# The Hawking Singularity Theorem
PREPARATIONS

Hauke Vogt

## 1 What Is a Singularity?

When we investigated the Schwarzschild solution of the Einstein equation, we found that at $r = 0$ the metric is ill-defined. An analogous result appeared in the case of the Robertson Walker solution at $t = 0$. A further computation would have revealed unboundedness of the Kretschmann curvature scalar [Wal09, ch.6.4]. We considered these to be *singularities*. But there was a flaw in the reasoning:

> Are singularities artefacts of the high degree of symmetry of the solutions or true features of the spacetime?

This question is addressed by the *singularity theorems*. These provide criteria for the (non-)existence of singularities in a given spacetime. But they do not contain any information about the properties of these singularities (e.g. their location).

The definition of a singular spacetime is a bit tricky.

*Intuition* 1.1. One could be tempted to think of a singularity as a „place“ in spacetime where the curvature „blows-up“ or the metric is ill-defined in any other sense. An unbounded curvature would correspond to arbitrarily large gravity tearing every object apart. But there are two obstructions:

1. The word „place“ is inappropriate because in GR the spacetime is not *a priori* given but the object one is trying to solve for from the matter content of the theory. If one tries to define the spacetime as a manifold up to these problematic points, a notion of a manifold with singular boundary is needed, which has not been satisfactorily developed until today.

2. The „blow-up“ of the curvature should be understood as a blow-up of some scalar formed out of the curvature tensors to serve as a diffeomorphism invariant quantity. But if unboundedness occurs e.g. as an observer moves to infinity, this should not be considered singular. Or the curvature tensor itself might be singular although all scalar combinations identically vanish. Furthermore a spacetime may be singular although all curvature scalars are finite.

Fortunately there is a way out. If one would remove the singularities from the spacetime, the „holes“ they leave behind should be detectible by geodesics which are endless in one direction but cannot be continued past a specific point. The motion of an observer passing this point would not be predictable anymore.

**Definition 1.2.** A geodesic $c$ is called *incomplete* if it is inextendible in at least one direction but has finite range of affine parameter. A spacetime $(M, g)$ is called *singular* if it is geodesically incomplete, i.e. it possesses at least one incomplete geodesic.

*Remark* 1.3.

1. For a Riemannian manifold $(M, g)$ the Hopf-Rinow theorem [O'N10, Thm.5.21.] established an equivalence of metric and geodesic completeness. Since the proof relies on the fact that the positive-definite metric induces a distance function this fails in the Lorentzian case. In particular a compact spacetime may be geodesically incomplete and the exponential map may fail to be surjective in a geodesically complete spacetime.

2. Since a spacetime can artificially be made singular by removing points by hand one restricts to *inextendible* or *maximal* [O'N10, Def.5.44.] spacetimes, i.e. spacetimes which are not isometric to a proper open submanifold of another spacetime.

3. The presence of a hole in a spacetime might affect only a certain class of geodesics. Since physically reasonable motions are given by timelike and null curves we restrict to timelike and null geodesic incompleteness.

**Example 1.4.** 1. The Clifton-Pohl torus is compact but geodesically incomplete. Let $(\bar{M}, \bar{g}) = (\mathbb{R}^2 \setminus \{0\}, \frac{2\mathrm{d}u \otimes \mathrm{d}v}{u^2+v^2})$. $(\mathbb{Z}, +)$ as a Lie group acts freely and properly on $\bar{M}$ by $n.(u,v) = (2^n u, 2^n v)$ and therefore $g$ is a Lorentzian metric on $M = \bar{M}/\mathbb{Z} \cong T^2$. The Christoffel symbols are easily computed and the geodesic equations turn out to be $\ddot{u} - \frac{2u}{u^2+v^2}\dot{u}^2 = 0$ and analogously for $v$. For $v = 0$ one finds the solution $c(t) = \left(\frac{1}{at+b}, 0\right)$ which obviously has finite affine parameter although it is inextendible with image $\{(u,0) \,|\, u > 0\}$.

2. 2-dimensional Anti-de Sitter universe is geodesically complete but $\exp_p$ is nowhere surjective.

