Spacetime Geometry

Beijing International Mathematics Research Center

2007 Summer School*

Gregory J. Galloway
Department of Mathematics
University of Miami

*Notes last modified: October 30, 2008. Please send typos to: galloway@math.miami.edu

Contents

1 Pseudo-Riemannian Geometry

We begin with a brief introduction to pseudo-Riemmanian geometry.

1.1 Manifolds

Let M^n be a smooth n-dimensional manifold. Hence, M^n is a topological space (Hausdorff, second countable), together with a collection of coordinate charts $(U, x^i) = (U, x^1, ..., x^n)$ (U open in M) covering M such that on overlapping charts (U, x^i), (V, y^i), $U \cap V \neq \emptyset$, the coordinates are smoothly related

$$y^i = f^i(x^1, ..., x^n), \quad f^i \in C^\infty ,$$

$i = 1, ..., n$.

For any $p \in M$, let T_pM denote the tangent space of M at p. Thus, T_pM is the collection of tangent vectors to M at p. Formally, each tangent vector $X \in T_pM$ is a *derivation* acting on real valued functions f, defined and smooth in a neighborhood of p. Hence, for $X \in T_pM$, $X(f) \in \mathbb{R}$ represents the directional deriviative of f at p in the direction X.

If p is in the chart (U, x^i) then the coordinate vectors based at p,

$$\frac{\partial}{\partial x^1}|_p, \frac{\partial}{\partial x^2}|_p, ..., \frac{\partial}{\partial x^n}|_p$$

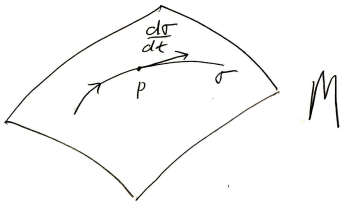
form a basis for T_pM. I.e., each vector $X \in T_pM$ can be expressed uniquely as,

$$X = X^i \frac{\partial}{\partial x^i}|_p , \quad X^i \in \mathbb{R}.$$

Here we have used the Einstein summation convention: If, in a coordinate chart, an index appears repeated, once up and once down, then summation over that index is implied.

Note: We will sometimes use the shorthand: $\partial_i = \frac{\partial}{\partial x^i}$.

Example. Tangent vectors to curves. Let $\sigma : I \to M$, $t \to \sigma(t)$, be a smooth curve in M.



The tangent vector to σ at $p = \sigma(t_0)$, denoted $\frac{d\sigma}{dt}(t_0) = \sigma'(t_0) \in T_pM$ is the derivation defined by,

$$\sigma'(t_0)(f) = \frac{d}{dt} f \circ \sigma(t)|_{t_0}$$

Suppose p lies in the coordinate chart (U, x^i). Then near p, σ can be expressed in terms of coordinate functions,

$$\sigma : x^i = x^i(t), \quad i = 1, ..., n.$$

Then, the chain rule implies,

$$\frac{d\sigma}{dt}(t_0) = \frac{dx^i}{dt}(t_0)\frac{\partial}{\partial x^i}|_p,$$

i.e., $\frac{dx^i}{dt}(t_0)$ are the components of $\frac{d\sigma}{dt}(t_0)$. In fact every vector $X \in T_pM$ can be expressed as the tangent vector to some curve through p.

The *tangent bundle* of M, denoted TM is, as a set, the collection of all tangent vectors,

$$TM = \bigcup_{p \in M} T_pM.$$

To each vector $V \in TM$, there is a natural way to assign to it $2n$ coordinates,

$$V \sim (x^1, ..., x^n, V^1, ..., V^n),$$

where $(x^1, ..., x^n)$ are the coordinates of the point p at which V is based, and $(V^1, ..., V^n)$ are the components of V wrt the coordinate basis vectors $\frac{\partial}{\partial x^1}|_p, ...\frac{\partial}{\partial x^n}|_p$. By this correspondence one sees that TM forms in a natural way a smooth manifold of dimension $2n$. Moreover, with respect to this manifold structure, the natural projection map $\pi : TM \to M$, $V_p \to p$, is smooth.

1.2 Vector fields

A vector field X on M is an assignment to each $p \in M$ of a vector $X_p \in T_pM$,

$$p \in M \xrightarrow{X} X_p \in T_pM.$$

If (U, x^i) is a coordinate chart on M then for each $p \in U$ we have

$$X_p = X^i(p)\frac{\partial}{\partial x^i}|_p.$$

This defines n functions $X^i : U \to \mathbb{R}$, $i = 1, ..., n$, the *components* of X on (U, x^i). If for a set of charts (U, x^i) covering M the components X^i are smooth ($X^i \in C^\infty(U)$) then we say that X is a *smooth* vector field.

Let $\mathfrak{X}(M)$ denote the set of smooth vector fields on M. Vector fields can be added pointwise and multiplied by functions; for $X, Y \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$,

$$(X + Y)_p = X_p + Y_p, \qquad (fX)_p = f(p)X_p.$$

From these operations we see that $\mathfrak{X}(M)$ is a module over $C^\infty(M)$.

Given $X \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, X acts on f to produce a function $X(f) \in C^\infty(M)$, defined by,

$$X(f)(p) = X_p(f).$$

With respect to a coordinate chart (U, x^i), $X(f)$ is given by,

$$X(f) = X^i \frac{\partial f}{\partial x^i}.$$

Thus, a smooth vector field $X \in \mathfrak{X}(M)$ may be viewed as a map

$$X : C^\infty(M) \to C^\infty(M), \quad f \to X(f)$$

that satisfies,

(1) $X(af + bg) = aX(f) + bX(g) \quad (a, b \in \mathbb{R})$,

(2) $X(fg) = X(f)g + fX(g)$.

Indeed, these properties completely characterize smooth vector fields.

Given $X, Y \in \mathfrak{X}(M)$, the *Lie bracket* $[X, Y]$ of X and Y is the vector field defined by

$$[X, Y] : C^\infty(M) \to C^\infty(M), \qquad [X, Y] = XY - YX,$$

i.e.

$$[X, Y](f) = X(Y(f)) - Y(X(f)).$$

In local coordinates one sees that the second derivatives cancel out.

Exercise. Show that with respect to a coordinate chart, $[X, Y]$ is given by

$$\begin{aligned}[X, Y] &= (X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}) \frac{\partial}{\partial x^j} \\ &= (X(Y^j) - Y(X^j)) \frac{\partial}{\partial x^j}.\end{aligned}$$

It is clear from the definition that the Lie bracket is skew-symmetric,

$$[X, Y] = -[Y, X].$$

In addition, the Lie bracket is linear in each slot over the *reals*, and satisfies,

(1) For all $f, g \in C^\infty(M)$, $X, Y \in \mathfrak{X}(M)$,

$$[fX, gY] = fg[X, Y] + fX(g)Y - gY(f)X.$$

(2) (Jacobi identity) For all $X, Y, Z \in \mathfrak{X}(M)$,

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0.$$

Exercise. Prove (1) and (2).

1.3 Co-vectors and 1-forms

A *co-vector* ω at $p \in M$ is a linear functional $\omega : T_pM \to \mathbb{R}$ on the tangent space at p. A *1-form* on M is an assignment to each $p \in M$ of a co-vector ω_p at p, $p \to \omega_p$. A *1-form* ω is *smooth* provided for each $X \in \mathfrak{X}(M)$, the function $\omega(X)$, $p \to \omega_p(X_p)$, is smooth. Equivalently, ω is smooth provided for each chart (U, x^i) in a collection of charts covering M, the function $\omega(\frac{\partial}{\partial x^i})$ is smooth on U, $i = 1, ..., n$.

Given $f \in C^\infty(M)$, the differential df is the smooth 1-form defined by

$$df(X) = X(f)\,, \qquad X \in \mathfrak{X}(M).$$

In a coordinate chart (U, x^i), df is given by,

$$df = \frac{\partial f}{\partial x^i} dx^i,$$

where dx^i is the differential of the i^{th}coordinate function on U.

Note: At each $p \in U$, $\{dx^1, ..., dx^n\}$ is the dual basis to the basis of coordinate vectors $\{\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}\}$.

1.4 Pseudo-Riemannian manifolds

Let V be an n-dimensional vector space over $\mathbb{R}$. A symmetric bilinear form $b : V \times V \to \mathbb{R}$ is

(1) positive definite provided $b(v, v) > 0$ for all $v \neq 0$,

(2) nondegenerate provided for each $v \neq 0$, there exists $w \in V$ such that $b(v, w) \neq 0$ (i.e., the only vector orthogonal to all vectors is the zero vector).

Note: 'Positive definite' implies 'nondegenerate'.

A *scalar product* on V is a nondegenerate symmetric bilinear form $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$. A scalar product space is a vector space V equipped with a scalar product $\langle\ ,\ \rangle$.

Let V be a scalar product space. An orthonormal basis for V is a basis $\{e_1, ..., e_n\}$ satisfying,

$$\langle e_i, e_j \rangle = \begin{cases} 0, & i \neq j \\ \pm 1, & i = j\,, \end{cases}$$

or in terms of the Kronecker delta,

$$\langle e_i, e_j \rangle = \varepsilon_i \delta_{ij} \quad \text{(no sum)}$$

where $\varepsilon_i = \pm 1$, $i = 1, ..., n$.

Fact. *Every scalar product space $(V, \langle\ ,\ \rangle)$ admits an orthonormal basis.*

The *signature* of an orthonormal basis is the n-tuple $(\varepsilon_1, \varepsilon_2, ..., \varepsilon_n)$. It is customary to order the basis so that the minus signs come first. The *index* of the scalar product

space is the number of minus signs in the signature. It can be shown that the index is well-defined, i.e., does not depend on the choice of basis. The cases of most importance are the case of index 0 and index 1, which lead to Riemannian geometry and Lorentzian geometry, respectively.

Definition 1.1. *Let M^n be a smooth manifold. A pseudo-Riemannian metric $\langle\ ,\ \rangle$ on a M is a smooth assignment to each $p \in M$ of a scalar product $\langle\ ,\ \rangle_p$ on T_pM,*

$$p \xrightarrow{\langle\ ,\ \rangle} \langle\ ,\ \rangle_p : T_pM \times T_pM \to \mathbb{R}.$$

such that the index of $\langle\ ,\ \rangle_p$ is the same for all p.

By 'smooth assignment' we mean that for all $X, Y \in \mathfrak{X}(M)$, the function $\langle X, Y\rangle$, $p \to \langle X_p, Y_p\rangle_P$, is smooth.
Note: We shall also use the letter g to denote the metric, $g = \langle\ ,\ \rangle$.

Definition 1.2. *A pseudo-Riemannian manifold is a manifold M^n equipped with a pseudo-Riemannian metric $\langle\ ,\ \rangle$. If $\langle\ ,\ \rangle$ has index 0 then M is called a Riemannian manifold. If $\langle\ ,\ \rangle$ has index 1 then M is called a Lorentzian manifold.*

If (U, x^i) is a coordinate chart then the *metric components* g_{ij} are the functions on U defined by,

$$g_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle, \qquad i, j = 1, ..., n\,.$$

If X, Y are vectors at some point in U then, by bilinearity,

$$\langle X, Y\rangle = g_{ij} X^i Y^j\,.$$

Thus, the metric components completely determine the metric on U.
Note: The metric $\langle\ ,\ \rangle$ is smooth iff for each chart (U, x^i), the g_{ij}'s are smooth.

Classically, one displays the metric components as

$$ds^2 = g_{ij} dx^i dx^j\,.$$

Ex. Euclidean space $\mathbb{E}^n$ as a Riemannian manifold. We equip $\mathbb{R}^n$ with the *Euclidean metric*: Let $(x^1, ..., x^n)$ be Cartesian coordinates on $\mathbb{R}^n$. Then for $X, Y \in T_p\mathbb{R}^n$,

$$X = X^i \frac{\partial}{\partial x^i}|p$$
$$Y = Y^i \frac{\partial}{\partial x^i}|p\,,$$

we have

$$\begin{aligned}\langle X, Y\rangle &= X \cdot Y \\ &= \sum_{i=1}^{n} X^i Y^i \\ &= \delta_{ij} X^i Y^j \,,\end{aligned}$$

where $\delta_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle$ is the Kronecker delta.

Ex. Minkowski space $\mathbb{M}^{n+1}$. This is the Lorentzian analogue of Euclidean space. We equip $\mathbb{R}^{n+1}$ with the Minkowski metric: Let $(x^0, x^1, ..., x^n)$ be Cartesian coordinates on $\mathbb{R}^{n+1}$. Then for $X, Y \in T_p\mathbb{R}^{n+1}$,

$$X = X^i \frac{\partial}{\partial x^i}|p \,, \qquad Y = Y^i \frac{\partial}{\partial x^i}|p \,,$$

we define,

$$\begin{aligned}\langle X, Y\rangle &= -X^0 Y^0 + \sum_{i=1}^{n} X^i Y^i \\ &= \eta_{ij} X^i Y^j \,,\end{aligned}$$

where $\eta_{ij} = \varepsilon_i \delta_{ij}$, and $(\varepsilon_0, \varepsilon_1, ..., \varepsilon_n) = (-1, 1, ..., 1)$.

1.5 Linear connections

We introduce the notion of covariant differentiation, which formalizes the process of computing the directional derivative of vector fields.

Definition 1.3. *A linear connection ∇ on a manifold M is an $\mathbb{R}$-bilinear map,*

$$\begin{aligned}\nabla : &\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M) \\ &(X, Y) \to \nabla_X Y\end{aligned}$$

satisfying for all $X, Y \in \mathfrak{X}(M)$, $f \in C^\infty(M)$,

(1) $\nabla_{fX} Y = f \nabla_X Y$,

(2) $\nabla_X fY = X(f)Y + f\nabla_X Y$.

$\nabla_X Y$ is called the *covariant* derivative of Y wrt X. It can be shown that for any $p \in M$, $\nabla_X Y|p$ depends only on the values of Y in a neighborhood of p and the value of X just at p. In particular, it makes sense to write $\nabla_X Y|_p$ as $\nabla_{X_p} Y$. This can be thought of as the directional derivative of Y at p in the direction of X_p.

In a coordinate chart (U, x^i) we introduce the *connection coefficients* Γ^k_{ij}, $1 \leq i, j, k \leq n$, which are smooth functions on U defined by,

$$\nabla_{\partial_i} \partial_j = \Gamma^k_{ij} \partial_k \,,$$

where, recall, $\partial_i = \frac{\partial}{\partial x^i}$.

Exercise. Show that with respect to a coordinate chart (U, x^i), $\nabla_X Y$ can be expressed as,

$$\nabla_X Y = (X(Y^k) + \Gamma^k_{ij} X^i Y^j) \partial_k \,, \tag{1.1}$$

where X^i, Y^i are the components of X and Y, respectively, wrt the coordinate basis ∂_i.

Note that this coordinate expression can also be written as,

$$\nabla_X Y = X^i Y^k{}_{;i} \, \partial_k$$

where we have introduced the classical notation,

$$Y^k{}_{;i} = \partial_i Y^k + \Gamma^k_{ij} Y^j \,.$$

1.6 The Levi-Civita connection

Definition 1.4. *A linear connection ∇ on M is symmetric provided for all $X, Y \in \mathfrak{X}(M)$,*

$$[X, Y] = \nabla_X Y - \nabla_Y X \,.$$

Using the coordinate expression (1.1) for $\nabla_X Y$, one easily checks that a linear connection ∇ is symmetric iff wrt each each coordinate chart, the connection coefficients satisfy,

$$\Gamma^k_{ij} = \Gamma^k_{ji} \,, \qquad \text{for } 1 \leq i, j, k \leq n \,.$$

Definition 1.5. *Let $(M, \langle \, , \, \rangle)$ be a pseudo-Riemannian manifold, and let ∇ be a linear connection on M. We say that ∇ is compatible with the metric provided for all $X, Y \in \mathfrak{X}(M)$,*

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle \,,$$

i.e., the metric product rule holds.

Remark: The standard linear connection on Euclidean space (and on Minkowski space) is symmetric and compatible with the metric.

Theorem 1.1 (Fundamental theorem of pseudo-Riemannian geometry)**.** *On a pseudo-Riemannian manifold there exists a unique linear connection ∇ that is symmetric and compatible with the metric.*

Comment on the proof. Using the symmetry and compatibility with the metric of the connection, one derives the *Kozul formula*,

$$\begin{aligned}\langle \nabla_X Y, Z\rangle = \frac{1}{2}[&X\langle Y,Z\rangle + Y\langle Z,X\rangle - Z\langle X,Y\rangle \\ &- \langle X,[Y,Z]\rangle + \langle Y,[Z,X]\rangle + \langle Z,[X,Y]\rangle].\end{aligned}$$

This formula implies uniqueness, and in fact can serve to define a linear connection that is symmetric and compatible with the metric. □

Using the Kozul formula one can show that the connection coefficients of a Levi-Civita connection are given by,

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})\,,$$

where $[g^{ij}] = [g_{ij}]^{-1}$.

1.7 Geodesics

Let $\sigma : I \to M$, $t \to \sigma(t)$, be a smooth curve in a pseudo-Riemannian manifold M. Let $\mathfrak{X}(\sigma)$ denote the collection of smooth vector fields X along σ,

$$t \xrightarrow{X} X(t) \in T_{\sigma(t)}M$$

In local coordinates (U, x^i), we have

$$\begin{aligned}\sigma : x^i &= x^i(t)\,, \quad i = 1, ..., n \\ X(t) &= X^i(t)\partial_i|_{\sigma(t)}\,,\end{aligned}$$

where the components $X^i(t)$ are smooth.

