universität wien

MASTERARBEIT / MASTER THESIS

Titel der Masterarbeit / Title of the Master Thesis

„Geometric Foundations of the Classical Singularity Theorems"

verfasst von / submitted by

Argam Ohanyan, BSc BSc MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 876
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Physik UG2002
Betreut von / Supervisor:	ao. Univ.-Prof. Mag. Dr. Roland Steinbauer
Mitbetreut von / Co-Supervisor:	Dr. Stefan Palenta, BSc MSc

Contents

Preface **8**

Notation and Conventions **10**

1 Submanifolds of semi-Riemannian Manifolds **11**
1.1 General semi-Riemannian submanifolds 11
1.1.1 The normal bundle 12
1.1.2 The induced connection and the second fundamental form 14
1.1.3 The normal connection and the Codazzi equation . . . 19
1.1.4 Curves in semi-Riemannian submanifolds 21
1.1.5 Totally geodesic and totally umbilic submanifolds . . . 24
1.2 Semi-Riemannian hypersurfaces 26
1.3 Topological hypersurfaces in spacetimes 30

2 Variational Theory of Curves **33**
2.1 Jacobi fields and geodesic variations 33
2.2 Variational formulas 41
2.3 Timelike index theory 46
2.3.1 The timelike index form and conjugate points along timelike geodesics 47
2.3.2 The timelike Morse index theorem 53
2.3.3 Jacobi and Lagrange tensors along timelike geodesics . 59
2.4 Null index theory 66
2.4.1 Quotient constructions along null geodesics 66
2.4.2 Jacobi classes and conjugate points 67
2.4.3 The null Morse index theorem 80

3 Singularity Theorems **83**
3.1 Causality theory 83
3.1.1 Cauchy surfaces, Cauchy developments and horizons . 83
3.1.2 Causally disconnected spacetimes 94
3.1.3 The genericity condition 95
3.1.4 Focal points 104

3.1.5 Trapped sets and trapped surfaces 110
3.2 The classical singularity theorems 111
3.2.1 The Hawking singularity theorem 112
3.2.2 The Penrose singularity theorem 116
3.2.3 The Hawking–Penrose singularity theorem 117

Abstract **123**

Zusammenfassung **125**

Bibliography **127**

Preface

The discovery of the singularity theorems in the 20th century marks a breakthrough in our understanding of General Relativity (GR). They form the basis of our mathematical understanding of black holes and the Big Bang. Singularity theorems were initiated by Roger Penrose, with significant later contributions by Stephen Hawking, George Ellis and others. It was for his work on singularity theorems and black holes that Penrose received the Nobel Prize for Physics in 2020.

The term "singularity theorem" refers to a set of results in Lorentzian geometry that all roughly have the following form: If a spacetime (M, g) satisfies a set of physically reasonable conditions (a causality condition, an energy condition, and an initial/boundary condition), then there is an incomplete causal geodesic. This is physically interpreted as the sudden end of an observer or a light ray meant to signify a "singularity".

Singularity theorems are applicable in a variety of scenarios: The Hawking theorem is useful in cosmological spacetimes, where one mostly uses its past version to conclude the existence of a Big Bang. The Penrose theorem assumes the existence of trapped surfaces, which can arise due to a concentration of matter or radiation (but also in a vacuum). The most refined of the classical results, the Hawking–Penrose theorem, is applicable in a plethora of situations under rather weak assumptions on the causality.

This thesis deals with the geometric underpinnings of the classical singularity theorems. It is divided into three Chapters: In Chapter 1, we study various aspects of semi-Riemannian submanifolds while putting an emphasis on their intrinsic and extrinsic curvature. In Chapter 2, we undertake a thorough investigation of variations of curves, and develop all the tools needed to understand them. The question we are interested in the most is whether/when a given geodesic stops maximizing the Lorentzian distance. These results will then be used in Chapter 3, where we prove the singularity theorems of Hawking, Penrose, and Hawking-Penrose.

The prerequisites for reading this thesis are a solid understanding of smooth manifolds as well as basic Riemannian and Lorentzian geometry. These requirements are fully met by a thorough study of [18, Ch. 1-16], as well as [19, Ch. 1-7] and [22, Ch. 1-3, 5, 14.1-14.5]. The latter two will actually serve as sources for many results we treat in this thesis. Most standard results, in particular those from smooth manifold theory, will be used without reference.

Acknowledgements: I want to thank my advisor Roland Steinbauer for the opportunity to work on this thesis, and for all the courses and seminars I attended under his supervision. His welcoming attitude and the amount of work he puts into teaching and conveying a passion for mathematics in general are greatly appreciated. By extension, I thank everyone in the DIANA research group at the University of Vienna for much the same reasons.

I want to thank Stefan Palenta from the Department of Physics of the University of Vienna for his co-supervision, which made this thesis, whose contents are at the intersection between mathematics and physics, possible.

I also want to thank all of my colleagues and friends from the Faculty of Physics at the University of Vienna for their support throughout the years. The thought-provoking discussions I have with them regularly have heavily contributed to my understanding of many aspects of physics and mathematics.

Finally, I want to thank my mother, Anush Ohanyan, who has supported me in all of my endeavors and who has motivated me to always pursue my interests and give it my all.

Notation and Conventions

In this section, we fix the notation and the conventions that will be used throughout the thesis. For the most part, we will be in adherence to standard modern differential geometry in the spirit of [18], [19], [22] and [1]. Topological notions are in line with [17]. Only the more nonstandard conventions and/or those differing from the mentioned standard sources will be addressed.

A *manifold* will always be a second countable Hausdorff space that is locally Euclidean and that possesses a smooth (i.e. C^∞) structure. More often than not in this thesis (and generally in most global studies of Riemannian and Lorentzian manifolds), connectedness will play no role and it will be assumed even when not mentioned explicitly.
All maps (provided it is possible) will be assumed to be smooth (or piecewise smooth in the context of curves) unless otherwise specified. Given a smooth map $f : M \to N$ between manifolds, the tangent map at $p \in M$ will be denoted $T_p f$.
If $f : M \to N$ is any smooth map between manifolds, then $\mathfrak{X}(f)$ will denote the vector fields over f, i.e. smooth maps $M \to TN$ such that $X(p) \in T_{f(p)}N$ for all $p \in M$. The most important occurrence of this will be when f is a curve, i.e. a smooth map from some interval into a manifold.
Submanifold without further specifiers will always mean embedded submanifold (this will also be the case for more specialized notions, e.g. hypersurfaces). A subtler differentiation between immersed and embedded submanifolds will be explicitly made if necessary.
For a vector field $X \in \mathfrak{X}(M)$, Fl^X will denote its flow map, and $\mathrm{Fl}^X_t(p) := \mathrm{Fl}^X(t, p)$, where t is a parameter and $p \in M$.
Given a semi-Riemannian manifold (M, g), one has $\mathfrak{X}(M) \cong \Omega^1(M)$ (similarly for tensor spaces of higher valence $\mathcal{T}^k_l(M)$) via the g-musical isomorphisms. These will be denoted via $X \mapsto X^\flat$ and $\omega \mapsto \omega^\sharp$ for $X \in \mathfrak{X}(M)$, $\omega \in \Omega^1(M)$. Given a smooth function $f : M \to \mathbb{R}$, $\operatorname{grad} f \in \mathfrak{X}(M)$ will denote the vector field $(df)^\sharp$. If ∇ denotes the Levi-Civita connection on (M, g) (which it always will throughout the thesis), then $df = \nabla f$.
A *geodesic* in a semi-Riemannian manifold (M, g) is a smooth curve satisfying the geodesic equation $\nabla_{\gamma'}\gamma' = 0$. Given an initial point and an initial vector, such a solution exists locally and is unique. We say (M, g) is *geodesically complete* if all maximally extended solutions of the geodesic equation exist on all of $\mathbb{R}$.
For notions concerning semi-Riemannian submanifolds of a semi-Riemannian manifold, we will stick to the notation in [22]. On the other hand, we will adopt the conventions in [19] concerning curvature quantities, e.g. $R_{XY}Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$ for the Riemannian curvature tensor, which would be $-R_{XY}Z$ in the convention of [22].

Given a *two parameter map* $f : \mathbb{R}^2 \supseteq D \to M$ (i.e. a smooth map defined on an open set $D \subseteq \mathbb{R}^2$ such that vertical and horizontal lines intersect D in intervals) and a vector field $X(t,s)$ over f, we will sometimes use the notation $\nabla_s X(t,s)$ to denote the covariant s-derivative, similarly $\nabla_t X(t,s)$ for the covariant t-derivative.

A *Lorentzian metric* will always be of signature $(-,+,\ldots,+)$. A *spacetime* is a connected Lorentzian manifold that possesses a time orientation, i.e. a timelike vector field $X \in \mathfrak{X}(M)$. X determines forward and backward lightcones as follows: If $v \in T_pM$ is causal, then v is (by definition) future-directed if and only if $g(v, X_p) \leq 0$.

We will assume knowledge of basic causality theory (following [22]) throughout this work. Let us recall important notation: If $U \subset M$, we say $x \ll_U y$ if there is a future directed timelike curve from x to y contained in U. If $U = M$, we omit the index. Similarly, we say $x \leq_U y$ if either $x = y$ or there is a future directed causal curve from x to y entirely in U. We write $x < y$ if $x \neq y$ and $x \leq y$. By $I^+_U(x)$ we denote all points y such that $x \ll_U y$. Similarly, $J^+_U(x)$ are those points y such that $x \leq_U y$. As before, we omit the index if $U = M$. Past variants of these notions are denoted in an analogous manner.

A *convex set* (or a *convex neighborhood)* U in a semi-Riemannian manifold M is an open set that is a normal neighborhood of each of its points, i.e. given any point $p \in U$, it is diffeomorphic via $\exp_p$ to some star-shaped neighborhood of 0 in T_pM.

The notation $K \subset\subset M$ means that K is compact and contained in M.

Chapter 1

Submanifolds of semi-Riemannian Manifolds

In this chapter, we aim to understand the theory of submanifolds of semi-Riemannian manifolds. In the first (and longest) section, following the expositions in [22, Ch. 4], [19, Ch. 8] and [14, Sec. 4.4], we develop general results about semi-Riemannian submanifolds. In particular, we try to get an understanding of the various notions of curvature for such submanifolds, both intrinsic to the submanifold itself and extrinsic, i.e. with respect to the ambient space. Decomposing (via the Gauss formula) the ambient Levi-Civita connection applied to vector fields tangent to a semi-Riemannian submanifold, one can define the second fundamental form as its normal part, and use this important object to understand the tangential part (cf. Gauss equation) and the normal part (cf. Codazzi equation) of the ambient Riemannian curvature tensor. In the next couple of sections, several special classes of submanifolds are treated and results that will be needed later are collected (the references are the same except for the section on C^0-hypersurfaces, which is from [22, p. 413-415]). A treatment of Cauchy surfaces in spacetimes is postponed until the theory of conjugate and focal points has been developed in the next chapter.

1.1 General semi-Riemannian submanifolds

In this section, we develop some general notions for semi-Riemannian submanifolds of semi-Riemannian manifolds following [22, Ch. 4], [19, Ch. 8] and [14, Sec. 4.4]. Throughout this section, let $(\tilde{M}, \tilde{g})$ be a semi-Riemannian manifold of signature (k, l), with $k + l = n = \dim \tilde{M}$.

Definition 1.1.1. (Semi-Riemannian submanifold)
An (embedded) submanifold $M \subseteq \tilde{M}$ is called *semi-Riemannian submanifold* of $\tilde{M}$ if $(M, \iota^*\tilde{g})$ is a semi-Riemannian manifold, i.e. $\iota^*\tilde{g}$ is a semi-Riemannian metric on M, where $\iota : M \hookrightarrow \tilde{M}$ denotes the inclusion. If $\iota^* g$

is a Riemannian metric on M, then M is called a *Riemannian submanifold* of $\tilde{M}$.

Unless stated otherwise, the general dimension of M will always be denoted $\dim M = m$.

Remark 1.1.2. (On semi-Riemannian submanifolds)

(1) We mostly consider embedded submanifolds, as this will be the only case that appears in the context of singularity theorems. In some contexts, it may be important to look at immersed submanifolds (e.g. images of certain types of curves). Most of the results we discuss in this chapter will also hold for *immersed semi-Riemannian submanifolds*, but some will not (e.g. the results on hypersurfaces where we choose normals, which requires the hypersurface to be embedded).

(2) The condition in Definition 1.1.1 is void if $(\tilde{M}, \tilde{g})$ is Riemannian, because in that case, any embedded submanifold of $\tilde{M}$ is a Riemannian submanifold with the induced metric. In the general semi-Riemannian case, it can happen that $\iota^*\tilde{g}$ is degenerate (and is hence not a semi-Riemannian metric).

(3) If (M, g) is a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$ and (k', l') is its signature, then $k' \leq k$ and $l' \leq l$.

Example 1.1.3. (Trivial (non)examples)
Let $\tilde{M} = \mathbb{R}^2_1$, i.e. 2-dimensional Minkowski space with the Minkowski metric $\eta = -dt^2 + dx^2$. Then $\{t = 0\}$ is an example of a Riemannian submanifold of $\tilde{M}$. On the other hand, $\{t = x\}$ is a submanifold, but not a semi-Riemannian one, because the pullback of η under the inclusion is identically zero.

1.1.1 The normal bundle

In this subsection, we also follow [19, Ch. 1] in addition to the aforementioned references.
It turns out that one can learn a lot about the geometry of a semi-Riemannian submanifold by relating its tangent spaces and their complements in the tangent spaces of the ambient manifold.

Definition 1.1.4. (Normal bundle)
If $M \subseteq \tilde{M}$ is a semi-Riemannian submanifold, then for any $p \in M$, we have the orthogonal (with respect to g) direct sum decomposition

$$T_p\tilde{M} \cong T_pM \oplus N_pM,$$

with $N_pM := (T_pM)^\perp$. We call

$$NM := \bigsqcup_{p \in M} N_pM$$

the *normal bundle* of M.

Definition and Lemma 1.1.5. (Adapted orthonormal frames)
Let $M \subseteq \tilde{M}$ be a semi-Riemannian submanifold. Then for any $p \in M$ there exists an orthonormal frame $(E_1, \dots, E_n)$ defined in a neighborhood U of p in $\tilde{M}$ such that $E_1|_{U\cap M}, \dots, E_m|_{U\cap M}$ are tangent to M. We call such an orthonormal frame *adapted* to M.

Proof. By Gram-Schmidt, there always exist local orthonormal frames. So let $(E_1, \dots, E_m)$ be a local orthonormal frame around p in M, and consider the E_i extended arbitrarily to local vector fields in a neighborhood of p in $\tilde{M}$. Now complete the orthonormal vectors $(E_1|_p, \dots, E_m|_p)$ to a basis $(E_1|_p, \dots, E_m|_p, v_{m+1}, \dots, v_n)$ of $T_p\tilde{M}$, and extend the v_i arbitrarily to local vector fields V_i around p in $\tilde{M}$. Now we can use Gram-Schmidt on the local frame $(E_1, \dots, E_m, V_{m+1}, \dots, V_n)$ to get an orthonormal frame $(\tilde{E}_1, \dots, \tilde{E}_n)$ around p in $\tilde{M}$. Since the E_i were tangent to M, so are the $\tilde{E}_i$, $i = 1, \dots, m$. □

Proposition 1.1.6. (The normal bundle is a vector bundle)
Let $M \subseteq \tilde{M}$ be a semi-Riemannian submanifold of dimension m. Then the normal bundle NM is a smooth rank-$(n-m)$-subbundle of the ambient tangent bundle $T\tilde{M}|_M$, and we have the following Whitney sum decomposition:

$$T\tilde{M}|_M \cong TM \oplus NM.$$

Associated to this decomposition, we have the natural smooth projection maps

$$\begin{aligned} \tan : T\tilde{M}|_M &\to TM, \\ \mathrm{nor} : T\tilde{M}|_M &\to NM, \end{aligned}$$

called *tangential* and *normal projection*, respectively. They restrict to orthogonal projections on each fiber.

Proof. Let $(E_1, \dots, E_n)$ be a local orthonormal frame in $\tilde{M}$ adapted to M. Then the restrictions of $(E_{m+1}, \dots, E_n)$ to M locally span NM, which shows that NM is a subbundle of $T\tilde{M}|_M$ of rank $n - m$ (cf. [18, Lem. 10.32]). Decomposing the restriction of $(E_1, \dots, E_n)$ to M into $(E_1, \dots, E_m)$ and $(E_{m+1}, \dots, E_n)$ (also restricted to M) proves the claimed Whitney sum decomposition. Since tan and nor are just the maps that project to the first m and last $n - m$ components with respect to the frame $(E_1, \dots, E_n)$, respectively, they are smooth maps by [18, Lem. 10.29]. □

Definition and Corollary 1.1.7. (Tangential and normal vector fields)
Associated to the vector bundle decomposition

$$T\tilde{M}|_M \cong TM \oplus NM$$

there is a decomposition of the space of sections

$$\Gamma(T\tilde{M}|_M) \cong \Gamma(TM) \oplus \Gamma(NM)$$

as $C^\infty(M)$-modules, where $\Gamma(TM) = \mathfrak{X}(M)$. We call the elements of

$$\tilde{\mathfrak{X}}(M) := \Gamma(T\tilde{M}|_M)$$

the *$\tilde{M}$-vector fields on M*, they are vector fields over the inclusion $M \hookrightarrow \tilde{M}$. Elements of $\mathfrak{X}(M)^\perp := \Gamma(NM)$ are called the *vector fields normal to M*. For $X \in \tilde{\mathfrak{X}}(M)$, the decomposition amounts to

$$X = \tan X + \operatorname{nor} X \equiv X^\top + X^\perp.$$

Remark 1.1.8. (Reminder: linear connections on vector bundles)
It the following, we will encounter different connections on various vector bundles, so let us recall the general definition: If (N, h) is a semi-Riemannian manifold and $E \to N$ is a vector bundle over N, then we call a map $\nabla : \mathfrak{X}(N) \times \Gamma(E) \to \Gamma(E)$ a *(linear) connection* on E if ∇ is $C^\infty(N)$-linear in the first slot, $\mathbb{R}$-linear in the second and it satisfies the usual product rule $\nabla_X(f\omega) = X(f)\omega + f\nabla_X\omega$ for all $f \in C^\infty(N), X \in \mathfrak{X}(N)$ and $\omega \in \Gamma(E)$. Thus, affine connections on N (in particular, the Levi-Civita connection) are connections on the tangent bundle $TN \to N$.

1.1.2 The induced connection and the second fundamental form

By considering the Levi-Civita connection on the ambient space $\tilde{M}$ in directions tangent to the semi-Riemannian submanifold M, we gain information about the way M curves in $\tilde{M}$.

Definition 1.1.9. (Induced connection)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. If $\tilde{\nabla}$ denotes the Levi-Civita connection on $(\tilde{M}, \tilde{g})$, we denote by the same symbol the map

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

where the expression $\tilde{\nabla}_X Y$ is defined at $p \in M$ via locally extending X, Y around p to vector fields on $\tilde{M}$, applying the Levi-Civita connection, and then evaluating at p. We call this map the *induced connection on M*.

Lemma 1.1.10. (The induced connection is well-defined)
The induced connection $(X, Y) \mapsto \tilde{\nabla}_X Y$, $X \in \mathfrak{X}(M), Y \in \tilde{\mathfrak{X}}(M)$, is a well-defined map.

Proof. Clearly, $\tilde{\nabla}_X Y$ is smooth by construction. Hence $\tilde{\nabla}_X Y \in \tilde{\mathfrak{X}}(M)$ if we can show that it is well-defined.
Let $p \in M$ and suppose $\tilde{X}, \tilde{Y}$ are extensions of $X \in \mathfrak{X}(M)$, $Y \in \tilde{\mathfrak{X}}(M)$ to local vector fields on $\tilde{M}$ in an $\tilde{M}$-neighborhood U around p. We may assume that U is a coordinate neighborhood with coordinates (x^i). Write $\tilde{Y} = f^i \frac{\partial}{\partial x^i}$ on U. Then

$$\nabla_{\tilde{X}} \tilde{Y} = \tilde{X}(f^i) \frac{\partial}{\partial x^i} + f^i \nabla_{\tilde{X}} \frac{\partial}{\partial x^i}.$$

Now if $q \in M \cap U$, then

$$\begin{aligned} \tilde{X}(f^i)|_q &= X_q(f^i) = X_q(f^i|_M), \\ \tilde{\nabla}_{\tilde{X}} \frac{\partial}{\partial x^i}|_q &= \tilde{\nabla}_{X_q} \frac{\partial}{\partial x^i}. \end{aligned}$$

This shows that the value of $\tilde{\nabla}_{\tilde{X}} \tilde{Y}$ at points of M only depends on X and Y and is hence independent of the choice of local extension. □

The induced connection on M is a connection on the ambient tangent bundle $T\tilde{M}|_M \to M$. In a suitable sense, it inherits the torsion-free and the metric property from the Levi-Civita connection on $\tilde{M}$, which is unsurprising from its definition.

Proposition 1.1.11. (Properties of the induced connection)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. Denote by $\tilde{\nabla}$ the induced connection on M. Let $V, W \in \mathfrak{X}(M)$, $X, Y \in \tilde{\mathfrak{X}}(M)$. Then:

(1) The induced connection is a connection on the ambient tangent bundle $T\tilde{M}|_M$ over M.

(2) $[V, W] = \tilde{\nabla}_V W - \tilde{\nabla}_W V$.

(3) $V\langle X, Y\rangle = \langle \tilde{\nabla}_V X, Y\rangle + \langle X, \tilde{\nabla}_V Y\rangle$.

Proof. All of these follow immediately from the corresponding properties of the Levi-Civita connection on $\tilde{M}$, upon locally extending the given vector fields to local vector fields on $\tilde{M}$. □

By uniqueness, the tangential projection of the induced connection, when considered only on vector fields tangent to M, must necessarily agree with the Levi-Civita connection on M.

Lemma 1.1.12. (Tangential part of the induced connection)
For $V, W \in \mathfrak{X}(M)$, it holds that

$$\tan \tilde{\nabla}_V W = \nabla_V W,$$

where ∇ is the Levi-Civita connection on M.

Proof. By Proposition 1.1.11, $(V, W) \mapsto \tan \tilde{\nabla}_V W$ is a torsion-free, metric connection on M. Hence by uniqueness, it must coincide with the Levi-Civita connection. □

For $X, Y \in \mathfrak{X}(M)$, $\tilde{\nabla}_X Y \in \tilde{\mathfrak{X}}(M)$ is fully understood once its tangential and normal parts are known. By the previous result, the tangential part is just $\nabla_X Y$, so no extrinsic information is necessary to understand it. The normal part, however, provides insight into the extrinsic geometry of M in $\tilde{M}$.

Definition 1.1.13. (Second fundamental form)
The *second fundamental form* of M is the map

$$\begin{aligned} II : \mathfrak{X}(M) \times \mathfrak{X}(M) &\to \mathfrak{X}(M)^\perp \\ (X, Y) &\mapsto II(X, Y), \end{aligned}$$

where $II(X, Y) := \operatorname{nor} \tilde{\nabla}_X Y$.

Proposition 1.1.14. (Properties of II)

(1) II is symmetric.

(2) II is $C^\infty(M)$-bilinear.

(3) $II(X, Y)|_p$ only depends on X_p and Y_p.

Proof.

(1) By Proposition 1.1.11(2), we have for $X, Y \in \mathfrak{X}(M)$ that

$$[X, Y] = \tilde{\nabla}_X Y - \tilde{\nabla}_Y X.$$

Applying the normal projection yields

$$0 = \operatorname{nor} \tilde{\nabla}_X Y - \operatorname{nor} \tilde{\nabla}_Y X = II(X, Y) - II(Y, X).$$

(2) By definition, $II(X, Y)$ is $C^\infty(M)$-linear in X, hence also in Y by symmetry.

(3) This follows from the general proof of well-definedness in Lemma 1.1.10 and symmetry.

□

The following equation holds by definition, but we note it for convenience.

Corollary 1.1.15. (Gauss formula)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. For $X, Y \in \mathfrak{X}(M)$, we have

$$\tilde{\nabla}_X Y = \nabla_X Y + II(X, Y). \tag{1.1}$$

At this point, let us introduce a notion that will become important only later.

Definition 1.1.16. (Mean curvature vector field)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. The *mean curvature vector field* of M is the normal vector field $H \in \mathfrak{X}(M)^{\perp}$ defined locally via

$$H = \frac{1}{m} \sum_{i=1}^{m} g(E_i, E_i) II(E_i, E_i), \tag{1.2}$$

where $(E_1, \ldots, E_m)$ is a local orthonormal frame in M.

Note that H is well-defined by the properties of II.

Let us now introduce the Weingarten map for general semi-Riemannian submanifolds. While it does not necessarily provide more information than the second fundamental form, it can be useful in computations, as we shall see. Its importance will be more evident in the case of hypersurfaces later on, where there are fewer choices for normal directions.

Definition 1.1.17. (Weingarten map)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. For $N \in \mathfrak{X}(M)^{\perp}$, consider the map

$$\begin{aligned} II_N : \mathfrak{X}(M) \times \mathfrak{X}(M) &\to C^{\infty}(M), \\ II_N(X, Y) &:= \langle N, II(X, Y) \rangle. \end{aligned}$$

Since $II_N \in \mathcal{T}_2^0(M)$, $X \mapsto II_N(X, .)^{\#}$ is a map $\mathfrak{X}(M) \to \mathfrak{X}(M)$. It is denoted by W_N and is called the *Weingarten map of M in direction N*.

The Weingarten map W_N is characterized by

$$\langle W_N(X), Y \rangle = II_N(X, Y) = \langle N, II(X, Y) \rangle \tag{1.3}$$

for all $X, Y \in \mathfrak{X}(M)$. In particular, W_N defines a bundle homomorphism $TM \to TM$.

Proposition 1.1.18. (Weingarten equation)
For all $X \in \mathfrak{X}(M)$, $N \in \mathfrak{X}(M)^{\perp}$, it holds that

$$\tan \tilde{\nabla}_X N = -W_N(X). \tag{1.4}$$

Proof. Since both sides are in $\mathfrak{X}(M)$, it suffices to show that their inner products with an arbitrary $Y \in \mathfrak{X}(M)$ agree. Using the properties of the induced connection (cf. Proposition 1.1.11), the definition of the Weingarten map (cf. Definition 1.1.17), the Gauss formula (cf. Corollary 1.1.15), and the fact that $\langle N, Y \rangle = 0$, we calculate

$$\begin{aligned} 0 = X\langle N, Y \rangle &= \langle \tilde{\nabla}_X N, Y \rangle + \langle N, \tilde{\nabla}_X Y \rangle \\ &= \langle \tilde{\nabla}_X N, Y \rangle + \langle N, \nabla_X Y + II(X, Y) \rangle \\ &= \langle \tan \tilde{\nabla}_X N, Y \rangle + \langle N, II(X, Y) \rangle \\ &= \langle \tan \tilde{\nabla}_X N, Y \rangle + \langle W_N(X), Y \rangle, \end{aligned}$$

which proves the claim. □

If $\tilde{R}$ denotes the Riemannian curvature tensor of $\tilde{M}$, then it can be defined for $\tilde{M}$-vector fields on M by applying the definition of $\tilde{R}$ and understanding the appearing connections as induced connections on M. The following result relates the tangential part of the extrinsic curvature of M in $\tilde{M}$ with its intrinsic curvature. It turns out that their difference is given by terms involving the second fundamental form.

Theorem 1.1.19. (Gauss equation)
Let $V, X, Y, Z \in \mathfrak{X}(M)$. Then

$$\langle \tilde{R}_{VX} Y, Z \rangle = \langle R_{VX} Y, Z \rangle - \langle II(V, Z), II(X, Y) \rangle + \langle II(V, Y), II(X, Z) \rangle.$$

Proof. We need to relate $\tilde{\nabla}$ to ∇ via the Gauss equation and keep track of the appearing second fundamental forms and their derivatives. This is done by a simple (if lengthy) calculation, using the fact that II is always normal to M and we may use its associated Weingarten map. The Weingarten equation (cf. (1.4)) is then used to simplify the terms:

$$\begin{aligned} \langle \tilde{R}_{VX} Y, Z \rangle &= \langle \tilde{\nabla}_V \tilde{\nabla}_X Y - \tilde{\nabla}_X \tilde{\nabla}_V Y - \tilde{\nabla}_{[V,X]} Y, Z \rangle \\ &= \langle \tilde{\nabla}_V (\nabla_X Y + II(X, Y)) \\ &\quad - \tilde{\nabla}_X (\nabla_V Y + II(V, Y)) - \tilde{\nabla}_{[V,X]} Y, Z \rangle \\ &= \langle \tilde{\nabla}_V \nabla_X Y, Z \rangle - \langle W_{II(X,Y)}(V), Z \rangle \\ &\quad - \langle \tilde{\nabla}_X \nabla_V Y, Z \rangle + \langle W_{II(V,Y)}(X), Z \rangle - \langle \tilde{\nabla}_{[V,X]} Y, Z \rangle \\ &= \langle \tilde{\nabla}_V \nabla_X Y, Z \rangle - \langle II(X, Y), II(V, Z) \rangle \\ &\quad - \langle \tilde{\nabla}_X \nabla_V Y, Z \rangle + \langle II(V, Y), II(X, Z) \rangle - \langle \tilde{\nabla}_{[V,X]} Y, Z \rangle \\ &= \langle \nabla_V \nabla_X Y, Z \rangle - \langle \nabla_X \nabla_V Y, Z \rangle - \langle \nabla_{[V,X]} Y, Z \rangle \\ &\quad - \langle II(X, Y), II(V, Z) \rangle + \langle II(V, Y), II(X, Z) \rangle \\ &= \langle R_{VX} Y, Z \rangle - \langle II(V, Z), II(X, Y) \rangle + \langle II(V, Y), II(X, Z) \rangle. \end{aligned}$$

□

1.1.3 The normal connection and the Codazzi equation

Having derived a formula that gives us concrete information about the tangential part of the ambient curvature tensor, we aim to do something similar for its normal part. To do this, it is important to consider the following connection on the normal bundle.

Definition 1.1.20. (Normal connection)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. The *normal connection on M* is the map

$$\begin{aligned} \nabla^\perp : \mathfrak{X}(M) \times \mathfrak{X}(M)^\perp &\to \mathfrak{X}(M)^\perp, \\ \nabla^\perp_X N &:= \operatorname{nor} \tilde{\nabla}_X N. \end{aligned}$$

By the Weingarten equation we have that

$$\tilde{\nabla}_X N = -W_N(X) + \nabla^\perp_X N \tag{1.5}$$

for all $X \in \mathfrak{X}(M)$, $N \in \mathfrak{X}(M)^\perp$.

Lemma 1.1.21. (Properties of the normal connection)
$\nabla^\perp$ is a connection on the normal bundle NM of M. Moreover, it is a metric connection in the sense that for all $N_1, N_2 \in \mathfrak{X}(M)^\perp$, we have

$$X\langle N_1, N_2\rangle = \langle \nabla^\perp_X N_1, N_2\rangle + \langle N_1, \nabla^\perp_X N_2\rangle.$$

Proof. By definition, $\nabla^\perp_X N \in \mathfrak{X}(M)^\perp$. It then follows from the properties of the induced connection that $\nabla^\perp$ is a connection on NM: $C^\infty(M)$-linearity in X is clear, and the product rule follows from that of the induced connection and the fact that the normal projection restricted to $\mathfrak{X}(M)^\perp$ is the identity:

$$\begin{aligned} \nabla^\perp_X(fN) = \operatorname{nor}(\tilde{\nabla}_X fN) &= \operatorname{nor}(X(f)N) + \operatorname{nor}(f\,\tilde{\nabla}_X N) \\ &= X(f)N + f\nabla^\perp_X N, \end{aligned}$$

for all $f \in C^\infty(M)$, $X \in \mathfrak{X}(M)$, $N \in \mathfrak{X}(M)^\perp$ The metric property is also an easy consequence of the corresponding property of $\tilde{\nabla}$ (cf. Proposition 1.1.11):

$$\begin{aligned} X\langle N_1, N_2\rangle &= \langle \tilde{\nabla}_X N_1, N_2\rangle + \langle N_1, \tilde{\nabla}_X N_2\rangle \\ &= \langle \operatorname{nor} \tilde{\nabla}_X N_1, N_2\rangle + \langle N_1, \operatorname{nor} \tilde{\nabla}_X N_2\rangle \\ &= \langle \nabla^\perp_X N_1, N_2\rangle + \langle N_1, \nabla^\perp_X N_2\rangle \end{aligned}$$

for all $X \in \mathfrak{X}(M)$, $N_1, N_2 \in \mathfrak{X}(M)^\perp$. $\square$

Consider now the vector bundle $F := \operatorname{Hom}(TM \oplus TM, NM)$ over M. Its smooth sections correspond precisely to $C^\infty(M)$-bilinear maps $\mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)^\perp$, so in particular $II \in \Gamma(F)$.

Definition and Lemma 1.1.22. (A connection on F)
Let $B \in \Gamma(F)$. Then $\nabla^F : \mathfrak{X}(M) \times \Gamma(F) \to \Gamma(F)$, defined via

$$(\nabla^F_X B)(Y, Z) := \nabla^\perp_X (B(Y, Z)) - B(\nabla_X Y, Z) - B(Y, \nabla_X Z),$$

where $X, Y, Z \in \mathfrak{X}(M)$, is a connection on the vector bundle F over M.

Proof. Let B, X, Y, Z be as above. Clearly, $(\nabla^F_X B)(Y, Z) \in \mathfrak{X}(M)^\perp$ because all of the individual terms in the definition are themselves in $\mathfrak{X}(M)^\perp$. Moreover, $(\nabla^F_X B)(Y, Z)$ is $C^\infty(M)$-linear in X for the same reason. As for the product rule, observe that for $f \in C^\infty(M)$, we have (using the fact that $\nabla^\perp$ satisfies the product rule, cf. Lemma 1.1.21) that

$$\begin{aligned}
(\nabla^F_X fB)(Y, Z) &= \nabla^\perp_X (fB(Y, Z)) - fB(\nabla_X Y, Z) - fB(Y, \nabla_X Z) \\
&= X(f)B(Y, Z) + f\nabla^\perp_X (B(Y, Z)) \\
&\quad - fB(\nabla_X Y, Z) - fB(Y, \nabla_X Z) \\
&= X(f)B(Y, Z) + f(\nabla^F_X B)(Y, Z).
\end{aligned}$$

It remains to show that $\nabla^F_X B \in \Gamma(F)$, i.e. that it is C^∞-bilinear in its arguments. $\mathbb{R}$-bilinearity is trivial, so we only need to consider the case where the arguments are multiplied by an arbitrary $C^\infty(M)$-element f. Using the $C^\infty(M)$-bilinearity of B, we have

$$\begin{aligned}
(\nabla^F_X B)(fY, Z) &= \nabla^\perp_X (fB(Y, Z)) - B(\nabla_X fY, Z) - fB(Y, \nabla_X Z) \\
&= X(f)B(Y, Z) + f\nabla^\perp_X B(Y, Z) - X(f)B(Y, Z) \\
&\quad - fB(\nabla_X Y, Z) - fB(Y, \nabla_X Z) \\
&= f(\nabla^F_X B)(Y, Z),
\end{aligned}$$

proving $C^\infty(M)$-linearity in the first argument. An analogous calculation then shows the $C^\infty(M)$-linearity in the second argument, finishing the proof. □

With all of this notation, we can now express the normal part of the ambient curvature tensor in terms of ∇^F-derivatives of the second fundamental form.

Theorem 1.1.23. (Codazzi equation)
Let $W, X, Y \in \mathfrak{X}(M)$. Then

$$\operatorname{nor} \tilde{R}_{WX} Y = (\nabla^F_W II)(X, Y) - (\nabla^F_X II)(W, Y). \tag{1.6}$$

Proof. Since both sides of the claimed formula are in $\mathfrak{X}(M)^\perp$, it suffices to show that their inner products with an arbitrary $N \in \mathfrak{X}(M)^\perp$ agree. Clearly,

$$\langle \tilde{R}_{WX} Y, N \rangle = \langle \operatorname{nor} \tilde{R}_{WX} Y, N \rangle,$$

so we can use the Gauss formula (see Corollary1.1.15) to calculate as follows:

$$\begin{aligned}\langle \operatorname{nor} \tilde{R}_{WX}Y, N\rangle &= \langle \tilde{\nabla}_W(\nabla_X Y + II(X,Y)) - \tilde{\nabla}_X(\nabla_W Y + II(W,Y)) \\ &\quad - \tilde{\nabla}_{[W,X]}Y, N\rangle \\ &= \langle II(W, \nabla_X Y) + \nabla^{\perp}_W(II(X,Y)) \\ &\quad - II(X, \nabla_W Y) - \nabla^{\perp}_X(II(W,Y)) - II([W,X],Y), N\rangle.\end{aligned}$$

We now express the terms using the definition of ∇^F, e.g. $\nabla^{\perp}_W(II(X,Y)) - II(X, \nabla_W Y) = (\nabla^F_W II)(X,Y) + II(\nabla_W X, Y)$. Inserting these into the above calculation gives

$$\begin{aligned}\langle \operatorname{nor} \tilde{R}_{WX}Y, N\rangle &= \langle (\nabla^F_W II)(X,Y) + II(\nabla_W X, Y) \\ &\quad - (\nabla^F_X II)(W,Y) - II(\nabla_X W, Y) \\ &\quad - II([W,X],Y), N\rangle \\ &= \langle (\nabla^F_W II)(X,Y) - (\nabla^F_X II)(W,Y), N\rangle,\end{aligned}$$

where the last equality is due to vanishing torsion. □

Remark 1.1.24. (The tangent second fundamental form)
Recall from (1.5) that for $X \in \mathfrak{X}(M)$, $N \in \mathfrak{X}(M)^{\perp}$, we have that

$$\tilde{\nabla}_X N = -W_N(X) + \nabla^{\perp}_X N.$$

By definition, the Weingarten map satisfies $\langle W_N(X), Y\rangle = \langle N, II(X,Y)\rangle$ for $Y \in \mathfrak{X}(M)$. Interpreting the Weingarten map $W_N(X)$ as a map in both its entries, we get the *tangent second fundamental form*

$$\begin{aligned}II^{\tan} :&\mathfrak{X}(M) \times \mathfrak{X}(M)^{\perp} \to \mathfrak{X}(M), \\ &II^{\tan}(X,N) := -W_N(X).\end{aligned}$$

Hence

$$\tilde{\nabla}_X N = II^{\tan}(X,N) + \nabla^{\perp}_X N. \tag{1.7}$$

The tangent second fundamental form yields no new information, but it can sometimes be notationally useful. We note that, by definition,

$$\langle II^{\tan}(X,N), Y\rangle = -\langle W_N(X), Y\rangle = -\langle N, II(X,Y)\rangle. \tag{1.8}$$

1.1.4 Curves in semi-Riemannian submanifolds

Next, we shall consider curves in M. We want to understand their behavior when considered in M and when considered in the ambient space $\tilde{M}$. Before we go on, let us fix some notation: If (N, h) is a semi-Riemannian manifold,

γ is a smooth curve in N and $X \in \mathfrak{X}(\gamma)$, we denote the *covariant derivative* of X along γ (cf. [22, Prop. 3.18]) by

$$\frac{\nabla}{dt} X \equiv \nabla_t X,$$

where t is the parameter of the curve.

Corollary 1.1.25. (Gauss formula along a curve)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$, and let $\gamma : I \to M$ be a smooth curve in M. If $X \in \mathfrak{X}(\gamma)$ is everywhere tangent to M, then

$$\frac{\tilde{\nabla}}{dt} X = \frac{\nabla}{dt} X + II(\gamma', X). \tag{1.9}$$

Proof. Let $t_0 \in I$ and let $(E_1, \ldots, E_n)$ be an orthonormal frame around $\gamma(t_0)$ in $\tilde{M}$ adapted to M. Write $X(t) = X^j(t) E_j|_{\gamma(t)}$ for t near t_0, $j = 1, \ldots, m$. Then the Gauss formula yields

$$\begin{aligned}
\frac{\tilde{\nabla}}{dt} X|_{t=t_0} &= \dot{X}^j(t_0) E_j|_{\gamma(t_0)} + X^j(t_0)(\tilde{\nabla}_{\gamma'} E_j)_{\gamma(t_0)} \\
&= \dot{X}^j(t_0) E_j|_{\gamma(t_0)} + X^j(t_0) \nabla_{\gamma'(t_0)} E_j + X^j(t_0) II(\gamma'(t_0), E_j|_{\gamma(t_0)}) \\
&= \frac{\nabla}{dt} X|_{t=t_0} + II(\gamma'(t_0), X(t_0)).
\end{aligned}$$

□

Corollary 1.1.26. (Comparing accelerations)
If γ is a smooth curve in M, then the accelerations of γ with respect to M and with respect to $\tilde{M}$ satisfy

$$\frac{\tilde{\nabla}}{dt} \gamma' = \frac{\nabla}{dt} \gamma' + II(\gamma', \gamma'). \tag{1.10}$$

Proof. Set $X = \gamma'$ in (1.9). □

Note that the first derivative, γ', only depends on the differentiable structure and is hence independent of any notions of curvature in M or in $\tilde{M}$.

In the rest of this subsection, we want to discuss normal covariant derivatives along curves and obtain results on how to transport normal vectors along curves in a way that they stay normal throughout.

Definition 1.1.27. (Normal covariant derivative)
Let γ be a smooth curve in M and N a smooth vector field along γ in $\tilde{M}$ that is everywhere normal to M. Then its *normal covariant derivative* is

$$\frac{\nabla^{\perp}}{dt} N := \operatorname{nor} \frac{\tilde{\nabla}}{dt} N.$$

Similarly to the Gauss formula along curves, it can be derived that

$$\frac{\tilde{\nabla}}{dt}N = II^{\tan}(\gamma', N) + \frac{\nabla^{\perp}}{dt}N, \tag{1.11}$$

cf. Remark 1.1.24.

Definition 1.1.28. (Normal parallel vector fields)
Let γ be a smooth curve in M and N a vector field along γ that is everywhere normal to M. Then N is called *normal parallel* if

$$\frac{\nabla^{\perp}}{dt}N = 0.$$

Proposition 1.1.29. (Normal parallel transport)
Let $\gamma : I \to M \subseteq \tilde{M}$ be a smooth curve in M. If $y \in N_{\gamma(a)}M$, $a \in I$, then there is a unique normal parallel vector field N along γ such that $Y(a) = y$.