The logic in the proofs of the singularity theorems is the following:

1. Timelike curves of maximal proper time are – up to reparametrisation – timelike geodesics [O'N10, Cor.5.19]. But a timelike geodesic may not be maximising. The absence of *conjugate points* is a necessary and sufficient condition for a geodesic to be a maximum of proper time.
2. Using the Einstein equations on an inequality for the Ricci tensor one obtains so called *energy conditions*, providing a criterion for the existence of conjugate points.
3. In globally hyperbolic spacetimes it can be shown, that timelike curves(!) of maximal proper time do always exist by using compactness of the space of causal curves.
4. These cannot be geodesics yielding geodesic incompleteness of the spacetime.

The last two points will be worked out in the next talk.

# 2 Conjugate Points

Next we will recall a few basic concepts from differential geometry.

**Theorem 2.1.** *Let $c$ be a geodesic and $J$ a vector field along $c$. $J$ is a variation vector field along $c$ through geodesic variations iff it satisfies the* Jacobi equation *or* geodesic deviation equation

$$\nabla_{\dot{c}}^2 J = R(\dot{c}, J)\dot{c}, \tag{1}$$

*where $R$ denotes the Riemannian curvature tensor. $J$ is called a* Jacobi field.

*Proof.* see [O'N10, Lem.8.3.] □

*Intuition* 2.2. So one may think of a Jacobi field as the vector field modelling the variation of one geodesic into a neighbouring one, i.e. it creates a one-parameter family of geodesics. The equation provides a characterisation of curvature: Curvature accelerates geodesics toward or away from each other. Therefore the rhs is sometimes referred to as *tidal force operator* on $\dot{c}^\perp$ [O'N10, Def.8.8.].

**Poposition 2.3.** *Let $c : [a,b] \to M$ be a geodesic. Then $q = c(b)$ is called* conjugate *to $p = c(a)$ along $c$ if one of the following equivalent conditions is satisfied:*

1. *$q$ is a critical value of $\exp_p : T_pM \to M$, i.e. $\mathrm{d}\exp_p |_{b\dot{c}(a)}$ does not have full rank.*
2. *There exists a not identically vanishing Jacobi field $J$ along $c$ vanishing in $p$ and $q$.*

*Proof.* see [O'N10, Prop.10.10.] □

Especially the first characterisation will be of great importance since it shows that any normal coordinate system breaks down at conjugate points.

*Intuition* 2.4. Roughly speaking a conjugate point is a point, where two neighbouring geodesics (almost) meet. Some references (e.g. [O'N10, p.271]) add the word „almost“ because geodesics do not need to intersect but only do so „to first order“ while higher order variations may not vanish.

*Remark* 2.5. 1. The notion of conjugate points encodes global information about the spacetime since curvature determines conjugate points via the Jacobi equation and conjugate points affect the global structure, e.g. via the exponential map.

2. A sufficiently short geodesic does not admit conjugate points.

The notion of conjugate points can be readily generalised replacing one point by a submanifold, which we will take to be a spacelike hypersurface. To that aim we define a new exponential map.

**Definition 2.6.** Let $(M, g)$ be a spacetime, $S \subset M$ a spacelike hypersurface, $p \in S$ and $c_{n_p} : [a, b] \to M$ a timelike geodesic with future-pointing unit-length initial velocity vector $n_p \in N_pS$ normal to $S$ in $p$. We can define the *normal exponential map* by

$$\begin{aligned} \exp_p : N_pS \times [a, b] &\to M \\ (n_p, t) &\mapsto c_{n_p}(t). \end{aligned}$$

This induces a map $\exp : NS \to M$ on the normal bundle $NS$.

**Definition 2.7.** With the notions from above critical values $q$ of $\exp : NS \to M$ are *conjugate* or *focal* to $S$ , i.e. $\mathrm{d}\exp_p|_{b\dot{c}_{n_p}(a)}$ does not have full rank for some $p$.

*Remark* 2.8.

1. Alternatively one could transfer the notion of a Jacobi field to this situation and give a description analogous to 2.3 [O'N10, Prop.10.28., Def.10.29., Prop.10.30.].
2. The intuition behind this notion is obviously analogous to the one before.
3. Varying over the one-parameter family of geodesics orthogonal to $S$ we obtain a set of conjugate points forming a so called *caustic* [Pen72, p.58].