The Levi-Civita connection ∇ on M induces a covariant differentiation on vector field along σ,

$$\frac{D}{dt} : \mathfrak{X}(\sigma) \to \mathfrak{X}(\sigma)$$

Proposition 1.2. *Let $\sigma : I \to M$ be a smooth curve in a pseudo-Riemannian manifold M. Then there exists a unique $\mathbb{R}$-linear operator*

$$\frac{D}{dt} : \mathfrak{X}(\sigma) \to \mathfrak{X}(\sigma)$$

satisfying

(1) for $X \in \mathfrak{X}(\sigma)$, $f \in C^\infty(I)$,

$$\frac{D}{dt} fX = \frac{df}{dt} X + f\frac{DX}{dt}\,,$$

(2) for $X \in \mathfrak{X}(M)$,

$$\frac{D}{dt}X|_{\sigma(t)} = \nabla_{\sigma'(t)}X.$$

In local coordinates we find that,

$$\frac{DX}{dt} = (\frac{dX^k}{dt} + \Gamma^k_{ij}\frac{dx^i}{dt}X^j)\partial_k. \tag{1.2}$$

Also we note that the operator $\frac{D}{dt}$ obeys the metric product rule,

$$\frac{d}{dt}\langle X, Y\rangle = \langle \frac{DX}{dt}, Y\rangle + \langle X, \frac{DY}{dt}\rangle.$$

Given a smooth curve $t \to \sigma(t)$ in M, $\frac{D}{dt}\frac{d\sigma}{dt}$ is the *covariant acceleration* of σ. In local coordinates,

$$\frac{D}{dt}\frac{d\sigma}{dt} = (\frac{d^2x^k}{dt^2} + \Gamma^k_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt})\partial_k\,,$$

as follows by setting $X^k = \frac{dx^k}{dt}$ in Equation (1.2).

Definition 1.6. *A smooth curve $t \to \sigma(t)$ is a geodesic provided it has vanishing covariant acceleration,*

$$\frac{D}{dt}\frac{d\sigma}{dt} = 0 \qquad \textit{(Geodesic equation)}$$

The basic existence and uniqueness result for systems of ODE's guarantees the following.

Proposition 1.3. *Given $p \in M$ and $v \in T_pM$, there exists an interval I about $t = 0$ and a unique geodesic $\sigma : I \to M$, $t \to \sigma(t)$, satisfying,*

$$\sigma(0) = p\,, \qquad \frac{d\sigma}{dt}(0) = v\,.$$

In fact, by a more refined analysis it can be shown that each $p \in M$ is contained in a *(geodesically) convex* neighborhood U, which has the property that any two points in U can be joined by a unique geodesic contained in U. In fact U can be chosen so as to be a normal neighborhood of each of its points; cf. [22], p. 129. (Recall, a normal neighborhood of $p \in M$ is the diffeomorphic image under the exponential map of a star-shaped domain about $\mathbf{0} \in T_pM$.)

1.8 Riemann curvature tensor

Definition 1.7. *Let M^n be a pseudo-Riemannian manifold. The Riemann curvature tensor of M is the map $R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ given by*

$$R(X, Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z\,. \tag{1.3}$$

R so defined is trilinear wrt to $C^\infty(M)$. That R is $\mathbb{R}$-trilinear is clear. The key point is the following.

Proposition 1.4. *For $f \in C^\infty(M)$,*

$$R(fX, Y)Z = R(X, fY)Z = R(X, Y)fZ = fR(X, Y)Z\,.$$

Proof. Exercise.

Proposition 1.4 implies that R is indeed *tensorial*, i.e., that the value of $R(X, Y)Z$ at $p \in M$ depends only on the value of X, Y, Z at p; hence for $(R(X, Y)Z)_p$ we may write $R(X_p, Y_p)Z_p$.

From the analytic point of view, the Riemann curvature tensor R may viewed as a measure of the extent to which covariant differentiation fails to commute. This failure to commute may be seen as an obstruction to the existence of parallel vector fields. According to Riemann's theorem, a pseudo-Riemannian manifold is locally pseudo-Euclidean iff the Riemann curvature tensor vanishes.

Proposition 1.5. *The Riemann curvature tensor has the following symmetry properties.*

(1) $R(X, Y)Z + R(Y, X)Z = 0$,

(2) $R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0$ *(first Bianchi identity)* ,

(3) $\langle R(X, Y)Z, W\rangle + \langle R(X, Y)W, Z\rangle = 0$,

(4) $\langle R(X, Y)Z, W\rangle = \langle R(Z, W)X, Y\rangle$.

The components ${R^l}_{kij}$ of the Riemann curvature tensor R in a coordinate chart (U, x^i) are defined by the following equation,

$$R(\partial_i, \partial_j)\partial_k = {R^l}_{kij}\partial_l$$

All of the above symmetries can be expressed in terms of components.

The Ricci tensor is obtained by contraction,

$$R_{ij} = {R^l}_{ilj}$$

Symmetries of the Riemann curvature tensor imply that the Ricci tensor is symmetric, $R_{ij} = R_{ji}$. By tracing the Ricci tensor, we obtain the scalar curvature,

$$R = g^{ij} R_{ij} \,,$$

where $[g^{ij}] = [g_{ij}]^{-1}$. The Einstein equation, with cosmological term, is the tensor equation,

$$R_{ij} - \frac{1}{2} R g_{ij} + \Lambda g_{ij} = 8\pi T_{ij} \,, \tag{1.4}$$

where Λ is the cosmological constant and T_{ij} is the energy-momentum tensor.

1.9 Sectional curvature

Let $(M^n, \langle\,,\,\rangle)$ be a pseudo-Riemannian manifold. A 2-dimensional subspace Π of the tangent space T_pM is called a tangent plane to M at p. Π is said to be nondegenerate provided $\langle\,,\,\rangle_p$ restricted to Π is nondegenerate. Suppose the vectors $X, Y \in T_pM$ span (i.e., form a basis for) Π. Then the sectional curvature $K(\Pi)$ of the tangent plane Π, is defined as

$$K(\Pi) = \frac{\langle R(X,Y)Y, X\rangle}{\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2}.$$

This expression is easily seen to be independent of the spanning set X, Y. Moreover, nondegeneracy ensures that the denominator is nonzero.

Sectional curvature has a natural geometric interpretation based on the following fact. If M^2 is a surface in $\mathbb{R}^3$ with its induced metric, and Π is the tangent plane to M at p, then $K(\Pi) =$ the Gaussian curvature of M at p.

M^n is said to have constant curvature if there exists a constant K_0 such that for all $p \in M$, and for all nondegenerate tangent planes Π at p, $K(\Pi) = K_0$. Minkowski space, de Sitter space and anti-de Sitter space are Lorentzian manifolds of constant curvature (zero, positive and negative, respectively).

2 Lorentzian geometry and causal theory

2.1 Lorentzian manifolds

Let $(M^{n+1}, \langle\,,\,\rangle)$ be a Lorentzian manifold. Hence, for each $p \in M$, $\langle\,,\,\rangle : T_pM \times T_pM \to \mathbb{R}$ is a scalar product of signature $(-1, +1, ..., +1)$. Let $\{e_0, e_1, ..., e_n)$ be an orthonormal basis for T_pM. Set $g_{ij} = \langle e_i, e_j\rangle$. Then, as a matrix,

$$[g_{ij}] = [\eta_{ij}] = \mathrm{diag}(-1, 1, ..., 1) \,.$$

Hence, for $X, Y \in T_pM$,

$$\begin{aligned} \langle X, Y\rangle &= g_{ij} X^i Y^j = \eta_{ij} X^i Y^j \\ &= -X^0 Y^0 + \sum_{i=1}^{n} X^i Y^i \,. \end{aligned}$$

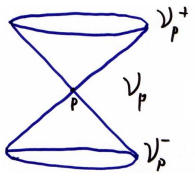
For $X \in T_pM$,

$$X \text{ is } \begin{cases} \text{timelike} & \text{if } \langle X, X\rangle < 0 \\ \text{null} & \text{if } \langle X, X\rangle = 0 \\ \text{spacelike} & \text{if } \langle X, X\rangle > 0\ . \end{cases}$$

Finally, we say X is causal (or nonspacelike) if it is timelike or null.

We see that the set of null vectors $X \in T_pM$,

$$\langle X, X\rangle = -(X^0)^2 + \sum_{i=1}^{n} (X^i)^2 = 0$$



forms a double cone $\mathcal{V}_p$ in the tangent space T_pM, called the null cone at p. Timelike vectors point inside the null cone and spacelike vectors point outside.

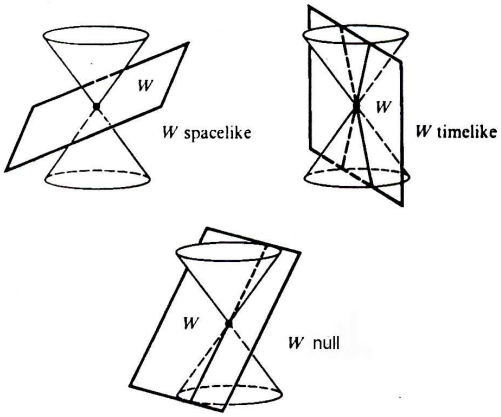
A subspace W of T_pM may be assigned a causal character as follows,

(1) W is *spacelike* if $\langle\ ,\ \rangle|_W$ has index 0, i.e., is positive definite.

(2) W is *timelike* if $\langle\ ,\ \rangle|_W$ has index 1.

(3) W is *null* (or *lightlike*) if $\langle\ ,\ \rangle|_W$ is degenerate.

(see the figure).



For $X \in T_pM$, $X \neq 0$, let $[X]^\perp = \{Y \in T_pM : \langle X, Y\rangle = 0\}$. Note that $[X]^\perp$ is spacelike, timelike, or null, depending on whether X is timelike, spacelike, or null, respectively.

For a causal vector $X \in T_pM$, define its length as, $|X| = \sqrt{-\langle X, X\rangle}$.

Proposition 2.1. *The following basic inequalities hold.*

(1) (Reverse Schwarz inequality) For all causal vectors $X, Y \in T_pM$,

$$|\langle X, Y\rangle| \geq |X||Y|$$

(2) (Reverse triangle inequality) For all causal vectors X, Y that point into the same half cone of the null cone at p,

$$|X + Y| \geq |X| + |Y| .$$

Proof. Exercise.

The Reverse triangle inequality is the source of the twin paradox.

Let $\gamma : I \to M$ be a smooth curve in M. γ is said to be timelike (resp., spacelike, null, causal) provided $\gamma'(t)$ is timelike (resp., spacelike, null, causal) for all $t \in I$. Heuristically, in accordance with relativity, information flows along causal curves, and so such curves are the focus of our attention. The notion of a causal curve extends in a natural way to piecewise smooth curves. The only extra requirement is that when two segments join, at some point p, say, the end point tangent vectors must point into the same half cone of the null cone $\mathcal{V}_p$ at p. We will normally work within this class of piecewise smooth causal curves. Finally, note since geodesics γ are constant speed curves ($\langle \gamma', \gamma' \rangle = const.$), each geodesic in a Lorentzian manifold is either timelike, spacelike or null.

The length of a causal curve $\gamma : [a, b] \to M$, is defined by

$$L(\gamma) = \text{Length of } \gamma = \int_a^b |\gamma'(t)| dt = \int_a^b \sqrt{-\langle \gamma'(t), \gamma'(t) \rangle}\, dt .$$

If γ is timelike one can introduce arc length parameter along γ. In general relativity, a timelike curve corresponds to the history of an observer, and arc length parameter, called proper time, corresponds to time kept by the observer.

Certain geometric and causal features of Minkowski space that may fail to hold in the large in a general Lorentzian manifold, nontheless hold locally. Let U be a convex neighborhood in a Lorentzian manifold. Hence for each pair of points $p, q \in U$ there exists a unique geodesic segment from p to q in U, which we denote by $\overline{pq}$.

Proposition 2.2 ([22], p. 146)**.** *Let U be a convex neighborhood in a Lorentzian manifold M^{n+1}.*

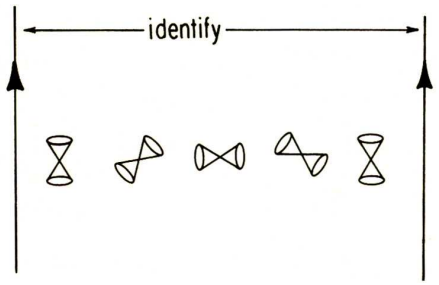
(1) If there is a timelike (resp., causal) curve in U from p to q then $\overline{pq}$ is timelike (causal).

(2) If $\overline{pq}$ is timelike then $L(\overline{pq}) \geq L(\gamma)$ for all causal curves γ in U from p to q. Moreover, the inequality is strict unless, when suitable parametrized, $\gamma = \overline{pq}$.

Thus, within a convex neighborhood, null geodesics are *achronal*, i.e., no two points can be joined by a timlike curve, and timelike geodesics are *maximal*, i.e., are causal curves of greatest length. Both of these features can fail in the large.

$\mathcal{V}_p$, the null cone at p, consists of two half-cones $\mathcal{V}_p^+$ and $\mathcal{V}_p^-$ (see the figure on the previous page). We may designate one of the half cones, $\mathcal{V}_p^+$, say, as the *future* null

cone at p, and the other half cone, $\mathcal{V}_p^-$, as the *past* null cone at p. The assignment of a future cone and past cone to each point of M^{n+1} can always be done in a continuous manner locally. If the assignment can made in a continuous manner over all of M then we say that M is *time-orientable*. The following figure illustrates a Lorentzian manifold that is *not* time-orientable.



There are various ways to make the phrase "continuous assignment" precise, but they all result in the following fact.

Fact 2.3. *A Lorentzian manifold M^{n+1} is time-orientable iff it admits a smooth timelike vector field U.*

If M is time-orientable, the choice of a smooth time-like vector field U fixes a time orientation on M. For any $p \in M$, a causal vector $X \in T_p$ is future directed (resp. past directed) provided $\langle X, U\rangle < 0$ (resp. $\langle X, U\rangle > 0$). Thus X is future directed if it points into the same null half cone at p as U.

By a *spacetime* we mean a connected time-oriented Lorentzian manifold $(M^{n+1}, \langle\ ,\ \rangle)$. Henceforth, we restrict attention to spacetimes.

2.2 Futures and pasts

Let $(M, \langle\ ,\ \rangle)$ be a spacetime. We introduce the standard causal relations '$\ll$' and '$<$'. A timelike (resp. causal) curve $\gamma : I \to M$ is said to be *future directed* provided each tangent vector $\gamma'(t)$, $t \in I$, is future directed. (*Past-directed* timelike and causal curves are defined in a time-dual manner.)

Definition 2.1. *For $p, q \in M$,*

(1) $p \ll q$ means there exists a future directed timelike curve in M from p to q (we say that q is in the timelike future of p),

(2) $p < q$ means there exists a future directed causal curve in M from p to q (we say that q is in the causal future of p),

We shall use the notation $p \leq q$ to mean $p = q$ or $p < q$.

The causal relations $\ll$ and $<$ are clearly transitive. Also, from variational considerations, it is heuristically clear that the following holds,

$$\text{if } p \ll q \text{ and } q < r \text{ or if } p < q \text{ and } q << r \text{ then } p \ll r\,.$$

The above statement is a consequence of the following fundamental causality result; see [22, p. 294] for a careful proof.

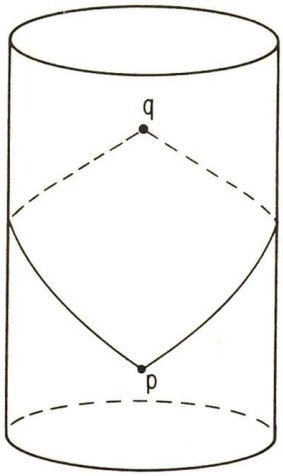
Proposition 2.4. *In a spacetime M, if q is in the causal future of p ($p < q$) but is not in the timelike future of p ($p \not\ll q$) then any future directed causal curve γ from p to q must be a null geodesic (when suitably parameterized).*

Given any point p in a spacetime M, the *timelike future* and *causal future* of p, denoted $I^+(p)$ and $J^+(p)$, respectively are defined as,

$$I^+(p) = \{q \in M : p \ll q\} \quad \text{and} \quad J^+(p) = \{q \in M : p \leq q\}\,.$$

Hence, $I^+(p)$ consists of all points in M that can be reached from p by a future directed timelike curve, and $J^+(p)$ consists of the point p and all points in M that can be reached from p by a future directed causal curve. The timelike and causal *pasts* of p, $I^-(p)$ and $J^-(p)$, respectively, are defined in a time-dual manner in terms of past directed timelike and causal curves. Note by Proposition 2.4, if $q \in J^+(p) \setminus I^+(p)$ ($q \neq p$) then there exists a future directed null geodesic from p to q.

If p is a point in Minkowski space $\mathbb{M}^{n+1}$, then $I^+(p)$ is open, $J^+(p)$ is closed and $\partial I^+(p) = J^+(p) \setminus I^+(p)$ is just the future null cone at p. $I^+(p)$ consists of all points inside the future null cone, and $J^+(p)$ consists of all points on and inside the future null cone. We note, however, that curvature and topology can drastically change the structure of 'null cones' in spacetime. Consider the example depicted in the following figure of a flat spacetime cylinder, closed in space. For any point p in this spacetime the future 'null cone' at p, $\partial I^+(p)$, is compact and consists of the two future directed null geodesic segments emanating from p that meet to the future at a point q. By extending these geodesics beyond q we enter $I^+(p)$.