Proof. Let $(E_1(t), \ldots, E_n(t))$ be an orthonormal frame along γ such that $(E_{n-m}(t), \ldots, E_n(t))$ span $N_{\gamma(t)}M$ for all $t \in I$. Make the ansatz $Y = Y^i E_i$, $i = n-m, \ldots, n$, and write $y = y^i E_i(a)$, $i = n-m, \ldots, n$. Then, since the normal covariant derivative satisfies a product rule that it inherits from the normal connection, we have

$$\frac{\nabla^{\perp}}{dt}Y = ((Y^j)' + Y^i Z_i^j)E_j,$$

where Z_i^j are smooth functions on I, $i, j = n-m, \ldots, n$. So the problem reduces to solving the initial value problem

$$(Y^j)' + Y^i Z_i^j = 0, \quad Y^j(a) = y^j.$$

As this is a system of linear ODEs, there is a unique solution Y that is the normal parallel vector field we were looking for. ◻

Remark 1.1.30. (Normal parallel transport is an isometry)
Normal parallel transport induces an isometry of normal spaces: Let $\gamma : I \to M$ with $a, b \in I$, $a < b$. Consider the map

$$P_{ab}^{\gamma,\perp} : N_{\gamma(a)}M \to N_{\gamma(b)}M, \quad y \mapsto Y(b),$$

where Y is the unique normal parallel vector field along γ satisfying $Y(a) = y$. Since Y was found as the solution of a linear system of ODEs, $P_{ab}^{\gamma,\perp}$ is a linear map. It is an isometry because (cf. (1.11))

$$\frac{d}{dt}\langle Y(t), Y(t)\rangle = 2\langle \frac{\nabla^{\perp}}{dt}Y(t), Y(t)\rangle = 0.$$

1.1.5 Totally geodesic and totally umbilic submanifolds

We now turn to a brief discussion of submanifolds whose external geometry agrees with that of the ambient manifold.

Definition 1.1.31. (Totally geodesic submanifolds)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$. Then (M, g) is called *totally geodesic* if $II = 0$.

Proposition 1.1.32. (Characterization of totally geodesic submanifolds)
For a semi-Riemannian submanifold (M, g) of $(\tilde{M}, \tilde{g})$ the following are equivalent:

(1) (M, g) is totally geodesic.

(2) A smooth curve in M is a g-geodesic if and only if it is a $\tilde{g}$-geodesic.

(3) If $v \in T_pM \subseteq T_p\tilde{M}$, then the unique $\tilde{g}$-geodesic $\gamma_v^{\tilde{g}}$ with $\gamma_v^{\tilde{g}}(0) = p$ and $(\gamma_v^{\tilde{g}})'(0) = v$ lies initially in M.

(4) If α is a smooth curve in M (defined on an interval I with $0 \in I$), and $v \in T_{\alpha(0)}M$, then the g-parallel translate and the $\tilde{g}$-parallel translate of v agree.

Proof. (1) $\Rightarrow$ (2): If γ is a smooth curve in M, then the assumption $II = 0$ and Corollary 1.1.26 give

$$\frac{\tilde{\nabla}}{dt}\gamma' = \frac{\nabla}{dt}\gamma',$$

proving (2).
(2) $\Rightarrow$ (1): Given any $p \in M$ and any $v \in T_pM$, let γ_v^g be the unique g-geodesic with $\gamma_v^g(0) = p$, $(\gamma_v^g)'(0) = v$. By assumption, it is also a $\tilde{g}$-geodesic, hence Corollary 1.1.26 yields

$$II(v, v) = 0.$$

Since v was arbitrary, it follows that $II = 0$ by polarization.
(2) $\Rightarrow$ (3): Let $v \in T_pM$, then the assumption and uniqueness give that $\gamma_v^g(t) = \gamma_v^{\tilde{g}}(t)$ as long as the left hand side is defined.
(3) $\Rightarrow$ (1): Let $v \in T_pM$. Then Corollary 1.1.26 yields

$$0 = \frac{\tilde{\nabla}}{dt}\gamma_v^{\tilde{g}}|_{t=0} = \frac{\nabla}{dt}\gamma_v^{\tilde{g}}|_{t=0} + II(v, v),$$

and since this is an orthogonal decomposition into tangent and normal part, both have to be equal to 0, thus $II(v, v) = 0$ and $II = 0$ by polarization.

(1) $\Rightarrow$ (4): Let V be the unique g-parallel vector field along α in M satisfying $V(0) = v$. By Corollary 1.1.25, V is also $\tilde{g}$-parallel, hence the claim follows by uniqueness.
(4) $\Rightarrow$ (2): If γ is a smooth curve in M, then the assumption yields that γ' is g-parallel along γ if and only if it is $\tilde{g}$-parallel, which means precisely that γ is a g-geodesic if and only if it is a $\tilde{g}$-geodesic. $\square$

Geodesically complete totally geodesic submanifolds are rather special and are in a way already determined by one of their tangent spaces, as the following result shows.

Proposition 1.1.33. (”Rigidity” of totally geodesic submanifolds)
Let $M, N \subseteq \tilde{M}$ be connected, geodesically complete, totally geodesic semi-Riemannian submanifolds. Suppose there is a point $p \in M \cap N$ such that $T_pM = T_pN$. Then $M = N$.

Proof. Suppose first that M is connected and N is complete. We will show that $M \subseteq N$. Since both M and N are connected and geodesically complete by assumption, this will yield $M = N$.
Let σ be a geodesic segment in M from p to $q \in M$. Since M is totally geodesic, σ is an $\tilde{M}$-geodesic by Proposition 1.1.32, and $\sigma'(0) \in T_pM = T_pN$. Again by Proposition 1.1.32, σ is initially in N and $\sigma \cap N$ is an N-geodesic. By geodesic completeness, $\sigma \subseteq N$, in particular $q \in N$. Parallel translation along σ, which is independent of whether it is done in M, N or $\tilde{M}$, yields $T_qM = T_qN$, hence the same arguments apply to geodesic segments in M starting at q. Since M is connected, any two points can be connected via a broken geodesic (cf. [22, Lem. 3.32]), so by the arguments above we get $M \subseteq N$. By the comments at the beginning of the proof, we are done. $\square$

To close off this section, let us introduce a type of submanifold whose extrinsic curvature points in a specific normal direction. We will have more to say on these when we discuss hypersurfaces.

Definition 1.1.34. (Totally umbilic submanifolds)
Let (M, g) be a semi-Riemannian submanifold of $(\tilde{M}, \tilde{g})$.

(1) A point $p \in M$ is called *umbilic* if there exists $z \in N_pM$ such that
$$II(v, w) = \langle v, w \rangle z \quad \text{for all } v, w \in T_pM.$$
In this case, z is called the *normal curvature vector* of M at p.

(2) M is called *totally umbilic* if every point is umbilic. In other words, M is totally umbilic if and only if there exists $Z \in \mathfrak{X}(M)^{\perp}$ such that
$$II(V, W) = \langle V, W \rangle Z \quad \text{for all } V, W \in \mathfrak{X}(M).$$

In particular, totally geodesic submanifolds are totally umbilic with $Z \equiv 0$.

1.2 Semi-Riemannian hypersurfaces

In this section, we are interested in *semi-Riemannian hypersurfaces* in a semi-Riemannian manifold $(\tilde{M}, \tilde{g})$, i.e. semi-Riemannian submanifolds of codimension 1. Let (M, g) be such a semi-Riemannian hypersurface. Then its normal spaces N_pM are 1-dimensional, hence the restriction of g_p to $N_pM \times N_pM$ is either positive or negative definite. In particular, any (local) normal field N satisfies $\langle N, N\rangle > 0$ or $\langle N, N\rangle < 0$. If M is obtained as a regular level set of some function f, then $\operatorname{grad} f$ spans NM and g restricted to the normal spaces is entirely determined by its value on $\operatorname{grad} f$.
After this initial discussion, the following result is a trivial consequence of the regular level set theorem ([18, Cor. 5.14]) and the definition of the gradient.

Proposition 1.2.1. (Hypersurfaces as level sets)
Let $(\tilde{M}, \tilde{g})$ be a semi-Riemannian manifold of signature (k, l). Let $f : \tilde{M} \to \mathbb{R}$ be smooth and let $c \in \mathbb{R}$ be a regular value of f. Then $M := f^{-1}(c)$ is a semi-Riemannian hypersurface of $\tilde{M}$. Its signature is $(k-1, l)$ if $\langle \operatorname{grad} f, \operatorname{grad} f\rangle < 0$ on M and it is $(k, l-1)$ is $\langle \operatorname{grad} f, \operatorname{grad} f\rangle > 0$ on M.

Definition 1.2.2. (Hypersurfaces in Lorentzian manifolds)
If $(\tilde{M}, \tilde{g})$ is Lorentzian (of dimension $n+1$), i.e. $(k, l) = (1, n)$, then hypersurfaces of signature $(0, n)$ are called *spacelike* and those of signature $(1, n-1)$ are called *timelike.*

From now on, we assume throughout this section that (M, g) is an n-dimensional semi-Riemannian hypersurface of an $(n+1)$-dimensional semi-Riemannian manifold $(\tilde{M}, \tilde{g})$.
Locally, we may always choose a unit normal field N for M (since M is a hypersurface, there are precisely two such (local) choices, namely N and $-N$). Having made this choice, we consider the Weingarten map of M (in direction N), $X \mapsto W_N(X)$, characterized by

$$\langle W_N(X), Y\rangle = \langle N, II(X, Y)\rangle \quad \text{for all } X, Y \in \mathfrak{X}(M).$$

Definition 1.2.3. (Scalar second fundamental form)
The symmetric $(0, 2)$-tensor $h : \mathfrak{X}(M) \times \mathfrak{X}(M) \to C^\infty(M)$, defined by

$$h(X, Y) := \langle W_N(X), Y\rangle = \langle N, II(X, Y)\rangle,$$

is called the *scalar second fundamental form* of M.

Note that if $-N$ instead of N is chosen as unit normal, then h switches sign.

Remark 1.2.4. (On the scalar second fundamental form)

(1) Note that by the Gauss formula,

$$h(X, Y) = \langle N, \tilde{\nabla}_X Y\rangle. \tag{1.12}$$

(2) Since N spans NM, we have

$$II(X,Y) = \langle N,N\rangle\langle N, II(X,Y)\rangle N = \langle N,N\rangle h(X,Y)N, \qquad (1.13)$$

where $\langle N,N\rangle = \pm 1$.

(3) h is the quantity that is often referred to as the second fundamental form by physicists (in physics literature, the letter k is often used to denote it).

Once N has been chosen, it is customary to refer to the Weingarten map W_N simply as S and call it the *shape operator* of M. In this notation, (1.13) reads

$$II(X,Y) = \langle N,N\rangle\langle W_N(X), Y\rangle N = \langle N,N\rangle\langle S(X), Y\rangle N. \qquad (1.14)$$

Lemma 1.2.5. (Form of the shape operator)
Let M be a semi-Riemannian hypersurface in $\tilde{M}$, and let S be the shape operator of M associated to some choice of unit normal N. Then

$$S = -\tilde{\nabla} N. \qquad (1.15)$$

Proof. As was noted in Definition 1.1.20, for any $X \in \mathfrak{X}(M)$ we have

$$\tilde{\nabla}_X N = -S(X) + \nabla^{\perp}_X N.$$

We claim that N is parallel with respect to the normal connection: Since $\langle N,N\rangle = \pm 1 = const.$, we have

$$0 = X\langle N,N\rangle \overset{1.1.21}{=} 2\langle \nabla^{\perp}_X N, N\rangle.$$

Because N spans the normal bundle, it follows that $\nabla^{\perp}_X N = 0$. Thus

$$\tilde{\nabla}_X N = -S(X),$$

as was claimed. □

Next, we want to characterize totally umbilic hypersurfaces (recall the notion of totally umbilic submanifold from Definition 1.1.34). The following result shows that a hypersurface is totally umbilic if and only if the shape operator applied to any vector field is proportional to that vector field.

Proposition 1.2.6. (Characterizing totally umbilic hypersurfaces)
Let M be a semi-Riemannian hypersurface in $\tilde{M}$. M is totally umbilic if and only if its shape operator is scalar, i.e. there is an open covering $\{U_i\}$ of M and smooth functions $f_i \in C^\infty(U_i)$ such that in U_i,

$$S(X) = f_i X \quad \text{for all } X \in \mathfrak{X}(M).$$

Proof. Suppose M is totally umbilic, $II(X,Y) = \langle X,Y\rangle Z$ for all $X,Y \in \mathfrak{X}(M)$, where $Z \in \mathfrak{X}(M)^{\perp}$ is its normal curvature vector field. Let N be a choice of local unit normal and let S be the corresponding shape operator. Then for all $X,Y \in \mathfrak{X}(M)$,

$$\langle S(X),Y\rangle = \langle N, II(X,Y)\rangle = \langle X,Y\rangle\langle N,Z\rangle = \langle\langle N,Z\rangle X,Y\rangle,$$

which implies $S(X) = \langle N,Z\rangle X$. Hence S is scalar with $f = \langle N,Z\rangle$.
To prove the converse direction, suppose that for any choice of local unit normal N the corresponding shape operator S is scalar, so $S(X) = f_N X$ where f_N is a locally defined smooth function depending on N. Observe now that by Remark 1.2.4,

$$II(X,Y) = \langle N,N\rangle\langle S(X),Y\rangle N = \langle N,N\rangle f_N\langle X,Y\rangle N.$$

Since the shape operator changes sign if $N \to -N$, we have that $f_{-N} = -f_N$. Hence we may around each point choose a local unit normal N and define a global normal vector field $Z \in \mathfrak{X}(M)^{\perp}$ locally via $Z := \langle N,N\rangle f_N N$. Since, locally, the only choices are N or $-N$ and Z is unaffected by these choices, it is indeed a well defined global normal vector field that is the normal curvature vector field of M. Hence, M is totally umbilic. $\square$

Remark 1.2.7. (Various notions of curvature)
There are several important notions of curvature for Riemannian hypersurfaces which we now briefly discuss. Let (M,g) be a Riemannian hypersurface of a Riemannian or Lorentzian manifold $(\tilde{M},\tilde{g})$, with $\dim\tilde{M} = n+1$. Recall that the shape operator $S : \mathfrak{X}(M) \to \mathfrak{X}(M)$ is self-adjoint because the second fundamental form is symmetric (cf. Proposition 1.1.14(1)):

$$\langle S(X),Y\rangle = \langle N, II(X,Y)\rangle = \langle N, II(Y,X)\rangle = \langle X, S(Y)\rangle$$

for all $X,Y \in \mathfrak{X}(M)$. Considering individual tangent spaces, for each $p \in M$, $S_p : T_pM \to T_pM$ is a self-adjoint endormorphism (with respect to the positive definite inner product g_p) of a finite-dimensional vector space, so S_p is orthogonally diagonalizable with real eigenvalues $\kappa_1,\ldots,\kappa_n$. Let $(b_1,\ldots,b_n)$ be an orthonormal basis of T_pM such that $S_p b_i = \kappa_i b_i$. Since the scalar second fundamental form is defined via $h(X,Y) = \langle S(X),Y\rangle$ for $X,Y \in \mathfrak{X}(M)$, one has for $v,w \in T_pM$

$$h_p(v,w) = \sum_{i=1}^{n}\kappa_i g_p(v,b_i)g_p(w,b_i).$$

The $\kappa_1,\ldots,\kappa_n$ are called *principal curvatures* of M at p, and the b_i are the corresponding *principal directions*. The principal curvatures depend on N only up to sign, just like h.

The determinant of S_p, $K := \det S_p = \kappa_1 \dots \kappa_n$, is called *the Gaussian curvature* of M at p. It is multiplied by $(-1)^n$ if one changes the normal from N to $-N$.
The famous Gauss-Bonnet Theorem says that if (M, g) is a smoothly triangulated compact Riemannian 2-manifold, then the integral over K is a purely topological quantity (cf. [19, Thm. 9.7]).
The object $H_p^{scal} := \langle H_p, N_p \rangle$ is called the *mean curvature* of M at p, where H_p is the mean curvature vector field of M at p, cf. Definition 1.1.16. One sees that $H_p^{scal} = (1/n)\,\mathrm{tr}(S_p) = (1/n)(\kappa_1 + \dots + \kappa_n)$.

Remark 1.2.8. (Sectional curvature)
For the sake of completeness, we briefly discuss the notion of sectional curvature. It will not play a role in the singularity theorems, but sectional curvature bounds are a crucial ingredient of many important results.

Let (M, g) be a semi-Riemannian manifold, and let Π be a nondegenerate (with respect to g_p) 2-dimensional vector subspace of T_pM (a "section"), $p \in M$. Then the *sectional curvature* of Π at p is

$$K(\Pi) \equiv K(v, w) := \frac{\langle R_{vw}w, v\rangle}{\langle v, v\rangle\langle w, w\rangle - \langle v, w\rangle^2},$$

where $\{v, w\}$ is a basis of Π. One can show that K does not depend on the choice of basis of Π ([22, Lem. 3.39]). K is easily seen to define a smooth map on the subbundle $G_2(M)^{nondeg}$ of the 2-Grassmannian bundle consisting of nondegenerate 2-planes. Important basic results on the sectional curvature and its relation to the curvature tensor can be found in e.g. [22, p. 77-80], [19, p. 250-255] (for the Riemannian case) and [14, p. 184-186].
We can use the Gauss equation (cf. Theorem 1.1.19) to derive a formula for the sectional curvature of a semi-Riemannian submanifold (M, g) in relation to that of the ambient space $(\tilde{M}, \tilde{g})$: If $T_pM \supseteq \Pi$ is a nondegenerate 2-plane spanned by $\{v, w\}$, then

$$\tilde{K}(v, w) = K(v, w) - \frac{\langle II(v, v), II(w, w)\rangle - \langle II(v, w), II(v, w)\rangle}{\langle v, w\rangle\langle w, w\rangle - \langle v, w\rangle^2}.$$

If (M, g) is a hypersurface with shape operator S corresponding to a normal N, then this formula simplifies (using the representation of II as given in Remark 1.2.4) to

$$\tilde{K}(v, w) = K(v, w) - \langle N_p, N_p\rangle \frac{\langle S_p(v), v\rangle\langle S_p(w), w\rangle - \langle S_p(v), w\rangle^2}{\langle v, v\rangle\langle w, w\rangle - \langle v, w\rangle^2}.$$

These formulas are very helpful in calculating sectional curvatures of submanifolds, especially those embedded in some (pseudo-)Euclidean space, since there $\tilde{K} = 0$.

Let us demonstrate this for the sphere $S^n(r)$ of radius r in Euclidean space. Namely, this space is of constant sectional curvature with

$$K = \begin{cases} 0 & n = 1, \\ \frac{1}{r^2} & n \geq 2. \end{cases}$$

The case $n = 1$ is immediate since any 1-dimensional semi-Riemannian manifold is flat (by the antisymmetry properties of the curvature tensor). For $n \geq 2$, consider on $\mathbb{R}^{n+1}$ the position vector field

$$x \mapsto P(x) := \frac{1}{r} \sum_{i=1}^{n+1} x^i \frac{\partial}{\partial x^i}.$$

This is everywhere normal to $S^n(r)$ (as can be seen by considering $S^n(r)$ as a level set of the function $x \mapsto \|x\|^2$ and then calculating the differential of that map). By Lemma 1.2.5, the shape operator of $S^n(r)$ corresponding to this normal satisfies ($\overline{\nabla}$ is the flat connection on $\mathbb{R}^{n+1}$)

$$S(v) = -\overline{\nabla}_v P = -\frac{1}{r} \sum_{i=1}^{n+1} v(x^i) \frac{\partial}{\partial x^i} = -\frac{1}{r} v,$$

for any tangent vector $v \in TS^n(r)$, where we suppressed basepoints in the notation. We may w.l.o.g. assume that a given tangent plane is spanned by orthogonal vectors v and w. Using all of this, the formula derived above with $\tilde{K} = 0$ gives the claim.

$S^n(r)$ is the prototypical example of a Riemannian manifold with constant positive sectional curvature. The Riemannian manifolds with constant vanishing and constant negative sectional curvature are $\mathbb{R}^n$ and hyperbolic space $\mathbb{H}^n$, respectively. In the Lorentzian case, the corresponding constant curvature spaces are de-Sitter space, Minkowksi space and anti-de Sitter space. See [19, Ch. 3] and [22, Ch. 4] for much more on these model manifolds.

1.3 Topological hypersurfaces in spacetimes

In this final section of the first chapter, we will focus exclusively on spacetimes (M, g), i.e. connected, time-oriented Lorentzian manifolds. We will consider subsets of M whose points are not timelike/causally related and we will give criteria for these sets to be topological hypersurfaces in M.

The reference for this section is [22, p. 413-415]. Throughout this section, let (M, g) be a spacetime of dimension $(n + 1)$.

Definition 1.3.1. (Achronal, acausal and future/past sets)
Let $A \subseteq M$.

(1) A is called *achronal* if there are no points $p, q \in A$ such that $p \ll q$.

(2) A is called *acausal* if there are no points $p \neq q$ in A such that $p < q$.

(3) A is called a *future set* if $I^+(A) \subseteq A$. *Past sets* are defined analogously.

Definition 1.3.2. (Edge of an achronal set)
Let $A \subseteq M$ be achronal. The *edge* of A, denoted $\mathrm{edge}(A)$, is the set of all $p \in \overline{A}$ such that for any neighborhood U of p there is a timelike curve from $I_U^-(p)$ to $I_U^+(p)$ that does not meet A.

Lemma 1.3.3. (Achronal identity)
Let $A \subseteq M$ be an achronal set. Then

$$\overline{A} \setminus A \subseteq \mathrm{edge}(A).$$

Proof. Note first that if A is achronal, then so is $\overline{A}$: Because if $p, q \in \overline{A}$ such that $p \ll q$ and $A \ni p_n \to p$, $A \ni q_n \to q$, then the openness of the $\ll$-relation (cf. [22, Lem. 14.3]) implies $p_n \ll q_n$ for large n, contradicting achronality of A.
Let now $q \in \overline{A} \setminus A$ and let γ be any timelike curve through q. By achronality, γ does not meet $\overline{A}$ at any other point, in particular it does not meet A. Hence $q \in \mathrm{edge}(A)$. □

Definition 1.3.4. (C^0-hypersurface)
A subset $S \subseteq M$ is called C^0*-hypersurface* (or *topological hypersurface*) if for any $p \in M$ there is a neighborhood U of p in M and a homeomorphism $\phi : U \to \phi(U) \subseteq \mathbb{R}^n$ onto an open subset of $\mathbb{R}^n$ such that $\phi(U \cap S)$ is the intersection of $\phi(U)$ with some hyperplane in $\mathbb{R}^n$.

Theorem 1.3.5. (Achronal sets as C^0-hypersurfaces)
An achronal set $A \subseteq M$ is a C^0-hypersurface if and only if $A \cap \mathrm{edge}(A) = \emptyset$.

Proof. Suppose first that A is a C^0-hypersurface. Let $p \in A$. We will show that $p \notin \mathrm{edge}(A)$. To this end, let U be a neighborhood of p as in Definition 1.3.4. Upon shrinking U, we may assume that U is connected and that $U \setminus A$ has exactly two connected components (this is possible because a ball without a plane in $\mathbb{R}^n$ has this property). Consider the relative timelike futures $I_U^-(p)$ and $I_U^+(p)$. These are open, disjoint sets that do not meet A by achronality. They are easily seen to be connected. Now, any timelike curve through p must meet both sets by definition, hence they are in different components of $U \setminus A$. Hence any timelike curve from $I_U^-(p)$ to $I_U^+(p)$ must meet A by connectedness. This shows that $p \notin \mathrm{edge}(A)$ and thus $A \cap \mathrm{edge}(A) = \emptyset$.
To prove the converse direction, suppose A is achronal and $A \cap \mathrm{edge}(A) = \emptyset$. Let $p \in M$ and let $(U, \phi = (x^0, \ldots, x^{n-1}))$ be a chart around p such that $\frac{\partial}{\partial x^0}$ is future-directed timelike. We may find a smaller neighborhood $V \subseteq U$ of p such that $\phi(V) = (a - \delta, b + \delta) \times N \subseteq \mathbb{R}^n$, $\{x^0 = a\} \subseteq I_U^-(p)$ and

$\{x^0 = b\} \subseteq I_U^+(p)$. Shrinking U further, due to $p \notin \mathrm{edge}(A)$ we may assume that $s \mapsto \phi^{-1}(s, y)$, $s \in [a, b]$ and $y \in N$ arbitrary, meets A. By achronality, this meeting point is unique. Denote by $h(y)$ the x^0-coordinate of this point. We will now show that $h : N \to (a, b)$ is continuous. The claim then follows from this, because $\Phi := (x^0 - h \circ (x^1, \ldots, x^{n-1}), x^1, \ldots, x^{n-1})$ is clearly a homeomorphism that takes $A \cap V$ to $\{x^0 = 0\} \cap \phi(V)$.
To show the continuity of h, let $y_n \in N$, $y_n \to y \in N$ and assume $h(y_n) \not\to h(y)$. Since the values of h are bounded, there exists a subsequence of $h(y_n)$ that converges to some $r \neq h(y)$. Then $\phi^{-1}(r, y) \in I_V^-(q) \cup I_V^+(q)$, where $q := \phi^{-1}(h(y), y) \in A$. Since the union of these timelike futures is an open set, it contains all $\phi^{-1}(h(y_n), y_n) \in A$ for n large due to convergence. But this contradicts achronality of A. $\square$

Corollary 1.3.6. (Closed achronal sets)
An achronal set $A \subseteq M$ is a (topologically) closed C^0-hypersurface if and only if $\mathrm{edge}(A) = \emptyset$.

Proof. Suppose A is a closed C^0-hypersurface. By Theorem 1.3.5, $A \cap \mathrm{edge}(A) = \emptyset$. Since A is closed and by definition $\mathrm{edge}(A)$ is contained in the closure of A, it must be empty.
Suppose now that $\mathrm{edge}(A) = \emptyset$. Again by Theorem 1.3.5, A is a C^0-hypersurface. By Lemma 1.3.3, $\overline{A} \setminus A \subseteq \mathrm{edge}(A) = \emptyset$, hence $A = \overline{A}$. $\square$

Corollary 1.3.7. (Boundaries of future/past sets)
Let $F \subseteq M$ be a future or past set. Then either $\partial F = \emptyset$ or ∂F is an achronal set and a closed C^0-hypersurface.

Proof. We may assume F to be a future set and $\partial F \neq \emptyset$. We first show achronality: Let $p \in \partial F$. For $q \in I^+(p)$, $I^-(q)$ is a neighborhood of p and hence $I^-(q) \cap F \neq \emptyset$, which means $q \in I^+(F) \subseteq F$ because F is a future set. This shows that $I^+(p) \subseteq F$. Analogously, $I^-(p) \subseteq M \setminus F$ (because the complement of a future set is a past set). In particular, we have shown that $I^+(\partial F) \cap I^-(\partial F) = \emptyset$, implying achronality of ∂F.
Note that ∂F is trivially closed. It has no edge points because for any $p \in \partial F$ the arguments above show $I^+(p) \subseteq F^\circ$ and $I^-(p) \subseteq (M \setminus F)^\circ$, hence any timelike curve from $I^-(p)$ to $I^+(p)$ must meet ∂F, thus $p \notin \mathrm{edge}(\partial F)$. So $\mathrm{edge}(\partial F) = \emptyset$ and the result follows from Corollary 1.3.6. $\square$

Chapter 2

Variational Theory of Curves

In this chapter, we want to study clusters of curves (called *variations*) in spacetimes and understand their focusing and defocusing behavior. This analysis is done by way of *Jacobi fields* which arise as transversal derivative vector fields of geodesic variations. Whether or not nearby curves focus is determined by the existence of *conjugate points*. It turns out that maximality properties of causal curves are intimately linked with the (non)existence of conjugate points.

We will separate the case of timelike and the case of null geodesics in order to better see the differences that arise in their treatments.

Our main reference for this chapter is [1, Ch. 10], while also some elements from [22, Ch. 10], [19, Ch. 6, 10] (especially for the treatment of variations and Jacobi fields) have been used. Several general results are (in my opinion) written down most elegantly in [15] and will be used in that form.

2.1 Jacobi fields and geodesic variations

In this section, we introduce the formalism behind variations of curves and give a treatment of Jacobi fields. These notions will be used later to understand conjugate points along timelike and null geodesics in spacetimes. Unless otherwise stated, let (M, g) be a semi-Riemannian manifold of signature (k, l) with $k + l = n = \dim M$.

Definition 2.1.1. (Variations)
Let $c_0 : [a, b] \to M$ be a smooth curve in M.

(1) A smooth map $c : [a, b] \times J \to M$, where J is an open interval with $0 \in J$, is called a *variation* of c if $c(t, 0) = c_0(t)$ for all $t \in [a, b]$. If $c(a, s) = c(b, s)$ for all $s \in J$, then c is called a *fixed endpoint variation* (or *FEP-variation*). If c_0 is a geodesic and each longitudinal curve $c(., s)$ is as well, then c is called a *geodesic variation.*

Variations for piecewise smooth curves are defined analogously: instead of smoothness one demands piecewise smoothness in the sense that c is continuous and smooth on subrectangles $[a_i, a_{i+1}] \times J$. In that case the transverse curves $s \mapsto c(t,s)$ are smooth but the longitudinal curves $t \mapsto c(t,s)$ are only piecewise smooth.

(2) Let c be a variation of c_0 as above. The *variation vector field* of c is the vector field $V \in \mathfrak{X}(c_0)$ defined by

$$V(t) := \left.\frac{\partial c(t,s)}{\partial s}\right|_{s=0}.$$

If c is only piecewise smooth, then V is also piecewise smooth. In this case we understand V' to be defined everywhere via left limits.

One may in general consider curves c_0 defined on an open interval I, a variation of c_0 will then be a smooth map $c : I \times J \to M$ as above.

Lemma 2.1.2. (Any vector field is a variation vector field)
Let $c_0 : [a,b] \to M$ be piecewise smooth and let V be a piecewise smooth vector field along c_0. Then there exists a variation c of c_0 with variation vector field V. If $V(a) = 0$ and $V(b) = 0$, then c can be chosen to be an FEP-variation.

Proof. Let $c(t,s) := \exp_{c_0(t)}(sV(t))$. By compactness of $[a,b]$, c is defined on $[a,b] \times (-\varepsilon, \varepsilon)$ for ε small. Clearly, if $a = a_0 < a_1 < \cdots < a_m = b$ is a subdivision of $[a,b]$ such that both c_0 and V are smooth on each $[a_i, a_{i+1}]$, then c is smooth on each $[a_i, a_{i+1}] \times (-\varepsilon, \varepsilon)$. Moreover,

$$\left.\frac{\partial c(t,s)}{\partial s}\right|_{s=0} = T_0 \exp_{\gamma(t)}(V(t)) = V(t).$$

If V vanishes at a and b, then c satisfies $c(a,s) = \gamma(a)$ and $c(b,s) = \gamma(b)$ for all $s \in (-\varepsilon, \varepsilon)$, hence c is an FEP-variation. $\square$

We present two general results on covariant derivatives that will be used throughout.

Lemma 2.1.3. (Symmetry of second derivatives)
Let $c : [a,b] \times J \to M$ be a variation of a piecewise smooth curve $c_0 = c(.,0)$. Then

$$\frac{\nabla}{\partial t}\frac{\partial c}{\partial s} = \frac{\nabla}{\partial s}\frac{\partial c}{\partial t},$$

where the above covariant derivatives are the ones along the longitudinal and transversal curves, respectively.

Proof. Fix a point $c(t_0, s_0)$ and consider coordinates (x^j) around it. It holds that

$$\frac{\partial c}{\partial t} = \frac{\partial(x^j \circ c)}{\partial t}\frac{\partial}{\partial x^j}, \quad \frac{\partial c}{\partial s} = \frac{\partial(x^j \circ c)}{\partial s}\frac{\partial}{\partial x^j}.$$

Hence

$$\begin{aligned}
\frac{\nabla}{\partial s}\frac{\partial c}{\partial t} &= \left(\frac{\partial^2(x^k \circ c)}{\partial s \partial t} + \frac{\partial(x^i \circ c)}{\partial t}\frac{(\partial x^j \circ c)}{\partial s}(\Gamma^k_{ji} \circ c)\right)\frac{\partial}{\partial x^k},\\
\frac{\nabla}{\partial t}\frac{\partial c}{\partial s} &= \left(\frac{\partial^2(x^k \circ c)}{\partial t \partial s} + \frac{\partial(x^i \circ c)}{\partial s}\frac{(\partial x^j \circ c)}{\partial t}(\Gamma^k_{ji} \circ c)\right)\frac{\partial}{\partial x^k},
\end{aligned}$$

cf. [19, Thm. 4.24] for the local coordinate formula of a covariant derivative. The result follows from the symmetry of the Christoffel symbols. □

Lemma 2.1.4. (Exchanging two covariant derivatives)
Let $f : I \times J \to M$ be a smooth map, denote the points in $I \times J$ by (t, s) and let $X \in \mathfrak{X}(f)$. Then

$$\nabla_s \nabla_t X - \nabla_t \nabla_s X = R(\partial_s f, \partial_t f)X.$$

Proof. Fix $(t, s) \in I \times J$ and consider coordinates $\phi = (x^i)$ around $f(t, s)$. Write $f^i := x^i \circ f$ and $X = X^j \partial_j$. Then by the usual formula for covariant derivation we get

$$\nabla_t X = \frac{\partial X^i}{\partial t}\partial_i + X^i \nabla_t \partial_i,$$

where $(\nabla_t \partial_i)(t, s) = \nabla_{\partial_t f(t,s)} \partial_i$. Now taking the covariant derivative with respect to s, we get

$$\nabla_s \nabla_t X = \frac{\partial^2 X^i}{\partial s \partial t}\partial_i + \frac{\partial X^i}{\partial_t}\nabla_s \partial_i + \frac{\partial X^i}{\partial s}\nabla_t \partial_i + X^i \nabla_s \nabla_t \partial_i.$$

If we interchange s and t and subtract, we get

$$\nabla_s \nabla_t X - \nabla_t \nabla_s X = X^i(\nabla_s \nabla_t \partial_i - \nabla_t \nabla_s \partial_i). \quad (*)$$

Note that $\partial_t f = \partial_t f^i \partial_i$ and similarly for $\partial_s f$. Extending $\partial_s f$ and $\partial_t f$ locally in an arbitrary way, we may write

$$\nabla_t \partial_i = \nabla_{\partial_t f} \partial_i = \partial_t f^j \nabla_{\partial_j} \partial_i$$

and hence

$$\begin{aligned}
\nabla_s \nabla_t \partial_i &= \partial_s \partial_t f^j \nabla_{\partial_j} \partial_i + \partial_t f^j \nabla_s \nabla_{\partial_j} \partial_i\\
&= \partial_s \partial_t f^j \nabla_{\partial_j} \partial_i + \partial_t f^j \partial_s f^k \nabla_{\partial_k} \nabla_{\partial_j} \partial_i.
\end{aligned}$$

Again, we exchange s and t and subtract to get

$$\begin{aligned}\nabla_s\nabla_t\partial_i - \nabla_s\nabla_t\partial_i &= \partial_t f^j \partial_s f^k (\nabla_{\partial_k}\nabla_{\partial_j}\partial_i - \nabla_{\partial_j}\nabla_{\partial_k}\partial_i) \\ &= \partial_t f^j \partial_s f^k R(\partial_k, \partial_j)\partial_i = R(\partial_s f, \partial_t f)\partial_i.\end{aligned}$$

Inserting this into $(*)$ gives the claim. $\square$

Now we come to the definition of Jacobi fields along geodesics. They arise as variation vector fields of geodesic variations, as will be discussed further below.

Definition 2.1.5. (Jacobi field)
Let $\gamma : I \to M$ be a geodesic. A vector field $J \in \mathfrak{X}(\gamma)$ is called a *Jacobi field* along γ, if it satisfies the *Jacobi equation* on I:

$$J'' + R_{J\gamma'}\gamma' = 0.$$

Proposition 2.1.6. (Variation fields of geodesic variations are Jacobi fields)
Let $\gamma : I \times J \to M$ be a geodesic variation of the geodesic $\gamma_0 = \gamma(., 0)$. Then the variation vector field J of γ is a Jacobi field along γ_0.

Proof. Let us abbreviate $T(t, s) = \partial_t\gamma(t, s)$ and $S(t, s) = \partial_s\gamma(t, s)$. Thus $t \mapsto S(t, 0)$ is the variation field of γ. Since each $\gamma(., s)$ is a geodesic, we have

$$\frac{\nabla}{\partial t}T = 0, \quad \text{hence trivially} \quad \frac{\nabla}{\partial s}\frac{\nabla}{\partial t}T = 0.$$

Since exchanging two covariant derivatives yields a Riemann tensor (cf. Lemma 2.1.4), we may calculate as follows:

$$0 = \nabla_s\nabla_t T = \nabla_t\nabla_s T + R_{ST}T \overset{2.1.3}{=} \nabla_t\nabla_t S + R_{ST}T.$$

Evaluating this at $s = 0$ and using that $S(t, 0) = J(t)$ and $T(t, 0) = \gamma_0'(t)$ gives the claim. $\square$

Recall that our convention for the indices of the curvature tensor is $R(E_i, E_j)E_k = R_{ijk}{}^l E_l$ for a given local frame (E_i).

Proposition 2.1.7. (Generating Jacobi fields from data)
Let $\gamma : I \to M$ be a geodesic, $p := \gamma(a)$. Then for any two vectors $v, w \in T_pM$ (initial values), there is a unique Jacobi field J along γ satisfying $J(a) = v$ and $\nabla_t J(a) = w$.

Proof. Let (E_i) be a parallel orthonormal frame along γ (obtained by choosing an ONB at a point and then parallel transporting along γ). Write $v = v^i E_i(a)$, $w = w^i E_i(a)$, $\gamma'(t) = z^i(t)E_i(t)$. Now let $J \in \mathfrak{X}(\gamma)$, $J(t) =$

$J^i(t)E_i(t)$. Since the E_i are parallel, $J''(t) = \nabla_t\nabla_t J(t) = (J^i)''(t)E_i(t)$. Moreover,

$$\begin{aligned} R(J(t), \gamma'(t))\gamma'(t) &= R\left(J^i(t)E_i(t), z^j(t)E_j(t)\right) z^k(t)E_k(t) \\ &= J^i(t)z^j(t)z^k(t){R_{ijk}}^l(\gamma(t))E_l(t), \end{aligned}$$

where ${R_{ijk}}^l(\gamma(t))$ are the components of the curvature tensor w.r.t. arbitrary local frames that extend the (E_i) around points of γ (since we only consider their values on γ, everything is well-defined). Hence the Jacobi equation for an unknown J reduces to the following equation in coordinates:

$$(J^i)''(t) + {R_{ljk}}^i(\gamma(t))J^l(t)z^j(t)z^k(t).$$

This is a system of n linear second-order ODEs in J^i. Let us convert this into a first order system via an auxiliary vector field $W(t) = W^i(t)E_i(t)$:

$$\begin{aligned} &(J^i)'(t) = W^i(t), \\ &W^i(t) = -R^i_{ljk}(\gamma(t))J^l(t)z^j(t)z^k(t). \end{aligned}$$

Specifying the initial conditions $J^i(a) = v^i$, $W^i(a) = w^i$ thus gives a unique solution J on all of I by usual linear ODE theory. Noting that $\nabla_t J(a) = (J^i)'(a)E_i(a) = W^i(a)E_i(a) = w$ finishes the proof. □

Corollary 2.1.8. (Dimension of the space of Jacobi fields)
Let γ be a geodesic in M. The space $\mathfrak{Jac}(\gamma) \subseteq \mathfrak{X}(\gamma)$ of all Jacobi fields along γ is a $2n$-dimensional vector subspace of $\mathfrak{X}(\gamma)$.

Proof. First of all, $\mathfrak{Jac}(\gamma)$ is a vector space because the Jacobi equation is linear. Its dimension is $2n$ by Proposition 2.1.7: After fixing a parameter value $a \in I$, where $\gamma : I \to M$, any Jacobi field is uniquely determined by $J(a)$, $\nabla_t J(a) \in T_{\gamma(a)}M$, thus $J \mapsto (J(a), \nabla_t J(a))$ is a linear isomorphism $\mathfrak{Jac}(\gamma) \to T_{\gamma(a)}M \oplus T_{\gamma(a)}M$. □

The next result gives a partial converse to Proposition 2.1.6 and shows that, under suitable circumstances, any Jacobi field is the variation field of some geodesic variation.

Proposition 2.1.9. (Jacobi fields come from geodesic variations)
Let $\gamma : I \to M$ be a geodesic. Suppose that I is a compact interval or that M is geodesically complete. Then every $J \in \mathfrak{Jac}(\gamma)$ is the variation field of some geodesic variation of γ.

Proof. We may assume that $0 \in I$. Let $p := \gamma(0)$, $v = \gamma'(0)$. Hence, $\gamma(t) = \exp_p(tv)$ for all $t \in I$ (by either of the assumptions). Let $\sigma : (-\varepsilon, \varepsilon) \to M$

be a smooth curve with $\sigma(0) = p$, $\sigma'(0) = J(0)$, and let $V \in \mathfrak{X}(\sigma)$ satisfying $V(0) = v$, $V'(0) = J'(0)$. Let

$$\tilde{\gamma}(t, s) := \exp_{\sigma(s)}(tV(s)).$$

If M is complete, $\tilde{\gamma}$ is certainly defined on all of $I \times (-\varepsilon, \varepsilon)$. If I is compact, then we may choose a minimal global $\varepsilon > 0$. Note that

$$\tilde{\gamma}(t, 0) = \exp_p(tv) = \gamma(t).$$

Thus $\tilde{\gamma}$ is a variation of γ. Fixing s, $t \mapsto \exp_{\sigma(s)}(tV(s))$ is clearly a geodesic by the definition of the exponential map. Thus $\tilde{\gamma}$ is a geodesic variation. By Proposition 2.1.6, its variation vector field $W(t) := \partial_s \tilde{\gamma}(t, s)|_{s=0}$ is a Jacobi field along $\gamma = \tilde{\gamma}(., 0)$. Since

$$\tilde{\gamma}(0, s) = \exp_{\sigma(s)}(0) = \sigma(s),$$

it follows that

$$W(0) = \frac{\partial}{\partial s}\tilde{\gamma}(0, s)|_{s=0} = \sigma'(0) = J(0).$$

If we show that $W'(0) = J'(0)$, then Proposition 2.1.7 implies $W = J$ on all of I, thus yielding the claim.
By elementary properties of the exponential map,

$$\partial_t \tilde{\gamma}(t, s)|_{t=0} = \partial_t \exp_{\sigma(s)}(tV(s))|_{t=0} = V(s), \quad (*)$$

and $\nabla_t \partial_s \tilde{\gamma} = \nabla_s \partial_t \tilde{\gamma}$ by Lemma 2.1.3. Finally, due to our choice of V, the claim above follows from

$$\begin{aligned} \nabla_t W(t)|_{t=0} &= \nabla_t(\partial_s \tilde{\gamma}(t, s)|_{s=0})|_{t=0} = \nabla_s(\partial_t \tilde{\gamma}(t, s)|_{t=0})|_{s=0} \\ &\overset{(*)}{=} \nabla_s V(s)|_{s=0} = \nabla_t J(t)|_{t=0}. \end{aligned}$$

□

Example 2.1.10. (Trivial examples)
Let $\gamma : I \to M$ be a geodesic.

