Universidade de São Paulo
Instituto de Física

Teoremas de singularidade e condições de energia

Rodrigo de Azeredo Orselli

João Carlos Alves Barata

Supervisor: Prof. Dr. João Carlos Alves Barata

Dissertação submetida ao Instituto de Física da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Banca Examinadora:
Prof. Dr. João Carlos Alves Barata - Orientador (IF-USP)
Prof. Dr. Claudio Gorodski (IME-USP)
Prof. Dr. André Gustavo Scagliusi Landulfo (CCNH-UFABC)

São Paulo
2023

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Orselli, Rodrigo de Azeredo

Teoremas de singularidade e condições de energia / Singularity theorems and energy conditions. São Paulo, 2023.

Dissertação (Mestrado) - Universidade de São Paulo. Instituto de Física. Depto. de Física Matemática.

Orientador(a): Prof. Dr. João Carlos Alves Barata

Área de Concentração: Relatividade e Gravitação

Unitermos: 1. Física Matemática; 2. Relatividade; 3. Geometria - Semi-Riemanniana; 4. Teoria Quântica de Campo.

USP/IF/SBI-006/2023

University of São Paulo
Physics Institute

Singularity theorems and energy conditions

Rodrigo de Azeredo Orselli

Supervisor: Prof. Dr. João Carlos Alves Barata

Dissertation submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Master of Science.

Examining Committee:
Prof. Dr. João Carlos Alves Barata - Supervisor (IF-USP)
Prof. Dr. Claudio Gorodski (IME-USP)
Prof. Dr. André Gustavo Scagliusi Landulfo (CCNH-UFABC)

São Paulo
2023

Acknowledgements

This work was made possible by the financial support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) *, via grant 130767/2019-0. The author would like to extend his gratitude for the funding received during the first two years of this work.*

Abstract

The singularity theorems proved by Penrose and Hawking between 1965 and 1970 settled a decades-long debate concerning the existence of singularities in General Relativity. However, they are of limited applicability when the quantum behaviour of matter is taken into account, as the Strong and Null Energy Conditions, which these theorems require, are known (since at least the work by Epstein, Glaser and Jaffe in 1965) not to be valid in Quantum Field Theories. Extensions of Hawking and Penrose's theorems with alternative energy conditions were found as early as the late 1970s by Tipler, Roman, Wald, Yurtsever and others. But it was not until 2011, with work by Fewster and Galloway, that singularity theorems with energy conditions inspired by the energy inequalities that quantum fields are believed to verify were first proved. This work was subsequently updated by Fewster and Kontou, who in a 2019 paper showed how essentially the same theorems could be obtained with a new, simpler strategy, which replaced the traditional way of detecting focal points via the Raychaudhuri inequality with a variational method. In this dissertation, we give a detailed review of these developments in the field of singularity theorems, with particular attention to the most recent results by Fewster and Kontou and the new mathematical approach utilised therein.

Keywords: General Relativity, Differential Geometry, Singularity Theorems, Quantum Energy Inequalities, Quantum Field Theory

Resumo

Os teoremas de singularidade provados por Penrose e Hawking entre 1965 e 1970 concluíram um debate de décadas a respeito da existência de singularidades na Relatividade Geral. No entanto, sua aplicabilidade é limitada quando o comportamento quântico da matéria é levado em consideração: as Condições Forte e Nula de Energia, requeridas por estes teoremas, não são válidas em Teorias Quânticas de Campos, como se sabe desde, pelo menos, o trabalho de Epstein, Glaser e Jaffe, em 1965. Extensões dos teoremas de Hawking e Penrose com condições de energia alternativas vêm sendo encontradas desde o final da década de 1970 por Tipler, Roman, Wald, Yurtsever e outros. Mas foi somente em 2011, em trabalho de Fewster e Galloway, que apareceram pela primeira vez teoremas de singularidade com condições de energia inspiradas pelas desigualdades do tipo verificado por campos quânticos. Posteriormente, este trabalho foi atualizado por Fewster e Kontou, que, em artigo de 2019, mostraram como essencialmente os mesmos teoremas podiam ser obtidos através de uma nova e mais simples estratégia, que consiste em substituir o método tradicional para detecção de pontos focais via a desigualdade de Raychaudhuri por uma técnica variacional. Nesta dissertação, apresentaremos uma revisão detalhada destes desenvolvimentos no campo dos teoremas de singularidade, com atenção especial aos resultados mais recentes de Fewster e Kontou e aos novos métodos matemáticos ali empregados.

Palavras-Chave: Relatividade Geral, Geometria Diferencial, Teoremas de Singularidade, Desigualdades Quânticas de Energia, Teoria Quântica de Campos

Contents

Chapter 1

Introduction

The singularity theorems published by Penrose and Hawking between 1965 and 1970 [1–5] were perhaps the most consequential developments in General Relativity since the introduction of the theory by Einstein, in 1915. They settled once and for all the question of the existence of singularities within the framework of General Relativity by establishing that, under a few mild hypotheses, these were generic features of the theory. Most crucially, their assumptions made no mention of symmetries of the spacetime, which were frequently pointed to as the reason why early solutions of the Einstein equations, such as the cosmological solutions of Friedman-Lemaître-Robertson-Walker and Schwarzschild's black hole solution, displayed singularities.

The influence of the works of Hawking and Penrose, especially the latter, were not limited, however, to the way General Relativity was perceived conceptually; it also changed the way the theory was approached and worked with mathematically. The most celebrated example of this is the formalisation of the concept of singularity as geodesic incompleteness, attributed to Penrose himself. Whereas previously the term was used broadly to describe some sort of catastrophic behaviour of a model—usually the divergence of some curvature scalar—, Penrose's mathematical description of singularities is more general, as the occurrence of incomplete causal geodesics may or may not be accompanied by curvature singularities. As other contributions of the singularity theorems to the theoretical framework of General Relativity, one could mention the application of techniques of differential topology to study the causal properties of spacetime in a much more structured way, and concepts such as *trapped surfaces*, which in Penrose's theorem expresses the notion of the "point of no return" in gravitational collapse [6].

In his 1998 review on the subject, Senovilla [7] proposes a logical structure which singularity theorems generally follow. Namely, to prove the existence of singularities, one needs three types of assumptions:

- **Causality conditions:** these exclude spacetimes with physically strange causal properties, such as closed causal curves;

- **Curvature conditions:** these control the focussing of geodesics due to gravitational effects. Using Einstein's equation, they can be translated into restrictions on the energy-momentum tensor of the theory, and for this reason they are more commonly referred to as **energy conditions**;

- **Initial conditions:** these describe the rate of convergence of a family of causal geodesics, or the extrinsic curvature of some spacelike region in spacetime, in an initial situation.

Then, the proof makes use of the hypotheses above as follows:

1. Assuming the geodesic completeness of spacetime, the causality condition implies that all (or at least one) of the curves in the family with which the initial condition deals is free of focal points;

2. Using the initial condition and the energy condition, one studies the evolution of that family of geodesics and concludes that one (or all) of its members eventually reaches a focal point;

3. Items 1. and 2. are in contradiction; therefore, either one or more of the hypotheses have to be discarded, or the initial assumption of geodesic completeness itself. In the latter case, the existence of a singularity is proven.

For a concrete example, we state below the Hawking and Penrose theorems to show how their hypotheses fall within this scheme.

Theorem 1 (Hawking)**.** *Let M be a globally hyperbolic spacetime and S a compact Cauchy hypersurface of M. If:*

1. $\mathrm{Ric}(t,t) \geq 0$ *for every timelike vector* t *(i.e., the* ***Strong Energy Condition*** *holds in* M*); and*

2. *The convergence* k *of* S *satisfies* $k > 0$*,*

then M *is future timelike geodesically incomplete.*

In the statement above, the assertion that M is globally hyperbolic with compact Cauchy surfaces is the causality condition. Naturally, the Strong Energy Condition is the energy condition required by Hawking's theorem. Finally, item 2 plays the role of the initial condition (the precise definition of the convergence of a submanifold will be seen in Chapter 2). The Hawking theorem predicts, under appropriate circumstances, the occurrence of a *big crunch* in cosmological models (by time-reversion, assuming instead that S has *negative* convergence, one obtains past geodesically incompleteness, i.e., a *big bang*). Note how the conclusions do not depend on any symmetry of M; in particular, it showed that the singularities of FLRW models were not an artifact of the assumption of isotropy.

Theorem 2 (Penrose)**.** *Let* M *be a globally hyperbolic spacetime with noncompact Cauchy hypersurfaces. If:*

1. $\mathrm{Ric}(l,l) \geq 0$ *for all lightlike* l *(i.e., the* ***Null Energy Condition*** *holds); and*

2. M *contains a trapped surface,*

then M *is future lightlike geodesically incomplete.*

In the Penrose theorem, item 1 is the energy condition, and item 2 is the initial condition (the full definition of a trapped surface is intricate and will not concern us at this point). The statement that M is globally hyperbolic and its Cauchy surfaces are noncompact is the causality condition. Penrose's theorem deals with gravitational collapse and the formation of black holes.

The question of whether the singularity theorems of Hawking and Penrose can be strengthened by relaxing one or more of its hypotheses is then one that arises naturally, and Senovilla's scheme can be used to classify these extensions according to which type of condition they seek to weaken. This text is concerned exclusively with singularity theorems with alternative energy conditions, as this class of extensions is of pressing physical relevance. For a list of references on work with causality and initial conditions, see [6].

In classical field theory, the Strong and Null Energy Conditions express the positivity of energy, and are thus considered a requirement for a physically reasonable model. It is remarkable, however, that even some very simple classical theories, like the nonminimally coupled Klein-Gordon field, these conditions can be violated [8]. The situation becomes even more complicated in the quantum setting. In 1965, it was observed by Epstein, Glaser and Jaffe [9] that pointwise conditions on the positivity of energy (in fact, of any observable), such as the Strong and Null Energy Conditions, cannot hold in any quantum field theory under the Wightman axioms. This is the main reason why energy conditions draw particular interest in attempts to extend the singularity theorems.

