Hawking's Incompleteness Theorem

Kostas Triamatakis

October 30, 2025

“...Remember to look up at the stars and not down at your feet.
While there is Life, there is Hope...”

Stephen Hawking

Contents

Introduction

In this small project our main goal is to present one of the great results of General Relativity proved by Stephen Hawking in his dissertation in 1966. The famous theorem, mostly known as Hawking's singularity theorem, asserts that under certain conditions our spacetime exhibits incompleteness along its timelike geodesics. Our goal, in these notes, is to provide all the necessary tools and bibliography for someone to comprehend the theorem and its proof.

I would like to express my gratitude to professor Grigorios Fournodavlos for trusting me with this project and for granting me a scholarship under the SINGinGR-Proj. 101078061 program which allows me to focus on my graduate studies.

Preliminaries from differential geometry

Smooth manifolds

In this subsection we mention necessary tools from the theory of smooth manifolds. It is based on the books of Lee and Tu, (JL) and (LT), respectively.
A very important result asserts the existence of partitions of unity subordinated to some cover. This key concept allows one to extend smooth functions, fields, forms etc.

Theorem 2.1. *If M is a smooth manifold then for any open cover $(U_a)_{a\in A}$ there exists a smooth partition of unity $(\psi_a)_{a\in A}$ subordinated to this cover. That is, that there exists a family of smooth functions $\psi_a : M \to \mathbb{R}$ which satisfies:*
a). $0 \leq \psi_a(x) \leq 1$ for all $x \in M$ and $a \in A$.
b). $supp(\psi_a) \subseteq U_a$ for each $a \in A$.
c). The set $\{supp(\psi_a) : a \in A\}$ is locally finite. That is, for every point $p \in M$ there exists a neighborhood of that point which intersects only finite elements of that set.
d). $\sum_{a\in A} \psi_a(x) = 1$, for all $x \in M$.
Notice that the sum on d). is a finite sum for every x because of c).

Another useful notion in differential topology that plays a central role in general relativity is the notion of submanifolds. Here we define the embedded submanifolds.

Definition 2.1. *An $S \subset M$ is called a smooth $k-$submanifold of M if for all $p \in S$ there is a smooth chart (U, φ) of $p \in M$ such that*

$$\varphi(U \cap S) = \{(x_1, \dots, x_n) \in \varphi(U) : x_{k+1} = \dots = x_n = 0\}.$$

(U, φ) is, then, called a slice-chart. The topology of S is the one that it inherits from M and the atlas of S that makes it a smooth $k-$ manifold is $\mathcal{A}_S \coloneqq \{(U \cap S, \varphi|_{U\cap S}) : (U, \varphi)$ is a slice-chart$\}$. An embedded submanifold of dimension $\dim(M) - 1$ is called a hypersurface.

A curve $\gamma : I \to M$ is said to be an integral curve of $X \in \mathfrak{X}(M)$ if and only if $\dot{\gamma}(t) = X_{\gamma(t)}$.The following result is immediate from ODE theory.

Theorem 2.2. *If $X \in \mathfrak{X}(M)$ then for all $p \in M$ there exists a unique integral curve $\gamma : (-\varepsilon, \varepsilon) \to M$ such that $\gamma(0) = p$. That is,*

$$\begin{cases} \dot{\gamma}(t) = X_{\gamma(t)} \\ \\ \gamma(0) = p \end{cases}.$$

The integral curve γ has a maximal extension. If all integral curves of $X \in \mathfrak{X}(M)$ can be extended to the whole real line we say the field is complete.

Similarly to integral curves, there exists a map which is called the flow of the vector field and it is defined by the following theorem.

Theorem 2.3. *If $X \in \mathfrak{X}(M)$ and $p \in M$ then there exists an $\varepsilon > 0$, a neighborhood U of p and map $\varphi : (-\varepsilon, \varepsilon) \times U \to M$ with the property:*

$$\begin{cases} \dfrac{\partial \varphi}{\partial t}(t, q) = X_{\varphi(t,q)} \\ \\ \varphi(0, q) = q \end{cases}, \ \forall \ (t, q) \in (-\varepsilon, \varepsilon) \times U.$$

From the above theorem one can immediately see that the set $\{\varphi_t\}_{t \in (-\varepsilon, \varepsilon)}$ defines a group of local diffeomorphisms. Indeed, the set satisfies the group axioms as follows:
a). $\varphi_t \circ \varphi_s = \varphi_{t+s}$ for when some handside is well defined (and so the operation is associative and commutative).
b). $\varphi_0 = id$.
c). $\varphi_t^{-1} = \varphi_{-t}$.
It can be proved that two fields commute if and only if their flows commute.

We conclude this section by introducing the notions of connections and geodesics. We follow (KA, p. 35).

Definition 2.2. *A connection on a smooth manifold M is a map $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ which we denote by $\mathfrak{X}(M) \times \mathfrak{X}(M) \ni (X, Y) \mapsto \nabla_X Y \in \mathfrak{X}(M)$ and satisfies:*
a). It is $\mathbb{R}$-bilinear.
b). It is $C^\infty(M)-$linear with respect to the first variable, that is $\nabla_{fX} Y = f \nabla_X Y$
c). It is a derivation with respect to the second variable, that is $\nabla_X (fY) = Xf \cdot Y + f \nabla_X Y$.
It can be proved that the field $\nabla_X Y$ depends locally on X and Y and pointswise with respect to X. In fact, it can be shown that $(\nabla_X Y)_p$ depends pointwise with respect to X (i.e. X_p) and on the values of Y along any smooth curve that satisfies $\dot{\gamma}(0) = Y_p$.
A connection is called symmetric if its torsion, $T(X, Y) := \nabla_X Y - \nabla_Y X - [X, Y]$, vanishes identically. Given any local frame $\{E_1, \ldots, E_n\}$ we define the Christoffel symbols of the connection with respect to that frame as $\nabla_{E_i} E_j := \sum_{k=1}^{n} \Gamma_{ij}^k E_k$.
The field $\nabla_X Y$ is called the covariant derivative of Y along X.

The local expression of the connection with respect to any local frame is

$$\nabla_X Y = \sum_k \left(X(Y^k) + \sum_{i,j} X^i Y^j \Gamma_{ij}^k \right) E_k.$$

A connection is symmetric, then, if and only if for any local and commutative frame [1] we have $\Gamma_{ij}^k = \Gamma_{ji}^k$.

[1] $[E_i, E_j] = 0$

Theorem 2.4. *Given a smooth curve $\gamma : I \to M$ we consider the vector space:*

$$\mathfrak{X}(\gamma) \coloneqq \{X : I \to TM : \text{ For all } t \in I,\ X_t \in T_{\gamma(t)}M\}$$

which is the space of all smooth vector fields defined on γ. Then there exists a unique, linear operator

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

with the following properties:

a). $\dfrac{D(fX)}{dt} = f'X + f\dfrac{DX}{dt}$, for all smooth functions $f : I \to \mathbb{R}$ and fields $X \in \mathfrak{X}(\gamma)$.

b). For any smooth extensions Γ, $\bar{X} \in \mathfrak{X}(M)$ of $\dot{\gamma}$ and X, respectively, it holds $\dfrac{DX}{dt} = \nabla_\Gamma \bar{X}$ and since the field $\nabla_\Gamma \bar{X}$ depends only on $\Gamma_{\gamma(t)} = \dot{\gamma}(t)$ and values of $\bar{X}$ along any curve with velocity $\dot{\gamma}(t)$ (take γ) we can write, without loss of generality, $\dfrac{DX}{dt} = \nabla_{\dot{\gamma}} X$ for each $t \in I$. Thus the local expression reads:

$$\frac{DX}{dt} = \sum_k \left(X_k' + \sum_{i,j} \gamma_i' X_j \Gamma_{ij}^k \right) \frac{\partial}{\partial x_k}.$$

The operator defined above is called the covariant derivative on γ and measures how much a field changes by moving on γ with respect to γ.

We call $X \in \mathfrak{X}(\gamma)$ parallel if $\dfrac{DX}{dt} = 0$ for all $t \in I$. Geometrically, that means that if an observer walks alongside X then they see no change of X as they move on γ. The following theorem is obvious from the theory of ODE's.

Proposition 2.1. *For all smooth curves $\gamma : (-\varepsilon, \varepsilon) \to M$ and all $v \in T_{\gamma(0)}M$ there exists a unique smooth and parallel vector field along γ, $X \in \mathfrak{X}(\gamma)$, with initial condition v. That is, X solves the ODE system*

$$\begin{cases} \dfrac{DX}{dt} = 0, \\ \\ X_0 = v \end{cases}.$$

X is called the parallel translation of v along γ. Thus an isomorphism of tangent spaces is induced by considering $\tau_{a,b} : T_{\gamma(a)}M \to T_{\gamma(b)}M$ such that $\tau_{a,b}(v) \coloneqq X_b$ where X is the parallel translation of v along γ.

Finally, we define geodesics. Intuitively, a geodesic is a curve in which a particle moving along γ feels no change of direction. We could say that the particle feels no force or that it moves along a "straight line".

Definition 2.3. *A geodesic $\gamma : I \to M$ is a smooth curve that satisfies $\dfrac{D\dot{\gamma}}{dt} = 0$. The expression $\dfrac{D\dot{\gamma}}{dt} = 0$. is, locally, equivalent to a system of ODE's that reads:*

$$\gamma_k'' + \sum_{i,j} \gamma_i' \gamma_j' \Gamma_{ij}^k = 0,\ k = 1, \ldots, n.$$

The next theorem asserts that locally, for all points and all possible initial velocities there exists a unique geodesic with this initial data.

Theorem 2.5. *For all $(p, v) \in TM$, that is for all $p \in M$ and $v \in T_pM$, there exists, locally in time, a unique geodesic $\gamma : (-\varepsilon, \varepsilon) \to M$ satisfying $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.*

A smooth manifold with a connection (M, ∇) is called complete if any geodesic can be extended (as a geodesic) to the whole real line. This definition of completeness is strongly connected to the completeness of a vector field defined on TM which we call the geodesic field. Its flow is called the geodesic flow. We say M is complete if and only if the geodesic field of M is complete. It can be proven that its integral curve on $(p, v) \in TM$ is of the form $(\gamma(t), \dot{\gamma}(t))$ where γ is the (p, v)-geodesic.
Finally, we consider the set $E \coloneqq \{(p, v) \in TM : \text{the } (p, v)\text{−geodesic is defined on } [0, 1]\}$ which is an open set of TM. We define the exponential map $\exp : E \to M$ as $exp(p, v) \coloneqq \gamma_{(p,v)}(1)$ [2] and the restriction of it on T_pM as $(\exp)_p(v) \coloneqq \gamma_{(p,v)}(1)$. It holds that $(\exp)_p(tv) = \gamma_{(p,v)}(t)$ and that $(\exp)_p$ is a local diffeomorphism at $0 \in T_pM$. By considering a linear isomorphism from T_pM to $\mathbb{R}^n$ we have that $h \circ (\exp)_p^{-1} : \exp_p(U) \to U \subseteq \mathbb{R}^n$ where $U \subseteq T_pM$ the neighborhood in which $(\exp)_p$ is a local diffeomorphism then we get what we call the geodesic or normal coordinates of M. In these coordinates the "center" of the neighborhood, p, is mapped to $0 \in \mathbb{R}^n$ and all geodesics emitting from p are mapped to straight lines which pass through the origin. From the geodesic equations it's immediate that at p we have $\Gamma^k_{ij}(p) = -\Gamma^k_{ji}(p)$ and thus if the connection is symmetric we get that the Christoffel symbols vanish at p. A neighborhood of $p \in M$, $U = \exp_p(V) \cong V$ where V is a star-shaped neighborhood of $0 \in T_pM$ and mapped diffeormorphically to U by the exponential map at p shall be called a normal neighborhood from now on.

Pseudo-Riemannian geometry

Our geometric setting is that of Pseudo-Riemannian/Semi-Riemannian manifolds. In this section we will mention almost all of the results that we are going to need. We will omit the proofs for the majority of the theorems mentioned but a bibliographical reference will always be included.
The biggest part of this section are notes taken from the books (BO) and (KA).

Definition 2.4. *A symmetric, bilinear form on a vector space V, $b : V \times V \to \mathbb{R}$, is called a positive/negative definite form (scalar product) when for all $v \in V$, $b(v, v) > 0$ ($b(v, v) < 0$). Positive/negative semidefinite when $b(v, v) \geq 0$ ($b(v, v) \leq 0$) while it is called non-degenerate if $b(v, w) = 0$, for all $w \in V$ implies $v = 0$. A vector space equipped with a non-degenerate form is called a scalar product space (SPS, for short) while a subspace will be called non-degenerate if the restriction of the form to the subspace is non-degenerate.*

We say v is orthogonal to w when $b(v, w) = 0$. The norm of v is denoted by $\|v\| \coloneqq |b(v, v)|^{\frac{1}{2}}$.

Definition 2.5. *In an SPS, (V, b), a vector v is called timelike if $b(v, v) < 0$, null/lightlike if $b(v, v) = 0$ and spacelike if $b(v, v) > 0$.*

It is a matter of simple linear algebra manipulations to prove the following lemmas and theorems (see (BO, p. 47)).

Lemma 2.6. *(V, b) is an SPS if and only if for all basis $(e_i)_{i=1}^n$ the matrix $(b_{ij})_{i,j=1,\dots,n}$ is invertible.*

Lemma 2.7. *For all $W \leq V$, where V is an SPS, it holds that $\dim(W) + \dim(W^\perp) = \dim(V)$ and so for dimensional reasons it holds $W^{\perp\perp} = W$. Then since $(W + W^\perp)^\perp = W \cap W^\perp$ we have W is non-degenerate if and only if $W \cap W^\perp = 0$ if and only if $V = W \oplus W^\perp$.*

A natural question that arises is whether we can have an orthonormal basis for our SPS. The answer is positive only with a slightly different technique to prove it from the positive definite case (in which we have an inner product and thus we can apply the Gram-Schmidt algorithm). Let V be non-trivial. Then since b is non-degenerate there exists a $w_1 \neq 0$ such that $b(w_1, w_1) \neq 0$. Consider $W_1 \coloneqq span\{w_1\}$ which is a non-degenerate subspace and thus $V = W_1 \oplus W_1^\perp$. Pick a non-zero w_2 such that $b(w_2, w_2) \neq 0$ (that is always possible unless the dimension of V is exactly 1). Apparently, w_1 is orthogonal to w_2. Consider then $W_2 \coloneqq span\{w_1, w_2\}$ which is, also, non-degenerate and $V = W_2 \oplus W_2^\perp$. Pick w_3 appropriately as above and let the algorithm run until we have run out of dimensions. Then we normalize our vectors. We conclude the following theorem.

[2] We denote by $\gamma_{(p,v)}$ the (p, v)−geodesic.

Theorem 2.8. *Any SPS has an orthonormal basis.*

It is immediate that an orthonormal basis can not contain null vectors. It contains only timelike and spacelike vectors. We denote the basis by $\{e_1^-, \dots, e_k^-, e_{k+1}^+, \dots, e_n^+\}$ to imply that e_j^- is timelike and e_i^+ is spacelike. The number of timelike vectors is invariant of the orthonormal basis and we call it the signature/index of the form/metric.

Proposition 2.2. *If (V, b) is an SPS then in any orthonormal basis of V the number of timelike vectors is constant. We call this number the index of the form. The index is the dimension of the maximal, in terms of dimension, subspace which is negative definite* [3].

We conclude our study of the linear algebra of scalar product spaces with the following characterization.

Theorem 2.9. *Two SPS are isometric if and only if they have the same dimension and index.*

Definition 2.6. *A pair (M, g), where M is a smooth manifold and a symmetric, non-degenerate $(0,2)-$ smooth tensor field $g : \mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M)$ of constant index* [4] *which we call a Pseudoriemannian metric, is called a Pseudoriemannian manifold. If g is positive definite then the manifold is called Riemanian. If the index is equal to 1 then we have a Lorentzian metric and a Lorentzian manifold.*

The most classical example of a Pseudoriemannian manifold is the ν-Minkowski space $\mathbb{R}^n_\nu$ endowed with the metric $g \coloneqq -\sum\limits_{j=1}^{\nu} dx_j^2 + \sum\limits_{j=\nu+1}^{n} dx_j^2$. The case $\nu = 1$ is called the Minkowski space.