(1) $J_1(t) := \gamma'(t)$ is a Jacobi field along γ, because $J_1'(t) = \gamma''(t) = 0$ and thus $J_1''(t) = 0$, and also $R(J_1(t), \gamma'(t))\gamma'(t) = R(\gamma'(t), \gamma'(t))\gamma'(t) = 0$ by the properties of the curvature tensor. If M is geodesically complete or I is compact, then we may follow the proof of Proposition 2.1.9 to construct the geodesic variation of which J_1 is the variation field: Clearly, $\sigma(s) = \gamma(s)$, and V may be chosen as $V(s) = \gamma'(s)$, thus the desired geodesic variation is

$$\tilde{\gamma}(t, s) = \exp_{\gamma(s)}(t\gamma'(s)) = \gamma(t + s).$$

(2) $J_2(t) := t\gamma'(t)$ is also a Jacobi field along γ by similarly easy arguments as above. We calculate the corresponding geodesic variation if the assumptions in Proposition 2.1.9 are met: Since $\sigma(0) = \gamma(0)$, $\sigma'(0) = J_2(0) = 0$, we have $\sigma(s) = \gamma(0)$ for all s. As for a choice of V: We require $V(0) = \gamma'(0)$ and $V'(0) = (J_2)'(0) = \gamma'(0)$, so we may choose $V(s) = (1+s)\gamma'(0)$. Thus,

$$\tilde{\gamma}(t,s) = \exp_{\sigma(s)}(tV(s)) = \exp_{\gamma(0)}(t(1+s)\gamma'(0)) = \gamma((1+s)t).$$

These are examples of Jacobi fields that are everywhere tangent to γ. As we shall discuss below, the space of such Jacobi fields is 2-dimensional and spanned by J_1 and J_2. Every curve in a variation defined by such a Jacobi field is simply the original curve in a different parametrization, hence there is no new insight to be gained.

We are now interested in Jacobi fields tangential or normal to the velocity vector at each point. For simplicity, let us introduce the following notation: For a curve $\gamma : I \to M$ with nowhere vanishing γ', we write $T^{\top}_{\gamma(t)}M := \operatorname{span}\{\gamma'(t)\} \subseteq T_{\gamma(t)}M$ and let $T^{\perp}_{\gamma(t)}M \subseteq T_{\gamma(t)}M$ denote its orthogonal complement. The notions *tangential* and *normal* for vector fields along γ are defined with respect to these subspaces of the tangent space. We will write $\mathfrak{X}^{\top}(\gamma)$ and $\mathfrak{X}^{\perp}(\gamma)$ for tangential and normal vector fields along γ, respectively. Moreover, if γ is a geodesic, we denote the tangential and normal Jacobi fields along γ by $\mathfrak{Jac}^{\top}(\gamma)$ and $\mathfrak{Jac}^{\perp}(\gamma)$.
The following result gives several characterizations of normal Jacobi fields.

Proposition 2.1.11. (Characterizing normal Jacobi fields)
Let $\gamma : I \to M$ be a geodesic and let $J \in \mathfrak{Jac}(\gamma)$. Then the following are equivalent:

(1) $J \in \mathfrak{Jac}^{\perp}(\gamma)$.

(2) There are $t_0 \neq t_1 \in I$ such that $J(t_0) \perp \gamma'(t_0)$ and $J(t_1) \perp \gamma'(t_1)$.

(3) There is $t \in I$ such that $J(t) \perp \gamma'(t)$ and $J'(t) \perp \gamma'(t)$.

(4) $J, J' \in \mathfrak{X}^{\perp}(\gamma)$.

Proof. Let $f : I \to \mathbb{R}$, $f(t) := \langle J(t), \gamma'(t)\rangle$. Then by the Jacobi equation and using that $\gamma'' = 0$, we get

$$f''(t) = \langle J''(t), \gamma'(t)\rangle = -\langle R(J(t), \gamma'(t))\gamma'(t), \gamma'(t)\rangle = 0,$$

where the vanishing is due to the symmetries of the curvature tensor. Hence, $f(t) = at + b$. Note that $f'(t) = \langle J'(t), \gamma'(t)\rangle = a$ is a constant.
Let us prove the equivalence of the claims above: Trivially (4) implies (1) and (1) implies (2). Since an affine function with two zeros is identically zero (and so is its derivative), (2) implies (3), while (3) implies that f above is the zero function, yielding (4) as a consequence. $\square$

Corollary 2.1.12. (Orthogonal decomposition of Jacobi fields)
Let $\gamma : I \to M$ be a nonconstant, nonnull geodesic. Then $\mathfrak{Jac}^{\perp}(\gamma) \subseteq \mathfrak{Jac}(\gamma)$ is a $(2n-2)$-dimensional vector subspace of the $2n$-dimensional space of Jacobi fields along γ, and $\mathfrak{Jac}^{\top}(\gamma) \subseteq \mathfrak{Jac}(\gamma)$ is a 2-dimensional vector subspace. Moreover, every Jacobi field $J \in \mathfrak{Jac}(\gamma)$ has a unique decomposition of the form $J = J^{\top} + J^{\perp}$ with $J^{\top} \in \mathfrak{Jac}^{\top}(\gamma)$ and $J^{\perp} \in \mathfrak{Jac}^{\perp}(\gamma)$.

Proof. Observe that by Corollary 2.1.8 for $t \in I$ we have that $J \mapsto (J(t), J'(t))$ is an isomorphism $\mathfrak{Jac}(\gamma) \to T_{\gamma(t)}M \oplus T_{\gamma(t)}M$. By Proposition 2.1.11(3), $\mathfrak{Jac}^{\perp}(\gamma)$ is thus isomorphic to the $(2n-2)$-dimensional subspace $T^{\perp}_{\gamma(t)}M \oplus T^{\perp}_{\gamma(t)}M$ of $T_{\gamma(t)}M \oplus T_{\gamma(t)}M$. (Note that, up to here, we did not need γ to be nonnull.)
By Example 2.1.10, $J^{\top}(\gamma)$ contains $\gamma'(t)$ and $t\gamma'(t)$, so its dimension is at least 2. Since γ is nonnull, $\mathfrak{Jac}^{\top}(\gamma) \cap \mathfrak{Jac}^{\perp}(\gamma) = \{0\}$, hence the dimension of $\mathfrak{Jac}^{\top}(\gamma)$ must be 2. The claims about the direct sum decomposition now follow from linear algebra. $\square$

Remark 2.1.13. (The case of null geodesics in 2.1.12)
If γ is null, then γ' is both a tangential and a normal Jacobi field, hence the claim in Corollary 2.1.12 about the direct sum decomposition does not go through. However, even if γ is null, the dimension of $\mathfrak{Jac}^{\top}(\gamma)$ can still be seen to be exactly 2 and not higher: If J is a tangential Jacobi field, then $J(t) = f(t)\gamma'(t)$ for some smooth function f on the interval I. Then $J''(t) = f''(t)\gamma'(t)$ and the Jacobi equation (using the symmetries of the curvature tensor) implies that $J''(t) = 0$, and since γ is assumed to be nonconstant, this means $f'' = 0$, so $f(t) = at + b$, proving that $J_1(t) = \gamma'(t)$ and $J_2(t) = t\gamma'(t)$ span $\mathfrak{Jac}^{\top}(\gamma)$.

We will now consider Jacobi fields that vanish at a point. This will become important later in the context of conjugate points.

Proposition 2.1.14. (Jacobi fields vanishing at a point)
Let $\gamma : I \to M$ be a geodesic, $0 \in I$, and let $p = \gamma(0)$, $v = \gamma'(0)$. For any $w \in T_pM$, the (unique) Jacobi field $J \in \mathfrak{Jac}(\gamma)$ with $J(0) = 0$, $J'(0) = w$ is given by

$$J(t) = T_{tv} \exp_p(tw).$$

The corresponding geodesic variation is given by

$$\tilde{\gamma}(t,s) = \exp_p(t(v + sw)).$$

Proof. We begin by proving the latter formula, which we will then use to derive the former. Suppose therefore that $J \in \mathfrak{Jac}(\gamma)$ is the unique Jacobi field with $J(0) = 0$, $J'(0) = w$. Then by the proof of Proposition 2.1.9, the corresponding geodesic variation is given by

$$\tilde{\gamma}(t, s) = \exp_{\sigma(s)}(tV(s)),$$

where σ is any smooth curve satisfying $\sigma(0) = p$, $\sigma'(0) = J(0) = 0$, and V is a smooth vector field along γ satisfying $V(0) = v$, $V'(0) = J'(0) = w$. Here, the constant curve $\sigma(s) \equiv p$ meets these conditions and V can be chosen as $V(s) = v + sw$. Thus, the formula for $\tilde{\gamma}$ is proven.
Since we know that J is the variation field of $\tilde{\gamma}$, we may use the chain rule to derive its explicit form:

$$J(t) = \left.\frac{\partial \tilde{\gamma}(t, s)}{\partial s}\right|_{s=0} = T_{tv} \exp_p(tw).$$

□

Remark 2.1.15. (Jacobi fields vanishing at a point in RNC)
If $\phi = (x^j)$ are normal coordinates, then the formula in Proposition 2.1.14 for the variation reads

$$\phi \circ \tilde{\gamma}(t, s) = (t(v^1 + sw^1), \ldots, t(v^n + sw^n)),$$

because the exponential map in normal coordinates is the identity. Hence, the Jacobi field J takes the form

$$J(t) = tw^j \left.\frac{\partial}{\partial x^j}\right|_{\gamma(t)}.$$

This also shows that, in a normal neighborhood U of p, for any point $q \in U \setminus \{p\}$, any vector $w \in T_qM$ can be obtained as the value of a Jacobi field along a radial geodesic from p to q in U vanishing at p.

2.2 Variational formulas

In this section, we want to collect a number of results on derivatives of length and energy functionals that will be of great use in this thesis. Throughout this section, let (M, g) be a semi-Riemannian manifold of dimension $n \geq 2$. Denote by $\mathfrak{X}_{\mathrm{pw}}(\gamma)$ the space of piecewise smooth vector fields along the curve γ, and denote by $\mathfrak{X}^0_{\mathrm{pw}}(\gamma)$ those piecewise smooth vector fields along γ that vanish at the endpoints (if γ is defined on a compact interval). As we noted before, we understand derivatives of piecewise smooth objects to be defined everywhere via left limits.

Proposition 2.2.1. (First variation of arclength)
Let $\alpha : [a,b] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth variation of a nowhere null curve $c : [a,b] \to M$. Then the first derivative of the arclength function $s \mapsto L(t \mapsto \alpha(t,s))$ is

$$L'(s) = \sigma \int_a^b \frac{\langle \partial_t \alpha, \nabla_s \partial_t \alpha \rangle}{|\partial_t \alpha|} dt,$$

where $|v| = \sqrt{|g(v,v)|}$ and $\sigma = \mathrm{sgn}(c', c')$. In particular, since $c = \alpha(.,0)$, we have

$$L'(0) = \sigma \int_a^b \frac{\langle c', V' \rangle}{|c'|} dt,$$

where $V = \partial_s \alpha|_{s=0}$ is the variation field of α.

Proof. Note that, by continuity, we may assume that
$\sigma = \mathrm{sgn}\langle \partial_t \alpha(t,s), \partial_t \alpha(t,s) \rangle$ for any s (see [1, Lem. 10.7] for details). The formulas follow from simple calculations:

$$\begin{aligned} L'(s) &= \frac{d}{ds} \int_a^b \sqrt{\sigma \langle \partial_t \alpha, \partial_t \alpha \rangle} dt = \int_a^b \frac{1}{2} (\sigma \langle \partial_t \alpha, \partial_t \alpha \rangle)^{-1/2} \cdot 2\sigma \langle \partial_t \alpha, \nabla_s \partial_t \alpha \rangle dt \\ &= \sigma \int_a^b \frac{\langle \partial_t \alpha, \nabla_s \partial_t \alpha \rangle}{|\partial_t \alpha|}. \end{aligned}$$

Thus the general formula is proven. Since $\nabla_s \partial_t \alpha = \nabla_t \partial_s \alpha$, the formula for $L'(0)$ also follows. $\square$

Remark 2.2.2. (Alternative formulas for $L'(0)$)
Consider the same situation as above and suppose $|c'|$ is constant (e.g. if c is a geodesic). Then

$$L'(0) = \frac{\sigma}{|c'|} \int_a^b \langle c', V' \rangle dt.$$

Let now $a = t_0 < t_1 < \cdots < t_k = b$ be a subdivision of $[a,b]$ such that c is smooth on the subintervals, and $t_1, \ldots, t_{k+1}$ are precisely the points of discontinuity of c'. Then

$$\int_{t_i}^{t_{i+1}} \langle c', V' \rangle = \langle c', V \rangle|_{t_i}^{t_{i+1}} - \int_{t_i}^{t_{i+1}} \langle c'', V \rangle.$$

Adding from $i = 0$ to k, we get

$$\int_a^b \langle c', V' \rangle dt = \langle c', V \rangle|_a^b - \sum_{i=1}^{k-1} \langle \delta c'(t_i), V(t_i) \rangle - \int_a^b \langle c'', V \rangle,$$

where $\delta c'(t_i) = \lim_{t\downarrow t_i} c'(t) - \lim_{t\uparrow t_i} c'(t)$ because c was assumed to only be piecewise smooth. (Note that V is everywhere continuous). Thus

$$L'(0) = \frac{\sigma}{|c'|}\left(-\int_a^b \langle c'', V\rangle dt - \sum_{i=1}^{k-1}\langle \delta c'(t_i), V(t_i)\rangle + \langle c', V\rangle|_a^b\right).$$

Let us now consider the special case that (M, g) is Lorentzian, $c : [a, b] \to M$ is a piecewise smooth unit-speed timelike curve and $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ is a piecewise smooth variation of c by timelike curves. In this case, $\sigma = -1$, $|c'| = 1$, and we may combine the boundary terms to get

$$L'(0) = \int_a^b \langle c'', V\rangle dt + \sum_{i=0}^{k}\langle \delta c'(t_i), V(t_i)\rangle, \tag{2.1}$$

where $\delta c'(a) := c'(a)$ and $\delta c'(b) := -c'(b)$. If we specialize even further and assume that c is a timelike unit-speed geodesic, then we get

$$L'(0) = -\langle V, c'\rangle|_a^b. \tag{2.2}$$

Proposition 2.2.3. (Geodesics are extrema of L)
Let $c : [a, b] \to M$ be a piecewise smooth curve of constant speed $|c'| \neq 0$. Then c is a (smooth) geodesic if and only if $L'(0) = 0$ for any piecewise smooth FEP-variation $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$.

Proof. If c is a geodesic, then $L'(0) = 0$ follows immediately from the formula derived in Remark 2.2.2, since $c'' = 0$, $\delta c(t_i) = 0$, and $V(a) = 0$, $V(b) = 0$ if the variation fixes the endpoints.
Conversely, suppose $L'(0) = 0$ for any piecewise smooth FEP-variation of c. Let $a = t_0 < \cdots < t_k = b$ be a subdivision of $[a, b]$ such that c is smooth on each subinterval. We first show that each $c|_{[t_i,t_{i+1}]}$ is a geodesic. To this end, fix $t \in [t_i, t_{i+1}]$ and let $f : [a, b] \to [0, 1]$ be a bump function with $f(t) = 1$ and $\mathrm{supp}(f) \subseteq [t - \delta, t + \delta] \subsetneq [t_i, t_{i+1}]$. Let $y \in T_{c(t)}M$ arbitrary, and let $Y \in \mathfrak{X}(c)$ be the parallel translate of y along c. Set $V := fY$. By construction $V(a) = 0$ and $V(b) = 0$, hence by Lemma 2.1.2,

$$\alpha(t, s) := \exp_{c(t)}(sV(t))$$

is an FEP-variation of c with variation field V. By assumption $L'(0) = 0$, and moreover $V(t) = 0$ outside of $[t-\delta, t+\delta]$. Hence, the formula in Remark 2.2.2 reduces to

$$0 = \int_a^b \langle c'', V\rangle dt = \int_{t-\delta}^{t+\delta} \langle c'', fY\rangle dt.$$

Taking $\delta \to 0$ gives $\langle c''(t), y\rangle = 0$, but $y \in T_{c(t)}M$ was arbitrary, hence $c''(t) = 0$. This shows that $c|_{[t_i,t_{i+1}]}$ is a geodesic.

It remains to show that there are no break points, i.e. $\delta c'(t_i) = 0$ for every i. For this, suppose $y \in T_{c(t_i)}M$ is arbitrary, let Y be its parallel translate, and let $f : [a,b] \to [0,1]$ be a bump function with $f(t_i) = 1$ and with support contained in $[t_i - \delta, t_i + \delta] \subsetneq [t_{i-1}, t_{i+1}]$. Consider again the formula derived in Remark 2.2.2, the first term vanishes since we have shown that $c''(t) = 0$ almost everywhere, and the last term vanishes since V is the variation field of an FEP-variation. Using $L'(0) = 0$, what remains is

$$0 = \langle \delta c'(t_i), y \rangle.$$

Since y was arbitrary, it follows that $\delta c'(t_i) = 0$, which shows that c is everywhere smooth. $\square$

Definition 2.2.4. (Energy function and energy functional)
Let $\gamma : [a,b] \to M$ be a smooth curve in a semi-Riemannian manifold (M, g). Then the *energy function* of γ is the smooth map $E_\gamma : [a,b] \to \mathbb{R}$ defined by

$$E_\gamma(t) := \frac{1}{2} \int_a^t |g(\gamma'(t), \gamma'(t))| dt.$$

Moreover, we refer to $E_\gamma(b) =: E(\gamma)$ as the *energy* of γ.

Lemma 2.2.5. (Energy and arclength)
Let $\gamma : [a,b] \to M$ be a smooth curve. Then

$$L(\gamma)^2 \leq 2(b-a)E(\gamma)$$

and equality holds if and only if $g(\gamma', \gamma')$ is constant.

Proof. By the Cauchy-Schwarz inequality,

$$\begin{aligned} L(\gamma)^2 = \left(\int_a^b 1 \cdot \sqrt{|g(\gamma'(s), \gamma'(s))|} ds \right) &\leq \int_a^b 1^2 ds \cdot \int_a^b |g(\gamma'(s), \gamma'(s))| ds \\ &= 2(b-a)E(\gamma). \end{aligned}$$

Again by the Cauchy-Schwarz inequality, equality holds above if and only if $\sqrt{|g(\gamma', \gamma')|}$ is proportional to 1, i.e. if and only if $g(\gamma', \gamma')$ is constant. $\square$

Proposition 2.2.6. (First variation of energy)
Let $\alpha : [a,b] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth variation with constant (in s) $\sigma = \operatorname{sgn}(g(\partial_t \alpha, \partial_t \alpha))$ in a semi-Riemannian manifold (M, g). Then the first derivative of the energy function $s \mapsto E(t \mapsto \alpha(t,s))$ is given by

$$\begin{aligned} E'(s) &= \sigma \int_a^b \langle \nabla_s \partial_t \alpha, \partial_t \alpha \rangle dt \\ &= \sigma \left(\langle \partial_s \alpha, \partial_t \alpha \rangle_a^b - \int_a^b \langle \partial_s \alpha, \nabla_t \partial_t \alpha \rangle dt \right). \end{aligned}$$

In particular, $E'(0) = 0$ if α is an FEP-variation and $\alpha(., 0)$ is a geodesic.

Proof. This is a straightforward calculation using Lemma 2.1.3 and integration by parts. ☐

Proposition 2.2.7. (Second variation of arc length)
Let (M, g) be a spacetime and let $c : [a, b] \to M$ be a g-unit speed timelike geodesic. Let $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth variation of c by timelike curves with variational vector field V. Let $N(t) := V(t) + \langle V(t), c'(t)\rangle c'(t)$, then $L(s) := L(t \mapsto \alpha(t, s))$ has the following second derivative at $s = 0$:

$$L''(0) = \int_a^b \langle N'' + R(V, c')c', N\rangle dt + \sum_{i=0}^{k} \langle N(t_i), \delta N'(t_i)\rangle - \langle \nabla_s \partial_s \alpha|_{(t,0)}, c'\rangle|_a^b,$$

where the $t_i \in [a, b]$ are the breakpoints of α, with $t_0 = a$ and $t_0 = b$.

Proof. This follows by a lengthy but elementary calculation, see [1, Prop. 12.26]. ☐

Proposition 2.2.8. (Second variation of energy)
Let $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth variation with constant (in s) $\sigma = \operatorname{sgn}\langle \partial_t \alpha, \partial_t \alpha\rangle$ in a semi-Riemannian manifold (M, g). Suppose further that $\alpha(., 0)$ is a geodesic. Then the second derivative of the energy $E(s) = E(\alpha(., s))$ at $s = 0$ is given by

$$\begin{aligned} E''(0) = \sigma \Bigg(&\langle \nabla_s \partial_s \alpha, \partial_t \alpha\rangle_a^b - \int_a^b \langle R(\partial_s \alpha, \partial_t \alpha)\partial_t \alpha, \partial_s \alpha\rangle dt \\ &+ \int_a^b \langle \nabla_t \partial_s \alpha, \nabla_t \partial_s \alpha\rangle \Bigg), \end{aligned}$$

where the right hand side is understood to be evaluated at $s = 0$.

Proof. This is elementary using Lemma 2.1.4 and integration by parts, see [15, Lem. 3.1.2]. ☐

Next, we are interested in the following scenario: Suppose (M, g) is a spacetime of dimension $n \geq 2$, and suppose $c : [a, b] \to M$ is a timelike curve with $c(a) \in H$, where H is a spacelike hypersurface of M. In our study of focal points, see Subsection 3.1.4, we will be interested in variations $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ of c by timelike curves such that all $\alpha(., s)$ start in H and end in $c(b)$. For these variations, the variational vector field V satisfies $V(a) \in T_{c(a)}H$ and $V(b) = 0$. In this case, the first variation formula (2.1) reduces to

$$L'(0) = \int_a^b \langle V, c''\rangle dt + \sum_{i=1}^{k-1} \langle V(t_i), \delta c'(t_i)\rangle + \langle V(a), c'(a)\rangle.$$

Now fix $q \in M \setminus H$, and consider all timelike curves from H to q. If there is a longest curve of this type, then it must be a geodesic (because it also locally maximizes arc length, hence it is a piecewise geodesic, and since in any FEP-variation of a timelike curve, all close enough curves are timelike, the result follows from Proposition 2.2.3). Suppose $c : [a, b] \to M$ is a unit-speed geodesic from H to q of maximal length, and let $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ be a variation of c by timelike curves such that $\alpha(a, s) \in H$ and $\alpha(b, s) = c(b) = q$ for every s. Since $V(b) = 0$, it follows from (2.2) that

$$L'(0) = \langle V(a), c'(a) \rangle.$$

Since c maximizes L, it follows that $L'(0) = 0$, hence $V(a) \perp c'(a)$. Since such variations may be defined for any initial tangent direction $v \in T_{c(a)}H$, it follows that $c'(a) \perp T_{c(a)}H$. We have hence proven the following result.

Corollary 2.2.9. (Maximizing curves to H are orthogonal geodesics)
Let H be a spacelike hypersurface in a spacetime (M, g). If $c : [a, b] \to M$ is a future timelike curve from H to $q \in I^+(H)$ of maximal length (i.e. $L(c) = d(H, q)$), then c is a timelike geodesic orthogonal to H at $c(a)$.

Proposition 2.2.10. (Second variation of arc length: HSF variations)
Let H be a spacelike hypersurface in a spacetime (M, g) and let $c : [a, b] \to M$ be a g-unit timelike geodesic starting orthogonal to H. If $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ is a piecewise smooth variation of c by timelike curves with variational vector field V, and $N := V + \langle V, c' \rangle c'$, then

$$\begin{aligned} L''(0) = &\int_a^b \langle N'' + R(V, c')c', N \rangle dt + \sum_{i=1}^{k-1} \langle N(t_i), \delta N'(t_i) \rangle + \langle N(a), N'(a) \rangle \\ &+ \langle W_{c'(a)}(N(a)), N(a) \rangle, \end{aligned}$$

where t_i are the breakpoints of α and the Weingarten map $W_{c'(a)}$ of H at $c(a)$ is well-defined because $c'(a) \perp T_{c(a)}H$.

Proof. In light of Proposition 2.2.7, the result is proven once

$$-\langle \nabla_s \partial_s \alpha|_{(t,0)}, c' \rangle|_a^b = \langle W_{c'(a)}(N(a)), N(a) \rangle$$

is shown. See [1, Cor. 12.27] for details. □

2.3 Timelike index theory

In this section, we will study conjugate points along timelike geodesics. They tell us whether or not a given timelike geodesic is maximizing or not. The treatment of the timelike case is analogous to that of conjugate points along geodesics in Riemannian manifolds, and tools like Jacobi fields and index forms will be used. At the end of the section, we shall give a proof of the Lorentzian Morse Index Theorem in the timelike case, which relates the index of a timelike geodesic with a certain sum involving dimensions of spaces of Jacobi fields. This is proven in complete analogy to the Riemannian case, for which we refer to [10]. Our exposition follows [1, Ch. 10]. The part about Jacobi tensors along timelike geodesics can be found in [1, Sec. 12.1].

Throughout this section, we fix a spacetime (M, g) of dimension ≥ 2. Unless stated otherwise, all timelike geodesics are assumed to be parametrized by g-arc length. Given a curve $\gamma : [a, b] \to M$, we denote by $\mathfrak{X}_{\mathrm{pw}}(\gamma)$ the piecewise smooth vector fields along γ, $\mathfrak{X}^0_{\mathrm{pw}}(\gamma)$ are those piecewise smooth vector fields that vanish at the endpoints; similarly, we write $\mathfrak{X}^{\perp}_{\mathrm{pw}}(\gamma)$ and $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(\gamma)$ for vector fields in the aforementioned two classes that are everywhere perpendicular to γ. For piecewise smooth objects, we understand their derivatives to be defined everywhere (at breakpoints, we understand the derivative to be a left limit).

2.3.1 The timelike index form and conjugate points along timelike geodesics

Definition 2.3.1. (Conjugate points along timelike geodesics)
Let $c : [a, b] \to M$ be a timelike geodesic. Then $c(t)$, $t \in (a, b]$, is said to be *conjugate to* $c(a)$ *along* c if there is a nontrivial Jacobi field $J \in \mathfrak{Jac}(c)$ with $J(a) = 0$, $J(t) = 0$. Similarly, the notion of $c(t_1)$, $c(t_2)$ being conjugate along c (with $t_1, t_2 \in [a, b]$) is defined.

Note that such a Jacobi field as well as its derivative are necessarily orthogonal to c by Proposition 2.1.11.

Proposition 2.3.2. (Conjugate points and exp)
Let $c : [a, b] \to M$ be a timelike geodesic. For $c(t_0)$ and $c(t_1)$ the following are equivalent:

(1) $c(t_0)$ and $c(t_1)$ are not conjugate along c.

(2) For each $v \in T_{c(t_0)}M$ and each $w \in T_{c(t_1)}M$, there is a unique Jacobi field $J \in \mathfrak{Jac}(c)$ with $J(t_0) = v$ and $J(t_1) = w$.

(3) $\exp_{c(t_0)}$ is a local diffeomorphism around $(t_1 - t_0)c'(t_0)$.

Proof. (1) $\Leftrightarrow$ (2) is immediate since $\mathfrak{Jac}(c)$ is $2n$-dimensional, and $c(t_0)$ and $c(t_1)$ are not conjugate if and only if $\mathfrak{Jac}(c) \to T_{c(t_0)}M \times T_{c(t_1)}M, J \mapsto (J(t_0), J(t_1))$ is injective, hence an isomorphism.
(1) $\Leftrightarrow$ (3): Let

$$\psi : T_{(t_1-t_0)c'(t_0)}T_{c(t_0)}M \to T_{c(t_1)}M,\ \psi(w) := T_{(t_1-t_0)c'(t_0)} \exp_{c(t_0)}(w).$$

Now (1) $\Leftrightarrow$ if $J \in \mathfrak{Jac}(c)$ satisfies $\mathfrak{Jac}(t_0) = 0$ and $J(t_1) = 0$, then $J = 0$ (in particular $J'(t_0) = 0$). A slight adaptation of Proposition 2.1.14 shows that the unique Jacobi field J with $J(t_0) = 0$ and $J'(t_0) = w$ is given by

$$J(t) = T_{(t-t_0)c'(t_0)} \exp_{c(t_0)}((t - t_0)w).$$

Given J like this, $J = 0$ is equivalent to $0 = J'(t_0) = w$, which in turn is equivalent to ψ being bijective, which is precisely the statement of (3). $\square$

Definition 2.3.3. (Timelike index form)
Let $c : [a, b] \to M$ be a timelike geodesic. The symmetric bilinear form $I : \mathfrak{X}^{\perp}_{\text{pw}}(c) \times \mathfrak{X}^{\perp}_{\text{pw}}(c) \to \mathbb{R}$ defined by

$$I(X, Y) := -\int_a^b \big(\langle X', Y'\rangle - \langle R(X, c')c', Y\rangle\big)\, dt$$

is called the *timelike index form along c.*

Remark 2.3.4. (On the timelike index form)
If $X \in \mathfrak{X}^{\perp}_{\text{pw}}(c)$ is indeed smooth, we may integrate the first term in the index form by parts to obtain the alternative formula

$$I(X, Y) = -\langle X', Y\rangle|_a^b + \int_a^b \langle X'' + R(X, c')c', Y\rangle dt$$

for any $Y \in \mathfrak{X}^{\perp}_{\text{pw}}(c)$. In particular, if $X \in \mathfrak{Jac}(c)$, then

$$I(X, Y) = -\langle X', Y\rangle|_a^b$$

for all $Y \in \mathfrak{X}^{\perp}_{\text{pw}}(c)$. Hence $I(X, Y) = 0$ for all $X \in \mathfrak{Jac}(c)$ and $Y \in \mathfrak{X}^{0,\perp}_{\text{pw}}(c)$. More generally, one may apply the above idea of integration by parts on subintervals for piecewise smooth vector fields to obtain

$$I(X, Y) = \sum_{i=0}^{k} \langle \delta X'(t_i), Y\rangle + \int_a^b \langle X'' + R(X, c')c', Y\rangle dt$$

for all $X, Y \in \mathfrak{X}^{\perp}_{\text{pw}}(c)$, where $a = t_0 < t_1 < \cdots < t_k = b$ is a subdivision of $[a, b]$ such that X is smooth on those subintervals, and

$$\delta X'(t_i) = \lim_{t \downarrow t_i} X'(t) - \lim_{t \uparrow t_i} X'(t),$$

noting that the "$t \downarrow b$"- and "$t \uparrow a$"-terms are 0 by convention.

In the following, we will be interested in piecewise smooth variations of timelike geodesics. Note that, by restricting the transversal parameter, any such variation can be assumed to be a variation consisting only of timelike curves by continuity (for a detailed argument, see [1, Lem. 10.7]).

Recall that any $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$, $c : [a,b] \to M$ a timelike geodesic, is the variation vector field of an FEP-variation, see Lemma 2.1.2. In the proof, we explicitly constructed one such variation, namely

$$\alpha(t,s) := \exp_{c(t)}(sY(t)).$$

We will refer to this variation as the *canonical FEP-variation* of c associated with Y.

Remark 2.3.5. (Second variation and the index form)
Let $c : [a,b] \to M$ be a timelike geodesic and let $\alpha : [a,b] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth FEP-variation by timelike curves. Looking back at the results from Section 2.2, we see that for $L(s) = L(\alpha(.,s))$, we have

$$L'(0) = 0, \quad L''(0) = I(Y,Y),$$

where $Y \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$ is the variation vector field of α. The idea here is to use basic calculus: Since $L'(0) = 0$, we have to rely on $L''(0)$ to tell us whether the Lorentzian arc length of c is maximal (among the neighboring curves $\alpha(.,s)$) or not. Hence if $I(Y,Y) > 0$, then the canonical FEP-variation of c with variation vector field Y provides longer curves. We conclude that if the timelike geodesic c is maximizing, then its index form is necessarily negative semidefinite.

We now wish to establish a connection between conjugate points and non-maximality of a given timelike geodesic.

Proposition 2.3.6. (Conjugate points and maximality)
Let $c : [a,b] \to M$ be a timelike geodesic, and suppose there is $t_0 \in (a,b)$ such that $c(t_0)$ is conjugate to $c(a)$ along c. Then there exists a piecewise smooth FEP-variation $\alpha : [a,b] \times (-\varepsilon, \varepsilon) \to M$ by timelike curves such that $L(\alpha(.,s)) > L(c)$ for all $s \neq 0$. In particular, c is not maximizing.

Proof. By what has been said up to now (see Remark 2.3.5), we only need to find a vector field $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ with $I(Y,Y) > 0$. For this, let $J \in \mathfrak{Jac}(c)$ be a nontrivial Jacobi field with $J(a) = 0$ and $J(t_0) = 0$. Then $J, J' \in \mathfrak{X}^{0,\perp}(c)$. Note that necessarily $J'(t_0) \neq 0$ since $J \neq 0$, and it is a spacelike vector since $J'(t_0) \perp c'(t_0)$. Denote by I^t_s the restriction of the index form to the interval $[s,t] \subseteq [a,b]$, i.e. the integration is only carried out over $[s,t]$. As we noted in Remark 2.3.4 (and this of course continues to hold true for the restricted index form), we have

$$I^s_a(J,Z) = -\langle J', Z\rangle|^s_a$$

for all $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$.

Fix now a smooth function $f : [a, b] \to \mathbb{R}$ with $f(a) = f(b) = 0$ and $f(t_0) = 1$. Let $\tilde{Z} \in \mathfrak{X}(c)$ be the parallel translate of $-J'(t_0)$, and let $Z := f\tilde{Z} \in \mathfrak{X}^{0,\perp}(c)$. Define now for $\varepsilon < 1$

$$Y_\varepsilon(t) := \begin{cases} J(t) + \varepsilon Z(t), & a \le t \le t_0 \\ \varepsilon Z(t), & t_0 \le t \le b. \end{cases}$$

Then $Y_\varepsilon \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ and, using the above-mentioned simplified formula for I_s in the case of a Jacobi field, we calculate as follows:

$$\begin{aligned} I(Y_\varepsilon, Y_\varepsilon) &= I_a^{t_0}(Y_\varepsilon, Y_\varepsilon) + I_{t_0}^b(Y_\varepsilon, Y_\varepsilon) \\ &= I_a^{t_0}(J, J) + 2\varepsilon I_a^{t_0}(J, Z) + \varepsilon^2 I_a^{t_0}(Z, Z) + \varepsilon^2 I_{t_0}^b(Z, Z) \\ &= -\langle J', J\rangle|_a^{t_0} - 2\varepsilon\langle J', Z\rangle|_a^{t_0} + \varepsilon^2 I(Z, Z). \end{aligned}$$

Since $J(a) = 0$, $J(t_0) = 0$ and $Z(a) = 0$, we get

$$\begin{aligned} I(Y_\varepsilon, Y_\varepsilon) &= -2\varepsilon\langle J'(t_0), Z(t_0)\rangle + \varepsilon^2 I(Z, Z) \\ &= 2\varepsilon \underbrace{\langle J'(t_0), J'(t_0)\rangle}_{>0} + \varepsilon^2 I(Z, Z) \end{aligned}$$

and thus we see that $I(Y_\varepsilon, Y_\varepsilon) > 0$ if ε is chosen small enough. $\square$

We saw in Remark 2.3.4 that if $J \in \mathfrak{Jac}(c)$ for a timelike geodesic $c : [a, b] \to M$, then $I(J, Y) = 0$ for all $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$. The next result shows that this property is characteristic for Jacobi fields.

Proposition 2.3.7. (Jacobi fields and the index form)
Let $c : [a, b] \to M$ be a timelike geodesic. A vector field $J \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ is a Jacobi field if and only if $I(J, Y) = 0$ for all $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$.

Proof. Since we already discussed the easy direction (see Remark 2.3.4), it remains to prove the converse. Suppose $I(J, Y) = 0$ for all $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$. We have to show that J is smooth and satisfies the Jacobi equation.
Note that we have

$$\langle J', c'\rangle = 0, \quad \langle J'' + R(J', c')c', c'\rangle = 0$$

everywhere except at the finitely many points $t_1, \ldots, t_{k-1}$ where J is not smooth, $a = t_0 < t_1 < \cdots < t_k = b$. This follows trivially from differentiating the equation $\langle J, c'\rangle = 0$ and the fact that $\langle R(J, c')c', c'\rangle = 0$ by the symmetries of the curvature tensor. But then these equalities (trivially) extend to the points t_i by taking left and right-handed limits. So the difference vectors $\delta J'(t_i)$ (see Remark 2.3.4) are orthogonal to $c'(t_i)$. By the same reference, we may write the index form in the following way: For $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$

arbitrary,

$$I(J,Z) = \sum_{i=0}^{k} \langle \delta J'(t_i), Z(t_i) \rangle + \int_a^b \langle J'' + R(J,c')c', Z \rangle dt.$$

Now let $f : [a,b] \to [0,1]$ be a smooth function vanishing at all t_i, $i = 0, \ldots, k$, and $f(t) > 0$ everywhere else. Then $\tilde{Z} := f(J'' + R(J,c')c') \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ and $\tilde{Z}(t_i) = 0$ for all i. Thus the first term in the above formula vanishes, and using the assumption we obtain that

$$0 = I(J,\tilde{Z}) = \int_a^b f(t) \| J'' + R(J,c')c' \|^2 dt,$$

noting that the integrand is nonnegative since J'' and $R(J,c')c'$ are orthogonal to c' (and are hence spacelike). The only way this integral vanishes is if the integrand vanishes, hence

$$J'' + R(J,c')c' = 0$$

for all $t \in [a,b]$ except possibly for $t = t_i$. We see that J satisfies the Jacobi equation everywhere except maybe at the points of nonsmoothness, so we are done once we show that J is smooth. For this, it is sufficient to show that $\delta J'(t_i) = 0$ for all i.
Note that in light of $J'' + R(J,c')c' = 0$ almost everywhere, the index form simplifies to

$$I(J,Z) = \sum_{i=0}^{k} \langle \delta J'(t_i), Z(t_i) \rangle$$

for all $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$. Let $\tilde{Z}_i$ be the parallel translate of $\delta J'(t_i)$ along c. Since $\delta J(t_i) \perp c'(t_i)$, also $\tilde{Z}_i(t) \perp c'(t)$ for all $t \in [a,b]$ and thus $Z_i := h\tilde{Z}_i \in \mathfrak{X}^{0,\perp}(c)$, where $h : [a,b] \to [0,1]$ is a smooth function with $h(0) = 0 = h(1)$, $h(t_i) = 1$, $h(t_j) = 0$ for $j \neq i$, and $h(t) > 0$ elsewhere. Using our assumption and the above formula for the index form, we get

$$0 = I(J, Z_i) = \|\delta J'(t_i)\|^2$$

which gives $\delta J'(t_i) = 0$ since it is spacelike. Doing this for all $i = 1, \ldots, k-1$ shows that J' has no discontinuities, and thus J is smooth, which concludes the proof. □

We showed in Proposition 2.3.6 that if c has conjugate points, then we can produce an FEP-variation with longer neighboring curves. We now show the converse statement: If c has no conjugate points, then it is the longest curve among close enough neighboring curves.

Proposition 2.3.8. (Conjugate points and maximality II)
Let $c : [a, b] \to M$ be a timelike geodesic such that no $c(t)$ is conjugate to $c(a)$ along c. Then for any piecewise smooth FEP-variation $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ there is a $\delta > 0$ such that the neighboring curves $\alpha(., s)$ are all timelike and satisfy $L(\alpha(., s)) \leq L(c)$ for all s with $|s| < \delta$, with equality holding if and only if $\alpha(., s)$ is a reparametrization of c.

Proof. For simplicity, assume that c is g-unit speed parametrized and defined on an interval $[0, b]$ and α is defined on $[0, b] \times (-\varepsilon, \varepsilon)$ such that all $\alpha(., s)$ are timelike curves. Set $p := c(0)$. Let $\phi : [0, b] \to T_pM$, $\phi(t) := tc'(0)$. Since no $c(t)$ is conjugate to $c(0)$ along c by assumption, it follows from Proposition 2.3.2 that $\exp_p$ is a local diffeomorphism around $\phi(t)$ for each $t \in [0, b]$. Hence by compactness, there is a subdivision $a = t_0 < t_1 < \cdots < t_k = b$ and neighborhoods $U_j \supseteq \phi([t_j, t_{j+1}])$ in T_pM such that $\exp_p$ maps U_j diffeomorphically onto $V_j := \exp_p(U_j) \subseteq M$. By continuity and compactness, there is some $\delta > 0$ such that $\alpha([t_j, t_{j+1}] \times (-\delta, \delta)) \subseteq V_j$ for each $j = 0, \dots, k - 1$.
Now let $\psi(t, s) := \exp_p|_{U_j}^{-1}(\alpha(t, s))$, where j is such that $t \in [t_j, t_{j+1}]$. Then ψ defined on $[0, b] \times (-\delta, \delta)$ is a well-defined piecewise smooth map into T_pM, and $\exp_p \circ \psi = \alpha$. Now for fixed s, $\exp_p \circ \psi(., s) = \alpha(., s)$, hence $L(\alpha(., s)) < L(\alpha(., 0)) = L(c)$ unless $\alpha(., s)$ is a reparametrization of c. The latter is seen as follows: The fact that we have covered c with a finite number of normal neighborhoods U_j such that the variation α stays in U_j for appropriate subintervals means that we can lift α to an FEP-variation on $\bigcup V_j \subseteq T_pM$. Fix $s \in (-\delta, \delta)$, and let $\tilde{\gamma}$ be the lift of $\gamma := \alpha(., s)$. We can use polar coordinates in T_pM to write $\tilde{\gamma}(t) = s(t)v(t)$ with $|v(t)| = -1$, hence $\gamma(t) = \exp_p(s(t)v(t))$. From here, the usual Gauss-lemma arguments show that γ must be shorter that c unless it is a reparametrization of c (see [22, Prop. 5.34]). $\square$

Theorem 2.3.9. (No conjugate points and definiteness of I)
Let $c : [a, b] \to M$ be a (future directed) timelike geodesic. Then the following are equivalent:

(1) No $c(t)$, $t \in (a, b]$, is conjugate to $c(a)$ along c.