In some situations it is convenient to restrict the causal relations $\ll$ and $<$ to open subsets U of a spacetime M. For example, $I^+(p, U)$, the chronological future of p within U, consists of all points q in U for which there exists a future directed timelike curve *within* U from p to q, etc. Note that, in general $I^+(p, U) \neq I^+(p) \cap U$.

In general the sets $I^+(p)$ in a spacetime M are open. This is heuristically rather clear: A sufficiently small smooth perturbation of a timelike curve is still timelike. A rigorous proof is based on the causality of convex neighborhoods.

Proposition 2.5. *Let U be a convex neighborhood in a spacetime M. Then, for each $p \in U$,*

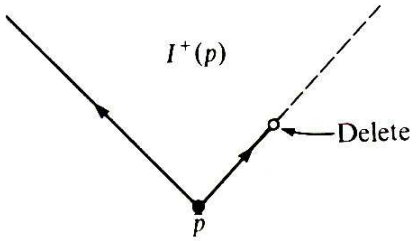
(1) $I^+(p, U)$ is open in U (and hence M),

(2) $J^+(p, U)$ is the closure in U of $I^+(p, U)$.

This proposition follows essentially from part (1) of Proposition 2.2.

Exercise: Prove that for each p in a spacetime M, $I^+(p)$ is open.

In general, sets of the form $J^+(p)$ need not be closed. This can be seen by removing a point from Minkowski space, as illustrated in the figure below.



Points on the dashed line are not in $J^+(p)$, but are in the closure $\overline{J^+(p)}$.

For any subset $S \subset M$, we define the timelike and causal future of S, $I^+(S)$ and $J^+(S)$, respectively by

$$I^+(S) = \bigcup_{p \in S} I^+(p) \quad \text{and} \quad J^+(S) = \bigcup_{p \in S} J^+(p)\,.$$

Thus, $I^+(S)$ consists of all points in M reached by a future directed timelike curve starting from S, and $J^+(S)$ consists of the points of S, together with the points in M reached by a future directed causal curve starting from S. Since arbitrary unions of open sets are open, it follows that $I^+(S)$ is always an open set. $I^-(S)$ and $J^-(S)$ are defined in a time-dual manner.

Although in general $J^+(S) \neq \overline{I^+(S)}$, the following relationships always hold between $I^+(S)$ and $J^+(S)$.

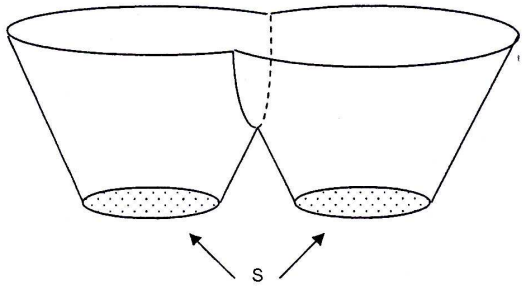
Proposition 2.6. *For all subsets $S \subset M$,*

(1) $\operatorname{int} J^+(S) = I^+(S)$,

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

Proof. Exercise.

Achronal sets play an important role in causal theory. A subset $S \subset M$ is *achronal* provided no two of its points can be joined by a timelike curve. Of particular importance are *achronal boundaries*. By definition, an achronal boundary is a set of the form $\partial I^+(S)$ (or $\partial I^-(S)$), for some $S \subset M$. We wish to describe several important structural properties of achronal boundaries. The following figure illustrates nicely the properties to be discussed. It depicts the achronal boundary $\partial I^+(S)$ in Minkowski 3-space $\mathbb{M}^3$, where S is the disjoint union of two spacelike disks; $\partial I^+(S)$ consists of S and the merging of two future light cones.



Proposition 2.7. *An achronal boundary $\partial I^+(S)$, if nonempty, is a closed achronal C^0 hypersurface in M.*

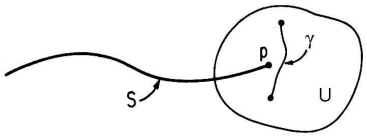
We discuss the proof of this proposition, beginning with the following basic lemma.

Lemma 2.8. *If $p \in \partial I^+(S)$ then $I^+(p) \subset I^+(S)$, and $I^-(p) \subset M \setminus \overline{I^+(S)}$.*

Proof. To prove the first part of the lemma, note that if $q \in I^+(p)$ then $p \in I^-(q)$, and hence $I^-(q)$ is a neighborhood of p. Since p is on the boundary of $I^+(S)$, it follows that $I^-(q) \cap I^+(S) \neq \emptyset$, and hence $q \in I^+(S)$. The second part of the lemma, which can be proved similarly, is left as an exercise. □

Next, we need to introduce the notion of an *edge point* of an achronal set.

Definition 2.2. *Let $S \subset M$ be achronal. Then $p \in \overline{S}$ is an edge point of S provided every neighborhood U of p contains a timelike curve γ from $I^-(p, U)$ to $I^+(p, U)$ that does* not *meet S (see the figure).*



We denote by edge S the set of edge points of S. Note that $\overline{S} \setminus S \subset \text{edge}\, S \subset \overline{S}$. If edge $S = \emptyset$ we say that S is edgeless.

Claim: An achronal boundary $\partial I^+(S)$ is achronal and edgeless.

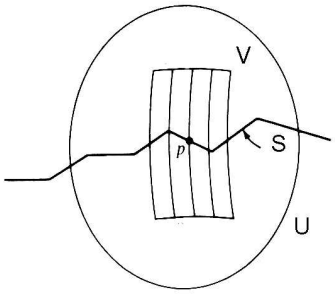
Proof of the claim: Suppose there exist $p, q \in \partial I^+(S)$, with $q \in I^+(p)$. By Lemma 2.8, $q \in I^+(S)$. But since $I^+(S)$ is open, $I^+(S) \cap \partial I^+(S) = \emptyset$. Thus, $\partial I^+(S)$ is achronal. Moreover, Lemma 2.8 implies that for any $p \in \partial I^+(S)$, any timelike curve from $I^-(p)$ to $I^+(p)$ must meet $\partial I^+(S)$. It follows that $\partial I^+(S)$ is edgless. □

Proposition 2.7 now follows from the following basic result.

Proposition 2.9. *Let S be achronal. Then $S \setminus \text{edge}\, S$, if nonempty, is a C^0 hypersurface in M.*

Proof. We sketch the proof; for details, see [22, p. 413]. It suffices to show that in a neighborhood of each $p \in S \setminus \text{edge}\, S$, $S \setminus \text{edge}\, S$ can be expressed as a C^0 graph over a smooth hypersurface.

Fix $p \in S \setminus \text{edge}\, S$. Since p is not an edge point there exists a neighborhood U of p such that every timelike curve from $I^-(p, U)$ to $I^+(p, U)$ meets S. Let X be a future directed timelike vector field on M, and let $\mathcal{N}$ be a smooth hypersurface in U transverse to X near p. Then, by choosing $\mathcal{N}$ small enough, each integral curve of X through $\mathcal{N}$ will meet S, and meet it exactly once since S is achronal. Using the flow generated by X, it follows that there is a neighborhood $V \approx (t_1, t_2) \times \mathcal{N}$ of p such that $S \cap V$ is given as the graph of a function $t = h(x)$, $x \in \mathcal{N}$ (see the figure below)



One can now show that a discontinuity of h at some point $x_0 \in \mathcal{N}$ leads to an achronality violation of S. Hence h must be continuous. □

The next result shows that, in general, large portions of achronal boundaries are ruled by null geodesics. A future (resp., past) directed causal curve $\gamma : (a, b) \to M$ is said to be *future (resp., past) inextendible* in M if $\lim_{t \to b^-} \gamma(t)$ does not exist. A future directed causal curve $\gamma : (a, b) \to M$ is said to be inextendible if γ and $-\gamma$ are future and past inextendible, respectively.

Proposition 2.10. *Let $S \subset M$ be closed. Then each $p \in \partial I^+(S) \setminus S$ lies on a null geodesic contained in $\partial I^+(S)$, which either has a past end point on S, or else is past inextendible in M.*

The proof uses a standard tool in causal theory, namely that of taking a limit of causal curves. A technical difficulty arises however in that a limit of smooth causal curves need not be smooth. Thus, we are lead to introduce the notion of a C^0 causal curve.

Definition 2.3. *A continuous curve $\gamma : I \to M$ is said to be a future directed C^0 causal curve provided for each $t_0 \in I$, there is an open subinterval $I_0 \subset I$ about t_0 and a convex neighborhood U of $\gamma(t_0)$ such that given any $t_1, t_2 \in I_0$ with $t_1 < t_2$, there exists a smooth future directed causal curve in U from $\gamma(t_1)$ to $\gamma(t_2)$.*

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

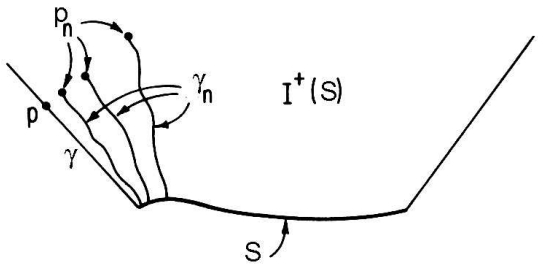
We now give a version of the *limit curve lemma* (cf., [3, p. 511]). For its statement it is convenient to introduce a background complete Riemannian (positive definite) metric h on M. Observe that any future inextendible causal γ will have infinite length to the future, as measured in the metric h. Hence, if parameterized with respect to h-arc length, γ will be defined on the interval $[0, \infty)$.

Lemma 2.11 (Limit curve lemma)**.** *Let $\gamma_n : [0, \infty) \to M$ be a sequence of future inextendible causal curves, parameterized with respect to h-arc length, and suppose that $p \in M$ is an accumulation point of the sequence $\{\gamma_n(0)\}$. Then there exists a future inextendible C^0 causal curve $\gamma : [0, \infty) \to M$ such that $\gamma(0) = p$ and a subsequence $\{\gamma_m\}$ which converges to γ uniformly with respect to h on compact subsets of $[0, \infty)$.*

The proof of this lemma is an application of Arzela's theorem; see especially the proof of Proposition 3.31 in [3]. There are analogous versions of the limit curve lemma for past inextendible, and (past and future) inextendible causal curves.

Remark: We note that C^0 causal curves obey a local Lipschitz condition, and hence are rectifiable. Thus, in the limit curve lemma, the γ_n's could be taken to be C^0 causal curves. We also note that the "limit" parameter acquired by the limit curve γ need not in general be the h-arc length parameter.

Proof of Proposition 2.10. Fix $p \in \partial I^+(S) \setminus S$. Since $p \in \partial I^+(S)$, there exists a sequence of points $p_n \in I^+(S)$, such that $p_n \to p$. For each n, let $\gamma_n : [0, a_n] \to M$ be a past directed timelike curve from p_n to $q_n \in S$, parameterized with respect to h-arc length. Extend each γ_n to a past inextendible timelike curve $\tilde{\gamma}_n : [0, \infty) \to M$, parameterized with respect to h-arc length. By the limit curve lemma, there exists a subsequence $\tilde{\gamma}_m : [0, \infty) \to M$ that converges to a past inextendible C^0 causal curve $\gamma : [0, \infty) \to M$ such that $\gamma(0) = p$. By taking a further subsequence if necessary we can assume $a_m \uparrow a$, $a \in (0, \infty]$. We claim that $\gamma|_{[0,a]}$ (or $\gamma|_{[0,a)}$ if $a = \infty$) is the desired null geodesic (see the figure).



Fix $t \in (0, a)$. Eventually $a_m > t$, and so for large m we have $\tilde{\gamma}_m(t) = \gamma_m(t) \in I^+(S)$. Hence, since $\gamma(t) = \lim_{m\to\infty} \gamma_m(t)$, it follows that $\gamma(t) \in \overline{I^+(S)}$. Suppose $\gamma(t) \in I^+(S)$. Then there exists $x \in S$ such that $x \ll \gamma(t) < p$. This implies $p \in I^+(S)$, contradicting that it is on the boundary. It follows that $\gamma(t) \in \partial I^+(S)$. Thus we have shown that $\gamma|_{[0,a)} \subset \partial I^+(S)$. Suppose for the moment $\gamma|_{[0,a)}$ is piecewise smooth. Since $\partial I^+(S)$ is achronal, no two points of γ can be joined by a timelike curve. It then follows from Proposition 2.4 that γ is a null geodesic. But using the fact that C^0 causal curves can be approximated by piecewise smooth causal curves, one can show in the general case that $\gamma|_{[0,a)}$ is a null geodesic. (Exercise: Show this.)

Finally, we consider the two cases $a < \infty$ and $a = \infty$. If $a < \infty$, then by the uniform convergence, $\gamma(a) = \lim_{m\to\infty} \gamma_m(a_m) = \lim_{m\to\infty} q_m \in S$, since S is closed. Thus, we have a null geodesic from p contained in $\partial I^+(S)$ that ends on S. If $a = \infty$ then we have a null geodesic from p in $\partial I^+(S)$ that is past inextendible in M. □

2.3 Causality conditions

A number of results in Lorentzian geometry and general relativity require some sort of causality condition. It is perhaps natural on physical grounds to rule out the occurrence of closed timelike curves. Physically, the existence of such a curve signifies the existence of an observer who is able to travel into his/her own past, which leads to variety of paradoxical situations. A spacetime M satisfies the *chronology condition* provided there are no closed timelike curves in M. Compact spacetimes have limited interest in general relativity since they all violate the chronology condition.

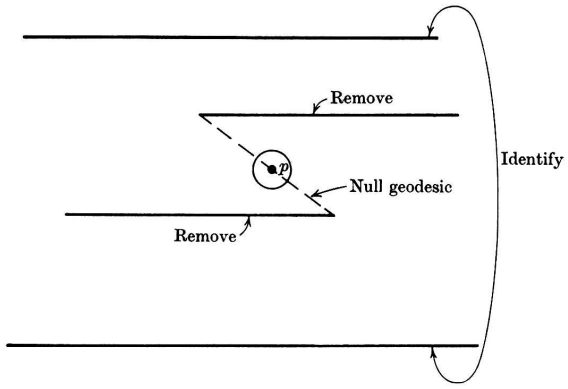
Proposition 2.12. *Every compact spacetime contains a closed timelike curve.*

Proof. The sets $\{I^+(p); p \in M\}$ form an open cover of M from which we can abstract a finite subcover: $I^+(p_1)$, $I^+(p_2)$, ..., $I^+(p_k)$. We may assume that this is the minimal number of such sets covering M. Since these sets cover M, $p_1 \in I^+(p_i)$ for some i. It follows that $I^+(p_1) \subset I^+(p_i)$. Hence, if $i \neq 1$, we could reduce the number of sets in the cover. Thus, $p_1 \in I^+(p_1)$ which implies that there is a closed timelike curve through p_1. □

A somewhat stronger condition than the chronology condition is the *causality condition*. A spacetime M satisfies the causality condition provided there are no closed (nontrivial) causal curves in M.

Exercise: Construct a spacetime that satisfies the chronology condition but not the causality condition.

A spacetime that satisfies the causality condition can nontheless be on the verge of failing it, in the sense that there exist causal curves that are "almost closed", as illustrated by the following figure.



Strong causality is a condition that rules out almost closed causal curves. An open set U in spacetime M is said to be *causally convex* provided no causal curve in M meets U in a disconnected sets. Given $p \in M$, strong causality is said to hold at p provided p has arbitrarily small convex neighborhoods, i.e., for each neighborhood V of p there exists a causally neighborhood U of p such that $U \subset V$. Note that strong causality fails at the point p in the figure above. In fact strong causality fails at all points along the dashed null geodesic. It can be shown that the set of points at which strong causality holds is open.

M is said to be strongly causal if strong causality holds at all of its points. This is the "standard" causality condition in spacetime geometry, and, although there are even stronger causality conditions, it is sufficient for most applications. There is an interesting connection between strong causality and the so-called *Alexandrov topology.* The sets of the form $I^+(p) \cap I^-(q)$ form the base for a topology on M, which is the Alexandrov topology. This topology is in general more coarse than the manifold topology of M. However it can be shown that the Alexandrov topology agrees with the manifold topology iff M is strongly causal.

The following lemma is often useful.

Lemma 2.13. *Suppose strong causality holds at each point of a compact set K in a spacetime M. If $\gamma : [0, b) \to M$ is a future inextendible causal curve that starts in K then eventually it leaves K and does not return, i.e., there exists $t_0 \in [0, b)$ such that $\gamma(t) \notin K$ for all $t \in [t_0, b)$.*

Proof. Exercise.

In referring to the property described by this lemma, we say that a future inextendible causal curve cannot be "imprisoned" or "partially imprisoned" in a compact set on which strong causality holds.

We now come to a fundamental condition in spacetime geometry, that of *global hyperbolicity*. Mathematically, global hyperbolicity is a basic 'niceness' condition that often plays a role analogous to geodesic completeness in Riemannian geometry. Physically, global hyperbolicity is connected to the notion of (strong) cosmic censorship introduced by Roger Penrose. This is the conjecture that, generically, spacetime solutions to the Einstein equations do not admit *naked singularities* (singularities visible to some observer).

Definition 2.4. *A spacetime M is said to be globally hyperbolic provided*

(1) M is strongly causal.

(2) (Internal Compactness) The sets $J^+(p) \cap J^-(q)$ are compact for all $p, q \in M$.