Definition 2.7. *We say a connection ∇ in a Pseudoriemannian manifold with a connection, (M, g, ∇), is compatible with the metric when for all smooth curves γ and all times a, $b \in D(\gamma)$ the parallel transport $\tau_{a,\,b} : T_{\gamma(a)}M \to T_{\gamma(b)}M$ is a linear isometry.*

Theorem 2.10. *The following are equivalent:*
a). The connection ∇ is compatible with the metric.
b). For all X, Y, $Z \in \mathfrak{X}(M)$ smooth vector fields it holds:

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

c). For all smooth curves γ and all V, $W \in \mathfrak{X}(\gamma)$ smooth vector fields on γ it holds:

$$\frac{d}{dt}(g(V, W)) = g\left(\frac{DV}{dt}, W\right) + g\left(V, \frac{DW}{dt}\right).$$

Proof. See (KA, p. 51). ☐

The fundamental theorem of Pseudoriemannian geometry asserts that given a Pseudoriemannian manifold there always exists a unique symmetric and compatible connection. We call this connection the Levi-Civita connection.

Theorem 2.11. *Let (M, g) a Pseudoriemannian manifold. There exists a unique symmetric and compatible with the metric connection which we call the Levi-Civita connection. The Christoffel symbols, in local coordinates, read :*

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{m=1}^{n} g^{mk} \left(\frac{\partial g_{mi}}{\partial x_j} + \frac{\partial g_{mj}}{\partial x_i} - \frac{\partial g_{ij}}{\partial x_m} \right).$$

Proof. See (BO, p. 62). ☐

From now on all our Pseudoriemannian manifolds will be equipped with the Levi-Civita connection. Recall that geodesics are defined through the connection. Therefore the definition of geodesics remains the same. An easy exercise comes as the following corollary.

[3] The restriction of the form to this subspace is negative definite.

[4] That is g assigns smoothly to every point $p \in M$, g_p which is a symmetric, $C^\infty(M)$-bilinear and non-degenarate form of index 1 on the tangent space of p. Smooth means that for all smooth vector fields X, Y $g(X, Y)$ is smooth.

Corollary 2.1. *In a Pseudoriemannian manifold (M, g) all geodesics have constant speed. That is, $g(\dot{\gamma}, \dot{\gamma})$ is constant.*

It can be proved that local isometries preserve the Levi-Civita connections between manifolds. Thus, a local isometry maps geodesics to geodesics.
Since our metric is a non-degenarate, symmetric, bilinear form it assigns a scalar product to every tangent space. Thus we can pick $h : T_pM \to \mathbb{R}^n_\nu$ an isometry of scalar product spaces and then consider normal coordinates with respect to this isomorphism. It follows that, since the connection is symmetric, $\Gamma^k_{ij}(p) = 0$, where p is the center. Then, if $g_{ij} \coloneqq g(\partial_i, \partial_j)$ we get

$$\frac{\partial g_{ij}}{\partial x_k} = g(\nabla_{\partial_k}\partial_i, \partial_j) + g(\partial_i, \nabla_{\partial_k}\partial_j).$$

From the definition of Christoffel symbols we get $\dfrac{\partial g_{ij}}{\partial x_k}(p) = 0$. Finally, if ε_i is the sign of g_{ii} then we get $g_{ij}(p) = \varepsilon_i \delta_{ij}$. See (BO, p. 73).

Proposition 2.3. *In normal neighborhoods all points can be connected to the center with a (radial) geodesic. Also, our manifold is connected if and only if any two points can be connected through a broken* [5] *geodesic.*

Proof. See (BO, p. 73). ☐

There is a very natural isomorphism between the space of vector fields on a Pseudoriemannian manifold and $1-$forms. Consider $\mathfrak{X}(M) \ni X \mapsto X^* \coloneqq g(X, \cdot)$. Since the metric is non-degenerate we have that the mapping is injective. Applying the Riesz Representation theorem we get that for any $1-$form $\omega \in \Omega^1(M)$ there exists a vector field X in M such that $\omega = g(X, \cdot)$. X reads in local coordinates $X_k = \displaystyle\sum_l g^{lk}\omega_l$ and thus is smooth which means the above mapping is an isomorphism. Note that X^* in local coordinates reads $X^* = \displaystyle\sum_l g_{lk}X_l$.
In the definition of what a pseudoriemannian metric is we used the phrase "$(0, 2)-$tensor field". An $(r, s)-$tensor field of type a is a $C^\infty(M)$-multilinear form $A : (\Omega^1(M))^r \times \mathfrak{X}(M)^s \to C^\infty(M)$ while a $(r, s)-$tensor field of type b is a $C^\infty(M)$-multilinear form $A : (\Omega^1(M))^r \times \mathfrak{X}(M)^s \to \mathfrak{X}(M)$. In the relativistic literature one will encounter the notation $A^{i_1,\dots,i_r}_{j_1,\dots,j_s}$ for an (r, s)-tensor field (of type a, mostly). The formal explanation of this notation is that in local coordinates the tensor reads:

$$A(\theta_1, \dots, \theta_r, X_1, \dots, X_s) = \sum_{\substack{i_1,\dots,i_r \\ j_1,\dots,j_s}} A^{i_1,\dots,i_r}_{j_1,\dots,j_s} \frac{\partial}{\partial x_{i_1}} \otimes \cdots \otimes \frac{\partial}{\partial x_{i_r}} \otimes dx_{j_1} \otimes \cdots \otimes dx_{j_s}(\theta_1, \dots, X_s),$$

where we denote by $\otimes$ the tensor product of two tensors which are (r_1, s_1) and (r_2, s_2) tensors of type a respectively as $A \otimes B$ which is a $(r_1 + r_2, s_1 + s_2)$-tensor as follows:

$$(A \otimes B)(\theta_1, \dots, \theta_{r_1+r_2}, X_1, \dots, X_{s_1+s_2}) = A(\theta_1, \dots, \theta_{r_1}, X_1, \dots, X_{s_1}) \cdot B(\theta_1, \dots, \theta_{r_2}, X_1, \dots, X_{s_2}).$$

We will be denoting the space of all $(r, s)-$tensor fields of type a as $\mathfrak{X}^r_s(M)$. Notice that a vector field $X \in \mathfrak{X}(M)$ can be considered a $(1, 0)-$tensor field since in the following way: write $X(\omega) \coloneqq \omega(X)$. $1-$forms are by definition $(0, 1)-$tensor fields.

Lemma 2.12. *There exists a unique $C^\infty(M)$-linear operator $C : \mathfrak{X}^1_1(M) \to C^\infty(M)$ which satisfies $C(X \otimes \theta) = \theta(X)$. We call this operator contraction and, in a sense, it is a trace operator. In local coordinates we have*

$$C(A) = \sum_i A_{ii} = \sum_i A(dx_i, \partial x_i).$$

[5] Piecewise smooth.

The contraction operator could not have any other local representation since

$$
\begin{aligned}
C(A) = C\Big(\sum_{i,j} A^i_j \frac{\partial}{\partial x_i} \otimes dx_j\Big) &= \sum_{i,j} A^i_j \cdot C\Big(\frac{\partial}{\partial x_i} \otimes dx_j\Big) \\
&= \sum_{i,j} A^i_j \delta_{ij} \\
&= \sum_{i} A^i_i.
\end{aligned}
$$

The method to construct these operators is very standard in differential topology. One begins by noticing that the properties of the operator demand for a standard local representation formula. Then one considers, locally, the operator with this formula with respect to some chart (U, φ). Finally, one shows that the definition is independent of the chart or more specifically that in the intersection of the two open sets the two representations agree.
We generalize, now, this operator to higher order tensors as follows:
We consider $C^i_j : \mathfrak{X}^r_s(M) \to \mathfrak{X}^{r-1}_{s-1}(M)$ such that:

$$
\begin{aligned}
C^i_j(A)(\theta_1, \dots, \theta_{r-1}, X_1, \dots, X_{s-1}) &:= C(A(\theta_1, \dots, \theta_{i-1}, \cdot, \theta_{i+1}, \dots, \theta_{r-1}, X_1, \dots, X_{j-1}, \cdot, X_j, \dots, X_{s-1})). \\
&= \sum_a A(\theta_1, \dots, \theta_{i-1}, dx_a, \theta_i, \dots, \theta_{r-1}, X_1, \dots, X_{j-1}, \partial x_a, X_j, \dots, X_{s-1}).
\end{aligned}
$$

Something that we will, also, need is the raising/lowering indices operator. We define $\Big\uparrow^{a}_{b} : \mathfrak{X}^r_s(M) \to \mathfrak{X}^{r+1}_{s-1}(M)$ such that:

$$
\Big(\Big\uparrow^{a}_{b} A\Big)(\theta_1, \dots, \theta_{r+1}, X_1, \dots, X_{s-1}) := A(\theta_1, \dots, \theta_{a-1}, \theta_{a+1}, \dots, X_1, \dots, X_{b-1}, \theta^*_a, X_b, \dots, X_{s-1}),
$$

where by θ^* we denote the dual vector field of the 1-form. In the same way we define $\Big\downarrow^{a}_{b} : \mathfrak{X}^r_s(M) \to \mathfrak{X}^{r-1}_{s+1}(M)$ by sending the extra X_b vector field to the $a-$th forms-slot as its dual form X^*_b. Finally, we consider $C_{a,b} := C^a_{b-1} \circ \Big\uparrow^{a}_{a}$.

Definition 2.8. *Given any $f \in C^\infty(M)$ there exists a unique smooth vector field on M, which we denote by $grad(f)$ and call the gradient of f, that satisfies $df = g(grad(f), \cdot)$. That means that $grad(f)$ is nothing else but the dual vector field of the differential of f. We then define the divergence of a vector field $X \in \mathfrak{X}(M)$ as $div(X) := C_{1,2}(\nabla X)$, where ∇X is the $(1,1)-$tensor field which reads $\nabla X(\omega, Y) = \omega(\nabla_Y X)$. For any covariant [6] tensor field we can define for T which is a $(0,s)-$tensor the $(0,s)$-tensor $\nabla_X T$ as:*

$$
(\nabla_X T)(Y_1, \dots, Y_s) := X(T(Y_1, \dots, Y_s)) - \sum_i T(Y_1, \dots, \nabla_X Y_i, \dots, Y_s).
$$

Thus we define the divergence of a symmetric $(0,2)$-tensor of type a, A, as $div(A) := C_{1,3}(\nabla A)$. Finally, we define the laplacian $\Delta : C^\infty(M) \to C^\infty(M)$ as $\Delta := div \circ grad$.

Let us write some of these operators locally. First, it is immediate from a previous discussion that the gradient is written in local coordinates as

$$
grad(f) = \sum_{i,j} g^{ij} \frac{\partial f}{\partial x_j} \frac{\partial}{\partial x_i}.
$$

[6] That is, for any $(0,s)$-tensor field of type a while $(r,0)$-tensor fields of type a are called contravariant

Next the divergence of a vector field can be computed from the trace formula we wrote above as

$$\begin{aligned} div(X) = C(\nabla X) &= \sum_i \nabla X(dx_i, \partial x_i) \\ &= \sum_i dx_i(\nabla_{\partial x_i} X) \\ &= \sum_i dx_i \Big(\sum_j \frac{\partial X_j}{\partial x_i} + \sum_{l,j} \Gamma^l_{i,j} X_j \frac{\partial}{\partial x_l} \Big) \\ &= \sum_i \Big(\frac{\partial X_i}{\partial x_i} + \sum_l \Gamma^j_{li} X_l \Big). \end{aligned}$$

Notice that in local coordinates the Christoffel symbols vanish and so we get (only at the center) the usual formula for the divergence of a field.
The divergence of a symmetric $(0,2)$-tensor reads in local coordinates

$$\begin{aligned} div(A) = C_{1,3}(\nabla A) = C^1_2 \circ \Big\uparrow^1_1 (\nabla A) &= \sum_i \Big\uparrow^1_1 (\nabla A)(dx_i, \cdot, \partial x_i) \\ &= \sum_i \nabla A((dx_i)^*, \cdot, \partial x_i) \\ &= \sum_{i,j,k} g^{ij} (\nabla A)(\partial x_i, \partial x_j, \partial x_k)\, dx_k \end{aligned}$$

and so:

$$div(A)^k = \sum_{i,j} g^{ij} \Big(\frac{\partial A_{ik}}{\partial x_j} - \sum_l \Big(\Gamma^l_{ji} A_{lk} + \Gamma^l_{jk} A_{li} \Big) \Big).$$

Finally, by combining the results above we get the laplacian locally as:

$$\begin{aligned} \Delta f &= \sum_{i,j} \Big(g^{ij} \frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{\partial f}{\partial x_j} \Big(\frac{\partial g^{ij}}{\partial x_i} + \sum_l \Gamma^i_{li}\, g^{lj} \Big) \Big) \\ &= \frac{1}{\sqrt{|\det(g)|}} \sum_{i,j} \frac{\partial}{\partial x_i} \Big(\sqrt{|\det(g)|} g^{ij} \frac{\partial f}{\partial x_j} \Big). \end{aligned}$$

As one can notice the formula in normal coordinates agrees with the usual laplacian (modulo some signs). For all of the above see (BO, p. 35).

One of the great protagonists in the theory of general relativity is curvature. We follow (BO, p. 74). See, also, (KA, p. 75).

Definition 2.9. *The curvature tensor is a* $(0,3)-$*tensor of type* b, $R : \mathfrak{X}(M)^3 \to \mathfrak{X}(M)$ *given by the formula:*

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

The Riemann curvature tensor is the $(0,4)-$*tensor of type* a *given by the formula:*

$$Rm(X,Y,Z,W) \coloneqq g(R(X,Y)Z, W).$$

An obvious property of the two tensors is that they are antisymmetric on the first two variables. The first Bianchi identity asserts that:

$$R(X,Y)Z + R(Z,X)Y + R(Y,Z)X = 0,\ \forall\, X,\, Y,\, Z \in \mathfrak{X}(M).$$

Theorema Egregium, proved by Gauss, in elementary differential geometry asserts that the gaussian curvature of a surface is an intrinsic quantity, that is it can be measured by beings that live on the surface. We have a similar result here:

Lemma 2.13. *Set R^l_{ijk} the components of $R(\partial x_i, \partial x_j, \partial x_k)$ with respect to the basis $(\partial x_i)_{i=1}^n$. Then:*

$$R^l_{ijk} = \frac{\partial \Gamma^l_{jk}}{\partial x_i} - \frac{\partial \Gamma^l_{ik}}{\partial x_j} + \sum_m \Big(\Gamma^m_{jk}\, \Gamma^l_{im} - \Gamma^l_{ik}\, \Gamma^l_{mj} \Big).$$

Theorem 2.14 (Second Bianchi identity)**.** *It holds that:*

$$(\nabla_X R)(Y, Z, W) + (\nabla_Z R)(X, Y, W) + (\nabla_Y R)(Z, X, W) = 0, \ \forall\, X,\, Y,\, Z,\, W \in \mathfrak{X}(M).$$

In normal coordinates the above formula reads (where $p \in M$ is the center of the coordinate system):

$$\frac{\partial R^m_{ijk}}{\partial x_l}(p) + \frac{\partial R^m_{lik}}{\partial x_j}(p) + \frac{\partial R^m_{jlk}}{\partial x_i}(p) = 0.$$

The Riemann curvature tensor is, also, antisymmetric with respect to the third and fourth variables. The interesting fact here is that is symmetric with respect to interchanging the couples of first and second variable and third and fourth variable. That is, $Rm(X, Y, Z, W) = Rm(Z, W, X, Y)$.

Definition 2.10. *Let $\Pi \coloneqq span\{w, v\} \leq T_pM$ a two-dimensional, non-degenerate subspace. The sectional curvature along Π is the quantity:*

$$K_p(\Pi) \coloneqq \frac{Rm(v, w, w, v)}{g(v, v)g(w, w) - g(v, w)^2}.$$

The definition is independent of the basis we pick for Π.

The fact that Π is non-degenerate asserts that the denominator is non-zero. Although the sectional curvature is defined only for non-degenerate planes, it can be shown that for any degenerate plane there exists a non-degenerate plane arbitrarily close to our initial plane (see (BO, p.78)).