(2) The index form $I : \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c) \times \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c) \to \mathbb{R}$ is negative definite.

Proof. (1) $\Rightarrow$ (2): In Remark 2.3.5 we saw that if there is some $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ with $I(Y, Y) > 0$, then the corresponding FEP-variation produces longer curves, which contradicts Proposition 2.3.8. Hence I is negative semidefinite. It remains to argue that if $I(Y, Y) = 0$, then $Y = 0$. For this, let $Z \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ arbitrary, then negative semidefiniteness and the assumption $I(Y, Y) = 0$ give (for any $t \in \mathbb{R}$)

$$0 \geq I(Y - tZ, Y - tZ) = -2tI(Y, Z) + t^2 I(Z, Z).$$

Since this holds for all $t \in \mathbb{R}$, we get that $I(Y, Z) = 0$. Since $Z \in \mathfrak{X}_{\mathrm{pw}}^{0,\perp}(c)$ was arbitrary, we have that Y is a Jacobi field by Proposition 2.3.7. Since $c(b)$ is not conjugate to $c(a)$ along c, it follows that $Y = 0$, hence I is negative definite.
(2) $\Rightarrow$ (1): Suppose $c(t_0)$ is conjugate to $c(a)$ along c. Let $J \in \mathfrak{Jac}(c)$ with $J(a) = 0$ and $J(t_0) = 0$. Define

$$Y(t) := \begin{cases} J(t) & a \leq t \leq t_0, \\ 0 & t_0 \leq t \leq b. \end{cases}$$

Then $Y \in \mathfrak{X}_{\mathrm{pw}}^{0,\perp}(c)$ is nontrivial and $I(Y, Y) = -\langle J', J\rangle|_a^{t_0} = 0$, contradicting negative definiteness of I. □

From this result, we can derive a maximality result (with respect to the index form) for Jacobi fields along timelike geodesics without conjugate points.

Proposition 2.3.10. (Maximality of Jacobi fields)
Let $c : [a, b] \to M$ be a future directed timelike geodesic with no conjugate points to $c(a)$. Let $J \in \mathfrak{Jac}(c)$. Then for any $Y \in \mathfrak{X}_{\mathrm{pw}}^{\perp}(c)$, $Y \neq J$, with $Y(a) = J(a)$ and $Y(b) = J(b)$, we have that

$$I(J, J) > I(Y, Y).$$

Proof. $Z := J - Y \in \mathfrak{X}_{\mathrm{pw}}^{0,\perp}(c)$ and $Z \neq 0$, hence by Theorem 2.3.9 $I(Z, Z) < 0$. Using Remark 2.3.4 and $\langle J', Y\rangle|_a^b = \langle J', J\rangle|_a^b$, we calculate

$$\begin{aligned} 0 > I(J, J) - 2I(J, Y) + I(Y, Y) &= -\langle J', J\rangle|_a^b + 2\langle J', Y\rangle|_a^b + I(Y, Y) \\ &= I(Y, Y) + \langle J', J\rangle|_a^b \\ &= I(Y, Y) - I(J, J). \end{aligned}$$

□

2.3.2 The timelike Morse index theorem

This subsection is devoted to the proof of the timelike Lorentzian Morse Index Theorem. We will follow the proof of the Riemannian case along the lines of [10, p. 150-152], cf. [1, Ch. 10].

Definition 2.3.11. (Index of a timelike geodesic)
Let $c : [a, b] \to M$ be a timelike geodesic. Its *index* $\mathrm{Ind}(c)$ is the supremum over dimensions of subspaces of $\mathfrak{X}_{\mathrm{pw}}^{0,\perp}(c)$ on which the index form I is positive definite, and the *extended index* $\mathrm{Ind}_0(c)$ is defined similarly, considering subspaces on which I is merely positive semidefinite.

Clearly, $\mathrm{Ind}_0(c) \geq \mathrm{Ind}(c)$.

Let us also fix the following notation: If $c : [a,b] \to M$ is a timelike geodesic, we write $\mathfrak{Jac}_t(c)$ for the space of Jacobi fields J along c with $J(a) = 0$ and $J(t) = 0$. Note that $\mathfrak{Jac}_t(c) \subseteq \mathfrak{X}^{\perp}(c)$ for all $t \in (a,b]$ by Proposition 2.1.11.

Proposition 2.3.12. (Index, extended index and Jacobi fields)
Let $c : [a,b] \to M$ be a future directed timelike geodesic. Then $\mathrm{Ind}(c)$ and $\mathrm{Ind}_0(c)$ are finite. Moreover,

$$\mathrm{Ind}_0(c) = \mathrm{Ind}(c) + \dim \mathfrak{Jac}_b(c).$$

Proof. Preparations: We begin by choosing a subdivision $a = t_0 < t_1 < \cdots < t_k = b$ such that $c|_{[t_i,t_{i+1}]}$ contains no conjugate points to $c(t_i)$ along itself. This can always be done: In a convex set around any point on c, the segment of c contained in that set is the unique geodesic connecting its points. No two points on that restriction of c are conjugate because exp is a diffeomorphism along that entire segment, see Proposition 2.3.2. Covering $c([a,b])$ by finitely many such convex sets gives the claim.
Let us write $\mathfrak{Jac}\{t_i\} \subseteq \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ for those piecewise smooth normal vector fields Y along c vanishing at a, b whose restriction to each $[t_i, t_{i+1}]$ is a Jacobi field along $c|_{[t_i,t_{i+1}]}$. Then it follows from Proposition 2.3.2 that such a Y is uniquely determined by specifying $Y(t_i)$, $i = 1, \ldots, k-1$. And since Y is normal to c and $Y(a) = 0$, $Y(b) = 0$, we see that $\dim \mathfrak{Jac}\{t_i\} = (n-1)(k-1)$.
The main idea of the proof is to approximate vector fields in $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ by those in $\mathfrak{Jac}\{t_i\}$. To this end, define a map $\phi : \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c) \to \mathfrak{Jac}\{t_i\}$ as follows: For $X \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$, let $\phi(X)|_{[t_i,t_{i+1}]}$ be the unique Jacobi field along $c|_{[t_i,t_{i+1}]}$ with $\phi(X)(t_i) = X(t_i)$ and $\phi(X)(t_{i+1}) = X(t_{i+1})$. Note that $\phi|_{\mathfrak{Jac}\{t_i\}}$ is the identity map. Also, if $X \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c) \setminus \mathfrak{Jac}\{t_i\}$, then applying Proposition 2.3.10 to each subinterval gives

$$I(X,X) < I(\phi(X), \phi(X)).$$

Reducing Ind *and* Ind_0 *to* $\mathfrak{Jac}\{t_i\}$: We will now show the following: If Ind' and Ind'_0 denote the index and extended index of c with respect to the restricted index form $I|_{\mathfrak{Jac}\{t_i\} \times \mathfrak{Jac}\{t_i\}}$ (i.e. they are defined as suprema with respect to subspaces of $\mathfrak{Jac}\{t_i\}$), then

$$\mathrm{Ind}(c) = \mathrm{Ind}'(c), \quad \mathrm{Ind}_0(c) = \mathrm{Ind}'_0(c). \tag{2.3}$$

In particular $\mathrm{Ind}(c)$ and $\mathrm{Ind}_0(c)$ are finite, because $\mathfrak{Jac}\{t_i\}$ has finite dimension.
To see this, first note that ϕ is $\mathbb{R}$-linear, thus $\phi(\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c))$ is a vector subspace of $\mathfrak{Jac}\{t_i\}$. Since we trivially have $\mathrm{Ind}(c) \geq \mathrm{Ind}'(c)$ and $\mathrm{Ind}_0(c) \geq \mathrm{Ind}'_0(c)$, we only need to establish "$\leq$". For this, let $A \subseteq \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ be a vector subspace

such that $I|_{A\times A}$ is positive semidefinite, and let $X \in A\backslash\mathfrak{Jac}\{t_i\}$ with $\phi(X) = 0$. Then by above,

$$0 = I(\phi(X), \phi(X)) > I(X, X),$$

hence the assumption on A implies $X = 0$. This, together with $\phi(X) = X$ for $X \in \mathfrak{Jac}\{t_i\}$ shows that $\phi|_A$ is injective. This has the following implication: If A is a subspace of $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ on which I is positive semidefinite, then $\phi(A)$ is a subspace of $\mathfrak{Jac}\{t_i\}$ of equal dimension on which I is also positive semidefinite due to $I(X, X) < I(\phi(X), I(\phi(X)))$. The same argument goes through word for word if we replace "semidefinite" with "definite". This establishes (2.3).

Proof of the main claim: We subdivide $[a, b]$ a second time into $a = s_0 < s_1 < \ldots s_m = b$ such that $\{s_i\}$ has no interior points in common with $\{t_i\}$ and such that $c|_{[s_i,s_{i+1}]}$ has no conjugate points to $c(s_i)$ along itself. It follows that

$$\mathfrak{Jac}\{s_i\} \cap \mathfrak{Jac}\{t_i\} = \mathfrak{Jac}_b(c).$$

We apply (2.3) to $\{s_i\}$: There is a vector subspace $B_0' \subseteq \mathfrak{Jac}\{s_i\}$ of maximal dimension such that I is positive semidefinite on it, i.e. $\mathrm{Ind}_0(c) = \dim B_0'$. Since for any $J \in \mathfrak{Jac}_b(c)$ we have $I(J, J) = -\langle J', J\rangle|_a^b = 0$, I is in particular positive semidefinite on $\mathfrak{Jac}_b(c) \times \mathfrak{Jac}_b(c)$, hence $\mathfrak{Jac}_b(c) \subseteq B_0'$ by maximality. By the proof of (2.3) above, $\phi|_{B_0'} : B_0' \to \mathfrak{Jac}\{t_i\}$ is injective. Let $B_0 := \phi(B_0')$. Since $\phi|_{\mathfrak{Jac}_b(c)}$ is the identity, $\mathfrak{Jac}_b(c) \subseteq B_0$. Complement it by some subspace $B \subseteq B_0$, i.e. $B_0 = B \oplus \mathfrak{Jac}_b(c)$.
Claim: $I|_{B\times B}$ is positive definite. By construction, $I|_{B_0'\times B_0'}$ is positive semidefinite. If $0 \neq Z \in B$, we may write it as $Z = \phi(X)$ for $X \in B_0'$ (since $B_0 = B \oplus \mathfrak{Jac}_b(c)$ and $B_0 = \phi(B_0')$). Then $X \notin \mathfrak{Jac}_b(c)$, because otherwise $X = \phi(X) = Z \in \mathfrak{Jac}_b(c)$, contradicting the direct sum decomposition. In particular,

$$I(Z, Z) = I(\phi(X), \phi(X)) > I(X, X) \geq 0,$$

where the last inequality follows from the fact that $I|_{B_0'\times B_0'}$ is positive semidefinite. Hence $I|_{B\times B}$ is positive definite.
Since $B \subseteq \mathfrak{Jac}\{t_i\}$, we have

$$\mathrm{Ind}(c) = \mathrm{Ind}'(c) \geq \dim B.$$

Since $B_0 = B \oplus \mathfrak{Jac}_b(c)$, we have

$$\mathrm{Ind}_0(c) = \dim B_0' = \dim B_0 = \dim B + \dim \mathfrak{Jac}_b(c),$$

so the claim of the the Proposition is proven once we establish $\dim B = \mathrm{Ind}'(c) = \mathrm{Ind}(c)$, and we already know that "$\leq$" holds. To show the other inequality, let $A \subseteq \mathfrak{Jac}\{t_i\}$ be a vector subspace with $\mathrm{Ind}(c) = \mathrm{Ind}'(c) = \dim A$ and $I|_{A\times A}$ is positive definite, and suppose that $\dim A > \dim B$. Then, since $I_{\mathfrak{Jac}_b(c)\times\mathfrak{Jac}_b(c)}$ is identically 0, $A\cap\mathfrak{Jac}_b(c) = \{0\}$ and I is positive semidefinite on $A \oplus \mathfrak{Jac}_b(c)$, which would imply

$$\mathrm{Ind}_0(c) \geq \dim A + \dim \mathfrak{Jac}_b(c) > \dim B + \dim \mathfrak{Jac}_b(c) = \mathrm{Ind}_0(c),$$

a contradiction. This concludes the proof. □

We are now ready to prove a Lorentzian version of the famous Morse Index Theorem from Riemannian geometry.

Theorem 2.3.13. (Lorentzian Morse Index Theorem, timelike version)
Let $c : [a, b] \to M$ be a future directed timelike geodesic. Then c only has finitely many conjugate points, and the index and extended index of c with respect to the index form $I : \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)\times\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c) \to \mathbb{R}$ are given by the formulas

$$\mathrm{Ind}(c) = \sum_{t\in(a,b)} \dim \mathfrak{Jac}_t(c), \tag{2.4}$$

$$\mathrm{Ind}_0(c) = \sum_{t\in(a,b]} \dim \mathfrak{Jac}_t(c). \tag{2.5}$$

Proof. This is a long proof, so we proceed in several steps.
The above-given sums are finite: Note that $\dim \mathfrak{Jac}_t(c) \geq 1$ if and only if $c(t)$ is conjugate to $c(a)$ along c, $t \in (a, b]$. We can embed $i : \mathfrak{Jac}_t(c) \hookrightarrow \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ by extending $J \in \mathfrak{Jac}_t(c)$ by 0 past the parameter value t. By Proposition 2.3.12, $\mathrm{Ind}_0(c) < \infty$. Our strategy to show this first claim is the following: If $\{t_1, \ldots, t_k\}$ is a finite set of parameter values $t_i \in (a, b]$ such that $c(t_i)$ is conjugate to $c(a)$ along c, then $k \leq \mathrm{Ind}_0(c)$, which would show that there can only be finitely many conjugate points to $c(a)$ along c, which in turn implies the finiteness of the sums.
To show this, let $A_j := i(\mathfrak{Jac}_{t_j}(c))$ and $A := \oplus_{i=1}^k A_i$. Then $A \subseteq \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$, and given $Z \in A$ we can write $Z = \sum_{i=1}^k \lambda_i Z_i$ with $Z_i \in A_i$. Then

$$I(Z, Z) = \sum_{i,j=1}^{k} \lambda_i\lambda_j I(Z_i, Z_j).$$

This is identically 0: If $t_i \leq t_j$, then by definition of the embeddings of the Jacobi fields into $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$ we have (cf. the proof of Proposition 2.3.6 for the notation of the restricted index form)

$$I(Z_i, Z_j) = I_a^{t_i}(Z_i, Z_j) + 0 = -\langle Z_j', Z_i\rangle|_a^{t_i} = 0.$$

Thus $I|_{A\times A}$ is identically 0 and, in particular, it is positive semidefinite, which shows that

$$k \leq \dim A \leq \mathrm{Ind}_0(c).$$

We have established the fact that $c(a)$ only has finitely many conjugate points along c in $(a,b]$, which we denote by $c(t_1),\dots,c(t_r)$. Hence $\mathfrak{Jac}_t(c) = \{0\}$ if $t \notin \{t_1,\dots,t_r\}$ and thus, the sums are finite.

We show (2.5): Once we do this, Proposition 2.3.12 automatically implies (2.4). To this end, we define the $\mathbb{Z}$-valued functions

$$f, f_0 : (a,b] \to \mathbb{Z}$$

defined by $f(t) := \mathrm{Ind}(c|_{[a,t]})$ and $f_0(t) := \mathrm{Ind}_0(c|_{[a,t]})$. We show that (2.5) holds if f is left continuous and f_0 is right continuous. Indeed, if $t \notin \{t_1,\dots,t_r\}$, then by Proposition 2.3.12

$$f(t) - f_0(t) = -\dim \mathfrak{Jac}_t(c) = 0.$$

Also, by left continuity of f and right continuity of f_0, $f(t_{j+1}) = f_0(t_j)$. This can be seen as follows: f_0 is easily seen to be nondecreasing, we will show the argument further below. And since $f_0(b) = \mathrm{Ind}_0(c)$ is finite and f_0 is $\mathbb{Z}$-valued, f_0 can only change a finite number of times. So we may choose a finer subdivision $a = s_1 < \cdots < s_m = b$ such that the t_i appear among the s_j and such that $f_0|_{(s_j,s_{j+1})}$ is constant (we demand constancy on the corresponding half-open intervals for $s_1 = a$ resp. $s_m = b$). Since the $\{s_j\}$ are finer than the $\{t_i\}$, we have $f = f_0$ on each (s_j, s_{j+1}). From here, one may use left and right continuity of f and f_0 respectively, to show the claim: Say $t_j = s_k$ and $t_{j+1} = s_l$. Then

$$f(t_{j+1}) = \lim_{t\uparrow t_{j+1}} f(t) = \lim_{t\uparrow t_{j+1}} f_0(t) = \lim_{t\downarrow s_{l-1}} f_0(t) = f_0(s_{l-1})$$

because f_0 is constant on (s_{l-1}, s_l) and right continuous. Since $s_{k+1},\dots,s_{l-1} \in (t_j,t_{j+1})$, and $f = f_0$ is continuous on (t_j,t_{j+1}) it is thus constant on (t_j,t_{j+1}) because it is continuous on each interval $(t_j = s_k, s_{k+1}),\dots,\ (s_{l-1}, s_l = t_{j+1})$ and there are no breakpoints due to continuity, so $f_0(s_{l-1}) = f_0(s_{k+1}) = f_0(s_k) = f_0(t_j)$ where the second-to-last equality is again due to right continuity. Hence

$$\begin{aligned}\sum_{t\in(a,b]} \dim \mathfrak{Jac}_t(c) &= \sum_{t\in(a,b]} (f_0(t) - f(t)) = \sum_{j=1}^{r} (f_0(t_j) - f(t_j)) \\ &= f_0(t_r) - f(t_1).\end{aligned}$$

Note that $f(t_1) = 0$ since $c|_{[a,t]}$ has no conjugate points to $c(a)$ if $t < t_1$ (cf. Theorem 2.3.9), hence $f(t_1) = \lim_{t\uparrow t_1} f(t) = 0$. Also, since f_0 is constant on

$[t_r, b]$ (see above; it is indeed constant on $[t_r, b]$ and not just $(t_r, b]$ due to right continuity), we have that $f_0(t_r) = f_0(b)$, hence the above calculation yields precisely (2.5).

Left continuity of f and right continuity of f_0: We already noted above that f and f_0 are nondecreasing, this is seen as follows: Let $s \leq t \in (a, b]$. Then we can embed $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,s]})$ into $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,t]})$ via extension by 0. This embedding preserves the index form, hence if $A \subseteq \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,s]})$ is a vector subspace on which I is positive (semi)definite, then it can be seen as a vector subspace of $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,t]})$ and I is positive (semi)definite on it, hence $f(s) \leq f(t)$ and $f_0(s) \leq f_0(t)$.
Now fix $\tilde{t} \in (a, b]$. Moreover, fix $\delta > 0$ such that for any $s_1, s_2 \in [a, b]$ with $|s_1 - s_2| < \delta$, $s_1 \leq s_2$, $c|_{[s_1,s_2]}$ has no conjugate points to $c(s_1)$ along itself. Choose a subdivison $a = t_0 < t_1 < \ldots t_k = \tilde{t}$ (this has nothing to do with the t_i that were used before for the conjugate points!) with $|t_i - t_{i+1}| < \delta$. Let $J \subseteq [a, b] \cap (t_{k-1} - \delta, t_{k-1} + \delta)$ be an open interval with $\tilde{t} \in J$. For $t \in J$, let us write $\tilde{\mathfrak{Jac}}(c_t)$ for the (finite dimensional) subspace of $\mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,t]})$ consisting of vector fields $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c|_{[a,t]})$ such that $Y_{[t_j,t_{j+1}]}$, $j \in \{0, \ldots, k-2\}$, and $Y|_{[t_{k-1},t]}$ are Jacobi fields along the respective segments. By (2.3), $f(t)$ and $f_0(t)$ can be calculated with respect to the restriction of the index form I to $\tilde{\mathfrak{Jac}}(c_t) \times \tilde{\mathfrak{Jac}}(c_t)$. Now let

$$E := \bigsqcup_{i=1}^{k-1} T^{\perp}_{c(t_i)} M \subseteq TM,$$

where $T^{\perp}_{c(t_i)} M = \{v \in T_{c(t_i)} M : \langle v, c'(t_i) \rangle = 0\}$. Let us understand E as a vector space (via an outer direct sum; understand the topology on E to be the product topology, which is the subspace topology it inherits from TM), then we can define a positive definite scalar product on it componentwise. E is a closed subset of TM, and hence, since E only consists of spacelike vectors, $S := \{v \in E : \langle v, v \rangle = 1\}$ is compact.
Since $c|_{[t_i,t_{i+1}]}$ has no conjugate points, for any $v \in T^{\perp}_{c(t_i)} M$ and any $w \in T^{\perp}_{c(t_{i+1})} M$ there is a unique Jacobi field Y along c with $Y(t_i) = v$ and $Y(t_{i+1}) = w$ (see Proposition 2.3.2). From Proposition 2.1.11 it follows that $Y \perp c'$ everywhere. Such Jacobi fields can be put together at the t_i to get an element from $\tilde{\mathfrak{Jac}}(c_t)$. Thus, we see that

$$\phi_t : \tilde{\mathfrak{Jac}}(c_t) \to E, \quad \phi_t(Y) := (Y(t_1), \ldots Y(t_{k-1}))$$

is a vector space isomorphism. We define the following quadratic form:

$$Q_t : E \times E \to \mathbb{R}, \quad Q_t(u, v) := I(\phi_t^{-1}(u), \phi_t^{-1}(v)).$$

By (2.3), the index of Q_t (defined in the same manner as for I) is equal to $f(t)$ and the extended index is equal to $f_0(t)$. We further define

$$Q : E \times E \times J \to \mathbb{R}, \quad (v, w, t) \mapsto Q_t(v, w).$$

We now show that Q is continuous: Let $B := \{Y|_{[a,t_{k-1}]} : Y \in \tilde{\mathfrak{Jac}}(c_t)\}$. Then B is (via the restriction ϕ of ϕ_t to it) isomorphic to E (by precisely the same line of argument as for $\tilde{\mathfrak{Jac}}(c_t)$). Hence,

$$\begin{aligned} Q(u,v,t) &= I(\phi_t^{-1}(u), \phi_t^{-1}(v)) \\ &= I(\phi^{-1}(u), \phi^{-1}(v)) + I(\phi_t^{-1}(u)|_{[t_{k-1},t]}, \phi_t^{-1}(v)|_{[t_{k-1},t]}) \\ &= I(\phi^{-1}(u), \phi^{-1}(v)) - \langle \phi_t^{-1}(u)|_{[t_{k-1},t]}, \phi_t^{-1}(v)|'_{[t_{k-1},t]} \rangle|_{t_{k-1}}^{t}. \end{aligned}$$

The first term is certainly continuous. As for the right term, the Jacobi fields inside the scalar product are independent of t, because they vanish at t_{k-1} and are hence uniquely determined by their derivative at t_{k-1} (cf. Proposition 2.1.14), which in turn is determined by parameter values arbitrarily close to t_{k-1}. Hence, also the right term is continuous, which shows that Q is continuous.
We now show left continuity of f and right continuity of f_0: Recall that we chose some $\tilde{t} \in (a,b]$. By what was said above, there is a subspace $A \subseteq E$ such that $\dim A = f(\tilde{t})$ and $Q(u,u,\tilde{t}) > 0$ for all $0 \neq u \in A$. By continuity, there is some neighborhood $J_0 \ni \tilde{t}$, $J_0 \subseteq J$, such that $Q(u,u,t) > 0$ for all $t \in J_0$ and $u \in S \cap A$ (note that $S \cap A$ is compact). Hence, $Q_t|_{A\times A}$ is positive definite for all $t \in J_0$, thus $f(t) \geq f(\tilde{t})$ for all $t \in \tilde{t}$. Since f is nondecreasing, we have $f(t) = f(\tilde{t})$ for all $t \leq \tilde{t}$, $t \in J_0$, which shows that f is constant for a while left of $\tilde{t}$, and is thus in particular left continuous at $\tilde{t}$, which was arbitrarily chosen, so we are done.
It remains to show that f_0 is right continuous at $\tilde{t}$: For this, let $s_n \in J$, $s_n \downarrow \tilde{t}$. Since f_0 has values in $\mathbb{Z}$, we may assume that $f_0(s_n) = k$ for all n. Since it is nondecreasing, we have $f_0(\tilde{t}) \leq k$, so we only have to show $f_0(\tilde{t}) \geq k$. To this end, choose subspaces $A_n \subseteq E$ of dimension k such that $Q_{s_n}|_{A_n\times A_n}$ is (maximally) positive semidefinite there. Choosing orthonormal bases $\{e_1^n, \dots, e_k^n\}$ of A_n (with respect to the positive definite scalar product on the vector space E that was defined component-wise), then $e_i^n \in S$ for all $i = 1, \dots, k$ and all n. Since S is compact, $e_i^n \to e_i \in S$ for all i (up to subsequences). Then $\{e_1, \dots, e_k\}$ are an orthonormal basis for a k-dimensional subspace A of E. For $u \in A$, write $u = \sum_{j=1}^k \lambda_j e_j$. Then $u^n \to u$, where $u^n = \sum_{j=1}^k \lambda_j e_j^n \in A_n$. Since Q is continuous, we get

$$Q_{\tilde{t}}(u,u) = \lim_{n\to\infty} Q_{s_n}(u^n, u^n) \geq 0$$

since Q_{s_n} is by assumption positive semidefinite on $A_n \times A_n$. Hence, $Q_{\tilde{t}}$ is positive semidefinite on $A \times A$, which proves $f_0(\tilde{t}) \geq k$. This concludes our proof of the Morse Index Theorem. □

2.3.3 Jacobi and Lagrange tensors along timelike geodesics

In this final subsection concerning conjugate points on timelike geodesics, we would like to give an alternative way of describing conjugate points, namely via Jacobi tensors, which is as a collective matrix formalism of the usual vector formalism that one deals with in the context of Jacobi fields. A Jacobi tensor can be seen as a matrix whose entries are Jacobi fields, and the Jacobi tensor equation is just a sort of matrix version of the usual Jacobi equation. While these methods were not necessary in the context of conjugate points along timelike geodesics, they will become invaluable in the context of null geodesics and of singularity theorems, so we may as well introduce them now.

Suppose $c : [a, b] \to M$ is a timelike geodesic. Then the image of c is an immersed submanifold (with boundary) of M. For immersed submanifolds, it is still possible to define the normal bundle in a similar manner as in Definition 1.1.4, which we will denote Nc.

Definition 2.3.14. (Jacobi tensors along timelike geodesics)
Let $c : [a, b] \to M$ be a timelike geodesic. A *Jacobi tensor* A along c is a smooth map $A : [a, b] \to \mathrm{Hom}(Nc, Nc)$ such that $\pi \circ A = c$ (we will refer to such tensors along c as $(1, 1)$-*tensors normal to* c) satisfying the *Jacobi tensor equation*

$$A'' + RA = 0$$

and the nontriviality condition

$$\ker A(t) \cap \ker A'(t) = \{0\} \quad \text{for all } t \in [a, b],$$

where A' is the covariant derivative of the tensor field A and R is the $(1, 1)$-tensor normal to c defined via $R(v) := R(v, c'(t))c'(t)$ if $v \in T^{\perp}_{c(t)}M$, the so-called *tidal force operator*.

It is not hard to see that if A is any $(1, 1)$-tensor normal to c, then its pointwise adjoint $A^{\dagger}$ (with respect to g) is also a $(1, 1)$-tensor normal to c: To see this, note that the restriction of g to $T^{\perp}_{c(t)}M$ is positive definite, hence for any t, $A(t)^{\dagger}$ is a well-defined endomorphism $T^{\perp}_{c(t)}M \to T^{\perp}_{c(t)}M$. Smoothness of $t \mapsto A(t)^{\dagger}$ is clear. Also, we note that by linearity, $(A^{\dagger})' = (A')^{\dagger}$.

Definition 2.3.15. (Lagrange tensors along timelike geodesics)
Let $c : [a, b] \to M$ be a timelike geodesic. A Jacobi tensor A along c is called *Lagrange tensor* if

$$(A')^{\dagger}A - A^{\dagger}A' = 0.$$

Lemma 2.3.16. (Characterizing Lagrange tensors)
Let A be a Jacobi tensor along the timelike geodesic $c : [a, b] \to M$. If $A(t_0) = 0$ for some $t_0 \in [a, b]$, then A is a Lagrange tensor.

Proof. For any two Jacobi tensors A, B,

$$W(A,B) := (A')^\dagger B - A^\dagger B'$$

is a $(1,1)$-tensor normal to c, commonly referred to as the *Wronskian* of A and B. By definition, a Jacobi tensor A is Lagrange if and only if $W(A,A)=0$. The tidal force operator R is self-adjoint by the symmetries of the curvature tensor, because

$$\langle R(v), w\rangle = \langle R(v,c')c', w\rangle = \langle R(c',v)w, c'\rangle = \langle R(w,c')c', v\rangle = \langle v, R(w)\rangle. \quad (*)$$

We can use this and the fact that $(A')^\dagger = (A^\dagger)'$ (since taking the adjoint is a linear operation) to show that the Wronskian of any two Jacobi tensors is always constant (we call a $(1,1)$-tensor C normal to c *constant* along c if the following holds: Given $t \in [a,b]$, $v \in T^\perp_{c(t)}M$ and let Y be the parallel translate of v along c, then CY is the parallel translate of $C(t)v$ along c; such a $(1,1)$-tensor C is constant in this sense if and only if $C'=0$):

$$\begin{aligned}(W(A,B))' &= (A'^\dagger B - A^\dagger B')' = (A'')^\dagger B + (A')^\dagger B' - (A')^\dagger B' - A^\dagger B'' \\ &= (-RA)^\dagger B + A^\dagger RB \overset{(*)}{=} -A^\dagger RB + A^\dagger RB = 0.\end{aligned}$$

Since by assumption $A(t_0)=0$, also $A^\dagger(t_0)=0$, hence $W(A,A)(t_0)=0$, which implies $W(A,A)=0$ everywhere, which is precisely the characteristic property of Lagrange tensors. $\square$

Lemma 2.3.17. (Jacobi tensors and Jacobi fields)
Let $c:[a,b]\to M$ be a timelike geodesic. Let A be a Jacobi tensor along c and let $Y \in \mathfrak{X}(c)$ be parallel. Then $A(Y)$ is a Jacobi field along c. If $Y \neq 0$, then $A(Y)\neq 0$.

Proof. Since Y is parallel and A satisfies the Jacobi tensor equation, we have

$$(A(Y))'' = A''(Y) = -RA(Y) = -R(A(Y),c')c',$$

which means that $A(Y)$ satisfies the Jacobi equation. If $A(Y)$ were trivial, then for any t we would have $Y(t)\in\ker A(t)$, but also, since $(A(Y))' = A'(Y)$, $Y(t)\in\ker A'(t)$ for all t. But we demanded $\ker A(t)\cap\ker A'(t)=\{0\}$ for Jacobi tensors, hence $Y=0$. $\square$

Remark 2.3.18. (On Jacobi tensors and conjugate points)
Let $c:[a,b]\to M$ be a timelike geodesic and suppose A is a Jacobi tensor along c such that $A(a)=0$ and such that it has a nontrivial kernel somewhere along c except at $t=a$, i.e. there is some $t_0\in(a,b]$ and $0\neq v\in T^\perp_{c(t)}M$ such that $A(t_0)v=0$. Let $Y\in\mathfrak{X}(c)$ be the parallel translate of v along c, then by above $A(Y)$ is a nontrivial Jacobi field with $A(Y)(a)=0$ and $A(Y)(t_0)=A(t_0)v=0$, hence $c(t_0)$ is conjugate to $c(a)$ along c.

Lemma 2.3.19. (On the kernel intersection condition for Jacobi tensors)
Let A be a $(1,1)$-tensor along a timelike geodesic $c : [a,b] \to M$ satisfying the Jacobi tensor equation $A'' + RA = 0$. Then A is a Jacobi tensor if and only if there is some parameter $t_0 \in [a,b]$ such that

$$\ker A(t_0) \cap \ker A'(t_0) = \{0\}.$$

Proof. One direction is clear by definition. Now suppose there is some $t_0 \in [a,b]$ such that the kernel intersection is trivial. We need to show that the kernel intersection is trivial everywhere. For this, suppose there is some $s \in [a,b]$, $s \neq t_0$, such that there exists $0 \neq v \in \ker A(s) \cap \ker A'(s)$. Parallel translating v along c gives a nontrivial parallel vector field Y. Then AY is a nontrivial Jacobi field. Noting that $(AY)' = A'Y$, we have $AY(s) = A(s)(v) = 0$ and $(AY)'(s) = A'(s)(v) = 0$ due to our assumption on s. Since a Jacobi field is uniquely determined by its value and derivative at a parameter, it follows that $AY = 0$, a contradiction. □

Proposition 2.3.20. (Some properties of Jacobi tensors)
Let $c : [a,b] \to M$ be a timelike geodesic.

(1) Let $t_0 \in [a,b]$. Given any two linear maps $L_1, L_2 : T^{\perp}_{c(t_0)}M \to T^{\perp}_{c(t_0)}M$, there is a unique $(1,1)$-tensor A normal to c satisfying the Jacobi tensor equation such that $A(t_0) = L_1$, $A'(t_0) = L_2$. If L_1 or L_2 are nonsingular, then A is a Jacobi tensor.

(2) Two points $c(t_0)$ and $c(t_1)$ along c are conjugate along c if and only if the unique Jacobi tensor A along c with $A(t_0) = 0$, $A'(t_0) = \mathrm{id}$ satisfies $\ker A(t_1) \neq \{0\}$.

(3) The boundary value problem for the Jacobi tensor equation for any given linear maps $L_1 : T^{\perp}_{c(t_0)}M \to T^{\perp}_{c(t_0)}M$ and $L_2 : T^{\perp}_{c(t_1)}M \to T^{\perp}_{c(t_1)}M$, is uniquely solvable if and only if $c(t_0)$ and $c(t_1)$ are not conjugate along c. If either L_1 or L_2 is nonsingular, then the corresponding solution of the boundary value problem is a Jacobi tensor.

Proof.

(1) Using a parallel orthonormal frame along c, one can rewrite the Jacobi tensor equation as a linear second order matrix ODE, similarly as it was done for the Jacobi equation, cf. Proposition 2.1.7, so providing two initial data gives a unique solution. If one of the initial data is nonsingular, then the kernel intersection property is satisfied at the initial parameter, hence everywhere by Lemma 2.3.19.

(2) One direction was discussed in Remark 2.3.18. For the other direction, suppose $c(t_0)$ and $c(t_1)$ are conjugate along c, so there is some nontrivial Jacobi field $J_1 \in \mathfrak{Jac}(c)$ with $J(t_0) = 0$ and $J(t_1) = 0$, which

is then necessarily orthogonal to c. Since J_1 is nontrivial, we have $J_1'(t_0) \neq 0$, so we may parallel translate it along c to get a parallel field E_1 along c that is normal to c. We may complete E_1 to a parallel frame $E_1, \dots, E_{n-1}$ along c normal to c. Let $J_2, \dots, J_{n-1}$ be the unique Jacobi fields along c with $J_i(t_0) = 0$, $J_i'(t_0) = E_i(t_0)$. We can define a $(1,1)$-tensor A normal to c by setting

$$A(E_i) := J_i$$

and extending linearly. Then $A(t_0)E_i(t_0) = J_i(t_0) = 0$ for all i, hence $A(t_0) = 0$, and $A'(t_0)E_i(t_0) = (AE_i)'(t_0) = J_i'(t_0) = E_i(t_0)$, hence $A'(t_0) = id$. Moreover, $A(t_1)E_1(t_1) = J_1(t_1) = 0$, hence $\ker A(t_1) \neq \{0\}$. It remains to argue that A satisfies the Jacobi equation, because the kernel intersection property is clear by Lemma 2.3.19. For this, we check that $A'' + RA$ gives 0 on all E_i, which follows from parallelity of the latter and the Jacobi equation for the J_i:

$$A''(E_i) = (A(E_i))'' = J_i'' = -R(J_i, c')c' = -RA(E_i).$$

(3) The proof is similar to the corresponding proof for the boundary value problem of the Jacobi equation. First we assume that $c(t_0)$ and $c(t_1)$ are not conjugate along c. Let $\mathcal{J}$ denote the space of $(1,1)$-tensor fields normal to c satisfying the Jacobi tensor equation. We define the linear map

$$\begin{aligned} \phi &: \mathcal{J} \to \mathrm{End}(T^{\perp}_{c(t_0)}M) \times \mathrm{End}(T^{\perp}_{c(t_1)}M) \\ \phi(A) &:= (A(t_0), A(t_1)). \end{aligned}$$

ϕ is injective: Suppose $\phi(A) = 0$, and let Y be a parallel vector field along c. Then $A(Y)$ is a Jacobi field with $AY(t_0) = 0$ and $AY(t_1) = 0$, hence $AY = 0$ because $c(t_1)$ is not conjugate to $c(t_0)$ along c. Since Y was an arbitrary parallel vector field, it follows that $A = 0$. Moreover, $\mathcal{J}$ and $\mathrm{End}(T^{\perp}_{c(t_0)}M) \times \mathrm{End}(T^{\perp}_{c(t_1)}M)$ have the same dimension, hence ϕ is an isomorphism.

Conversely, suppose $c(t_0)$ and $c(t_1)$ are conjugate along c. Let J_1 be a nontrivial Jacobi field realizing this conjugacy. Let $E_1, \dots, E_{n-1}$ be a parallel frame along c normal to c. We define A by $A(E_1) := J_1$ and $A(E_j) = 0$ for all $j \neq 1$. Then A is a nontrivial $(1,1)$-tensor satisfying the Jacobi tensor equation and $A(t_0) = 0$, $A(t_1) = 0$. This shows that the map ϕ defined above is not injective, which means the boundary value problem is not uniquely solvable for some boundary data.

□

Definition 2.3.21. (Expansion, vorticity and shear)
Let A be a Jacobi tensor along a timelike geodesic $c : [a, b] \to M$. Let $B := A'A^{-1}$ wherever A is nonsingular. We define the *expansion, vorticity and shear* of A along c. The first is a smooth function $[a, b] \to \mathbb{R}$ and the latter two are $(1, 1)$-tensors normal to c.

(1) *Expansion* $\theta := \mathrm{Tr}(B)$.

(2) *Vorticity* $\omega := \frac{1}{2}(B - B^{\dagger})$.

(3) *Shear* $\sigma := \frac{1}{2}(B + B^{\dagger}) - \frac{\theta}{n-1}E$,
where E is the $(1, 1)$-tensor normal to c defined by $E(t) = \mathrm{id}$ for all t.

Remark 2.3.18 already hints at the fact that a Jacobi tensor can be viewed as a matrix of Jacobi fields, an observation that we will discuss in more detail later. With this, a Jacobi tensor is associated to geodesic variations that together form a *congruence* of geodesics. The expansion θ, the vorticity ω and the shear σ of this Jacobi tensor have an optical interpretation for the congruence suggested by their names. We refer to [9, p. 13] for a detailed discussion.

Remark 2.3.22. (Alternative formula for θ)
One can use Jacobi's formula from linear algebra (cf. [12, Par. 0.8.10]) to deduce the following formula for the expansion $\theta(t) = \mathrm{Tr}\big(A'(t)A^{-1}(t)\big)$:

$$\theta(t) = (\det A)'(\det A)^{-1}.$$

Proposition 2.3.23. (Raychaudhuri equation)
The expansion, vorticity and shear of a Jacobi tensor A satisfy the *Raychaudhuri equation*

$$\theta' = -\operatorname{Ric}(c', c') - \mathrm{Tr}\big(\omega^2\big) - \mathrm{Tr}\big(\sigma^2\big) - \frac{\theta^2}{n-1}.$$

Proof. Note that since $AA^{-1} = E$, we have $0 = (AA^{-1})' = A'A^{-1} + A(A^{-1})'$, so

$$(A^{-1})' = -A^{-1}A'A^{-1}. \tag{2.6}$$

Hence, by the fact that A is a Jacobi tensor, we get

$$B' = (A'A^{-1})' = A''A^{-1} - A'A^{-1}A'A^{-1} = -R - B^2.$$

We can write $B = \omega + \sigma + \frac{\theta}{n-1}E$, so it follows that

$$\theta' = \mathrm{Tr}\big(B'\big) = -\mathrm{Tr}(R) - \mathrm{Tr}\big(B^2\big) \tag{2.7}$$

$$= -\mathrm{Tr}(R) - \mathrm{Tr}\left(\left(\omega + \sigma + \frac{\theta}{n-1}E\right)^2\right). \tag{2.8}$$

Note that

$$\operatorname{Tr}(\omega) = \frac{1}{2}(\operatorname{Tr}(B) - \operatorname{Tr}(B)) = 0,$$
$$\operatorname{Tr}(\sigma) = \operatorname{Tr}(B) - \frac{\operatorname{Tr}(B)}{n-1}\operatorname{Tr}(E) = \operatorname{Tr}(B) - \operatorname{Tr}(B) = 0,$$

$$\begin{aligned}\operatorname{Tr}(\omega\sigma) = &\frac{1}{4}\left(\operatorname{Tr}\big(B^2\big) - \operatorname{Tr}\Big(B^\dagger B\Big) + \operatorname{Tr}\Big(BB^\dagger\Big) - \operatorname{Tr}\Big((B^\dagger)^2\Big)\right)\\ &+ \frac{\operatorname{Tr}(B)}{2(n-1)}\left(\operatorname{Tr}(B) - \operatorname{Tr}\Big(B^\dagger\Big)\right) = 0.\end{aligned}$$

So we may continue our calculation at (2.7) as follows:

$$\theta' = -\operatorname{Tr}(R) - \operatorname{Tr}\big(\omega^2\big) - \operatorname{Tr}\big(\sigma^2\big) - \frac{\theta^2}{n-1}.$$

To calculate $\operatorname{Tr}(R)$, choose an orthonormal frame $E_1, \dots, E_n$ along c with $E_n = c'$, then $E_1, \dots, E_{n-1}$ are spacelike and span the normal spaces to c', hence

$$\begin{aligned}\operatorname{Tr}(R) &= \sum_{i=1}^{n-1}\langle RE_i, E_i\rangle = \sum_{i=1}^{n-1}\langle R(E_i, c')c', E_i\rangle = \sum_{i=1}^{n}\langle E_i, E_i\rangle\langle R(E_i, c')c', E_i\rangle\\ &= \operatorname{Ric}(c', c').\end{aligned}$$

This concludes our proof. □

Corollary 2.3.24. (Vorticity-free Raychaudhuri equation)
If A is a Lagrange tensor along a timelike geodesic $c : [a, b] \to M$, then $B = A'A^{-1}$ is self-adjoint (wherever it is defined). Hence $\omega = 0$ and the vorticity-free Raychaudhuri equation holds:

$$\theta' = -\operatorname{Ric}(c', c') - \operatorname{Tr}\big(\sigma^2\big) - \frac{\theta^2}{n-1}.$$

Proof. Since A is Lagrange, we have $(A')^\dagger A = A^\dagger A'$, so $A' = (A^\dagger)^{-1}(A')^\dagger A$. Hence,

$$B = A'A^{-1} = (A^\dagger)^{-1}(A')^\dagger = (A^{-1})^\dagger(A')^\dagger = B^\dagger,$$

where $(A^{-1})^\dagger = (A^\dagger)^{-1}$ by elementary linear algebra. The rest follows trivially. □

Remark 2.3.25. (From Raychaudhuri to Riccati)
In the proof of Proposition 2.3.23, we saw that if A is a Jacobi tensor, then $B := A'A^{-1}$ satisfies the *matrix Riccati equation*

$$B' + B^2 + R = 0.$$

This is an important fact e.g. for the application of Riccati comparison techniques (see e.g. [8, Sec. 4]).