Condition (2) says roughly that M has no holes or gaps. For example Minkowski space $\mathbb{M}^{n+1}$ is globally hyperbolic but the spacetime obtained by removing one point from it is not.

We consider a few basic consequences of global hyperbolicity.

Proposition 2.14. *Let M be a globally hyperbolic spacetime. Then,*

(1) The sets $J^{\pm}(A)$ are closed, for all compact $A \subset M$.

(2) The sets $J^+(A) \cap J^-(B)$ are compact, for all compact $A, B \subset M$.

Proof. We prove $J^{\pm}(p)$ are compact for all $p \in$M, and leave the rest as an exercise. Suppose $q \in \overline{J^+(p)} \setminus J^+(p)$ for some $p \in M$. Choose $r \in I^+(q)$, and $\{q_n\} \subset J^+(p)$, with $q_n \to q$. Since $I^-(r)$ is an open neighborhood of q, $\{q_n\} \subset J^-(r)$ for n large. It follows that $q \in \overline{J^+(p) \cap J^-(r)} = J^+(p) \cap J^-(r)$, since $J^+(p) \cap J^-(r)$ is compact and hence closed. But this contradicts $q \notin J^+(p)$. Thus, $J^+(p)$ is closed, and similarly so is $J^-(p)$. □

Analogously to the case of Riemannian geometry, one can learn much about the global structure of spacetime by studying its causal geodesics. Locally, causal geodesics maximize Lorentzian arc length (cf., Proposition 2.2). Given $p, q \in M$, with $p < q$, we wish to consider conditions under which there exists a maximal future directed causal geodesic γ from p to q, where by maximal we mean that for any future directed causal curve σ from p to q, $L(\gamma) \geq L(\sigma)$.

For this purpose it is convenient to introduce the Lorentzian *distance function*, $d : M \times M \to [0, \infty]$. For $p < q$, let $\Omega_{p,q}$ denote the collection of future directed causal curves from p to q. Then, for any $p, q \in M$, define

$$d(p,q) = \begin{cases} \sup\{L(\sigma) : \sigma \in \Omega_{p,q}\}, & \text{if } p < q \\ 0, & \text{if } p \not< q \end{cases}$$

While the Lorentzian distance function is not a distance function in the usual sense of metric spaces, and may not even be finite valued, it does have a few nice properties. For one, it obeys a *reverse triangle inequality*,

$$\text{if } p < r < q \quad \text{then} \quad d(p,q) \geq d(p,r) + d(r,q)\,.$$

Exercise: Prove this.

We have the following basic fact.

Proposition 2.15. *The Lorentzian distance function is lower semi-continuous.*

Proof. Fix $p, q \in M$. Given $\epsilon > 0$ we need to find neighborhhoods U and V of p and q, respectively, such that for all $x \in U$ and all $y \in V$, $d(x,y) > d(p,q) - \epsilon$.

If $d(p,q) = 0$ there is nothing to prove. Thus, we assume $p < q$ and $0 < d(p,q) < \infty$. We leave the case $d(p,q) = \infty$ as an exercise. Let σ be a future directed timelike curve from p to q such that $L(\sigma) = d(p,q) - \epsilon/3$. Let U and V be convex neighborhoods of p and q, respectively. Choose p' on σ close to p and q' on σ close to q. Then $U' = I^-(p', U)$ and $V' = I^+(q', V)$ are neighborhoods of p and q, respectively. Moreover, by choosing p' sufficiently close to p and q' sufficiently close to q, one verifies that for all $x \in U'$ and $y \in V'$, there exists a future directed timelike curve α from x to y, containing the portion of σ from p' to q', having length $L(\alpha) > d(p,q) - \epsilon/2$. □

Though the Lorentzian distance function is not continuous in general , it is continuous (and finite valued) for globally hyperbolic spacetimes; cf., [22, p. 412].

Given $p < q$, note that a causal geodesic segment γ having length $L(\gamma) = d(p,q)$ is maximal. Global hyperbolicity is the standard condition to ensure the existence of maximal causal geodesic segments.

Proposition 2.16. *Let M be a globally hyperbolic spacetime. Given $p, q \in M$, with $p < q$, there exists a maximal future directed causal geodesic γ from p to q ($L(\gamma) = d(p,q)$).*

Proof. The proof involves a standard limit curve argument, together with the fact that the Lorentzian arc length functional is upper semi-continuous; see [25, p. 54].

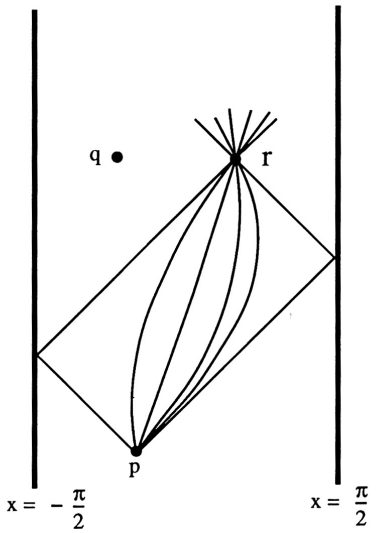
As usual, let h be a background complete Riemannian metric on M. For each n, let $\gamma_n : [0, a_n] \to M$ be a future directed causal curve from $p = \gamma_n(0)$ to $q = \gamma_n(a_n)$, parameterized with respect to h-arc length, such that $L(\gamma_n) \to d(p,q)$. Extend each γ_n to a future inextendible causal curve $\tilde{\gamma}_n : [0, \infty) \to M$, parameterized with respect to h-arc length. By the limit curve lemma, there exists a subsequence $\tilde{\gamma}_m : [0, \infty) \to M$ that converges to a future inextendible C^0 causal curve $\gamma : [0, \infty) \to M$ such that $\gamma(0) = p$. By taking a further subsequence if necessary we can assume $a_m \uparrow a$. Since each γ_m is contained in the compact set $J^+(p) \cap J^-(q)$, it follows that $\gamma|_{[0,a)} \subset \overline{J^+(p) \cap J^-(q)} = J^+(p) \cap J^-(q)$. Since M is strongly causal, it must be that $a < \infty$, otherwise, γ would be imprisoned in $J^+(p) \cap J^-(q)$, contradicting Lemma 2.13. Then, $\gamma(a) = \lim_{m \to \infty} \gamma_m(a_m) = q$.

Let $\bar{\gamma} = \gamma|_{[0,a]}$. $\bar{\gamma}$ is a future directed C^0 causal curve from p to q. Moreover, by the upper semi-continuity of L,

$$L(\bar{\gamma}) \geq \limsup_{m\to\infty} L(\gamma_m) = d(p,q)\,,$$

and so $L(\bar{\gamma}) = d(p,q)$. Hence, $\bar{\gamma}$ has maximal length among all future directed causal curves from p to q. This forces each sub-segment of $\bar{\gamma}$ to have maximal length. Using Proposition 2.2 (part (2) of which remains valid for C^0 causal curves) and Proposition 2.4, one can then argue that each sufficiently small segment of $\bar{\gamma}$ is a causal geodesic. (Exercise: Argue this.) □

Remarks: There are simple examples showing that if either of the conditions (1) or (2) fail to hold in the definition of global hyperbolicity then maximal segments may fail to exist. Moreover, contrary to the situation in Riemannian geometry, geodesic completeness does not guarantee the existence of maximal segments, as is well illustrated by anti-de Sitter space which is geodesically complete. The figure below depicts 2-dimensional anti-de Sitter space. It be can be represented as the strip $M = \{(t,x) : -\pi/2 < x < \pi/2\}$, equipped with the metric $ds^2 = \sec^2 x(-dt^2 + dx^2)$. Because the anti-de Sitter metric is conformal to the Minkowski metric on the strip, pasts and futures of both space times are the same. It can be shown that all future directed timelike geodesics emanating from p refocus at r. The points p and q are timelike related, but there is no timelike geodesic segment from p to q.



Global hyperbolicity is closely related to the existence of certain 'ideal initial value hypersurfaces', called *Cauchy surfaces*. There are slight variations in the literature in the definition of a Cauchy surface. Here we adopt the following definition.

Definition 2.5. *A Cauchy surface for a spacetime M is an achronal subset S of M which is met by every inextendible causal curve in M.*

From the definition it is easy to see that if S is a Cauchy surface for M then $S = \partial I^+(S) = \partial I^-(S)$. It follows from Proposition 2.7 that a Cauchy surface S is a closed achronal C^0 hypersurface in M.

Theorem 2.17. *Consider a spacetime M.*

(1) If M is globally hyperbolic then M has a Cauchy surface S (Geroch, [17]).

(2) If S is a Cauchy surface for M then M is homeomorphic to $\mathbb{R} \times S$.

Proof. We make a couple of comments about the proof. To prove (1), one introduces a measure μ on M such that $\mu(M) = 1$. Consider the function $f : M \to \mathbb{R}$ defined by

$$f(p) = \frac{\mu(J^-(p))}{\mu(J^+(p))}\,.$$

Internal compactness is used to show that f is continuous, and strong causality is used to show that f is strictly increasing along future directed causal curves. Moreover, if $\gamma : (a, b) \to M$ is a future directed inextendible causal curve in M, one shows $f(\gamma(t)) \to 0$ as $t \to a^+$, and $f(\gamma(t)) \to \infty$ as $t \to b^-$. It follows that $S = \{p \in M : f(p) = 1\}$ is a Cauchy surface for M. To prove (2), one introduces a future directed timelike vector field X on M. X can be scaled so that the time parameter t of each integral curve of X extends from $-\infty$ to ∞, with $t = 0$ at points of S. Each $p \in M$ is on an integral curve of X that meets S in exactly one point q. This sets up a correspondence $p \leftrightarrow (t, q)$, which gives the desired homeomorphism. □

As we discuss in the next subsection, the converse to (1) above holds. Thus, *a spacetime M is globally hyperbolic iff it admits a Cauchy surface S.* Along similar lines to (2) above, one has that any two Cauchy surfaces in a given globally hyperbolic spacetime are homeomorphic. Hence, in view of Theorem 2.17, any nontrivial topology in a globally hyperbolic spacetime must reside in its Cauchy surfaces.

The following fact is often useful.

Proposition 2.18. *If S is a compact achronal C^0 hypersurface in a globally hyperbolic spacetime M then S must be a Cauchy surface for M.*

The proof will be discussed in the next subsection.

2.4 Domains of dependence

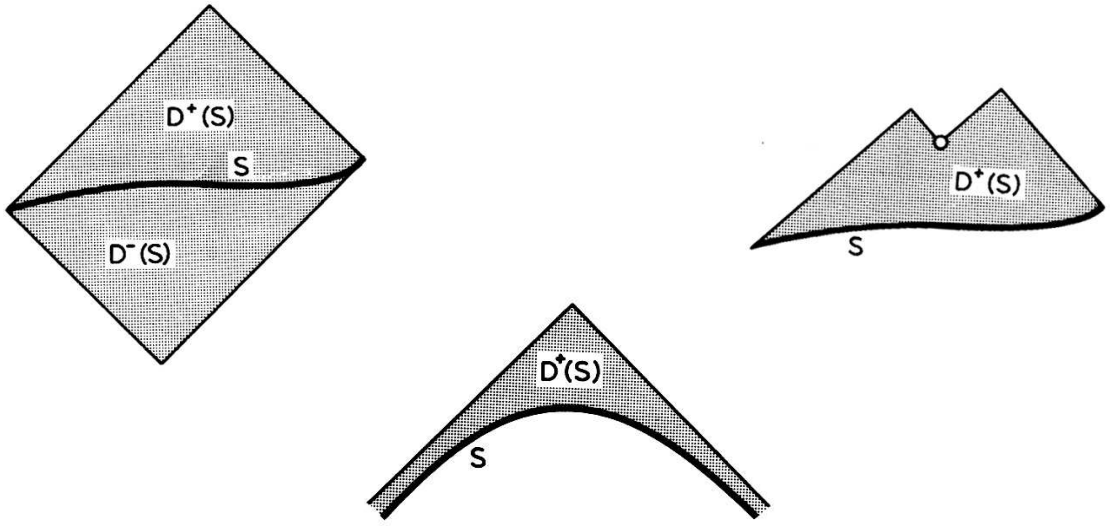
Definition 2.6. *Let S be an achronal set in a spacetime M. We define the future and past domains of dependence of S, $D^+(S)$ and $D^-(S)$, respectively, as follows,*

$$D^+(S) = \{p \in M : \text{ every past inextendible causal curve from } p \text{ meets } S\},$$
$$D^-(S) = \{p \in M : \text{ every future inextendible causal curve from } p \text{ meets } S\}.$$

The (total) domain of dependence of S is the union, $D(S) = D^+(S) \cup D^-(S)$.

In physical terms, since information travels along causal curves, a point in $D^+(S)$ only receives information from S. Thus if physical laws are suitably causal, initial data on S should determine the physics on $D^+(S)$ (in fact on all of $D(S)$).

Below we show a few examples of future and past domains of dependence.



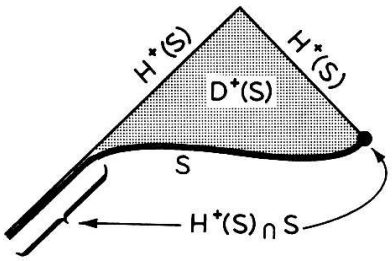
The figure in the top right shows the effect of removing a point from M. The bottom figure shows the future domain of dependence of the spacelike hyperboloid $t^2 - x^2 = 1$, $t < 0$, in the Minkowski plane.

If S is achronal, the future Cauchy horizon $H^+(S)$ of S is the future boundary of $D^+(S)$. This is made precise in the following definition.

Definition 2.7. *Let $S \subset M$ be achronal. The future Cauchy horizon $H^+(S)$ of S is defined as follows*

$$\begin{aligned} H^+(S) &= \{p \in \overline{D^+(S)} : I^+(p) \cap D^+(S) = \emptyset\} \\ &= \overline{D^+(S)} \setminus I^-(D^+(S))\,. \end{aligned}$$

The past Cauchy horizon $H^-(S)$ is defined time-dually. The (total) Cauchy horizon of S is defined as the union, $H(S) = H^+(S) = H^-(S)$.



We record some basic facts about domains of dependence and Cauchy horizons.

Proposition 2.19. *Let S be an achronal subset of M. Then the following hold.*

(1) $S \subset D^+(S)$.

(2) If $p \in D^+(S)$ *then* $I^-(p) \cap I^+(S) \subset D^+(S)$.

(3) $\partial D^+(S) = H^+(S) \cup S$, *and* $\partial D(S) = H(S)$.

(4) $H^+(S)$ *is achronal.*

(5) $\operatorname{edge} H^+(S) \subset \operatorname{edge} \mathrm{S}$, *with equality holding if* S *is closed.*

The achronality of $H^+(S)$ follows almost immediately from the definition: Suppose $p, q \in H^+(S)$ with $p \ll q$. Since $q \in \overline{D^+(S)}$, and $I^+(p)$ is a neighborhood of q, $I^+(p)$ meets $D^+(S)$, contradicting the definition of $H^+(S)$.

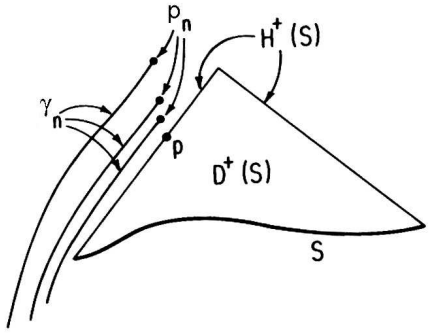
Cauchy horizons have structural properties similar to achronal boundaries, as indicated in the next two results. From Proposition 2.9 and Parts (4) and (5) of Proposition 2.19, we obtain the following.

Proposition 2.20. *Let $S \subset M$ be achronal. Then $H^+(S) \setminus \operatorname{edge} S$, if nonempty, is an achronal C^0 hypersurface in M.*

In a similar vein to Proposition 2.10, we have the following.

Proposition 2.21. *Let S be an achronal subset of M. Then $H^+(S)$ is ruled by null geodesics, i.e., every point of $H^+(S) \setminus \operatorname{edge} S$ is the future endpoint of a null geodesic in $H^+(S)$ which is either past inextendible in M or else has a past end point on* $\operatorname{edge} S$.

Comments on the proof. The proof uses a limit curve argument. Consider the case $p \in H^+(S) \setminus S$. Since $I^+(p) \cap D^+(S) = \emptyset$, we can find a sequence of points $p_n \notin D^+(S)$, such that $p_n \to p$. For each n, there exists a past inextendible causal curve γ_n that does not meet S. By the limit curve lemma there exists a subsequence γ_m that converges to a past inextendible C^0 causal curve γ starting at p. Near p this defines the desired null geodesic (see the figure).



The case $p \in S \setminus \operatorname{edge} S$ is handled somewhat differently; for details see [28, p. 203]. □

The following basic result ties domains of dependence to global hyperbolicity.

Proposition 2.22. *Let $S \subset M$ be achronal.*

(1) Strong causality holds on $\operatorname{int} D(S)$.

