Cauchy. This follows immediately from Theorem 3.3.4, the proof of which uses Lemma 4.1.1 to show that an inextendible curve in B must be inextendible in M as well.

Exercise: Let A be a spacelike hypersurface in the Schwarzschild exterior N. Show there exists an inextendible curve in the spacetime $N \cup B_L$ that does not meet A; B_L is the Schwarzschild interior (or black hole). Show that N and B_L are two globally hyperbolic spacetimes.

Global hyperbolicity is a strong condition to place on a spacetime and it is not required in the singularity theorems presented in this exposition. Instead we need only consider subsets of M that are globally hyperbolic. $\mathcal{H} \subset M$ is a *globally hyperbolic subset* provided *(i)* the strong causality condition holds on $\mathcal{H}$, and *(ii)* if $p, q \in \mathcal{H}$ with $p < q$ then $J(p, q)$ is compact and contained in $\mathcal{H}$. The subset we are interested in is the Cauchy development $D(S)$ of a hypersurface $S \subset M$. The typical approach is to show first that for S achronal, $\text{int}D(S)$ is globally hyperbolic. Then, using Lemma 4.2.5 below, one can prove the following:

Theorem 4.2.4 *([O] 14.43) If S is an acausal regular hypersurface in M, then $D(S)$ is open (hence globally hyperbolic).*

Lemma 4.2.5 *([O] 14.42) An achronal spacelike hypersurface S is acausal.*

Proof. Assume there exists a future-pointing causal curve segment α with endpoints $\alpha(0)$ and $\alpha(1)$ in S. If α is not a null geodesic then it admits a fixed endpoint deformation to a timelike curve – contradicting the achronality of S. On the other hand if α is a null geodesic, we can find a timelike curve arbitrarily near α because $\alpha'(0)$ is not normal to S. This also contradicts the achronality of S. □

For the purposes of this project, we need not show int$D(S)$ is globally hyperbolic and can thereby sidestep the proof of Theorem 4.2.4. Instead, we prove the following:

Theorem 4.2.6 *Let A be an acausal regular hypersurface of M. Then the Cauchy development $D(A)$ of A is globally hyperbolic.*

Lemma 4.2.7 *([O] 14.37) If A is acausal and $p \in D(A)$, then every inextendible causal curve through p meets both $I^-(A)$ and $I^+(A)$.*

Proof. By Lemma 1 of Proposition 3.2.4, $D^+(A) \subset A \cup I^+(A)$, so we can pick a point $p \in A \cup I^+(A)$. Let α be a past-inextendible causal curve starting at p with $\alpha(0) = p$, $\alpha(1) \ll p$, $\alpha(2) \ll p$ etc. We can find points $p_1, p_2, \ldots$ so that $\alpha(1) \ll p_1 \ll p$, $\alpha(2) \ll p_2 \ll p_1, \ldots$ By induction, for all $n \geq 1$, $\alpha(n) \ll p_n \ll p_{n-1}$. Joining each p_{n-1} to p_n by timelike segments will form a past-pointing causal curve β with $\beta(0) = p$.

In this construction, we can choose p_n close enough to $\alpha(n)$ that $\{p_n\}$ does not converge (i.e. $d(p_n, \alpha(n)) < \frac{1}{n}$). Ergo, β is an inextendible causal curve starting in $D(A) \cap I^+(A) \subset D^+(A)$ such that each $\beta(s)$ has a point of α in $I^-(\beta(s))$. Every past-inextendible causal curve starting in $D^+(A)$ must meet A, so β meets A and α must meet $I^-(A)$.

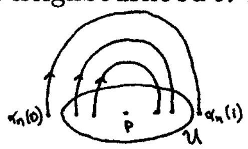
The same construction holds for past versions of these sets simply by reversing the time-orientation. The result follows. □

Proof of theorem 4.2.6. In order to show $D(A)$ satisfies the definition of a globally hyperbolic set, we break this proof up into proving four claims:

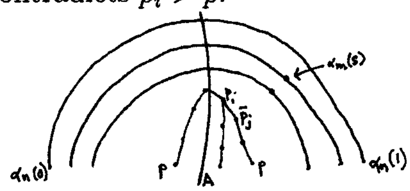
Claim 1: The causality condition holds on $D(A)$; there are no closed curves in $D(A)$.

Assume there exists a causal loop γ at $p \in D(A)$. Traversing γ repeatedly yields an inextendible causal curve $\tilde{\gamma}$, which must meet A. But $\tilde{\gamma}$ will meet A repeatedly and this contradicts the acausality of A.

Claim 2: The strong causality condition holds on $D(A)$.
Suppose in order to derive a contradiction, there exist future-pointing causal curve segments $\alpha_n : [0,1] \to M$ such that $\{\alpha_n(0)\} \to p$, $\{\alpha_n(1)\} \to p$, and every α_n leaves a fixed neighbourhood $\mathcal{U} \ni p$.

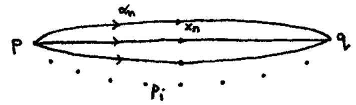


Let $\{p_i\}$ be the future-directed limit sequence of $\{\alpha_n\}$ starting at p. If $\{p_i\}$ is finite, the sequence ends at $\lim \alpha_n(1) = p$, implying $p < p$ and contradicting claim 1. Therefore $\{p_i\}$ must be infinite and the quasi-limit λ future inextendible. By Lemma 4.2.7, λ will enter and remain in $I^+(A)$, so there is a vertex $p_i \in I^+(A)$. There is a subsequence $\{\alpha_m\}$ with $\alpha_m(s) \in I^+(A)$ for $s \in [0,1]$ (after reparametrization) so that $\lim_{m\to\infty} \alpha_m(s) = p_i$. Since $p_i \neq p$, Proposition 4.0.6 applies and $\{\alpha_m|_{[s,1]}\}$ has a past-directed limit sequence $\{\bar{p_j}\}$ starting at p. If $\{\bar{p_j}\}$ is finite, it ends at $\lim \alpha_m(s) = p_i$, and as before this contradicts $p_i > p$.



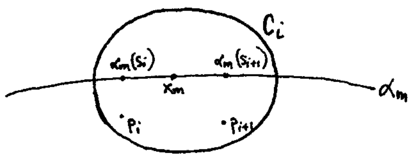
An infinite $\{\bar{p_j}\}$ has a past-inextendible quasi-limit $\bar{\lambda}$ that starts at p and meets $I^-(A)$. Consequently, for some $t \in [s,1]$ $\alpha_m(t)$ is in $I^-(A)$. This contradicts the acausality of A since α_m is future-pointing and $\alpha_m(s) \in I^+(A)$.