2.4 Null index theory

In this section, we want to conduct an analysis of Jacobi fields, conjugate points and related notions along null geodesics. The results for timelike geodesics do not directly carry over, however, because the fact that a null geodesic is normal to itself causes some issues. The trick is to quotient out the span of the tangent vector from the normal bundle, and from then on one can proceed in much the same way as in the timelike case. Because many of the proofs are near replicas of the timelike results, we shall only give an outline and reference the proofs that have already been carried out. Throughout this section, unless explicitly stated otherwise, let (M, g) be a spacetime of dimension $n \geq 3$ (this is assumed because null geodesics have no conjugate points in dimension $n = 2$, see [1, Lem. 10.45]), and let $\beta : [a, b] \to M$ be a future directed null geodesic. Our presentation once again follows [1, Ch. 10].

2.4.1 Quotient constructions along null geodesics

Remark 2.4.1. (Correct setting for null geodesics)
A null geodesic $\beta : [a, b] \to M$ defines an immersed submanifold of M, so we may define its normal bundle $N\beta$, which is a rank $(n-1)$-subbundle of TM. Then $[\beta'] := \operatorname{span}\{\beta'\} \subseteq N\beta$ is a rank 1-subbundle, so we may consider the quotient bundle $Q\beta := N\beta/[\beta']$, which is a rank $(n-2)$-vector bundle over (the image of) β.

(Piecewise smooth) sections of this vector bundle, denoted $\overline{\mathfrak{X}}^{\perp}_{\text{pw}}(\beta)$, may be seen as projections of normal vector fields along β, i.e. they are of the form $Y(t) + [\beta'(t)] \in Q\beta(t) = T^{\perp}_{\beta(t)}M/[\beta'(t)]$ with $Y \in \mathfrak{X}^{\perp}_{p.w.}(\beta)$. Furthermore, we write $\overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$ for those piecewise smooth sections Y of $Q\beta$ that satisfy $Y = [\beta']$ at the initial and final parameter. By $\overline{\mathfrak{X}}(\beta)$ we denote the smooth sections of $Q\beta$, other objects associated with $Q\beta$ will be denoted with bars on them in a similar way.

An alternative way to consider $Q\beta$ (without resorting to quotient bundles) is as follows: Consider a null vector field $\eta \in \mathfrak{X}(\beta)$ with $\langle \eta, \beta' \rangle = -1$ everywhere (obtained e.g. by parallel translation). Extend $\{\eta, \beta'\}$ to a parallel frame along β by taking $(n-2)$ additional spacelike parallel fields $E_1, \dots, E_{n-2}$ that are orthonormal among each other and orthogonal to η and β. Then the $\{E_1, \dots, E_{n-2}\}$ span an $(n-2)$ subbundle $S\beta$ of $N\beta$, and it is easily seen that

$$N\beta \cong [\beta'] \oplus S\beta \quad \text{and hence} \quad Q\beta \cong S\beta.$$

Furthermore, the metric g, which can be seen as a smooth map $N\beta \oplus N\beta \to \mathbb{R}$, descends to a metric $\overline{g} : Q\beta \oplus Q\beta \to \mathbb{R}$. Via the above identification $Q\beta \cong S\beta$, it is easy to see that $\overline{g}$ is positive definite on each fiber of $Q\beta$.

There is a natural induced notion of covariant derivative on $Q\beta$: If $Y \in \overline{\mathfrak{X}}^{\perp}_{\mathrm{pw}}(\beta)$, we may write $Y(t) = V(t) + [\beta'(t)]$, then

$$Y'(t) = V'(t) + [\beta'(t)].$$

This is easily seen to be well-defined: If $V_1, V_2 \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta)$ are such that $V_1(t) = V_2(t) + f(t)\beta'(t)$, then $V_1'(t) = V_2'(t) + f'(t)\beta'(t)$, so they project to the same element in $\overline{\mathfrak{X}}^{\perp}_{\mathrm{pw}}(\beta)$. This also shows, assuming V not to have any components in the β'-direction, that the covariant differentiations in $Q\beta$ and in $S\beta$ are compatible with their identification: If η is a parallel null field as above with $\langle \eta, \beta' \rangle = -1$ and V is a section of $S\beta$, then since $\langle \beta', V \rangle = 0$ and $\langle \eta, V \rangle = 0$, we get

$$\langle \beta', V' \rangle = \langle \eta, V' \rangle = 0,$$

so V' is again a section of $S\beta$. Note that covariant differentiation in $Q\beta$ is compatible with the projected metric $\overline{g}$ in the usual way.

Since we want to make sense of Jacobi fields in this quotient setting, we also project the curvature endomorphism (i.e. the tidal force operator)

$$R(., \beta'(t))\beta'(t) : T^{\perp}_{\beta(t)}M \to T^{\perp}_{\beta(t)}M$$

to the quotient in the natural way, namely if $v \in T^{\perp}_{\beta(t)}M$ and $v = x + [\beta'(t)]$, then

$$\overline{R}(v, \beta'(t))\beta'(t) := \pi(R(x, \beta'(t))\beta'(t)),$$

where $\pi : N\beta \to Q\beta$ denotes the quotient map. This is easily seen to be well-defined by the symmetries of the curvature tensor. Hence we obtain the projected curvature endomorphism (or projected tidal force operator)

$$\overline{R}(., \beta'(t))\beta'(t) : T^{\perp}_{\beta(t)}M/[\beta'(t)] \to T^{\perp}_{\beta(t)}M/[\beta'(t)],$$

or, in bundle language, a bundle homomorphism $\overline{R}(., \beta')\beta' : Q\beta \to Q\beta$. The projected operator inherits the obvious symmetries with respect to the projected metric $\overline{g}$.

2.4.2 Jacobi classes and conjugate points

In this subsection, we consider projections of Jacobi fields and we study the notion of conjugate points along null geodesics.

Definition 2.4.2. (Jacobi classes)
A smooth section $W \in \overline{\mathfrak{X}}(\beta)$ is a *Jacobi class* along β if it satisfies the *projected Jacobi equation* along β:

$$W'' + \overline{R}(W, \beta')\beta' = [\beta'].$$

We will denote the space of all Jacobi classes along β by $\overline{\mathfrak{Jac}}(\beta)$.

The main reason why we resort to these quotient constructions is that the two Jacobi fields that span the space of tangential Jacobi fields (cf. Remark 2.1.13, namely $\beta'(t)$ and $t\beta'(t)$, are in fact both normal Jacobi fields as well since β is null.

Lemma 2.4.3. (Jacobi fields and Jacobi classes)
Let $W \in \overline{\mathfrak{Jac}}(\beta)$ be a Jacobi class along β. Then there is a Jacobi field $J \in \mathfrak{Jac}(\beta)$ such that $W = J + [\beta']$. Conversely, for any Jacobi field $J \in \mathfrak{Jac}(\beta)$, $W := J + [\beta']$ is a Jacobi class.

Proof. Let W be a Jacobi class and let Y be the corresponding section of $S\beta$. By assumption, we have

$$W'' + \overline{R}(W, \beta')\beta' = [\beta'],$$

so $Y'' + R(Y, \beta')\beta' = f\beta'$ for some smooth function $f : [a, b] \to \mathbb{R}$. Let $h : [a, b] \to \mathbb{R}$ smooth with $h'' = f$, and set $J := Y - h\beta'$ (note that J is generally no longer a section of $S\beta$ since it has nonzero components in the β' direction). Then $W = Y + [\beta'] = J + [\beta']$, and hence

$$\begin{aligned} f\beta' &= Y'' + R(Y, \beta')\beta' = J'' + h''\beta' + R(J, \beta')\beta' \\ &= J'' + f\beta' + R(J, \beta')\beta', \end{aligned}$$

hence $J \in \mathfrak{Jac}(\beta)$.
Conversely, if $J \in \mathfrak{Jac}(\beta)$, then $J'' + R(J, \beta')\beta' = 0$ and thus
$W'' + \overline{R}(W, \beta')\beta' = [\beta']$ with $W := J + [\beta']$. $\square$

As we noted in the proof, for a Jacobi class $W \in \overline{\mathfrak{Jac}}(\beta)$, the corresponding section of $S\beta$ need not be a Jacobi field. But it is always possible to obtain W as the projection of a Jacobi field along β, and all Jacobi classes are obtained in this way, as Lemma 2.4.3 shows. In fact, we have the following result.

Lemma 2.4.4. (Jacobi classes and Jacobi fields in $S\beta$)
Let $W \in \overline{\mathfrak{Jac}}(\beta)$. Then the section J of $S\beta$ corresponding to W is a Jacobi field if and only if $R(J, \beta')\beta'|_t \in S\beta|_t$ for all $t \in [a, b]$, where $S\beta|_t$ is the fiber of $S\beta$ at $\beta(t)$, i.e. it is spanned by $(n-2)$ spacelike parallel orthonormal vectors $E_1(t), \ldots, E_{n-2}(t)$ as discussed in Remark 2.4.1.

Proof. Suppose first that $J \in \Gamma(S\beta)$ is a Jacobi field. Write $J = \sum_{j=1}^{n-2} J^j E_j$, then

$$R(J(t), \beta'(t))\beta'(t) = -J''(t) = -\sum_{j=1}^{n-2} (J^j)''(t) E_j(t) \in S\beta|_t.$$

Conversely, if $R(J(t), \beta'|(t))\beta'(t) \in S\beta|_t$, then, since W satisfies the projected Jacobi equation, we have

$$J'' + R(J, \beta')\beta' = f\beta',$$

but since neither J'' nor $R(J, \beta')\beta'$ have components in the β'-direction, it follows that $f = 0$. Hence $J \in \mathfrak{Jac}(\beta)$. $\square$

Lemma 2.4.5. (Jacobi classes vanishing at $t = a$ and $t = b$)
Let $W \in \overline{\mathfrak{Jac}}(\beta)$ with $W(a) = [\beta'(a)]$ and $W(b) = [\beta'(b)]$. Then there is a unique Jacobi field $J \in \mathfrak{Jac}(\beta)$ with $J(a) = 0$, $J(b) = 0$, such that $W = J + [\beta']$.

Proof. By Lemma 2.4.3, there is certainly some Jacobi field J_1 with $W = J_1 + [\beta']$. For $c_1, c_2 \in \mathbb{R}$, $J(t) := J_1 + c_1\beta'(t) + c_2 t\beta'(t)$ is also a Jacobi field. Since J_1 projects onto W, and $W(a) = [\beta'(a)]$ and $W(b) = [\beta'(b)]$, we have $J_1(a) = C_1\beta'(a)$ and $J_2(b) = C_2\beta'(b)$ for some $C_1, C_2 \in \mathbb{R}$. Setting

$$\begin{aligned} c_1 &:= \frac{C_2 a - C_1 b}{b - a}, \\ c_2 &:= \frac{1}{b}\left(\frac{C_1 b - C_2 a}{b - a} - C_2\right), \end{aligned}$$

we see that $J(a) = 0$ and $J(b) = 0$.
It remains to show that J is unique: Suppose $\tilde{J}$ is another Jacobi field with $\tilde{J}(a) = 0$ and $\tilde{J}(b) = 0$ such that $W = \tilde{J} + [\beta']$. Then $X := J - \tilde{J} = h\beta'$ for some smooth $h : [a, b] \to \mathbb{R}$. The Jacobi equation for X gives

$$0 = X'' + R(X, \beta')\beta' = h''\beta',$$

hence $h'' = 0$ and thus $h(t) = c_1 t + c_2$. But $X(a) = 0$ and $X(b) = 0$, implying $h = 0$, which means $J = \tilde{J}$. $\square$

We now define conjugate points along null geodesics fully in parallel to the timelike case, cf. Definition 2.3.1.

Definition 2.4.6. (Conjugate points along null geodesics)
Let $\beta : [a, b] \to M$ be a future directed null geodesic. We say that $\beta(t)$, $t \in (a, b]$ is conjugate to $\beta(a)$ along β if there exists a nontrivial Jacobi field $J \in \mathfrak{Jac}(\beta)$ such that $J(a) = 0$ and $J(b) = 0$. Similarly, the conjugacy of two points $\beta(t_1)$ and $\beta(t_2)$ along β is defined.

Note that for the definition of conjugate points we did not require Jacobi classes. The reason for this is the following: Suppose $\beta(t)$ is conjugate to $\beta(a)$, and J is a Jacobi field with $J(a) = 0$ and $J(t) = 0$. We can write

$$J(t) = J_1(t) + c_1\beta'(t) + c_2 t\beta'(t).$$

Then $J_1(t) \neq 0$, because if $J(t) = c_1\beta'(t) + c_2 t\beta'(t)$, then having two zeros would mean that the affine function $c_1 + c_2 t$ vanishes, hence J would be trivial. So there is no worry of J only having a component in β'-direction and the corresponding Jacobi class being trivial.

Recall that we write $\mathfrak{Jac}_t(\beta)$ for Jacobi fields with $J(a) = 0$ and $J(t) = 0$. Furthermore, we write $\overline{\mathfrak{Jac}}_t(\beta)$ for Jacobi classes W with $W(a) = [\beta'(a)]$ and $W(t) = [\beta'(t)]$.

Lemma 2.4.7. ($\mathfrak{Jac}_t(\beta) \cong \overline{\mathfrak{Jac}}_t(\beta)$)
The projection map $\mathfrak{Jac}_t(\beta) \to \overline{\mathfrak{Jac}}_t(\beta)$ is an isomorphism. In particular, two points $\beta(t_1)$ and $\beta(t_2)$ along β are conjugate along β if and only if there is a nontrivial Jacobi class $W \neq [\beta']$ with $W(t_1) = [\beta'(t_1)]$ and $W(t_2) = [\beta'(t_2)]$.

Proof. The projection map is an isomorphism by Lemma 2.4.5. By this result and what was discussed above, there is a 1-1 correspondence between nontrivial Jacobi fields $J \neq 0$ vanishing at two points and nontrivial Jacobi classes $W \neq [\beta']$ which are equal to $[\beta']$ at those two points. □

For null geodesics, we may also define an index form along the lines of Definition 2.3.3, but it will turn out that we will have to resort to its quotient form to get meaningful results.

Definition 2.4.8. (Null index form)
Let $\beta : [a, b] \to M$ be a null geodesic. Then the symmetric bilinear form $I : \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta) \times \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta) \to \mathbb{R}$ defined by

$$I(X, Y) := -\int_a^b \big(\langle X', Y'\rangle - \langle R(X, \beta')\beta', Y\rangle\big)\, dt$$

is called *null index form* along β.

Remark 2.4.9. (Alternative formulas for the null index form)
Just like for the timelike index form (see Remark 2.3.4), it holds that

$$I(X, Y) = \int_a^b \langle X'' + R(X, \beta')\beta', Y\rangle dt + \sum_{i=0}^{k} \langle \delta X'(t_i), Y\rangle,$$

where $a = t_0 < \cdots < t_k = b$ are such that X is smooth on the subintervals $[t_i, t_{i+1}]$.

In the context of null geodesics β, we are more interested in the energy functional of variations of β rather than the arc-length (because given a variation of β, the arc length $L(s)$ would not be differentiable at $s = 0$). However, since we are interested in results about maximality among neighboring curves etc., we will be primarily interested in the following type of variations.

Definition 2.4.10. (Admissible variations)
A piecewise smooth variation $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ is called *admissible* if all $\alpha(., s)$, $s \neq 0$, are timelike.

Remark 2.4.11. (Second variation of energy and the null index form)
Let $\beta : [a, b] \to M$ be a null geodesic. From the results on the variation of energy in Section 2.2, we conclude the following: If $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ is an admissible variation of β with variation field $W \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta)$, then

$$\frac{d}{ds}|_{s=0} E(\alpha(., s)) = 0, \quad \frac{d^2}{ds^2}|_{s=0} E(\alpha(., s)) = I(W, W).$$

Since $E(\beta) = 0$ and $E(\alpha(., s)) > 0$, we must have $I(W, W) \geq 0$ by elementary calculus. Hence for any $W \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta)$, $I(W, W) \geq 0$ is a necessary condition for W to be the variation vector field of an admissible variation of β.

Defining the index form for null geodesics in the same way as we did for timelike geodesics has obvious drawbacks: If $X = f\beta'$ for $f : [a, b] \to \mathbb{R}$ smooth with $f(a) = 0 = f(b)$, then $I(X, Y) = 0$ for all $Y \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(\beta)$. Hence, we cannot obtain an analogue of Theorem 2.3.9 for null geodesics in this way. Instead, we shall project the index form to sections on $Q\beta$.

Definition 2.4.12. (Quotient null index form)
The symmetric bilinear form $I : \overline{\mathfrak{X}}^{\perp}_{\mathrm{pw}}(\beta) \times \overline{\mathfrak{X}}^{\perp}_{\mathrm{pw}}(\beta) \to \mathbb{R}$ defined by

$$\overline{I}(V, W) := -\int_a^b \left(\overline{g}(V', W') - \overline{g}(\overline{R}(V, \beta')\beta', W)\right) dt,$$

is called the *quotient index form* of β.

Remark 2.4.13. (Alternative formulas for the quotient index form)
Using the definition of the quotient objects (see Remark 2.4.1), it is straightforward to obtain alternative formulas for the quotient null index form from the respective formulas for the usual null index form (see Remark 2.4.9):

$$\overline{I}(V, W) = \int_a^b \overline{g}(V'' + \overline{R}(V, \beta')\beta', W)dt + \sum_{i=0}^{k} \overline{g}(\delta V'(t_i), W(t_i)).$$

For $V \in \overline{\mathfrak{X}}^{\perp}(\beta)$ (i.e. smooth), we obtain

$$\overline{I}(V, W) = -\overline{g}(V', W)|_a^b + \int_a^b \overline{g}(V'' + \overline{R}(V, \beta')\beta', W)dt.$$

Moreover, if $V \in \overline{\mathfrak{Jac}}(\beta)$ is a Jacobi class along β, then

$$\overline{I}(V, W) = -\overline{g}(V', W)|_a^b.$$

Also, by the discussion preceding Definition 2.4.12, we have

$$\overline{I}(X + [\beta'], Y + [\beta']) = I(X, Y)$$

for all $X, Y \in \mathfrak{X}^{\perp}_{\text{pw}}(\beta)$, hence the quotient index form is compatible with the quotient construction.

We have the following characterization of Jacobi classes akin to Proposition 2.3.7 in the timelike case.

Proposition 2.4.14. (Jacobi classes and the quotient index form)
Let $\beta : [a, b] \to M$ be a null geodesic. A piecewise smooth vector class $W \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$ is a Jacobi class (hence smooth) if an only if $\overline{I}(W, Z) = 0$ for all $Z \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$.

Proof. If W is a Jacobi class, then it is smooth and satisfies the projected Jacobi equation, hence the result follows from the alternative formula for the index form (see Remark 2.4.13).
Conversely, suppose $I(W, Z) = 0$ for all $Z \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$. Let $W = Y + [\beta']$ with Y being the unique normal vector field along β with no β'-direction (recall the isomorphism $Q\beta \cong S\beta$). Let $a = t_0 < \cdots < t_k = b$ be such that Y is smooth on each $[t_j, t_{j+1}]$ and let $f : [a, b] \to \mathbb{R}$ be smooth such that $f(t_j) = 0$ for each j and $f > 0$ otherwise. Let $Z \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$ be the projection of $f(Y'' + R(Y, \beta')\beta') \in \mathfrak{X}^{0,\perp}_{\text{pw}}(\beta)$. Then the alternative representation of the index form (cf. Remark 2.4.13) together with the assumption gives

$$0 = I(W, Z) = \int_a^b f(t)\overline{g}(W'' + \overline{R}(W, \beta')\beta', W'' + \overline{R}(W, \beta')\beta')dt.$$

Since the projected metric $\overline{g}$ is positive definite (by the isomorphism $Q\beta \cong S\beta$), it follows that W satisfies the projected Jacobi equation everywhere except maybe at the t_j. So it remains to show that W is smooth, i.e. $\delta W'(t_i) = 0$ for $i = 1, \ldots, k - 1$. First of all, note that the assumption reduces to

$$0 = \overline{I}(W, Z) = \sum_{i=0}^{k} \overline{g}(\delta W'(t_i), Z(t_i))$$

for all $Z \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$. Since, by choice, Y is a piecewise smooth section of $S\beta$, we have $\langle Y, \beta' \rangle = 0 = \langle Y', \eta \rangle$ (cf. Remark 2.4.1 for the notation). Then also $\langle Y', \beta' \rangle = 0 = \langle Y', \eta \rangle$ except maybe at t_i, and hence, taking left and right limits, $\langle \delta Y'(t_i), \beta'(t_i) \rangle = 0 = \langle \delta Y'(t_i), \eta(t_i) \rangle$. We can easily construct a smooth vector field X_j with $X_j(a) = 0$, $X_j(b) = 0$, $X_j(t_j) = \delta Y'(t_j)$, $X_j(t_i) = 0$ if $i \neq j$, such that X_j is normal to β' and η (i.e. X_j is a section of $S\beta$) and $X(t_j) = \delta Y'(t_j)$ (e.g. via parallel translation and scaling, see the proof of Proposition 2.3.7). Let $Z_j \in \overline{\mathfrak{X}}^{0,\perp}_{\mathrm{pw}}(\beta)$ be the projection of X_j. Then, since the quotient index form agrees with the usual null index form applied to a representative (see again Remark 2.4.13), we get

$$0 = \overline{I}(W, Z_j) = \langle \delta Y'(t_j), \delta Y'(t_j) \rangle,$$

which implies $\delta Y'(t_j) = 0$ by positive definiteness. Since j was arbitrary, we are done. □

We can define the notion of $(1,1)$-*tensor classes normal to* β in a similar way as in the timelike case as smooth maps $\overline{A} : [a,b] \to \mathrm{Hom}(Q\beta, Q\beta)$ such that $\pi \circ A = \beta$. They are precisely projections of usual $(1,1)$-tensors normal to β that satisfy $A\beta' = f\beta'$. In particular, covariant differentiation and similar notions make sense for such tensor classes.
Since the projected metric is positive definite, one may also define the adjoint. It is just the projection of an adjoint of a representative.

Definition 2.4.15. (Jacobi tensor class)
Let $\beta : [a,b] \to M$ be a null geodesic. A $(1,1)$-tensor class $\overline{A}$ normal to β is called *Jacobi tensor class* if it satisfies the *projected Jacobi tensor equation*

$$\overline{A}'' + \overline{RA} = 0$$

and has the kernel intersection property

$$\ker \overline{A}(t) \cap \ker \overline{A}'(t) = \{[\beta'(t)]\}$$

for all $t \in [a,b]$.

Most results for Jacobi tensors along timelike geodesics (see Subsection 2.3.3) continue to hold with more or less the same proofs for Jacobi tensor classes along null geodesics. We give a summary in the following.

Remark 2.4.16. (Jacobi and Lagrange tensor classes: properties)
Let $\beta : [a,b] \to M$ be a null geodesic.

(1) If $\overline{A}$ is a Jacobi tensor class and $Y \in \overline{\mathfrak{X}}^{\perp}(\beta)$ is parallel, then $\overline{A}(Y)$ is a Jacobi vector class along β. If $Y \neq [\beta']$, then $\overline{A}(Y) \neq [\beta']$.

(2) The *Wronskian* $W(\overline{A}, \overline{B})$ for two Jacobi tensor classes $\overline{A}, \overline{B}$ is defined via

$$W(\overline{A}, \overline{B}) := (\overline{A}')^{\dagger}\overline{B} - \overline{A}^{\dagger}\overline{B}'.$$

The Wronskian is always constant in the sense that $(W(\overline{A}, \overline{B}))' = 0$. A Jacobi tensor class $\overline{A}$ is called *Lagrange tensor class* if $W(\overline{A}, \overline{A}) = 0$.

(3) The initial value problem for the projected Jacobi tensor equation is uniquely solvable given initial data $\overline{A}(t_0)$ and $\overline{A}'(t_0)$. If one of the initial data is nonsingular, then the unique solution is a Jacobi tensor class.

(4) Two points $\beta(t_0)$ and $\beta(t_1)$ on β are conjugate along β if and only if the unique Jacobi tensor class $\overline{A}$ with $\overline{A}(t_0) = 0$, $\overline{A}'(t_0) = \mathrm{id}$ satisfies $\ker \overline{A}(t_1) \neq \{[\beta'(t_1)]\}$.

(5) The boundary value problem for the projected Jacobi tensor equation at parameters t_0 and t_1 is uniquely solvable for any two given boundary data if and only if $\beta(t_0)$ and $\beta(t_1)$ are not conjugate along β.

(6) A $(1,1)$-tensor class satisfying the projected Jacobi tensor equation is already a Jacobi tensor class if it satisfies the kernel intersection property at one parameter value.

(7) If $\overline{A}$ is a Jacobi tensor class vanishing at some parameter, then $\overline{A}$ is a Lagrange tensor class.

(8) Let $\overline{A}$ be a Jacobi tensor class and let $\overline{B} := \overline{A}'\overline{A}^{-1}$ wherever defined. Then $\overline{B}$ satisfies the projected *matrix Riccati equation*

$$\overline{B}' + \overline{B}^2 + \overline{R} = 0.$$

The *expansion* $\overline{\theta}$, *vorticity tensor* $\overline{\omega}$ and *shear tensor* $\overline{\sigma}$ are defined precisy as in the timelike case (see Definition 2.3.21), and the *Raychaudhuri equation* holds:

$$\overline{\theta}' = -\operatorname{Ric}(\beta', \beta') - \operatorname{Tr}(\overline{\omega}^2) - \operatorname{Tr}(\overline{\sigma}^2) - \frac{\overline{\theta}}{n-2}. \tag{2.9}$$

Note that the different factor in the last term is due to $n - 2 = \dim Q\beta|_t$. Moreover, if $\overline{A}$ is a Lagrange tensor class, then the vorticity $\overline{\omega}$ vanishes and the *vorticity-free Raychaudhuri equation* holds:

$$\overline{\theta}' = -\operatorname{Ric}(\beta', \beta') - \operatorname{Tr}(\overline{\sigma}^2) - \frac{\overline{\theta}}{n-2}. \tag{2.10}$$

Proposition 2.4.17. (No conjugate points imply definiteness of $\overline{I}$)
Let $\beta : [a,b] \to M$ be a null geodesic such that no $\beta(t)$, $t \in (a,b]$, is conjugate to $\beta(a)$ along β. Then the quotient index form $\overline{I} : \overline{\mathfrak{X}}^{0,\perp}_{\mathrm{pw}}(\beta) \times \overline{\mathfrak{X}}^{0,\perp}_{\mathrm{pw}}(\beta) \to \mathbb{R}$ is negative definite.

Proof. Let $\overline{A}$ be the unique Jacobi tensor class satisfying $\overline{A}(a) = 0$, $\overline{A}'(a) = id$. Since $\beta(a)$ has no conjugate points along β by assumption, $\overline{A}$ is everywhere nonsingular.

Let $W \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$, $W \neq [\beta']$. Since $\overline{A}(a) = 0$ and $\overline{A}$ is nonsingular elsewhere, we may find $Z \in \overline{\mathfrak{X}}^{\perp}_{\text{pw}}(\beta)$ with $\overline{A}(Z) = W$. Using product rules for covariant derivatives and the projected Jacobi tensor equation, we find

$$\begin{aligned}
\overline{I}(W,W) &= -\int_a^b \big(\overline{g}(W',W') - \overline{g}(\overline{R}(W,\beta')\beta',W)\big)\,dt \\
&= -\int_a^b \big(\overline{g}((\overline{A}(Z))',(\overline{A}(Z))') - \overline{g}(\overline{RA}(Z),\overline{A}(Z))\big)\,dt \\
&= -\int_a^b \Big(\overline{g}(\overline{A}'(Z),\overline{A}'(Z)) + 2\overline{g}(\overline{A}'(Z),\overline{A}(Z')) \\
&\qquad + \overline{g}(\overline{A}(Z'),\overline{A}(Z')) + \overline{g}(\overline{A}''(Z),\overline{A}(Z))\Big)\,dt.
\end{aligned}$$

We "simplify" the last term in the integral as follows:

$$\begin{aligned}
\overline{g}(\overline{A}''(Z),\overline{A}(Z)) = \Big(\overline{g}(\overline{A}'(Z),\overline{A}(Z))\Big)' &- \overline{g}(\overline{A}'(Z'),\overline{A}(Z)) \\
&- \overline{g}(\overline{A}'(Z),\overline{A}(Z')) - \overline{g}(\overline{A}'(Z),\overline{A}'(Z)).
\end{aligned}$$

Using this, we get

$$\begin{aligned}
\overline{I}(W,W) = -\overline{g}(\overline{A}'(Z),\overline{A}(Z))|_a^b &- \int_a^b \big(\overline{g}(\overline{A}(Z'),\overline{A}(Z')) \\
&+ \overline{g}(\overline{A}'(Z),\overline{A}(Z')) - \overline{g}(\overline{A}'(Z'),\overline{A}(Z))\Big)\,dt.
\end{aligned}$$

The first term vanishes since $\overline{A}(Z) = W \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$. The third and fourth $\overline{g}$-terms can be put together via adjoints, so in total we get

$$\overline{I}(W,W) = -\int_a^b (\overline{g}(\overline{A}(Z'),\overline{A}(Z')) + \int_a^b \overline{g}(\overline{A}^\dagger\overline{A}'(Z) - (\overline{A}')^\dagger\overline{A}(Z),Z')dt,$$

where the second term vanishes because $A(a) = 0$ and hence $\overline{A}$ is a Lagrange tensor class. Thus we remain only with

$$\overline{I}(W,W) = -\int_a^b \overline{g}(\overline{A}(Z'),\overline{A}(Z'))dt.$$

Note that Z cannot be parallel because otherwise $\overline{A}(Z) = W \in \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta)$ would be a nontrivial Jacobi tensor class, which is impossible since $\beta(b)$ is not conjugate to $\beta(a)$. Hence Z' is nonzero on some interval, and since $\overline{A}$ is nonsingular and $\overline{g}$ is positive definite, we get $\overline{I}(W,W) < 0$. $\square$

Theorem 2.4.18. ($\overline{I}$ definite iff no conjugate points)
Let $\beta : [a,b] \to M$ be a null geodesic. Then there are no conjugate points to $\beta(a)$ along β if and only if $\overline{I} : \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta) \times \overline{\mathfrak{X}}^{0,\perp}_{\text{pw}}(\beta) \to \mathbb{R}$ is negative definite.

Proof. One direction is precisely Proposition 2.4.17. For the converse direction, extend a nontrivial Jacobi vector class trivially to get an element in $\overline{\mathfrak{X}}^{0,\perp}_{\mathrm{pw}}(\beta)$ on which $\overline{I}$ vanishes, this is completely analogous to the proof of Theorem 2.3.9. $\square$

The following result is again proven exactly as in the timelike case (cf. Proposition 2.3.10).

Proposition 2.4.19. (Maximality of Jacobi vector classes)
Let $\beta : [a, b] \to M$ be a null geodesic without conjugate points to $\beta(a)$. Let $J \in \overline{\mathfrak{X}}^{\perp}(\beta)$ be a Jacobi vector class along β. Then for any $Y \in \overline{\mathfrak{X}}^{\perp}_{\mathrm{pw}}(\beta)$, $Y \neq J$, satisfying $Y(a) = J(a)$ and $Y(b) = J(b)$, we have

$$\overline{I}(J, J) > \overline{I}(Y, Y).$$

In the timelike case, cf. Proposition 2.3.6, we showed that if a timelike geodesic has a conjugate point then there is an FEP-variation with longer curves, and this was simply done finding a vector field on which the index form is positive.
The situation is more complicated for null geodesics. Unlike in the case of timelike geodesics, the nearby geodesics in a variation will have all kinds of causal characters. Since we are interested in maximality, we introduced the notion of *admissible variations* for null geodesics, see Definition 2.4.10. But for a given vector field $V \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(\beta)$, the standard variation it defines need not be an admissible one. So the problem of maximality of a null geodesic, i.e. whether or not there is an admissible FEP-variation, requires a more detailed analysis. We start by describing sufficient conditions for the existence of an admissible variation.

Proposition 2.4.20. (Sufficient conditions for admissible variations)
Let $\beta : [0, 1] \to M$ be a null geodesic. Let $\alpha : [0, 1] \times (-\varepsilon, \varepsilon) \to M$ be a piecewise smooth FEP-variation of β. We write $V(t_0, s_0) := \partial_s|_{s=s_0}\alpha(t_0, s)$, $T(t_0, s_0) := \partial_t|_{t=t_0}\alpha(t, s_0)$, and $V(t) = V(t, 0)$, $T(t) = T(t, 0)$. Suppose the following conditions hold:

(1) $\langle V(t), \beta'(t)\rangle = 0$ for all $t \in [0, 1]$.

(2) There is $c > 0$ such that for all $t \in (0, 1)$ for which $V(t)$ is smooth,

$$\begin{aligned}&\frac{d}{dt}\left(\langle \nabla_s V(t, s), \beta'(t)\rangle|_{s=0} + \langle V(t), V'(t)\rangle\right)\\ &- \langle V(t), V''(t) + R(V(t), \beta'(t))\beta'(t)\rangle < -c.\end{aligned}$$

Then α is an admissible variation if restricted to a small enough subinterval of $(-\varepsilon, \varepsilon)$.

Proof. The idea is to use basic calculus: We need to show that there is a strict local maximum of $\langle T, T\rangle$ at $s = 0$, then our claim is proven, noting that $T(t,s) = \partial_t\alpha(t,s)$ is the velocity vector of the curve $\alpha(.,s)$ at parameter t. Let us derive the necessary formulas for the derivatives. We repeatedly make use of Lemma 2.1.3 and Lemma 2.1.4.

$$\begin{aligned}\frac{d}{ds}\langle T,T\rangle &= 2\langle \nabla_s T, T\rangle = 2\langle \nabla_s \partial_t \alpha(t,s), T\rangle \\ &= 2\langle \nabla_t \partial_s \alpha(t,s), T\rangle = 2\langle \nabla_t V, T\rangle \\ &= 2\frac{d}{dt}\langle V, T\rangle - 2\langle V, \nabla_t T\rangle.\end{aligned}$$

Since $\nabla_t T(t,0) = \beta''(t) = 0$, taking $s = 0$ here gives via condition (1)

$$\left.\frac{d}{ds}\right|_{s=0}\langle T,T\rangle = 2\frac{d}{dt}\langle V,T\rangle|_{s=0} = 0.$$

Note that requiring $\langle V,T\rangle|_{s=0} = 0$ or the vanishing of its derivative with respect to t are equivalent, since $V(t)$ vanishes as $t = 0$ and $t = 1$. Taking the second derivative, we obtain

$$\left.\frac{d^2}{ds^2}\right|_{s=0}\langle T,T\rangle = 2\frac{d}{dt}\left.\frac{d}{ds}\right|_{s=0}\langle V,T\rangle - 2\langle \nabla_s V, \nabla_t T\rangle|_{s=0} - 2\langle V, \nabla_s\nabla_t T\rangle|_{s=0}.$$

Note that by Lemma 2.1.4

$$\nabla_s\nabla_t T - \nabla_t\nabla_s T = R(\partial_s\alpha, \partial_t\alpha)T = R(V,T)T$$

and moreover

$$\nabla_s\nabla_t T - \nabla_t\nabla_s T = \nabla_s\nabla_t T - \nabla_t\nabla_t V = R(V,T)T,$$

hence

$$\nabla_s\nabla_t T = \nabla_t\nabla_t V + R(V,T)T.$$

Finally, we note that $\nabla_t T|_{s=0} = 0$ since β is null. Combining all of these observations, we can continue our calculation of the second derivative:

$$\begin{aligned}\left.\frac{d^2}{ds^2}\right|_{s=0}\langle T,T\rangle &= 2\frac{d}{dt}\left(\langle \nabla_s V, T\rangle + \langle V, \nabla_s T\rangle\right)|_{s=0} \\ &\quad - 2\langle V, \nabla_t\nabla_t V + R(V,T)T\rangle|_{s=0}.\end{aligned}$$

Using $\nabla_s T = \nabla_t V$, we see that our condition (2) precisely says that the second derivative is negative at $s = 0$, which means that $T(t,s)$ is timelike for small $|s|$, $s \neq 0$, hence α is an admissible variation. $\square$

Theorem 2.4.21. (Conjugate points and maximality of null geodesics) Let $\beta : [0,1] \to M$ be a null geodesic. If there are conjugate points to $\beta(0)$ along β, then there is a timelike curve from $\beta(0)$ to $\beta(1)$. In particular, β is not maximizing if it has conjugate points.