However, the quantum violations of the positivity of energy must be somehow constrained. Otherwise, according to an argument originally due to Ford [10], macroscopic violations of the Second Law of Thermodynamics would be observable, which contradicts experience. It is thus expected on physical grounds that every quantum field satisfies some *Quantum Energy Inequality*, and the search for these is an active area of research. In rough terms, what these conditions are expected to enforce is that a negative energy density in a region of spacetime should be balanced out by positive energy densities in its neighbouring regions. This sort of compensation is sometimes referred to as “quantum interest”.

Unfortunately, the derivation of quantum energy inequalities is a very technical problem, and, as far as is currently known, must be done in a largely case-by-case basis; a general rule for obtaining an inequality for a given quantum field theory does not exist. This is even more so in quantum field theories on curved spacetimes, where one runs into some problems which are not yet fully solved in this setting, such as renormalisation and the inexistence of a preferred vacuum state. Nevertheless, several isolated examples exist; references [11–23] constitute a brief list. For an thorough review of the current state of knowledge on quantum energy inequalities and a more exhaustive bibliography, see [24].

The first singularity theorems with weakened energy conditions were proved by Tipler [25, 26] and Roman [27, 28], between 1978 and 1988, followed by Wald and Yurtsever [29] in 1991. The energy hypotheses of these theorems were essentially averaged versions of the Strong, Weak and Null Energy Conditions, of the general form

$$\int_\gamma F\big(\gamma'(u), \gamma'(u)\big)\, \mathrm{d}u \geq 0\,,$$

where, according to which energy condition is being replaced, γ can be a timelike or lightlike geodesic and F can be the Ricci tensor or the energy-momentum tensor.

In spite of the alternative energy conditions that they employ, all of the extensions of the Hawking-Penrose theorems mentioned above are of a purely classical character. Indeed, a complete answer to the question of whether singularities still appear when the quantum behaviour of matter is taken into account is evidently not possible in the absence of a quantum theory of gravitation itself. However, one can look for indications in semiclassical gravity, where the Einstein equation is substituted by

$$G = 8\pi\langle T\rangle_\omega,\,. \tag{1.1}$$

In this equation, the classical energy-momentum tensor is replaced with the expected value of the energy-momentum of a *quantum* field theory, calculated in a given state ω. Thus, the gravitational field retains its classical character from General Relativity, while the other matter fields are treated via a quantum theory.

It is in this context that the first singularity theorems with energy conditions inspired by quantum energy inequalities appeared. This was achieved by C. J. Fewster and G. J. Galloway in 2011 [30]. The curvature assumption in their work is that the “energy density” $\rho(u) = \mathrm{Ric}\,\big(\gamma'(u), \gamma'(u)\big)$ verifies

$$\int_{\gamma'} \rho(u) f^2(u)\, \mathrm{d}u \geq -\sum_{i=0}^{N} Q_i \left\| f^{(i)} \right\|^2, \tag{1.2}$$

for any test function $f \in \mathcal{C}_0^\infty(\mathbb{R})$, where $Q_0, \ldots, Q_N$ are positive constants and $\|\cdot\|$ denotes the usual L^2-norm. A condition of this kind can be reasonably expected to hold for a Ricci tensor that is linked to the energy-momentum tensor of a quantum theory via (1.1); that is because the integral in the left hand side of (1.2) closely resembles the form of an expected value in quantum field theory. Note also that (1.2) does not require $\rho(u)$ to be positive for all u, but it does impose limits on the occurrence of negative values.

Subsequently, the results of [30] were expanded and updated by Fewster and E.-A. Kontou in 2019 [31]. The main feature of this most recent development is the usage of a new criterion for the detection of focal points, which, as discussed at the beginning of this chapter, are a central element in the proofs of singularity theorems. All of the singularity theorems previously mentioned had their proofs based on the Raychaudhuri equation, which rules the evolution of the expansion θ of a geodesic congruence. Under the initial and energy conditions, θ can be shown to diverge within finite time, which, in this approach, is the indicator of the presence of a focal point. Meanwhile, the criterion which detects focal points in the new approach is a functional inequality of the form: there exists a function f such that

$$J[f] \leq K\,, \tag{1.3}$$

where J is a functional on f which takes into account the energy density $\rho(u)$ and K is a quantity related to the convergence of a certain spacelike submanifold, which is where the initial condition comes in. Despite requiring a certain amount of mathematical background, the new mathematical method employed is much simpler to work with. Its form also makes it very convenient to work with energy conditions of the form (1.2), since both are stated in terms of a "test" function f.

The text is organised as follows. In Chapter 2, we present in detail the variational approach to the characterisation of geodesics, conjugate points and focal points, following closely O'Neill's textbook on Semi-Riemannian Geometry [32]. Chapter 3 deals with the application of the methods discussed in the previous chapter to the deduction of singularity theorems of [31]. Appendix A collects some definitions and results that are frequently used throughout the text and that may not be covered in standard General Relativity texts, mostly concerning semi-Riemannian submanifolds and Jacobi fields. There is also a second Appendix aimed at completing the proofs of two theorems in Chapter 2.

We conclude this introductory chapter with some conventions and notations that are used throughout this text. The pair (M, g), where M is an m-dimensional differentiable manifold and g a metric tensor on M will denote a semi-Riemannian manifold. Spacetime is modelled as a Lorentz manifold, i.e., a semi-Riemannian manifold with metric signature $(-, +, \ldots, +)$ (or, equivalently, with index $\nu = 1$). Given ∇, the Levi-Civita connection of the metric g, the curvature tensor R is defined as

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

The Ricci tensor is given by the trace over the first and fourth entries of the curvature: if $x, y \in T_pM$ and $e_1, \ldots, e_m$ are an orthonormal basis for T_pM, with $\varepsilon_i = \langle e_i \, , \, e_i \rangle$,

$$\mathrm{Ric}(x, y) = \sum_{i=1}^{m} \varepsilon_i \langle R(e_i, x)y \, , \, e_i \rangle \, .$$

With these conventions, the Einstein equation reads, in natural units,

$$\mathrm{Ric} - \frac{S}{2} g = 8\pi T \, ,$$

where S, the scalar curvature, denotes the trace of the Ricci tensor, and T is the energy-momentum tensor. In Misner-Thorne-Wheeler's [33] sign conventions table, our conventions correspond to $[+ + +]$.

We employ Einstein's summing notation, meaning that repeated up- and down-indices are implicitly summed over; for example,

$$g_{ij} u^i v^j = \sum_i \sum_j g_{ij} u^i v^j \, .$$

Chapter 2

Variational methods for the characterisation of geodesics, conjugate and focal points

The aim of this chapter is to show how variational methods can be used to characterise geodesics and locate conjugate or focal points on a semi-Riemannian manifold. The exposition follows closely the developments in Chapter 10 of O'Neill's book on Semi-Riemannian Geometry [32], along with some preliminary concepts and results from other chapters. However, the approach used here is slightly different, both for the sake of brevity and to make the treatment of the timelike and lightlike cases as unified as possible. Notations and conventions also differ somewhat from that reference; this will emphasised whenever it may lead to confusion.

In Section 2.1, we construct the concept of variation of a curve, which is the central object of the variational theory. Section 2.2 introduces the functional E, to which our variational methods will be applied, and it is used to derive a characterisation of geodesics. In Section 2.3, the Hessian, a bilinear form related to the second variation of E, is defined, and then used in Sections 2.4 and 2.5 to obtain characterisations of conjugate and focal points. Finally, in Section 2.6, the inequalities which are used in [31] as criteria for the detection of focal points are extracted as a consequence of the main theorem in the previous section. We refer the reader to Appendix A for several definitions and results which find frequent use throughout this chapter, concerning Jacobi fields and the theory of semi-Riemannian submanifolds.

2.1 Variation of a curve

The basic ingredient of the variational method that will be introduced in this text is the concept of variation of a curve:

Definition 3. Let M be a differentiable manifold and $\gamma : [a, b] \to M$ a smooth, regular curve in M. A **variation** of γ is a smooth function

$$\Phi : [a, b] \times (-\varepsilon, \varepsilon) \longrightarrow M\,,$$

where $\varepsilon > 0$, such that $\Phi(u, 0) = \gamma(u)$ for all u in $[a, b]$.

For each fixed $v \in (-\varepsilon, \varepsilon)$, we will refer to the curve $u \mapsto \Phi(u, v)$ as a **longitudinal curve** of the variation Φ; in particular, the longitudinal curve with $v = 0$ is just γ. Likewise, given $u \in [a, b]$, we will call $v \mapsto \Phi(u, v)$ a **transverse curve** of Φ. The transverse curves with $u = a$ and $u = b$ are called the **first** (or **initial**) and **last** (or **final**) transverse curves of Φ, respectively. When the initial and final transverse curves of Φ are constant at $\gamma(a)$ and $\gamma(b)$, respectively, Φ is said to be a **fixed endpoint variation**.

We will also refer to u and v as the longitudinal and transverse parameters of the variation, respectively. It is useful to think of Φ as a one-parameter family of curves around γ. When interpreted in this manner, the curves in the family are the longitudinal ones, and they are indexed by the transverse parameter v.

The variation naturally associates to each pair (u_0, v_0) two vectors tangent to M at $\Phi(u_0, v_0)$: the velocities of the longitudinal and transverse curves of Φ through that point. We will denote them, respectively, by U and V. As usual, they act on a function $f \in \mathcal{C}^\infty(M)$ as

$$U(u_0, v_0)f = \left.\frac{\partial(f \circ \Phi)}{\partial u}\right|_{(u_0,v_0)} \quad \text{and} \quad V(u_0, v_0)f = \left.\frac{\partial(f \circ \Phi)}{\partial v}\right|_{(u_0,v_0)}. \tag{2.1}$$

Note that, by allowing the point (u_0, v_0) to vary in the expressions above, the resulting objects Uf and Vf are not smooth functions on M, but rather on $[a,b] \times (-\varepsilon, \varepsilon)$.

Clearly, U and V need not define vector fields on M (nor on an open subset of it). In the first place, the dimension of M may be greater than 2, in which case U and V could at best be *extended* to vector fields on M. Besides, no assumption is made on the injectiveness of the variation. Therefore, different points (u, v) may map to the same point in M, and assign different values for the velocity vectors U and V at that point. In fact, U and V are more aptly described as vector fields *on the mapping* Φ itself:

Definition 4. Let M and N be differentiable manifolds and $\varphi : N \to M$ a smooth map between them. A **vector field on** φ is a smooth mapping $X : N \to TM$ such that, for each $p \in N$, $X(p)$ is in $T_{\varphi(p)}M$. The set of vector fields on φ is denoted by $\mathfrak{X}(\varphi)$.