Claim 3: For $p \leq q$ in $D(A)$, $J(p,q)$ is compact.
When $p = q$, $J(p,q) = \{p\}$ and we are done by part 1. Assume $p < q$ and let $\{x_n\} \subset J(p,q)$. We need to find a subsequence of $\{x_n\}$ which converges to a point in $J(p,q)$.



Let α_n be a future-pointing causal curve segment from p to q through x_n. Let $\mathfrak{R}$ be a convex covering of M for which $C_i \subset \mathfrak{R}$, and $\bar{C_i}$ is compact and contained in a convex open set. Note, all limit sequences will be with respect to $\mathfrak{R}$.

Suppose $p = p_0 < p_1 < \ldots < p_k = q$ is a limit sequence and $\{\alpha_m\}$ a subsequence as in (L1) of Definition 4.0.5. There exists $i < k$ such that for infinitely many m the point x_m lies on the i^{th} segment $\alpha|_{[s_{m_i}, s_{m_{i+1}}]}$ of α_m.



By (L1) $\alpha|_{[s_{m_i}, s_{m_{i+1}}]}$ lies in C_i and hence so do the points x_m. Thus the sequence $\{x_m\}$ converges to a point $x \in C_i$. Since the relation $\leq$ is closed on C_i, $p_i \leq x \leq p_{i+1}$ implies $p \leq x \leq q$. Thus $x \in J(p,q)$.

Next we must derive a contradiction to the statement: every limit sequence for $\{\alpha_n\}$ starting at p is infinite. Let λ be the corresponding quasi-limit. We can find a subsequence $\{\alpha_m\}$ and (by reparametrizing) an $s \in [0,1]$ so that

$$\{\alpha_m(s)\} \to p_i \in I^+(A).$$

Since $p_i \neq q$ we can construct a past directed limit sequence $\{q_i\}$ for $\alpha_m|_{[s,1]}$ starting at q. If $\{q_i\}$ is finite it ends at $\lim \alpha_m(s) = p_i$, and $p < p_1 < \ldots < p_i < \ldots < q_1 < q$ is a finite limit sequence for $\{\alpha_n\}$ starting at p – contradicting the statement above.

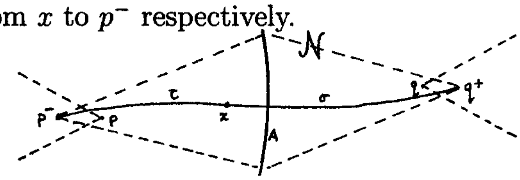
Hence, $\{q_i\}$ is an infinite sequence with quasi-limit μ. By Lemma 4.2.7 μ reaches $I^-(A)$, and so there must be some $q_i \in I^-(A)$. This implies that $\alpha_m(s) \in I^-(A)$ for some s. This contradicts the way we chose α_m and the acausality of A.

Claim 4: If $p \leq q \in D(A)$ then $J(p,q) \subset D(A)$.
Without loss of generality, assume $p < q$. There are two cases:

(i) $p, q \in I^+(A)$.
Pick a point $q^+ \in I^+(q) \cap D(A) \subset D^+(A)$. The set $\mathcal{N} = I^+(A) \cap I^-(q^+)$ is open and contains $J(p,q)$. Let σ be a past-pointing causal curve from q^+ to some point $y \in \mathcal{N}$. A is acausal and $y \in I^+(A)$, so σ does not meet A and $y \in D^+(A)$. Hence $\mathcal{N} \subset D(A)$.

(ii) $p \in J^-(A)$ and $q \in J^+(A)$.
Let $p^- \in I^-(p) \cap D^-(A)$ and $q^+ \in I^+(q) \cap D(A)$. The set $\mathcal{N} = I^+(p^-) \cap I^-(q^+)$ is an open neighbourhood of $J(p,q)$ contained in $D(A)$. If $x \in \mathcal{N}$, let σ and τ be past-pointing causal curve segments from q^+

to x and from x to p^- respectively.



Since $A \subset D(A)$ suppose $x \notin A$. The acausality of A implies at least one of the curves σ, τ does not meet A. If σ does not, $x \in D^+(A)$. If τ does not, $x \in D^-(A)$. In either case, $\mathcal{N} \subset D(A)$.

Therefore $D(A)$ satisfies the properties of a globally hyperbolic subset. □

4.3 Time Separation

Recall from the discussion in Chapter 1 that a timelike curve $\alpha : I \to M$ can be reparametrized by its own "proper time" τ. We can broaden this notion of time to describe the separation of two points in an arbitrary time-oriented Lorentz manifold.

Definition 4.3.1 *If $p, q \in M$, the* time separation $\tau(p,q)$ *from p to q is:*

$$\sup\{L(\alpha) : \alpha \text{ is a future-pointing causal curve segment from } p \text{ to } q\}.$$

$\tau(p,q) = \infty$ if the set of lengths is unbounded and $\tau(p,q) = 0$ if $q \notin J^+(p)$. If S is a subset of M, then $\tau(S,q) = \sup\{\tau(p,q) : p \in S\}$.

One can think of time separation as a dual to Riemannian distance; while d minimizes, τ maximizes. Thus τ satisfies the reverse triangle inequality,

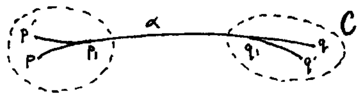
$$\tau(p,q) + \tau(q,r) \leq \tau(p,r) \quad \forall\, p \leq q \leq r.$$

The time separation $\tau(p,q)$ represents the proper time of a slowest trip in M from p to q.

Lemma 4.3.2 *([O] 14.17) The time separation function $\tau : M \times M \to [0,\infty]$ is lower semi-continuous.*

Proof. If $\tau(p,q) = 0$, $\sup\{L(\alpha)\} = \emptyset$ and there is nothing to prove. Suppose $q \in I^+(p)$ and $0 < \tau(p,q) < \infty$. Given $\delta > 0$ we want to find neighbourhoods $\mathcal{U}$ and $\mathcal{V}$ of p and q respectively, such that if $p' \in \mathcal{U}$ and $q' \in \mathcal{V}$, then $\tau(p',q') > \tau(p,q) - \delta$.