(2) Internal compactness holds on $\operatorname{int} D(S)$*, i.e., for all* $p, q \in D(S)$*,* $J^+(p) \cap J^-(p)$ *is compact.*

Comments on the proof. With regard to (1), first observe that the chronology condition holds on $D(S)$. For instance, suppose there exists a timelike curve γ passing through $p \in D^+(S)$, and take γ to be past directed. By repeating loops we obtain a past inextendible timelike curve $\tilde{\gamma}$, which hence must meet S. In fact, it will meet S infinitely often, thereby violating the achronality of S. A similar argument shows that the causality condition holds on $\operatorname{int} D(S)$. Suppose for example that γ is a past directed closed causal curve through $p \in \operatorname{int} D^+(S)$. By repeating loops we obtain a past inextendible causal curve $\tilde{\gamma}$ starting at p. Thus $\tilde{\gamma}$ meets S, and since $p \in \operatorname{int} D^+(S)$, will enter $I^-(S)$. This again leads to an achronality violation. By more refined arguments, using the limit curve lemma, one can show that strong causality holds on $\operatorname{int} D(S)$. With regard to (2), suppose there exist $p, q \in \operatorname{int} D(S)$, such that $J^+(p) \cap J^-(p)$ is noncompact. We want to show that every sequence q_n in $J^+(p) \cap J^-(p)$ has a convergent subsequence. Without loss of generality we may assume $\{q_n\} \subset D^-(S)$. For each n, let γ_n be a future directed causal curve from p to q passing through q_n. As usual, extend each γ_n to a future inextendible causal curve $\tilde{\gamma}_n$. By the limit curve lemma, there exists a subsequence $\tilde{\gamma}_m$ that converges to a future inextendible C^0 causal curve γ starting at p. One can then show that either the sequence of points q_m converges or γ does not enter $I^+(S)$. □

Putting several previous results together we obtain the following.

Proposition 2.23. *Let S be an achronal subset of a spacetime M. Then, S is a Cauchy surface for M iff $D(S) = M$ iff $H(S) = \emptyset$. Hence, if S is a Cauchy surface for M then M is globally hyperbolic.*

Proof. Exercise.

We now give a proof of Proposition 2.18 from the previous subsection.

Proof of Proposition 2.18. It suffices to show that $H(S) = H^+(S) \cup H^-(S) = \emptyset$. Suppose there exists $p \in H^+(S)$. Since S is edgeless, it follows from Proposition 2.21 that p is the future endpoint of a past inextendible null geodesic $\gamma \subset H^+(S)$. Then since $\gamma \subset D^+(S) \cap J^-(p)$ (exercise: show this), we have that γ is contained in the set $J^+(S) \cap J^-(p)$, which is compact by Proposition 2.14). By Lemma 2.13 strong causality must be violated at some point of $J^+(S) \cap J^-(p)$. Thus $H^+(S) = \emptyset$, and time-dually, $H^-(S) = \emptyset$.

We conclude this subsection by stating several lemmas that are useful in proving some of the results described here, as well as other results concerning domains of dependence.

Lemma 2.24 ([22], p. 416)**.** *Let γ be a past inextendible causal curve starting at p that does not meet a closed set C. If $p_0 \in I^+(p, M \setminus C)$, there exists a past inextendible* timelike *curve starting at p_0 that does not meet C.*

Proof. Exercise.

Lemma 2.25. *Let S be achronal. If $p \in \operatorname{int} D^+(S)$ then every past inextendible causal curve from p enters $I^-(S)$.*

Proof. This follows from the *proof* of the preceding lemma.

Lemma 2.26. *Let S be achronal. Then $p \in \overline{D^+(S)}$ iff every past inextendible* timelike *curve meets S.*

Proof. Exercise.

3 The geometry of null hypersurfaces

A smooth submanifold V of a spacetime $(M, \langle\ ,\ \rangle)$ is said to be spacelike (resp, timelike, null) if each of its tangent spaces T_pV, $p \in V$, is spacelike (resp., timelike, null). Hence if V is spacelike (resp., timelike) then, with respect to its *induced metric*, i.e., the metric $\langle\ ,\ \rangle$ restricted to the tangent spaces of V, V is a Riemannian (resp., Lorentzian) manifold. On the other hand, if V is a null submanifold then $\langle\ ,\ \rangle$ is degenerate when restricted to the tangent spaces of V, and so does not define a pseudo-Riemannian metric on V. Nontheless, null hypersurfaces have an interesting geometry, and play an important role general relativity, as they represent *horizons* of various sorts.

Let S be a smooth null hypersurface in a spacetime $(M, \langle\ ,\ \rangle)$. Thus, S is a smooth co-dimension one submanifold of M, such that at each $p \in M$, $\langle\ ,\ \rangle : T_pS \times T_pS \to \mathbb{R}$ is degenerate. This means that there exists a nonzero vector $K_p \in T_pS$ such that

$$\langle K_p, X\rangle = 0 \quad \text{for all} \quad X \in T_pS$$

In particular,

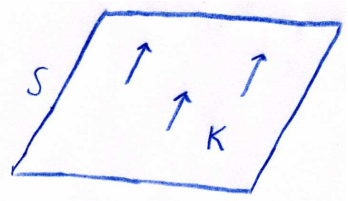
(1) K_p is a null vector, $\langle K_p, K_p\rangle = 0$, which we can choose to be future pointing, and

(2) $[K_p]^\perp = T_pS$.

(3) Moreover, every vector $X \in T_pS$ that is not a multiple of K_p is spacelike.

Thus, every null hypersurface S gives rise to a future directed null vector field K,

$$p \in S \xrightarrow{K} K_p \in T_pS,$$



which will be smooth, $K \in \mathfrak{X}(S)$, provided it is normalized in a suitably uniform way. Furthermore, the null vector field K is unique up to a positive pointwise scale factor.

As simple examples, in Minkowski space $\mathbb{M}^{n+1}$, the past and future cones, $\partial I^-(p)$ and $\partial I^+(p)$, respectively, are smooth null hypersurfaces away from the vertex p. Each nonzero null vector $X \in T_p\mathbb{M}^{n+1}$ determines a null hyperplane $\Pi = \{q \in \mathbb{M}^{n+1} : \langle \overline{pq}, X\rangle = 0$.

The following fact is fundamental.

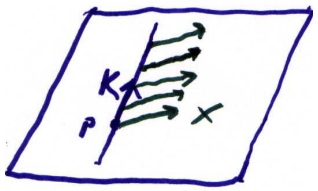
Proposition 3.1. *Let S be a smooth null hypersurface and let $K \in \mathfrak{X}(S)$ be a smooth future directed null vector field on S. Then the integral curves of K are null geodesics (when suitably parameterized),*

Remark: The integral curves of K are called the null generators of S. Apart from parameterziations, the null generators are intrinsic to the null hypersurface.

Proof. It suffices to show that $\nabla_K K = \lambda K$, for then the integral curves are in general *pre-geodesics* (i.e., are geodesics after a suitable reparameterization). To show this it suffice to show that at each $p \in S$, $\nabla_K K \perp T_pS$, i.e., $\langle \nabla_K K, X\rangle = 0$ for all $X \in T_pS$.

Extend $X \in T_pS$ by making it invariant under the flow generated by K,

$$[K, X] = \nabla_K X - \nabla_X K = 0$$



X remains tangent to S, so along the flow line through p, $\langle K, X\rangle = 0$. Differentiating we obtain,

$$0 = K\langle K, X\rangle = \langle \nabla_K K, X\rangle + \langle K, \nabla_K X\rangle\,,$$

and hence,

$$\langle \nabla_K K, X\rangle = -\langle K, \nabla_K X\rangle = -\langle K, \nabla_X K\rangle = -\frac{1}{2} X\langle K, K\rangle = 0\,.$$

□

To study the 'shape' of the null hypersurface S we study how the null vector field K varies along S. Since K is actually orthogonal to S, this is somewhat analogous

to how we study the shape of a hypersurface in a Riemannian manifold, or spacelike hypersurface in a Lorentzian manifold, by introducing the shape operator (or Weingarten map) and associated second fundamental form. We proceed to introduce null analogues of these objects. For technical reasons one works "mod K", as described below.

We introduce the following equivalence relation on tangent vectors: For $X, X' \in T_pS$,

$$X' = X \text{ mod } K \quad \text{if and only if} \quad X' - X = \lambda K \text{ for some } \lambda \in \mathbb{R}\,.$$

Let $\overline{X}$ denote the equivalence class of X. Let $T_pS/K = \{\overline{X} : X \in T_pS\}$, and $TS/K = \cup_{p\in S} T_pS/K$. TS/K, the mod K tangent bundle of S, is a smooth rank $n-1$ vector bundle over S. This vector bundle does not depend on the particular choice of null vector field K.

There is a natural positive definite metric h on TS/K induced from $\langle\,,\rangle$: For each $p \in S$, define $h : T_pS/K \times T_pS/K \to \mathbb{R}$ by $h(\overline{X}, \overline{Y}) = \langle X, Y\rangle$. A simple computation shows that h is well-defined: If $X' = X \text{ mod } K$, $Y' = Y \text{ mod } K$ then

$$\begin{aligned}\langle X', Y'\rangle &= \langle X + \alpha K, Y + \beta K\rangle \\ &= \langle X, Y\rangle + \beta\langle X, K\rangle + \alpha\langle K, Y\rangle + \alpha\beta\langle K, K\rangle \\ &= \langle X, Y\rangle\,.\end{aligned}$$

The *null Weingarten map* $b = b_K$ of S with respect to K is, for each point $p \in S$, a linear map $b : T_pS/K \to T_pS/K$ defined by $b(\overline{X}) = \overline{\nabla_X K}$.

Exercise: Show that b is well-defined. Show also that that if $\widetilde{K} = fK$, $f \in C^\infty(S)$, is any other future directed null vector field on S, then $b_{\widetilde{K}} = f b_K$. It follows that the Weingarten map $b = b_K$ at a point p is uniquely determined by the value of K at p.

Proposition 3.2. *b is self adjoint with respect to h, i.e., $h(b(\overline{X}), \overline{Y}) = h(\overline{X}, b(\overline{Y}))$, for all $\overline{X}, \overline{Y} \in T_pS/K$.*

Proof. Extend $X, Y \in T_pS$ to vector fields tangent to S near p. Using $X\langle K, Y\rangle = 0$ and $Y\langle K, X\rangle = 0$, we obtain,

$$\begin{aligned}h(b(\overline{X}), \overline{Y}) &= \langle \nabla_X K, Y\rangle = -\langle K, \nabla_X Y\rangle = -\langle K, \nabla_Y X\rangle + \langle K, [X, Y]\rangle \\ &= \langle \nabla_Y K, X\rangle = h(\overline{X}, b(\overline{Y}))\,.\end{aligned}$$

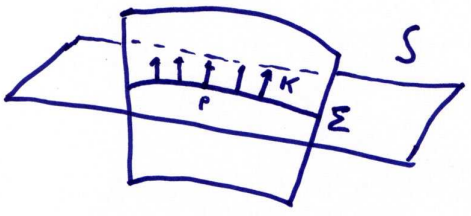
The *null second fundamental form* $B = B_K$ of S with respect to K is the bilinear form associated to b via h: For each $p \in S$, $B : T_pS/K \times T_pS/K \to \mathbb{R}$ is defined by,

$$B(\overline{X}, \overline{Y}) = h(b(\overline{X}), \overline{Y}) = \langle \nabla_X K, Y\rangle\,.$$

Since b is self-adjoint, B is symmetric. We say that S is *totally geodesic* iff $B \equiv 0$. This has the usual geometric meaning: If S is totally geodesic then any geodesic in M starting tangent to S stays in S. This follows from the fact that, when S is

totally geodesic, the restriction to S of the Levi-Civita connection of M defines a linear connection on S. Null hyperplanes in Minkowski space are totally geodesic, as is the event horizon in Schwarzschild spacetime.

The *null mean curvature* (or *null expansion scalar*) of S with respect to K is the smooth scalar field θ on S defined by, $\theta = \operatorname{tr} b$. θ has a natural geometric interpretation. Let Σ be the intersection of S with a hypersurface in M which is transverse to K near $p \in S$; Σ will be a co-dimension two spacelike submanifold of M, along which K is orthogonal.



Let $\{e_1, e_2, \cdots, e_{n-1}\}$ be an orthonormal basis for $T_p\Sigma$ in the induced metric. Then $\{\overline{e}_1, \overline{e}_2, \cdots, \overline{e}_{n-1}\}$ is an orthonormal basis for T_pS/K. Hence at p,

$$\begin{aligned} \theta = \operatorname{tr} b &= \sum_{i=1}^{n-1} h(b(\overline{e}_i), \overline{e}_i) = \sum_{i=1}^{n-1} \langle \nabla_{e_i} K, e_i \rangle. \\ &= \operatorname{div}_\Sigma K\,. \end{aligned} \tag{3.5}$$

where $\operatorname{div}_\Sigma K$ is the divergence of K along Σ. Thus, θ measures the overall expansion of the null generators of S towards the future.

It follows from the exercise on the preceeding page that if $\widetilde{K} = fK$ then $\widetilde{\theta} = f\theta$. Thus the null mean curvature inequalities $\theta > 0$, $\theta < 0$, etc., are invariant under positive rescaling of K. In Minkowski space, a future null cone $S = \partial I^+(p) \setminus \{p\}$ (resp., past null cone $S = \partial I^-(p) \setminus \{p\}$) has positive null mean curvature, $\theta > 0$ (resp., negative null mean curvature, $\theta < 0$).

We now study how the null Weingarten map propagates along the null geodesic generators of S. Let $\eta : I \to M$, $s \to \eta(s)$, be a future directed affinely parameterized null geodesic generator of S. For each $s \in I$, let

$$b(s) = b_{\eta'(s)} : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s) \tag{3.6}$$

be the Weingarten map based at $\eta(s)$ with respect to the null vector $K = \eta'(s)$. We show that the one parameter family of Weingarten maps $s \to b(s)$, obeys a certain Ricatti equation.

We first need to make a few definitions. Let $s \to \mathcal{Y}(s)$ be a TS/K vector field along η, i.e., for each $s \in I$, $\mathcal{Y}(s) \in T_{\eta(s)}S/K$. We say that $s \to \mathcal{Y}(s)$ is smooth if, at least locally, there is a smooth (in the usual sense) vector field $s \to Y(s)$ along η, tangent to S, such that $\mathcal{Y}(s) = \overline{Y(s)}$. Then define the covariant derivative of $s \to \mathcal{Y}(s)$ along η by, $\mathcal{Y}'(s) = \overline{Y'(s)}$, where Y' is the usual covariant differentiation.

Exercise: Show that $\mathcal{Y}'$ is independent of the choice of Y.

Then the covariant derivative of b along η is defined by requiring a natural product rule to hold. If $s \to X(s)$ is a vector field along η tangent to S, b' is defined by,

$$b'(\overline{X}) = b(\overline{X})' - b(\overline{X'})\,. \tag{3.7}$$

Proposition 3.3. *The one parameter family of Weingarten maps $s \to b(s)$, obeys the following Ricatti equation,*

$$b' + b^2 + R = 0\,, \tag{3.8}$$

where $R : T_{\eta(s)}S/\eta'(s) \to T_{\eta(s)}S/\eta'(s)$ is the curvature endomorphism defined by $R(\overline{X}) = \overline{R(X, \eta'(s))\eta'(s)}$.

Proof. Fix a point $p = \eta(s_0)$, $s_0 \in (a, b)$, on η. On a neighborhood U of p in S we can scale the null vector field K so that K is a geodesic vector field, $\nabla_K K = 0$, and so that K, restricted to η, is the velocity vector field to η, i.e., for each s near s_0, $K_{\eta(s)} = \eta'(s)$. Let $X \in T_pM$. Shrinking U if necessary, we can extend X to a smooth vector field on U so that $[X, K] = \nabla_X K - \nabla_K X = 0$. Then,

$$R(X, K)K = \nabla_X \nabla_K K - \nabla_K \nabla_X K - \nabla_{[X,K]} K = -\nabla_K \nabla_K X\,.$$

Hence along η we have, $X'' = -R(X, \eta')\eta'$ (which implies that X, restricted to η, is a *Jacobi field* along η). Thus, from Equation 3.7, at the point p we have,

$$\begin{aligned} b'(\overline{X}) &= \overline{\nabla_X K}\,' - b(\overline{\nabla_K X}) = \overline{\nabla_K X}\,' - b(\overline{\nabla_X K}) \\ &= \overline{X''} - b(b(\overline{X})) = -\overline{R(X, \eta')\eta'} - b^2(\overline{X}) \\ &= -R(\overline{X}) - b^2(\overline{X}), \end{aligned}$$

which establishes Equation 3.8. □

By taking the trace of (3.8) we obtain the following formula for the derivative of the null mean curvature $\theta = \theta(s)$ along η,

$$\theta' = -\mathrm{Ric}(\eta', \eta') - \sigma^2 - \frac{1}{n-1}\theta^2, \tag{3.9}$$

where $\sigma := (\mathrm{tr}\,\hat{b}^2)^{1/2}$ is the *shear scalar*, $\hat{b} := b - \frac{1}{n-1}\theta \cdot \mathrm{id}$ is the trace free part of the Weingarten map, and $\mathrm{Ric}(\eta', \eta') = R_{ij}(\eta^i)'(\eta^j)'$ is the Ricci tensor contracted on the tangent vector η'. Equation 3.9 is known in relativity as the Raychaudhuri equation (for an irrotational null geodesic congruence) . This equation shows how the Ricci curvature of spacetime influences the null mean curvature of a null hypersurface.

The following proposition is a standard application of the Raychaudhuri equation.