Since the dependence of $\nabla_X Y$ on X is only pointwise, it is possible to define the covariant derivative of a vector field on M with respect to a vector field on φ. Let $X \in \mathfrak{X}(\varphi)$, as in the definition above, and let $Y \in \mathfrak{X}(M)$. Given a point $p \in N$, let Ω be a neighbourhood of $\varphi(p)$ in M with an associated coordinate system. Then, the components of X and Y on the induced basis are functions

$$\begin{aligned} X^j : \varphi^{-1}(\Omega) &\longrightarrow \mathbb{R} \quad (1 \le j \le m) \\ Y^k : \quad \Omega &\longrightarrow \mathbb{R} \quad (1 \le k \le m)\,, \end{aligned}$$

and

$$(\nabla_X Y)(p) = \Big[X(p)Y^i + \Gamma^i_{jk}\big(\varphi(p)\big)X^j(p)Y^k\big(\varphi(p)\big)\Big]\partial_i(p)\,,$$

where Γ^i_{jk} are the Christoffel symbols of the connection of M in the given coordinates. Note that the result is in $\mathfrak{X}(\varphi)$, like X, rather than in $\mathfrak{X}(M)$, like Y.

If one tries to replace Y with a vector field on φ, one quickly runs into trouble with the fact that the first term in the expression above may try to take derivatives of Y^i along directions which are not tangent to the image of φ, and these are undefined. However, the covariant derivative of $Y \in \mathfrak{X}(\varphi)$ *can* be defined in the special case where X is the pushforward through φ of a vector field Z on N: $X = \varphi_* Z$. In this case, the components of Y are defined on $\varphi^{-1}(\Omega)$, like those of X in the previous example, and so we can act on them using Z instead of X. The definition then becomes

$$(\nabla_X Y)(p) = \left[Z(p) Y^i + \Gamma^i_{jk}\big(\varphi(p)\big) X^j(p) Y^k(p) \right] \partial_i(p) \,.$$

Put in different terms, when X is a pushforward, it only takes derivatives of Y along the image of φ, and the definition extends to $Y \in \mathfrak{X}(\varphi)$.

The situation of the last paragraph is true, in particular, for U and V: being the velocities of the longitudinal and transverse curves of Φ, they can be written as $\Phi_*\left(\frac{\partial}{\partial u}\right)$ and $\Phi_*\left(\frac{\partial}{\partial v}\right)$, respectively. For future reference, the derivatives of a vector field $Y \in \mathfrak{X}(\Phi)$ along U and V are

$$\nabla_U Y = \left[\frac{\partial Y^i}{\partial u} + \left(\Gamma^i_{jk} \circ \Phi\right) U^j Y^k \right] \partial_i \tag{2.2}$$

and

$$\nabla_V Y = \left[\frac{\partial Y^i}{\partial v} + \left(\Gamma^i_{jk} \circ \Phi\right) V^j Y^k \right] \partial_i \,. \tag{2.3}$$

If we use the equations above to calculate $\nabla_U V - \nabla_V U$, the symmetry of the Christoffel symbols makes it so that the second terms of either expression cancel out. If we denote the coordinate functions associated with the patch Ω by $x^1, \ldots, x^m$, the components of U and V are given by

$$U^i = \frac{\partial\left(x^i \circ \Phi\right)}{\partial u} \quad \text{and} \quad V^i = \frac{\partial\left(x^i \circ \Phi\right)}{\partial v} \,. \tag{2.4}$$

Therefore, we are left with

$$\nabla_U V - \nabla_V U = \left(\frac{\partial^2 (x^i \circ \Phi)}{\partial u\, \partial v} - \frac{\partial^2 (x^i \circ \Phi)}{\partial v\, \partial u} \right) \partial_i \,,$$

and, by commutation of partial derivatives, we conclude

$$\nabla_U V = \nabla_V U \,. \tag{2.5}$$

Equations (2.2) and (2.3) can be applied consecutively to obtain an expression for $\nabla_V \nabla_U Y$. Writing $\Gamma^i_{jk} \circ \Phi$ as $\hat{\Gamma}^i_{jk}$ for conciseness,

$$\begin{aligned}\nabla_V \nabla_U Y = \Bigg[& \frac{\partial^2 Y^i}{\partial v \partial u} + \frac{\partial \hat{\Gamma}^i_{jk}}{\partial v} U^j Y^k + \hat{\Gamma}^i_{jk} \frac{\partial U^j}{\partial v} Y^k + \\ & + \hat{\Gamma}^i_{jk} U^j \frac{\partial Y^k}{\partial v} + \hat{\Gamma}^i_{jk} V^j \frac{\partial Y^k}{\partial u} + \hat{\Gamma}^i_{jk} \hat{\Gamma}^k_{lm} V^j U^l Y^m \Bigg] \partial_i \,.\end{aligned}$$

Writing the corresponding expression for $\nabla_U \nabla_V Y$ and subtracting them, the first term cancels out by commutation of the partial derivatives, the fourth and fifth terms cancel as they appear on both expressions, and the third term cancels out by using (2.4). Finally, after making the replacements

$$\frac{\partial \hat{\Gamma}^i_{jk}}{\partial u} = U^l \frac{\partial \Gamma^i_{jk}}{\partial x^l} \circ \Phi \quad \text{and} \quad \frac{\partial \hat{\Gamma}^i_{jk}}{\partial v} = V^l \frac{\partial \Gamma^i_{jk}}{\partial x^l} \circ \Phi$$

(which are just applications of the chain rule) and shuffling around some indices, we find

$$\begin{aligned}\nabla_U \nabla_V Y - \nabla_V \nabla_U Y &= U^j V^k Y^l \left[\left(\frac{\partial \Gamma^i_{kl}}{\partial x^j} - \frac{\partial \Gamma^i_{jl}}{\partial x^k} + \Gamma^i_{jm} \Gamma^m_{kl} - \Gamma^i_{km} \Gamma^m_{jl} \right) \circ \Phi \right] \partial_i \\ &= \left(R_{jkl}{}^i \circ \Phi \right) U^j V^k Y^l \partial_i \,.\end{aligned}$$

Written in a component-free manner,

$$\nabla_U \nabla_V Y - \nabla_V \nabla_U Y = R(U, V) Y \,. \tag{2.6}$$

Equations (2.5) and (2.6) will play a fundamental role in the variational theory.

We end this section with a collection of definitions that will be used frequently in what follows. The first one is the **variation vector field** of a variation Φ. This is simply the vector field $W(u)$ along the base curve γ which gives the velocity of the transverse curve through $\gamma(u)$. More explicitly,

$$W(u) = V(u, 0) \qquad (\forall u \in [a, b]) \, .$$

Similarly, the mapping $u \mapsto (\nabla_V V)(0, t)$, giving the proper acceleration of the transverse curve through $\gamma(u)$, will be called the **acceleration vector field** of Φ and denoted $A(u)$.

Given a curve γ and a vector field W along γ, it is always possible to find a variation of γ whose variation vector field is W. Indeed, it is enough to define

$$\Phi(u, v) = \exp_{\gamma(u)} \left(v W(u) \right) ,$$

with $\varepsilon > 0$ chosen small enough that the exponential on the right hand side is well-defined for every $v \in (-\varepsilon, \varepsilon)$ at every point $\gamma(u)$. As we shall see below, the variation vector field W and the acceleration vector field A encode basically all of the information necessary for the variational theory.

The definition below extends the notion of variation to the case of piecewise smooth curves:

Definition 5. Let $\gamma : [a, b] \to M$ be a piecewise smooth curve with breaks $u_1, \ldots, u_n$, where $a < u_1 < \ldots < u_n < b$. For convenience, we will also denote $a = u_0$ and $b = u_{n+1}$. A **piecewise smooth variation of** γ is a map

$$\Phi : [a, b] \times (-\varepsilon, \varepsilon) \longrightarrow M \, ,$$

where $\varepsilon > 0$, such that the restriction of Φ to $[u_i, u_{i+1}] \times (-\varepsilon, \varepsilon)$ is smooth, for each $0 \leq i \leq n$. As before, it is required that $\Phi(u, 0) = \gamma(u)$.

Note that the definition implies that all the transverse curves of Φ are smooth, whereas the longitudinal ones are piecewise smooth. We introduce a notation for the discontinuities of γ at the breaks. For each $1 \leq i \leq n$, define

$$\Delta\gamma'(u_i) = \gamma'(u_i^+) - \gamma'(u_i^-) \, .$$

It will also be useful to think of γ' as identically zero for u outside $[a, b]$, so that we can include the following expressions to the ones above:

$$\Delta\gamma'(u_0) = \gamma'(a) \quad \text{and} \quad \Delta\gamma'(u_{n+1}) = -\gamma'(b)\,.$$

We note that the variation vector field is continuous, and smooth everywhere except possibly at the breaks.

A variation is said to be **nontrivial** when its variation vector field is not identically zero. When all the longitudinal curves of a variation are geodesics, we call it a **variation of γ by geodesics** (note that this can only happen when γ itself is a geodesic). The following is an important property of variations by geodesics that will be frequently used in our treatment of conjugate and focal points:

Proposition 6. *The variation vector field of a variation of γ by geodesics is a Jacobi field*[1].

Proof. By the definition of the curvature tensor,

$$R(W,\gamma')\gamma' = [R(V,U)U]|_{v=0} = (\nabla_V\nabla_U U - \nabla_U\nabla_V U)|_{v=0}\,.$$

Because the longitudinal curves are geodesics, $\nabla_U U = 0$. On the second term, we can use $\nabla_V U = \nabla_U V$ to get

$$R(W,\gamma')\gamma' = (-\nabla_U\nabla_U V)|_{v=0} = -W''\,,$$

so that W verifies the Jacobi equation. □

2.2 Variational theory of the E functional

Given a curve γ and Φ a variation of γ, we can define the following functional:

$$E_\Phi(v) = \frac{1}{2}\int_a^b \langle U(u,v)\,,\, U(u,v)\rangle\,\mathrm{d}u\,.$$

[1] See Appendix A.