Equipped with this exponential map we can construct a coordinate system in which the distance between a hypersurface and a point is measured by the proper time of a geodesic [Pen72, Def.7.13].

*Construction* 2.9. Let $q = \exp_p(t_0) \in M$ be a point not(!) conjugate to $S \subset M$ and denote the local coordinates around $p \in S$ by $(x^1, x^2, x^3)$. For a sufficiently small neighbourhood $V_q \subset M$ of $q$ we label each point $r$ by the proper time of the geodesic it lies on and the coordinates of $q$. Hence $\phi : V_q \times \mathbb{R} \to M$ defined by $r \mapsto (t, x^1, x^2, x^3)$ defines a chart.

**Definition 2.10.** A coordinate obtained from the above construction is called *Gaussian normal* or *synchronous* coordinate system.

The reason for this terminology is the following

**Lemma 2.11.** *In synchronous coordinates geodesics remain orthogonal to all hypersurfaces $S_t$ defined by $t = const.$. As a consequence the metric tensor is of the form*

$$g = -\mathrm{d}t \otimes \mathrm{d}t + \gamma_{ij} \mathrm{d}x^i \otimes \mathrm{d}x^j, \tag{2}$$

$\gamma_{ij} := \langle X_i, X_j \rangle$ *being the positive definite spatial metric.*

*Proof.* The claim is true by construction for $S_0 = S$. To show that this remains true on $V_q$, we have to show that the geodesic tangent field $n$ remains orthogonal to the coordinate basis fields $X_1, X_2, X_3$ generating the tangent space to $S_t$. For any of these fields we have

$$\nabla_n \langle n, X \rangle = \langle n, \nabla_n X \rangle = \langle n, \nabla_X n \rangle = \frac{1}{2} \nabla_X \langle n, n \rangle = 0,$$

where the first step uses metricity and the second symmetry of the Levi-Civita connection together with the vanishing of the Lie bracket of the coordinate vector fields $X$ and $n$. □

As a next step we want to derive a necessary and sufficient condition for the existence of conjugate points. Therefore we need to have a closer look at the behaviour of the geodesics. Since in a conjugate point neighbouring geodesics almost intersect and the Jacobi equation (1) connects this question to curvature, it is natural to compute the Riemann and Ricci curvature tensor. A short computation yields

$$\Gamma^0_{00} = \Gamma^0_{0i} = 0 \quad \text{and} \quad \Gamma^i_{0j} = \gamma^{ik} \frac{1}{2} \partial_t \gamma_{jk} =: \gamma^{ik} \beta_{kj}$$

($(\gamma_{ij})^{-1} = \gamma^{ij}$ as shorthand) and

$$R_{00} = R_{i00}{}^i = \partial_i \Gamma^i_{00} - \partial_t \Gamma^i_{i0} + \Gamma^j_{00} \Gamma^i_{ij} - \Gamma^j_{i0} \Gamma^i_{0j} = -\partial_t (\gamma^{ij} \beta_{ij}) - \gamma^{jk} \gamma^{il} \beta_{ki} \beta_{lj}.$$

The quantity $\gamma^{ij} \beta_{ij}$ has a special interpretation in synchronous coordinates, namely

$$\gamma^{ij} \beta_{ij} = \frac{1}{2} \operatorname{tr} \left( (\gamma_{ij})^{-1} \partial_t (\gamma_{ij}) \right) = \frac{1}{2} \partial_t \log \det(\gamma_{ij}) = \partial_t \log \sqrt{\det(\gamma_{ij})}.$$

where we used $(\log(\det(A)))' = \operatorname{tr}(A^{-1} A')$ . This justifies the following terminology.

**Definition 2.12.** $\theta := \gamma^{ij}\beta_{ij} =$ is the *expansion* of a family of geodesics.

*Remark* 2.13.