Definition 2.11. *The symmetric $(0, 2)-$tensor given by the formula $Ric_p(u, v) \coloneqq Tr(R_p(\cdot, u)v)$ is called the Ricci tensor of (M, g). The Ricci curvature of M at p along the unit vector field $v \in T_pM$ is the quantity $Ric_p(v, v)$. The scalar curvature is the trace or the contraction of the Ricci tensor, i.e. $Sc \coloneqq C_{1,2}(Ric)$.*

Set $R_{ij} \coloneqq Ric(\partial_i, \partial_j)$. Then, $R_{ij} = \sum_m R^m_{mij}$ and $Sc = \sum_{i,j} g^{ij} R_{ij}$. The next formula will assert, in the next section, that the Einstein tensor is divergence-free.

Theorem 2.15 (Contracted Bianchi identity)**.**

$$dSc = 2\, div(Ric).$$

An Einstein manifold is a connected Pseudoriemannian manifold such that $Ric = \Lambda g$ where Λ is a constant. If $Ric = fg$ for some smooth function f then by taking the divergence on both sides we get $div(Ric) = div(fg) = df = \frac{1}{2} dSc$. Also, $Sc = \sum_{ij} g^{ij} R_{ij} = \sum_{ij} g^{ij} f g_{ij} = f Tr(I_n) = nf$, where $n = \dim(M)$. By differentiating, we have $dSc = n\, df$. Comparing the two relationships we get $dSc = 0$ for $n \geq 3$. That means, since M is connected, Sc is constant. Also, $Ric = \frac{Sc}{n} g$ and thus it is an Einstein manifold.
The following results will play a significant role for our main theorem.

Lemma 2.16 (Gauss)**.** *If $p \in M$, $0 \neq x \in T_pM$ and v, $w \in T_x(T_pM)$ with $v = a \cdot x$[7] then:*

$$g((\exp_p)_{*,x}(v), (\exp_p)_{*,x}(w)) = v \cdot w$$

where $v \cdot w$ implies the inner product in $T_x(T_pM)$.

Proof. See (BO, p. 127). ☐

[7] By making the usual convention $T_pM \simeq T_x(T_pM)$.

Definition 2.12. *A geodesically convex neighborhood is a normal neighborhood for all of its points.*

Theorem 2.17. *Every point admits a geodesically convex neighborhood.*

Proof. See (BO, p. 130). ☐

Corollary 2.2. *A geodesic $\gamma : [0, \alpha) \to M$ is extendible past α if and only if it is continuously extendible up to α, i.e.* $\lim_{t \to a} \gamma(t)$ *exists.*

Proof. We follow (BO, p. 130). One direction is obvious. Conversely, if we can continuously extend the geodesic up to time α then we consider a geodesically convex neighborhood of $\gamma(\alpha)$, U. Take t_0 so close to α such that $\gamma([t_0, \alpha]) \subseteq U$. But then $\gamma|_{[t_0,\alpha]}$ is a radial geodesic for $\gamma(t_0)$ and thus it can be extended till it hits the boundary of the neighborhood. But $\gamma(\alpha) \in U$ and so we can extend it pass α. ☐

Recall, now, that a Lorentzian manifold (M, g) is a Pseudoriemannian manifold of index 1. If $p \in M$, its tangent space T_pM is a Lorentz vector space, that is an SPS of index 1. A subspace of a Lorentz vector space is called spacelike if the scalar product, restricted to the subspace, is positive definite. Similary, it is called timelike if it is non-degenerate and lightlike if it is degenerate. For the succeeding we follow (BO, p. 126).

Lemma 2.18. *If z is a timelike vector in a Lorentz vector space then $\{z\}^\perp$ is a spacelike subspace. More generally, if W is timelike then $W^\perp$ is spacelike and conversely if W is spacelike then $W^\perp$ is timelike. Also, two independent null vectors can not be orthogonal.*

The first claim comes from the invariance of the index (Sylvester's inertia law) while the last claim comes from the Cauchy-Schwartz inequality. We have the following useful characterization:

Lemma 2.19. *If $W \leq V$ with $\dim(W) \geq 2$ then the following are equivalent:*
a). W is timelike.
b). There exist two independent null vectors in W.
c). There exists a timelike vector in W.

Similarly, we have a characterization for lightlike subspaces:

Lemma 2.20. *If $W \leq V$ then the following are equivalent:*
a). W is lightlike.
b). W contains a null vector but no timelike vectors.
c). If $\Lambda := q^{-1}(0) - \{0\}$ the null cone, where $q(v) := g(v, v)$ the quadratic form of the scalar product, then $W \cap \Lambda = L - \{0\}$ where L is a one-dimensional submanifold.

We define, now, $\mathcal{T} := \{\text{timelike vectors of } V\}$. We call the set $C(u) := \{v \in \mathcal{T} : \langle v, w\rangle < 0\}$ the timecone of $u \in V$. Both of these sets are open! We can, also, define $C'(u) := \{v \in \mathcal{T} : \langle v, w\rangle > 0\} = C(-u)$. Two timelike vectors can not be orthogonal and thus for two timelike vectors v and u either $v \in C(u)$ or $v \in C(-u)$.

Lemma 2.21. *If $v, w \in \mathcal{T}$ live in the same timecone if and only if $\langle v, w\rangle < 0$. Timecones are convex and open sets. Also, for timelike vectors the inverse Cauchy-Schwartz inequality, i.e. $|\langle v, w\rangle| \geq \|v\| \cdot \|w\|$. If, moreover, they live in the same timecone then the inverse triangle inequality holds, i.e. $\|v+w\| \geq \|v\|+\|w\|$.*

By picking a timecone, in the vector space V, we set, what we call, a time-orientation for V.

Definition 2.13. *If (M, g) is a Lorentz manifold then we say it is time-orientable with time-orientation $\tau : M \to \bigcup_{p\in M} \{\text{time-orientations on } T_pM\}$ such that for every point $p \in M$, $\tau_p \in \{\text{time-orientations on } T_pM\}$ and there exists a neighborhood U_p of p and a local vector field $X \in \mathfrak{X}(U_p)$ such that $X_q \in \tau_q$, for all $q \in U_p$.*

In a sense, time-orientation of a manifold is a "continuous" choice of timecones on each tangent space. The local vector field of the definition is obviously timelike.

Theorem 2.22. *M is time-orientable if and only if there exists a global timelike vector field $X \in \mathfrak{X}(M)$.*

The only if part is obvious. Just choose the timecone generated by $X_p \in T_pM$. For the converse, the method is very commonly-used in such constructions. For any $p \in M$ there exists a neighborhood U_p and a X^{U_p} local timelike vector field at U_p. The set $\{U_p\}_{p\in M}$ is an open covering of M and thus there exists a partition of unity ψ_p subordinated to this cover. Set

$$Z \coloneqq \sum_{p\in M} \psi_p X^{U_p}.$$

Then, $Z \in \mathfrak{X}(M)$ and $g(Z, Z) = \sum_{p,\, q\in M} \psi_p\psi_q g(X^{U_p}, X^{U_q})$. But from Lemma 2.21 since X^{U_p}, X^{U_q} live in the same timecone (on the intersection $U_p \cap U_q$) we get $g(X^{U_p}, X^{U_q}) < 0$. Since a partition of unity constitutes of non-negative functions we get Z is timelike.

Now if (M, g) is a Riemannian manifold admitting a unit vector field U then by setting $\tilde{g} \coloneqq g - 2(U^* \otimes U^*)$ (where U^* is the dual form of the vector field U) we get that $(M, \tilde{g})$ is a Lorentzian manifold. With some tools from algebraic topology one can prove the following beautiful characterization[8]:

Theorem 2.23. *If M is a smooth manifold then the following are equivalent:*
a). M admits a Lorentz metric.
b). M admits a time-orientable metric.
c). There exists a non-vanishing vector field on M.
d). M is either non-compact or compact with its Euler characteristic zero, i.e. $\chi(M) = 0$.

An example where a Lorentzian metric is not possible is for $M \coloneqq \mathbb{S}^{2n}$ (the $2n-$sphere). It is a known fact from algebraic topology that in a sphere of even dimension every vector field vanishes at some point[9].

We say a curve $\gamma : I \to M$ is spacelike if $\dot{\gamma}$ is everywhere spacelike, timelike if $\dot{\gamma}$ is timelike and null if $\dot{\gamma}$ is everywhere null. We call γ a causal curve if it satisfies $g(\dot{\gamma}, \dot{\gamma}) \leq 0$. A broken timelike curve is a piecewise smooth curve that satisfies $g(a'(t_i^+), a'(t_i^-)) < 0$ at the points that it breaks, that is, it remains in a single timecone at the breaking point.
We will use the next lemma to prove that in normal neighborhoods timelike curves maximize proper time, that is the length functional $L(\gamma) \coloneqq \int_a^b \sqrt{-g(\dot{\gamma}, \dot{\gamma})}\, dt$ where $\gamma : [a, b] \to M$ is a timelike smooth curve.

Lemma 2.24. *If $b : [0, 1] \to T_pM$ is a (piecewise) smooth curve with $b(0) = 0$ and such that $a \coloneqq \exp_p \circ b$ is timelike then b remains in a single timecone on* $(0, 1]$.

Before the theorem we define some auxiliary functions. Let $\tilde{q} : T_pM \to \mathbb{R}$ be the quadratic form $\tilde{q}(v) \coloneqq g_p(v, v)$ and $q \coloneqq \tilde{q} \circ (\exp_p)^{-1}$. Let $r \coloneqq \sqrt{-\tilde{q}}$ and the position vector field $\tilde{P} : T_pM \to T(T_pM)$ such that $\tilde{P}(v) \coloneqq \sum_j v_j \left.\frac{\partial}{\partial v_j}\right|_v$ where the global frame $\left\{\frac{\partial}{\partial v_j}\right\}_{j=1}^n$ indicates the global frame of the tangent bundle of T_pM and is derived from the fact that every tangent vector on $p \in M$ can be written as $v = \sum_j v_j \partial_j|_p$.

We, then, transfer this map to the manifold via the exponential map as $P \coloneqq (\exp_p)_{*,\cdot} \circ \tilde{P}$.

Theorem 2.25. *Let U be a normal neighborhood $p \in M$. If there exists a timelike curve from p to q lying entirely in U then the radial geodesic from p to a is, also, timelike and maximizes the length functional acting on timelike curves* [10] *and any other timelike maximizer is a monotone parametrization of the radial geodesic.*

Proof. We follow (BO, p. 147). Indeed, let $\sigma : p \to q$ be a timelike curve lying entirely in U and let it be defined on $[0, 1]$. Since $b \coloneqq (\exp_p)^{-1} \circ \sigma$ is defined and, according to the lemma 2.24, since $b(0) = 0$

[8] A smooth manifold M always admits a Riemannian metric.
[9] Known as the **hairy ball theorem**.
[10] When the length functional acts on a timelike curve we call it the proper time of the curve and we denote it by τ.

and σ is timelike, remains in the same timecone for all times. If $\gamma : p \to q$ is the radial geodesic from p to q then $\sigma(1) = \gamma(1)$ and $\sigma(0) = \gamma(0)$. But $\gamma(t) = \exp_p(tv)$ where $v = \dot{\gamma}(0)$. Then $v = \exp_p^{-1}(\gamma(t))$ and $b(1) = \exp_p^{-1}(\sigma(1)) = \exp_p^{-1}(\gamma(1)) = v$. But $b(1)$ is timelike and thus v is timelike. From corollary 2.1 we have that the radial geodesic γ is timelike. Let now $U_t := P(b(t))/r(b(t))$ which is a timelike, unit vector field on the curve σ from Gauss lemma (lemma 2.16). Therefore, we can write $\sigma'(t) = -g(\sigma'(t), U_t)U_t + N_t$ where $N \perp U$ a spacelike vector field. Then,

$$\begin{aligned} \|\sigma'\|^2 &= |g(\sigma', U)|^2 - g(N, N) \\ &\leq |g(\sigma', U)|^2 \end{aligned}$$

which implies $\|\sigma'\| \leq -g(a', U)$ since a' and U live in the same timecone (again from Gauss Lemma). Notice now

$$\begin{aligned} \frac{d(r \circ b)}{dt} &= \frac{d}{dt}\left(\sqrt{-g(b, b)}\right) \\ &= \frac{1}{r(b)}\langle b', \tilde{P}(b)\rangle \\ &= -g(a', U) \end{aligned}$$

where $\langle \cdot, \cdot' \rangle$ implies the scalar product on $T.(T_pM)$ which is, in a sense, the metric at the point p. We conclude

$$L(\sigma) \leq \int_0^1 \frac{d(r \circ b)}{dt}\, dt = r(b(1) - r(b(0)) = r(v) - 0 = L(\gamma).$$

Now if the above inequality becomes equality we must have $\sigma' = -g(\sigma', U)U$. That gives us $\sigma' = \frac{(r \circ b)'}{r \circ b} P(r(b))$ which implies $(\exp_p)_{*,b}(b') = \frac{(r \circ b)'}{r \circ b}(\exp_p)_{*,b}(\tilde{P}(b))$ which implies (since the exponential map is a local diffeomorphism in the normal neighborhood it induces an isomorphism at the tangent spaces) $b_i' = \frac{(r \circ b)'}{r \circ b} b_i$ and so, $b_i = A_i\, r \circ b$ where A_i is a constant. By setting $t = 1$ we get $A_i = v_i/\|v_i\|$ and so $\sigma(t) = \exp_p\left(\frac{r \circ b}{\|v\|} v\right)$. Moreover, $\|\sigma'\| = \|(r \circ b)'\| > 0$ and thus $(r \circ b)'$ preserves its sign and thus $r \circ b$ is monotone. We have proved that σ is a monotone parametrization of γ (see more: (BO, p. 147)). ☐

A particular case of submanifolds, we are very much interested in, is the case of hypersurfaces (i.e. submanifolds of codimension 1). The following lemma provides a condition so that a Pseudoriemannian hypersurface (that is, if we restrict the metric to the hypersurface then it remains non-degenerate) is orientable.

Theorem 2.26. *If $M \subseteq N$ is a Pseudoriemannian hypersurface of the orientable Pseudoriemannian manifold N then it is orientable if and only if it admits a global, orthogonal and unit vector field $X \in \mathfrak{X}(N)$, that is, $X_p \perp T_pM$ for all $p \in M$ and $\|X\| = 1$.*

Proof. See (BO, p. 189). ☐

Before we proceed we mention one useful tool to do calculations locally. We want to prove that for any point there always exists a local orthonormal frame in a neighborhood of the point. Indeed, take a normal neighborhood of the point and at the tangent space of the space consider an orthonormal basis (that is possible from theorem 2.8). Then, parallel transport this basis along all possible radial geodesics of the neighborhood. Since parallel transport is an isometry we have that it creates a local frame. It is smooth from ODE theory.

Definition 2.14. *A top-form $\omega \in \Omega^{\dim(M)}(M)$ in a Pseudoriemannian manifold M is called a volume element if for every local orthonormal frame $(E_1, \ldots, E_n)$ we have $\omega(E_1, \ldots, E_n) = \pm 1$.*

The following lemma asserts that, at least, locally we always have a volume element. The global case requires one more condition.

Lemma 2.27. *For every chart (U, φ) there exists a (local) volume element $\omega \in \Omega^{\dim(M)}(M)$ and is given by the formula $\omega_{(U,\varphi)} \coloneqq |\det(g_{ij})|^{1/2}\, dx_1 \wedge \cdots \wedge dx_n$.*

Recall now that if a manifold is orientable then it admits an oriented atlas, that is that for any two charts that have no trivial intersection, the Jacobian of the transition map has positive determinant. But then for the local volume elements it holds:

$$\omega_{(U,\varphi)} = \frac{|\det(g_{ij})|^{1/2}}{|\det(\tilde{g_{ij}})|^{1/2}} \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \omega_{(V,\psi)}$$

and the scalar on the RHS is positive we have that they agree on every local orthonormal frame. But forms are $C^\infty(M)-$multilinear and thus they agree everywhere. Thus, we can define $\omega \in \Omega^{\dim(M)}(M)$ as $\omega|_U \coloneqq \omega_{(U,\varphi)}$ and it is a global volume element. Conversely, every global volume element is apparently non-vanishing and thus an orientation form. We have proved the following result:

Theorem 2.28. *The Pseudoriemannian manifold M admits a global volume element if and only if it is orientable. The local representation of the volume element is $\omega|_U = |\det(g_{ij})|^{1/2}\, dx_1 \wedge \cdots \wedge dx_n$. We shall denote μ_g for the volume element from now on.*

Proof. For the preceding, see (BO, p. 194). ☐

For the following see (BO, p. 263) or (KA, p. 95).