We will often omit the subscript Φ when the particular variation to which we are referring is understood. The quantity measured by $E(v)$ is closely related to the length of the longitudinal curve with parameter v, which would be given by

$$L_\Phi(v) = \int_a^b \sqrt{\sigma\langle U\,,\,U\rangle}\,\mathrm{d}u\,,$$

where σ is $+1$ for spacelike U and -1 for timelike U. Note that, while E cannot be given a geometric meaning like the length L —it is not even reparameterisation-invariant—, it has the advantage of being differentiable in regions where the character of the curves changes from timelike to spacelike, which L does not, because of the factor σ that is inserted to make the square root well-defined. This makes L ill-suited to work with variations of a lightlike curve, for example. The absence of the square root also makes E slightly less cumbersome to deal with. In spite of this, the variational theory of E gives the same information regarding focal and conjugate points as does L. References [31, 32] work with L for the timelike and spacelike cases, and then switch to E for the lightlike case. Here, for the sake of brevity and unification, we will use E throughout.

The following simple proposition gives an expression for the first and second derivatives of E with respect to v:

Proposition 7. *Let $\gamma : [a,b] \to M$ be a piecewise smooth curve on a semi-Riemannian manifold M and Φ a piecewise smooth variation of γ. Then,*

$$E'_\Phi(0) = -\sum_{i=0}^{n+1}\langle W(u_i)\,,\,\Delta\gamma'(u_i)\rangle - \int_a^b \langle W\,,\,\gamma''\rangle\,\mathrm{d}u \tag{2.7}$$

and

$$E''_\Phi(0) = -\sum_{i=0}^{n+1}\langle A(u_i)\,,\,\Delta\gamma'(u_i)\rangle + \\ + \int_a^b \left[\langle W'\,,\,W'\rangle + \langle R(W,\gamma')W\,,\,\gamma'\rangle - \langle A\,,\,\gamma''\rangle\right]\mathrm{d}u\,, \tag{2.8}$$

where W denotes the variation vector field of Φ. Equations (2.7) and (2.8) are known as the ***first*** *and* ***second variation formulas****, respectively.*

Proof. Let $h(u,v) = \frac{1}{2}\langle U\,,\,U\rangle$, so that $E_\Phi(v)$ is given by integration of h over $u \in [a,b]$. At first, we will work separately on each subinterval $[u_i, u_{i+1}]$, where Φ is smooth. For that purpose, we define

$$E_i(v) = \int_{u_i}^{u_{i+1}} h(u,v)\,\mathrm{d}u \qquad (0 \le i \le n)\,.$$

Taking derivatives with respect to v, we find

$$\frac{\partial h}{\partial v} = \langle \nabla_V U\,,\,U\rangle = \langle \nabla_U V\,,\,U\rangle$$

and

$$\begin{aligned}\frac{\partial^2 h}{\partial v^2} &= \langle \nabla_V \nabla_U V\,,\,U\rangle + \langle \nabla_U V\,,\,\nabla_V U\rangle \\ &= \langle \nabla_U \nabla_V V\,,\,U\rangle + \langle R(V,U)V\,,\,U\rangle + \langle \nabla_U V\,,\,\nabla_U V\rangle\,,\end{aligned}$$

where the following facts have been used:

- differentiation with respect to v is equivalent to application of the vector V;
- the compatibility of the Levi-Civita connection with the metric; and
- Equations (2.5) and (2.6).

When v is set to 0, U and V reduce to $\gamma'(u)$ and $W(u)$, respectively, whereas $\nabla_V V$ becomes $A(u)$. Therefore,

$$\left.\frac{\partial h}{\partial v}\right|_{v=0} = \langle W'\,,\,\gamma'\rangle = \langle W\,,\,\gamma'\rangle' - \langle W\,,\,\gamma''\rangle\,,$$

and

$$\begin{aligned}\left.\frac{\partial^2 h}{\partial v^2}\right|_{v=0} &= \langle A'\,,\,\gamma'\rangle + \langle R(W,\gamma')W\,,\,\gamma'\rangle + \langle W'\,,\,W'\rangle \\ &= \langle A\,,\,\gamma'\rangle' - \langle A\,,\,\gamma''\rangle + \langle R(W,\gamma')W\,,\,\gamma'\rangle + \langle W'\,,\,W'\rangle\,,\end{aligned}$$

where W' and A' denote covariant derivatives along γ (so that the steps that resemble applications of the Leibniz rule actually amount to the Levi-Civita property). Integrating the expressions above on $[u_i, u_{i+1}]$ gives $E_i'(0)$ and $E_i''(0)$:

$$E_i'(0) = \left\langle W(u_{i+1})\,,\, \gamma'\left(u_{i+1}^-\right)\right\rangle - \left\langle W(u_i)\,,\, \gamma'\left(u_i^+\right)\right\rangle - \int_{u_i}^{u_{i+1}} \langle W\,,\, \gamma''\rangle \, \mathrm{d}u$$

$$\begin{aligned} E_i''(0) = {} & \left\langle A(u_{i+1})\,,\, \gamma'(u_{i+1}^-)\right\rangle - \left\langle A(u_i)\,,\, \gamma'(u_i^+)\right\rangle + \\ & + \int_{u_i}^{u_{i+1}} [\langle W'\,,\, W'\rangle + \langle R(W,\gamma')W\,,\, \gamma'\rangle - \langle A\,,\, \gamma''\rangle] \, \mathrm{d}u\,. \end{aligned}$$

Finally, summing over all the subintervals gives the intended formulas for $E_\Phi'(0)$ and $E_\Phi''(0)$. $\square$

It is a well-known fact that, on sufficiently small neighbourhoods, geodesics are length-minimising curves. Therefore, it comes as no surprise that the variation of the L functional can be applied the characterisation of geodesics. The following theorem shows that the variation of E can be used to obtain the same information.

Theorem 8. *Let $\gamma : [a, b] \to M$ be a piecewise smooth curve on a semi-Riemannian manifold M. Then, γ is an unbroken geodesic if, and only if, $E_\Phi'(0) = 0$ for every fixed endpoint variation Φ of γ.*

Proof. From the first variation formula proved in Proposition 7, it is evident that, if γ is an unbroken geodesic, then $E_\Phi'(0) = 0$ for every variation Φ, since $\gamma'' = 0$, all the breaks at $u_1, \ldots, u_n$ are trivial and $W(a) = W(b) = 0$.

For the converse, let γ have breaks at $u_1, \ldots, u_n$. We begin by showing that γ must be a geodesic segment on each subinterval $[u_i, u_{i+1}]$. Pick any point $t \in (u_i, u_{i+1})$ and any nonzero vector w in $T_{\gamma(t)}M$. Extend w by parallel translation to a vector field W_0 along γ defined on some interval $I \subset [u_i, u_{i+1}]$. Take a smooth function $f : [a, b] \to \mathbb{R}$ whose support is contained in I and define the vector field along γ

$$W(u) = \begin{cases} f(u)W_0(u) & \text{for } u \in I \\ 0 & \text{otherwise} \end{cases} .$$

Now, using the procedure shown in Equation (2.7), one can obtain a variation Φ of γ which has W as its variation vector field. Then, by the first variation formula,

$$E'_\Phi(0) = -\int_I f\langle\gamma'' \,,\, W_0\rangle \,\mathrm{d}u \,.$$

But, by hypothesis, $E'_\Phi(0) = 0$. Since t and w are arbitrary, we conclude that $\gamma'' = 0$ on (u_i, u_{i+1}), i.e., γ is a geodesic on that subinterval.

To show that γ is unbroken, we employ the same strategy to construct, for each break point u_i, a variation Φ whose variation vector field W is trivial except on a small neighbourhood of u_i. Then, since we already know γ is a geodesic on both subintervals surrounding u_i, the equation $E'_\Phi(0) = 0$ reduces to

$$\langle W(u_i) \,,\, \Delta\gamma'(u_i)\rangle = 0 \,,$$

and the freedom in the choice of W, along with the nondegeneracy of the metric, implies $\Delta\gamma'(u_i) = 0$. □

If one is searching for curves which locally maximise or minimise the E functional, then, in light of the theorem above, the second variation formula acquires particular interest in the case where γ is an unbroken geodesic. That is because, as we now know, such curves are precisely the critical points of E, and thus the question of whether they correspond to local maxima, minima or neither passes through the second derivative test. With that in mind, we now specialise the second variation formula to the case of an unbroken geodesic segment:

Corollary 9. *Let $\gamma : [a, b] \to M$ be a geodesic segment on a semi-Riemannian manifold M and Φ a variation of γ. Then,*

$$E''_\Phi(0) = \langle A\,,\, \gamma' \rangle|_a^b + \langle W'\,,\, W \rangle|_a^b - \int_a^b \langle W'' + R(W,\gamma')\gamma'\,,\, W \rangle \, \mathrm{d}u\,.$$

If, in addition, Φ has a fixed endpoint at a or b, the corresponding boundary terms vanish.

Proof. This follows immediately from (2.8) by setting γ'' and the jumps of γ' to zero and by a rearrangement of the other two terms in the integral. □

Notice that the first factor in the inner product in the integrand is the quantity that is set to zero in the Jacobi equation. Thus, much in the same way as being a geodesic relates to γ being a critical point of E, the existence of conjugate points of $\gamma(a)$ in $(a, b]$ will give information on whether or not γ extremises E. How this comes about will be seen in detail in the following sections.

2.3 The Hessian

Given a manifold M and points p and q in M, we denote by $\Omega(p, q)$ the set of all piecewise smooth curves $\gamma : [a, b] \to M$ such that $\gamma(a) = p$ and $\gamma(b) = q$. $\Omega(p, q)$ has the structure of an ∞-dimensional Frèchet manifold. The precise definitions shall not concern us here, but this observation will motivate some of the terminology to be used from this point onward.