Let α be a timelike curve as in the following diagram:



$\mathcal{C}$ is a convex neighbourhood of q containing the point q_1, and $L(\alpha) > \tau(p,q) - \frac{\delta}{3}$.
Denote the geodesic segment from r to r' in $\mathcal{C}$ by $[r, r']$. Clearly, the length of $[r, r']$ depends continuously on the endpoints r and r'. Pick a neighbourhood $\mathcal{V}$ of q such that $[q_1, q]$ a causal curve for every $q' \in \mathcal{V}$ and where $L([q_1, q']) > L([q_1, q]) - \frac{\delta}{3}$. Since $[q_1, q]$ is a geodesic, $L([q_1, q])$ is greater or equal to the length of α between q_1 and q.
A corresponding construction at the endpoint p produces a similar neighbourhood $\mathcal{U}$ of p. The points $p' \in \mathcal{U}$ and $q' \in \mathcal{V}$ can be joined by a causal curve of length

$$L = \tau(p', q') > L(\alpha) - \frac{\delta}{3} > \tau(p,q) - \delta.$$

When $\tau(p,q) = \infty$ the same argument applies to show that for any $B > 0$ there are neighbourhoods as above such that $\tau(p', q') > B$. □

4.4 Maximal Geodesics

We can now use some of the tools developed so far to show that for a (realistic) spacetime it is always possible to find a timelike geodesic of maximal length.

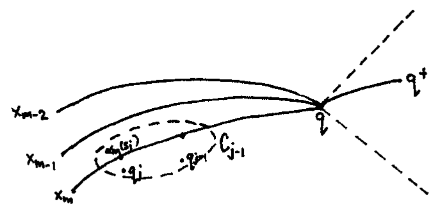
Theorem 4.4.1 *([O] 14.44) Let S be a closed achronal spacelike hypersurface in M. If $q \in D^+(S)$ then there exists a geodesic γ from S to q of length $\tau(S,q)$. Hence, γ is normal to S and has no focal points of S before q. (γ is timelike except in the trivial case $q \in S$).*

Outline of proof: The time separation function τ is continuous on the compact set of points $J^-(p) \cap D^+(S)$ and hence attains a maximum at a point $p \in S$. To determine a curve with this maximum length, find a sequence of causal curves whose lengths approach $\tau(p,q)$. The quasi-limit of this sequence will have length equal to $\tau(p,q)$.

Prior to the proof, we establish two important lemmas.

Lemma 4.4.2 *([O] 14.40) For S and q as in Theorem 4.4.1, the set $J^-(q) \cap D^+(S)$ is compact.*

Proof. If $q \in S$, there is nothing to show.
Suppose $q \in I^+(S) \cap D(S)$, and let $\{x_n\}$ be an infinite sequence in $J^-(p) \cap D^+(S)$. Let α_n be a past-pointing causal curve segment from q to x_n. If any subsequence of $\{x_n\}$ converges to q we are done. Otherwise, there is a past-directed limit sequence $\{q_i\}$ for $\{\alpha_n\}$ starting at q. If $\{q_i\}$ is infinite, it is nonconvergent by definition. Thus there must be some $x_n \in I^-(S)$, contradicting the choice of $\{x_n\}$. On the other hand, if $\{q_i\}$ is finite, we can find a subsequence $\{x_m\} \subset \{x_n\}$ converging to $x \in J^-(q)$.



Let σ be a timelike curve from $q^+ \in D^+(S) \cap I^+(q)$ to x. If σ meets S then either $x \in S \subset D^+(S)$ or $x \in I^-(S)$ (which implies there is an $x_n \in I^-(S)$). If σ avoids S, then $x \in D^+(S)$. □

Lemma 4.4.3 *([O] 14.21) If $\mathcal{U}$ is an open globally hyperbolic set and $q \in \mathcal{U}$, then the function $x \mapsto \tau(x, q)$ is continuous on $\mathcal{U}$.*

Proof. Previously we proved τ is lower semi-continuous, so it suffices to show that τ is upper semi-continuous. Suppose τ is not upper semi-continuous at the point $(p, q) \in \mathcal{U} \times \mathcal{U}$. Then there is a $\delta > 0$ and sequences $\{p_n\}$ and $\{q_n\}$ that converge to p and q respectively, with $\tau(p_n, q_n) \geq \tau(p, q) + \delta$. Since $\tau(p_n, q_n) > 0$ we can find a causal curve segment α_n with $L(\alpha_n) > \tau(p_n, q_n) - \frac{1}{n}$. $\mathcal{U}$ is open, so it contains points $p^- \ll p$ and $q^+ \gg q$ such that $\{p_n\} \subset I^+(p^-)$ and $\{q_n\} \subset I^-(q^+)$. Thus α_n is contained in the compact set $J(p^-, q^+)$. By Lemma 4.1.2 the quasi-limit λ satisfies

$$\lim_{m \to 0} L(\alpha_m) \leq L(\lambda).$$

Then $L(\alpha_n) > \tau(p_n, q_n) - \frac{1}{n} \geq \tau(p, q) + \delta - \frac{1}{n}$. Applying Lemma 4.1.2 again gives

$$L(\lambda) \geq lim_{n \to 0} L(\alpha_n) \geq \tau(p, q) + \delta,$$

contradicting the definition of τ. □

Remark 4.4.4 The argument in the proof of Lemma 4.4.3 is similar to the one required to prove Proposition 4.2.2. As an exercise, prove Proposition 4.2.2.

Proof of Theorem 4.4.1. Consider the set $J^-(q) \cap S$. By Lemma 4.4.2 $J^-(q) \cap S$ is compact. Lemma 4.4.3 applies to the globally hyperbolic set $D(S)$, asserting that $x \mapsto \tau(x, q)$ is continuous on $J^-(q) \cap S$ and takes a maximum value at a point $p \in S$. This maximum value is $\tau(p, q) = \tau(S, q)$. Let $\{\alpha_n\}$ be future pointing causal curve segments from p to q whose lengths approach $\tau(p, q)$. $\{\alpha_n\} \subset J(p, q)$ and by Lemma 4.1.2 there exists a broken geodesic γ from p to q with $L(\gamma) = \tau(p, q)$. If γ breaks, there is a longer causal curve from p to q, hence γ must be unbroken.