Definition 2.15. *A smooth function $\Gamma : (-\varepsilon, \varepsilon) \times [0,1] \to M$ is called a smooth variation of the smooth curve $\gamma : [0,1] \to M$ if $\Gamma(0,t) = \gamma(t)$. We say that it fixes endpoints if $\Gamma(s,0) = \gamma(0)$ and $\Gamma(s,1) = \gamma(1)$. We say it is by geodesics if for all $s \in (-\varepsilon, \varepsilon)$, $\Gamma(s,t) = \Gamma_s(t)$ is a geodesic, that is $\frac{D}{dt}\left(\frac{\partial \Gamma}{\partial t}\right)(s,t) = 0$. The variation field of the smooth variation Γ is the $V \in \mathfrak{X}(\gamma)$ given by the formula $V_t \coloneqq \frac{\partial \Gamma}{\partial s}(0,t)$. A Jacobi field on the smooth curve γ, $J \in \mathfrak{X}(\gamma)$, is one that solves the Jacobi equation:*

$$\frac{D^2 J}{dt^2} + R(J, \dot{\gamma})\dot{\gamma} = 0.$$

Γ fixes endpoints if and only if the variation field vanishes at the endpoints[11].

Lemma 2.29. *Some useful computational formulas:*
a). If $\Gamma : A \to M$ is smooth, where $A \subseteq \mathbb{R}^2$ is open, then:

$$\frac{D}{dt}\left(\frac{\partial \Gamma}{\partial s}\right) = \frac{D}{ds}\left(\frac{\partial \Gamma}{\partial t}\right).$$

b). However, the same does not hold for fields. If $V \in \mathfrak{X}(\Gamma)$, i.e. $V(s,t) \in T_{\Gamma(s,t)}M$ then:

$$\frac{D}{dt}\left(\frac{DV}{ds}\right) - \frac{D}{ds}\left(\frac{DV}{dt}\right) = R\left(\frac{\partial \Gamma}{\partial s}, \frac{\partial \Gamma}{\partial t}\right)V$$

We turn our attention, now, to Jacobi fields.

Lemma 2.30. *If Γ is a smooth variation by geodesics then the variation field of Γ is a Jacobi field.*

This is a straight-forward computation using the lemma 2.29.
From ODE theory now we have the following proposition.

[11] If a field vanishes at endpoints then it is a variation field of some smooth variation fixing endpoints. Set $\tilde{\Gamma}(s,t) \coloneqq (\exp)_{\gamma(t)}(sV_t)$.

Proposition 2.4. *For all v, $w \in T_pM$, $p = \gamma(0)$ there exists a unique Jacobi field such that $J(0) = v$, $\frac{DJ}{dt}(0) = w$. Equivalently, there exists a unique solution to the following initial value problem:*

$$\begin{cases} \dfrac{D^2 J}{dt^2} + R(J, \dot{\gamma})\dot{\gamma} = 0, \\ \\ J(0) = v, \ \dfrac{DJ}{dt}(0) = w. \end{cases}$$

Lemma 2.31. *If J is a Jacobi field with $J(0) = 0$ then it is a variation field of some smooth variation by geodesics.*

Jacobi fields play an important role in the study of conjugate points.

Definition 2.16. *If $\gamma(t) \coloneqq \exp_p(tv)$ is a geodesic, $\gamma : [0, 1] \to M$ with $\dot{\gamma}(0) = v$ then we call the point $q = \gamma(t_0)$ a conjugate point of $p = \gamma(0)$ along γ if and only if $(\exp_p)_{*,t_0 v}$ is not an isomorphism. We call the, non-trivial, dimension of its kernel the multiplicity of the conjugate point q.*

Theorem 2.32. *The point $q = \gamma(t_0)$ is a conjugate point of $p = \gamma(0)$ along γ if and only if there exists a non-trivial Jacobi field $J \neq 0$ along γ that satisfies $J(0) = 0$ and $J(t_0) = 0$.*

Corollary 2.3. *If $q = \gamma(t_0)$ is not conjugate then for all $v \in T_{\gamma(0)}M$ and all $w \in T_{\gamma(t_0)}M$ there exists a unique Jacobi field along γ that satisfies $J(0) = v$ and $J(t_0) = w$.*

This gives us the following very interesting corollary.

Corollary 2.4. *a). If (M, g) is a Riemannian manifold with non-positive sectional curvature then there are no conjugate points.*
b). If (M, g) is a Lorentzian manifold with non-negative sectional curvature along timelike two-dimensional planes then there are no conjugate points along timelike geodesics.

Proof. See (BO, p. 277). □

General Relativity

In this section we study some causality theory for 4-Lorentzian manifolds (spacetimes) and we conclude with the main result, Hawking's theorem.

Causality theory

Definition 3.1. *A spacetime, (M, g), is a connected and orientable Lorentzian $4-$manifold. We say it is singular when it is not geodesically complete.*

We define the future of a point $p \in M$ as:

$$I^+(p) \coloneqq \{q \in M | \exists\, \gamma : p \longrightarrow q \text{ future-directed smooth timelike curve}\}.^{12}$$

We define the causal future as:

$$J^+(p) \coloneqq \{q \in M | \exists\, \gamma : p \longrightarrow q \text{ future-directed smooth causal curve.}\}$$

Similarly, $I^+(S) \coloneqq \bigcup_{p \in S} I^+(p)$ and $J^+(S) \coloneqq \bigcup_{p \in S} J^+(p)$. For the rest of this section we will extensively use the fact that connecting two points with a broken-piecewise smooth timelike curve is equivalent to connecting them with a smooth timelike curve. This is, in reality, an approximation theorem.

Lemma 3.1. *If γ is a timelike piecewise smooth then arbitrarily close to γ there exists a smooth timelike curve with the same causal character.*

Proof. We follow (RP, p. 15). Let us assume it breaks at a point. We take a normal neighborhood at the breaking point and send it through a diffeomorphism that includes the exponential map to the Minkowski space to the curve $(x, |x|, 0, 0)$. Then we can consider a smooth variation $\gamma_s(t)$ that approximates this curve such that it is smooth everywhere except at $(0, 0)$ where it is discontinuous. Taking it back to our manifold we have a variation $\Gamma(s, t)$ smooth everywhere except at $(0, t_0)$ where it is not even continuous. Recall that a piecewise smooth curve being timelike means $g(\dot\gamma(t_0^-), \dot\gamma(t_0^+)) < 0$ where t_0 is the breaking point. For $t \neq t_0$ we have

$$g\left(\frac{\partial\Gamma}{\partial t}(0, t), \frac{\partial\Gamma}{\partial t}(0, t)\right) < 0 \implies g\left(\frac{\partial\Gamma}{\partial t}(s, t), \frac{\partial\Gamma}{\partial t}(s, t)\right) < 0,\ \forall\ s < s_0(t),$$

where the above holds for all small s by the definition of the limit. With the same arguement Γ_s remains in the same timecone for all small s. Also because of the compactness of $[0, 1]$ we can choose $s < s_0$ where s_0 does not depend on t. It remains to prove it for $t = t_0$. The compatibility condition $g(\dot\gamma(t_0^-), \dot\gamma(t_0^+)) < 0$ asserts that the desired result is valid via the same arguement. □

Lemma 3.2. *If γ is a causal curve with Γ a smooth variation of γ and V its variation field then if $g\left(\frac{DV}{dt}, \dot a\right) < 0$ then Γ_s is timelike for all small s. The result remains true for piecewise smooth causal curves.*

Proof. We follow (BO, p. 294). We have:

$$\frac{\partial}{\partial s}\left(g\left(\frac{\partial\Gamma}{\partial t}, \frac{\partial\Gamma}{\partial t}\right)\right) = g\left(\frac{D}{ds}\left(\frac{\partial\Gamma}{\partial t}\right), \frac{\partial\Gamma}{\partial t}\right) = g\left(\frac{D}{dt}\left(\frac{\partial\Gamma}{\partial s}\right), \frac{\partial\Gamma}{\partial t}\right).$$

Then,

$$\frac{\partial}{\partial s}\left(g\left(\frac{\partial\Gamma}{\partial t}, \frac{\partial\Gamma}{\partial t}\right)\right)(0) = g\left(\frac{DV}{dt}, \dot\gamma\right) < 0$$

[12] We denote by $\gamma : p \longrightarrow q$ a curve that connects p with q and by $\gamma : p \overset{+}{\longrightarrow} q$ to imply that the time-orientation is future-directed.

and $\partial_s g\left(\frac{\partial\Gamma}{\partial t}, \frac{\partial\Gamma}{\partial t}\right)(s) < 0$ for all small s because of continuity. This asserts that Γ_s is timelike for all small s since:

$$g\left(\frac{\partial\Gamma_s}{\partial t}, \frac{\partial\Gamma_s}{\partial t}\right) < g(\dot\gamma,\dot\gamma) \leq 0$$

□

The following theorem will be used extensively from now on.

Theorem 3.3. *If γ is a causal curve that is not a null (pre)geodesic* [13] *and connects p to q then we can find arbitrarily close to γ a smooth timelike that connects p to q.*

Proof. We follow (BO, p. 294). We begin with the case that $\dot\gamma(0)$ or $\dot\gamma(1)$ is timelike. Let's take for example that $\dot\gamma(1)$ is timelike. Since $g(\dot\gamma(1),\dot\gamma(1)) < 0$ by continuity there exists a $\delta > 0$ such that $g(\dot\gamma(t),\dot\gamma(t)) < -\delta$ for $t \in [1-\delta, 1]$. We consider W the parallel vector field on γ that is constructed by taking the parallel transport of $\dot\gamma(1)$ along γ. Since W remains timelike and $\dot\gamma$ is timelike then $g(W,\dot\gamma)$ has a constant sign (since both of these fields are timelike) and thus by evaluating at $t = 1$ we have $g(W,\dot\gamma) < 0$. We consider a smooth $f \in C^\infty([0,1])$ that vanishes on both ends and $f' > 0$ on $[0, 1-\delta]$. Consider now $V := fW$. Then, $g\left(\frac{DV}{dt}, \dot\gamma\right) = f' g(W,\dot\gamma) < 0$ on $[0, 1-\delta]$. But then from lemma 3.2 (and the fact that the vector field vanishes at both endpoints) we have that there exists a variation $\Gamma_s(t)$ that fixes endpoints and is timelike on $[0, 1-\delta]$ for all small s. Since $[1-\delta, 1]$ is compact and $g(\dot\gamma(t),\dot\gamma(t)) < -\delta$ on this interval we have that Γ_s is timelike for all small s and all $t \in [0,1]$. If the curve is timelike on $(0,1)$ then we can do the same procedure on $[0, t_0]$ and $[t_0, 1]$ and then smoothly approximate by lemma 3.1.
If the curve is null, but not null geodesic, then $g(\dot\gamma,\dot\gamma) = 0$ and thus by differentiating we have $\ddot\gamma \perp \dot\gamma$ but in order for γ to not be a null pregeodesic we must have $g(\ddot\gamma,\ddot\gamma)$ is not identically zero and thus be spacelike at some $t \in [0,1]$ [14]. The reason for this is that $\ddot\gamma$ is null everywhere if and only if $\ddot\gamma = f\dot\gamma$ for some smooth function f. But that implies that γ can be parametrised to become a geodesic. We proceed by considering a timelike parallel on γ vector field W such that $g(\dot\gamma, W) < 0$ [15]. Set $h := g(\ddot\gamma,\ddot\gamma)/g(W,\dot\gamma)$ which is not identically zero and take φ a smooth function such that $\int_0^1 \varphi h \, dx = -1$. Consider then $f(x) := \int_0^x (\varphi h + 1)\, dt$. We have $f' = \varphi h + 1$. Now consider $V := fW + \varphi\ddot\gamma$. Then,

$$\begin{aligned} g\left(\frac{DV}{dt}, \dot\gamma\right) &= f' g(W,\dot\gamma) - \varphi\, g(\ddot\gamma,\ddot\gamma) \\ &= (f' - \varphi h) \cdot g(W,\dot\gamma) \\ &= g(W,\dot\gamma) < 0. \end{aligned}$$

If the null curve breaks at a point $\dot\gamma(t_0)$ then consider $\Delta\dot\gamma(t_0) := \dot\gamma(t_0^+) - \dot\gamma(t_0^-)$ and parallel transport along both paths. By an easy calculation (evaluating at $t = t_0^\pm$) we have $g(W,\dot\gamma(t)) > 0$ for $t \geq t_0^+$ and $g(W,\dot\gamma(t)) < 0$ for $t \leq t_0^-$. We, then, consider a piecewise smooth f such that $f' > 0$ on $[0, t_0^-]$ and $f' < 0$ on $[t_0^+, 1]$ that vanishes on endpoints and set $V := fW$. □

Corollary 3.1. *Let S be a non-empty set. Then:*
a). $I^+(p)$ is an open set and thus $I^+(S)$ is open.
b). If $q \in J^+(p) - I^+(p)$ then every causal curve connecting the two points must be a null pregeodesic.
c). $\overline{I^+(I^+(S))} = \overline{S}$.
d). $\overline{J^+(S)} = \overline{I^+(S)}$
e). $Int(J^+(S)) = I^+(S)$ which implies $\partial I^+(S) = \partial J^+(S)$.

[13] Pregeodesic, here, means that it can be parametrized to become a geodesic.
[14] Remember a timelike and a null vector cannot be perpendicular.
[15] For example, take v a timelike vector in the same timecone with $\dot\gamma(t)$ for some t and then parallel transport along the curve.

Proof. For a). if $q \in I^+(p)$ then there exists a future-directed timelike curve $\gamma : p \xrightarrow{+} q$ and we consider a geodesically convex neighborhood of q, U_q. We consider a point $r \in \gamma$ that lies between p and q such that $r \in U_q$. Then we consider the positive-oriented timecone on T_rM and we project it with the exponential map to $V_r \subseteq U_q$ and thus we have $q \in V_r$ since there exists a timelike curve that connects r and q inside this neighborhood (Theorem 2.25). Then if $l \in V_r$ then we can find a future-directed timelike radial geodesic that connects r to l. Then we have $p \xrightarrow{+} r \xrightarrow{+} l$ and thus by smoothing out the broken future-directed geodesic (lemma 3.1) we have $l \in I^+(p)$ and thus $V_r \subseteq I^+(p)$.
For b). if $p \longrightarrow q$ with a causal non-null (pre)geodesic then from theorem 3.3 we have that the curve can be approximated by a timelike curve of same orientation and thus $q \in I^-(p)$ which is a contradiction. Thus the causal curve is indeed a null geodesic.
For c). if $r \in I^+(S)$ there exists a point $a \in S$ and a future-directed timelike $\gamma : a \xrightarrow{+} r$. We consider $l \in \gamma$ that lies between a and r and so we have $\gamma : a \xrightarrow{+} l \xrightarrow{+} r$ and thus $l \in I^+(S)$ and $r \in I^+(l) \subseteq I^+(I^+(S))$. Conversely, $r \in I^+(I^+(S))$ there exists a $l \in I^+(S)$ with $\gamma : l \xrightarrow{+} r$. But then there exists a $s \in S$ and a timelike $\delta : s \xrightarrow{+} l$. By concatenating δ and γ, $\delta * \gamma$, and smoothing it out, as in lemma 3.1, we get $s \xrightarrow{+,\text{timelike}} r$ and so $r \in I^+(S)$.
At last we prove d). and leave the rest as an exercise. It's immediate that $\overline{I^+(S)} \subseteq \overline{J^+(S)}$ since $I^+(S) \subseteq J^+(S)$. If, now, $a \in \overline{J^+(S)}$ then for all open neighborhoods of a, O_a, we have $O_a \cap J^+(S) \neq \emptyset$. Consider O_a to be a convex neighborhood [16]. Take a $l \in O_a \cap J^+(S)$ and an $r \in O_a$ such that $l \longrightarrow r$ is a future-directed timelike. But there exists a future-directed causal $S \ni s \longrightarrow l$ and thus the future-directed concatenated curve $s \longrightarrow l \longrightarrow r$ can be approximated by a future-directed timelike since the concatenation is not a null geodesic and has a constant time-orientation (implication of theorem 3.3). Then $r \in O_a \cap I^+(S)$ which gives us the desired inclusion. □

Corollary 3.2. *If $q \in J^+(p) \setminus I^+(p)$ then if $\lambda : p \xrightarrow{+} q$ causal then it must be a null pregeodesic.*

Proof. We follow (RW, p. 191). If this was not the case then there would exist a timelike curve connecting p to q which is a contradiction. □

Definition 3.2. *A set $S \neq \emptyset$ is called achronal if $I^+(S) \cap S = \emptyset$ and so two points of an achronal set S cannot be connected to one another by a timelike curve. We define the edge of the set S as the points of S, let's say $q \in S$, such that for all their open neighborhoods, O_q, there exist points $p \in I^+(q) \cap O_q$ and $r \in I^-(q) \cap O_q$ and a timelike curve $\lambda : p \longrightarrow r$ that remains in O_q but does not intersect S. A set with empty edge is called edgeless.*

Closed, achronal and edgeless sets are very interesting.