Take γ in $\Omega(p, q)$ and Φ a fixed endpoint variation of γ. Then, each longitudinal curve of Φ is a "point" in $\Omega(p, q)$; therefore, seen as a function of the transverse parameter v, the variation defines a *curve* in $\Omega(p, q)$. The variation vector field of Φ is the *velocity* of that curve at $v = 0$, and thus is to be interpreted as a tangent vector of $\Omega(p, q)$ at γ. We denote by $T_\gamma\Omega$ (omitting the specification of the endpoints p and q for simplicity) the set of all piecewise smooth vector fields X along γ such that $X(a) = X(b) = 0$.

The functional E assigns a real number to each curve connecting p and q; it is therefore a function on $\Omega(p, q)$. Hence, the quantity $E'_\Phi(0)$ associated with a variation Φ that we previously calculated is the application of the variation vector field of Φ to the function E, $W[E]$.

Now let γ be a *geodesic* in M, and consider the second variation formula (2.8) for fixed endpoint variations of γ. Notice that it does not depend on any particular feature of Φ other than the variation vector field W. This leads us to define

$$H_\gamma(X,Y) = \int_a^b [\langle X' \,, Y' \rangle + \langle R(X,\gamma')Y \,, \gamma' \rangle]\, \mathrm{d}u \,,$$

for all X and Y in $T_\gamma\Omega$. Then, by properties of the metric and the curvature, H_γ is *symmetric* and *bilinear* in X and Y (the homogeneity is with respect to real numbers, not functions on $[a,b]$). The second variation formula states that

$$E''_\Phi(0) = H_\gamma(W,W) \,,$$

for any variation Φ which has W as its variation vector field. Because of these properties, we shall refer to the bilinear form H_γ as the **Hessian** of E at γ.

Occasionally, it will prove useful to work with a modified version of the Hessian, which we denote $H^\perp_\gamma(X,Y)$. All this does is restrict H_γ to $T^\perp_\gamma\Omega$, i.e., the subspace composed of all those $X \in T_\gamma\Omega$ which are everywhere orthogonal to γ'. Note that, since γ may be lightlike, the tangential direction could itself be orthogonal. On the other hand, when γ is not lightlike, any vector field along γ has a unique decomposition into tangential and orthogonal components.

2.4 Conjugate points

Definition 10. Let $\gamma : [a,b] \longrightarrow M$ be a geodesic and c a number in $(a,b]$. Then, $\gamma(c)$ is said to be **conjugate** to $\gamma(a)$ if there exists a nonzero Jacobi field[2] Y on γ such that $Y(a) = Y(c) = 0$.

The following theorem gives two alternative characterisations of conjugate points, one in terms of properties of the exponential map at $\gamma(a)$ and one in terms of variations of γ:

[2]See Appendix A.

Theorem 11. *Let $\gamma : [a,b] \to M$ be a geodesic and $p = \gamma(a)$. Then, the following assertions are equivalent:*

i. *$\gamma(b)$ is conjugate to p.*

ii. *The exponential map $\exp_p : T_pM \to M$ is singular at $(b-a)\gamma'(a)$.*

iii. *There exists a nontrivial variation of γ by geodesics starting at p such that $W(b) = 0$.*

Proof. Before proceeding to the proof of the implications between *i*, *ii* and *iii*, we will construct a class of variations which will play an important role in the following steps.

Let $x \in T_pM$ and assume $\exp_p[(b-a)x]$ is well-defined. If not, x can be rescaled so that this condition holds true. Let y_x be a tangent vector to T_pM at x. Being a vector space, T_pM is canonically isomorphic to each of its tangent spaces. Denote by y the vector in T_pM which corresponds to y_x. Then, define

$$\Phi(u,v) = \exp_p[(u-a)(x+vy)] \, ,$$

for each $u \in [a,b]$ and each v in an interval $(-\varepsilon,\varepsilon)$ chosen small enough for definiteness. By elementary properties of the exponential map, Φ is a variation of the geodesic γ_x which departs from p at $u = a$ with velocity x. Likewise, its longitudinal curves are the geodesics γ_{x+vy}, so that Φ is a variation by geodesics with a fixed starting point p.

We will show that W, the variation vector field of Φ, has the following properties:

$$W(a) = 0 \, , \quad W'(a) = y \, ; \tag{2.9}$$

$$W(b) = \mathrm{d}\exp_p\left[(b-a)y_{(b-a)x}\right] , \tag{2.10}$$

where the subscript $(b-a)x$ labels the vector in $T_{(b-a)x}(T_pM)$ which canonically corresponds to y and y_x.

At $u = a$, we have

$$U(a, v) = x + vy \quad \text{and} \quad V(a, v) = 0\,;$$

these come from the facts that the initial velocities of the longitudinal curves are $x + vy$ and that the first transverse curve of Φ is constant at p. In particular, the variation vector field vanishes at a, which proves the first part of Equation (2.9). Using the formulas above and Equations (2.5) and (2.3), we calculate (using any local coordinate system),

$$(\nabla_U V)(a, v) = (\nabla_V U)(a, v) = \frac{\partial}{\partial v}(x + vy) = y\,.$$

This is true for any $v \in (-\varepsilon, \varepsilon)$; in particular, setting $v = 0$ gives $W'(a) = y$.

As for (2.10), we note that the exponential in the definition of Φ carries the curve

$$\beta(v) = (b - a)(x + vy)$$

in T_pM to the final transverse curve of the variation. Therefore, by the definition of the differential, $\mathrm{d}\exp_p$ maps the velocity of β at $v = 0$ (which is $(b - a)y_{(b-a)x}$) to the velocity of $v \mapsto \Phi(b, v)$ at $v = 0$ (which is $W(b)$).

Now we are in a position to prove all the necessary logical relations between (i), (ii) and (iii):

($i \Rightarrow ii$) We will construct a variation Φ of γ as above with appropriate choices of x and y. First, we let $x = \gamma'(a)$. The condition that $\exp_p[(b - a)x]$ be well-defined is then automatically satisfied, since the geodesic γ extends over $[a, b]$ by hypothesis.

From the definition of conjugate points, (i) implies the existence of a Jacobi field Y on γ which vanishes at a and b. Because Y is nonzero and $Y(a) = 0$, we must have $Y'(a) \neq 0$. We let $y = Y'(a)$.

The variation of γ thus constructed is a variation by geodesics, and hence its variation vector field is a Jacobi field. Furthermore, from (2.9), we see that it satisfies the same initial conditions as Y. The two fields must therefore coincide, and we have $W(b) = Y(b) = 0$. Equation (2.10) then shows that the exponential map is singular at $(b - a)\gamma'(a)$.

(*ii* $\Rightarrow$ *iii*) By hypothesis, there exists a nonzero vector z tangent to T_pM at $(b - a)\gamma'(a)$ such that $\mathrm{d}\exp_p(z) = 0$. Let y be the vector in T_pM canonically corresponding to z, and construct the variation Φ as before with this y and with $x = \gamma'(a)$. Then, the vector $(b - a)y_{(b-a)x}$ appearing in formula (2.10) is a multiple of z, and the linearity of the differential map allows us to conclude that $W(b) = 0$. Therefore, the variation Φ thus constructed has all the properties required in assertion (iii).

(*iii* $\Rightarrow$ *i*) The variation vector field W of such a variation already vanishes at a and b; since it is a variation by geodesics, W is also a Jacobi field. Thus, $\gamma(b)$ is conjugate to p. □

The presence of the term γ'' inside the integral in the First Variation Formula (2.7) hints that all unbroken geodesics are critical points of the E functional, and Theorem 8 shows that a converse holds in a certain sense. Likewise, the term $W'' + R(W, \gamma')\gamma'$ in the Second Variation Formula (in the form of Corollary 9), suggests that Jacobi fields annihilate the Hessian. The next result, much like Theorem 8—and proven using similar techniques—, will provide a converse to this observation.

Before we state the Theorem, we introduce some notation and terminology. Firstly, the Hessian H_γ is not a linear transform, but a bilinear form, and thus by its **kernel** we mean the set of all $X \in T_\gamma\Omega$ such that $H_\gamma(X, Y) = 0$ for all $Y \in T_\gamma\Omega$. If $\gamma : [a, b] \to M$ is a geodesic connecting p to q, we denote the set of all Jacobi fields on γ which vanish at a and b by $\mathfrak{J}(a, b)$.

Theorem 12. *For a geodesic $\gamma \in \Omega(p, q)$ parametrised on $[a, b]$, the kernel of H_γ is precisely $\mathfrak{J}(a, b)$.*

Proof. That $\mathfrak{J}(a, b) \subseteq \ker H_\gamma$ is immediate from Corollary 9. Let $X \in \ker H_\gamma$, with breaks $a = u_0 < u_1 < \ldots < u_n < u_{n+1} = b$. We will show that X satisfies the Jacobi equation on each subinterval of the form (u_i, u_{i+1}). Let $t \in (u_i, u_{i+1})$ and $y \in T_{\gamma(t)}M$. Extend y by parallel translation to an interval $I \subset (u_i, u_{i+1})$, and call this extension Y_0. Take a smooth function f defined on $[a, b]$ with support contained in I. Then, for the vector field

$$Y(u) = \begin{cases} f(u)Y_0(u) & \text{if } u \in I \\ 0 & \text{otherwise} \end{cases}$$

along γ, the Second Variation Formula gives

$$0 = H_\gamma(X, Y) = f \int_I \langle X'' + R(X, \gamma')\gamma' \, , \, Y_0 \rangle \, \mathrm{d}u \, .$$

Because the choices of t and y are arbitrary, this means that X must satisfy the Jacobi equation on (u_i, u_{i+1}).

Repeating this reasoning around the break points $u_1, \ldots, u_n$ shows that $\Delta X(u_i) = 0$ at each of them. Since $X \in T_\gamma\Omega$, it automatically vanishes at a and b. With this, we conclude that $X \in \mathfrak{J}(a, b)$, and so also $\ker H_\gamma \subseteq \mathfrak{J}(a, b)$. □

Notice that, starting with $X \in \ker H_\gamma^\perp$ and y orthogonal to $\gamma'(t)$, the proof of the theorem works without further modifications to show that $\ker H_\gamma^\perp$ is contained within the set of orthogonal Jacobi fields in $\mathfrak{J}(a, b)$. The converse is also immediate from Corollary 9, so that these sets actually coincide.