If $q \notin S$ then $p \ll q$ implies $L(\gamma) = \tau(p, q) > 0$. Thus γ is timelike (if γ were a null geodesic, there would be a fixed endpoint deformation of γ to a timelike curve with longer length).

To prove γ is normal to S look at the First Variation Formula for a fixed endpoint variation $\boldsymbol{x}$ of an unbroken geodesic:

$$L'(0) = \frac{-\epsilon}{c} \int_0^b \langle \gamma'', V \rangle \, du - \frac{\epsilon}{c} \sum \langle \Delta\gamma'(u_i), V_i \rangle + \frac{\epsilon}{c} \langle \gamma', V \rangle \Big|_0^b$$

Clearly $\gamma'' = 0$ and $\Delta\gamma'(u_i) = 0$. For variations with two fixed endpoints $V(0) = V(b) = 0$. Therefore $L'(0) = 0$. Now let $\widetilde{\boldsymbol{x}}$ be a fixed endpoint variation from S to q with variation vector field $J(t)$, where $J(0) \in T_p(S)$. Then $0 = L'(0) = \langle \gamma', J(0) \rangle$ and γ is orthogonal to S.
It follows from Theorem 2.1.1(2) that γ has no focal points. □

Example 4.4.5 *Timelike Radial Geodesics in Schwarzschild Spacetime.*
Recall the metric for Schwarzschild spacetime:

$$ds^2 = -\left(1 - \frac{2m}{r}\right)dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2\left(d\theta^2 + sin^2\theta d\phi^2\right).$$

In this example we parametrize an inextendible timelike radial geodesic with zero angular momentum by the affine parameter proper time and show that it is incomplete (inextendible with finite length). Relativistically, the example illustrates how a material particle will appear to a distant observer to take an infinite amount of time to reach the horizon ($r = 2m$) of the Schwarzschild black hole, while by its own proper time crossing the horizon and reaching the singularity ($r = 0$) in finite time.

Without loss of generality let $\theta = \pi/2$ and consider the geodesics that lie in this equatorial plane. A timelike geodesic parametrized by proper time satisfies the equation:

$$-\left(1 - \frac{2m}{r}\right)\left(\frac{dt}{d\tau}\right)^2 + \left(1 - \frac{2m}{r}\right)^{-1}\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 = -1. \qquad (4.1)$$

We can rewrite equation (4.1) by considering the constants of motion:

$$E = \left(1 - \frac{2m}{r}\right)\frac{dt}{d\tau} \quad \text{and} \quad L = r^2\frac{d\phi}{d\tau}$$

([W] p. 139). L is the angular momentum of a particle and by assumption is equal to zero. Substituting into equation (4.1) gives

$$\left(\frac{dr}{d\tau}\right)^2 = \frac{2m}{r} - \left(1 - E^2\right). \tag{4.2}$$

To determine an expression for τ we can rearrange (4.2) and integrate.

$$\begin{aligned}
\int d\tau &= \int \frac{dr}{\sqrt{\frac{2m}{r} - (1 - E^2)}} \\
&= \int \frac{rdr}{\sqrt{1 - E^2}\sqrt{\frac{2m}{1-E^2}r - r^2}} \\
&= \frac{1}{\sqrt{1 - E^2}} \int \frac{rdr}{\sqrt{(\frac{m}{1-E^2})^2 - (r - \frac{m}{1-E^2})^2}} .
\end{aligned}$$

To simplify the calculation, let $a = \frac{m}{1-E^2}$ and make the substitution $r - a = a\cos\psi$. The integral now becomes:

$$\begin{aligned}
\tau &= \frac{1}{\sqrt{1 - E^2}} \int \frac{a(1 + \cos\psi)}{\sqrt{a^2(1 - \cos^2\psi)}}(-a\sin\psi)d\psi \\
&= \frac{-1}{\sqrt{1 - E^2}} \int a(1 + \cos\psi)d\psi \\
&= \frac{a}{\sqrt{1 - E^2}}(-\psi - \sin\psi) \\
&= \frac{m}{(1 - E^2)\sqrt{1 - E^2}}\Big(- \psi + \sin(-\psi)\Big) \\
&= \frac{m}{(1 - E^2)^{3/2}}(\eta + \sin\eta) \qquad (4.3)
\end{aligned}$$

It is convenient to write τ and r in terms of η as in (4.3) and as in our substitution for $r = \frac{m}{1-E^2}(1 + \cos\eta) = \frac{m}{1-E^2}\cos^2\frac{1}{2}\eta$ (4.4) because it allows us to easily see what happens to τ as r approaches 0 and $2m$. Specifically, as $r \longrightarrow 2m$, $\eta \longrightarrow \eta_H = 2\sin^{-1}E$ and at $r \longrightarrow 0$, $\eta \longrightarrow \eta_0 = \pi$. This indicates the following effect on τ:

As

$$r \longrightarrow 2m, \quad \tau \longrightarrow \tau_H = \frac{m}{(1-E^2)^{3/2}}(\eta_H + \sin\eta_H) < \infty,$$

and as

$$r \longrightarrow 0, \quad \tau \longrightarrow \tau_0 = \frac{m}{(1-E^2)^{3/2}}\pi < \infty.$$

To see the effects on t as r approaches the horizon we would like to express t as a function of η. Taking similar steps to the above and using (4.4) we determine

$$\frac{dt}{d\tau} = \frac{E\cos^2\frac{1}{2}\eta}{\cos^2\frac{1}{2}\eta - \cos^2\frac{1}{2}\eta_H} \ .$$

Apply chain rule to get:

$$\frac{dt}{d\eta} = \frac{E\ (\cos^4\frac{1}{2}\eta)\ m}{(1-E^2)^{3/2}\Big(\cos^2\frac{1}{2}\eta - \cos^2\frac{1}{2}\eta_H\Big)} \ .$$

This can be integrated to yield

$$t = \frac{Em}{(1-E^2)^{3/2}}\left[\frac{1}{2}(\eta + \sin\eta) + (1-E^2)\eta\right] + 2m\ln\left[\frac{\tan\frac{1}{2}\eta_H + \tan\frac{1}{2}\eta}{tan\frac{1}{2}\eta_H - \tan\frac{1}{2}\eta}\right]$$

[C]. As $\eta \longrightarrow \eta_H$ one can see that $t \longrightarrow \infty$.