Theorem 3.4. *If S is a closed, achronal and edgeless set then it is a C^0-hypersurface (i.e. a topological 3-dimensional submanifold).*

Proof. We follow (RP, p. 23). Consider $q \in S$ and take a geodesically convex and precompact neighborhood of this point, U_q, such that it satisfies the negation of the definition of edge and so that ∂_t (in normal coordinates) is timelike (which is, of course, possible since $g_q(\partial_t|_q, \partial_t|_q) = -1 < 0$ thus by continuity such a neighborhood exists)[17]. Take now $V_q \subset\subset U_q$ a precompact subneighborhood and we consider it such that the normal chart at this neighborhood maps it diffeomorphically to the set

$$\left\{(t, x, y, z) : |t| < \rho,\ x^2 + y^2 + z^2 < \frac{1}{2}\rho^2\right\}$$

where $\rho << 2$ is small enough such that the flow of ∂_t is defined in the neighborhood U_q for time $t \in (-\rho, \rho)$. The $(\pm\rho, x, y, z)$ are chronologically related by a timelike curve which is defined in the obvious way. The integral curves of ∂_t are of the form $s \mapsto (s, x, y, z)$ with (x, y, z) fixed and are obviously timelike (since ∂_t is) and remain in V_q. From now on such a timelike curve will be denoted by $\eta_{(x,y,z)}$. Then without loss of generality $(-\rho, x, y, z) \in I^+(q) \cap U_q$ while $(\rho, x, y, z) \in I^-(q) \cap U_q$ by considering

[16] We can find arbitrarily small convex neighborhoods see (BO, p. 130)
[17] We consider the normal coordinates as (t, x, y, z).

the timelike curve $s \mapsto (s(\pm\rho), sx, sy, sz)$. But these points belong in the integral curve $\eta_{(x,y,z)}$ which lives inside the neighborhood U_q and negates the definition of edge. Since it connects to points from the future and the past of q it must intersect S. The intersection happens at exactly one point since S is achronal and since two different integral curves do not intersect one another we have that the mapping $(x, y, z) \mapsto \beta(x, y, z) \in \eta_{(x,y,z)} \cap S$ is well-defined and injective. Therefore, an injection is induced from $B(0, 1/2\rho) \subseteq \mathbb{R}^3$ to $U_q \cap S$. The continuity is proved as follows: if $\beta(x, y, z)$ and $\beta(x+x_0, y+y_0, z+z_0)$ have a difference of times Δt bigger than $2(x_0^2 + y_0^2 + z_0^2)^{1/2}$ then by considering $\tilde{t} = t + 2(x_0^2 + y_0^2 + z_0^2)^{1/2} + \varepsilon$ for $\varepsilon > 0$ we set

$$\gamma(s) \coloneqq (t + s(2(x_0^2 + y_0^2 + z_0^2)^{1/2} + \varepsilon), x + sx_0, y + sy_0, z + sz_0)$$

which is well-defined (it is inside the neighborhood because of convexity) and timelike. But then the two points are chronologically related which is a contradiction due to achronality. Then $\Delta t \leq 2(x_0^2+y_0^2+z_0^2)^{1/2}$ and thus the map is continuous. The homeomorphism is now obvious since if by $\pi : \mathbb{R}^4 \to \mathbb{R}^3$ we denoted he natural projection and φ the normal chart then

$$\pi|_{\varphi\circ\beta(B(0,1/2\rho))} \circ \varphi \circ \beta = id_{B(0,1/2\rho)}$$

and thus the inverse of $\varphi \circ \beta$ is $\pi|_{\varphi\circ\beta(B(0,1/2\rho))}$ which is apparently continuous. □

Corollary 3.3. *A chronological boundary $S = \partial I^+(C)$ of a set C is a closed, achronal and edgeless set thus a C^0-hypersurface.*

Proof. We follow (RW, p. 192). A boundary is always closed so we proceed with the other two. First we prove achronality. If $q \in \partial I^+(C)$ and take $p \in I^+(q)$ then $q \in I^-(p)$ and thus, since the past is open, there exists an open neighborhood of q such that $O_q \subseteq I^-(p)$. But $O_q \cap I^+(C) \neq \emptyset$ so take $r \in O_q \cap I^+(C)$. We have $r \in I^-(p)$ and so $p \in I^+(r) \subseteq I^+(C)$. Thus $I^+(q) \subseteq I^+(C)$ if $q \in \partial I^+(C)$. Also, if $I^-(q) \cap I^+(C) \neq \emptyset$ then if r belongs in this set then there exists a timelike which is future-directed timelike from r to q. Also, there exists a $c \in C$ and a future-directed timelike from c to r. Thus there exists a future-directed timelike from c to q and thus $q \in I^+(C)$. But this is a contradiction since $I^+(C)$ is open and therefore has a trivial intersection with its boundary. Thus $I^-(q) \subseteq M - I^+(C)$. If two q, $r \in \partial I^+(C)$ that can be connected through a future-directed timelike curve exist then for example it holds that $r \in I^+(q) \subseteq I^+(C)$ which is a contradiction because r belongs in the boundary of the latter set. Thus the chronological boundary is achronal. We, now, prove that it is edgeless. Let O_q an open neighborhood of q in the chronological boundary, $p \in I^+(q) \cap O_q$, $r \in I^-(q) \cap O_q$ and a timelike curve $\lambda : p \longrightarrow r$ that remains in O_q. Then it must intersect $\partial I^+(C)$ since $\lambda(0) = p \in I^+(q) \subseteq I^+(C)$ and $\lambda(1) = r \in I^-(q) \subseteq M - I^+(C)$. Thus $\partial I^+(C)$ can have no edge. □

Definition 3.3. *If $\gamma : I(\gamma) \to M$, where $I(\gamma)$ is an open interval of $\mathbb{R}$, is a future-directed causal curve then a $p \in M$ is called a future-endpoint of γ if for all of its open neighborhoods there exists a $t_0 \in I(\gamma)$ we have that if $t > t_0$ then $\gamma(t)$ lives in this neighborhood. A past endpoint is defined similarly for past-directed causal curves. A future/past-directed causal is called future/past-inextendible if it admits no future/past endpoints. Finally, a continuous curve λ is called future-directed causal if for all of its points $\lambda(\tau)$ and all $O_{\lambda(\tau)}$ open neighborhoods of $\lambda(\tau)$ we have that if $t_1 < t_2$ with $\lambda(t_1)$, $\lambda(t_2) \in O_{\lambda(\tau)}$ then $\lambda(t_2) \in I^+(\lambda(t_1))$.*

Notice that if the curve admits a future-endpoint then we can extend it to the future by taking a normal neighborhood and gluing it with a radial future-directed geodesic. Of course, we can not expect the regularity to remain the same (but we will have a continuous extension).

Before diving into more theorems let us make a simple remark. Any causal curve will have a past/future inextendbile extension. This is merely an application of Zorn's lemma. If $\gamma : I(\gamma) \to M$ is a past-directed causal curve then we consider the set:

$$\mathcal{A} \coloneqq \{\delta : I(\delta) \to M : I(\delta) \text{ is open, } \delta \text{ is past-directed causal and extends } \gamma\}$$

and we make it a poset by considering $\delta_1 \leq \delta_2 \iff \delta_2$ extends δ_1. If $(\delta_i : I(\delta_i) \to \dots)_{i\in I}$ is a chain of $\mathcal{A}$ then the curve $\delta : I(\delta) \to M$ such that $I(\delta) \coloneqq \bigcup_{i\in I} I(\delta_i)$ and $\delta|_{I(\delta_i)} \coloneqq \delta_i$ is well defined and an upper-bound of the chain. From Zorn's lemma there must exist a maximal element of this set. This maximal element

ought to be an inextendible curve for otherwise we would be able to extend it a bit further (as mentioned earlier) which is a contradiction...
We, only, mention the following technical lemma.

Lemma 3.5. *If λ is a past-inextendible causal that passes through $p \in M$ then for any $q \in I^+(p)$ there exists $\gamma \subseteq I^+(\lambda)$ which is a past-inextendible timelike that passes through q.*

Definition 3.4. *A $p \in M$ is a limit point of the sequence $(\lambda_n)_{n=1}^{\infty}$ if for all open neighborhoods of p, U_p, if there exist infinite $n \in \mathbb{N}$ such that $\lambda_n \cap U_p \neq \emptyset$. λ is a limit curve of $(\lambda_n)_{n=1}^{\infty}$ if there exists a subsequence for which the points of the curve λ are convergence points of this subsequence.*

The following lemma is also very technical and belongs in the family of "limit curve theorems". See (EM) for more.

Lemma 3.6 (Limit-curve theorem)**.** *If $(\lambda_n)_{n=1}^{\infty}$ is a sequence of future-directed causal curves which admits a limit point $p \in M$ then there exists a future-directed causal curve λ which passes through p and is a limit curve of the sequence.*

The regularity of the sequence is not inherited by the curve. Thus we consider the limit curve to be, merely, continuous. A typical application of this lemma is theorems of the following form:

Theorem 3.7. *If C is closed and $p \in \partial I^+(C) \setminus C$ then there exists a past-directed null geodesic segment starting at p which is either past-inextendible or has an endpoint at C.*

Proof. We follow (RW, p. 194). Since $p \in \partial I^+(C) \setminus C$ then there exists a sequence $(q_n)_n \subseteq I^+(C)$ such that it converges to p. We consider $c_n \in C$ such that $c_n \xrightarrow{+} q_n$ or past-directed timelike curves $\lambda_n : q_n \xrightarrow{-} c_n$. We may assume $q_n \notin C$ for all n, without loss of generality, since there can no exist infinite n's such that $q_n \in C$ (otherwise, since C is closed, p would live inside C). Then we consider the manifold $N \coloneqq M \setminus C$ which inherits the Lorentz metric from M since C is closed (thus N is open). Then λ_n become past-inextendible timelike curves in N with p being a limit point of theirs. From lemma 3.6 we have that there exists a causal λ which is past-inextendible in N and passes through p. Since λ_n are past-directed timelike that end up in C we have that $\lambda_n \subseteq I^+(C)$ and so $\lambda \subseteq \overline{I^+(C)}$. If we suppose that λ intersects $I^+(C)$ then there exists a $t > 0$ and a $c \in C$ such that $c \xrightarrow{+,\,\text{timelike}} \lambda(t) \xrightarrow{+,\,\text{causal}} p$ and thus $p \in I^+(C) \cap \partial I^+(C)$ which is impossible since $I^+(C)$ is an open set. Thus, in the past direction, we have $\lambda \in \partial I^+(C)$. Now from the fact that $\partial I^+(C)$ is achronal (corollary 3.3) we have that λ may only be a null geodesic. Now gluing back C, we will have that either it remains past-inextendible or it has an endpoint on C. □

One of the most popular concepts of physics in pop-culture is undeniably the notion of time-travel. Although, it would be fun to travel back in time [18] such a concept would create paradoxes (the grandfather paradox, for example). We would like to avoid pathological spacetimes where we can have timelike curves that are, or tend to become [19], closed and so we will try to impose natural conditions on our spacetime that prevents such anomalies.
For starters, we know, for sure, that our spacetime can not be compact.

Proposition 3.1. *If (M, g) is a compact spacetime then it admits closed timelike curves.*

Proof. We follow (BO, p. 407). Since for all $p \in M$, $I^+(p)$ is open, the set $(I^+(p))_{p\in M}$ is an open cover. Thus we have a finite subcover $(I^+(p_i))_{i=1}^N$ such that, without loss of generality, $I^+(p_1) \not\subseteq \bigcup_{i=2}^{N} I^+(p_i)$. Now if $p_1 \notin I^+(p_1)$ (that is, we do not have a closed timelike curve around p_1) we have that $p_1 \in \bigcup_{i=2}^{N} I^+(p_i)$ therefore there exists a $j \in \{2, \ldots, N\}$ such that $p_1 \in I^+(p_j)$. But then if $q \in I^+(p_1)$ then $q \in I^+(I^+(p_j)) = I^+(p_j)$, that is $I^+(p_1) \subseteq I^+(p_j)$, a contradiction. We conclude $p_1 \in I^+(p_1)$ and so we have a closed timelike curve around p_1. □

[18]Since, a travel forward in time is, in a sense, possible in special relativity (twin paradox).
[19]And so a small perturbation of the metric would create closed timelike curves.

Definition 3.5. *A spacetime (M, g) is called strongly-causal if for all $p \in M$ and O_p open neighborhoods of p there exists a $V_p \subseteq O_p$ such that every causal curve enters V_p at most once.*

Of course in such a space a closed timelike would not be possible since this curve would have a periodic extension to $\mathbb{R}$ making it possible to find neighborhoods that our curve enters infinitely many times. One could argue that there is a risk of having a curve entering the above neighborhood once and then going around and around the neighborhood, creating almost-closed curves. This is not possible due to the following lemma:

Lemma 3.8. *If $K \subseteq M$ is a compact subset of our spacetime and λ an inextendible causal curve that lies entirely on K then it admits future and past endpoints.*

Proof. We follow (RW, p. 197). Take any sequence $t_i \to \infty$ and set $a_i \coloneqq \lambda(t_i) \in K$. Since K is compact $(a_i)_{i=1}^{\infty}$ has an accumulation point in K. Call it $p \in M$. If p is not an endpoint of λ then there exists an open neighborhood of p, O_p, such that for a sequence $t_l \to \infty$ it holds that $\lambda(t_l) \notin O_p$. But then this neighborhood has a subneighborhood $p \in V_p \subseteq O_p$ such that λ can enter it at most once. But it is obvious that λ enters and exits the neighborhood an infinite amount of times (since $\lambda(t_{i_j}) \to p$ and $\lambda(t_l) \notin V_p$). This violates the strong-causality condition and therefore we have arrived at a contradiction. $\square$

A stronger causality condition is the stable causality of a spacetime. Geometrically, we consider a new metric with bigger timecones and we expect no closed timelike curves to appear. In a sense, we allow more curves to be considered timelike and thus we assert that we are safe from perturbing our metric to catch timelike loops.

Definition 3.6. *A spacetime (M, g) is called stably-causal if there exists a global timelike vector field $U \in \mathfrak{X}(M)$ such that the metric $\tilde{g} \coloneqq g - U^* \otimes U^*$ admits no closed timelike curves* [20].

We make a few remarks before the characterization of stable-causality. Firstly, if X is a timelike vector (field) for the metric g then the same goes for the modified metric $\tilde{g}$. Indeed,

$$\tilde{g}(X, X) = g(X, X) - g(U, X)g(U, X) = g(X, X) - g(U, X)^2 < 0.$$

Moreover, if X, Y live in the same timecone in g then the same goes for $\tilde{g}$ (this is why we claim $\tilde{g}$ posseses bigger timecones). Indeed,

$$\tilde{g}(X, Y) = g(X, Y) - g(U, X)g(U, Y) < 0, \text{ since } g(U, X) \text{ and } g(U, Y) \text{ have the same sign.}$$

Theorem 3.9. *A spacetime (M, g) is stably-causal if and only if there exists a smooth function $f \in C^\infty(M)$ such that $grad(f)$ is timelike.*

Proof. We follow (RW, p. 198) for the one direction and (HE, p. 199) for the converse. If such a smooth function exists then we set $U \coloneqq grad(f)$ and we find that the inverse of the metric has components

$$\tilde{g}^{ab} = g^{ab} - \frac{1}{1 - g(U, U)} U_a U_b.$$

[20] Where U^* is the dual form of the field U, i.e. $U^* = g(U, \cdot)$.