The next theorem details in which cases it is possible for the bilinear form H_γ to be semidefinite, given the causal character of γ and the semi-Riemannian index of M. In order to prove that result, we will need the following proposition, which shows that, under certain assumptions, we can control the sign of $H_\gamma(W, W)$ by appropriately choosing the vector field W.

Proposition 13. *Let $\gamma \in \Omega(p, q)$ be a geodesic parameterised on $[a, b]$. Suppose a vector $y \in T_pM$ can be chosen which is linearly independent from $\gamma'(a)$ and either unit timelike or unit spacelike, i.e., $\langle y \, , \, y \rangle = \varepsilon$, with $\varepsilon = \pm 1$. Then, we can construct a vector field $X \in T_\gamma\Omega$ such that $H_\gamma(X, X)$ has the same sign as ε.*

Proof. Take the parallel transport of y along γ, and call it Y. Let

$$X = \frac{1}{\omega} \sin\big(\omega(u - a)\big) Y \, ,$$

where $\omega = n\pi/(b-a)$ for some $n \in \mathbb{N}$, so that $X(a) = X(b) = 0$. Then,

$$\langle X' , X' \rangle = \cos^2\big(\omega(u-a)\big)\langle Y , Y \rangle = \varepsilon \cos^2\big(\omega(u-a)\big) .$$

For the curvature term,

$$\langle R(X,\gamma')X , \gamma' \rangle = \frac{1}{\omega^2}\sin^2\big(\omega(u-a)\big)\langle R(Y,\gamma')Y , \gamma' \rangle .$$

Then,

$$H_\gamma(X,X) = \varepsilon \int_a^b \cos^2\big(\omega(u-a)\big)\,\mathrm{d}u + \frac{1}{\omega^2}\int_a^b \langle R(Y,\gamma')Y , \gamma' \rangle \sin^2\big(\omega(u-a)\big)\,\mathrm{d}u .$$

By our choice of ω,

$$\int_a^b \cos^2\big(\omega(u-a)\big)\,\mathrm{d}u = \int_a^b \sin^2\big(\omega(u-a)\big)\,\mathrm{d}u = \frac{b-a}{2} .$$

Because $\langle R(Y,\gamma')Y , \gamma' \rangle$ is a continuous function of u, there exists $L > 0$ such that

$$|\langle R(Y,\gamma')Y , \gamma' \rangle| < \frac{2L}{b-a}$$

for all $u \in [a,b]$, so that

$$\left| \frac{1}{\omega^2}\int_a^b \langle R(Y,\gamma')Y , \gamma' \rangle \sin^2\big(\omega(u-a)\big)\,\mathrm{d}u \right| < \frac{L}{\omega^2} .$$

Therefore, choosing ω large enough makes it so that $H_\gamma(X,X)$ has the same sign as that of the first term, which is ε. □

Theorem 14. *Let $\gamma : [a,b] \to M$ be a geodesic on a semi-Riemannian manifold M of dimension m and index ν.*

i. If γ is lightlike, H_γ is neither positive nor negative semidefinite.

ii. If γ is spacelike, then:

   a. *If H_γ is positive semidefinite, $\nu = 0$.*

   b. *If H_γ is negative semidefinite, $\nu = m - 1$.*

iii. *If γ is timelike, then:*

   a. *If H_γ is negative semidefinite, $\nu = m$.*

   b. *If H_γ is positive semidefinite, $\nu = 1$.*

Proof. The previous proposition makes this result trivial to prove:

i. If γ is lightlike, there exist both a unit timelike and a unit spacelike vectors in $T_{\gamma(a)}M$ independent from $\gamma'(a)$. Therefore, H_γ takes on both positive and negative values.

We will prove items (ii) and (iii) simultaneously, by noting that they can be written in a unified way. Indeed, let σ denote the sign of $\langle \gamma' , \gamma' \rangle$. Then,

a'. If σH_γ is positive semidefinite, then either $\nu = 0$ or $\nu = m$.

b'. If σH_γ is negative semidefinite, then either $\sigma = +1$ and $\nu = m - 1$ or $\sigma = -1$ and $\nu = 1$.

Since there are no timelike geodesics on Riemannian manifolds nor spacelike geodesics on anti-Riemannian manifolds, item a' above correctly summarises (ii.a) and (iii.a) in the original phrasing of the theorem. Let us now prove statements (a') and (b'):

a'. Suppose $1 \leq \nu \leq m-1$. Then we can choose a unit vector $y \in T_{\gamma(a)}M$ orthogonal to $\gamma'(a)$ and with the opposite causal character. Thus, the construction of Proposition 13 with $\varepsilon = -\sigma$ gives $\sigma H_\gamma(X, X) < 0$, a contradiction.

b'. If the conclusions do not hold, we can choose $y \in T_{\gamma(a)}M$ orthogonal to and with the same causal character as $\gamma'(a)$. In this case, the procedure above gives $\sigma H_\gamma(X, X) > 0$, which contradicts the hypothesis of negative semidefiniteness. □

Theorem 14 tells us that there are basically only two interesting cases when studying the definiteness of the Hessian (and hence the possibility of extremising the functional E): geodesics on a Riemannian manifold and timelike geodesics on a Lorentzian manifold. The other two cases where the Hessian can be definite—geodesics on an anti-Riemannian manifold and spacelike geodesics on an anti-Lorentzian manifold—can be mapped onto the previous two cases by simply reversing the sign of the metric. This is the motivation behind the following definition:

Definition 15. A geodesic $\gamma : [a, b] \to M$ is said to be **cospacelike** if the subspace $\gamma'(u)^{\perp}$ of $T_{\gamma(u)}M$ is spacelike for some (and hence all) $u \in [a, b]$.

Theorem 18 below will establish a connection between the definiteness of the Hessian and the existence and location of conjugate points along a cospacelike geodesic. Before that, we prove two auxiliary results.

Lemma 16. *If X and Y are Jacobi fields along a geodesic γ, then $\langle X' , Y\rangle - \langle X , Y'\rangle$ is a constant.*

Proof. This is a simple consequence of the Levi-Civita condition and symmetries of the curvature:

$$\begin{aligned}\langle X' , Y\rangle' &= \langle X' , Y'\rangle + \langle X'' , Y\rangle = \langle X' , Y'\rangle - \langle R(X, \gamma')\gamma' , Y\rangle = \\ &= \langle X' , Y'\rangle - \langle R(Y, \gamma')\gamma' , X\rangle = \langle X' , Y'\rangle + \langle Y'' , X\rangle = \langle Y' , X\rangle'\end{aligned}$$ □

Lemma 17. *Let $Y_1, \ldots, Y_k$ be Jacobi fields on a geodesic γ such that*

$$\left\langle Y_i' , Y_j \right\rangle = \left\langle Y_i , Y_j' \right\rangle \qquad (\forall i, j)\,,$$

$f^1, \ldots, f^k$ smooth functions on γ and $X = f^i Y_i$. Then,

$$\langle X' , X'\rangle + \langle R(X, \gamma')X , \gamma'\rangle = \langle A , A\rangle + \langle X , B\rangle'\,, \tag{2.11}$$

where $A = \left(f^i\right)' Y_i$ and $B = f^i Y_i'$.

Proof. From the Levi-Civita property,

$$\langle X , B\rangle' = \langle X' , B\rangle + \langle X , B'\rangle\,.$$

Inserting $X' = A + B$ and $B' = (f^i)'Y_i' + f^iY_i''$,

$$\langle X\,,\,B\rangle' = \langle A\,,\,B\rangle + \langle B\,,\,B\rangle + \left\langle X\,,\,(f^i)'Y_i'\right\rangle + \left\langle X\,,\,f^iY_i''\right\rangle.$$

Now, using the definition of X in the third term and the fact that the Y_i are Jacobi fields in the fourth,

$$\langle X\,,\,B\rangle' = \langle A\,,\,B\rangle + \langle B\,,\,B\rangle + (f^i)'f^j\left\langle Y_j\,,\,Y_i'\right\rangle - \left\langle X\,,\,f^iR(Y_i,\gamma')\gamma'\right\rangle.$$

Next we apply the remaining hypothesis on the Y_i and the expression for X in terms of the Y_i:

$$\langle X\,,\,B\rangle' = \langle A\,,\,B\rangle + \langle B\,,\,B\rangle + (f^i)'f^j\left\langle Y_j'\,,\,Y_i\right\rangle - \langle X\,,\,R(X,\gamma')\gamma'\rangle.$$

Finally, recognising the formulas of A and B in the third term and rearranging, we obtain

$$2\langle A\,,\,B\rangle + \langle B\,,\,B\rangle = \langle X\,,\,B\rangle' - \langle R(X,\gamma')X\,,\,\gamma'\rangle.$$

Now, using $X' = A + B$ and the formula above,

$$\begin{aligned}\langle X'\,,\,X'\rangle &= \langle A\,,\,A\rangle + 2\langle A\,,\,B\rangle + \langle B\,,\,B\rangle\\ &= \langle A\,,\,A\rangle + \langle X\,,\,B\rangle' - \langle R(X,\gamma')X\,,\,\gamma'\rangle,\end{aligned}$$

which leads to the intended formula (2.11). □

Theorem 18. *Let $\gamma \in \Omega(p,q)$ be a cospacelike geodesic with sign σ.*

1. *If there are no points along γ conjugate to $p = \gamma(a)$, then $H_\gamma^\perp$ is positive definite.*

2. *If $q = \gamma(b)$ is the only point along γ which is conjugate to p, then $H_\gamma^\perp$ is semidefinite, but not definite.*

3. *If there is a point $\gamma(r)$ conjugate to p with $a < r < b$, then $H_\gamma^\perp$ is not semidefinite.*

Proof.

1. Let $v_1, \dots, v_{m-1}$ be a basis for $\gamma'(a)^\perp$. Then, for each i in $\{1, \dots, m-1\}$, let Y_i be the unique Jacobi field on γ with initial conditions $Y_i(a) = 0$, $Y_i'(a) = v_i$. Because there are no conjugate points of p on $(a, b]$, $\{Y_i(u)\}$ is a basis for $\gamma'(u)^\perp$ for every u in that interval (if the Y_i were not linearly independent at u, one could find coefficients α^i such that $\alpha^i Y_i(u) = 0$, and then the linear combination $\alpha^i Y_i$ would be a Jacobi field vanishing at both a and u).