Proof. Suppose $\beta(t_0)$ is the first conjugate point to $\beta(0)$ along β (note that there is a first conjugate point by Proposition 2.3.2, which is easily seen to hold for null geodesics as well, because β lies in a convex set around $\beta(0)$ for a short while, where exp is a diffeomorphism). Due to the push-up principle, it is sufficient to show that there is some $t_2 \in (t_0, 1]$ such that there exists a timelike curve from $\beta(0)$ to $\beta(t_2)$.
Let W be a nontrivial Jacobi vector class with $W(0) = [\beta'(0)]$ and $W(t_0) = [\beta'(t_0)]$. Since the projected metric $\overline{g}$ is positive definite, we may write

$$W(t) = f(t)\hat{W}(t)$$

where $\hat{W} \in \overline{\mathfrak{X}}^{\perp}(\beta)$ with $\overline{g}(\hat{W}, \hat{W}) = 1$ and $f : [0,1] \to \mathbb{R}$ is smooth. Note that $f(0) = f(t_0) = 0$ and $f \neq 0$ in $(0, t_0)$, we may assume that $f > 0$ there. Since W is nontrivial, $W'(t_0) \neq [\beta'(t_0)]$. Since

$$W'(t_0) = f'(t_0)\hat{W}(t_0) + f(t_0)\hat{W}'(t_0) = f'(t_0)\hat{W}(t_0),$$

thus $f'(t_0) \neq 0$, so there is some $t_1 \in (t_0, 1]$ such that $f < 0$ on $(t_0, t_1]$, in particular $W \neq [\beta']$ on $(t_0, t_1]$. For the rest of the proof, we intend to construct an admissible FEP-variation for $\beta|_{[0,t_2]}$, which would prove our claim.
We want to make use of Proposition 2.4.20. Consider the following ansatz for a vector class:

$$\overline{Z}(t) := (b(e^{at} - 1) + f(t))\hat{W}(t).$$

Here, $a > 0$ is such that for $h(t) := \overline{g}(\hat{W}'' + \overline{R}(\hat{W}, \beta')\beta', \hat{W})$,

$$a^2 > -\min_{t \in [0,t_1]} \{h(t)\},$$

and b is the constant

$$b := -\frac{f(t_1)}{e^{at_1} - 1}.$$

From the projected Jacobi equation for W and $\overline{g}(\hat{W}, \hat{W}) = 1$ (and hence $\overline{g}(\hat{W}', \hat{W}) = 0$), we obtain

$$\begin{aligned} 0 &= \overline{g}(W'' + \overline{R}(W, \beta')\beta', \hat{W}) = \overline{g}(f''\hat{W} + 2f'\hat{W}' + f\hat{W}'' + f\overline{R}(\hat{W}, \beta')\beta', \hat{W}) \\ &= f'' + 2f'\overline{g}(\hat{W}, \hat{W}) + fh = f'' + fh. \end{aligned}$$

Next, observe that $\overline{Z}$ vanishes at 0 and t_1 by choice of the constant b. Set $r(t) := b(e^{at}-1)+f(t)$. Thus $\overline{Z} = r\hat{W}$, and thus $\overline{Z}'' = r''\hat{W}+2r'\hat{W}'+r\hat{W}''$. This yields

$$\begin{aligned}
\overline{g}(\overline{Z},\overline{Z}'' + \overline{R}(\overline{Z},\beta')\beta') &= \overline{g}(r\hat{W}, r''\hat{W} + 2r'\hat{W} + r\hat{W}'' + r\overline{R}(\hat{W},\beta')\beta') \\
&= rr'' + r^2h = r(r''+h) \\
&= r(ba^2e^{at} + f'' + b(e^{at}-1)h + fh) \\
&= r(be^{at}(a^2+h) - bh + f'' + fh) \\
&= rb(e^{at}(a^2+h) - h).
\end{aligned}$$

Note that $b > 0$ since $f(t_1) < 0$. Also, $e^{at}(a^2+h) - h > a^2 > 0$ on $[0,t_1]$ by choice of a, so if we show that $r(t) > 0$, we would have that

$$\overline{g}(\overline{Z},\overline{Z}'' + \overline{R}(\overline{Z},\beta')\beta') > 0. \tag{2.11}$$

Since $f > 0$ on $(0,t_0)$, we have $r > 0$ there as well (since the first summand is anyway always nonnegative). In fact $r > 0$ even on $(0,t_0]$ and hence on some $(0,t_2)$ for $t_2 > t_0$ and $r(t_2) = 0$. If $t_2 \geq t_1$, then $t_2 = t_1$ since $r(t_1) = 0$. If $t_2 < t_1$, we can just consider the restriction $\overline{Z}|_{[0,t_2]}$ which will then satisfy $\overline{Z}(0) = [\beta'(0)]$, $\overline{Z}(t_2) = [\beta'(t_2)]$, and $\overline{g}(\overline{Z},\overline{Z}'' + \overline{R}(\overline{Z},\beta')\beta') > 0$ on $(0,t_2)$. Either way, we henceforth consider $\beta|_{[0,t_2]}$.
Let $\tilde{Z} \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(\beta|_{[0,t_2]})$ be such that it projects down to the vector class $\overline{Z}$. Then $\tilde{Z}(0) = \mu\beta'(0)$ and $\tilde{Z}(t_2) = \lambda\beta'(t_2)$, $\mu, \lambda \in \mathbb{R}$. Setting

$$Z := \tilde{Z} - \mu\beta' + \frac{\mu-\lambda}{t_2}t\beta',$$

Z vanishes at 0 and t_2 and projects down to $\overline{Z}$ as well. Since (2.11) holds for $\overline{Z}$ on $(0,t_2)$, it follows that

$$p(t) := g(Z'' + R(Z,\beta')\beta', Z) > 0 \quad \text{on } (0,t_2).$$

Hence we may choose $\varepsilon > 0$ such that

$$\varepsilon < \min_{t\in[t_2/4, 3t_2/4]}\{p(t)\}.$$

We define the following function:

$$\rho(t) := \begin{cases} -\varepsilon t & 0 \leq t \leq \frac{t_2}{4}, \\ \varepsilon\left(t - \frac{t_2}{2}\right) & \frac{t_2}{4} \leq t \leq \frac{3t_2}{4}, \\ \varepsilon(t_2 - t) & \frac{3t_2}{4} \leq t \leq t_2. \end{cases}$$

Then ρ is continuous, piecewise smooth, and vanishes at 0 and t_2. Recall from Remark 2.4.1 that we fixed a parallel null field η along β with $\langle \eta, \beta' \rangle =$

-1 and spacelike parallel fields $E_1, \ldots, E_{n-2}$ that are orthonormal among each other and normal to η and β'. We can now find an FEP-variation $\alpha : [0, t_2] \times (-\varepsilon, \varepsilon) \to M$ by requiring

$$\partial_s|_{s=0}\alpha = Z(t), \quad (\nabla_s \partial_s \alpha)|_{s=0} = \big(\langle Z, Z' \rangle - \rho\big)\, \eta.$$

Writing out the above equations in terms of $E_1, \ldots, E_{n-2}, \eta, \beta'$, ODE theory yields a solution $s \mapsto \alpha(t, s)$ for every $t \in [0, t_2]$, and by smooth dependence on initial data, these fit together to give an FEP variation $\alpha(t, s)$.
Let $T := \partial_t \alpha$, $V := \partial_s \alpha$, i.e. $Z(t) = V(t, 0)$. Then by the initial conditions,

$$\langle \partial_s V, \beta' \rangle|_{s=0} + \langle V(t,s), \partial_t V(t,s) \rangle|_{s=0} = \rho(t).$$

Thus,

$$\frac{d}{dt}\big(\langle \partial_s V, \beta' \rangle| + \langle V(t,s), \partial_t V(t,s) \rangle\big)\,|_{s=0} = \rho'(t).$$

Since ρ' is $-\varepsilon$ on the first and third subinterval of its definition, the choice of ε that we made, together with the fact that $p(t) > 0$ on $(0, t_2)$ imply that condition (2) in Proposition 2.4.20 holds. Also,

$$\langle \partial_s \alpha, \beta' \rangle|_{s=0} = \langle Z, \beta' \rangle = 0$$

by construction of Z. Hence condition (1) in Proposition 2.4.20 is also satisfied, and α is (for small enough s-parameters) an admissible FEP-variation of $\beta|_{[0,t_2]}$. □

2.4.3 The null Morse index theorem

There is also a version of the Lorentzian Morse Index Theorem for null geodesics (see Subsection 2.3.2 for a treatment of the timelike version). The proofs are completely analogous to the timelike case, replacing any space of (piecewise) Jacobi fields that appear there with the corresponding space of (piecewise) Jacobi classes. There is no reason to replicate the proofs here.

Definition 2.4.22. (Index of a null geodesic)
Let $\beta : [a, b] \to M$ be a null geodesic. Then the *index* $\mathrm{Ind}(\beta)$ is the supremum of dimensions $\dim A$, where A is a vector subspace of $\overline{\mathfrak{X}}^{0,\perp}_{\mathrm{pw}}(\beta)$ such that the quotient index form $\overline{I}$ is positive definite on $A \times A$. The *extended index* $\mathrm{Ind}_0(\beta)$ is defined analogously, except that one considers subspaces A such that $\overline{I}$ is positive semidefinite on $A \times A$.

Proposition 2.4.23. (Index, extended index and Jacobi classes)
The index and extended index of β are both finite. Moreover, it holds that

$$\mathrm{Ind}_0(\beta) = \mathrm{Ind}(\beta) + \dim \overline{\mathfrak{Jac}}_b(\beta).$$

Proof. The proof can be carried over word for word as in Proposition 2.3.12, if one replaces $\mathfrak{Jac}_t$ with $\overline{\mathfrak{Jac}}_t$ and $\mathfrak{Jac}\{t_i\}$ with $\overline{\mathfrak{Jac}}\{t_i\}$. $\square$

Theorem 2.4.24. (Lorentzian Morse Index Theorem, null version)
Let $\beta : [a, b] \to M$ be a null geodesic. Then β only has finitely many conjugate points. The index and extended index of β are given by

$$\begin{aligned} \mathrm{Ind}(\beta) &= \sum_{t\in(a,b)} \dim \mathfrak{Jac}_t(\beta), \\ \mathrm{Ind}_0(\beta) &= \sum_{t\in(a,b]} \dim \mathfrak{Jac}_t(\beta). \end{aligned}$$

Proof. Completely analogous to the proof of Theorem 2.3.13, if one works with the corresponding quotient objects. One then gets the above formulas for Ind and Ind_0 only with $\dim \overline{\mathfrak{Jac}}_t(\beta)$-terms on the right hand side, but by Lemma 2.4.7, $\dim \mathfrak{Jac}_t(\beta) = \dim \overline{\mathfrak{Jac}}_t(\beta)$. $\square$

Chapter 3

Singularity Theorems

In this final chapter, we prove the three classical singularity theorems of general relativity, namely the theorems of Hawking, Penrose, and Hawking-Penrose. They yield (causal/timelike/null) geodesic incompleteness under fairly generic circumstances. Before proving the singularity theorems, we provide a collection of further preparatory results, where we sometimes refer to the literature if the intricate proofs would lead us too far afield. For the singularity theorems of Hawking and Penrose, we follow [22, Ch. 14], and for the Hawking-Penrose theorem we follow [1, Ch. 12]. The preparatory results have been collected from [14], with the occasional input from [11] and [3].

3.1 Causality theory

In this preparatory first section of the final chapter, we give a treatment of various important topics and notions from causality theory that will play key roles in the formulation and interpretation of the singularity theorems. Throughout this section, let (M, g) be a spacetime.

3.1.1 Cauchy surfaces, Cauchy developments and horizons

Cauchy surfaces are subsets of spacetime that divide spacetime disjointly into past, present and future. It can be deduced that they are topological hypersurfaces that allow for little topological variance: In fact, any two Cauchy surfaces in a given spacetime are homeomorphic, Moreover, the existence of one is equivalent to global hyperbolicity of the spacetime, which represents the upper end of the causal ladder.

Definition 3.1.1. (Cauchy surface)
A subset $C \subseteq M$ is called *Cauchy surface* if every inextendible causal curve in M meets C exactly once.

Lemma 3.1.2. (Basic properties of Cauchy surfaces)
Let $C \subseteq M$ be a Cauchy surface. Then C is an acausal (i.e. no two distinct points in C are causally related), topologically closed C^0-hypersurface. Moreover, M is the disjoint union of $I^-(C), C$ and $I^+(C)$, and $C = \partial I^-(C) = \partial I^+(C)$.

Proof. If $p, q \in C$ such that $p < q$, extend a causal curve connecting p and q to an inextendible causal curve meeting C twice, a contradiction. Hence C is acausal.
If we show that $\mathrm{edge}(C) = \emptyset$, then C is a closed C^0-hypersurface by Corollary 1.3.6.
To show that $\mathrm{edge}(C) = \emptyset$, we first show that M is the disjoint union of $I^-(C)$, C and $I^+(C)$. Disjointness is clear from achronality. Now let $p \in M$, and let c be an inextendible timelike curve through p. By assumption, c meets C exactly once at some $q \in c \cap C$. Then either $p \in C$, $p \in I^-(C)$ or $p \in I^+(C)$.
Next, we argue that $C = \partial I^{\pm}(C)$. $\subseteq$ is generally true since C is achronal, so we only need to show $\supseteq$. From the disjoint union of M as above, we gather

$$\partial I^+(C) = \overline{I^+(C)} \cap \overline{M \setminus I^+(C)} \subseteq (I^+(C) \cup C) \cap (I^-(C) \cup C) = C.$$

An analogous calculation shows $C = \partial I^-(C)$.
These results imply $\mathrm{edge}(C) = \emptyset$, because any timelike curve from $I^-(C)$ to $I^+(C)$ has to meet C because C is their (common) boundary. ☐

Proposition 3.1.3. (Cauchy surfaces are retracts of spacetime)
Let $X \in \mathfrak{X}(M)$ be a timelike vector field and let $C \subseteq M$ be a Cauchy surface. Let $\rho : M \to C$ be defined as follows: For $p \in M$, let $\rho(p)$ be the (unique) point on C and the maximal integral curve of X through p. Then ρ is a continuous, open map with $\rho|_C = id_C$. In particular, any Cauchy surface in M is a retract of M. Moreover, any Cauchy surface is connected (because M is).

Proof. Clearly ρ is well-defined because any such timelike integral curve has a unique meeting point with C. Let Fl^X be the flow map of X, which is defined on some open subset D of $\mathbb{R} \times M$. Clearly, $(\mathbb{R} \times C) \cap D$ is a C^0-hypersurface in D on which Fl^X is continuous and bijective. Since $(\mathbb{R} \times C) \cap D$ and M are both topological manifolds of the same dimension, it follows by Brouwer's invariance of domain theorem (see [4, Thm. 4.5]) that Fl^X gives a homeomorphism $(\mathbb{R} \times C) \cap D \to M$. For $\pi_2 : \mathbb{R} \times M \to M$, $\rho := \pi_2 \circ \mathrm{Fl}_X |^{-1}_{(\mathbb{R} \times C) \cap D}$ is the desired retraction. ☐

Corollary 3.1.4. (Any two Cauchy surfaces are homeomorphic)
Any two Cauchy surfaces $C_1, C_2 \subseteq M$ are homeomorphic.

Proof. Let $\rho_{C_i} : M \to C_i$, $i = 1, 2$, be the retraction maps corresponding to C_1, C_2 according to Proposition 3.1.3. By construction via a timelike vector field (see the proof of Proposition 3.1.3), it is easy to see that $\rho_{C_1}|_{C_2}$ and $\rho_{C_2}|_{C_1}$ are inverse homeomorphisms. □

Definition 3.1.5. (Strong causality and global hyperbolicity)
A spacetime (M, g) is called *strongly causal* at $p \in M$ if for every neighborhood $U \ni p$ there is a neighborhood $V \subseteq U$ with $p \in V$ such that every causal curve which starts and ends in V is entirely in U.
The spacetime is *strongly causal* if it satisfies this property at every point.
(M, g) is called *globally hyperbolic* if it is strongly causal and all *causal diamonds* $J^+(p) \cap J^-(q)$, $p, q \in M$, are compact.

Remark 3.1.6. (On Cauchy surfaces and global hyperbolicity)

(1) It is a matter of convention how one defines a Cauchy surface: E.g. in [22], a Cauchy surface is a set that is met exactly once by every inextendible *timelike* curve. We were slightly more restrictive in our definition, but this will not play much of a role.

(2) It may be shown that a spacetime is globally hyperbolic if and only if it contains a Cauchy surface. In this context it does not matter if one defines Cauchy surfaces as we did or as is done in [22]. For details on global hyperbolicity and Cauchy surfaces in a very general setting, see [23]. Globally hyperbolic spacetimes enjoy many nice properties that mirror some of the properties of complete Riemannian manifolds, e.g. any two causally related points may be connected by a maximizing geodesic and the Lorentzian distance is finite and continuous. Moreover, the causal relation $\leq$ is closed in globally hyperbolic spacetimes. We will use many of these standard facts and refer to the above reference and also [22] for details.

(3) In [2], it was shown that any *causal* spacetime (M, g) (i.e. there are no closed causal curves) with compact causal diamonds is already globally hyperbolic. The more recent paper [13] improves on this in the physically reasonable case of non-totally vicious spacetimes (i.e. there is some point through which no closed timelike curve passes) of dimension ≥ 3, showing that such a spacetime is globally hyperbolic if and only if its causal diamonds are compact.

Next, we define the important notion of *Cauchy development* of an achronal set A. Physically, it is interpreted as the portion of spacetime that can be predicted by A. For Cauchy surfaces, we will see that their Cauchy development is all of spacetime.

Definition 3.1.7. (Cauchy development)
Let $A \subseteq M$ be an achronal set. Then its *future Cauchy development*, denoted $D^+(A)$, is the set of all $p \in M$ such that every past inextendible causal curve through p meets A. The *past Cauchy development* $D^-(A)$ is defined similarly via future inextendible causal curves. The *Cauchy development* of A is $D(A) := D^+(A) \cup D^-(A)$.

Lemma 3.1.8. (Basic properties of Cauchy developments)
Let $A \subseteq M$ be achronal.

(1) $A \subseteq D^\pm(A) \subseteq A \cup I^\pm(A) \subseteq J^\pm(A)$.

(2) $D^+(A) \cap I^-(A) = \emptyset$.

(3) $A = D^+(A) \cap D^-(A)$.

(4) $D(A) \cap I^\pm(A) = D^\pm(A) \setminus A$.

Proof.

(1) $A \subseteq D^\pm(A)$ is immediate. Now let $q \in D^+(A) \setminus A$, and let c be a past inextendible timelike curve through q. Then c meets A, hence $q \in I^+(A)$.

(2) Since A is achronal, $A \cap I^\pm(A) = \emptyset$ and $I^+(A) \cap I^-(A) = \emptyset$. Hence by above, $D^+(A) \cap I^-(A) \subseteq (A \cup I^+(A)) \cap I^-(A) = \emptyset$.

(3) $A \subseteq D^+(A) \cap D^-(A) \subseteq (A \cup I^+(A)) \cap (A \cup I^-(A)) = A$ by achronality.

(4) By the above, $D(A) \cap I^\pm(A) = D^\pm(A) \cap I^\pm(A) = D^\pm(A) \setminus A$.

□

Example 3.1.9. (Cauchy development of Cauchy surfaces)
If $C \subseteq M$ is a Cauchy surface, then $D(C) = M$. Indeed, we showed in Lemma 3.1.2 that $M = I^-(C) \cup C \cup I^+(C)$ and this union is disjoint. Since it is generally true that $D^\pm(C) \subseteq I^\pm(C) \cup C$ by Lemma 3.1.8(1), we only need to show the converse inclusion. Suppose $p \in C \cup I^+(C)$ and let γ be any inextendible timelike curve with $p \in \gamma$. Then γ meets C since it is a Cauchy surface. Since $p \in C \cup I^+(C)$, γ meets C in the past by achronality of C, hence $p \in D^+(C)$. This shows that

$$M = I^-(C) \cup C \cup I^+(C) = D^-(C) \cup D^+(C) = D(C).$$

Lemma 3.1.10. (From $D^+(A)^\circ$ into $I^-(A)$)
Let A be achronal. Any past inextendible causal curve passing through some $x \in D^+(A)^\circ$ meets $I^-(A)$.

Proof. Let γ be a past inextendible causal curve starting in $x \in D^+(A)^\circ$. Suppose γ does not meet $I^-(A)$. Then $I^+(\gamma) \cap A = \emptyset$. Consider a sequence $x_j \in \gamma$ such that $x_{j+1} \in J^-(x_j)$, moreover assume that the x_j do not accumulate. Let $x_0 := x$ and $y_0 \in I^+(x_0) \cap D^+(A)^\circ$, the latter being a nonempty open set. There is some $y_1 \in I^-(y_0) \cap I^+_{B_1(x_1)}(x_1)$, where B_ε are balls with respect to some Riemannian metric. Inductively find $y_j \in I^-(y_{j-1}) \cap I^+_{B_{1/j}(x_j)}(x_j)$ and connect the y_j by timelike curves. Concatenating all of them (in the sense of limit curves) gives a past inextendible timelike curve $\mu \subseteq I^+(\gamma)$. Since μ starts from $y_0 \in D^+(A)^\circ$, it must meet A at some y. Then there is some y_j on μ before y, i.e. $y_j \ll y$, and since $y_j \in I^+(x_j)$, we have $x_j \ll y$. Since $x_j \in \gamma$ and $y \in A$, we get $y \in I^+(\gamma) \cap A$, a contradiction to $I^+(\gamma) \cap A = \emptyset$. $\square$

For a brief sketch of the proof of the next result, we follow [21, Thm. 3.45], to which we refer for more details.

Theorem 3.1.11. (Global hyperbolicity of Cauchy developments)
If $A \subseteq M$ is closed and achronal, then $D(A)^\circ$ is globally hyperbolic if it is nonempty. Moreover, if A is a closed, acausal topological hypersurface, then $D(A)$ is open and hence globally hyperbolic.

Proof. Suppose that $A \subseteq M$ is closed and achronal. Let $p, q \in D(A)^\circ$. We begin by showing that the causal diamond $J^+(p) \cap J^-(q)$ *in* M is compact. First, suppose it is not relatively compact, then there are $r_n \in J^+(p) \cap J^-(q)$ escaping every compact set. By connecting p to r_n and r_n to q we get a sequence σ_n of future causal curves from p to q, and we can use a limit curve theorem (see [21, Thm. 2.53]) to get a future inextendible causal curve σ^p starting at p and a past inextendible causal curve σ^q ending at q both of which are uniform limits of a subsequence of σ_n. From Lemma 3.1.10 we may conclude that there are points $\tilde{p} \in \sigma^p \cap I^+(A)$ and $\tilde{q} \in \sigma^q \cap I^-(q)$. Invoking the limit curve theorem [21, Thm. 2.53] once again, we get that $\tilde{p} \ll \tilde{q}$, contradicting achronality of A. Thus, $J^+(p) \cap J^-(q)$ is relatively compact.
Now suppose that $J^+(p) \cap J^-(q)$ is not closed. Then there is a sequence $r_n \in J^+(p) \cap J^-(q)$ and a point $r \in \overline{J^+(p) \cap J^-(q)} \setminus J^+(p) \cap J^-(q)$ such that $r_n \to r$. By connecting p to r_n and then r_n to q we obtain a sequence of future causal curves σ_n, and we may use arguments similar to the ones above to contradict achronality of A, since no sublimit of σ_n can join p to q, otherwise $r \in J^+(p) \cap J^-(q)$. This establishes compactness of $J^+(p) \cap J^-(q)$.
By [21, Prop. 3.43], $D(A)^\circ$ is *causally convex*, i.e. for any $p, q \in D(A)^\circ$ any causal curve connecting p and q must lie entirely in $D(A)^\circ$. From this, we conclude that for $p, q \in D(A)^\circ$ it holds that $J^+(p) \cap J^-(q) \subseteq D(A)^\circ$, i.e. the $D(A)^\circ$-causal diamonds agree with the ones in M.
It remains to argue that strong causality holds at points in $D(A)^\circ$. Suppose there exists $p \in D(A)^\circ$ such that strong causality does not hold at p. Then

there is a neighborhood U of p such that for every neighborhood V of p with $V \subseteq U$ there is a causal curve starting and ending in V which leaves U. Let $x_n \in I^+(p)$ and $y_n \in I^-(p)$ be two sequences converging to p, then $V_n := I^+(y_n) \cap I^-(x_n)$ is a neighborhood of p such that $V_n \subset U$ for large n. By assumption, there is a causal curve with endpoints in V_n that leaves U. Concatenating this curve with a causal curve from y_n to its starting point and from its endpoint to x_n, we may again use limit curve methods to conclude that A is not achronal, a contradiction. This concludes our sketch of the proof that $D(A)^\circ$ is globally hyperbolic if it is nonempty.
For the case of A being a closed, acausal topological hypersurface, we refer to [22, Lem. 14.43]. $\square$

Next, we define the Cauchy horizon of an achronal set A. It is interpreted as the boundary of predictable events if one starts in A.

Definition 3.1.12. (Cauchy horizons)
Let $A \subseteq M$ be achronal. Then the *future Cauchy horizon* is defined as

$$H^+(A) := \overline{D^+(A)} \setminus I^-(D^+(A)) = \{p \in \overline{D^+(A)} : I^+(p) \cap D^+(A) = \emptyset\}.$$

Similarly, the *past Cauchy horizon* is $H^-(A) := \overline{D^-(A)} \setminus I^+(D^-(A))$.

Lemma 3.1.13. (Basic properties of Cauchy horizons)
Let $A \subseteq M$ be achronal.

(1) $H^\pm(A)$ are closed, achronal sets.

(2) If A is closed, then $\overline{D^+(A)}$ is precisely the set of all $p \in M$ such that every past inextendible timelike curve through p meets A.

(3) If A is closed, then $\partial D^\pm(A) = A \cup H^\pm(A)$.

Proof.

(1) Closedness is clear by definition. For achronality, note that $I^+(H^+(A))$ is open and does not meet $D^+(A)$ by definition of $H^+(A)$. Hence it also does not meet $\overline{D^+(A)}$. Since $H^+(A) \subseteq \overline{D^+(A)}$, it follows that $H^+(A) \cap I^+(H^+(A)) = \emptyset$, which is equivalent to achronality of $H^+(A)$.

(2) Suppose now that A is closed. Let T be the set of all $p \in M$ such that every past inextendible timelike curve through p meets A. We show $T = \overline{D^+(A)}$.
$\overline{D^+(A)} \subseteq T$: Suppose $p \in \overline{D^+(A)} \setminus T$, Then there is some past inextendible (future directed) timelike curve $\alpha : (-b, 0]$ with $\alpha(0) = p$ that does not meet A. Let $p_k \in \overline{D^+(A)}$ such that $p_k \to p$. Choose relatively compact convex neighborhoods $U_k \ni p$ with $U_k \to \{p\}$. We may assume that $p_l \in U_k$ for all $l \geq k$ and that $\alpha(-1/k) \ll_{U_k} p_l$ for

all $l \geq k$. Let now α_k be the curve α up to $\alpha(-1/k)$ followed by a timelike curve from $\alpha(-1/k)$ to p_k that stays entirely in U_k. Then the α_k are past inextendible timelike curves from p_k and must hence meet A at some $q_k \in A$. But α does not meet A, so the q_k have to be on the replaced pieces and hence in U_k. Since $U_k \to \{p\}$, $q_k \to p \in A$, since A is closed. But trivially $A \subseteq T$, a contradiction to the assumption $p \notin T$.

$\overline{D^+(A)} \supseteq T$: If $q \notin \overline{D^+(A)}$, choose some $q_0 \in I^-_{M\setminus\overline{D^+(A)}}(q)$. By [22, Lem. 14.30], there is some past inextendible timelike curve from q_0 not meeting $\overline{D^+(A)}$, and hence there is a past inextendible timelike curve from p not meeting $\overline{D^+(A)}$. Since $A \subseteq \overline{D^+(A)}$, we see that $q \notin T$.

(3) We prove this in several steps:

$A \subseteq \partial D^+(A)$: We have $A \subseteq D^+(A)$, so suppose there is some $p \in A \cap D^+(A)^\circ$ and let $q \in I^-(p) \cap D^+(A)^\circ$. Let c be a past inextendible timelike curve starting at q, then c has to meet A at some point r. But then $r \ll q \ll p$ and $p, r \in A$, a contradiction to achronality. $H^+(A) \subseteq \partial D^+(A)$: By definition, $H^+(A) \subseteq \overline{D^+(A)}$. If there were a point $p \in H^+(A) \cap D^+(A)^\circ$, then $I^+(p) \cap D^+(A)$ would be nonempty, which cannot happen since $p \in H^+(A)$.

$\partial D^+(A) \subseteq A \cup H^+(A)$: Suppose $p \in \partial D^+(A) \setminus (A \cup H^+(A))$. Since $p \in \overline{D^+(A)} \setminus A$, then every past inextendible timelike curve from p meets A, so $p \in I^+(A)$. Since $p \notin H^+(A)$, there is some $q \in I^+(p) \cap D^+(A)$, so $p \in I^+(A) \cap I^-(q)$ which is an open set. We now show that $I^-(q) \cap I^+(A) \subseteq D^+(A)$, which would give the contradiction $p \in D^+(A)^\circ$. Indeed, let $r \in I^+(A) \cap I^-(q)$. Let c be a past inextendible causal curve from r and let γ be a past directed timelike curve from q to r. Since $r \in I^+(A)$, $\gamma \subseteq I^+(A)$ and $\gamma \cap A = \emptyset$ by achronality. Since $q \in D^+(A)$, the past inextendible causal curve $\gamma \cup c$ from q meets A, and thus c meets A since γ does not, which implies $r \in D^+(A)$, as desired.

□

Lemma 3.1.14. (On $\partial J^+(A) \setminus \overline{A}$)
For any (achronal) set A, $\partial J^+(A)$ is an achronal C^0-hypersurface. Any $p \in \partial J^+(A) \setminus \overline{A}$ is the future end point of a null geodesic in $\partial J^+(A)$ without conjugate points that is either past inextendible or has a past endpoint in $\mathrm{edge}(A) \subseteq \overline{A}$.

Proof. It follows from Corollary 1.3.7 that $\partial J^+(A)$ is an achronal topological hypersurface, since it is the boundary of the future set $J^+(A)$.
Let $x_k \in I^+(A)$, $x_k \to p$. Choose past-directed timelike curves $\gamma_k : [0, b_k] \to M$ from x_k to $\gamma_k(b_k) \in A$. Since $p \notin \overline{A}$, the γ_k leave a fixed neighborhood

of p, hence by the limit curve theorem (see [20, Thm. 3.1(1)]) we get a limit curve γ from p that is either past inextendible (if b_k are unbounded) or ends in $\gamma(b) = \lim_k \gamma_k(b_k) \in \overline{A}$ if $b_k \to b < \infty$. Since all γ_k lie entirely in $J^+(A)$, it follows that $\gamma \subseteq \overline{J^+(A)}$. γ has no point in common with $I^+(A)$, because if that were not the case, then we would get $p \in I^+(A)$ which is absurd. Hence $\gamma \subseteq \partial J^+(A)$, in particular γ is a null geodesic without conjugate points. (Since it is in $\partial J^+(A)$, it has to be maximizing between any of its points. A maximizing null curve is a null geodesic by [22, Prop. 10.46], and by Theorem 2.4.21 γ cannot have any conjugate points.)
It remains to argue that $\gamma(b) \in \mathrm{edge}(A)$ if γ does have a past endpoint. Consider γ maximally extended to the past and suppose $y := \gamma(b) \in \overline{A} \setminus \mathrm{edge}(A)$: By definition of $\mathrm{edge}(A)$, there is a convex neighborhood U of y such that for every $z^\pm \in I^\pm_U(y)$, any timelike curve from z^- to z^+ intersects A. Note that by construction of γ, there is a neighborhood W of $\gamma \setminus \{y\}$ that does not meet $\overline{A}$. Let $z^\pm$ be as before, then we may choose

$$z^0 \in I^+_U(z^-) \cap I^-_U(z^+) \cap (\gamma \setminus \{y\}),$$

and let λ_1 be timelike from z^- to z^0. We may suppose that $z^+ \in I^+_U(y) \cap W$, and choose a timelike curve λ_2 from z^0 to z^+ entirely in W. Then $\lambda_1\lambda_2$ is a timelike curve from z^- to z^+ and must hence meet A. By choice of z^0 and λ_2, we see that λ_2 does not meet $\overline{A}$ and hence it also does not meet A, so λ_1 meets A in some $z \in \lambda_1 \cap A$. There are two possibilites: Either z^0 comes before y on γ or after. In the first case, we have that

$$z^- \ll z \ll z^0 \leq y \leq p$$

and hence $p \in I^+(A)$, contradicting $p \in \partial J^+(A)$. If z^0 comes after y, then one similarly concludes that $p \in I^+(A)$. □

Lemma 3.1.15. ($H^+(A) \setminus \overline{A} \subseteq I^+(A)$)
Let A be an achronal set, then $H^+(A) \setminus \overline{A} \subseteq I^+(A)$.

Proof. Let $x \in H^+(A) \setminus \overline{A}$, then there is a neighborhood U of x not meeting $\overline{A}$. Let $y \in I^-_U(x)$, then $x \in I^+_U(y)$ and since $x \in \overline{D^+(A)}$, the open neighborhood $I^+_U(y)$ intersects $D^+(A)$ in some z.
Suppose that $y \notin D^+(A)$, then let γ be a past inextendible causal curve from y that does not meet A. Let σ be a timelike curve from y to z entirely in U. Then $\gamma\sigma$ has to meet A since $z \in D^+(A)$, but γ does not meet A and neither does σ, because $\sigma \subseteq U$ and U does not meet $\overline{A}$, a contradiction.
So $y \in D^+(A) \subseteq A \cup I^+(A)$, which implies that $x \in I^+(A)$. □

Proposition 3.1.16. (Generators of H^+)
Let A be closed and achronal. Then every $p \in H^+(A)$ is the future endpoint of a maximal null geodesic entirely in $\partial J^+(A)$ that is either past inextendible or has past endpoint in $\mathrm{edge}(A)$.

Proof. First, note that $P := D^+(A) \cup I^-(A)$ is a past set since $I^-(D^+(A)) \subseteq D^+(A) \cup I^-(A)$: Indeed, suppose $x \in I^-(D^+(A))$ and $x \notin D^+(A)$. Hence, there is a past inextendible causal curve γ with future endpoint x not meeting A. Since $x \in I^-(D^+(A))$, there is some $y \in D^+(A)$ such that $x \ll y$. Connect x to y by a future causal curve σ, then $\eta := \gamma\sigma$ is a past inextendible causal curve from y, hence it meets A at some $z \in A$. Since γ does not meet A, we have $z \in \sigma$ and hence $x \ll z$. Thus $x \in I^-(A)$. Since $H^+(A) \subseteq \partial P$, it is thus a subset of a closed topological hypersurface by Corollary 1.3.7.
Next, let $x \in H^+(A) \setminus \mathrm{edge}(A)$. We may choose a sequence $x_n \in I^+(A) \setminus D^+(A)$ with $x_n \to x$. For each n, let γ_n be a past inextendible causal curve that does not meet A. Parametrizing the γ_n to be h-unit speed, where h is an arbitrary complete Riemannian metric on M, the limit curve theorem implies that (up to a choice of subsequence) the γ_n converge locally uniformly to a past inextendible causal curve γ with future endpoint x. We will show that there is a convex neighborhood U of x such that $\gamma \cap U \subseteq H^+(A)$.
Since A is closed, Lemma 3.1.15 implies that $H^+(A) \subseteq A \cup I^+(A)$. Hence, for x as before there are the possibilities $x \in A \setminus \mathrm{edge}(A)$ or $x \in I^+(A)$. If $x \in A \setminus \mathrm{edge}(A)$, there exists a neighborhood V of x such that all timelike curves from $I^-_V(x)$ to $I^+_V(x)$ meet A. Choose U to be a convex neighborhood of x contained in V, then certainly every timelike curve from $I^-_U(x) \subseteq I^-_V(x)$ to $I^+_U(x) \subseteq I^+_V(x)$ meets A. In the case $x \in I^+(A)$, let U be a convex neighborhood of x contained in $I^+(A)$.
We claim that in either case, $\gamma \cap U \subseteq A \cup I^+(A)$. If $x \in I^+(A)$, this is clear by our choice of the convex neighborhood $U \subseteq I^+(A)$. Now suppose $x \in A \setminus \mathrm{edge}(A)$ and $y \in (\gamma \cap U) \setminus A$. Since A is closed, we may find a convex neighborhood V of y with $V \subseteq U$. Since $x \in A$ and $y \in \gamma \subseteq J^-(x)$, V intersects $I^-(x)$. Also, V intersects the curves γ_n for large enough n and, since V is convex and does not meet A, there is a timelike curve from $I^-(x)$ to γ_n entirely in V not meeting A. Concatenating this curve with the rest of γ_n to the future, which also does not meet A, we get a timelike curve from $I^-_U(x)$ to $I^+_U(x)$ which does not meet A, a contradiction to $x \in A \setminus \mathrm{edge}(A)$. Thus, we have shown $\gamma \cap U \subseteq A \cup I^+(A)$.
Via similar lines of argument, one then shows that $\gamma \cap U \subseteq \overline{D^+(A)}$ and $(\gamma \cap U) \cap I^-(D^+(A)) = \emptyset$, thus $\gamma \cap U \subseteq H^+(A)$, see [14, Prop. 8.3.1] for details.
Since $H^+(A)$ is closed, the curve $\gamma \cap U$ has a past endpoint $\hat{x}$ in $H^+(A)$. If $\hat{x} \notin \mathrm{edge}(A)$, we can repeat our construction to obtain a causal curve $\hat{\gamma} \subseteq H^+(A)$ with future endpoint $\hat{x}$. If this construction stops after finitely many iterations, we get a causal curve in $H^+(A)$ with future endpoint x and past endpoint in $\mathrm{edge}(A)$. If it does not, we get a past inextendible causal curve contained in $H^+(A)$ with future endpoint x. In either case, due to achronality of $H^+(A)$ such a causal curve must be everywhere null and maximizing, thus a (reparametrization) of a maximizing null geodesic. $\square$

Definition 3.1.17. (Horismos)
Let $A \subseteq M$ be any subset. Then its *future/past horismos* is

$$E^{\pm}(A) := J^{\pm}(A) \setminus I^{\pm}(A).$$

Lemma 3.1.18. (Basic properties of the horismos)
Let $A \subset M$ be achronal. Then:

(1) $\overline{A}$ is achronal.

(2) $A \subset E^{\pm}(A)$.

(3) If $E^{\pm}(A)$ is closed, then $E^{\pm}(A) = E^{\pm}(\overline{A})$.

(4) $E^{\pm}(A) \subset \partial J^{\pm}(A)$.

Proof.

(1) Suppose there are $p, q \in \overline{A}$ with $p \ll q$. Choose sequences $p_n, q_n \in A$ with $p_n \to p$ and $q_n \to q$. Then for large n, $p_n \ll q_n$ by openness of the chronological relation, contradicting the achronality of A.

(2) Since $A \subseteq J^{\pm}(A)$ for any set, and by achronality $A \cap I^{\pm}(A) = \emptyset$, the claim follows.

(3) We argue this only for E^+, with the E^--case being analogous. Suppose that $E^+(A)$ is closed. We first claim that $I^+(A) = I^+(\overline{A})$: $I^+(A) \subseteq I^+(\overline{A})$ is trivial. For the converse inclusion, suppose $q \in I^+(\overline{A})$. Then there is $p \in \overline{A}$ with $p \ll q$. Choose a sequence $p_n \in A$ with $p_n \to q$, then $p_n \ll q$ for large n, hence $q \in I^+(A)$. So we conclude that

$$E^+(A) = J^+(A) \setminus I^+(A) = J^+(A) \setminus I^+(\overline{A}) \subset J^+(\overline{A}) \setminus I^+(\overline{A}) = E^+(\overline{A}).$$

Conversely, since $A \subset E^+(A)$ by achronality and $E^+(A)$ is closed, we get $\overline{A} \subset E^+(A)$. Noting that $J^+(A) = J^+(E^+(A))$, we get

$$\begin{aligned} E^+(\overline{A}) &= J^+(\overline{A}) \setminus I^+(\overline{A}) \subset J^+(E^+(A)) \setminus I^+(A) \\ &= J^+(A) \setminus I^+(A) = E^+(A). \end{aligned}$$

(4) This is clear because $J^{\pm}(A)^{\circ} = I^{\pm}(A)$.

$\square$

Lemma 3.1.19. (Inclusion relation for H^+)
Let A be a closed, achronal set. Then

$$H^+(E^+(A)) \subseteq H^+(\partial J^+(A)).$$

Proof. Suppose $x \in \overline{H^+(E^+(A))} \setminus H^+(\partial J^+(A))$. Since $E^+(A) \subseteq \partial J^+(A)$, clearly $\overline{D^+(E^+(A))} \subseteq \overline{D^+(\partial J^+(A))}$, and since $x \notin H^+(\partial J^+(A))$, the definition gives $x \in I^-(D^+(\partial J^+(A)))$. Let $y \in I^+(x) \cap D^+(\partial J^+(A))$.
We show that $I^+(x) \cap I^-(y) \cap \partial J^+(A) = \emptyset$: Suppose not, and let z be in that set. Then $I^-(z) \ni x$, but $x \in \overline{D^+(E^+(A))}$, hence $I^-(z)$ meets $D^+(E^+(A))$. Let $z_0 \in I^-(z) \cap D^+(E^+(A))$. Let λ be a past inextendible timelike curve from z_0, then λ meets $E^+(A)$ at some z_1, hence $z_1 \ll z_0 \ll z$. But $z_1 \in E^+(A) \subseteq \partial J^+(A)$, hence $z_1 \in I^-(z) \cap \partial J^+(A)$, which is a contradiction to the achronality of $\partial J^+(A)$ since $z \in \partial J^+(A)$.
Now the neighborhood $I^-(y)$ of x contains points not in $\overline{D^+(E^+(A))}$, choosing such a point we get a past inextendible timelike curve not meeting $E^+(A)$. Since $I^+(x) \cap I^-(y)$ does not meet $\partial J^+(A) \supseteq E^+(A)$, we get a past inextendible timelike curve γ from y not meeting $E^+(A)$. Since $y \in D^+(\partial J^+(A))$, it follows that γ meets $\partial J^+(A)$ at some $\tilde{z}$. Let μ be a null geodesic from $\tilde{z}$ that is either past inextendible or ends in edge(A) (see Lemma 3.1.14). We show that both cases lead to a contradiction.
If μ ends in $z_0 \in \text{edge}(A)$, then $z_0 \in A$ since A is closed. Hence $\mu \subseteq J^+(A)$, and since $\tilde{z}$ is the future endpoint of μ, we get $\tilde{z} \in J^+(A) \cap \partial J^+(A) = E^+(A)$, a contradiction since $\tilde{z} \in \gamma$ and γ does not meet $E^+(A)$.
Suppose μ is past inextendible (and hence does not meet A). Since γ is timelike with future endpoint $y \in D^+(\partial J^+(A))$, γ must thus meet $(D^+(\partial J^+(A)))^\circ$. Concatenating γ from that point with μ after $\tilde{z}$, we see that μ intersects $I^-(\partial J^+(A))$. But $\mu \subseteq \partial J^+(A)$, so this is a contradiction by achronality of $\partial J^+(A)$. $\square$

Lemma 3.1.20. (Cauchy horizon of horismos)
Let (M, g) be a strongly causal spacetime and let $A \subseteq M$ be a closed, achronal set such that $E^+(A)$ is closed. Then $H^+(E^+(A))$ is noncompact or empty.

Proof. Suppose $H^+(E^+(A))$ is compact and nonempty. By strong causality, $H^+(E^+(A))$ can be covered by finitely many relatively compact convex sets U_i such that no causal curve leaving U_i ever returns. Let $z_1 \in H^+(E^+(A))$ and let U_1 be such a convex set with $z_1 \in U_1$. By Lemma 3.1.19, $H^+(E^+(A)) \subseteq H^+(\partial J^+(A))$, so the neighborhood U_1 of $z_1 \in H^+(\partial J^+(A)) \subseteq \partial D^+(\partial J^+(A))$ must contain a point x_1 not in $\overline{D^+(\partial J^+(A))}$, and this point may be chosen to be in $J^+(A)$. By Lemma 3.1.13, there is a past inextendible timelike curve α_1 from x_1 not meeting $\partial J^+(A)$, and it does not even meet $\overline{D^+(\partial J^+(A))}$ (indeed: if α_1 met $\overline{D^+(\partial J^+(A))}$, then it would constitute an inextendible timelike curve from that point on and would thus have to meet $\partial J^+(A)$). Since $x_1 \in J^+(A)$ and α_1 does not meet $\partial J^+(A)$, it is entirely contained in $I^+(A)$. Since inextendible curves cannot be trapped in a (relatively) compact set (this is due to strong causality, see [21, Thm. 2.80]), α_1 leaves U_1, so there is some $y_1 \in \alpha_1 \setminus U_1 \subseteq I^+(A) \subseteq I^+(E^+(A))$. Let

$\beta_1 : [0,1] \to M$ be a past timelike curve from $y_1 \notin \overline{D^+(E^+(A))}$ to $E^+(A)$. We claim that β_1 meets $H^+(E^+(A))$: Since $\beta_1(1) \in E^+(A) \subseteq \overline{D^+(E^+(A))}$, there is some $t_0 > 0$ such that $\beta_1(t) \notin \overline{D^+(E^+(A))}$ for $t < t_0$ and $\beta_1(t_0) \in \partial D^+(E^+(A)) = E^+(A) \cup H^+(E^+(A))$ (since $E^+(A)$ is closed and achronal). But $\beta_1(1) \in E^+(A)$, so $\beta_1(t_0) \in H^+(E^+(A)) \setminus E^+(A)$ by achronality of $E^+(A)$.
Now let $z_2 \in \beta_1 \cap H^+(E^+(A))$ and let U_2 be another one of the finitely many convex neighborhoods from above such that $z_2 \in U_2$. Retracing the steps of the constructions so far, we see that $x_1 \ll y_1 \ll z_2$, and $y_1 \notin U_1$, hence by strong causality $z_2 \notin U_1$. Continuing by induction, we see that no finite number of the U_i can cover $H^+(E^+(A))$, a contradiction to compactness. □

Corollary 3.1.21. (Inextendible timelike curve in D^+)
Let (M, g) be strongly causal and let $S \subseteq M$ be a nonempty, achronal set such that $E^+(S)$ is compact. Then there is a future inextendible timelike curve contained in $D^+(E^+(S))$.