Since $A \coloneqq \dfrac{1}{1-g(U,U)} > 0$ and $U_a = \displaystyle\sum_k g^{ka}\frac{\partial f}{\partial x_k}$ then if $\overline{grad}(f)$ is the induced $\tilde{g}-$gradient, we have:

$$\begin{aligned}
\tilde{g}(\overline{grad}(f),\overline{grad}(f)) &= \overline{grad}(f)(f) \\
&= \sum_{a,b}\tilde{g}^{ab}\frac{\partial f}{\partial x_a}\frac{\partial f}{\partial x_b} \\
&= \sum_{a,b} g^{ab}\frac{\partial f}{\partial x_a}\frac{\partial f}{\partial x_b} - A\sum_{a,b} U_a U_b \frac{\partial f}{\partial x_a}\frac{\partial f}{\partial x_b} \\
&= g(grad(f),grad(f)) - A\sum_{\substack{a,b\\k,l}} g^{ka}g^{lb}\frac{\partial f}{\partial x_a}\frac{\partial f}{\partial x_b}\frac{\partial f}{\partial x_k}\frac{\partial f}{\partial x_l} \\
&= g(grad(f),grad(f)) - A\sum_{k,a} g^{ka}\frac{\partial f}{\partial x_a}\frac{\partial f}{\partial x_k}\sum_{l,b} g^{lb}\frac{\partial f}{\partial x_l}\frac{\partial f}{\partial x_b} \\
&= g(grad(f),grad(f) - Ag(grad(f),grad(f))^2 < 0,
\end{aligned}$$

and so $\overline{grad}(f)$ is timelike in $\tilde{g}$ as well. If $grad(f)$ is past-directed (if it is not then take $-f$ as your function) then we calculate

$$\begin{aligned}
\tilde{g}(grad(f),\overline{grad}(f)) &= grad(f)(f) \\
&= g(grad(f),grad(f)) < 0
\end{aligned}$$

and so $\overline{grad}(f)$ and $grad(f)$ live in the same timecone. That means $\overline{grad}(f)$ is also past-directed. If γ is future-directed causal curve we have $g(\overline{grad}(f),\dot{\gamma}) = \dot{\gamma}f = (f\circ\gamma)' > 0$ and thus $f\circ\gamma$ is strictly increasing. This prohibits γ from being a closed causal curve since we would have $(f\circ\gamma)(\tau_1) = (f\circ\gamma)(\tau_2)$ for $\tau_1 \neq \tau_2$.
For the converse, we introduce a new probability measure for M as follows: (we follow (HR, p. 114) for this one).
We consider $(U_n,\varphi_n)_{n=1}^{\infty}$ a countable smooth oriented atlas and $(\psi_n)_{n\in\mathbb{N}}$ a smooth partition of unity subordinated to this atlas. Then we consider a Riemannian metric h and $m_i \coloneqq \displaystyle\int_M \psi_i\,\mu_h$ where μ_h is the volume element of (M,h). Set as ω the top-form $\omega \coloneqq \displaystyle\sum_{n=1}^{\infty}\frac{1}{m_n 2^n}\psi_n\mu_h$. We define the linear functional $\Lambda : C_c^\infty(M)\to\mathbb{R}$ by the formula $\Lambda(f) \coloneqq \displaystyle\int_M f\,\omega$. Riesz's representation theorem asserts that there exists a measure $(M,\mathcal{M},\mu)$ where $\mathcal{M}$ is a σ-algebra that contains the borel σ-algebra such that for all $C_c^\infty(M)$ it holds $\displaystyle\int_M f\omega = \int f d\mu$. By setting $p_N \coloneqq \displaystyle\sum_{j=1}^{N}\psi_j$ we have that $p_j(x)\to 1$ for all $x\in M$ and it is increasing, positive sequence. By Lebesgue's monotone convergence theorem we have that

$$\mu(M) = \int d\mu = \lim_{j\to\infty}\int p_j\,d\mu = 1.$$

Also, for all open sets U, $\mu(U) > 0$. Indeed, consider V a precompact subset of U such that $\overline{V}\subseteq U$ and consider a bump function ρ on $\overline{V}$ with support on U [21]. Then,

$$\mu(U) = \int_U d\mu \overset{0\le\rho\le 1}{\geq} \int_U \rho\,d\mu = \int_M \rho\,d\mu = \int_M \rho\,\omega > 0.$$

Consider $a\in[0,3]$ and the family of metrics $g_a \coloneqq g - a\cdot U^*\otimes U^*$. We then consider the function $\theta(p,a) \coloneqq I^-(p,g_a)$[22]. Our function θ has some nice properties and some bad properties. For example, it is easy to see that our function is bounded and , for fixed a, it increases along future-directed causal curves

[21] This is possible due to the existence of partitions of unity.
[22] The past set of p with respect to the metric g_a.

(since if $t_1 < t_2$ then if $p \in I^-(\gamma(t_1), g_a)$ then $p \xrightarrow{+,\,\text{timelike}} \gamma(t_1) \xrightarrow{+,\,\text{causal}} \gamma(t_2)$ and so $p \in I^-(\gamma(t_2), g_a)$ that is $I^-(\gamma(t_1), g_a) \subseteq I^-(\gamma(t_2), g_a) \implies \theta(\gamma(t_1), a) \leq \theta(\gamma(t_2), a)$). But it is not necessarily continuous. We go around this by considering

$$\tilde{\theta}(p) := \int_1^2 \theta(p, a)\, da$$

which is continuous (see (HE, p. 200)) and preserves the nice properties of θ. We then smooth it out via usual techniques from analysis (by considering mollifiers). □

Corollary 3.4. *Stable-causality implies strong-causality.*

Proof. See (SM, p. 39). □

Definition 3.7. *The future domain of dependence of a closed and achronal set S is defined as*

$$D^+(S) := \{q \in M : \textit{every past inextendible causal curve passing through } q \textit{ intersects } S\}.$$

The past domain of dependence $D^-(S)$ is defined similarly.
A spacetime that satisfies $D(\Sigma) := D^+(\Sigma) \cup D^-(\Sigma) = M$ for some closed and achronal Σ is said to be a globally hyperbolic spacetime. Σ is called a Cauchy hypersurface of M.

It is easy to see that a Cauchy hypersurface must be also edgeless. If $q \in \Sigma$ and we consider $p \in I^+(q)$, $r \in I^-(q)$ and $\lambda : p \to r$ a timelike curve then this curve has an inextendible extension. Without loss of generality, let λ be past-directed. Then λ cannot intersect Σ after it passes through r since then we would have $r \xrightarrow{-} \lambda(t) \in \Sigma$ and $r \in I^-(\Sigma)$ thus $\lambda(t) \in \Sigma \cap I^-(\Sigma)$ violating the achronality of Σ. Similarly, it cannot intersect Σ before passing through p. Therefore the intersection happens in between the two points thus Σ can have no edge. From theorem 3.4 we conclude :

Theorem 3.10. *A Cauchy hypersurface is a closed, achronal and edgeless set such that every inextendible curve intersects it. Therefore it is a C^0-hypersurface of M.*

It is, also, easy to see that $I^-(S) \cap D^+(S) = \emptyset$ for all closed and achronal sets S. The following theorem characterizes the closure of $D^+(S)$.

Proposition 3.2. *$p \in \overline{D^+(S)}$ if and only if every past-inextendible timelike curve passing through p intersects S.*

Proof. We follow (RW, p. 202). Indeed if $p \in \overline{D^+(S)}$ and there exists a past-inextendbile timelike λ that passes through p and does not intersect S then it is certain that $p \notin S$. Since S is closed we can consider a geodesically convex neighborhood of p, U, such that $U \subseteq M \setminus S$. Consider, then, $r \in \lambda \cap U$. Then, $p \in I^+(r) \cap U$ and so if we considered as V_p the projection (through the exponential map) of the positively-oriented timecone in the tangent space of r (such that it remains inside $I^+(r) \cap U$) we would have $p \in V_p \subseteq I^+(r) \cap U$ and every other $q \in V_p$ can be connexted to r via a radial geodesic (which will be timelike because of proposition 2.25 and remain in $U \subseteq M \setminus S$) and then be extended to infinity via λ. By making a smooth approximation of this curve we conclude that $V_p \cap D^+(S) = \emptyset$ which is impossible since $p \in \overline{D^+(S)}$.
Conversely, if every past-inextendible timelike intersects S then either $p \in I^+(S)$ or $p \in S$. If the latter holds we are done. If $p \in I^+(S)$ then every neighborhood of p intersects $I^-(p) \cap I^+(S)$ since of O_p is open neighborhood of p then $O_p \cap I^+(S)$ is also an open neighborhood of p and every open neighborhood of p intersects its past and future. We will show that $I^-(p) \cap I^+(S) \subseteq D^+(S)$. If $q \in I^-(p) \cap I^+(S)$ and there exists a past-inextendible causal from q that doesn't intersect S then we consider $\gamma \subseteq I^+(\lambda)$ a past-inextendible timelike that passes through p given by lemma 3.5. Since $q \in I^+(S)$, λ remains in $I^+(S)$ for some time. If it remains there forever then $\gamma \subseteq I^+(S)$ and therefore it is not possible for γ to intersect S because of the achronality of S. If λ leaves $I^+(S)$ then it should cross $\partial I^+(S) \setminus S$ (since it does not intersect S). Let $r \in \partial I^+(S) \setminus S$ the point of intersection. We consider a timelike $p \xrightarrow{-} q \xrightarrow{-,\lambda} r$. The segment $p \to q$ cannot intersect S because $q \in I^+(S)$ and S is achronal. The segment $q \xrightarrow{\lambda} r$, also, does not intersect S by our hypothesis. We can approximate this concatenated curve, sufficiently close, with a

smooth timelike $p \xrightarrow{-} r$ such that it does not intersect S. We then consider its inextendible (maximal) extension. Since $r \in \partial I^+(S)$ an intersection with S is not possible since then we would have $r \in I^+(S)$ which is a contradiction. $\square$

By definition, every inextendible causal curve, in a globally hyperbolic space, must intersect the Cauchy hypersurface Σ at some point. The point need not be unique since our curve might be a null geodesic which does not violate the achronality (see (BS, p. 3)). An inextendible timelike curve intersects Σ at exactly one point, though. The following proposition asserts that an inextendible causal curve intersects Σ and both $I^{\pm}(\Sigma)$.

Proposition 3.3. *If Σ is a Cauchy hypersurface then every inextendible causal curve intersects Σ, $I^+(\Sigma)$ and $I^-(\Sigma)$.*

Proof. We follow (RW, p. 202). By definition they intersect Σ. Let λ be an inextendible causal curve that does not intersect $I^-(\Sigma)$ then since $M = \Sigma \cup I^+(\Sigma) \cup I^-(\Sigma)$ we have $\lambda \subseteq \Sigma \cup I^+(\Sigma)$. Consider $\gamma \subseteq I^+(\lambda) \subseteq I^+(\Sigma)$ a past-inextendible timelike given by lemma 3.5. It is obvious that γ cannot intersect Σ in the past and thus it must intersect it in the future. If by γ^{-1} we denote the time-reversal curve of γ then $\gamma^{-1} : \gamma(t) \xrightarrow{+} \sigma \in \Sigma$ where $t > 0$ (that is $\gamma(t) \in I^-(\gamma(0))$). Then, $\sigma \in \Sigma \cap I^+(I^+(\Sigma)) = \Sigma \cap I^+(\Sigma)$ which is impossible. We have constructed an inextendible timelike curve that does not intersect Σ and thus arrived at a contradiction. $\square$

We strengthen the above proposition into a characterization of Σ with the following definition and proposition.

Definition 3.8. *We define the future Cauchy horizon of the closed and achronal set S as*

$$H^+(S) \coloneqq \overline{D^+(S)} \setminus I^-(D^+(S)).$$

It is immediate that $I^-(H^+(S)) \cap H^+(S) = \emptyset$ and thus $H^+(S)$ is achronal.

Proposition 3.4. *If $p \in H^+(S) \setminus edge(S)$ then p lies on a null geodesic segment contained on $H^+(S)$ which is either a part of a past-inextendible null geodesic that lies entirely on $H^+(S)$ or reaches a past-endpoint at the edge of S.*

Proof. See (EM, p.). $\square$

It can be proved that $H(S) \coloneqq H^+(S) \cup H^-(S) = \partial D(S)$. It follows (since we consider our spacetimes to be connected) that a closed and achronal set $\Sigma \neq \emptyset$ is a Cauchy hypersurface if and only if $H(\Sigma) = \emptyset$. Indeed, $H(\Sigma) = \emptyset$ if and only if $D(\Sigma)$ is clopen which is if and only if $D(\Sigma) = M$ (since $\Sigma \subseteq D(\Sigma)$ and $\Sigma \neq \emptyset$ thus $D(\Sigma)$ is not empty). A more interesting characterization that strengthens proposition 3.3.

Theorem 3.11. *If $\Sigma \neq \emptyset$ is closed, achronal and edgeless then it is a Cauchy hypersurface if and only if every inextendible null geodesic intersects Σ, $I^+(\Sigma)$ and $I^-(\Sigma)$.*

Proof. We follow (RW, p. 205). If $H^+(\Sigma) \neq \emptyset$ and $p \in H^+(\Sigma)$ then since Σ is edgeless there exists a past-inextendible null geodesic that lies entirely on $H^+(\Sigma)$ and starts at p. Since $\Sigma \subseteq D^+(\Sigma)$ the null geodesic λ does not intersect $I^-(\Sigma)$ in the past. Therefore, it must intersect it in the future. But $p \in I^+(\Sigma) \cup \Sigma$ and if $\lambda(t') \in I^-(\Sigma)$ for $t' < 0$ then there exists a $\sigma \in \Sigma$ such that $\sigma \in I^+(\lambda(t'))$. But, $\lambda(t') \in J^+(p)$ and thus $p \xrightarrow{+,\text{causal}} \lambda(t') \xrightarrow{+,\text{timelike}} \sigma$ thus $\sigma \in I^+(\Sigma)$ which violates the achronality of Σ. We have arrived at a contradiction. The case $H^-(\Sigma) \neq \emptyset$ is treated similarly. $\square$

Globally hyperbolic spacetimes exhibit great causal behaviour.

Theorem 3.12. *Globally hyperbolic implies strongly causal.*

Proof. See (RW, p. 205). $\square$

Our final goal is to prove that globally hyperbolic actually implies stably causal. To do so, we follow (RW, p. 206) and we define $C(p,q)$ the set of all future-directed causal curves from p to q modulo reparametrizations. We endow it with the unique topology having the sets $O(U)$, $U \subseteq M$ open, as its basis where $O(U) := \{\lambda \in C(p,q) : \lambda \subseteq U\}$. It can be proved that in strongly causal spaces the convergence in $C(p,q)$, say $\lambda_n \xrightarrow{C(p,q)} \lambda$, is equivalent to λ being a convergent curve of λ_n.