   By the observations above, any $X \in T_\gamma^\perp \Omega$ can be uniquely written as $f^i Y_i$ on $(a, b]$, where the f^i are continuous, piecewise smooth functions. Because all the Y_i vanish at a, Lemma 16 implies that $\left\langle Y_i', Y_j \right\rangle - \left\langle Y_i, Y_j' \right\rangle = 0$ on $[a, b]$. Then, Lemma 17 applies, and

   $$H_\gamma^\perp(X, X) = \int_a^b \langle A, A \rangle \, \mathrm{d}u + \langle X, B \rangle |_a^b .$$

   The boundary terms vanish, because $X(a) = X(b) = 0$. Since γ is cospacelike and each Y_i is perpendicular to γ, $\langle A, A \rangle \geq 0$, and hence $H_\gamma^\perp(X, X) \geq 0$. Besides, if $H_\gamma^\perp(X, X) = 0$, then $\langle A, A \rangle$ must vanish identically, implying each f^i is constant. But, because X vanishes at q, the f^i are all zero at b. Then, $X = 0$.

2. By hypothesis, there exists a nontrivial Jacobi field on γ which vanishes at a and b. By Proposition 44 in Appendix A, the perpendicular component of this field (which can be taken, since γ is cospacelike) is also a Jacobi field vanishing at a and b. Then, Theorem 12 proves that $H_\gamma^\perp$ has a nontrivial kernel, and hence cannot be definite.

   We now proceed to prove that $H_\gamma^\perp$ is positive semidefinite, i.e., for any $X \in T_\gamma^\perp \Omega$, $H_\gamma^\perp(X, X) \geq 0$. First, observe that if there exists $r \in (a, b)$ such that X vanishes identically on $[r, b]$, then item 1 of this theorems applies with γ replaced with $\gamma|_{[a,r]}$ to show that $H_\gamma^\perp(X, X) > 0$, and so there is nothing to prove. We therefore assume that such a point r does not exist.

We now construct a sequence X_n in $T_\gamma^\perp\Omega$ which approaches X as $n \to +\infty$. Let $r_n = b - 1/n$, for each $n \in \mathbb{N}$ large enough that r_n lies in the last smooth segment of X. Let Y_n denote the parallel transport of $X(r_n)$ along γ. Then, define

$$X_n(u) = \begin{cases} X(u)\,, & \text{for } u \in [a, r_n] \\ \left(1 - 2\dfrac{u - r_n}{b - r_n}\right) Y_n(u)\,, & \text{for } u \in \left(r_n, \frac{b+r_n}{2}\right) \\ 0\,, & \text{for } u \in \left[\frac{b+r_n}{2}, b\right] \end{cases}\,.$$

Because X_n is nontrivial on $[a, r_n]$ and vanishes identically on $\left[\frac{b+r_n}{2}, b\right]$, item 1 implies that $H_\gamma^\perp(X_n, X_n) > 0$. Then, the fact that $H_\gamma^\perp(X_n, X_n) \to H_\gamma^\perp(X, X)$ shows that $H_\gamma^\perp(X, X) \geq 0$.[3]

3. By Proposition 13, using the fact that γ is cospacelike, we can always obtain $X \in T_\gamma^\perp\Omega$ such that $H_\gamma^\perp(X, X) > 0$. So it remains to show that the opposite inequality also occurs.

   By hypothesis, there exists a nontrivial Jacobi field Y on γ which vanishes at a and r. Let $X = Y$ on $[a, r)$ and $X = 0$ on $[r, b]$. Since Y is not identically zero, $X'(r^-) \neq 0$. Let $Z \in T_\gamma^\perp\Omega$ with $Z(r) = -X'(r^-)$ and $\delta > 0$.

$$H_\gamma^\perp(X + \delta Z, X + \delta Z) = H_\gamma^\perp(X, X) + 2\delta H_\gamma^\perp(X, Z) + \delta^2 H_\gamma^\perp(Z, Z)\,.$$

   The first term vanishes, since X is a piecewise Jacobi field. Meanwhile, because X vanishes identically on $[r, b]$, the second one reduces to

$$2\delta H_\gamma^\perp(X, Z) = 2\delta \int_a^r [\langle X'\,,\, Z'\rangle + \langle R(X, \gamma')Z\,,\, \gamma'\rangle]\, \mathrm{d}u\,.$$

[3] While the statement that $H_\gamma^\perp(X_n, X_n) \to H_\gamma^\perp(X, X)$ may seem harmless enough at first sight, note that there is an integral involved, and therefore the pointwise convergence of X_n to X does not immediately imply the convergence of $H_\gamma^\perp(X_n, X_n)$. A complete proof of this point can be found in Appendix B.

Using $\langle X', Z'\rangle = \langle X', Z\rangle' - \langle X'', Z\rangle$,

$$\begin{aligned} 2\delta H_\gamma^\perp(X,Z) &= 2\delta \int_a^r \left[\langle X', Z\rangle' - \langle X'' + R(X,\gamma')\gamma', Z\rangle\right] \mathrm{d}u \\ &= 2\delta\langle X'(r^-), Z(r)\rangle = -2\delta\langle X'(r^-), X'(r^-)\rangle < 0, \end{aligned}$$

the inequality stemming from the fact that γ is cospacelike. Then, for sufficiently small δ, $H_\gamma^\perp(X+\delta Z, X+\delta Z) < 0$. □

2.5 Focal points

In the previous section, by studying the variational theory of the functional E on the set of curves $\Omega(p,q)$, we were able to derive a characterisation of geodesics joining p and q, as well as establish a connection between the properties of the Hessian of E and the location of conjugate points along these geodesics. Now, by replacing the endpoint p with a higher-dimensional submanifold P of M, we will show how the same type of technique can be used to study *focal points* of P.

Definition 19. Let M be an m-dimensional semi-Riemannian manifold, q a point in M and P a p-dimensional semi-Riemannian submanifold of M. The set of all piecewise smooth curves $\gamma : [a,b] \to M$ such that $\gamma(a) \in P$ and $\gamma(b) = q$ shall be denoted by $\Omega(P,q)$.

The notion of fixed endpoint variation then extends as follows: a variation Φ of a curve $\gamma \in \Omega(P,q)$ is said to be a (P,q)-**variation** if $\Phi(a,v) \in P$ and $\Phi(b,v) = q$ for all $v \in (-\varepsilon,\varepsilon)$. As before, these variations can be regarded as curves in $\Omega(P,q)$.

The tangent space of $\Omega(P,q)$ at γ consists of the piecewise smooth vector fields X along γ which verify $X(a) \in T_{\gamma(a)}P$ and $X(b) = 0$. It is denoted $T_\gamma\Omega$, once again omitting P and q for short.

In this more general setting, when we attempt to extremise the E functional on curves in $\Omega(P,q)$, one extra condition appears in comparison with the fixed endpoint case. Namely, the critical points of E are not just geodesics from P to

q, but specifically those which are *orthogonal* to P. The proposition below more formally expresses this generalisation of Theorem 8:

Proposition 20. *A curve $\gamma \in \Omega(P, q)$ is an unbroken geodesic orthogonal to P if, and only if, $E'_\Phi(0) = 0$ for every (P, q)-variation Φ of γ.*

Proof. The direct implication is immediate from the First Variation Formula, since the boundary term at $u_0 = a$ becomes

$$-\langle W(a)\,,\,\gamma'(a)\rangle\,,$$

which vanishes, because, for a (P, q)-variation, $W(a)$ is tangent to P, whereas $\gamma'(a)$ is by hypothesis orthogonal to it.

For the opposite direction, the proof proceeds exactly like that of Theorem 8 to show that γ must be a geodesic on each subinterval (u_i, u_{i+1}) and that the breaks $\Delta\gamma'(u_1), \dots, \Delta\gamma'(u_{n+1})$ are all trivial. Special attention is only required at $u_0 = a$. Here, since $W(a)$ is constrained by hypothesis to be tangent to P, we cannot conclude that $\gamma'(a) = 0$, but only that $\gamma'(a)$ is orthogonal to P. □

From now on, we will assume $\gamma \in \Omega(P, q)$ is a geodesic orthogonal to P, and look for the appropriate generalisation for the Hessian of E. As before, this should be a bilinear form H_γ defined on $T_\gamma\Omega$ such that, whenever W is the variation vector field of a variation Φ, $H_\gamma(W, W)$ coincides with $E''_\Phi(0)$. By the Second Variation Formula, the only complication with respect to the fixed endpoint case is the boundary term at a, which no longer vanishes trivially. If we let $\alpha(v)$ denote the initial transverse curve $v \mapsto \Phi(a, v)$,

$$\langle A(a)\,,\,\gamma'(a)\rangle = \langle \alpha''(0)\,,\,\gamma'(a)\rangle\,.$$

Now, $\alpha''(0)$ can be canonically decomposed as the sum of a vector in T_pP and a vector orthogonal to P (see Appendix A); only the latter contributes to the inner product above, since $\gamma'(a)$ is also perpendicular to P. But the normal component of $\alpha''(0)$ can itself be rewritten as

$$\alpha''(0)^\perp = \mathrm{II}\left(\alpha'(0), \alpha'(0)\right) = \mathrm{II}\left(W(a), W(a)\right),$$

where II denotes the Second Fundamental Form of P. This motivates the following generalisation of the Hessian:

Definition 21. Let P be a semi-Riemannian submanifold of M, $\gamma : [a,b] \to M$ a P-normal geodesic in $\Omega(P,q)$ and $X, Y \in T_\gamma\Omega$. Then, we define the **Hessian** of E at γ as the bilinear form

$$H_\gamma(X,Y) = -\big\langle \gamma'(a)\,,\, \mathrm{II}\left(X(a), Y(a)\right)\big\rangle + \\ + \int_a^b \left[\langle X'\,,\, Y'\rangle + \langle R(X,\gamma')Y\,,\,\gamma'\rangle\right] \mathrm{d}u\,. \qquad (2.12)$$

Proposition 22. *Let P be a semi-Riemannian submanifold of M, γ a P-normal geodesic and Y a Jacobi field on γ. Then, Y is the variation vector field of a variation Φ of γ by P-normal geodesics if, and only if,*

1. *$Y(a)$ is tangent to P, and*
2. *$Y'(a)^\top = \widetilde{\mathrm{II}}\big(Y(a), \gamma'(a)\big)$.*

Proof. First, assume Y is the variation vector field of a variation Φ as above. $Y(a)$ is the initial velocity of the initial transverse curve of Φ, which is constrained to remain in P. Therefore, $Y(a)$ is tangent to P. Now, $U(a,v)$ is a vector field along the initial transverse curve of Φ which is everywhere normal to P, whereas $V(a,v)$ is everywhere tangent. Then,

$$\begin{aligned} Y'(a)^\top &= [\nabla_U V(a,0)]^\top \\ &= [\nabla_V U(a,0)]^\top \\ &= \widetilde{\mathrm{II}}\big(V(a,0), U(a,0)\big) \\ &= \widetilde{\mathrm{II}}\big(Y(a), \gamma'(a)\big)\,, \end{aligned}$$

proving that the second condition is satisfied.