Proposition 3.4. *Let M be a spacetime which obeys the null enery condition (NEC), $\mathrm{Ric}\,(X, X) \geq 0$ for all null vectors X, and let S be a smooth null hypersurface in M. If the null generators of S are future geodesically complete then S has nonnegative null mean curvature, $\theta \geq 0$.*

Proof. Suppose $\theta < 0$ at $p \in S$. Let $s \to \eta(s)$ be the null generator of S passing through $p = \eta(0)$, affinely parametrized. Let $b(s) = b_{\eta'(s)}$, and take $\theta = \mathrm{tr}\, b$. By the invariance of sign under scaling, one has $\theta(0) < 0$. Raychaudhuri's equation and the NEC imply that $\theta = \theta(s)$ obeys the inequality,

$$\frac{d\theta}{ds} \leq -\frac{1}{n-1}\theta^2\,, \tag{3.10}$$

and hence $\theta < 0$ for all $s > 0$. Dividing through by θ^2 then gives,

$$\frac{d}{ds}\left(\frac{1}{\theta}\right) \geq \frac{1}{n-1}\,, \tag{3.11}$$

which implies $1/\theta \to 0$, i.e., $\theta \to -\infty$ in finite affine parameter time, contradicting the smoothness of θ. □

Exercise. Let Σ be a local cross section of the null hypersurface S, as depicted on p. 34, with volume form ω. If Σ is moved under flow generated by K, show that $L_K\omega = \theta\,\omega$, where $L =$ Lie derivative.

Thus, Proposition 3.4 implies, under the given assumptions, that cross sections of S are nondecreasing in area as one moves towards the future. Proposition 3.4 is the simplest form of Hawking's black hole area theorem [19]. For a recent study of the area theorem, with a focus on issues of regularity, see [6].

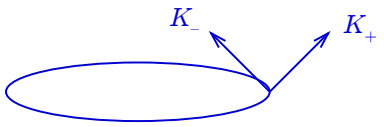
4 Trapped surfaces

In this section we introduce the important notions of trapped and marginally trapped surfaces, which are associated with gravitational collapse and black hole formation. As applications of these notions, we present the classical Penrose singularity theorem and discuss the topology of black holes.

4.1 Trapped and marginally trapped surfaces

Let $(M^{n+1}, \langle\,,\,\rangle)$ be an $(n+1)$-dimensional spacetime, with $n \geq 3$. Let Σ^{n-1} be a compact co-dimension two spacelike submanifold of M. Each normal space of Σ, $[T_p\Sigma]^\perp$, $p \in \Sigma$, is timelike and 2-dimensional, and hence admits two future directed null directions orthogonal to Σ.

Thus, under suitable orientation assumptions, Σ admits two smooth nonvanishing future directed null normal vector fields K_+ and K_-.



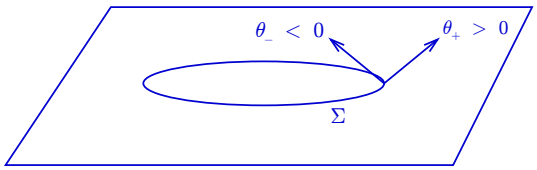
By convention, we refer to K_+ as outward pointing and K_- as inward pointing.

Let S_+ and S_- be the null hypersurfaces, defined and smooth near Σ, generated by the null geodesics with initial tangents K_+ and K_-, respectively. Let θ_+ (resp., θ_-) be the null expansion of S_+ (resp., S_-) *restricted to* Σ. Thus, as in Equation 3.5

$$\theta_+ = \mathrm{div}_\Sigma K_+ \quad \text{and} \quad \theta_- = \mathrm{div}_\Sigma K_- \,.$$

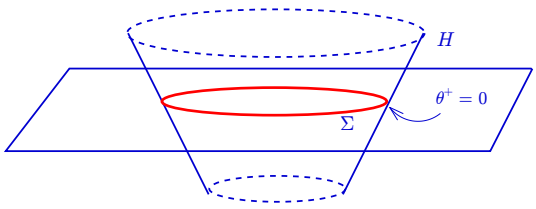
Hence, θ_+ and θ_- are smooth scalars on Σ that measure the overall expansion of the outward going and inward going light rays, respectively, emanating from Σ.

For round spheres in Euclidean slices of Minkowski space, with the obvious choice of inside and outside, one has $\theta_- < 0$ and $\theta_+ > 0$.

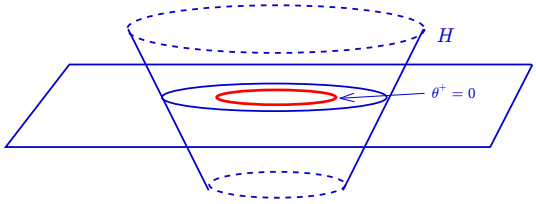


In fact, this is the case in general for large "radial" spheres in *asymptotically flat* spacelike hypersurfaces. However, in regions of spacetime where the gravitational field is strong, one may have both $\theta_- < 0$ and $\theta_+ < 0$, in which case Σ is called a *trapped surface*. As discussed in the following subsection, under appropriate energy and causality conditions, the occurrence of a trapped surface signals the onset of gravitational collapse [24].

Focussing attention on just the outward null normal, we say that Σ is an outer trapped surface if $\theta_+ < 0$, and is a marginally outer trapped surface (MOTS) if $\theta_+ = 0$. MOTSs arise in a number of natural situations. For example, compact cross sections of the event horizon in *stationary* (steady state) black hole spacetimes are MOTSs. (Recall, from Proposition 3.4, that in general one has $\theta \geq 0$ on the event horizon, but in the steady state limit this goes to zero.)



For dynamical black hole spacetimes, MOTS typically occur in the black hole region, i.e., the region inside the event horizon.



While there are heuristic arguments for the existence of MOTSs in this situation, based on looking at the boundary of the 'trapped region' [19, 28] within a given spacelike slice, a recent result of Schoen [27] rigorously establishes their existence under natural conditions.

There has been a lot of recent work done concerning properties of MOTSs. In large measure, this is due to renewed interest in quasi-local notions of black holes, such as *dynamical horizons* [2], and to connections between MOTSs in spacetime and *minimal surfaces* in Riemannian manifolds. In fact, if Σ^{n-1} is a hypersurface in a *time-symmetric* (i.e, totally geodesic) spacelike hypersurface V^n, then, with K^+ suitably normalized, $\theta^+ = H$, where H is the mean curvature of Σ within V. Thus, a MOTS contained in a totally geodesic spacelike hypersurface $V^n \subset M^{n+1}$ is simply a minimal hypersurface in V. Despite the absence of a variational characterization of MOTs like that for minimal surfaces, MOTS have been shown to satisfy a number of properties analogous to those of minimal surfaces. As a case in point, in Subsection 4.3 we describe recent work with Rick Schoen [14], in which we generalize to higher dimensions a classical theorem of Hawking on the topology of black holes.

4.2 The Penrose singularity theorem

The Penrose singularity theorem [24] is the first of the famous singularity theorems of general relativity. The singularity theorems establish, under generic circumstances, the existence in spacetime of incomplete timelike or null geodesics. Such incompleteness indicates that spacetime has come to an end either in the past or future. In specific models past incompleteness is typically associated with a "big bang" beginning of the universe, and future incompleteness is typically associated with a "big crunch" (time dual of the big bang), or, of a more local nature, gravitational collapse to a black hole. The Penrose singularity theorem is associated with the latter.

All the classical singularity theorems require *energy conditions*. The Penrose singularity theorem requires that the null energy condition (NEC) holds, namely that $\mathrm{Ric}(X, X) \geq 0$ for all null vectors X. If a spacetime M satisfies the Einstein equations (1.4), then one can express the NEC in terms of the energy momentum tensor: M obeys the NEC iff $T_{ij}X^iX^j \geq 0$ for all null vectors X.

In studying an *isolated* gravitating system, such as the collapse of a star and formation of a black hole, it is customary to model this situation by a spacetime which is asymptotically flat (i.e., asymptotically Minkowskian). In this context, the assumption of the Penrose singularity theorem that spacetime admit a noncompact Cauchy surface is natural.

The key concept introduced by Penrose in this singularity theorem is that of the trapped surface (discussed in the previous subsection). What Penrose proved is that once the gravitational field becomes sufficiently strong that trapped surfaces appear (as they do in the Schwarzschild solution) then the development of singularities is inevitable.

Theorem 4.1. *Let M be a globally hyperbolic spacetime with noncompact Cauchy surfaces satisfying the NEC. If M contains a trapped surface Σ then M is future null geodesically incomplete.*

Proof. Suppose that M is future null geodesically complete. We show that the achronal boundary $\partial I^+(\Sigma)$ is compact. Since $\partial I^+(\Sigma)$ is closed, if $\partial I^+(\Sigma)$ is noncompact, there exists a sequence of points $\{q_n\} \subset \partial I^+(\Sigma)$ that *diverges to infinity* in M, i.e., that does not have a convergent subsequence in M. Since, by Proposition 2.14, $J^+(\Sigma)$ is closed, we have,

$$\partial I^+(\Sigma) = \partial J^+(\Sigma) = J^+(\Sigma) \setminus I^+(\Sigma)\,. \tag{4.12}$$

Hence, by Proposition 2.4, there exists a future directed null geodesic $\eta_n; [0, a_n] \to M$ from some point $p_n \in \Sigma$ to q_n, which is contained in $\partial I^+(\Sigma)$. In particular, η_n must meet Σ orthogonally at p_n (otherwise $q_n \in I^+(\Sigma)$). By passing to a subsequence if necessary, we may assume that each η_n is 'outward pointing' ($\eta_n'(0) = K^+_{p_n}$).

Since Σ is compact there exists a subsequence $\{p_m\}$ of $\{p_n\}$, such that $p_m \to p \in \Sigma$. It follows that the sequence $\{\eta_m\}$ converges in the sense of geodesics to a future complete outward pointing normal null geodesic $\eta : [0, \infty) \to M$, starting at p, which is contained in $\partial I^+(\Sigma)$. By Equation (4.12), there can be no timelike curve from a point of Σ to a point of η. This implies that no outward pointing null normal geodesic can meet η, for they would have to meet in a corner. A point further out on η would then be timelike related to Σ. On similar grounds, there can be no *null focal point* to Σ along η, i.e., no point on η where nearby outward pointing null normal geodesics cross η "to first order" ([22, Prop. 48, p. 296]). This implies that the exponential map, restricted to the null normal bundle of Σ, is nonsingular along η (see [22], Prop. 30, p. 283 and Cor. 40, p. 290). It follows that for any $a > 0$, the segment $\eta|_{[0,a]}$, is contained in a smooth null hypersurface S, generated by the outward pointing null normal geodesics emanating from a sufficiently small neighborhood of p in Σ. Since Σ is a trapped surface, $\theta^+(p) < 0$. Choose $a > \frac{n-1}{|\theta^+(p)|}$.

Let $s \to \theta(s)$ be the null mean curvature of S along η. By assumption, $\theta(0) = \theta^+(p) < 0$. As in the proof of Proposition 3.4, the Raychaudhuri equation (3.9) and the NEC imply the differential inequality (3.11), from which it follows that $\theta \to -\infty$

in an affine parameter time $\leq \frac{n-1}{|\theta^+(p)|} < a$, contradicting the smoothness of S in a neighborhood of $\eta|_{[0,a]}$.

Thus we have shown that if M is future null geodesically complete then $\partial I^+(\Sigma)$ is compact. It now follows from Propositions 2.7 and 2.18 that $\partial I^+(\Sigma)$ is a *compact* Cauchy surface for M, contrary assumption. □

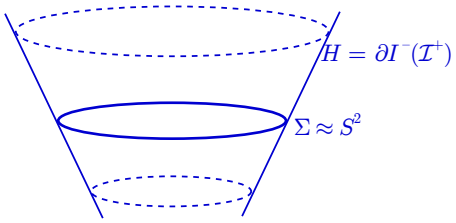
4.3 The topology of black holes

One of the remarkable achievements of the mathematical theory of black holes is the discovery and proof of the black hole uniqueness theorems - the so-called 'no hair theorems'. The basic version asserts that every $3+1$-dimensional asymptotically flat stationary black hole spacetime solving the vacuum Einstein equations is uniquely determined by its mass M and angular momentum J, and in fact must be the Kerr black hole solution for the given M and J. Thus, regardless of the nature of the collapse of two disparate stellar objects, the resulting steady state configuration will be the same, provided the mass and angular momentum are the same.

Recent developments in physics inspired by string theory (e.g., the conjectured AdS/CFT correspondence, braneworld scenarios, etc.) have increased interest in the study of black holes in higher dimensions. In fact there has been a great deal of activity in this area in recent years. One of the first questions to be addressed was: *Does black hole uniqueness hold in higher dimensions?* As it turns out, it does not. In fact, one does not even have topological uniqueness, as we now explain.

A basic step in the proof of the uniqueness of the Kerr solution in $3+1$ dimensions is Hawking's black hole topology theorem.

Theorem 4.2 (Hawking's black hole topology theorem)**.** *Suppose M is a $3+1$-dimensional AF stationary black hole spacetime obeying the dominant energy condition (DEC). Then cross sections of the event horizon are topologically 2-spheres.*



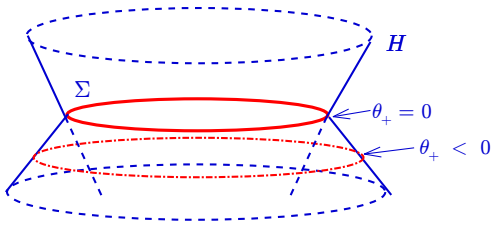
Remark: Let M be a spacetime that satisfies the Einstein equations (1.4) with $\Lambda = 0$. Then we say M obeys the DEC if $\mathcal{T}(X,Y) = T_{ij}X^iY^j \geq 0$ for all future directed causal vectors X, Y.

With impetus coming from the development of string theory, Myers and Perry in a 1986 paper [21] constructed a natural higher dimensional generalization of the Kerr solution, which, in particular, has spherical horizon topology. Perhaps one might have expected black hole uniqueness to extend to these higher dimensional models.

But any such expectations were quelled by the remarkable example of Emparan and Reall [7], published in 2002, of a $4+1$-dimensional AF vacuum stationary black hole spacetime with horizon topology $S^2 \times S^1$, the so-called"black ring".

The question then naturally arises as to what, if any, are the restrictions on the topology of black holes in higher dimensions. This was addressed in a recent paper with Rick Schoen [14], which I would like to describe here. We obtained a natural generalization of Hawking's black hole topology theorem to higher dimensions. Our result implies many well-known restrictions on the topology, some of which we shall review here.

I want to recall briefly the idea behind Hawking's proof of Theorem 4.2. The proof is variational in nature. As in the following figure,



let Σ be a cross section of the event horizon H. Thus Σ is a co-dimension two compact spacelike submanifold contained in H. The null generators of H are orthogonal to Σ at points of intersection. Since the spacetime is stationary, the null generators have vanishing expansion. It follows that Σ is a MOTS, $\theta_+ = 0$.

If Σ is not topologically a 2-sphere, i.e., if it has genus $g \geq 1$ then using the Gauss-Bonnet theorem and the DEC, Hawking shows how to deform Σ along a past null hypersurface to a strictly outer trapped surface, $\theta_+ < 0$, outside the black hole region. But the existence of an outer trapped surface outside the black hole region is forbidden by standard results.

Remarks:

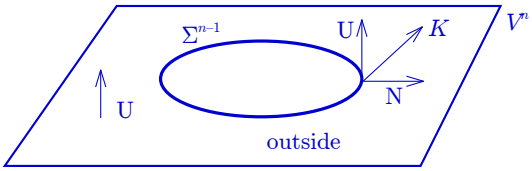
(1) Actually, the torus ($g = 1$) is borderline for Hawking's argument. But this can occur only under special circumstances.

(2) Hawking showed by a variation of his original argument, that the conclusion of his theorem also holds for 'outer apparent horizons' in black hole spacetimes that are not necessarily stationary. This will be the context for the generalization of Hawking's theorem described below.

(3) In higher dimensions, one cannot appeal to the Gauss-Bonnet theorem. This is one of the complicating the issues.

We now present a generalization of Hawking's black hole topology theorem. Let V^n be an n-dimensional, $n \geq 3$, spacelike hypersurface in a spacetime $(M^{n+1}, \langle\,,\,\rangle)$.

Let Σ^{n-1} be a closed hypersurface in V^n, and assume that Σ^{n-1} separates V^n into an "inside" and an "outside". Let N be the outward unit normal to Σ^{n-1} in V^n, and let U be the future directed unit normal to V^n in M^{n+1}. Then $K = U + N$ is an outward null normal field to Σ^{n-1}, unique up to scaling.



We shall say Σ^{n-1} is an *outer apparent horizon* in V^n provided, (i) Σ is marginally outer trapped, i.e., $\theta = 0$, and (ii) there are no outer trapped surfaces outside of Σ in V homologous to Σ. Heuristically, Σ is the "outer limit" of outer trapped surfaces in V. Note that any cross section of the event in a *stationary* black hole spacetime arising from the intersection with a spacelike hypersurface V is necessarily an outer apparent horizon in V.

Theorem 4.3 ([14]). *Let $(M^{n+1}, \langle\,,\,\rangle)$, $n \geq 3$, be a spacetime satisfying the dominant energy condition. If Σ^{n-1} is an outer apparent horizon in V^n then Σ^{n-1} is of* **positive Yamabe type***, i.e., admits a metric of positive scalar curvature, unless Σ^{n-1} is Ricci flat (flat if $n = 3, 4$) in the induced metric, and both B and $\mathcal{T}(U, K) = T_{ab}U^aK^b$ vanish on Σ.*

Theorem 4.3 may be viewed as a spacetime analogue of earlier results of Schoen and Yau [26] concerning minimal hypersurfaces in manifolds of positive scalar curvature.