Theorem 3.13. *If (M,g) is globally hyperbolic then $C(p,q)$ is a compact, Hausdorff and second-countable topological space.*

Proof. See (RW, p. 206). ☐

Corollary 3.5. *If (M,g) is globally hyperbolic and p, $q \in M$ then $J^+(p) \cap J^-(q)$ is compact. It follows that $J^+(x)$ is closed for any $x \in M$.*

Proof. We follow (RW, p. 207). Take $(r_n)_{n=1}^{\infty} \subseteq J^+(p) \cap J^-(q)$. There exist $\lambda_n : p \to q$ future directed causal that pass through r_n and thus $\lambda_n \in C(p,q)$. Since $C(p,q)$ is compact there exists a $\lambda \in C(p,q)$ which is a limit of some subsequence of λ_n. Replace the subsequence with λ_n. If U is any precompact set containing λ then there exists a n_0 such that for $n \geq n_0$ it holds $\lambda_n \subseteq U$ and therefore $r_n \in U \subseteq \overline{U}$ and $\overline{U}$ is compact [23]. But then there exists a subsequence r_{n_k} such that it converges to some $r \in \overline{U}$. It follows that if V is a precompact set containing λ then $\overline{V}$ contains r. If $r \notin \lambda$ then there exist O_r for r and W_λ for λ disjoint open sets such that $r \in O_r$ and $\lambda \subseteq W_\lambda$. Take V_λ precompact such that $\lambda \subseteq V_\lambda \subset \overline{V_\lambda} \subset W_\lambda$. Then from the previous arguements we have $r \in \overline{V_\lambda}$ which is a contradiction. Thus $r \in \lambda$ and therefore $r \in J^+(p) \cap J^-(q)$ which asserts that it is a compact set. We leave as an exercise the claim that $J^+(x)$ is implied closed from the fact that $J^+(p) \cap J^-(q)$ is compact (thus closed). ☐

Theorem 3.14. *Let (M,g) be a globally hyperbolic spacetime with a Cauchy hypersurface Σ and X be the time-orientation global vector field. We consider $r : M \to \Sigma$ such that $r(x) \in \Sigma$ is the unique intersection point of the integral curve passing through $x \in M$ with the Cauchy hypersurface Σ. The function r is continuous, onto with $r|_\Sigma = id_\Sigma$ and thus a retraction. It follows that any Cauchy hypersurface is connected and any two Cauchy hypersurfaces are homeomorphic.*

Proof. We follow (BO, p. 417). Let $\psi : D \to M$ be the flow of X. Recall that D is an open subset of $\mathbb{R} \times M$ and since Σ is a C^0-hypersurface it follows that the set $D(\Sigma) := (\Sigma \times \mathbb{R}) \cap D$ is a C^0-hypersurface of D. The restriction $\psi|_{D(\Sigma)} : D(\Sigma) \to M$ is a continuous bijection since if $x \in M$ then the maximal integral curve passing through x must be an inextendible timelike curve (because otherwise by considering the limit we could extend it a little furter thus violating maximality) and thus it intersects Σ. This proves that $\psi|_{D(\Sigma)}$ is onto. It is also one-to-one since if $\psi(t,x) = \psi(t',x')$ for x, $x' \in \Sigma$ then since two different integral curves cannot intersect we have $x' = \psi(s,x)$. But $\psi_x(s)$ is timelike and thus in order not to violate the achronality of Σ it follows that $s = 0$ and $x' = x$. Then, $\psi(t,x) = \psi(t',x)$ and if $t \neq t'$ then we would have a closed timelike curve which is impossible in strongly causal (thus in globally hyperbolic) spaces. By the invariance of domain theorem, $\psi|_{D(\Sigma)}$ is a homeomorphism since $\dim(D(\Sigma)) = \dim(M)$. For $x \in M$ there exists a unique $(t,y) \in \mathbb{R} \times \Sigma$ such that $x = \psi(t,y)$. Then, $r(x) = r(\psi(t,y)) = y$ which implies that if $\pi_\Sigma : \mathbb{R} \times \Sigma \to \Sigma$ is the natural projection to Σ then $r \circ \psi|_{D(\Sigma)} = \pi_\Sigma$ and so since $\psi|_{D(\Sigma)}$ is a homeomorphism it follows that $r = \pi_\Sigma \circ \psi|_{D(\Sigma)}^{-1}$ which proves that r is continuous, onto and since $r(y) = r(\psi(0,y)) = y$ for $y \in \Sigma$ it is also a retraction.
Now two Cauchy hypersurfaces Σ_1, Σ_2 are homeomorphic since we can consider $r_{\Sigma_1} : M \to \Sigma_1$ and $r_{\Sigma_2} : M \to \Sigma_2$ and take their restrictions $r_{\Sigma_1}|_{\Sigma_2} : \Sigma_2 \to \Sigma_1$ and $r_{\Sigma_2}|_{\Sigma_1} : \Sigma_1 \to \Sigma_2$ which will be inverse maps. ☐

Of course, if the two Cauchy hypersurfaces are smooth hypersurfaces of M then the above proof asserts that they are diffeomorphic.
We conclude this section with a huge theorem first proved by Geroch in 1970.

[23] Such a precompact set always exists since λ is compact. A manifold is locally compact and thus the construction of such a set is immediate.

Theorem 3.15 (Geroch)**.** *Globally hyperbolic implies stably causal. Moreover, if (M, g) is globally hyperbolic then there exists a continuous time function $\tau : M \to \mathbb{R}$ which is onto and if $\gamma : (T_-, T_+) \to M$ is an inextendible causal curve then $\lim_{t\to T_\pm} \tau \circ \gamma = \mp\infty$ with the sign to be minus if the curve is past-directed and plus if it is future-directed. Finally, every level set of τ, say $\tau^{-1}(c)$, is a Cauchy hypersurface and thus M is foliated by Cauchy hypersurfaces. By considering the map $M \ni x \mapsto (\tau(x), r(x)) \in \mathbb{R} \times \Sigma$, where r is the map constructed in theorem 3.14, we have that M is homeomorphic to $\mathbb{R} \times \Sigma$.*

Proof. We follow (RW, p. 209). For the last part see (BS, p. 4). We consider the measure mentioned in the proof of theorem 3.9 such that $\mu(M) = 1$. Then we consider the function $f^-(p) := \mu(J^-(p))$. It can be proved that f^- is continuous and as we have seen it increases along future-directed causal curves. By smoothing it out we get a smooth $\tilde{f^-}$ which has past-directed timelike gradient. This proves stable causality. Returning now to f^- we prove that it goes to zero over past-directed inextendible causal curves. Indeed, if γ is past-directed inextendible causal and $f^- \circ \gamma$ does not go to zero then there exists a sequence $t_i \to \infty$ and an $\varepsilon > 0$ such that $\mu(J^-(\gamma(t_i))) \geq \varepsilon$. It is immediate that $\limsup_{i\to\infty} \mu(J^-(\gamma(t_i))) > 0$ and so $\mu(\limsup_{i\to\infty} J^-(\gamma(t_i))) > 0$ since the measure is finite and thus $\limsup_{i\to\infty} J^-(\gamma(t_i)) \neq \emptyset$. Therefore, there exists an $r \in J^-(\gamma(t))$ for infinite and thus for all t. If $\gamma(0) = q$ this implies $\gamma(t) \in J^+(r) \cap J^-(q)$ which is compact and thus γ must have a past-endpoint in this set, by lemma 3.8, a contradiction. Similarly, if we consider $f^+(p) := \mu(J^+(p))$ then it goes to zero along future-directed inextendible causal curve. The function f^- increases along future-directed causal curves and so f^-/f^+ goes to infinity along future-directed inextendible causal curves and to zero along past-directed inextendible causal curves. The desired time function is $\tau := \log(f^-/f^+)$. Its level sets are closed and two p, $q \in \tau^{-1}(c)$ cannot be connected through a causal curve since $\tau \circ \gamma$ increases over future-directed causal curves. That is, the level sets are achronal. From the fact that $\tau \circ \gamma \to \pm\infty$ for any causal curve it is immediate that its level sets are Cauchy hypersurfaces. $\square$

The existence of smooth spacelike Cauchy hypersurfaces and smooth time function $\tau : M \to \mathbb{R}$ were rigorously proved by Bernal and Sanchez in (BS).

Main result

In this subsection we prove the main theorem. We follow (GN) and (AC).

As we discussed earlier, a geodesically incomplete spacetime is said to be singular. Stephen Hawking found some very natural conditions so that the spacetime admits incomplete future-directed timelike geodesics. Before we state the theorem we explain what the strong energy condition is.

Definition 3.9. *We say that our spacetime (M, g) satisfies the strong energy condition when for all timelike vector fields $X \in \mathfrak{X}(M)$ the Ricci tensor satisfies $Ric(X, X) \geq 0$.*

Although this could be considered a curvature condition we notice that if the metric g solves the Einstein equations $Ric = 8\pi T$ where T is the reduced stress-energy tensor then it is equivalent to say that $T(X, X) \geq 0$ for all timelike vector fields $X \in \mathfrak{X}(M)$. Of course, the strong energy condition is trivially satisfied in vacuum, i.e. $Ric(g) = 0$.

Theorem 3.16 (Hawking)**.** *If (M, g) is a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ satisfying the strong energy condition and such that the mean curvature of Σ, say H, satisfies either $H \leq H_0 < 0$ or $H \geq H_0 > 0$ then (M, g) is singular.*

Perhaps, the only condition that we should be cautious about is the mean curvature condition on the grounds that it may be too strict to be realistic. As we will see, the mean curvature is proportional to the derivative with respect to time (on some suitable coordinate system) of the logarithm of the volume element of Σ. Now, if our universe came into being from a Big Bang then this condition is natural since, in an early stage after the explosion, our universe would expand with a great acceleration and thus its volume would get bigger and bigger.

Let (M, g) be a globally hyperbolic spacetime with a smooth spacelike and orientable Cauchy hypersurface Σ and a smooth time function $t : M \to \mathbb{R}$. There exists, then, a unit normal vector field N to Σ and we consider an open set $U \subseteq \mathbb{R} \times \Sigma$ such that $\{0\} \times \Sigma \subseteq U$. We consider a modification of the exponential map (which we will, also, denote by exp), $\exp : U \to M$ such that $\exp(t, x) := \gamma_x(t)$ where γ_x is the $(x, N_x)-$timelike geodesic [24]. If we identify the tangent space of U with $T_t\mathbb{R} \times T_x\Sigma$ then it is obvious that for $(0, x) \in U$ the modified exponential map induces a diffeomorphism in a neighborhood of the point since its differential $\exp_{*,\,(0,x)}$ maps ∂_t to N_x and $\partial_{x_i} \in T_x\Sigma$ to $\partial_{x_i} \in T_xM$ (by considering some local slice chart on Σ). A critical point of the exponential map (that is, a point where $\exp(\cdot, x)$ ceases to be a diffeomorphism in some neighborhood of $(\cdot, x)$) will be called a conjugate point of x along γ_x.

Suppose, now, that the point $q = \exp(t_0, x)$ is not conjugate to x along γ_x and for all $0 \leq t \leq t_0$ the curve γ_x does not admit conjugate points. We can, then, cover $\gamma_x|_{[0,t_0]}$ with finite neighborhoods such that exp is a diffeomorphism there or equivalently that it is a local diffeomorphism in the union of these neighborhoods. By considering a local slice chart on $x_0 \in \Sigma$, say $(V \cap \Sigma, \varphi)$, we have that a local chart of the form $(t, x_1, x_2, x_3) = (id \times \varphi) \circ \exp^{-1}(y)$ is induced in each of these neighborhoods. Also, t and x_i for $i = 1, 2, 3$ can be defined in the union of these neighborhoods by the gluing lemma (for example, consider $t|_{V_i} := \pi_{\mathbb{R}} \circ (id \times \varphi) \circ \exp^{-1}|_{\exp(V_i)}$ where V_i is one of these neighborhoods [25]). Set as V the union of these neighborhoods. Then,

$$\partial_t|_{\exp(t,x)} = \partial_t|_{(t,x)}(\cdot \circ \exp) = \dot{\gamma}_x(t)$$

and so $g_{00} = g(\partial_t, \partial_t) = -1$ on V and also

$$\begin{aligned}
\frac{\partial g_{0i}}{\partial t} &= g(\nabla_t \partial_t, \partial_{x_i}) + g(\partial_t, \nabla_t \partial_{x_i}) \\
&= 0 + g(\partial_t, \nabla_i \partial_t) \\
&= \partial_{x_i}(g_{00}) \\
&= 0.
\end{aligned}$$

[24] The timelike geodesic with initial conditions (x, N_x).

[25] It shouldn't be confused with the time function.

Since $g_{0i}|_\Sigma = 0$ we have $g_{0i} = 0$ on V. It follows that $g = -dt^2 + \tilde{g}$ where $\tilde{g} = \sum_{i,j=1}^{3} g_{ij}\, dx_i \otimes dx_j$. It follows that $\tilde{g}$ is a Riemmanian metric on Σ since Σ is spacelike. We proved the following useful lemma:

Lemma 3.17. *If γ_{x_0} admits no conjugate points for $0 \le t \le t_0$ then for some neighborhood of $\gamma_{x_0}|_{[0,t_0]}$ the metric takes the form $g = -dt^2 + \tilde{g}$ where $\tilde{g}$ is some spatial metric such that $\tilde{g}|_\Sigma$ is positive-definite.*

Proof. See (AC, p. 12). □

The Christoffel symbols read:

$$\Gamma^l_{00} = 0, \text{ for } l = 0, 1, 2, 3$$

and

$$\Gamma^i_{0j} = \frac{1}{2}\sum_{k=1}^{3} \tilde{g}^{ik}\frac{\partial \tilde{g}_{jk}}{\partial t}.$$

It follows that

$$R_{00} = Ric(\partial_t, \partial_t) = -\frac{\partial \theta}{\partial t} - \sum_{i,j,k,l=1}^{3} \tilde{g}^{ik}\tilde{g}^{jl} b_{jk} b_{il}$$

for $b_{ij} = \frac{1}{2}\frac{\partial(\tilde{g}_{ij})}{\partial_t}$ and $\theta = \sum_{i,j=1}^{3} \tilde{g}^{ij} b_{ij}$.

Notice that

$$\begin{aligned}
\theta &= \frac{1}{2}\sum_{i,j=1}^{3} \tilde{g}^{ij}\frac{\partial \tilde{g}_{ij}}{\partial_t} \\
&= \frac{1}{2} tr((\tilde{g}^{ij})(\partial_t(\tilde{g}_{ij}))) \\
&= \frac{1}{2}\partial_t(\log(\det(\tilde{g}))) \\
&= \partial_t(\log(\sqrt{\det(\tilde{g})})).
\end{aligned}$$

Now we will prove that θ is in fact the mean curvature of Σ (modulo a sign). Recall that the mean curvature of a hypersurface is the trace of its shape operator, where its shape operator is defined as the unique, symmetric linear operator $S : \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ such that

$$g(S(X), Y) = g(II(X, Y), N)$$

where N is the normal vector field on Σ and $II(X, Y) \coloneqq (\nabla_X Y)^\perp$ the second fundamental form (see (BO, p. 107)). Since $N = \partial_t$ in the coordinates we defined above, we compute

$$\begin{aligned}
g(S(\partial_{x_i}), \partial_{x_j}) &= g(\nabla_i \partial_{x_j}, \partial_t) \\
&= -g(\nabla_i \partial_t, \partial_{x_j}) \\
&= -g(\nabla_t \partial x_i, \partial_{x_j}) \\
&= -\frac{\partial \tilde{g}_{ij}}{\partial t} + g(\partial_{x_i}, \nabla_t \partial_{x_j}) \\
&= -\frac{\partial \tilde{g}_{ij}}{\partial t} - g(S(\partial_{x_j}), \partial_{x_i}) \\
&= -\frac{1}{2}\frac{\partial \tilde{g}_{ij}}{\partial t}.
\end{aligned}$$

Since $H = tr(S)$ we conclude $H = -\frac{1}{2}\sum_{i,j=1}^{3} g^{ij}\frac{\partial g_{ij}}{\partial t} = -\theta$.

We now make the first steps towards the proof of our main theorem.