1. Physically speaking it is the (logarithmic) rate of change of a volume – e.g. seen from the divergence theorem – measured by a family of synchronised free falling observers. Mathematically it is the trace of the second fundamental form of constant time hypersurfaces.
2. In general the curvature dependent behaviour of a family of timelike geodesics is governed by the*Raychaudhuri equation* [Wal09, ch.9.2 and 9.3]

$$\dot{\theta} = -\frac{1}{3}\theta^2 - \mathrm{tr}(\sigma^2) + \mathrm{tr}(\omega^2) - \mathrm{Ric}(\dot{c},\dot{c}) \tag{3}$$

with quantities representing trace, antisymmetric and symmetric trace-free part of a tensor field $B$ measuring the failure of a Jacobi field $J$ along a geodesic $c$ to be parallelly transported, i.e. acting by $J \mapsto \nabla_{\dot{c}} J$

$$\begin{aligned} \theta &:= \mathrm{tr}(B) && (\textit{expansion}) \\ \omega &:= \frac{1}{2}(B - B^*) && (\textit{twist}) \\ \sigma &:= \frac{1}{2}(B + B^*) - \frac{1}{3}\theta\gamma && (\textit{shear}). \end{aligned}$$

Since a singularity in $\theta$ indicates a conjugate point, we investigate the positivity of the Raychaudhuri equation. By virtue of Frobenius' theorem [War83, Thm.2.32] it can be shown that in synchronous coordinates, i.e. for a hypersurface-orthogonal family of geodesics, $\mathrm{tr}(\omega^2)$ vanishes [Wal09, Thm.B.3.2]. Furthermore $-\mathrm{tr}(\sigma^2)$ is manifestly negative. The last term is the only place where additional input is needed.

**Definition 2.14.** A spacetime $(M, g)$ is said to satisfy the *strong energy* or *timelike convergence* condition if $\mathrm{Ric}(X, X) \geq 0$ holds for every timelike vector field $X \in TM$.

The reason for the first name is

*Intuition* 2.15. Using the Einstein equation $\mathrm{Ric} = 8\pi T$ we find that

$$\mathrm{Ric}(X,X) = 8\pi T(X,X) = 8\pi\left[E(X,X) - \frac{1}{2}\mathrm{tr}(E)\langle X,X\rangle\right] = 8\pi\left[E(X,X) + \frac{1}{2}\mathrm{tr}(E)\right]$$

for all $X$ timelike unit vector fields. Hence $E(X,X) \geq -\frac{1}{2}\mathrm{tr}(E)$. This is can be transcribed as „gravity is attractive on average“, since the Ricci tensor is a trace of the Riemann tensor.

**Example 2.16.** For a perfect pressureless fluid with unit velocity vector field $U \in TM$ the strong energy condition implies

$$T(X,X) = \rho\left(\nu \otimes \nu + \frac{1}{2}g\right)(X,X) = \rho\left(\langle U,X\rangle^2 + \frac{1}{2}\langle X,X\rangle\right) \geq 0.$$

The term in parentheses is positive. This can be seen from orthogonally decomposing $X = aU + V$, hence having spacelike $V$ (so $\langle V,V\rangle \geq 0$). Thus

$$\begin{aligned} \langle X,X\rangle &= \langle aU+V, aU+V\rangle = a^2\langle U,U\rangle + \langle V,V\rangle = -a^2 + \langle V,V\rangle \\ \langle U,X\rangle^2 &= \langle U, aU+V\rangle^2 = a^2. \end{aligned}$$

The sum of these is positive and strictly larger than the term in parentheses and therefore $\rho \geq 0$.

Returning to (3) we know now that the last term is negative if the Einstein equation and the strong energy condition hold. We can now condition for the existence of conjugate points.

**Lemma 2.17.** *Let $(M, g)$ be a spacetime satisfying the strong energy condition, $S \subset M$ a spacelike hypersurface and $p \in S$ a point where $\theta = \theta_0 < 0$. Then $c_{n_p}$ contains a point $q \in M$ conjugate to $S$ at a distance of at most $-\frac{3}{\theta_0}$, provided it can be extended that far.*

*Proof.* From (3) and our assumptions we obtain

$$\partial_t\theta + \frac{1}{3}\theta^2 \leq 0 \iff \partial_t\theta^{-1} \geq \frac{1}{3}.$$

Integrating that inequality then yields

$$\frac{1}{\theta} \geq \frac{1}{\theta_0} + \frac{t}{3}$$

and hence $\frac{1}{\theta}$ possesses a zero and $\theta \to -\infty$ at a value no greater than $-\frac{3}{\theta_0}$. □

*Remark* 2.18. This is not a spacetime singularity but merely one of the synchronous coordinates!