Chapter 5

Singularity Theorems

We now have the tools required to state and prove Hawking's singularity theorems. The main focus will be on the theorems which affirm the existence of incomplete timelike geodesics in a spacetime with nonnegative Ricci curvature. Two theorems by Penrose and Penrose-Hawking are presented (without proof) in an attempt to acquaint the reader with some of the earliest singularity theorems designed to show the existence of a black hole singularity. A theorem by Birkhoff is also mentioned. It shows that locally the Schwarzschild solution is the only C^2 solution of the Ricci flat Einstein field equation that is spherically symmetric.

5.1 Hawking's Theorems

Definition 5.1.1 *Let S be a spacelike hypersurface in M with future-pointing unit normal U and mean curvature vector field H. The* future convergence *k of S is the real-valued function*

$$k = \langle U, H \rangle = \frac{1}{n-1} \textit{ trace } II_S.$$

Remark Recall from example 1.5.1 the general evolution equation for a Robertson-Walker model of our universe: $3a''/a = -4\pi(\rho + 3\mathfrak{p})$. Provided $\rho > 0$ and $\mathfrak{p} \geq 0$ ($a'' < 0$), this equation predicts that the universe is not "standing still". Consequently, it is either always expanding ($a' > 0$) or contracting ($a' < 0$) (though there may be an exception at the instant of time when expansion changes to contraction). This expansion can be explained by the spacelike Cauchy hypersurface $S(t_0) = \{(t_0, p) : p \in S,\ t_0 \textit{ constant}\}$

having negative past convergence, i.e. $\boldsymbol{k} = \langle U, H \rangle < 0$ when U is past-pointing. These hypersurfaces bend outwards as time keeps on ticking into the future, and the distance between galaxies (which travel on timelike geodesics) is extended. Likewise, as one goes backwards in time and the rate of expansion gets faster ($a'' < 0, a' > 0$) the distance between galaxies is much less. Thus, general relativity predicts that at a finite time ago the universe was singular; the *big bang*. Some physical evidence for expansion comes from the observed cosmological redshift of light emitting from distant galaxies. At the time of emission, the wavelength of light is assumed to be the same for each galaxy. However as light travels long distances through space toward an earthbound observer the universe expands and the wavelengths elongate. The observed light is shifted toward the red end of the spectrum (hence the name *redshift*).

Theorem 5.1.2 *([O] 14.55A) Suppose $Ric(v,v) \geq 0$ for every timelike tangent vector to M. Let S be a spacelike future Cauchy hypersurface with future convergence $\boldsymbol{k} \geq b > 0$. Then every future-pointing timelike curve starting in S has length at most $\frac{1}{b}$.*

Proof. Consider the chronological future of S, $I^+(S)$. It is the set of points for which there is a future pointing timelike curve that starts in S. The objective is to show $I^+(S) \subset \{p \in M : \tau(S,p) \leq \frac{1}{b}\}$ since $\tau(S,p)$ by definition is the supremum of the length of all future pointing causal curves from S to p. By Theorems 4.4.1 and 2.1.2, $D^+(S) \subset \{p \in M : \tau(S,q) \leq \frac{1}{b}\}$. This follows as there is a normal timelike geodesic γ from S to any point $q \in D^+(S)$ that has no focal points before q and whose length $L(\gamma) = \tau(S,q)$. Furthermore, $L(\gamma) \leq \frac{1}{b}$ (otherwise there would be focal points along γ before q).

$I^+(S) \subset D^+(S)$ follows immediately from the fact that S is a future Cauchy hypersurface and hence satisfies $J^+(S) \subset D^+(S)$. Recall, $I^+(S) = \text{int}J^+(S)$. □

In some versions of Theorem 5.1.2, M is required to be globally hyperbolic. This places undesirable and stringent conditions on the global structure of M. While the hypothesis of a future Cauchy hypersurface is not as restrictive, it is a strong condition nonetheless. Hawking strengthened the above theorem by removing any condition on the global causality of M.

Theorem 5.1.3 *([H] p. 192) M cannot be timelike geodesically complete if:*

1. *The energy momentum tensor obeys the inequality,* $T_{ab}\omega^a\omega^b \geq \frac{1}{2}\omega_a\omega^a T$ *for any timelike vector* ω^a. *(This is satisfied by a perfect fluid of density* ρ *and pressure* $\mathfrak{p}$ *if* $\rho + \mathfrak{p} \geq 0$, $\rho + 3\mathfrak{p} \geq 0$.*)*

2. *There is a compact, imbedded, three-dimensional, spacelike submanifold S (i.e. S is a compact spacelike hypersurface without edges).*

3. *The contraction of the second fundamental form of S is either everywhere positive or everywhere negative. (This means that the unit normals to S are everywhere diverging or everywhere converging.)*

The version of Hawking's Singularity Theorem that we will prove has been refined slightly and appears in [O].

Theorem 5.1.4 *([O] 14.55B) Suppose* $Ric(v,v) \geq 0$ *for every* $v \in TM$. *Let* S *be a compact spacelike hypersurface with future convergence* $\boldsymbol{k} > 0$. *Then* M *is future timelike incomplete.*

Proof. We may suppose S is connected. Let $b > 0$ be the minimum of $\boldsymbol{k}$ on S.
Claim: There exists an inextendible future-pointing normal geodesic starting in S that has length $\leq \frac{1}{b}$.
Without loss of generality, assume that S is achronal. It is always possible to find a covering of Lorentz manifolds $\pi : \tilde{M} \to M$ such that $\tilde{S} \subset \tilde{M}$ is achronal and isometric to S under π. By Lemma 4.2.5 an achronal hypersurface is acausal and thus, as in Theorem 5.1.2,

$$D^+(S) \subset \{p \in M : \ \tau(S,p) \leq \frac{1}{b}\}.$$

There are two cases to consider: either (i) the future Cauchy horizon of S is empty, $H^+(S) = \emptyset$, or (ii) $H^+(S) \neq \emptyset$. If the former holds, S is a future Cauchy hypersurface and Theorem 5.1.2 applies. Consider case (ii) and assume for contradiction that M is timelike complete. We will need some lemmas.