Lemma 3.18. *Let (M, g) be a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ, satisfying the strong energy condition and suppose that there exists a point $x \in \Sigma$ such that $\theta(0, x) = \theta_0 < 0$. Then the orthogonal (to Σ) timelike geodesic γ_x admits a conjugate point for time $T \leq -3/\theta_0$.*

Proof. We follow (GN, p. 301). Suppose that γ_x admits no conjugate points for $t \leq -3/\theta_0$ and that it can be extended that far. Then the coordinate system we established above is well-defined for all times $0 \leq t \leq -3/\theta_0$. The strong energy condition asserts that $R_{00} \geq 0$ and so

$$\frac{\partial \theta}{\partial t} + \sum_{i,j,k,l=1}^{3} \tilde{g}^{ik} \tilde{g}^{jl} b_{jk} b_{il} \leq 0.$$

Notice that

$$\sum_{i,j,k,l=1}^{3} \tilde{g}^{ik} \tilde{g}^{jl} b_{jk} b_{il} = \sum_{i,j=1}^{3} a_{ij} a_{ji} = tr((a_{ij})(a_{ij})^T) \geq \frac{1}{3} tr(a_{ij})^2$$

where $a_{ij} = \sum_{k=1}^{3} \tilde{g}^{ik} b_{kj}$ and we used the inequality $tr(AA^T) \geq \frac{1}{n} tr(A)^2$ for A an $n \times n$ matrix. Notice that $tr(a_{ij}) = \theta$ and so the inequality becomes

$$\frac{\partial \theta}{\partial t} + \frac{\theta^2}{3} \leq 0 \implies \frac{\partial \theta}{\partial t} \leq -\frac{\theta^2}{3} \leq 0.$$

Thus, θ is decreasing along γ_x and by integrating the above inequality we have

$$-\frac{1}{\theta} + \frac{1}{\theta_0} + \frac{t}{3} \leq 0 \implies \frac{1}{\theta_0} + \frac{t}{3} \leq \frac{1}{\theta} < 0,$$

since $\theta \leq \theta_0 < 0$ along γ_x. By sending $t \to -3/\theta_0$ we arrive at a contradiction since the coordinate system should be well defined there and we found that θ blows up while approaching this value. $\square$

Lemma 3.19. *Let (M, g) be a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ and γ a timelike geodesic that intersects Σ orthogonally, connects Σ with $p \in M$ and admits a conjugate point q before reaching p. Then γ does not maximize proper time.*

Proof. We follow (GN, p. 301). Since q is a conjugate point we can consider an orthogonal to Σ timelike geodesic $\tilde{\gamma}$ that connects Σ with q and such that the proper time of $\tilde{\gamma}$, $\tau(\tilde{\gamma})$, equals the proper time of $\gamma|_{[\Sigma \to q]}$, $\tau(\gamma|_{[\Sigma \to q]})$. Consider a geodesically convex neighborhood of q, U_q, and take $r \in \tilde{\gamma} \cap U_q$ and $s \in \gamma \cap U_q$. Consider the concatenation

$$\delta := \tilde{\gamma}|_{[\Sigma \to r]} * [r \to s] * \gamma|_{[s \to p]}$$

where $[r \to s]$ is the radial geodesic from r to s inside U_q. We claim $\tau(\delta) > \tau(\gamma)$. Indeed,

$$\tau(\gamma|_{[\Sigma \to q]}) = \tau(\tilde{\gamma}) = \tau(\tilde{\gamma}|_{[\Sigma \to r]}) + \tau(\tilde{\gamma}|_{[r \to q]})$$

and so

$$\begin{aligned}
\tau(\delta) &= \tau(\tilde{\gamma}|_{[\Sigma \to r]}) + \tau([r \to s]) + \tau(\gamma|_{[s \to p]}) \\
&> \tau(\tilde{\gamma}|_{[\Sigma \to r]}) + \tau(\tilde{\gamma}|_{[r \to q]}) + \tau(\gamma|_{[q \to s]}) + \tau(\gamma|_{[s \to p]}) \\
&= \tau(\tilde{\gamma}) + \tau(\gamma|_{[q \to p]}) \\
&= \tau(\gamma).
\end{aligned}$$

The curve δ is a piecewise smooth timelike curve with $\tau(\delta) > \tau(\gamma)$. We can, then, smooth it out to find a smooth timelike curve with this property.
In a more rigorous manner, if q is a conjugate point along $\gamma : [0, a] \to M$ (set $q = \gamma(t_0)$) then we consider a non-trivial Jacobi field $\tilde{Y}$ such that $\tilde{Y}(0) = 0$ and $\tilde{Y}(t_0) = 0$ (and thus it is orthogonal to γ). Next, consider

as Y the vector field that coincides with $\tilde{Y}$ between $\Sigma \ni \gamma(0)$ and q and vanishes from q to p. We, also, consider the parallel (to γ) vector field $\tilde{Z}$ given by the initial condition $\tilde{Z}(t_0) = -\nabla_{\dot{\gamma}(t_0)}\tilde{Y}(t_0) = -D_t(\tilde{Y})(t_0)$ (and thus necessarily spacelike) and $Z := \theta\tilde{Z}$ where θ is a smooth function that vanishes at endpoints and $\theta(t_0) = 1$. Finally, set $Y_\varepsilon := Y + \varepsilon Z$ and consider Γ_ε to be the smooth variation of the curve γ with variation field Y_ε. By considering the proper-time functional

$$\tau_{\Gamma_\varepsilon}(s) := \int_0^a \sqrt{-g(\partial_t(\Gamma_{\varepsilon,s}), \partial_t(\Gamma_{\varepsilon,s}))}\, dt$$

one can prove (see for example (KA, p. 99)) that

$$\tau''_{\Gamma_\varepsilon}(0) = -\int_0^a \Big(g(D_t(Y_\varepsilon), D_t(Y_\varepsilon)) + Rm(\dot{\gamma}, Y_\varepsilon, \dot{\gamma}, Y_\varepsilon) \Big)\, dt = I(Y_\varepsilon, Y_\varepsilon).$$

We call the bilinear form $I(V, W) := -\int_0^a \Big(g(D_t V, D_t W) + Rm(\dot{\gamma}, V, \dot{\gamma}, W) \Big)\, dt$ the index form and it is clearly symmetric. Therefore,

$$\tau''_{\Gamma_\varepsilon}(0) = I(Y, Y) + 2\varepsilon I(Y, Z) + \varepsilon^2 I(Z, Z)$$

and since Y is a Jacobi field (see definition 2.15) between Σ and q and zero elsewhere, we get $I(Y, Y) = 0$. Moreover,

$$\begin{aligned} I(Y, Z) &= -\int_0^{t_0} \Big(g(D_t(Y), D_t(Z)) + g(R(\dot{\gamma}, Y)\dot{\gamma}, Z) \Big)\, dt \\ &= -g(D_t(Y), Z)\Big|_0^{t_0} + \int_0^{t_0} \Big(g(D_t^2(Y), Z) + g(R(Y, \dot{\gamma})\dot{\gamma}, Z) \Big)\, dt \\ &= g(D_t(\tilde{Y})(t_0)), D_t(\tilde{Y})(t_0)) > 0, \end{aligned}$$

where the integral vanishes since Y is a Jacobi field for $t \in [0, t_0]$. This asserts that for sufficiently small $\varepsilon > 0$ we can make $\tau''_{\Gamma_\varepsilon}(0) > 0$ which means that γ can not maximize proper time for otherwise the second variation would be non-positive at $t = 0$ for all $\varepsilon > 0$ (since the functional with this variation would attain a maximum for $t = 0$). For the complete details see (JN, p. 71). □

Lemma 3.20. *Let (M, g) be a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ and $p \in D^+(\Sigma)$. Then, $D^+(\Sigma) \cap J^-(p)$ is compact.*

Proof. See (GN, p. 302). □

Lemma 3.21. *Let (M, g) be a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ and $p \in D^+(\Sigma)$. Then, there exists a timelike geodesic γ from Σ to p which intersects Σ orthogonally and maximizes proper time among all timelike curves from Σ to p.*

Proof. We follow (GN, p. 304). Set $A := D^+(\Sigma) \cap J^-(p)$ (which is compact) and $C(A)$ the compact subsets of A. Also, set $T(\Sigma, p)$ the timelike curves from Σ to p which is a subset of $C(A)$. Consider any distance metric d on M and the Hausdorff metric d_H on $C(A)$ defined by the formula

$$d_H(K, L) := \inf\{\varepsilon > 0 |\ K \subseteq B(L, \varepsilon),\ L \subseteq B(K, \varepsilon)\}.$$

It can be proved that $(C(A), d_H)$ is a compact metric space. Consider $\tau : T(\Sigma, p) \to \mathbb{R}$ the proper time functional

$$\tau(\gamma) := \int_0^{t(p)} \|\dot{\gamma}\|\, ds$$

where we have parametrised by the time-function $t : M \to \mathbb{R}$ and $\mathbb{R}$ is endowed with the topology $\mathcal{T} := \{(-\infty, a) : a \in \bar{\mathbb{R}}\}$. The functional τ is continuous with respect to the topology of $C(A)$ and $(\mathbb{R}, \mathcal{T})$. We continuously extend it to $C(\Sigma, p) := \overline{T(\Sigma, p)}$, which is compact in $C(A)$, by the formula

$$\tau(\delta) := \lim_{\varepsilon \to 0} \sup\{\tau(\gamma) | \gamma \in B_\varepsilon(\delta) \cap T(\Sigma, p)\}.$$

The set $C(\Sigma, p)$ is the set of all continuous causal curves from Σ to p. The compact sets of $(\mathbb{R}, \mathcal{T})$ attain a maximum and so by continuity τ attains a maximum on a causal curve γ. Cover γ with a finite number of geodesically convex neighborhoods V_i, $i = 1, \ldots, N$ and take $p_1, \ldots, p_N$ such that $p_1 = \gamma(0)$, p_i and p_{i+1} live in the same V_i and $p_N = p$. Take, now, a sequence of timelike curves in $T(\Sigma, p)$, say γ_n, that converge, in $C(A)$, to γ. Since the sets $\gamma_n \cap t^{-1}(t(p_i))$ are singletons (t, here, is the time function) consider $p_{i,n} \in \gamma_n \cap t^{-1}(t(p_i))$. It is immediate that $p_{i,n}$ converge to p_i since if W_i is a neighborhood of p_i then there exist $W_{\gamma(t)} \subseteq M \setminus t^{-1}(t(p_i))$ since $\gamma(t) \notin t^{-1}(t(p_i))$ because $t \circ \gamma$ is strictly monotone. Then since we can cover γ by finite of these sets $W_1, \ldots, W_K$ then for all large n we have $\gamma_n \subseteq \bigcup_{i=1}^{K} W_i$ (since γ_n converge to γ with respect to d_H). It follows that $p_{i,n} \in \gamma_n \cap t^{-1}(t(p_i)) \subseteq W_i \cap t^{-1}(t(p_i))$ and so $p_{i,n} \in W_i$ for all large n. For all large n, consider δ_n to be the concatenation of the radial geodesics connecting $p_{i,n}$ to $p_{i+1,n}$ inside V_i. It is obvious that $\tau(\delta_n) \geq \tau(\gamma_n)$. Since $p_{i,n}$ converge to p_i it follows that δ_n converge to δ (the concatenation of radial geodesics connecting p_i to p_{i+1} inside V_i). By sending n to infinity we get $\tau(\delta) \geq \tau(\gamma)$ and so $\tau(\delta) = \tau(\gamma)$. The curve δ is, in fact, a smooth timelike geodesic (and not piecewise smooth) since if it was not smooth we could consider a geodesically convex neighborhood around the points that it breaks and then consider a radial geodesic that would increase its length. Therefore, δ is a smooth timelike geodesic. The following arguement asserts that it intersects Σ orthogonally (we follow (HE, p. 105) here):
Consider $\Gamma : (-\varepsilon, \varepsilon) \times [0, t(p)] \to M$ a smooth variation of δ such that $\Gamma(s, 0)$ is a smooth curve lying entirely on Σ and $\Gamma(s, t(p)) = p$ for all $s \in (-\varepsilon, \varepsilon)$. Since δ maximizes proper time it follows that

$$\tau_\Gamma(s) := \int_0^{t(p)} \sqrt{-g(\partial_t \Gamma_s, \partial_t \Gamma_s)}\, dt$$

attains a maximum for $s = 0$. Therefore, $\tau_\Gamma'(0) = 0$. We compute

$$\begin{aligned}
\tau_\Gamma'(s) &= \int_0^{t(p)} \frac{-1}{\sqrt{-g(\partial_t \Gamma_s, \partial_t \Gamma_s)}} g(D_s(\partial_t \Gamma_s), \partial_t \Gamma_s)\, dt \\
&= \int_0^{t(p)} \frac{-1}{\sqrt{-g(\partial_t \Gamma_s, \partial_t \Gamma_s)}} g(D_t(\partial_s \Gamma_s), \partial_t \Gamma_s)\, dt \\
&= \int_0^{t(p)} \frac{-1}{\sqrt{-g(\partial_t \Gamma_s, \partial_t \Gamma_s)}} \left(\frac{d}{dt} g(\partial_s \Gamma_s, \partial_t \Gamma_s) - g(\partial_s \Gamma_s, D_t(\partial_t \Gamma_s)) \right) dt.
\end{aligned}$$

By integrating by parts and setting $s = 0$ (since $\Gamma(\cdot, t(p)) = p$ =constant) we get $g(V_0, \dot{\delta}(0)) = 0$ where V is the variation field of the smooth variation Γ. It follows that V_0 is spacelike (since δ is timelike) and so $\dot{\delta}(0)$ is orthogonal to Σ since the curve $\Gamma(s, \cdot)$ was a random path on Σ. □

Theorem 3.22 (Hawking)**.** *Let (M, g) be a globally hyperbolic spacetime with a spacelike Cauchy hypersurface Σ satisfying the strong energy condition. Assume that the expansion of the mean curvature satisfies $\theta \leq \theta_0 < 0$ on Σ. Then, (M, g) is singular.*

Proof. We follow (GN, p. 307). We will prove that no future-directed timelike geodesic extends (to the future) beyond the time $-3/\theta_0$. If this was not the case then there would exist a timelike geodesic with a proper time parametrization such that it could be defined for $\tau_0 = -3/\theta_0 + \varepsilon$ for some $\varepsilon > 0$. Set $p := \gamma(\tau_0)$. Then, there exists an orthogonal to Σ timelike geodesic δ that reaches p and maximizes proper time among all timelike curves from Σ to p (lemma 3.21). Therefore, $\tau(\delta) \geq -3/\theta_0 + \varepsilon$. Then, from lemma 3.18, δ must admit a conjugate point which contradicts lemma 3.19 since δ maximizes proper time. Contradiction. □

If the expansion θ satisfies $\theta \geq \theta_0 > 0$ then the result remains true but for past-directed timelike geodesics (see (GN, p. 307)).

References

[KA] K. Athanasopoulos, "An Introduction to Riemannian Geometry", Course Notes, University of Crete, 2022.

[BS] A. N. Bernal, M. Sanchez, "On smooth Cauchy hypersurfaces and Geroch's splitting theorem", arXiv:gr-qc/0306108v2, 2003.

[AC] A. Chatzikaleas, "Hawking's Singularity Theorem", Lecture Notes, 15th Panhellenic Geometry Conference, University of Ioannina, 2022.

[GN] L. Godinho, J. Natario, "Introduction to Riemannian Geometry and applications to Relativity and Mechanincs", Springer, Universitext, 2014.

[SH] S. W. Hawking, The occurrence of singularities in cosmology. iii. causality and singularities, Proc. Roy. Soc. London Ser. A 300 (1967), no. 1461, 187–201.

[HE] S. W. Hawking, G. Ellis, "The Large Scale Structure of Spacetime (Cambridge Monographs on Mathematical Physics)", Cambridge University Press, 1975.

[HP] S. W. Hawking, R. Penrose, "The Nature of Space and Time", Princeton University Press, 1996.

[JL] J. M. Lee, "Introduction to Smooth Manifolds", Springer, Graduate Texts in Mathematics, 2012.

[EM] E. Minguzzi, "Lorentzian causality theory", Living Rev Relativ 22, 3, 2019.

[SM] E. Minguzzi, M. Sanchez, "The causal hierarchy of spacetimes", arXiv:gr-qc/0609119v3, 2006.

[JN] J. Natario, "Mathematical Relativity", arXiv:2003.02855v1 [gr-qc] 5 Mar 2020.

[BO] B. O' Neil, "Semi-Riemannian Geometry, with Applications to Relativity", Volume 103 (Pure and Applied Mathematics), Academic Press, 1983.

[RP] R. Penrose, "Techniques of Differential Topology in Relativity", CBMS-NSF Regional Conference Series in Applied Mathematics, 1972.

[HR] H. Ringstrom, "The Cauchy Problem in General Relativity", EMS, ESI Lectures in Mathematics and Physics, 2009.

[LT] L. W. Tu, "Introduction to Manifolds", Springer, Universitext, 2011.

[RW] R. M. Wald, "General Relativity", University Of Chicago Press, First Edition, 1984.