# 3 Non-Maximising Geodesics

As advertised in the beginning conjugate points indicate whether a geodesic is maximising proper time or not. This is made precise by the following

**Poposition 3.1.** *Let $(M, g)$ be a spacetime, $S \subset M$ a spacelike hypersurface, $p \in M$ and $c : [a, b] \to M$ a smooth timelike geodesic orthogonal to $S$ connecting $p$ to $S$. Then $c$ does not maximises proper time between $S$ and $p$ over a one-parameter family of timelike geodesics if there is a point conjugate to $S$ between $S$ and $p$.*

*Heuristic Proof.* Let $q$ be the first conjugate point between $S$ and $p$ along $c$. Before reaching $q$ we may use synchronous coordinates. Since $q$ is conjugate to $S$ we can find a geodesic $\tilde{c}$ starting orthogonally to $S$ from a different point and meeting $c$ at $q$. Let $V_q$ be a geodesically convex neighbourhood of $q$ and choose two points $r, s \in V_q$ on $\tilde{c}$ between $S$ and $q$ and $q$ and $p$ respectively. Inside $V_q$ we can connect $r$ and $s$ by a unique timelike geodesic segment $\bar{c}$ and the piecewise smooth timelike curve obtained by $\tilde{c}|_{[S,r]} \wedge \bar{c}|_{[r,s]} \wedge c|_{[s,p]}$ connects $S$ and $p$ as well but has strictly larger proper time by the generalised triangle inequality (twin paradox).
„Rounding of the corners“ one obtains a smooth curve of larger proper time [HE08, Prop.4.5.8.]. □

*Formal Proof.* This time we show, that there exists a normal vector field $X$ along $c$ such that $I_c(X, X) > 0$. Let again $q = c(t_1)$ be the first conjugate point between $S$ and $p$ along $c$. This gives rise to a non-trivial normal Jacobi field $J$ along $c|_{[a,t_1]}$ vanishing in $t = a$ and $t = t_1$. This can trivially be extended to

$$V(t) = \begin{cases} J(t) & \text{if } t \in [a, t_1] \text{ and} \\ 0 & \text{if } t \in [t_1, b]. \end{cases}$$

Define a vector field $W$ along $c$ such that $W(b) = \Delta\nabla_{\dot{c}}V|_{t=b}$. The jump cannot be zero since $\Delta\nabla_{\dot{c}}V|_{t=b} = -\nabla_{\dot{c}}J|_{t=b} = 0$ and $J(b) = 0$ would imply that the Jacobi field is trivial.
For small $\delta > 0$ small we have

$$\begin{aligned} -I_c(V + \delta W, V + \delta W) &= -I_c(V, V) - 2\delta I_c(V, W) - \delta^2 I_c(W, W) \\ &= -2\delta\langle W(b), W(b)\rangle - \delta^2 I_c(W, W) < 0 \end{aligned}$$

for sufficiently small $\delta > 0$. The first term vanished because $V$ is a Jacobi field on both segments and is zero at the break. The second term is seen by plugging into (4). □

# 4 Towards a Topology on $C(S, p)$

In order to motivate the next step, we should advertise what will be coming: Proving the existence of maximising timelike curves involves compactness of the space of continuous causal curves $C(S, p)$. In order to topologise this space compactness of $D^+(S) \cap J^-(p)$ will be needed. Therefore we are going to prove

**Poposition 4.1.** *Let $(M, g)$ be a globally hyperbolic spacetime, $S \subset M$ a Cauchy hypersurface and $p \in D^+(S)$. Then $D^+(S) \cap J^-(p)$ is compact.*

*Intuition* 4.2. In words: The space of all points, which can be influenced by events in $S$ and have influenced the event $p$, is compact.

*Remark* 4.3. Of course $D^-(S) \cap J^+(p)$ is compact, too.

**Definition 4.4.** Let $(M, g)$ be a spacetime and $U \subset M$ a geodesically convex open subset diffeomorphic to some open ball whose boundary forms a compact submanifold of a larger geodesically convex open set. Then $U$ is said to be a *simple neighbourhood.*

This means $\partial U$ is diffeomorphic to $S^3$ and $\bar{U}$ is compact.