Lemma 1: *If* $q \in H^+(S)$ *there exists a normal geodesic from* S *to* q *of length* $\tau(S,q) \leq \frac{1}{b}$.
Let $\{q_n\}$ be a sequence of events in $D^+(S)$ that converge to q. For each q_n there is a normal geodesic that starts in S and has length $\tau(S,q_n) \leq \frac{1}{b}$. Define the set B in the normal bundle NS of S as

$$B = \{v \in NS : v = 0 \text{ or } v \text{ is future pointing with } |v| \leq \frac{1}{b}\}.$$

It is a compact set since S is. The normal exponential map $\exp^{\perp} : NS \longrightarrow M$, which takes $v \mapsto \gamma_v(1)$, where γ_v is a geodesic with $\gamma_v'(0) = v$, acts on B in the following way: for every q_n there is a vector $v_n \in B$ such that $\exp^{\perp}(v_n) = q_n$. Since B is compact, $\{v_n\}$ converges to some vector $v \in B$, and by continuity, $\{q_n\}$ converges to $\exp^{\perp}(v) = \gamma_v(1)$. Thus $\gamma_v(1) = q$.

By construction, $\tau(S, q_n) = |v_n|$ converges to $|v| \leq \frac{1}{b}$. The map $p \mapsto \tau(S,p)$ is lower semi-continuous, so $|v| \geq \tau(S,q)$. In fact, by the premise that M is complete, the two are equal. The geodesic γ_v is defined on [0,1], and therefore $\tau(S,q) = L(\gamma_v) = |v| \leq \frac{1}{b}$.

Lemma 2: *$q \mapsto \tau(S,q)$ is strictly decreasing on past-pointing generators of $H^+(S)$. (cf Proposition 3.2.4)*

Let $\alpha : I \longrightarrow M$ be such a generator and let $s < t$ be in I. By Lemma 1 there exists a past-pointing timelike geodesic σ with $L(\sigma) = \tau(S, \alpha(t))$. α is null and the broken causal curve $\alpha|_{[s,\ t]} + \sigma$ can be lengthened by a small fixed endpoint deformation so that,

$$\tau(S, \alpha(s)) > L(\alpha|_{[s,\ t]} + \sigma) \geq L(\sigma) = \tau(S, \alpha(t)).$$

Now, returning to the proof of the theorem, we derive the contradiction. By assumption M is (timelike) complete, so $\exp^{\perp} : NS \longrightarrow M$ is defined on all of B. Since $H^+(S)$ is closed and contained in the continuous image of the compact set B, $H^+(S)$ is compact. When restricted to $H^+(S)$, the lower semi-continuous map $p \mapsto \tau(S,p)$ will take on a finite minimum at some point. This contradicts Lemma 2, since there is a generator extending pastward from each point of $H^+(S)$. Therefore γ_v is inextendible. □

Example 5.1.5 *The big bang*

Consider the Robertson-Walker model of a flat, dust-filled universe:

$$ds^2 = -dt^2 + (3t)^{\frac{4}{3}}\Big(dr^2 + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\Big).$$

This is the warped product $M = \mathbb{R} \times_a S$ with nonzero warping function $a = (3t)^{2/3}$. In Example 4.2.3(2), we saw that the hypersurface $S(t_0)$ is Cauchy, thus we can apply Theorem 5.1.2 to demonstrate a singular state in the spacetime M.

(i) $Ric(U,U) \geq 0$.
Let $U = -\partial_t$ be a past-pointing unit normal on $S(t_0)$. $Ric(U,U) = -3a''/a$ (cf ex. 1.5.1), which by the Remark above is non-negative.

Explicitly:

$$\begin{aligned} Ric(U,U) &= (-3)\frac{(-2/9)(3t)^{\frac{-4}{3}}}{(3t)^{\frac{2}{3}}} \\ &= \frac{2}{3}(3t)^{-2} \\ &> 0 \end{aligned}$$

(ii) U is past-pointing, so the past convergence $\boldsymbol{k} = \langle U, H \rangle$ should be negative.
Since $S(t_0)$ is a hypersurface, the mean normal curvature vector can be calculated from

$$H = \frac{1}{3}\sum_{i=1}^{3} II(e_i, e_i)$$

where e_1, e_2, e_3 is an orthonormal basis for $T_p(S)$. Recalling example 1.5.4, $II(X,Y) = \langle X, Y \rangle\,(a'/a)U$, for all $X, Y \in T_p(S)$. Thus,

$$\begin{aligned} II(e_i, e_i) &= \langle e_i, e_i \rangle \frac{(2/3)(3t)^{-1/3}}{(3t_0)^{2/3}}U \\ &= \frac{2}{9}(t_0^{-1})U. \end{aligned}$$

Inputting this into the formula for past convergence yields:

$$\begin{aligned} \boldsymbol{k} &= \left\langle -\partial_t, -\frac{1}{3}\sum_{i=1}^{3}(\frac{2}{9}\ t^{-1})\partial_t \right\rangle \\ &= (\frac{2}{9}\ t_0^{-1})\,\langle \partial_t, \partial_t \rangle \\ &= -(\frac{2}{9}\ t_0^{-1})\ <\ 0 \end{aligned}$$

Thus $t = 0$ is a physical singularity of M.

Example 5.1.6 *The Schwarzschild Black Hole*
In this application of Theorem 5.1.2 we consider the interior region ($r < 2m$) of the Schwarzschild black hole with metric

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta\ d\phi^2).$$

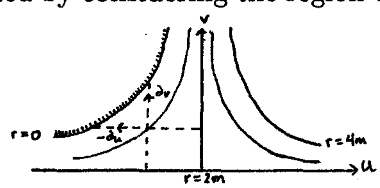
As one may infer from the metric when $0 < r < 2m$, ∂_r is a future-pointing timelike vector. Since the Schwarzschild model of a spacetime is Ricci flat, to apply Hawking's theorem we need to verify (i) $\Sigma = \{r = \text{constant}\}$ is a Cauchy hypersurface, and (ii) the future convergence $\boldsymbol{k}$ of Σ is positive. To do this we use the Kruskal null coordinates (u, v, θ, ϕ) for the extension $K = B \times_r S^2$. The metric takes the form null coordinates (u, v, θ, ϕ) for the extension $K = B \times_r S^2$. The metric takes the form

$$ds^2 = \frac{16m^2}{r}\, e^{1-r/2m}\, dudv + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

Recall from Chapter 1 the relationship between coordinates $(r - 2m)e^{r/2m-1} = uv$. The spacelike hypersurface Σ in the black hole (interior) region coincides with the hypersurface $\{uv = -c,\ c > 0\} \times S^2 \subset K$.

(i) $\{uv = -c,\ c > 0\} \times S^2$ is a Cauchy hypersurface.