Proof. Since $E^+(S) = E^+(\overline{S})$, we may assume that S is closed. Let X be a global timelike vector field on M. Since $H^+(E^+(S)) \subseteq \overline{D^+(E^+(S))}$, it follows from Lemma 3.1.13 that each maximal integral curve of X through a point of $H^+(E^+(S))$ meets $E^+(S)$. Suppose now that every maximal integral curve of X that meets $E^+(S)$ also meets $H^+(E^+(S))$. Define a continuous surjection $E^+(S) \to H^+(E^+(S))$ by sending $p \in E^+(S)$ to the (unique) point on the unique integral curve of X through p that is on $H^+(E^+(S))$. Then compactness of $E^+(S)$ would imply the compactness of $H^+(E^+(S))$ (unless $H^+(E^+(S))$ is empty, in which case the claim is trivial). □

3.1.2 Causally disconnected spacetimes

We intend to make use of the notion of causal disconnectedness in the proof of the Hawking-Penrose theorem, so let us introduce this concept now. Under some additional causality assumptions, causally disconnected spacetimes can be seen as the Lorentzian analogues of Riemannian manifolds that are not simply connected at infinity. The latter then always have a line, i.e. an inextendible geodesic that realizes the distance everywhere. We will show that strongly causal, causally disconnected spacetimes contain (causal) lines.

Definition 3.1.22. (Rays and lines)
Let (M, g) be a spacetime. A *future causal ray* is a causal, future inextendible geodesic maximizing the distance between each of its points. A *past causal ray* is defined dually. A *causal line* is an inextendible causal geodesic maximizing the distance between each of its points.

Definition 3.1.23. (Causal disconnectedness)
A spacetime (M, g) is *causally disconnected* if there is a compact set K and two sequences p_n, q_n that each leave every compact set such that $p_n < q_n$ for all n, and all future causal curves from p_n to q_n meet K.

Example 3.1.24. (Causally disconnected spacetime)
Consider the cylinder $M := \mathbb{R} \times S^1$ with the Lorentzian metric $-dt^2 + d\theta^2$. Let $K := \{0\} \times S^1$, and choose any inextendible timelike spiral $\gamma : \mathbb{R} \to M$. Choosing $p_n := \gamma(-n)$ and $q_n := \gamma(n)$, we see that M is causally disconnected.

Proposition 3.1.25. (Lines in causally disconnected spacetimes)
Let (M, g) be a strongly causal, causally disconnected spacetime. Then (M, g) has a causal line.

Proof. Let K, p_n, q_n be the data of the causal disconnection of M. Connect p_n to q_n by a causal curve $\gamma_n : [a_n, b_n] \to M$ that is parametrized by h-arc length for some fixed complete Riemannian metric h on M. Moreover, assume that

$$L_g(\gamma_n) \geq d_g(\gamma_n(a_n), \gamma_n(b_n)) - \varepsilon_n$$

where $\varepsilon_n \downarrow 0$. Then each γ_n meets K at some $\gamma(t_n)$, we may assume $t_n = 0$. By compactness, $\gamma_n(0) \to p \in K$. Since $d_g(p_n, q_n) \to \infty$, the a_n and b_n are necessarily unbounded and hence we may assume that $a_n \to -\infty$, $b_n \to \infty$. In particular, the γ_n (eventually) leave a fixed neighborhood of p. By the version of the limit curve theorem in [20, Thm. 3.1(1)], we get a g-inextendible, h-unit speed causal limit curve $\gamma : \mathbb{R} \to M$. Due to our assumption on $L_g(\gamma_n)$, γ is everywhere maximizing and hence a causal line. $\square$

Lemma 3.1.26. Let (M, g) be strongly causal. If there is a null line in M, then (M, g) is causally disconnected.

Proof. Let $c : (a, b) \to M$ be a null line. Let $d \in (a, b)$ and $p := c(d)$. Set $K := \{p\}$. Choose $s_n \uparrow a$ and $t_n \downarrow b$, and $p_n := c(s_n)$, $q_n := c(t_n)$. Then (after maybe dropping some s_n, t_n) we have $p_n < p < q_n$. Since c is inextendible, it leaves every compact set, and hence p_n, q_n leave every compact set. Now let $\sigma : [0, 1] \to M$ be any causal curve from p_n to q_n. Then σ has to be null, and hence a reparametrization of $c|_{[s_n, t_n]}$, which meets K. Thus, (M, g) is causally disconnected by K, p_n, q_n. $\square$

3.1.3 The genericity condition

The idea behind the genericity condition is that there is some nontrivial dynamics happening at least somewhere along any causal curve. This is a

reasonable assumption to make in nontrivial, physically relevant spacetimes. Together with Ricci estimates, the genericity condition implies the existence of conjugate points on complete, causal geodesics, as we will show.
Throughout this subsection, let (M, g) be a spacetime of dimension $n \geq 2$, except when we consider null geodesics, then we will assume $n \geq 3$ (since null geodesics do not have conjugate points in dimension $n = 2$, as mentioned in the introduction to Section 2.4).

Definition 3.1.27. (Genericity condition)
Let $c : (a, b) \to M$ be a timelike geodesic. (M, g) is said to satisfy the *genericity condition along* c if there is some $t_0 \in (a, b)$ such that the *tidal force operator*

$$R(., c'(t_0))c'(t_0) : T^{\perp}_{c(t_0)}M \to T^{\perp}_{c(t_0)}M$$

is not the zero map. Similarly, if $\beta : (a, b) \to M$ is a null geodesic, (M, g) is said to satisfy the *genericity condition along* β if there is some $t_0 \in (a, b)$ such that the *quotient tidal force operator*

$$\overline{R}(, \beta'(t_0))\beta'(t_0) : Q\beta(t_0) \to Q\beta(t_0)$$

is not the zero map (see Remark 2.4.1 for the notation). (M, g) is said to satisfy the *genericity condition* if it satisfies the genericity condition along any inextendible causal geodesic.

Definition 3.1.28. (Timelike and null energy conditions)
(M, g) is said to satisfy the *timelike energy condition* if $\mathrm{Ric}(v, v) \geq 0$ for all timelike $v \in TM$. It is said to satisfy the *null energy condition* if $\mathrm{Ric}(v, v) \geq 0$ for all null $v \in TM$. If it satisfies both energy conditions, we say that (M, g) satisfies the *causal energy condition*.

Note that, by continuity of Ric, the timelike energy condition implies the null energy condition. In particular, the timelike and causal conditions are equivalent. This is, however, not the case if one lowers the regularity of g (as is done in many recent works, see e.g. [7]). We will keep the different terminologies depending on the causal character of the vectors we want to emphasize.

Remark 3.1.29. (Some remarks on the expansion θ)
Let $c : [a, b] \to M$ be a timelike geodesic and let A be a Jacobi tensor along c. Then the *expansion* is $\theta(t) = \mathrm{Tr}\big(A'(t)A^{-1}(t)\big)$ (see Definition 2.3.21). Recall from Remark 2.3.22 the following formula:

$$\theta = (\det A)'(\det A)^{-1}.$$

So if $|\theta| \to \infty$ for $t \to t_0$, then $\det A(t_0) = 0$. Determining where $\det A(t_0) = 0$ is of relevance for the following reason: In the proof of Proposition 2.3.20, we showed how to understand A as an $(n-1)\times(n-1)$-matrix whose columns are Jacobi fields vanishing at some fixed initial point, say a. If $\det A(t_0) = 0$, then there is a nontrivial linear combination of the Jacobi fields at t_0 which is equal to 0, i.e. $\sum_i \lambda_i J_i(t_0) = 0$. But then $J(t) := \sum_i \lambda_i J_i(t)$ is a Jacobi field with $J(a) = 0$ and $J(t_0) = 0$, implying that $c(a)$ and $c(t_0)$ are conjugate along c.

Proposition 3.1.30. (Negative and positive expansion)
Let $c : I \to M$ be an inextendible timelike geodesic.
Suppose that $\operatorname{Ric}(c'(t), c'(t)) \geq 0$ for all $t \in I$. Moreover, suppose there is a Lagrange tensor A along c such that the expansion $\theta(t) = \operatorname{Tr}\big(A'(t)A^{-1}(t)\big)$ is negative (resp. positive) somewhere, i.e. there is some $t_1 \in J$ such that

$$\theta(t_1) < 0 \quad \text{resp. } \theta(t_1) > 0.$$

Then $\det A(t) = 0$ for some $t \in \left[t_1, t_1 - \frac{n-1}{\theta(t_1)}\right]$ resp. $t \in \left[t_1 - \frac{n-1}{\theta(t_1)}, t_1\right]$ provided that $t \in I$.

Proof. As noted in Remark 3.1.29, it suffices to show $|\theta| \to \infty$ for $t \to t_0$ with t_0 in the above intervals (depending on the case of negative or positive expansion). Let $T_1 := \frac{n-1}{\theta(t_1)}$. We only consider the case $T_1 < 0$, the case $T_1 > 0$ is similar. Since $\operatorname{Tr}\big(\sigma^2\big) \geq 0$, where σ is the shear, the vorticity-free Raychaudhuri equation (see Corollary 2.3.24) implies that

$$\theta' \leq -\frac{\theta^2}{n-1}.$$

We can integrate this from t_1 to $t > t_1$:

$$\int_{\theta(t_1)}^{\theta(t)} \frac{d\theta}{\theta^2} \leq \int_{t_1}^{t} -\frac{dt}{n-1}.$$

Upon calculating the integrals, some basic manipulations yield

$$\theta(t) \leq \frac{n-1}{t + T_1 - t_1}.$$

Note that the right hand side is negative for $t \in [t_1, t_1 - T_1)$. Hence

$$|\theta(t)| \geq -\frac{n-1}{t + T_1 - t_1}$$

and the right hand side goes to $+\infty$ for $t \to t_1 - T_1$, which proves the claim. $\square$

In the following, we want to establish that the genericity condition together with the timelike energy condition force a timelike geodesic to develop conjugate points. We will need some preparations for this.

Remark 3.1.31. (One-parameter family of Lagrange tensors)
Let $c : [t_1, \infty) \to M$ be a timelike geodesic without conjugate points to $c(t_1)$, and let $s \in (t_1, \infty)$. Then there is a unique Jacobi tensor D_s along c satisfying the boundary conditions $D_s(t_1) = \mathrm{id}$ and $D_s(s) = 0$ (note that the unbounded interval is of no significance and we can still use Proposition 2.3.20). Since $D_s(s) = 0$, D_s is even a Lagrange tensor (see Lemma 2.3.16). We will maintain this notation throughout this subsection.

Remark 3.1.32. (Integration of $(1,1)$-tensors)
Suppose c is a unit-speed timelike geodesic and A is a $(1,1)$-tensor normal to c, and let $E_1, \ldots, E_n$ be a parallel orthonormal frame along c with $E_n = c'$. Then A is represented as a matrix with entries A^i_j with respect to this frame. We can define for any parameter t_0 the *integral* $\int_{t_0}^t A(\tau)d\tau$ as the $(1,1)$-tensor along c whose components (at parameter value t) with respect to $E_i(t)$ are

$$\int_{t_0}^t A^i_j(\tau)d\tau.$$

This construction is independent of the parallel frame, since any two parallel frames along c are related by a constant base change matrix. The usual rules and conventions for integrals apply for the integration of $(1,1)$-tensors. A similar construction method can be applied to integrate $(1,1)$-tensor classes along null geodesics. (See [5] for more details.)

Lemma 3.1.33. (Formula for D_s)
Let $c : [t_1, \infty) \to M$ be a timelike geodesic without conjugate points to $c(t_1)$. Let A be the unique Lagrange tensor along c with $A(t_1) = 0$ and $A'(t_1) = \mathrm{id}$. For each $s \in (t_1, \infty)$, let D_s be the unique Lagrange tensor along c with $D_s(t_1) = \mathrm{id}$, $D_s(s) = 0$. Then

$$D_s(t) = A(t) \int_t^s (A^\dagger A)^{-1}(\tau)d\tau$$

for $t \in (t_1, s]$. In particular, D_s is nonsingular in (t_1, s).

Proof. First note that A is nonsingular on (t_0, ∞) by Proposition 2.3.20 because c has no conjugate points to $c(t_0)$. Let $X(t)$ be the $(1,1)$-tensor on the right hand side of the equation above. We need to show that X satisfies the Jacobi tensor equation, and $X(s) = 0 = D_s(s)$, $X'(s) = D_s'(s)$, because then $X = D_s$ by uniqueness.
We first show that X satisfies the Jacobi tensor equation, i.e. $X'' + RX = 0$. The first derivative yields

$$\begin{aligned} X'(t) &= A'(t) \int_t^s (A^\dagger A)^{-1}(\tau)d\tau - A(t)(A^\dagger A)^{-1}(t) \\ &= A'(t) \int_t^s (A^\dagger A)^{-1}(\tau)d\tau - (A^\dagger)^{-1}(t). \end{aligned}$$

Thus, using the formula (2.6), the second derivative of X is

$$\begin{aligned} X''(t) &= A''(t)\int_s^t (A^\dagger A)^{-1}(\tau)d\tau - A'(t)(A^\dagger A)^{-1}(t) - ((A^\dagger)^{-1})'(t) \\ &= A''(t)\int_t^s (A^\dagger A)^{-1}(\tau)d\tau - A'(t)A^{-1}(t)(A^\dagger)^{-1}(t) \\ &+ (A^\dagger)^{-1}(t)(A^\dagger)'(t)(A^\dagger)^{-1}(t). \end{aligned}$$

Since A is a Lagrange tensor, by definition (see Definition 2.3.15) we have that $(A^\dagger)' = A^\dagger A' A^{-1}$, so the second and third terms in X'' cancel, leaving

$$X''(t) = A''(t)\int_t^s (A^\dagger A)^{-1}(\tau)d\tau.$$

From this, since A satisfies the Jacobi tensor equation, it is evident that X does so as well. Clearly, $X(s) = 0$ and by the formula derived above for X',

$$X'(s) = -(A^\dagger)^{-1}(s).$$

We are done if we show that $D_s'(s) = -(A^\dagger)^{-1}(s)$. To this end, recall that $R = R^\dagger$ (as a $(1,1)$-tensor) and observe that

$$((A^\dagger)'D_s - A^\dagger D_s')' = (A^\dagger)''D_s - A^\dagger D_s'' = -A^\dagger R D_s + A^\dagger R D_s = 0.$$

Since $(A^\dagger)'(t_1)D_s(t_1) - A^\dagger(t_1)D_s'(t_1) = \mathrm{id} - 0 = \mathrm{id}$, it follows that

$$(A^\dagger)'D_s - A^\dagger D_s' = E,$$

where E is the Jacobi tensor along c that equals id for every t. Evaluating at s, we get (using $D_s(s) = 0$)

$$-A^\dagger(s)D_s'(s) = \mathrm{id},$$

and thus $D_s'(s) = -(A^\dagger)^{-1}(s)$ as claimed. Thus $X = D_s$ as was discussed at the beginning of the proof.
Finally D_s is nonsingular in $(t_1, s]$ because A is nonsingular and $(A^\dagger A)^{-1}$ is positive definite and self-adjoint, hence so is its integral, which is then in particular nonsingular. $\square$

Lemma 3.1.34. (The limit $\lim_{s\to\infty} D_s$)
Let $c : [a,\infty) \to M$ be a timelike geodesic such that no two points on c are conjugate along c. For $t_1 > a$ and $s \in [a,\infty) \setminus \{t_1\}$, let D_s be the unique Lagrange tensor along c with $D_s(t_1) = \mathrm{id}$ and $D_s(s) = 0$. Then for $v \in T^\perp_{c(t)}M$,

$$D(t)v := \lim_{s\to\infty} D_s(t)v$$

(with the limit understood in the finite-dimensional vector space $T^\perp_{c(t)}M$) defines a Lagrange tensor $D = D(t)$ along c. Moreover, $D(t)$ is nonsingular for $t \in (t_1, \infty)$.

Proof. In the first step, we show that $D_s'(t_1)$ has a self-adjoint limit for $s \to \infty$. Since D_s is a Lagrange tensor, we have (note that $(A^\dagger)' = (A')^\dagger$ due to linearity of taking the adjoint)

$$((D_s)')^\dagger(t_1)D_s(t_1) = D_s^\dagger D_s'(t_1).$$

Since $D_s(t_1) = \mathrm{id}$ and hence also $D_s^\dagger(t_1) = \mathrm{id}$, we have $((D_s)')^\dagger(t_1) = D_s'(t_1)$. The pointwise limit $D'(t_1)$ of the $D_s'(t_1)$ will then necessarily be self-adjoint by continuity, if it exists. If we show that for each $y \in T_{c(t_1)}^\perp M$, $\langle D_s'(t_1)y, y\rangle$ converges to some number $\langle D'(t_1)y, y\rangle$, then this completely determines the self-adjoint linear map $D'(t_1)$ by polarization. To this end, we establish that $\langle D_s'(t_1)y, y\rangle$ (for fixed $y \in T_{c(t_1)}^\perp M$) is monotonically increasing for $s \in (t_1, \infty)$ and is bounded above by $\langle D_a'(t_1)y, y\rangle$. Suppose $r \in (t_1, s)$. Then by (the proof of) Lemma 3.1.33, for $t \in (t_1, s]$ we have

$$D_s'(t) = A'(t)\int_t^s (A^\dagger A)^{-1}(\tau)d\tau - (A^\dagger)^{-1}(t),$$

where A is the unique Lagrange tensor with $A(t_1) = 0$ and $A'(t_1) = \mathrm{id}$. Denote by Y the parallel translate of y along c such that $Y(t_1) = y$. Using the formula above, we get for $t \in (t_1, s)$

$$\begin{aligned}\langle D_s'(t)Y(t), Y(t)\rangle = &\left\langle A'(t)\int_t^s (A^\dagger A)^{-1}(\tau)d\tau Y(t), Y(t)\right\rangle\\ &- \langle (A^\dagger)^{-1}(t)Y(t), Y(t)\rangle.\end{aligned}$$

For $t \in (t_1, r)$, we get

$$\begin{aligned}&\langle D_s'(t)Y(t), Y(t)\rangle - \langle D_r'(t)Y(t), Y(t)\rangle\\ =&\left\langle A'(t)\int_r^s (A^\dagger A)^{-1}(\tau)d\tau Y(t), Y(t)\right\rangle.\end{aligned}$$

Taking the limit $t \downarrow t_1$ and using $Y(t_1) = y$ and $A'(t_1) = id$, we get

$$\langle D_s'(t_1)y, y\rangle - \langle D_r'(t_1)y, y\rangle = \left\langle \int_r^s (A^\dagger A)^{-1}(\tau)d\tau Y(t_1), Y(t_1)\right\rangle.$$

Here, the integral on the right hand side is understood to be a constant (i.e. parallel) $(1,1)$-tensor along c. Since Y is parallel, the expression on the right hand side equals

$$\int_r^s \langle (A^\dagger A)^{-1}(\tau)Y(\tau), Y(\tau)\rangle d\tau.$$

Expanding the left factor, this can be rewritten as

$$\int_r^s \langle (A^\dagger)^{-1}(\tau)Y(\tau), (A^\dagger)^{-1}(\tau)Y(\tau)\rangle d\tau.$$

Since Y is nowhere trivial and A is everywhere nonsingular (except at $t = t_1$), the above expression is positive (since the restriction of the metric to $T^{\perp}_{c(\tau)}M$ is positive definite). We find that

$$\langle D'_r(t_1)y, y\rangle < \langle D'_s(t_1)y, y\rangle,$$

i.e. $s \mapsto \langle D'_s(t_1)y, y\rangle$ is monotonically increasing for $s > t_1$. Now we show that $\langle D'_s(t_1)y, y\rangle < \langle D'_a(t_1)y, y\rangle$ for $s > t_1$. Let Y be as before and define

$$J(t) := \begin{cases} D_a(t)Y(t) & a \le t < t_1, \\ D_s(t)Y(t) & t_1 \le t \le s. \end{cases}$$

Since $D_s(t_1) = id = D_a(t_1)$, J is continuous at t_1. Moreover, J is a piecewise Jacobi field since D_s and D_a are Lagrange tensors and Y is parallel, and it holds that $J(a) = 0$ and $J(s) = 0$. Denote by J_a and J_s the two smooth pieces of J. If I_a^s is the timelike index form of $c|_{[a,s]}$, then

$$\begin{aligned} I_a^s(J,J) &= I_a^{t_1}(J,J) + I_{t_1}^s(J,J) = -\langle J'_a(t_1), J_a(t_1)\rangle + \langle J'_s(t_1), J_s(t_1)\rangle \\ &= -\langle D'_a(t_1)y, y\rangle + \langle D'_s(t_1)y, y\rangle, \end{aligned}$$

where we used the formula from Remark 2.3.4 and the fact that $D_a(t_1) = id = D_s(t_1)$. By Theorem 2.3.9, $I(J,J) < 0$ because c has no conjugate points and $J(a) = 0$, $J(s) = 0$. This yields the claim

$$\langle D'_s(t_1)y, y\rangle < \langle D'_a(t_1)y, y\rangle$$

for all $s > t_1$. By the discussion at the beginning of the proof, we conclude that the self-adjoint linear map $D'(t_1) : T^{\perp}_{c(t_1)}M \to T^{\perp}_{c(t_1)}M$ exists and it is the pointwise limit of the $D'_s(t_1)$. Now define D as the unique Jacobi tensor along c with $D(t_1) = id$ and $D'(t_1)$ as the limit obtained above. Since the initial conditions of D_s approach those of D for $s \to \infty$ and D_s, D are both Jacobi tensors, it follows from ODE theory (continuous dependence on initial data) that D is the pointwise limit of the D_s. By definition of Lagrange tensors, it is clear that this property is preserved under limits, hence D is also Lagrange. The nonsingularity claim is clear from the integral representation as was proven for D_s in Lemma 3.1.33. □

Lemma 3.1.35. (Two classes of Lagrange tensors)
Let $c : \mathbb{R} \to M$ be a complete timelike geodesic such that $\mathrm{Ric}(c', c') \ge 0$ along c and such that M satisfies the genericity condition along c at some $t_1 \in \mathbb{R}$. Moreover, let

$$\begin{aligned} L_+ &:= \{\text{Lagrange tensors } A \text{ with } A(t_1) = \mathrm{id}, \theta(t_1) \ge 0\}, \\ L_- &:= \{\text{Lagrange tensors } A \text{ with } A(t_1) = \mathrm{id}, \theta(t_1) \le 0\}. \end{aligned}$$

Then for each $A \in L_-$, there is some $t \in (t_1, \infty)$ such that $\det A(t) = 0$, and for each $A \in L_+$, there is some $t \in (-\infty, t_1)$ such that $\det A(t) = 0$.

Proof. We only prove the claim for $A \in L_-$, the case $A \in L_+$ is analogous. If $A \in L_-$, then $A(t_1) = id$ and $\theta(t_1) = \mathrm{Tr}\big(A'(t_1)A^{-1}(t_1)\big) = \mathrm{Tr}(A'(t_1)) \leq 0$. The vorticity-free Raychaudhuri equation, cf. Corollary 2.3.24 implies that

$$\theta'(t) = -\operatorname{Ric}(c'(t), c'(t)) - \mathrm{Tr}\big(\sigma^2(t)\big) - \frac{\theta^2(t)}{n-1} \leq 0.$$

If $\theta(t_0) < 0$ for some $t \geq t_1$, then Proposition 3.1.30 would give the claim. Suppose not, i.e. in light of $\theta' \leq 0$ suppose $\theta(t) = 0$ for all $t \geq t_1$. Thus $\theta'(t) = 0$ for $t \geq t_1$. Reinserting this back into the vorticity-free Raychaudhuri equation, we get for $t \geq t_1$

$$0 = -\operatorname{Ric}(c'(t), c'(t)) - \mathrm{Tr}\big(\sigma^2(t)\big) - 0 \leq -\mathrm{Tr}\big(\sigma^2(t)\big),$$

but $\mathrm{Tr}\big(\sigma^2\big) \geq 0$, hence $\mathrm{Tr}\big(\sigma^2(t)\big) = 0$ for $t \geq t_1$. Since the shear σ is self-adjoint (cf. Definition 2.3.21), it follows that $\sigma(t) = 0$ for all $t \geq t_1$. Now $B = A'A^{-1}$ is self-adjoint because A is Lagrange, cf. the proof of Corollary 2.3.24. This, together with the assumption that $\theta(t) = 0$ on $[t_1, \infty)$ implies that

$$B(t) = \sigma(t) = 0 \quad \text{on } [t_1, \infty).$$

But then the Riccati equation $R = -B' - B^2$ (cf. Remark 2.3.25) implies that the tidal force operator vanishes at t_1, which is a contradiction to our assumptions. □

We are now ready to formulate and prove the main result of this subsection, which shows that the genericity condition forces a (complete) timelike geodesic to focus (i.e. to develop conjugate points), if the Ricci curvature along the geodesic is nonnegative (e.g. if the spacetime satisfies the timelike energy condition).

Theorem 3.1.36. (Genericity and energy conditions imply focusing)
Let (M, g) be a spacetime of dimension $n \geq 2$. Let $c : \mathbb{R} \to M$ be a complete timelike geodesic such that $\operatorname{Ric}(c', c') \geq 0$ for all $t \in \mathbb{R}$. If M satisfies the genericity condition along c at some $t_1 \in \mathbb{R}$, then c has a pair of conjugate points.

Proof. Suppose c has no conjugate points. Then the Lagrange tensor $D(t) = \lim_s D_s(t)$ constructed in Lemma 3.1.34 with $D(t_1) = id$ is nonsingular for all $t \geq t_1$. Then $D \notin L_-$ in the notation of Lemma 3.1.35, so necessarily $D \in L_+ \setminus L_-$, which means

$$\theta(t_1) = \mathrm{Tr}\big(D'(t_1)D^{-1}(t_1)\big) = \mathrm{Tr}\big(D'(t_1)\big) > 0.$$

By continuity, there is some $s > t_1$ such that

$$\mathrm{Tr}\big(D_s'(t_1)\big) > 0.$$

Again by Lemma 3.1.35, $D_s(t_2)$ is singular for some $t_2 < t_1$, so let $v \in T^{\perp}_{c(t_2)}M \setminus \{0\}$ such that $D_s(t_2)v = 0$. Since D_s is a nontrivial Jacobi tensor, letting Y be the parallel translate of v along c, $D_s(Y)$ is a nontrivial Jacobi field along c vanishing at t_2 and s (because $D_s(s) = 0$), a contradiction to the assumption that c has no conjugate points. $\square$

Corollary 3.1.37. (Incompleteness or focusing)
Let (M, g) be a spacetime of dimension $n \geq 2$. Suppose (M, g) satisfies the timelike energy condition and the genericity condition along each inextendible timelike geodesic. Then each timelike geodesic in M is either incomplete or has a pair of conjugate points.

Results similar to these may be established for null geodesics in much the same manner, except that one has to work with the corresponding quotient objects and equations. The most important ingredient, which is a consequence of the vorticity-free Raychaudhuri equation for null geodesics, is the following analogue of Proposition 3.1.30:

Proposition 3.1.38. (Negative and positive expansion: null case)
Let (M, g) be a spacetime of dimension $n \geq 3$ and let $\beta : I \to M$ be an inextendible null geodesic. Suppose $\mathrm{Ric}(\beta'(t), \beta'(t)) \geq 0$ for all $t \in I$. Let $\overline{A}$ be a Lagrange tensor class along β such that the expansion

$$\overline{\theta}(t) = \mathrm{Tr}\Big(\overline{A}'(t)\overline{A}^{-1}(t)\Big) = (\det \overline{A}(t))'(\det \overline{A}(t))^{-1})$$

has a negative resp. positive value at some $t_1 \in I$. Then $\det \overline{A}(t) = 0$ for some $t \in \Big[t_1, t_1 - \frac{n-2}{\overline{\theta}(t_1)}\Big]$ resp. $t \in \Big[t_1 - \frac{n-2}{\overline{\theta}(t_1)}, t_1\Big]$.

By establishing quotient versions of all the lemmas discussed so far, one eventually proves the following analogue of Theorem 3.1.36:

Theorem 3.1.39. (Genericity and energy imply focusing: null case)
Let (M, g) be a spacetime of dimension $n \geq 3$ and let $\beta : \mathbb{R} \to M$ be a complete null geodesic such that $\mathrm{Ric}(\beta'(t), \beta'(t)) \geq 0$ for all $t \in \mathbb{R}$. If M satisfies the genericity condition along β at some $t_1 \in \mathbb{R}$, then β has a pair of conjugate points.

Corollary 3.1.40. (Incompleteness or focusing: null case)
Let (M, g) be a spacetime of dimension $n \geq 3$ satisfying the null energy condition and the genericity condition along any inextendible null geodesic. Then each null geodesic in M is either incomplete or has a pair of conjugate points.

We can combine Corollary 3.1.37 and Corollary 3.1.40 to get the following main result of this subsection (recall that the causal energy condition and timelike energy condition are equivalent; we still sometimes refer to the causal energy condition to put an emphasis on both timelike and null vectors).

Theorem 3.1.41. (Incompleteness or conjugate points)
Let (M, g) be a spacetime of dimension $n \geq 3$ satisfying the causal energy condition and the genericity condition. Then every causal geodesic in M is either incomplete or has a pair of conjugate points.

3.1.4 Focal points

In this subsection, we will be interested in geodesics orthogonal to a spacelike hypersurface H of a spacetime (M, g) and their focusing behavior. All the operations that appear (the Levi-Civita connection, covariant derivatives) are understood as operations in M (unless stated otherwise).
Recall from Section 1.2 (and generally Chapter 1) that for a given spacelike hypersurface H, we (locally) have two choices of unit timelike normal fields. We will choose the normal N to be future directed. For the shape operator $S = W_N$ (we will use these notations interchangeably), it holds by Lemma 1.2.5 that

$$S(X) = W_N(X) = -\nabla_X N$$

for all $X \in \mathfrak{X}(H)$, where $\nabla_X N$ is understood via local extensions (see Definition 1.1.9).

Proposition 3.1.42. (Geodesic variations orthogonal to H)
Let H be a spacelike hypersurface in a spacetime (M, g) and let N be a local unit normal for H. Let $c : [a, b] \to M$ be a timelike geodesic with $c(a) \in H$, $\dot{c}(a) = N(c(a)) \perp H$. Let $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ be a geodesic variation of c with $\alpha(a, s) \in H$, $\partial_t \alpha(t, s)|_{t=a} = N_{c(a,s)} \perp H$. Then the variation vector field J of α is a Jacobi field along c and satisfies

$$\tan J'(a) = -W_{N(c(a))} J(a),$$

where $\tan J'(a)$ is the component of $J(a)$ tangent to $T_{c(a)}H$.

Proof. Since α is a geodesic variation, J is a Jacobi field (cf. Proposition 2.1.6). We now show that the claimed equation holds. Note first that since $s \mapsto \alpha(a, s) \in H$, $J(a) = \partial_s|_0 \alpha(a, s) \in TH$. Now let $X \in \mathfrak{X}(H)$ be arbitrary. We will use the notations $q := c(a)$ and $\alpha_s = \alpha(., s)$. Then, since $\dot{\alpha}_s(0) = \partial_t \alpha(t, s)|_{t=a} \perp H$, we have

$$\begin{aligned}
0 &= \frac{d}{ds}|_0 \langle X_{\alpha_s(a)}, \dot{\alpha}_s(a) \rangle \\
&= \langle \nabla^M_s|_0 X_{\alpha_s(a)}, \dot{c}(a) \rangle + \langle X_q, \nabla^M_s|_0 \partial_t|_a \alpha(t, s) \rangle \\
&= \langle II_q(X_q, J(a)), N_q \rangle + \langle X_q, \nabla^M_t|_a J(t) \rangle,
\end{aligned}$$

where we used $\dot{c}(a) = N_q$, the Gauss formula (see Corollary 1.1.25) along the transversal curve $s \mapsto \alpha(a, s)$ and Lemma 2.1.3. Therefore, with the understanding that $J'(a) = \nabla_t^M|_a J(t)$,

$$\langle J'(a), X_q \rangle = -\langle II_q(X_q, J(a)), N_q \rangle = -\langle W_{N_q} J, X_q \rangle,$$

which proves the claim. $\square$

Remark 3.1.43. (Arbitrary hypersurface variations)
The proof of Proposition 3.1.42 shows in fact that if α is *any* (not necessarily geodesic) variation of c with the properties described there, then its variational vector field V will satisfy

$$\tan V'(a) = -W_{N(c(a))} V(a).$$

Moreover, if V is a Jacobi field along c with $V(a) \in T_{c(a)}H$, $\tan V'(a) = -W_{N(c(a))} V(a)$, then there is a geodesic variation as in Proposition 3.1.42 with variation vector field V (see [14, Lem. 4.6.12]).

Definition 3.1.44. (H-Jacobi fields)
Let $H \subseteq M$ be a spacelike hypersurface with local unit normal N and let $c : [a, b] \to M$ be a timelike geodesic with $c(0) \in H$, $\dot{c}(0) = N_{c(0)} \perp H$. A Jacobi field $J \in \mathfrak{X}(c)$ is called *H-Jacobi field* if $J \perp \dot{c}$ and

$$\tan J'(a) = \nabla_t^M|_0 J(t) = -W_{N(c(a))} J.$$

Definition 3.1.45. (Focal points)
Let $H \subseteq M$ be a spacelike hypersurface with local normal field N and let $c : [a, b] \to M$ be a timelike geodesic with $c(a) \in H$, $\dot{c}(a) \perp H$. Then $c(t)$, $t \in (a, b]$, is called *focal point* of H along c if there is a nontrivial H-Jacobi field $J \in \mathfrak{X}(c)$ with $J(t) = 0$.

We will henceforth refer to geodesics c as above as geodesics orthogonal to H (at $c(a)$).

Remark 3.1.46. (Focal points and Jacobi tensors)
Let $c : [a, b] \to M$ be a timelike geodesic orthogonal to a spacelike hypersurface $H \subseteq M$. Note that $T_{c(a)}^{\perp} M = T_{c(a)}H$. There is a unique Jacobi tensor A along c satisfying

$$A(a) = \mathrm{id}, \quad A'(a) = -W_{N(c(a))} A(a) = -W_{N(c(a))},$$

where $W_{N(a)} : T_{c(a)}^{\perp} M \to T_{c(a)}^{\perp} M$ is well-defined on tangent vectors (cf. Definition 1.1.17). Now if $Y \in \mathfrak{X}(c)$ is a nontrivial parallel field orthogonal to c, then AY is a nontrivial H-Jacobi field along c. Since $J'(a) = -W_{N(a)} J = -W_{N(a)} J(a)$, the dimension of H-Jacobi fields along c orthogonal to c is $n - 1$. Since AY is a nontrivial H-Jacobi field for every nontrivial parallel field Y along c orthogonal to c (of which there are $n-1$ linearly independent directions), it follows that every H-Jacobi field can be expressed in the form AY.

Lemma 3.1.47. (A is Lagrange)
Let A be as in Remark 3.1.46. Then A is a Lagrange tensor.

Proof. Since the Weingarten map is self-adjoint, it follows that $A'(a)^\dagger = A'(a)$. Together with $A(a) = id$, it follows that

$$A'(a)^\dagger A(a) = A(a)^\dagger A'(a).$$

Since the Wronskian is constant along c, it follows that A is a Lagrange tensor. $\square$

With this observation, we may prove the following analogue of Proposition 3.1.30 and Proposition 3.1.38

Corollary 3.1.48. (Negative and positive expansion: hypersurface case)
Let (M, g) be a spacetime of dimension $n \geq 2$ and let $c : I \to M$ be an inextendible timelike geodesic orthogonal to a spacelike hypersurface $H \subseteq M$ at $q =: c(t_0)$. Suppose that $\mathrm{Ric}(c'(t), c'(t)) \geq 0$ for all $t \in I$. If $-Tr(W_N)$ (with N a choice of unit normal for H) has a negative resp. positive value T_0 at q, then there is a focal point $c(t)$ to H along c for some $t \in \left[t_0, t_0 - \frac{n-1}{T_0}\right]$ resp. $t \in \left[t_0 - \frac{n-1}{T_0}, t_0\right]$.

Proof. We saw above that the unique Jacobi tensor along c with $A(t_0) = id$ and $A'(t_0) = -W_{N(c(t_0))}$ is Lagrange[1]. Note that

$$\theta(t_0) = \mathrm{Tr}\big(A'(t_0)A^{-1}(t_0)\big) = \mathrm{Tr}\big(A'(t_0)\big) = -\,\mathrm{Tr}\big(W_{N(q)}\big).$$

From Proposition 3.1.30 it follows that A becomes singular in the given intervals depending on whether $T_0 < 0$ or $T_0 > 0$. But if A becomes singular at some t, we may take a nonzero vector in its kernel, parallel transport it to get an H-Jacobi field AY that vanishes at t, which means that $c(t)$ is a focal point of H along c. $\square$

Definition 3.1.49. (Hypersurface index form)
Let $c : [a, b] \to M$ be a g-unit timelike geodesic orthogonal to a spacelike hypersurface $H \subseteq M$ at $c(a)$. Let $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$ with $Z(b) = 0$. We define the *H-index form* of Z as

$$I_H(Z, Z) := I(Z, Z) + \langle W_{c'(a)} Z(a), Z(a)\rangle,$$

where I is the usual timelike index form along c. If $Z \in \mathfrak{X}^{0,\perp}_{\mathrm{pw}}(c)$, then $I_H(Z, Z) = I(Z, Z)$.

[1] We argued with some initial parameter a but of course it does not matter at which parameter the curve intersects H.

Remark 3.1.50. (Hypersurface variations and I_H)
Let $c : [a,b] \to M$ be a unit speed timelike geodesic orthogonal to H at $c'(a)$ and let $\alpha : [a,b] \times (-\varepsilon, \varepsilon) \to M$ be a variation of c by timelike curves such that the variation field V is orthogonal to c' and satisfies $V(b) = 0$. Then Proposition 2.2.10 and Remark 2.3.4 yield

$$L''(0) = I_H(V, V).$$

Proposition 3.1.51. (Maximality and I_H)
Let $c : [a,b] \to M$ be a unit speed timelike geodesic starting orthogonally to a spacelike hypersurface H. If there is a focal point $c(t_0)$, $t_0 \in (a,b)$, of H along c, then there is a vector field $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$ with $0 \neq Z(a) \in T_{c(a)}H$ and $Z(b) = 0$ such that $I_H(Z,Z) > 0$. In this case, any variation of c with variational field Z will produce longer timelike curves from H to $c(b)$ (for small enough transversal parameters).

Proof. By assumption, there is a nontrivial H-Jacobi field J_1 along c vanishing at t_0. Let J be defined as J_1 up to t_0 and 0 after, then J is a piecewise smooth Jacobi field. Note that J' has a true breakpoint at t_0 since $J_1'(t_0) \neq 0$. Now choose any smooth vector field $V \in \mathfrak{X}(c)$ with $V'(a) = 0$, $V(a) = 0$, $V(b) = 0$ and $\langle V(t_0), \delta J'(t_0)\rangle = -1$ and define for $r \in \mathbb{R} \setminus \{0\}$ the vector field

$$Z := \frac{1}{r}J - rV.$$

Then $Z \in \mathfrak{X}^{\perp}_{\mathrm{pw}}(c)$ and

$$\begin{aligned} Z'(a) &= \frac{1}{r}J_1'(a) = \frac{1}{r}(-W_{c'(a)}J_1(a)) = -W_{c'(a)}\left(\frac{1}{r}J(a) + rV(a)\right) \\ &= -W_{c'(a)}Z(a). \end{aligned}$$

Moreover,

$$\begin{aligned} I_H(Z,Z) &= I(Z,Z) + \langle W_{c'(a)}Z(a), Z(a)\rangle \\ &= I(Z,Z) + \left\langle W_{c'(a)}\big(\frac{1}{r}J(a) - rV(a)\big), \frac{1}{r}J(a) - rV(a)\right\rangle \\ &= I(Z,Z) + \frac{1}{r^2}\langle W_{c'(a)}J(a), J(a)\rangle \\ &= I(Z,Z) + \frac{1}{r^2}\langle -J'(a), J(a)\rangle \\ &= \frac{1}{r^2}I(J,J) + r^2 I(V,V) - 2I(J,V) + \frac{1}{r^2}\langle -J'(a), J(a)\rangle. \end{aligned}$$

Now, since J is a piecewise Jacobi field, it follows from Remark 2.3.4 that

$$I(J,J) = \sum_{i=0}^{k}\langle \delta J'(t_i), J(t_i)\rangle$$

where $t_0 = a$, $t_k = b$, and the t_i are the breakpoints of J'. But J' only has one breakpoint, namely t_0, and since $J(t_0) = 0$, this breakpoint does not contribute. What remains from the above are the boundary terms, and together with $J(b) = 0$, we are finally left with

$$I(J, J) = \langle J'(a), J(a) \rangle.$$

Inserting this in the above calculation for I_H, it cancels the last term and we get

$$I_H(Z, Z) = r^2 I(V, V) - 2I(J, V).$$

Similarly to our arguments for $I(J, J)$, one can show that

$$I(J, V) = \langle \delta J'(t_0), V(t_0) \rangle = -1.$$

Hence

$$I_H(Z, Z) = r^2 I(V, V) + 2.$$

Since $I(V, V)$ attains a minimum on the compact interval $[a, b]$, we may choose r to be so small that $I_H(Z, Z) > 0$. Now consider any variation α with variation field Z. Since $Z(a) = J(a) = 1/rJ_1(a) \in T_{c(a)}H$, we have that $\alpha(a, s) \in H$ for small s, and $\alpha(., s)$ will be timelike (after shrinking the s-interval even more if necessary). Since $Z'(a) = -W_{c'(a)}Z(a)$ and c starts orthogonally to H, we may read the calculation in the proof of Proposition 3.1.42 backwards to see that $\partial_t|_a \alpha(t, s) \perp TH$ for all s. Hence, $I_H(Z, Z) > 0$ together with Remark 3.1.50 proves that $\alpha(., s)$, s small, is a timelike curve from H to $c(b)$ that is longer than c. □

Having dealt with timelike geodesics orthogonal to spacelike hypersurfaces, we are now interested in focal points of codimension-2 spacelike submanifolds along orthogonal null geodesics.