**Lemma 4.5.** *Simple neighbourhoods form a basis of the topology of $M$.*

*Proof.* Simple neighbourhoods are diffeomorphic to open balls. □

**Lemma 4.6.** *Any open cover has a countable locally finite refinement by simple neighbourhoods.*

*Proof.* Since simple neighbourhoods form a basis of the topology of $M$, every open cover $\{V_\alpha\}_{\alpha\in A}$ admits a refinement $\{U_\beta\}_{\beta\in B}$ by simple neighbourhoods, not *a priori* countable or locally finite. For $V := \bigcup_{\alpha\in A} V_\alpha$ compact there is a finite subcover obviously countable and locally finite.
For $V$ not compact we may compactly exhaust $V$, i.e. choose a sequence $\{K_i\}_{i\in\mathbb{N}}$ of compact subsets of $V$ with $K_i \subset K_{i+1}$ and $V = \bigcup_{i=1}^{\infty} K_i$. Since $K_1$ is compact and covered by $\{U_\beta\}_{\beta\in B}$ we may choose a finite subcover $\{U_{\beta_1},\ldots,U_{\beta_k}\}$. Inductively we find a finite collection of neighbourhoods $\{U_{\beta_1^i},\ldots,U_{\beta_{k_i}^i}\}$ covering the compact set $K_i \setminus \operatorname{int} K_{i-1}$. Taking appropriately small simple neighbourhoods we can adjust $\bigcup_{j=1}^{k_i} U_{\beta_i^j} \subset \operatorname{int} K_{i+1} \setminus K_{i-2}$. Hence the countable cover $\{U_{\beta_j^i}\}_{i\in\mathbb{N},\,1\le j\le k_i}$ is locally finite. □

*Proof of the proposition.* Suppose $A := D^+(S) \cap J^-(p)$ were not compact. Then there would exist a countable and locally finite open cover by simple neighbourhoods $\{U_n\}_{n\in\mathbb{N}}$ of $A$ not admitting a finite subcover. Let $\{q_n\}_{n\in\mathbb{N}}$ be a sequence with $q_n \in A \cap U_n$ with $q_m \neq q_n$ if $m \neq n$. This cannot have accumulation points since $M$ has neighbourhoods intersecting only finitely many $U_n$. Especially $U_n$ cannot contain infinitely many $q_n$.
Set $p_1 = p$. $p_1 \in A$ and hence $p_1 \in U_{n_1}$ for some $n_1 \in \mathbb{N}$. Choose $q_n \notin U_{n_1}$. Having $q_n \in J^-(p_1)$ it can be connected to $p_1$ by a future-directed causal curve $c_n$. $c_n$ intersects $\partial U_{n_1}$ in, say, $r_{1,n}$. There are infinitely many of these intersection points since $U_{n_1}$ contains only finitely points from $\{q_n\}_{n\in\mathbb{N}}$. By compactness of $\partial U_{n_1}$ the sequence $\{r_{1,n}\}_{n\in\mathbb{N}}$ will accumulate at, say, $p_2 \in \partial U_{n_1}$.
$p_2 \in J^-(p_1)$: Because $\bar{U}_{n_1}$ is contained in a geodesically convex open set, we can find a sequence $\{\gamma_{1,n}\}_{n\in\mathbb{N}}$ of unique causal geodesics connecting $p_1$ to $r_{1,n}$. Choosing a subsequence of $\{r_{1,n}\}_{n\in\mathbb{N}}$ converging to $p_2$ we get a subsequence of $\{\gamma_{1,n}\}_{n\in\mathbb{N}}$ converging to $\gamma_1$ connecting $p_1$ and $p_2$.
$p_2 \in D^+(S)$: If $S = t^{-1}(0)$ is the „initial surface“, then $t(r_{1,n}) \geq 0 \Rightarrow t(p_2) \geq 0$ hence $p_2 \in D^+(S)$ and consequently in $A$.
Since $p_2 \notin U_{n_1}$ it lies in some different $U_{n_2}$. This again only contains a finite number of points in the sequence $\{q_n\}_{n\in\mathbb{N}}$ and so again an infinite number of curves $c_n$ intersects $\partial U_{n_2}$ past to $r_{1,n}$, which we call $r_{2,n}$. By compactness of $\partial U_{n_2}$ the sequence $\{r_{2,n}\}_{n\in\mathbb{N}}$ converges to some point $p_3 \in \partial U_{n_2}$.
By the same arguments as above one can now connect $r_{1,n}$ and $r_{2,n}$ by causal geodesics converging to a causal geodesic $\gamma_2$ connecting $p_2$ and $p_3$. And because $J^-(p_2) \subset J^-(p_1)$ $p_3$ also lies in $J^-(p)$. So we find $p_3 \in A$.
Iterating this procedure one finds a sequence of causal geodesic segments $\gamma_i$ connecting points $p_i \in U_{n_i}$ and $p_{i+1} \in U_{n_{i+1}}$ of a sequence $\{p_i\}_{i\in\mathbb{N}}$. None of these segments can intersect $S$. And therefore the past-directed causal curve $\gamma$ starting at $p_1$ obtained by gluing the segments together smoothly cannot intersect $S$, either. But this curve is past-inextendible because the sequence $\{p_i\}_{i\in\mathbb{N}}$ cannot accumulate. But since $p_1 \in D^+(S)$ by assumption, $\gamma$ needs to intersect $S$. Contradiction! □