By Theorem 3.3.4 it is enough to show that $\{uv = -c\}$ is Cauchy in B. This can be visualized by considering the region $v > 0$ in the uv-plane.



The curves $uv = -c$ form hyperbolas as depicted in the diagram except that $r = 2m$ gives the coordinate axes. Every particle (material or lightlike) in the interior region moves inward toward the singularity ([O] 13.30), and so every curve $\{uv = -c\}$ will be intersected by both of the future-pointing null vectors ∂_v and $-\partial_u$. Thus, $\{uv = -c\}$ is a Cauchy hypersurface by definition.

(ii) $\boldsymbol{k} = \langle H, \nu \rangle > 0$, where ν is the future-pointing unit normal on Σ. The orthonormal frame for Σ is given by

$$e_1 = \frac{1}{\sqrt{-2uvF}}(v\partial_v - u\partial_u), \quad e_2 = \frac{1}{r}\partial_\theta, \quad e_3 = \frac{1}{r\sin\theta}\partial_\phi$$

and the future-pointing unit normal is

$$\nu = \frac{1}{\sqrt{-2uvF}}(v\partial_v + u\partial_u).$$

In order to determine $\boldsymbol{k}$, we may compute

$$\boldsymbol{k} = \frac{1}{3}\sum_{i=1}^{3}\langle\overline{\nabla}_{e_i}e_i, \nu\rangle$$

Recalling example 1.5.4, we first consider

$$\begin{aligned}\langle\overline{\nabla}_{e_2}e_2, \nu\rangle &= \frac{-1}{r}\langle \mathrm{grad} r, \nu\rangle \\ &= \frac{-1}{4mr\sqrt{-2uvF}}\langle v\partial_v + u\partial_u, v\partial_v + u\partial_u\rangle \\ &= \frac{\sqrt{-2uvF}}{4mr}\end{aligned}$$

Similarly,

$$\begin{aligned}\langle\overline{\nabla}_{e_3}e_3, \nu\rangle &= \frac{-1}{r}\langle \mathrm{grad} r, \nu\rangle \\ &= \frac{\sqrt{-2uvF}}{4mr}\end{aligned}$$

Then,

$$\begin{aligned}\langle\overline{\nabla}_{e_1}e_1, \nu\rangle &= \frac{1}{(-2uvF)^{3/2}}\langle\overline{\nabla}_{v\partial_v - u\partial_u}(v\partial_v - u\partial_u), v\partial_v + u\partial_u\rangle \\ &= \frac{-1}{(-2uvF)^{3/2}}\langle v\partial_v - u\partial_u, \overline{\nabla}_{v\partial_v + u\partial_u}(v\partial_v - u\partial_u)\rangle \\ &= \frac{1}{(-2uvF)^{3/2}}\overline{\nabla}_{v\partial_v + u\partial_u}(uvF) \\ &= \frac{uv}{(-2uvF)^{3/2}}\Big(2F + v\frac{\partial F}{\partial v} + u\frac{\partial F}{\partial u}\Big)\end{aligned}$$

To compute $\frac{\partial F}{\partial v} + \frac{\partial F}{\partial u}$ recall from example 1.5.3, $F(u,v) = h(r) = \frac{8m^2}{r}e^{1-r/2m}$ and $uv = f(r) = (r-2m)e^{r/2m-1}$. Then $\frac{\partial F}{\partial v} = h'(r)\frac{\partial r}{\partial v} = u\frac{h'(r)}{f'(r)}$ and similarly for $\frac{\partial F}{\partial u}$. Substituting these into the above equality gives

$$\begin{aligned}\langle\overline{\nabla}_{e_1}e_1, \nu\rangle &= \frac{uv}{(-2uvF)^{3/2}}\Big(2F - 2uv\ \frac{h'(r)}{f'(r)}\Big) \\ &= \frac{1}{\sqrt{-2uvF}}\Big(-1 + \frac{uv}{r^2}\ e^{1-r/2m}\ (2m + r)\Big)\end{aligned}$$

Finally, the future convergence of Σ is:

$$\begin{aligned} k &= \frac{1}{3}\sum_{i=1}^{3}\langle \overline{\nabla}_{e_i} e_i, \nu\rangle \\ &= \frac{3m-2r}{3r^{3/2}\sqrt{2m-r}} \\ &> 0 \text{ provided } r < \frac{3m}{2}. \end{aligned}$$

Therefore for a small enough value of r it is possible to find a Cauchy hypersurface whose future convergence is positive. Hence, $r = 0$ is an actual singularity rather than a coordinate singularity as $r = 2m$ is.

5.2 Penrose-Hawking Theorems

The previous two theorems establish the existence of spacetime singularities in a broad cosmological context. The results by Penrose and Penrose-Hawking examine specifically the case of a black hole singularity. The reader is directed to [HE] for their proofs.

Penrose's theorem ascertains sufficient conditions for the existence of singularities in a context relevant to complete gravitational collapse without any assumption of symmetry. The criterion used by Penrose to generalize models such as Schwarzschild's is the possession of a *trapped surface* by a spacetime M.

Definition 5.2.1 *A spacelike submanifold of M is called a* trapped surface *provided its mean curvature vector field H is past-pointing.*

Lemma 5.2.2 *Let P be a spacelike submanifold of M with codimension ≥ 2. The following are equivalent:*

1. $k(v) = \langle H, v\rangle > 0$ *for all future-pointing null vectors v normal to P.*

2. $k(w) = \langle H, w\rangle > 0$ *for all future-pointing causal vectors w normal to P.*

3. *H is past-pointing timelike.*

Typically in spacetime models the trapped surface is S^2, though this need not be the case. A strictly causal definition of a trapped surface is given in [W]. Intuitively, it describes the surface as being in such a strong gravitational field that even the "outgoing" lightrays (null geodesics) are dragged back toward the singularity. Since nothing can travel faster than light, the material particles are also trapped and forced into surfaces of smaller and smaller area.

Theorem 5.2.3 *(Penrose) Let M be a globally hyperbolic spacetime with noncompact Cauchy hypersurface S. Suppose $Ric(v, v) \geq 0$ for all null vectors tangent to M. Suppose further that M contains a closed achronal trapped surface T with codimension 2. Then M is future null incomplete.*

The unwanted condition that M is globally hyperbolic can be eliminated with some additional assumptions as in the theorem by Penrose-Hawking below.