Theorem 4.3 says that, apart from certain exceptional circumstances, Σ is of positive Yamabe type. This implies many well-known restrictions on the topology. Assume for the discussion that Σ is orientable.

In the standard case: $\dim \Sigma = 2$ ($\dim M = 3 + 1$), Σ admits a metric of positive Gaussian curvature, so $\Sigma \approx S^2$ by Gauss-Bonnet, and hence one recovers Hawking's theorem.

Let's now focus on the case $\dim \Sigma = 3$ ($\dim M = 4 + 1$). If Σ is positive Yamabe then by well-known results of Schoen-Yau [26] and Gromov-Lawson [18] we know that, Σ must be diffeomorphic to

(1) a spherical space (i.e., a homotopy 3-sphere, perhaps with identifications) or,

(2) $S^2 \times S^1$, or

(3) a connected sum of the above two types.

This topological conclusion may be understood as follows. By the prime decomposition theorem, Σ can be expressed as a connected sum of spherical spaces, $S^2 \times S^1$'s,

and $K(\pi,1)$ manifolds (manifolds whose universal covers are contractible). But as Σ admits a metric of positive scalar curvature, it cannot have any $K(\pi,1)$'s in its prime decomposition.

Thus, the basic horizon topologies in $\dim M = 4+1$ are S^3 and $S^2 \times S^1$, both of which are realized by nontrivial black hole spacetimes.

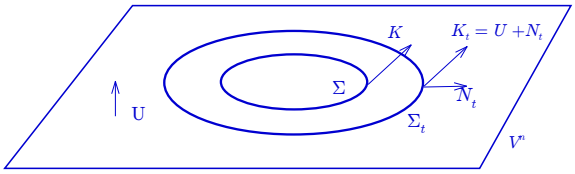
Proof of Theorem 4.3. We consider normal variations of Σ in V, i.e., variations $t \to \Sigma_t$ of $\Sigma = \Sigma_0$ with variation vector field

$$V = \frac{\partial}{\partial t}|_{t=0} = \phi N, \quad \phi \in C^\infty(\Sigma)\,.$$

Let

$$\theta(t) = \text{ the null expansion of } \Sigma_t,$$

where $K_t = U + N_t$ and N_t is the unit normal field to Σ_t in V (see the following figure).



A computation shows [5, 1]

$$\left.\frac{\partial\theta}{\partial t}\right|_{t=0} = L(\phi)\,, \tag{4.13}$$

where,

$$L(\phi) = -\triangle\phi + 2\langle X, \nabla\phi\rangle + \left(Q + \operatorname{div} X - |X|^2\right)\phi\,, \tag{4.14}$$

$$Q = \frac{1}{2}S - \mathcal{T}(U,K) - \frac{1}{2}|b|^2, \qquad X = \tan\left(\nabla_N U\right), \tag{4.15}$$

S is the scalar curvature of Σ, and $\langle\,,\,\rangle$ now denotes the the induced metric on Σ.

L is a second order linear elliptic operator, associated with variations in θ, analogous to the stability operator of minimal surface theory [1]. In fact, in the time-symmetric case (V totally geodesic) the vector field X vanishes and L reduces to the stability operator of minimal surface theory. Note, however, that L is *not* in general self-adjoint (with respect to the standard L^2 inner product on Σ).

Although L is not in general self adjoint, its principal eigenvalue (eigenvalue with smallest real part) $\lambda_1(L)$ is real, and one can choose a principal eigenfunction ϕ which is strictly positive, $\phi > 0$. Using the eigenfunction ϕ to define our variation, we have from (4.13),

$$\left.\frac{\partial\theta}{\partial t}\right|_{t=0} = \lambda_1(L)\phi\,. \tag{4.16}$$

The eigenvalue $\lambda_1(L)$ cannot be negative, for otherwise (4.13) would imply that $\frac{\partial\theta}{\partial t} < 0$ on Σ. Since $\theta = 0$ on Σ, this would mean that for $t > 0$ sufficiently small, Σ_t would be outer trapped, contrary to our assumptions. Hence, $\lambda_1(L) \geq 0$.

Now consider the "symmetrized operator",

$$L_0(\phi) = -\triangle\phi + Q\,\phi\,, \tag{4.17}$$

obtained formally by setting $X = 0$ in (4.14)

The following claim is the heart of the proof.

Claim: $\lambda_1(L_0) \geq \lambda_1(L)$. Hence, $\lambda_1(L_0) \geq 0$.

Proof of the claim. Completing the square on the right hand side of (4.14), and using $L(\phi) = \lambda_1(L)\phi$ gives,

$$-\triangle\phi + (Q + \operatorname{div} X)\,\phi + \phi|\nabla \ln\phi|^2 - \phi|X - \nabla\ln\phi|^2 = \lambda_1(L)\phi \tag{4.18}$$

Setting $u = \ln\phi$, we obtain,

$$-\triangle u + Q + \operatorname{div} X - |X - \nabla u|^2 = \lambda_1(L)\,. \tag{4.19}$$

Absorbing the Laplacian term $\triangle u = \operatorname{div}(\nabla u)$ into the divergence term gives,

$$Q + \operatorname{div}(X - \nabla u) - |X - \nabla u|^2 = \lambda_1(L). \tag{4.20}$$

Setting $Y = X - \nabla u$, we arrive at,

$$-Q + |Y|^2 + \lambda_1(L) = \operatorname{div} Y\,. \tag{4.21}$$

Given any $\psi \in C^\infty(\Sigma)$, we multiply through by ψ^2 and derive,

$$\begin{aligned} -\psi^2 Q + \psi^2|Y|^2 + \psi^2\lambda_1(L) &= \psi^2 \operatorname{div} Y \\ &= \operatorname{div}(\psi^2 Y) - 2\psi\langle\nabla\psi, Y\rangle \\ &\leq \operatorname{div}(\psi^2 Y) + 2|\psi||\nabla\psi||Y| \\ &\leq \operatorname{div}(\psi^2 Y) + |\nabla\psi|^2 + \psi^2|Y|^2\,. \end{aligned}$$

Integrating the above inequality yields,

$$\lambda_1(L) \leq \frac{\int_\Sigma |\nabla\psi|^2 + Q\psi^2}{\int_\Sigma \psi^2} \quad \text{for all } \psi \in C^\infty(\Sigma),\ \psi \not\equiv 0\,. \tag{4.22}$$

The claim now follows from the well-known Rayleigh formula for the principal eigenvalue applied to the operator (4.17). □

Thus, we have that $\lambda_1(L_0) \geq 0$ for the operator (4.17), where Q is given in (4.15). We have *in effect* reduced the situation to the time-symmetric (or Riemannian) case, where standard arguments become applicable.

Let $f \in C^\infty(\Sigma)$ be an eigenfunction associated to $\lambda_1(L_0)$; f can be chosen to be strictly positive. Consider Σ in the conformally related metric $\tilde{h} = f^{2/n-2}h$, where h is the induced metric. The scalar curvature $\tilde{S}$ of Σ in the metric $\tilde{h}$ is given by,

$$\begin{aligned}\tilde{S} &= f^{-n/(n-2)}\left(-2\triangle f + Sf + \frac{n-1}{n-2}\frac{|\nabla f|^2}{f}\right) \\ &= f^{-2/(n-2)}\left(2\lambda_1(L_0) + 2\mathcal{T}(U,K) + |B|^2 + \frac{n-1}{n-2}\frac{|\nabla f|^2}{f^2}\right), \qquad (4.23)\end{aligned}$$

where, in the second equation, we have used (4.17), with $\phi = f$, and (4.15).

Since all terms in the parentheses above are nonnegative, (4.23) implies that $\tilde{S} \geq 0$. If $\tilde{S} > 0$ at some point, then by well known results [20] one can conformally rescale $\tilde{h}$ to a metric of strictly positive scalar curvature. If, on the other hand, $\tilde{S}$ vanishes identically, then (4.23) implies: $\lambda_1(L_0) = 0$, $\mathcal{T}(U,K) \equiv 0$, $B \equiv 0$ and f is constant. Equations (4.17) and (4.15) then imply that $S \equiv 0$. One can then deform h in the direction of the Ricci tensor of Σ to obtain a metric of positive scalar curvature, unless (Σ, h) is Ricci flat (see [20]). □

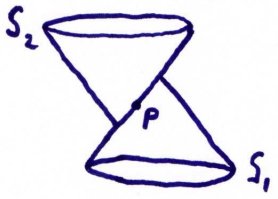
Remark: A drawback of Theorem 4.3 is that it allows certain 'exceptional circumstances'. For example note that Theorem 4.3 does not rule out the possibility of a vacuum black hole spacetime with toroidal horizon topology. More recently, we have succeded in ruling out these exceptional cases in a number of natural situations; see [13].

5 The null splitting theorem

5.1 Maximum principle for null hypersurfaces

There is a well-known geometric maximum principle for hypersurfaces in Riemannian geometry and spacelike hypersurfaces in Lorentzian geometry, which extends to null hypersurfaces. This maximum principle for null hypersurfaces, which we would now like to discuss, is a key ingredient in the proof of the null splitting theorem.

Consider two null hypersurfaces S_1 and S_2 in spacetime meeting tangentially at a point p, with S_2 to the future of S_1.



Because S_2 lies to the 'future side' of S_1, we must have (assuming a compatible scaling) $\theta_2 \geq \theta_1$ at p, where θ_i is the null mean curvature of S_i, $i = 1, 2$. The maximum principle for null hypersurfaces examines what happens when the reverse inequalities hold.

Theorem 5.1. *Let S_1 and S_2 be smooth null hypersurfaces in a spacetime M. Suppose,*

(1) S_1 and S_2 meet at $p \in M$ and S_2 lies to the future side of S_1 near p, and

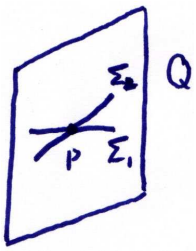
(2) the null mean curvature scalars θ_1 of S_1, and θ_2 of S_2, satisfy, $\theta_2 \leq 0 \leq \theta_1$.

Then S_1 and S_2 coincide near p and this common null hypersurface has null mean curvature $\theta = 0$.

The heuristic idea is that since the generators of S_1 are (weakly) diverging, and the generators of S_2, which lie to the future of S_1 are (weakly) converging, the two sets of generators are forced to agree and form a nonexpanding congruence.

Proof. We give a sketch of the proof; for details, see [10]. S_1 and S_2 have a common null direction at p. Let Q be a timelike hypersurface in M passing through p and transverse to this direction. By taking Q small enough, the intersections,

$$\Sigma_1 = S_1 \cap Q \quad \text{and} \quad \Sigma_2 = S_2 \cap Q$$



will be smooth spacelike hypersurfaces in Q, with Σ_2 to the future side of Σ_1 near p.

Σ_1 and Σ_2 may be expressed as graphs over a fixed spacelike hypersurface V in Q (with respect to normal coordinates around V), $\Sigma_1 = \text{ graph}\,(u_1)$, $\Sigma_2 = \text{ graph}\,(u_2)$. Let,

$$\theta(u_i) = \theta_i|_{\Sigma_i = \text{ graph}\,(u_i)}\,, \quad i = 1, 2\,.$$

By suitably normalizing the null vector fields $K_1 \in \mathfrak{X}(S_1)$ and $K_2 \in \mathfrak{X}(S_2)$ determining θ_1 and θ_2, respectively, a computation shows,

$$\theta(u_i) = H(u_i) + \text{ lower order terms}\,,$$

where H is the mean curvature operator on spacelike graphs over V in Q. (The lower order terms involve the second fundamental form of Q.) Thus θ is a second order quasi-linear elliptic operator. In the present situation we have:

(1) $u_1 \leq u_2$, and $u_1(p) = u_2(p)$.

(2) $\theta(u_2) \leq 0 \leq \theta(u_1)$.

Then Alexandrov's strong maximum principle for second order quasi-linear elliptic PDEs implies that $u_1 = u_2$. Thus, Σ_1 and Σ_2 agree near p. The null normal geodesics to Σ_1 and Σ_2 in M will then also agree. This implies that S_1 and S_2 agree near p. □

The usefulness of Theorem 5.1 is somewhat limited by the fact that the most interesting null hypersurfaces arising in general relativity, e.g., event horizons, Cauchy horizons, and observer horizons, are in general *rough*, i.e., are C^0, but in general not C^1. The key point, however, is that Theorem 5.1 extends to C^0 null hypersurfaces, suitably defined. Roughly, a C^0 null hypersurface is a locally achronal C^0 hypersurface in spacetime that is ruled, in a suitable sense, by null geodesics. The *null portions* of achronal boundaries, $\partial I^{\pm}(S) \setminus S$, are the basic models for C^0 null hypersurfaces (recall Propositions 2.7 and 2.10). Although C^0 null hypersurfaces do not in general have null mean curvature in the classical sense, they, nonetheless may obey null mean curvature inequalities in a certain weak sense, namely in the sense of *support null hypersurfaces* [10, 12]. Thus, the null mean curvature inequalities, $\theta_2 \leq 0 \leq \theta_1$, can hold for C^0 null hypersurfaces in the support sense.

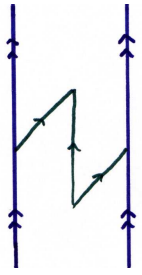
The upshot of these comments is that Theorem 5.1 extends, in an appropriate manner, to C^0 null hypersurfaces; see [10, Theorem III.2]. It is this maximum principle for C^0 null hypersurfaces that is actually needed to prove the null splitting theorem.

5.2 The null splitting theorem

The null splitting theorem is a descendant of the famous Cheeger-Gromoll splitting theorem of Riemannian geometry, and the more recent Lorentzian splitting theorem, its direct Lorentzian analogue. The problem of establishing a Lorentzian analogue of the Cheeger-Gromoll splitting theorem was posed by S.-T. Yau [29] in the early 80's as an approach to establishing the rigidity of the Hawking-Penrose singularity theorems[1], and was solved in a series of papers towards the end of the 80's; see [3, Chapter 14] for a nice treatment.

The Lorentzian splitting theorem is concerned with the structure of spacetimes that admit a *timelike line*, which, by definition, is an inextendible timelike geodesic, each segment of which is maximal. The null splitting theorem is concerned with the structure of spacetimes that admit a *null line*.

By definition a null line is an inextendible null geodesic that is globally achronal, i.e., no two points can be joined by a timelike line. (From the point of view of the Lorentzian distance function, each segment of a null line is maximal.) We know from Proposition 2.2 that null geodesics are locally achronal, but they may not be achronal in the large, even in globally hyperbolic spacetimes. Consider, for example a null geodesic winding around a flat spacetime cylinder (closed in space); eventually points on the null geodesic are timelike related.



[1]Establishing this rigidity remains an important open problem; see, for example the discussion in [3, p. 503ff].

Null lines arise naturally in causal arguments; recall, for example, that sets of the form $\partial I^{\pm}(S) \setminus S$, S closed, are ruled by null geodesics which are necessarily achronal. Null lines have arisen in various situations in general relativity, for example in the proofs of the Hawking-Penrose singularity theorem, topological censorship and certain versions of the positivity of mass.

Every null geodesic in Minkowski space and de Sitter space is a null line. At the same time, both of these spacetimes obey the null energy condition. In general, it is difficult for complete null lines to exist in spacetimes which obey the null energy condition. The null energy condition tends to focus congruences of null geodesics, which can lead to the occurence of *null conjugate points*. But a null geodesic containing a pair of conjugate points cannot be achronal. Thus we expect a spacetime which satisfies the null energy condition and which contains a complete null line to be special in some way, to exhibit some sort of *rigidity*. The null splitting theorem addresses what this rigidity is.

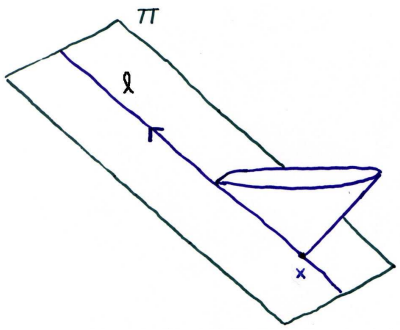
Theorem 5.2 ([10])**.** *Let M be a null geodesically complete spacetime which obeys the NEC. If M admits a null line η then η is contained in a smooth properly embedded achronal* **totally geodesic** *null hypersurface S.*

Recall from Section 3, 'totally geodesic' means that the null second form of S vanishes, $B \equiv 0$, or equivalently that the null expansion and shear, θ and σ, repectively, vanish on S. This implies that the metric h defined on the vector bundle TS/K is invariant under the flow generated by K; it is in this sense that S 'splits'.

The simplest illustration of Theorem 5.2 is Minkowski space: Each null line ℓ in Minkowski space is contained in a unique null hyperplane Π.

Proof. The proof is an application of the maximum principle for C^0 null hypersurfaces. For simplicity we shall assume M is strongly causal; this however is not required; see [10] for details.

By way of motivation, note that the null plane Π in Minkowski space determined by the null line ℓ can be realized as the limit of the future null cone $\partial I^+(x)$ as x goes to past null infinity along the null line ℓ.



Π can also be realized as the limit of the past null cone $\partial I^-(x)$ as x goes to future null infinity along the null line ℓ. In fact, one sees that $\Pi = \partial I^+(\ell) = \partial I^-(\ell)$.