**Corallary 4.7.** *$J^+(p)$ is closed and $J^+(p) \cap J^-(p)$ is compact.*

*Proof.* Let $\{q_n\}_{n\in\mathbb{N}}$ be a sequence in $J^+(p)$ converging to $q \in M$. We need to show that $q \in J^+(p)$. Take $S$ a Cauchy surface with $t(S) > t(q)$ and $q \in D^-(S)$. Then for $n \in \mathbb{N}$ large enough we have $q_n \in D^-(S) \cap J^+(p)$. Since this is compact we get $q \in D^-(S) \cap J^+(p) \subset J^+(p)$.
By the same argument $J^-(q)$ is closed and so $J^+(p) \cap J^-(q)$ is closed. We want to show that it is a subset of the compact set $D^-(S) \cap J^+(p)$ for some Cauchy surface $S$. This time taking a Cauchy surface $S$ such that $t(S) = t(q)$. □

# A The Second Variation on Geodesics

The more formal proof of Prop. 3.1 uses the *second variation formula* for the *proper time functional.* Therefore we collect a few concepts from variational calculus here. Write

$$\Gamma : [a,b] \times (-\delta,\delta) \to M$$
$$(t,s) \mapsto c_s(t)$$

for a one-parameter family of piecewise smooth timelike curves with variation and affine parameters $s$ and $t$.

**Definition A.1.** Take $(M,g)$ to be a spacetime. The *proper time* of a piecewise smooth timelike curve (segment) $c : [a,b] \to M$ is given by

$$L(c) := \int_a^b \sqrt{|\langle \dot{c}, \dot{c}\rangle|}\,\mathrm{d}t.$$

For a one-parameter family this yields a real-valued function

$$\begin{aligned} L_\Gamma : (-\delta, \delta) &\to \mathbb{R} \\ s &\mapsto L(c_s) \end{aligned}$$

satisfying $L_\Gamma(0) = L(c_0)$. Hence $L$ is a functional. Furthermore first and second variation are denoted by $L'(0) := \partial_s L_{\Gamma(t,s)}|_{s=0}$ and $L''(0) := \partial_s^2 L_{\Gamma(t,s)}|_{s=0}$.
Since the first variation is less important for us, we only state a simple corollary of the *first variation formula* that one can probably already guess:

**Corallary A.2.** *An affinely parametrised piecewise smooth curve $c : [a, b] \to M$ is a geodesic iff $L'' = 0$ for every fixed endpoint one-parameter family with $c_0 = c$.*

*Proof.* see [O'N10, Cor.10.3.] □

**Theorem A.3.** *Let $c = c_0 : [a, b] \to M$ be a unit speed smooth timelike geodesic (segment) on a spacetime $(M, g)$, $\Gamma$ its one-parameter family with fixed endpoints and $V(t) = \Gamma(t, 0)$ the variation vector field, then*

$$L''(0) = -\int_a^b \langle \nabla_{\dot{c}} V, \nabla_{\dot{c}} V \rangle - \langle R(V, \dot{c})\dot{c}, V \rangle \mathrm{d}t.$$

*Proof.* see [O'N10, Thm.10.4.] □

*Remark* A.4.