Let $H \subseteq M$ be a spacelike submanifold of codimension 2 and let $p \in H$. We may choose an orthonormal basis $E_1(p), \ldots, E_n(p)$ of T_pM such that $E_1(p), \ldots, E_{n-2}(p)$ are a spacelike orthonormal basis for T_pH, and $E_{n-1}(p)$ and $E_n(p)$ are future null vectors with $\langle E_{n-1}(p), E_n(p) \rangle = -1$. Considering the $E_i(p)$ locally extended to vector fields E_i around $p \in H$ such that all of these conditions continue to hold, it follows that E_{n-1} and E_n are (local) vector fields orthogonal to H, hence we may consider their Weingarten maps $W_{E_{n-1}}$ and W_{E_n}.

Now let $\beta : [a, b] \to M$ be a future null geodesic orthogonal to H at $\beta(a) = p$, with $\beta'(a) = E_n(p)$. We may parallel translate $E_1(p), \ldots, E_n(p)$ along β to get an orthonormal basis along β, which we will also denote by $E_1, \ldots, E_n$, where $E_n = \beta'$. Using the quotient notations and constructions from Remark 2.4.1, we see that the restriction of the projection $N\beta(a) \to Q\beta(a)$ to

T_pH is an isomorphism (by construction). Using this isomorphism, we may project the (pointwise) Weingarten maps $W_{E_i(p)} : T_pH \to T_pH$, $i = n-1, n$, to obtain the *projected Weingarten maps* $\overline{W}_{E_i(p)} : Q\beta(a) \to Q\beta(a)$.

Lemma 3.1.52. (Orthogonal null variations)
Let $\beta : [a, b] \to M$ be a future null geodesic starting orthogonally to a codimension-2 spacelike hypersurface H. Let $\alpha : [a, b] \times (-\varepsilon, \varepsilon) \to M$ be a variation of β such that the neighboring curves start in H, and moreover $\partial_t\alpha(t, s)|_{t=a} = E_n(\alpha(a, s))$, i.e. the neighboring curves start with initial null direction E_n. Moreover, assume that $\partial_s\alpha(a, s)$ is orthogonal to H for each s. Then the variational vector field V of α satisfies

$$V'(a) = -W_{\beta'(a)}V(a) + \lambda\beta'(a)$$

for some $\lambda \in \mathbb{R}$.

Proof. Since $\partial_t\alpha(t, s)|_{t=a} = E_n(\alpha(a, s))$ is null, it follows by Lemma 2.1.3 that

$$\begin{aligned} 0 &= \frac{d}{ds}|_0\langle \partial_t|_a\alpha(t, s), \partial_t|_a\alpha(t, s)\rangle = 2\langle \nabla_s|_0\partial_t|_a\alpha(t, s), \partial_t|_a\alpha(t, 0)\rangle \\ &= 2\langle V'(a), \beta'(a)\rangle. \end{aligned}$$

Hence $V'(a) \perp \beta'(a)$, which means that it can uniquely be written as

$$V'(a) = v + \lambda\beta'(a)$$

for some $\lambda \in \mathbb{R}$ and $v \in T_pH$. We want to show $v = -W_{\beta'(a)}V(a)$, and for this it suffices to show that

$$\langle W_{\beta'(a)}V(a), y\rangle = -\langle V'(a), y\rangle$$

for all $y \in T_pH$. For this, observe that

$$V'(a) = \nabla^M_t|_a\partial_s|_0\alpha(t, s) = \nabla^M_s|_0\partial_t|_a\alpha(t, s).$$

Now, $X(s) := \partial_t|_a\alpha(t, s)$ is a vector field along the transversal curve $\gamma(s) := \alpha(a, s)$ that is in H, and X is everywhere orthogonal to H by assumption. Hence, the results in Definition 1.1.27 yield

$$V'(a) = II^{\tan}(\gamma'(0), X(0)) + \nabla^\perp_s|_0X(s).$$

The second term does not matter in an inner product with $y \in T_pH$. As for the first term, $\gamma'(0) = \partial_s|_0\alpha(a, s) = V(a)$ and $X(0) = \partial_t|_a\alpha(t, 0) = \beta'(a)$, Hence by definition of $II^{\tan}$ (see Remark 1.1.24) it follows that

$$V'(a) = -W_{\beta'(a)}V(a) + \nabla^\perp_s|_0X(s)$$

and thus $\langle V'(a), y\rangle = -\langle W_{\beta'(a)}V(a)\rangle$ as claimed. $\square$

Definition 3.1.53. (H-Jacobi class)
Let $\beta : [a,b] \to M$ be a null geodesic starting orthogonally to a codimension-2 spacelike submanifold H. A Jacobi class $\overline{J} \in \overline{\mathfrak{Jac}}(\beta)$ is called *H-Jacobi class* if

$$\tan \overline{J}'(a) = -\overline{W}_{\beta'(a)} \overline{J}(a),$$

where the tangential part of the class $J'(a)$ is the class of the tangential part of a representative.

Definition 3.1.54. (Focal points along null geodesics)
Let $\beta : [a,b] \to M$ be a null geodesic starting orthogonally to a codimension-2 spacelike submanifold H. Then $\beta(t_0)$, $t_0 \in (a,b]$, is called *focal point* of H along β if there is a nontrivial H-Jacobi class $\overline{J} \in \overline{\mathfrak{Jac}}(\beta)$ with $\overline{J}(t_0) = [\beta'(t_0)]$.

Using these quotient versions of the notions we used before, the following result is proven in a similar manner to what has been done up to here, see [14, Lem. 4.6.15] for details.

Theorem 3.1.55. (Focal points and Maximality of null geodesics)
Let $\beta : [a,b] \to M$ be a null geodesic starting orthogonally to a codimension-2 spacelike submanifold H. If there is a focal point $\beta(t_0)$ to H along β, then there is a timelike curve from H to $\beta(b)$. In particular, β is not maximizing from H to $\beta(b)$.

The following is proven in the same way as Corollary 3.1.48.

Proposition 3.1.56. (Positive and negative expansion: codimension 2)
Let (M,g) be a spacetime of dimension $n \geq 3$ and let $H \subseteq M$ be a codimension-2 spacelike submanifold. Let $\beta : I \to M$ be an inextendible null geodesic orthogonal to H at $p := \beta(t_0)$. Suppose that $\mathrm{Ric}(\beta'(t), \beta'(t)) \geq 0$ for all $t \in I$. If $-\operatorname{Tr}\big(W_{\beta'(t_0)}\big)$ has a negative resp. positive value T_0 at t_0, then there is a focal point $\beta(t)$ to H along β for some $t \in \left[t_0, t_0 - \frac{n-2}{T_0}\right]$ resp. $\left[t_0 - \frac{n-2}{T_0}, t_0\right]$.

3.1.5 Trapped sets and trapped surfaces

Trapped surfaces are thought of as codimension-2 submanifolds that make outgoing lightrays focus. This can be formulated in analytic terms via conditions on the traces of Weingarten maps in normal null directions, and in terms of causality theory by demanding the compactness of the horismos $E^{\pm} = J^{\pm} \setminus I^{\pm}$, which leads to the notion of trapped sets. In this subsection, we investigate the relationship between these two notions.

Definition 3.1.57. (Trapped surface)
A *trapped surface* H in a spacetime (M, g) is a compact codimension-2 spacelike submanifold such that if $E_1, \ldots, E_n$ is a local frame in TM around a point in H such that $E_1, \ldots, E_{n-2}$ reduce to an orthonormal frame for TH, and E_{n-1} and E_n are null with $\langle E_{n-1}, E_n \rangle = -1$, and $E_1, \ldots, E_{n-2}$ are normal to E_{n-1}, E_n, then $\mathrm{Tr}\big(W_{E_{n-1}}\big)$ and $\mathrm{Tr}(W_{E_n})$ are either both positive (*future trapped surface*) or both negative (*past trapped surface*).

Definition 3.1.58. (Trapped set)
A nonempty achronal set $A \subseteq M$ is said to be a *future trapped set* if $E^+(A) := J^+(A) \setminus I^+(A)$ is compact. *Past trapped sets* are defined analogously.

In the proof of the following result, we follow [8, Prop. 6.7].

Proposition 3.1.59. (Trapped sets and trapped surfaces)
Let (M, g) be a spacetime of dimension $n \geq 3$ satisfying the null energy condition. Suppose $H \subseteq M$ is a trapped surface. Then at least one of the following holds:

(1) $E^+(H)$ or $E^-(H)$ is compact.

(2) (M, g) is null geodesically incomplete.

Proof. Suppose (M, g) is null geodesically complete and that $\mathrm{Tr}\big(W_{E_{n-1}}\big) > 0$ and $\mathrm{Tr}(W_{E_n}) > 0$. Consider future directed null geodesics from points of H with initial direction E_{n-1} or E_n. Each such geodesic reaches a focal point of H by Proposition 3.1.38. Given such a geodesic, we can write it as $\exp_p(tE_i(p))$ for $i \in \{n-1, n\}$, and then there is some t_p such that for both $i = n-1$ and $i = n$ we have $\exp_p(t_p E_i(p)) \in I^+(H)$. By openness of $I^+(H)$ and continuity of exp, we have $\exp_q(t_p E_i(q)) \in I^+(H)$ for all q close to p. Covering H with finitely many such neighborhoods, we may then take the maximum of such t_p, call it T, to see that $E^+(H) = J^+(H) \setminus I^+(H) \subseteq \exp([0, T] \times K)$, where $K := \{E_i(p) : p \in H, i = n-1, n\}$, hence $E^+(H)$ is relatively compact.
To see that it is closed, let $x_n \in E^+(H)$ with $x_n \to x$. Then $x_n \to x \in \exp([0, T] \times K)$, hence $x \in J^+(H)$. If $x \in I^+(H)$, then $x_n \in I^+(H)$ for large n, but this cannot happen. Hence $x \in E^+(H)$, thus $E^+(H)$ is compact. $\square$

3.2 The classical singularity theorems

In this final section, we will prove the classical singularity theorems of Hawking, Penrose, and Hawking-Penrose. They are results of very similar flavor, which can be summarized as follows:

Prototype singularity theorem: Let (M, g) be a spacetime satisfying

(E) an *energy condition*, e.g. the timelike/null/causal energy conditions or the genericity condition;

(C) a *causality condition*, e.g. (M, g) is chronological/causal/globally hyperbolic;

(I) an *initial condition*, e.g. there exists a trapped surface.

Then (M, g) is timelike/null/causal geodesically incomplete.

The singularity theorems have various applications in general relativity. Predicting, under physically reasonable assumptions, the existence of an incomplete causal geodesic, they hint at a problem of the spacetime manifold. Indeed, causal geodesic completeness is usually taken as a minimal requirement for a spacetime to be free of singularities (for a detailed discussion of the definition of a singularity in GR, see [11, Sec. 8.1]). Important physical phenomena, such as the Big Bang or the Big Crunch, can be understood by studying these theorems. We refer to Senovilla's seminal work [25] for a detailed analysis of many facets of singularity theorems (see also [26] for a recent review article).
While classically, the singularity theorems were always studied for smooth Lorentzian metrics, over the past decade efforts have been made to prove analogous results for metrics of lower regularity (see [8], [7], [16]).

3.2.1 The Hawking singularity theorem

We start our discussion of the singularity theorems with Hawking's singularity theorem. This important result has two variations: One of them assumes global hyperbolicity and an initial condition on a (not necessarily compact) Cauchy surface, the other simply requires the existence of a compact acausal spacelike hypersurface with positive convergence. Either form of Hawking's theorem predicts, if applied to the past, a Big Bang-like event.

Definition 3.2.1. (Convergence)
Let (M, g) be a spacetime and let $P \subseteq M$ be a semi-Riemannian submanifold. The *convergence* of P is the function $k : NP \to \mathbb{R}$ defined by

$$k_P(z) := \langle z, H_x \rangle, \quad z \in N_x P,$$

where H is the mean curvature vector field of P (see Definition 1.1.16).

If P is spacelike and $l = \dim P$ and $e_1, \dots, e_l$ are an orthonormal basis of T_pP, then

$$k_P(z) = \frac{1}{\dim P} \sum_{i=1}^{l} \langle z, II_p(e_i, e_i) \rangle = \frac{1}{\dim P} \sum_{i=1}^{l} \langle W_z(e_i), e_i \rangle = \frac{1}{\dim P} \operatorname{Tr}(W_z).$$

More specifically, if P is a spacelike hypersurface, then we may choose a future unit normal N and view k_P as $k_P = \langle N, H\rangle = \frac{1}{n-1}\operatorname{Tr}(W_N)$. In this context, since we chose a future unit normal, we call k_P the *future convergence* of P.
The following result and its proof are from [6, Lem. 3.4]. Recall that for any subset $C \subseteq M$ and any point $x \in M$, we may define the Lorentzian distance $d(C, x) := \sup_{y\in C} d(y, x)$.

Proposition 3.2.2. (Maximizers from causally complete subsets)
Let (M, g) be a globally hyperbolic spacetime and let $C \subseteq M$ be a closed subset such that $J^-(q)\cap C$ is compact for all $q \in J^+(C)$. Then $x \mapsto d(C, x)$ is a finite-valued, continuous function on M. Moreover, given any $q \in J^+(C)$, there is a maximizing geodesic α from C to q, i.e. $L(c) = d(C, q)$.

Proof. First note that $J^+(C)$ is closed by global hyperbolicity. Let $q \in J^+(C)$. By definition, here is a sequence $x_k \in J^-(q)\cap C$ with $\lim_k d(x_k, q) = d(C, q)$. By compactness of $J^-(q)\cap C$, there is some $x_0 \in J^-(q)\cap C$ to which a subsequence of the x_k, w.l.o.g. x_k itself, converges. It follows that

$$d(C, q) = d(x_0, q).$$

This establishes finiteness of $d(C, q)$. By global hyperbolicity, there is a maximizing geodesic from x_0 to q which thus maximizes the distance from C to q. It remains to show continuity: $x \mapsto d(C, x)$ is continuous on the open set $M \setminus J^+(C)$ because it equals 0 there. So we need to show continuity on $J^+(C)$. Let $q_k \to q \in J^+(C)$, we show that $d(C, q_k) \to d(C, q)$. Fix $q_0 \in I^+(q) \subseteq J^+(C)$, then by assumption $J^-(q_0) \cap C$ is compact. We will have $q_k \in J^-(q_0)$ for large k, hence $J^-(q_k) \cap C \subseteq J^-(q_0) \cap C$. By our arguments above, we may choose $p_k \in J^-(q_0) \cap C$ such that

$$d(C, q_k) = d(p_k, q_k).$$

By compactness, the p_k go to some $p \in J^-(q_0) \cap C$ and since the Lorentzian distance is continuous on $M \times M$ by global hyperbolicity, $d(p_k, q_k) \to d(p, q)$. Since $p \in C$,

$$d(p, q) \leq d(C, q) = d(x_0, q).$$

But also, since $x_0 \in C$,

$$d(x_0, q_k) \leq d(C, q_k) = d(p_k, q_k),$$

hence

$$\lim d(C, q_k) = \lim d(p_k, q_k) = \lim d(x_0, q_k) = d(x_0, q) = d(C, q),$$

which concludes the proof. $\square$

Subsets with the compactness property described in the result above are called *future causally complete*. *Past causal completeness* is defined analogously and the obvious time-dual of the result above holds. Note that compact sets and Cauchy surfaces are past and future causally complete.
In case one deals with a spacelike hypersurface in Proposition 3.2.2, a maximizer will always be focal-point free and start orthogonally (see Corollary 2.2.9 and Proposition 3.1.51).
The following analogous, though different, result can be proven, for which we give a reference.

Proposition 3.2.3. Let $H \subseteq M$ be a compact, acausal spacelike hypersurface and let $q \in D^+(H)$. Then there is a geodesic from H to q maximizing the distance $d(H, q)$. This geodesic necessarily starts orthogonal to H and has no focal points before q. It is timelike unless $q \in H$.

Proof. See [22, Thm. 14.44]. □

We are now ready to formulate and prove the first version of Hawking's singularity theorem. Assuming the timelike energy condition and a convergence condition on an initial Cauchy surface, future timelike geodesic incompleteness can be concluded.

Theorem 3.2.4. (Hawking Singularity Theorem, Version 1)
Let (M, g) be a globally hyperbolic spacetime satisfying the timelike energy condition and containing a spacelike Cauchy surface $C \subseteq M$ whose future convergence k_C satisfies

$$k_C \geq b > 0$$

for some $b > 0$. Then the arclength of every future timelike curve from C is bounded above by $\frac{1}{b}$. In particular, (M, g) is future timelike geodesically incomplete.

Proof. Let $q \in I^+(C)$, then there is a maximizing timelike g-unit speed geodesic $\gamma : [0, T] \to M$ from C to q by Proposition 3.2.2. γ starts normal to C and has no focal points of C before q. But by assumption,

$$k_C = \langle N, H \rangle = \frac{1}{n-1} \operatorname{Tr}(W_N) \geq b > 0,$$

which implies that γ will have a focal point if its length is greater than $\frac{1}{b}$ (see Corollary 3.1.48). It follows that

$$I^+(C) \subseteq \left\{ p \in M : d(C, p) \leq \frac{1}{b} \right\}.$$

This concludes the proof. □

Remark 3.2.5. (Achronality and acausality of hypersurfaces)
If H is a (closed, connected) spacelike hypersurface in M, we may w.l.o.g. assume that it is acausal for our purposes: There is some Lorentzian covering $\tilde{M} \to M$ and an *achronal* closed connected hypersurface $\tilde{H} \subseteq \tilde{M}$ that is isometric to H under the covering map (cf. [22, Prop. 14.48]). Moreover, achronal spacelike hypersurfaces are always acausal (cf. [22, Lem. 14.42]). In the context of singularity theorems, we are interested in geodesic (in)completeness of M, which is equivalent to incompleteness of any Lorentzian covering space of M.

We come to the second version of Hawking's theorem. Here, global hyperbolicity is not assumed. The existence of a spacelike hypersurface with positive future convergence that is compact turns out to be sufficient to yield future timelike geodesic incompleteness. We do not, however, get an explicit bound on the Lorentzian lengths of future timelike curves as we did in the first version of the theorem.

Theorem 3.2.6. (Hawking Singularity Theorem, Version 2)
Let (M, g) be a spacetime satisfying the timelike energy condition and let $H \subseteq M$ be a compact acausal spacelike hypersurface whose future convergence k_H satisfies

$$k_H > 0.$$

Then (M, g) is future timelike geodesically incomplete.

Proof. We may assume that H is connected. Since H is compact, k_H assumes a minimum $b > 0$. Using Proposition 3.2.3 and proceeding as in the proof of Theorem 3.2.4, it is easy to conclude that

$$D^+(H) \subseteq \left\{ p \in M : d(H, q) \leq \frac{1}{b} \right\}.$$

We may assume that $H^+(H)$ is nonempty, otherwise H is a (future) Cauchy surface and the same arguments as in Theorem 3.2.4 apply to give the result. In fact, we intend to show the following (stronger) claim: There is a future inextendible future directed timelike geodesic starting orthogonally from H that has arclength $1/b$. In particular, this geodesic is future incomplete and this gives the claim.
We will argue by contradiction, so assume there is no such geodesic. Let $q \in H^+(H)$, then there is a maximizing geodesic from H to q starting orthogonally to H, i.e. its length is $d(H, q) \leq \frac{1}{b}$ (where the inequality follows from above by continuity since $H^+(H) \subseteq \overline{D^+(H)}$): This is a simple compactness argument upon choosing $q_n \in D^+(H)$ with $q_n \to q$ and using Proposition 3.2.3 to connect H to q_n by a maximizing geodesic starting orthogonal to H. Observe that the limit is indeed defined up to q because, by indirect

assumption, there is no future inextendible normal geodesic from H that has arclength $\leq 1/b$.
Next, we consider (past) generators of $H^+(H)$ as described in Proposition 3.1.16 and we show that $p \mapsto d(H,p)$ is strictly decreasing along such generators. For this, let α be a past directed generator of $H^+(H)$ and let $s < t$ in its domain of definition. By Proposition 3.2.3, we know that we can find maximizers from H to $\alpha(s)$ resp. $\alpha(t)$, but the one to $\alpha(t)$ can be traveled up along α to $\alpha(s)$ to give a broken geodesic which cannot maximize, hence

$$d(H, \alpha(s)) > d(H, \alpha(t)).$$

We derive a contradiction: Let $B \subseteq TM|_H$ be the union of the zero section with all future pointing vectors v satisfying $\sqrt{|g(v,v)|} \leq 1/b$. Then B is compact since H is. By assumption, the normal exponential map of H is defined on B and $H^+(H)$ is contained in its compact image. Since $H^+(H)$ is closed, it follows that $H^+(H)$ is compact. By lower semicontinuity (cf. [22, Lem. 14.17]), $p \mapsto d(H,p)$ assumes a finite minimum on $H^+(H)$. But this is absurd, since we showed that $p \mapsto d(H,p)$ is strictly decreasing along past directed generators, and from each point in $H^+(H)$ we can find a past inextendible generator (see Proposition 3.1.16; such a generator is always past inextendible in our case because H is a hypersurface and hence edgeless). $\square$

3.2.2 The Penrose singularity theorem

The Penrose singularity theorem describes a gravitational collapse scenario. It gives null geodesic incompleteness under the assumption that the spacetime is globally hyperbolic with a *noncompact* Cauchy surface and a *compact* trapped surface. The interplay between these assumptions is essential for the proof.

Theorem 3.2.7. (Penrose Singularity Theorem)
Let (M,g) be a globally hyperbolic spacetime of dimension $n \geq 3$ with a noncompact Cauchy surface satisfying the null energy condition. If there is an achronal trapped surface H in M, then M is null geodesically incomplete.

Proof. Suppose M is null geodesically complete. We assume H is a future trapped surface, then by (the proof of) Proposition 3.1.59 $E^+(H)$ is compact. Since M is globally hyperbolic, $J^+(H)$ is closed and thus $E^+(H) = \partial J^+(H)$, in particular it is a closed, achronal, topological hypersurface by Corollary 1.3.7 (since $J^+(H)$ is a future set). Now let C be a Cauchy surface of M and $\rho : M \to C$ the corresponding retraction (see Proposition 3.1.3). Then $\rho|_{E^+(H)}$ is continuous and injective by the achronality of $E^+(H)$. But $E^+(H)$ is a topological hypersurface, hence by Brouwer's invariance of domain theorem (see [4, Thm. 4.5]) $\rho|_{E^+(H)}$ must be a homeomorphism. However, $E^+(H)$ is compact and C is not, a contradiction. $\square$

3.2.3 The Hawking–Penrose singularity theorem

The final subsection of this chapter is dedicated to proving the most general of the classical singularity theorems, the Hawking-Penrose theorem. It describes a vast array of scenarios under weak causality assumptions on the spacetime.

Proposition 3.2.8. (Conjugate points $\Rightarrow$ strong causality)
If (M, g) is a chronological spacetime such that each inextendible null geodesic has a pair of conjugate points, then (M, g) is strongly causal.

Proof. Suppose (M, g) fails to be strongly causal at $p \in M$, and fix a convex neighborhood U of p. Then there is a sequence of neighborhoods $V_k \to \{p\}$ of p, $V_k \subseteq U$, and future causal curves γ_k that start and end in V_k, but leave U. By the limit curve theorem (see [1, Prop. 3.31]), the γ_k converge uniformly on compact subsets to a limit causal curve γ. Since the V_k collapse to p, γ is closed. Since (M, g) is chronological, no two points on γ are timelike related and thus γ is a null geodesic. But then γ can be seen as a complete, everywhere maximizing null geodesic, contradicting the assumption that all inextendible null geodesics have conjugate points. $\square$

Proposition 3.2.9. (Strong causality or null incompleteness)
Let (M, g) be a chronological spacetime of dimension $n \geq 3$ satisfying the genericity condition and the causal energy condition. Then (M, g) is either strongly causal or null geodesically incomplete.

Proof. By Theorem 3.1.41, every complete null geodesic has a pair of conjugate points. So if (M, g) is null geodesically complete, it must be strongly causal by the previous result Proposition 3.2.8. $\square$

Proposition 3.2.10. (Causal disconnectedness and incompleteness)
Let (M, g) be a chronological spacetime of dimension $n \geq 3$ satisfying the genericity condition and the causal energy condition. If (M, g) is causally disconnected, then it is causal geodesically incomplete.

Proof. Suppose (M, g) is causal geodesically complete. Then it is strongly causal by above, and by Theorem 3.1.41 each causal geodesic has a pair of conjugate points. But since (M, g) is causally disconnected, strongly causal and causal geodesically complete, there exists a causal line in (M, g), which cannot have conjugate points, a contradiction. $\square$

Proposition 3.2.11. (Causal disconnection by horismos)
Let (M, g) be a chronological spacetime of dimension $n \geq 3$ such that each inextendible null geodesic contains a pair of conjugate points. If there exists a future resp. past trapped set $S \subseteq M$, then (M, g) is causally disconnected by the compact set $E^+(S)$ resp. $E^-(S)$.

Proof. Suppose S is future trapped, the other case is analogous. By Proposition 3.2.8, (M, g) is strongly causal. Moreover, there exists a future inextendible timelike curve γ in $D^+(E^+(S))$ (see Corollary 3.1.21). Extend γ maximally in M. Then γ meets $E^+(S)$ in a point r (which is unique due to achronality of $E^+(S)$). By strong causality, γ leaves every compact set, so we may take sequences p_n and q_n on γ going to infinity such that $p_n \ll r \ll q_n$. Let now λ be a causal curve from p_n to q_n. We may extend λ to an inextendible causal curve by traveling γ up to p_n, then following λ to q_n, and then taking γ again from q_n. Since $q_n \in D^+(E^+(S))$, this new inextendible causal curve must meet $E^+(S)$. But since γ meets $E^+(S)$ only at r and nowhere else, it follows that λ meets $E^+(S)$. Hence $E^+(S)$ (together with the sequences p_n and q_n) causally disconnects (M, g). $\square$

In Proposition 3.1.59 we saw that if (M, g) satisfies the null energy condition and has a trapped surface H, then one of the sets $E^{\pm}(H)$ is compact (depending on whether H is a future or past trapped surface), or (M, g) is null geodesically incomplete. Since a trapped surface need not be achronal in general, we cannot conclude that H is a trapped set in case of null geodesic completeness in this context. However, if (M, g) is strongly causal, we can still show the existence of a trapped set.

Proposition 3.2.12. (Trapped sets or null incompleteness)
Let (M, g) be a spacetime of dimension $n \geq 3$ that is strongly causal and that satisfies the null energy condition. If there is a trapped surface H in M, then (at least) one of the following holds:

(1) There is a trapped set in M.

(2) (M, g) is null geodesically incomplete.

Proof. Suppose (M, g) is null geodesically complete. Then by Proposition 3.1.59, we may assume that $E^+(H)$ is compact. We set

$$S := E^+(H) \cap H$$

and show that S is a future trapped set. Clearly S is achronal (since $E^+(H)$ is) and compact (since both $E^+(H)$ and H are).
Note that $S = E^+(H) \cap H = (J^+(H) \setminus I^+(H)) \cap H$. If we assume that S is empty, then $H \subseteq I^+(H)$. Since H is compact, cover it with a finite number of sets $I^+(p_i)$ with $p_i \in H$. From this, it is easy to conclude the existence

of a closed timelike curve, which contradicts strong causality. Hence S is nonempty.
We now argue that $E^+(S)$ is compact, more specifically we will show that $E^+(S) = E^+(H)$, which is known to be compact. For this, cover H by convex open sets $U_1, \dots, U_k$ such that no causal curve leaving U_i returns to it (which is possible by strong causality). Since H is spacelike, it is locally achronal, so we may assume that $U_i \cap H$ is achronal for all i.
Claim: $I^+(S) = I^+(H)$: Clearly $\subseteq$ holds because $S \subseteq H$, for $\supseteq$ suppose there is $q \in I^+(H) \setminus I^+(S)$. Choose $p_1 \in H$ with $p_1 \ll q$. Then $p_1 \in U_{\sigma(1)} \cap H$ for some permutation σ of k numbers. Since $q \notin I^+(S)$, $p_1 \notin S$ and due to $p_1 \in H$ we conclude $p_1 \notin E^+(H)$ (note that H is not contained in $E^+(H)$ because then H would be achronal and the result of this proposition would be trivial). Hence $p_1 \in H \cap I^+(H)$, so there is $p_2 \in H$ with $p_2 \ll p_1$. By achronality, $p_2 \notin U_{\sigma(1)} \cap H$, hence $p_2 \in U_{\sigma(2)} \cap H$ (the permutation σ being altered appropriately along the proof). Since $p_2 \ll p_1 \ll q$ and $q \notin I^+(S)$, we again have $p_2 \notin E^+(H)$ as before. Hence $p_2 \in H \cap I^+(H)$, so there is $p_3 \in H$ with $p_3 \ll p_2$ with $p_3 \notin U_{\sigma(1)} \cup U_{\sigma(2)}$. Continuing this way, we obtain an infinite number of points p_i and corresponding distinct convex sets $U_{\sigma(i)}$, which cannot happen. Hence $I^+(S) = I^+(H)$.
Claim: $J^+(S) = J^+(H)$: Again $\subseteq$ is immediate. For $\supseteq$ assume $q \in J^+(H) \setminus J^+(S)$. Since $I^+(S) = I^+(H) \subseteq J^+(S)$, it follows that there is a null curve from some $p \in H$ to q. Then $p \notin I^+(H)$ because $p \leq q$ and $q \notin I^+(H)$. Hence $p \in E^+(H)$, thus $p \in S$ and hence $q \in J^+(S)$, a contradiction. Hence $J^+(S) = J^+(H)$.
Altogether, we have shown that $E^+(S) = E^+(H)$, hence S is a future trapped set. □

Definition 3.2.13. (Trapped point)
Let (M, g) be a spacetime of dimension $n \geq 3$. A point $p \in M$ is called *trapped point* if for every future null geodesic $\beta : [0, a) \to M$ from p, the expansion $\overline{\theta}$ of the unique Lagrange tensor class $\overline{A}$ along β with $\overline{A}(0) = 0$ and $\overline{A}'(0) = \mathrm{id}$ becomes negative.

Proposition 3.2.14. (Trapped points and null incompleteness)
Let (M, g) be a spacetime of dimension $n \geq 3$ satisfying the null energy condition. Let $p \in M$ be a trapped point. Then at least one of the following holds:

(1) $\{p\}$ is a trapped set.

(2) (M, g) is null geodesically incomplete.

Proof. Suppose (M, g) is null geodesically complete. By Proposition 3.1.38, each null geodesic from p has a conjugate point to p. The set of (normalized) future null vectors at p is compact, hence we conclude that $E^+(p)$ is compact just as in Proposition 3.1.59. □

Theorem 3.2.15. (Mutually exclusive assumptions)
There is no spacetime (M, g) of dimension $n \geq 3$ that satisfies all of the following conditions.

(1) (M, g) is chronological.

(2) Every inextendible causal geodesic in (M, g) has a pair of conjugate points.

(3) There is a trapped set in M.

Proof. Suppose (M, g) is a spacetime of dimension $n \geq 3$ satisfying (1)-(3). Then by Proposition 3.2.8, (M, g) is strongly causal. By Proposition 3.2.11, (M, g) is causally disconnected. Finally, by Proposition 3.1.25, there exists a (not necessarily complete) causal line in M, which cannot be since any inextendible causal geodesic must possess conjugate points by assumption. □

Corollary 3.2.16. (Trapped set $\Rightarrow$ singularity)
Let (M, g) be a chronological spacetime of dimension $n \geq 3$ satisfying the causal energy and genericity conditions. If there is a trapped set in M, then (M, g) is causal geodesically incomplete.

Proof. Suppose (M, g) is such a spacetime and it is causal geodesically complete. Then each inextendible causal geodesic has a pair of conjugate points by Theorem 3.1.41. But then it satisfies (1)-(3) from Theorem 3.2.15, a contradiction. □

After these preparations, we are ready to state and prove the final result of this thesis, the Hawking-Penrose singularity theorem.

Theorem 3.2.17. (Hawking-Penrose Singularity Theorem)
Let (M, g) be a chronological spacetime of dimension $n \geq 3$ satisfying the causal energy condition and genericity condition. Suppose one of the following is satisfied:

(1) There is a trapped surface in M.

(2) There is a trapped point in M.

(3) There is a (nonempty) compact, achronal edgeless set in M.

Then (M, g) is causal geodesically incomplete.

Proof. Suppose (M, g) is causal geodesically complete. Then (1) and (2) imply that (M, g) has a trapped set, cf. Proposition 3.2.12 and Proposition 3.2.14. In the case of (3), if S is a compact, achronal, nonempty, edgeless set in M, then S is a topological hypersurface by Theorem 1.3.5, and we can assume that S is connected. We claim that $E^+(S) = S$: Suppose $q \in E^+(S) \setminus S$, and let β be any null geodesic from $p \in S$ to q. Since $q \notin \mathrm{edge}(A)$, β must enter $I^+(S)$ immediately after p (see [14, Cor. 8.3.1]), a contradiction. Hence S is itself a trapped set. In any of the cases, Corollary 3.2.16 gives a contradiction to the assumption that (M, g) is causal geodesically complete. $\square$

Abstract

The aim of this thesis is to give a self-contained and pedagogical presentation of the three classical singularity theorems of general relativity, namely the theorems of Hawking, Penrose, and Hawking-Penrose. These theorems state that, under physically reasonable assumptions, spacetime has to contain a singularity (i.e. an incomplete causal geodesic). Along the way, we develop the geometric background that is needed to formulate and prove them. This work is motivated by recent developments in the study of singularity theorems for metrics of lower regularity (see e.g. [8], [7], [24], [16]) and the subsequent desire to have a solid and comprehensive account of the classical results.

In Chapter 1, we study semi-Riemannian submanifolds. We collect various results on the relations between the intrinsic geometry of the submanifold and its extrinsic geometry that relates it to the ambient manifold. The results are mostly formulated for general semi-Riemannian submanifolds and semi-Riemannian ambient spaces, but later, we are only interested in Lorentzian ambient spaces and (mostly) Riemannian submanifolds. In this chapter, we have followed the presentation in [22, Ch. 4] and [19, Ch. 8], with the occasional input from [14, Sec. 4.4].

Chapter 2 is dedicated to variation theory of curves in Lorentzian manifolds. A variation is a collection of (piecewise) smooth nearby curves, and much can be said about a variation by studying its transverse derivative. If the nearby curves are geodesics, then the transverse derivative is a Jacobi field, whose vanishing at certain points gives us information about the focusing of geodesics. Such focusing results can then be used to deduce whether or not a given causal geodesic maximizes the Lorentzian distance or not. The tools developed in Chapter 2 are essential ingredients in the proofs of the singularity theorems. In our account on variation theory, we proceeded along the lines of [1, Ch. 10]. Some supplementary material was adapted from [14, Sec. 4.6], [22, Ch. 10] and [11, Ch. 6,8].

Chapter 3 is the final chapter of this work and is devoted to proving the aforementioned singularity theorems of Hawking, Penrose, and Hawking-Penrose. Some preparatory material preceeds the singularity theorems to put some causality notions into context and clear up the presentation. These results are collected from [1, Ch. 8,12], [14, Ch. 8,9] and [22, Ch. 14]. For the theorems of Hawking and Penrose, we follow the proofs in [22, Ch. 14]. The proof of the Hawking-Penrose Theorem is adapted from [1, Ch. 12].

Zusammenfassung

Das Ziel dieser Arbeit ist eine vollständige und eigenständige Präsentation der klassischen Singularitätentheoreme der allgemeinen Relativitätstheorie, d.h. der Theoreme von Hawking, Penrose, und Hawking-Penrose. Diese Theoreme implizieren die Existenz einer Singularität (d.h. einer unvollständigen kausalen Geodäte) in der Raumzeit unter physikalisch plausiblen Annahmen. Wir behandeln den zugehörigen geometrischen Hintergrund, welcher für ihre Formulierungen und Beweise notwendig ist. Die Motivation dafür liefern neue Entwicklungen, insbesondere aus dem Studium der Singularitätentheoreme für Metriken mit niedriger Regularität (siehe z.B. [8], [7], [24], [16]) und dem daraus folgenden Bedürfnis, ein umfassende Quelle der klassischen Resultate zu haben.

In Kapitel 1 studieren wir semi-Riemannsche Untermannigfaltigkeiten. Wir sammeln verschiedenste Resultate, welche die intrinsische Geometrie der Untermannigfaltigkeit und ihre extrinsischen Geometrie im Kontext der umgebenden Mannigfaltigkeit in Beziehung setzen. Die Ergebnisse sind meist für allgemeine semi-Riemannsche Mannigfaltigkeiten formuliert, für uns jedoch später meist nur im Kontext von Lorentzmannigfaltigkeiten und Riemannschen Untermannigfaltigkeiten von Interesse. Die Hauptreferenzen für dieses Kapitel sind [22, Ch. 4] und [19, Ch. 8], wobei sich auch einige Ideen aus [14, Sec. 4.4] finden lassen.

Kapitel 2 ist dem Studium der Variationstheorie von Kurven in Lorentzmannigfaltigkeiten gewidmet. Eine Variation ist eine (stückweise) glatte Schar von Kurven nahe einer Referenzkurve, und ihre transversale Ableitung beinhaltet viel Information über ihre Eigenschaften. Falls die Kurven einer Variation allesamt Geodäten sind, so ist ihre transversale Ableitung ein Jacobifeld, dessen Verschwinden an bestimmten Punkten Aufschluss über das Fokussieren von Geodäten gibt. Aus solchen Resultaten lässt sich wiederum bestimmen, ob eine gegebene Geodäte die Lorentzdistanz maximiert oder nicht. Die Werkzeuge, die wir im Laufe von Kapitel 2 entwickeln, spielen eine essentielle Rolle in den Beweisen der Singularitätentheoreme. In unserer Darstellung der Variationstheorie folgen wir [1, Ch. 10]. Einiges an Zusatzmaterial stammt aus [14, Sec. 4.6], [22, Ch. 10] und [11, Ch. 6,8].

Das abschließende 3. Kapitel ist den Beweisen der Singularitätentheoreme von Hawking, Penrose, und Hawking-Penrose, gewidmet. Wir beginnen mit einigen Vorbereitungen aus der Kausalitätstheorie, um die Präsentation etwas aufzulockern. Diese Ergebnisse stammen aus [1, Ch. 12], [14, Ch. 8,9] und [22, Ch. 14]. Unsere Beweise des Hawking- und des Penrose-Singularitätentheorems folgen [22, Ch. 14], während unser Beweis des Hawking-Penrose Theorems [1, Ch. 12] folgt.

Bibliography

[1] Beem, J. K., Ehrlich, P. E., Easley, K. L., et al. Global Lorentzian Geometry. Marcel Dekker, Inc., 1996.

[2] Bernal, A. N., and Sánchez, M. Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'. Classical and Quantum Gravity 24, 3 (2007).

[3] Bär, C. Lorentzgeometrie. Lecture Notes, https://www.math.uni-potsdam.de/fileadmin/user_upload/Prof-Geometrie/Dokumente/Lehre/Lehrmaterialien/skript-LorGeo.pdf, 2006. Accessed: 10.07.2021.

[4] Du, A. Homology, Brouwer's Fixed Point Theorem and Invariance of Domain. Lecture Notes, http://math.columbia.edu/~mmiller/TProjects/ADu20s.pdf, 2020. Accessed: 31.05.2020.

[5] Eschenburg, J.-H., and O'Sullivan, J. J. Growth of Jacobi fields and divergence of geodesics. Mathematische Zeitschrift 150, 3 (1976).

[6] Galloway, G. J., and Vega, C. Achronal limits, lorentzian spheres, and splitting. In Annales Henri Poincaré (2014), vol. 15, Springer.

[7] Graf, M. Singularity theorems for C^1-Lorentzian metrics. Communications in Mathematical Physics 378, 2 (2020).

[8] Graf, M., Grant, J. D., Kunzinger, M., and Steinbauer, R. The Hawking–Penrose singularity theorem for $C^{1,1}$-Lorentzian metrics. Communications in Mathematical Physics 360, 3 (2018).

[9] Griffiths, J. B., and Podolskỳ, J. Exact space-times in Einstein's general relativity. Cambridge University Press, 2009.

[10] Gromoll, D., Klingenberg, W., and Meyer, W. Riemannsche Geometrie im Grossen, vol. 55. Springer-Verlag, 2006.

[11] Hawking, S. W., and Ellis, G. F. R. The large scale structure of space-time, vol. 1. Cambridge University Press, 1973.

[12] Horn, R. A., and Johnson, C. R. Matrix analysis. Cambridge University Press, 2012.

[13] Hounnonkpe, R., and Minguzzi, E. Globally hyperbolic spacetimes can be defined without the 'causal' condition. Classical and Quantum Gravity 36, 19 (2019).

[14] Kriele, M. Spacetime: Foundations of General Relativity and Differential Geometry, vol. 59. Springer Science & Business Media, 1999.

[15] Kunzinger, M. Comparison Geometry. Lecture Notes, 2013.

[16] Kunzinger, M., Ohanyan, A., Schinnerl, B., and Steinbauer, R. The Hawking–Penrose singularity theorem for C^1-Lorentzian metrics. Communications in Mathematical Physics (2022). To appear.

[17] Lee, J. Introduction to Topological Manifolds, vol. 202. Springer Science & Business Media, 2010.

[18] Lee, J. M. Introduction to Smooth Manifolds. Springer, 2013.

[19] Lee, J. M. Introduction to Riemannian Manifolds. Springer, 2018.

[20] Minguzzi, E. Limit curve theorems in lorentzian geometry. Journal of mathematical physics 49, 9 (2008), 092501.

[21] Minguzzi, E. Lorentzian causality theory. Living reviews in relativity 22, 1 (2019).

[22] O'neill, B. Semi-Riemannian Geometry with Applications to Relativity. Academic press, 1983.

[23] Sämann, C. Global hyperbolicity for spacetimes with continuous metrics. In Annales Henri Poincaré (2016), vol. 17, Springer.

[24] Schinnerl, B., and Steinbauer, R. A note on the Gannon-Lee theorem. arXiv preprint arXiv:2101.04007 (2021).

[25] Senovilla, J. M. Singularity theorems and their consequences. General Relativity and Gravitation 30, 5 (1998).

[26] Senovilla, J. M. A critical appraisal of the singularity theorems. arXiv preprint arXiv:2108.07296 (2021).