Theorem 5.2.4 *(Penrose-Hawking) Suppose a spacetime satisfies the following 4 conditions:*

1. *$Ric(v, v) \geq 0$ for all timelike and null vectors.*

2. *The generic condition is satisfied, i.e. every non-spacelike geodesic contains a point at which*

   $$U_a R_{bcde} U_f U^c U^d - U_b R_{acde} U_f U^c U^d + U_b R_{acdf} U_e U^c U^d - U_a R_{bcdf} U_e U^c U^d \neq 0$$

   where U is the tangent vector to the geodesic.

3. *The chronology condition holds on M (no closed timelike curves).*

4. *At least one of the following properties hold:*

   i) M contains a compact achronal set without edge,

   ii) M contains a closed trapped surface,

   iii) there existsp $\in M$ such that the future convergence of future-directed null geodesics emanating from p is positive.

Then M is timelike or null incomplete.

Remark. Items 4(i) and 4(ii) are a bit redundant given the definition of trapped surfaces used here. However the purely causal definition offered in [W] and [HE] seems to be the popular choice among physicists and it does not make any claims about future convergence. Either definition can be used to satisfy the theorem.

Example 5.2.5 *Trapped Surfaces in Schwarzschild Black Hole.*
For the future-pointing null vector $-v\partial_u$ the future convergence of S^2 is

$$\begin{aligned} k &= \left\langle -v\partial_u, \frac{-1}{4mr}(u\partial_u + v\partial_v) \right\rangle \\ &= \frac{v^2}{4mr} \langle \partial_u, \partial_v \rangle \\ &> 0. \end{aligned}$$

Therefore S^2 is a trapped surface in K.

Exercise: Verify that the Schwarzschild black hole satisfies the generic condition.

5.3 Birkhoff's Theorem

Schwarzschild's exact solution to Einstein's field equations was the first to demonstrate properties of a spherically symmetric spacetime. His work was elaborated on by Birkhoff, who in 1923 proposed that any Ricci flat spherically symmetric spacetime is locally equivalent to the Schwarzschild model. As stated in [HE],

Theorem 5.3.1 *(Birkhoff) Any C^2 solution of Einstein's empty space equations which is spherically symmetric in an open set $\mathcal{V}$, is locally equivalent to part of the maximally extended Schwarzschild solution in $\mathcal{V}$.*

The proof in [HE] considers an open neighbourhood of an event in an arbitrary spherically symmetric spacetime with metric of the form

$$ds^2 = \frac{-dt^2}{F^2(t,r)} + X^2(t,r)dr^2 + Y^2(t,r)(d\theta^2 + \sin^2\theta\, d\phi^2).$$

For this open set, one can write Einstein's field equations as a system of differential equations in F, X, and Y, and then for a Ricci flat solution, derive the Schwarzschild metric as in (5.1.6).

Exercise: A stationary Ricci flat solution to Einstein's field equation need not be spherically symmetric. The Kerr metric

$$\begin{aligned} ds^2 = & - (\frac{\Omega - a^2\sin^2\theta}{\Sigma})dt^2 - \frac{2a\sin^2\theta(r^2+a^2-\Omega)}{\Sigma}dtd\phi \\ & + \left[\frac{(r^2+a^2)^2 - \Omega a^2\sin^2\theta}{\Sigma}\right]\sin^2\theta\, d\phi^2 + \frac{\Sigma}{\Omega}dr^2 + \Sigma d\theta^2 \quad \text{[W 12.3.1]} \end{aligned}$$

$$\begin{aligned} \text{where } \Sigma &= r^2 + a^2\cos^2\theta \\ \Omega &= r^2 + a^2 + e^2 - 2mr \end{aligned}$$

and e, a, and m are parameters, is the only known axisymmetric, stationary solution of Einstein's vacuum field equation.

(i) Show that the Kerr metric reduces to the Schwarzschild solution when $e = a = 0$.

(ii) Determine when the singularities of the Kerr solutions exist and whether or not they are true physical singularities. *Hint* use the curvature invariant $R_{ijkl}R^{ijkl}$.

(iii) Determine the trapped surface of the Kerr black hole. *Hint* a change in coordinates may be helpful, see [HE].

Bibliography

[BE] Beem, J. K., K. L. Easley, and P. E. Ehrlich. *Global Lorentzian Geometry*, Second Edition, Dekker, New York, 1996.

[C] Chandrasekhar, S. *The Mathematical Theory of Black Holes*, Claredon Press, Oxford, 1983.

[F] Faber, R. L. *Differential Geometry and Relativity Theory; An Introduction*, Dekker, New York, 1983.

[Fr] Frankel, T. *Gravitational Curvature; An Introduction to Einstein's Theory*, W. H. Freeman and Company, San Fransico, 1979.

[H] Hawking, S. W. *The occurrence of singularities in cosmology. III. Causality and singularities,* Proc. Roy. Soc. Lond., 300 (1967), 187-201.

[HE] Hawking, S. W. and G. F. R. Ellis. *The Large Scale Structure of Space-Time*, Cambridge Univ. Press, London and New York, 1973.

[HT] Hughston, L. P. and K. P. Tod. *An Introduction to General Relativity*, Cambridge Univ. Press, Cambridge and New York, 1990.

[K] Kruskal, M. D. *Maximal extension of Schwarzschild metric,* Phys. Rev., 119 (1960), 1743-1745.

[L] Lee, J. M. *Riemannian Manifolds; An Introduction to Curvature*, Springer-Verlag, New York, 1997.

[N] Naber, G. L. *Spacetime and Singularities; An Introduction*, Cambridge Univ. Press, Cambridge and New York, 1988.

[O] O'Neill, B. *Semi-Riemannian Geometry with Applications to Relativity*, Academic Press, San Diego and London, 1983.

[P] Penrose, R. *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. 14 (1965), 57-59.

[SW] Sachs, R. K. and H. Wu. *General Relativity for Mathematicians*, Springer-Verlag, New York and Berlin, 1977.

[S1] Spivak, M. *Calculus on Manifolds*, W. A. Benjamin, Inc., New York, 1965.

[S2] Spivak, M. *A Comprehensive Introduction to Differential Geometry*, Vol. I, Publish or Perish, Huston, Texas, 1999.

[T] Torretti, R. *Relativity and Geometry*, Pergamon Press, Oxford and New York, 1983.

[W] Wald, R. M. *General Relativity*, The Univ. of Chicago Press, Chicago, 1984.