1. $V$ is interpreted as the velocity vector field of $\Gamma$ when crossing $c_0$.

2. This formula only depends on $V$ but not on the choice of the one-parameter family.

3. Actually the formula depends only on the component $(\nabla_{\dot{c}} V)^\perp$ orthogonal to $c_0$: Decompose $V = V^\parallel + V^\perp := \langle V, \dot{c} \rangle \dot{c} + V - V^\parallel$ and note $\nabla_{\dot{c}} \dot{c} = 0$. Then

$$\nabla_{\dot{c}} V^\parallel = \langle \nabla_{\dot{c}} V, \dot{c} \rangle \dot{c} = (\nabla_{\dot{c}} V)^\parallel \qquad \nabla_{\dot{c}} V^\perp = (\nabla_{\dot{c}} V)^\perp$$

   and hence

$$\langle \nabla_{\dot{c}} V, \nabla_{\dot{c}} V \rangle = \langle \nabla_{\dot{c}} V, \dot{c} \rangle^2 + \langle \nabla_{\dot{c}} V^\perp, \nabla_{\dot{c}} V^\perp \rangle.$$

   Furthermore the symmetry $\langle R(\dot{c}, \dot{c})\cdot, \cdot \rangle = \langle R(\cdot, \cdot)\dot{c}, \dot{c} \rangle = 0$ implies

$$\langle R(V, \dot{c})\dot{c}, V \rangle = \langle R(V^\perp, \dot{c})\dot{c}, V^\perp \rangle.$$

   Intuitively the tangential component contributes to a reparametrisation only.

To establish a result analogous to the „second derivative test“ used in analysis, we regard the space of all (piecewise) smooth curves $c : [a, b] \to M$ from $p \in M$ to $q \in M$ as a manifold $C(p, q)$, which is to be made more precise in the next talk. This manifold also has a tangent space $T_c(C)$ at $c$ represented by all piecewise smooth vector fields $V$ along $c$ vanishing in $p$ and $q$ [O'N10, Def.10.5., Def.10.6., Cor.10.7.]. In analogy to the Hessian of a function we introduce

**Poposition A.5.** *The* index form $I_c$ *of a unit speed timelike geodesic $c \in C(p, q)$ is the unique symmetric bilinear form*

$$\begin{aligned} I_c : T_c(C) \times T_c(C) &\to \mathbb{R} \\ (V, W) &\mapsto \int_a^b \langle \nabla_{\dot{c}}^2 V - R(V, \dot{c})\dot{c}, W \rangle \mathrm{d}t + \sum_{i=1}^k \langle \Delta \nabla_{\dot{c}} V|_{t_i}, W(t_i) \rangle, \end{aligned} \tag{4}$$

*where $c$ and $V$ have breaks at $t_1 < \cdots < t_k$ and $\Delta \nabla_{\dot{c}} V$ denotes the jumps in $\nabla_{\dot{c}} V$. In particular $I_c(V, V) = L''_\Gamma(0)$.*

*Proof.* see [O'N10, Cor.10.8.] □

*Remark* A.6. Note that $I_c(V, V)$ is the second variation formula integrated by parts. Hence the previous theorem shows that if a geodesic $c$ is maximising, then $I_c(V, V) \leq 0$ for any variation vector field $V$ along $c$.

# References


[HE08] S. W. Hawking and G. F. R. Ellis. *The large scale structure of space-time.* Cambridge Univ. Press, Cambridge [u.a.], 21. printing edition, 2008.

[O'N10] Barrett O'Neill. *Semi-Riemannian geometry : with applications to relativity.* Pure and applied mathematics : a series of monographs and textbooks. Academic Press, San Diego [u.a.], 2010.

[Pen72] Roger Penrose. *Techniques of differential topology in relativity.* Regional conference series in applied mathematics. Soc. for Industrial and Applied Mathematics, Philadelphia, Pa., 1972.

[Wal09] Robert M. Wald. *General relativity.* Univ. of Chicago Press, Chicago [u.a.], [repr.] edition, 2009.

[War83] Frank Wilson Warner. *Foundations of differentiable manifolds and Lie groups.* Graduate texts in mathematics : GTM. Springer, New York, NY [u.a.], 1983.