Thus, in the setting of Theorem 5.2, consider the achronal boundaries $S_+ = \partial I^+(\eta)$ and $S_- = \partial I^-(\eta)$. By Proposition 2.7, S_+ and S_- are closed achronal C^0 hypersurfaces in M. Since η is achronal, it follows that S_+ and S_- both contain η. For simplicity, assume S_+ and S_- are connected (otherwise restrict attention to the component of each containing η). The proof then consists of showing that S_+ and S_- agree and form a smooth totally geodesic null hypersurface.

By Proposition 2.10, each point $p \in S_+ \setminus \eta$ lies on a null geodesic $\sigma \subset S_+$ which either is past inextendible in M or else has a past endpoint on η. In the latter case, σ meets η at an angle, and Proposition 2.4 then implies that there is a timelike curve from a point on η to a point on σ, violating the achronality of S_+. Thus, S_+ is ruled by null geodesics which are past inextendible in M, and hence, by the completeness assumption, past complete. In a similar fashion we have that S_- is ruled by null geodesics which are future complete.

Suppose for the moment that S_- and S_+ are smooth null hypersurfaces. Then, by Proposition 3.4 (and its time-dual), S_- and S_+ have null mean curvatures satisfying,

$$\theta_+ \leq 0 \leq \theta_- \,. \tag{5.24}$$

Let q be a point of intersection of S_+ and S_-. S_+ necessarily lies to the future side of S_- near q. We may now apply Theorem 5.1 to conclude that S_+ and S_- agree near q to form a smooth null hypersurface having null mean curvature $\theta = 0$. A fairly straightforward continuation argument shows that $S_+ = S_- = S$ is a smooth null hypersurface with $\theta = 0$. By setting $\theta = 0$ in the Raychaudhuri equation (3.9), and using the NEC, we see that the shear σ must vanish, and hence S is totally geodesic.

In the general case in which S_+ and S_- are merely C^0 null hypersurfaces, one can show that (5.24) holds *in the support sense*. Then the C^0 version of Theorem 5.1 ([10, Theorem III.2]) may be applied to arrive at the same conclusion. □

5.3 An application: Uniqueness of de Sitter space

In this subsection, as an application of the null splitting theorem, we present a uniqueness result for de Sitter space, dS^{n+1}, which is the the simply connected space form of constant positive curvature (which for the purposes of discussion we take to be $+1$). De Sitter space satisfies the vacuum ($T_{ij} = 0$) Einstein equations with cosmological constant $\Lambda = n(n-1)/2$,

$$\mathrm{Ric} = ng \tag{5.25}$$

where $g = \langle\,,\,\rangle$. dS^{n+1} can be explicitly realized as the hyperboloid of one sheet,

$$-(x^0)^2 + \sum_{i=1}^{n+1} (x^i)^2 = 1 \tag{5.26}$$

in $n+2$ dimensional Minkowski space. Introducing spherical type coordinates, dS^{n+1} can be expressed globally as,

$$M = \mathbb{R} \times S^n, \qquad ds^2 = -dt^2 + \cosh^2 t \, d\Omega^2 \,. \tag{5.27}$$

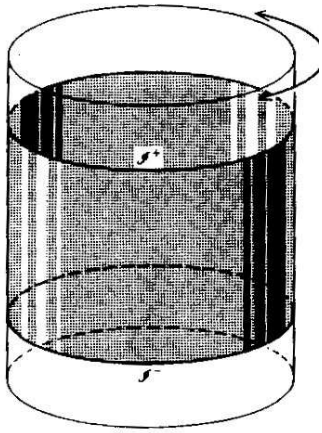
There has been increased interest in recent years in de Sitter space, and spacetimes which are asymptotically de Sitter, due, firstly, to observations supporting an accelerated rate of expansion of the universe, suggesting the presence in our universe of a positive cosmological constant, and due, secondly, to recent efforts to understand quantum gravity on such spacetimes (see [4] and references cited therein).

We use Penrose's notion of conformal infinity [23] to make precise what it means for spacetime to be asymptotically de Sitter. Recall, this notion is based on the way in which the standard Lorentzian space forms, Minkowski space, de Sitter space and anti-de Sitter space, conformally imbed into the Einstein static universe $(\mathbb{R} \times S^n, -du^2 + d\omega^2)$.

Under the transformation $u = \tan^{-1}(e^t) - \pi/4$, the metric (5.27) becomes

$$ds^2 = \frac{1}{\cos^2(2u)}(-du^2 + d\omega^2)\,. \tag{5.28}$$

Thus, de Sitter space conformally imbeds onto the region $-\pi/4 < u < \pi/4$ in the Einstein static universe.



Future conformal infinity $\mathcal{I}^+$ (resp., past conformal infinity $\mathcal{I}^-$) is represented by the *spacelike* slice $u = \pi/4$ (resp., $u = -\pi/4$). This serves to motivate the following definitions.

Definition 5.1. *A spacetime (M, g) is* **asymptotically de Sitter** *provided there exists a spacetime-with-boundary $(\tilde{M}, \tilde{g})$ and a smooth function Ω on $\tilde{M}$ such that*

(1) M is the interior of $\tilde{M}$; hence $\tilde{M} = M \cup \mathcal{I}$, $\mathcal{I} = \partial\tilde{M}$.

(2) $\tilde{g} = \Omega^2 g$, where (i) $\Omega > 0$ on M, and (ii) $\Omega = 0$, $d\Omega \neq 0$ along $\mathcal{I}$.

(3) $\mathcal{I}$ is spacelike.

In general, $\mathcal{I}$ decomposes into two disjoint sets, $\mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^-$ where $\mathcal{I}^+ \subset I^+(M, \tilde{M})$ and $\mathcal{I}^- \subset I^-(M, \tilde{M})$. $\mathcal{I}^+$ is future conformal infinity and $\mathcal{I}^-$ is past conformal infinity.

Definition 5.2. *An asymptotically de Sitter spacetime is* **asymptotically simple** *provided every inextendible null geodesic in M has a future end point on $\mathcal{I}^+$ and a past end point on $\mathcal{I}^-$.*

Thus, an asymptotically de Sitter spacetime is asymptotically simple provided each null geodesic extends to infinity both to the future and the past. In particular, such a spacetime is null geodesically complete.

It is a fact that every inextendible null geodesic in de Sitter space is a null line. As discussed below, this may be understood in terms of the causal structure of de Sitter space. As the following result shows, the occurrence of null lines is a very special feature of de Sitter space among asymptotically simple and de Sitter vacuum spacetimes.

Theorem 5.3 ([11])**.** *Let (M, g) be a 4-dimensional asymptotically simple and de Sitter spacetime satisfying the vacuum Einstein equations (5.25). If M contains a null line then M is isometric to de Sitter space.*

This theorem can be interpreted in terms of the initial value problem in the following way: Friedrich's work [9] on the nonlinear stability of de Sitter space shows that the set of asymptotically simple solutions to the Einstein equations with positive cosmological constant is open in the set of all maximal globally hyperbolic solutions with compact spatial sections. As a consequence, by slightly perturbing the initial data on a fixed Cauchy surface of dS^4 we get in general an asymptotically simple solution of the Einstein equations different from dS^4. Thus, by virtue of theorem 5.3, such a spacetime *has no null lines*. In other words, a small generic perturbation of the initial data destroys *all* null lines. This suggests that the so-called generic condition of singularity theory [19] is in fact generic with respect to perturbations of the initial data.

As discussed in [14], Theorem 5.3 may also be interpreted as saying that no other asymptotically simple and de Sitter solution of the vacuum Einstein equations besides dS^4 develops *eternal observer horizons*. By definition, an observer horizon $\mathcal{A}$ is the past achronal boundary $\partial I^-(\gamma)$ of a future inextendible timelike curve γ, thus $\mathcal{A}$ is ruled by future inextendible achronal null geodesics. In the case of de Sitter space, observer horizons are eternal, that is, all null generators of $\mathcal{A}$ extend from $\mathcal{I}^+$ all the way back to $\mathcal{I}^-$.

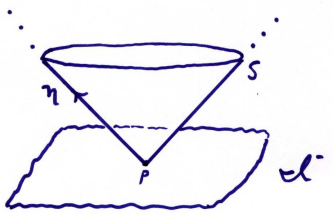
Since the observer horizon $\partial I^-(\gamma)$ is the boundary of the region of spacetime that can be observed by γ, the question arises as to whether at one point γ would be able to observe the whole of space. More precisely, we want to know if there exists $q \in \tilde{M}$ such that $I^-(q)$ would contain a Cauchy surface of spacetime. Gao and Wald [16] were able to answer this question affirmatively for globally hyperbolic spacetimes with compact Cauchy surfaces, assuming null geodesic completeness, the null energy condition and the null generic condition. Thus, as expressed by Bousso [4], asymptotically de Sitter spacetimes satisfying the conditions of the Gao and Wald result, have Penrose diagrams that are "tall" compared to de Sitter space.

Though no set of the form $I^-(q)$ in dS^4 contains a Cauchy surface, $I^-(q)$ gets arbitrarily close to doing so as $q \to \mathcal{I}^+$. However, notice that de Sitter space is not a counterexample to Gao and Wald's result, since dS^4 does not satisfy the null generic condition. Actually, the latter remark leads us to interpret theorem 5.3 as a rigid version of the Gao and Wald result in the asymptotically simple (and vacuum) context: by dropping the null generic hypothesis in [16] the conclusion will only fail if $(\tilde{M}, \tilde{g})$ is isometric to dS^4.

Proof of Theorem 5.3. We present some comments on the proof; see [11, 15] for further details. The main step is to show that M has constant curvature. Since M is Einstein, it is sufficient to show that M is conformally flat.

Let η be the assumed null line in M. By Theorem 5.2, η is contained in a smooth totally geodesic null hypersurface S in M. By asymptotic simplicity, η acquires a past end point p on $\mathcal{I}^-$ and a future end point q on $\mathcal{I}^+$. Let us focus attention on the situation near p. By the proof of Theorem 5.2, and the fact that p is the past end point of η, we have that,

$$S = \partial I^+(\eta) = \partial I^+(p, \tilde{M}) \cap M \,.$$



It follows that $N_p := S \cup \{p\}$ is a smooth null cone in $\tilde{M}$, generated by the future directed null geodesics emanating from p.

From the Riccati equation (3.8), one easily derives a propagation equation for $\hat{b}$, the trace free part of the Weingarten map involving the Weyl conformal tensor (exercise: derive this). But since S is totally geodesic, $\hat{b}$ vanishes identically, and then this propagation equation implies that the components C_{a0b0} of the conformal tensor (with respect to an appropriately chosen pseudo-orthonormal frame in which e_0 is aligned with the generators) vanish on $S = N_p \setminus \{p\}$. An argument of Friedrich [8], based on the conformal field equations, specifically the divergencelessness of the rescaled conformal tensor,

$$\tilde{\nabla}_i d^i_{jkl} = 0, \qquad d^i_{jkl} = \Omega^{-1} C^i_{jkl} \,,$$

in which N_p plays the role of an initial characteristic hypersurface, then shows that the conformal tensor of g vanishes on the future domain of dependence of N_p,

$$C^i_{jkl} = 0 \quad \text{on} \quad D^+(N_p, \tilde{M}) \cap M \,. \tag{5.29}$$

In a time-dual manner one obtains that C^i_{jkl} vanishes on $D^-(N_q, \tilde{M}) \cap M$. Since it can be shown that M is contained in $D^+(N_p, \tilde{M}) \cup D^-(N_q, \tilde{M})$, we conclude that M is conformally flat. Together with equation (5.25), this implies that M has constant curvature $= +1$. Moreover, further global arguments show that M is geodesically

complete and simply connected. It then follows from uniqueness results for Lorentzian space forms that M is isometric to de Sitter space. □

Remark: It has recently been shown that the conclusion of Theorem 5.3 applies under much more general circumstances. The assumption of asymptotic simplicity can be substantially weakened, and one can allow a priori for the presence of certain matter fields; see [15]. The arguments make use of the fact that the null splitting theorem does not require full null geodesic completeness. As the proof of the null splitting theorem shows, if η is the given null line, it is sufficient to require that the generators of $\partial I^-(\eta)$ be future geodesically complete and the generators of $\partial I^+(\eta)$ be past geodesically complete.

References

[1] Lars Andersson, Marc Mars, and Walter Simon, *Local existence of dynamical and trapping horizons*, Phys. Rev. Lett. **95** (2005), 111102.

[2] Abhay Ashtekar and Badri Krishnan, *Dynamical horizons and their properties*, Phys. Rev. D (3) **68** (2003), no. 10, 104030, 25. MR MR2071054 (2005c:83030)

[3] John K. Beem, Paul E. Ehrlich, and Kevin L. Easley, *Global Lorentzian geometry*, second ed., Monographs and Textbooks in Pure and Applied Mathematics, vol. 202, Marcel Dekker Inc., New York, 1996. MR MR1384756 (97f:53100)

[4] Raphael Bousso, *Adventures in de Sitter space*, The future of the theoretical physics and cosmology (Cambridge, 2002), Cambridge Univ. Press, Cambridge, 2003, pp. 539–569. MR MR2033285

[5] Mingliang Cai and Gregory J. Galloway, *On the topology and area of higher-dimensional black holes*, Classical Quantum Gravity **18** (2001), no. 14, 2707–2718. MR MR1846368 (2002k:83051)

[6] P. T. Chruściel, E. Delay, G. J. Galloway, and R. Howard, *Regularity of horizons and the area theorem*, Ann. Henri Poincaré **2** (2001), no. 1, 109–178. MR MR1823836 (2002e:83045)

[7] Roberto Emparan and Harvey S. Reall, *A rotating black ring solution in five dimensions*, Phys. Rev. Lett. **88** (2002), no. 10, 101101, 4. MR MR1901280 (2003e:83060)

[8] Helmut Friedrich, *Existence and structure of past asymptotically simple solutions of Einstein's field equations with positive cosmological constant*, J. Geom. Phys. **3** (1986), no. 1, 101–117. MR MR855572 (88c:83006)

[9] ———, *On the existence of n-geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure*, Comm. Math. Phys. **107** (1986), no. 4, 587–609. MR MR868737 (88b:83006)

[10] Gregory J. Galloway, *Maximum principles for null hypersurfaces and null splitting theorems*, Ann. Henri Poincaré **1** (2000), no. 3, 543–567. MR MR1777311 (2002b:53052)

[11] ———, *Some global results for asymptotically simple space-times*, The conformal structure of space-time, Lecture Notes in Phys., vol. 604, Springer, Berlin, 2002, pp. 51–60. MR MR2007041 (2004k:53105)

[12] ———, *Null geometry and the Einstein equations*, The Einstein equations and the large scale behavior of gravitational fields, Birkhäuser, Basel, 2004, pp. 379–400. MR MR2098922 (2006f:83015)

[13] ———, *Rigidity of outer horizons and the topology of black holes*, (2006), gr–qc/0608118.

[14] Gregory J. Galloway and Richard Schoen, *A generalization of Hawking's black hole topology theorem to higher dimensions*, Comm. Math. Phys. **266** (2006), no. 2, 571–576. MR MR2238889

[15] Gregory J. Galloway and Didier A. Solis, *Uniqueness of de sitter space*, Classical Quantum Gravity (2007), 3125–3138.

[16] Sijie Gao and Robert M. Wald, *Theorems on gravitational time delay and related issues*, Classical Quantum Gravity **17** (2000), no. 24, 4999–5008. MR MR1808809 (2001m:83077)

[17] Robert Geroch, *Domain of dependence*, J. Mathematical Phys. **11** (1970), 437–449. MR MR0270697 (42 #5585)

[18] Mikhael Gromov and H. Blaine Lawson, Jr., *Positive scalar curvature and the Dirac operator on complete Riemannian manifolds*, Inst. Hautes Études Sci. Publ. Math. (1983), no. 58, 83–196 (1984). MR MR720933 (85g:58082)

[19] S. W. Hawking and G. F. R. Ellis, *The large scale structure of space-time*, Cambridge University Press, London, 1973, Cambridge Monographs on Mathematical Physics, No. 1. MR MR0424186 (54 #12154)

[20] Jerry L. Kazdan and F. W. Warner, *Existence and conformal deformation of metrics with prescribed Gaussian and scalar curvatures*, Ann. of Math. (2) **101** (1975), 317–331.

[21] R. C. Myers and M. J. Perry, *Black holes in higher-dimensional space-times*, Ann. Physics **172** (1986), no. 2, 304–347. MR MR868295 (88a:83074)

[22] Barrett O'Neill, *Semi-Riemannian geometry*, Pure and Applied Mathematics, vol. 103, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983, With applications to relativity. MR MR719023 (85f:53002)

[23] R. Penrose, *Zero rest-mass fields including gravitation: Asymptotic behaviour*, Proc. Roy. Soc. Ser. A **284** (1965), 159–203. MR MR0175590 (30 #5774)

[24] Roger Penrose, *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14** (1965), 57–59. MR MR0172678 (30 #2897)

[25] ———, *Techniques of differential topology in relativity*, Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1972, Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 7. MR MR0469146 (57 #8942)

[26] R. Schoen and S. T. Yau, *On the structure of manifolds with positive scalar curvature*, Manuscripta Math. **28** (1979), no. 1-3, 159–183.

[27] Richard Schoen, Lecture at Miami Waves Conference (2004).

[28] Robert M. Wald, *General relativity*, University of Chicago Press, Chicago, IL, 1984. MR MR757180 (86a:83001)

[29] Shing Tung Yau, *Survey on partial differential equations in differential geometry*, Seminar on Differential Geometry, Ann. of Math. Stud., vol. 102, Princeton Univ. Press, Princeton, N.J., 1982, pp. 3–71. MR MR645729 (83i:53003)