For the converse, assume Y satisfies conditions in 1 and 2 in the statement. We will construct a variation Φ of γ with the desired properties and whose variation vector field W coincides with Y. Let $\sigma : (-\varepsilon, \varepsilon) \to P$ be a curve such that $\sigma'(0) = Y(a)$. We will first show that there exists a vector field Z along σ such that $Z(0) = \gamma'(a)$ and $Z'(0) = Y'(a)$. For that purpose, let A and B denote the *normal* parallel transports of $\gamma'(a)$ and $Y'(a)^\perp$ along σ, respectively. Then, define $Z(v) = A(v)+vB(v)$. It is clear that Z is normal to P and that $Z(0) = A(0) = \gamma'(a)$. Besides,

$$\begin{aligned} Z'(0) &= A'(0) + B(0) \\ &= \widetilde{\mathrm{II}}\big(\sigma'(0), \gamma'(a)\big) + Y'(a)^\perp \\ &= \widetilde{\mathrm{II}}\big(Y(a), \gamma'(a)\big) + Y'(a)^\perp \\ &= Y'(a)^\top + Y'(a)^\perp \\ &= Y'(a)\,, \end{aligned}$$

as intended.

Finally, we can construct the desired variation of γ using Z:

$$\Phi(u, v) = \exp_{\sigma(v)}[(u-a)Z(v)]\,.$$

Because of the use of the exponential map, it is immediate that this is a variation by geodesics. Furthermore, since the initial velocities of these geodesics are given by the vector field Z, they are all normal to P. It remains to show that W, the variation vector field of Φ, coincides with Y. In the first place, $W(a) = \sigma'(0) = Y(a)$. Secondly, because W is the variation vector field and Z the initial velocities of the longitudinal curves of Φ,

$$W'(a) = Z'(0) = Y'(a)\,.$$

By Theorem 43 on page 74, the two observations above imply $W \equiv Y$. □

Definition 23. Let $P \subset M$ be a semi-Riemannian submanifold and $\gamma : [a, b] \to M$ a geodesic perpendicular to P. A Jacobi field Y along γ is called a ***P*-Jacobi field** if $Y(a)$ is tangent to P and

$$Y'(a)^\top = \widetilde{\mathrm{II}}\big(Y(a), \gamma'(a)\big)\,,$$

i.e., if it verifies both conditions in Proposition 22. The set of P-Jacobi fields along

γ will be denoted $\mathfrak{J}_P$.

A point $\gamma(r)$ is called a **focal point** of P if there exists a P-Jacobi field along γ which vanishes at $\gamma(r)$. The set of all such vector fields is a subspace of $\mathfrak{J}_P$ whose dimension is called the **focal order** of $\gamma(r)$. This number is at most $m-1$, since:

1. $u \mapsto (u-a)\gamma'(u)$ is a P-Jacobi field along γ and does **not** vanish at $\gamma(r)$;

2. $\mathfrak{J}_P$ is isomorphic to $T_{\gamma(a)}M$, which is m-dimensional. (To see why, take $x \in T_{\gamma(a)}P$, $z \in T^{\perp}_{\gamma(a)}P$ and let X be the unique Jacobi field along γ with $X(a) = x$, $X'(a) = z + \widetilde{\mathrm{II}}\big(x, \gamma'(a)\big)$. By construction, X is a P-Jacobi field, and, since $T_{\gamma(a)}M$ decomposes as $T_{\gamma(a)}P \oplus T^{\perp}_{\gamma(a)}P$, the mapping $x \oplus z \mapsto X$ gives the required isomorphism.)

According to Proposition 44, page 76, a P-Jacobi field along γ vanishing at $\gamma(r)$ is always perpendicular to γ, since it is so at both r and the intersection between γ and P.

The geometrical interpretation of focal points is similar to that of conjugate points. If $\gamma(r)$ is a focal point of P along a normal geodesic, then there exists a family of geodesics around γ which are all normal to P and almost meet at $\gamma(r)$.

The theorem below gives a characterisation of focal points in a very similar manner as Theorem 11 does for conjugate points.

Theorem 24. *Let $\gamma : [a,b] \to M$ be a normal geodesic to P. Then, the following assertions are equivalent:*

i. $\gamma(b)$ is a focal point of P along γ.

ii. There is a nontrivial variation of γ by normal geodesics to P such that $W(b) = 0$.

iii. The normal exponential map of P, $\exp : NP \to M$, is singular at $(b-a)\gamma'(a)$.

Proof. Proposition 22 establishes the equivalence between (i) and (ii). To complete the proof, it will suffice to show that (ii) and (iii) are also equivalent.

(*ii* $\Rightarrow$ *iii*) Let Φ denote the variation of γ in question. Since its longitudinal curves are all normal geodesics to P, Φ can be written in terms of the normal exponential map of P:

$$\Phi(u, v) = \exp[(u - a)U(a, v)]\,.$$

Consider the curve $\varphi : (-\varepsilon, \varepsilon) \to NP$ given by

$$\varphi(v) = (b - a)U(a, v)\,.$$

From the two previous formulas, it is clear that φ is carried by the normal exponential map onto the last transverse curve of Φ, $v \mapsto \Phi(b, v)$. Likewise, it is mapped by the projection $\pi : NP \to P$ onto the first transverse curve of Φ, $v \mapsto \Phi(a, v)$. Therefore, by the definition of the differential map, $\mathrm{d}\exp$ and $\mathrm{d}\pi$ carry the initial velocity of φ to $W(b)$ and $W(a)$, respectively, since these are the initial velocities of the last and first transverse curves of Φ. But, by hypothesis, $W(b) = 0$. This means that, if the initial velocity of φ is nonzero, the exponential map is singular at its basepoint, $(b - a)\gamma'(a)$, as we intended to show. If, on the other hand, the initial velocity of φ is zero, then $W(a)$, its image through the linear map $\mathrm{d}\pi$, must also vanish. This implies that $\gamma(b)$ and $\gamma(a)$ are *conjugate points* along γ, and Theorem 11 guarantees that once again exp is singular at $(b - a)\gamma'(a)$.

(*iii* $\Rightarrow$ *ii*) The procedure is very similar to the one used in the previous item. Let x be a nonzero tangent vector to NP at $(b-a)\gamma'(a)$ such that $(\mathrm{d}\exp)(x) = 0$, which exists by the assumption that exp is singular. Let $Z : (-\varepsilon, \varepsilon) \to NP$ be a curve with initial velocity x (it is therefore a normal vector field to P with $Z(0) = (b - a)\gamma'(a)$). Define

$$\begin{aligned}\Phi : [a, b] \times (-\varepsilon, \varepsilon) &\longrightarrow M\\ (u, v) &\longmapsto \exp\Big(\frac{u - a}{b - a} Z(v)\Big)\,;\end{aligned}$$

then, Φ is a variation of γ by P-normal geodesics. Let $p = \gamma(a)$. We first assume that the curve $\pi\big(Z(v)\big)$ is not the constant curve at p. In that case,

Φ is a nontrivial variation. The last transverse curve of Φ, $v \mapsto \Phi(b, v)$, is the image of Z through exp; therefore, $\mathrm{d}\exp$ maps x onto $W(b)$. As a result, $W(b) = 0$.

Now, if $\pi \circ Z$ is constant at p, then x is tangent to $T_p^{\perp} P$ (and therefore also to $T_p M$). This means that not only the normal exponential map, but also $\exp_p$ is singular at $(b - a)\gamma'(a)$. By Theorem 11, $\gamma(b)$ is conjugate to p. We will show that it is also a focal point of P along γ. Since we already have equivalence between (i) and (ii), this will prove that (iii) implies (ii) also in this case. W, the variation vector field of Φ, is a Jacobi field along γ which vanishes at a and b; thus, $W(a)$ is trivially tangent to P. It remains to prove that $W'(a)^{\top} = \widetilde{\mathrm{II}}\big(W(a), \gamma'(a)\big)$. Because $W(a) = 0$ and $\widetilde{\mathrm{II}}$ is a bilinear form, this amounts to showing $W'(a)^{\top} = 0$. Using the longitudinal and transverse velocity fields of Φ, U and V, and noting that $U(a, v) = Z(v)$,

$$W'(a) = (\nabla_U V)(a, 0) = (\nabla_V U)(a, 0) = Z'(0)\,.$$

By construction, $Z'(0)$ is the vector in $T_p M$ canonically corresponding to x, which is normal to P; therefore, the tangential component of $W'(a)$ vanishes, as required. □

The following lemma characterises the kernel of $H_\gamma^{\perp}$, just as Theorem 12 does in the fixed endpoint case.

Lemma 25. *Let P be a semi-Riemannian submanifold of M and $\gamma \in \Omega(P, q)$ a geodesic issuing orthogonally from P. Then, the kernel of $H_\gamma^{\perp}$ is precisely the set of P-Jacobi fields along γ which vanish at q, which we denote $\mathfrak{J}(P, q)$.*

Proof. Starting from (2.12), we first rewrite the boundary term using

$$\Big\langle \gamma'(a)\,,\, \mathrm{II}\,\big(Y(a), X(a)\big) \Big\rangle = -\Big\langle \widetilde{\mathrm{II}}\big(Y(a), \gamma'(a)\big)\,,\, X(a) \Big\rangle\,.$$

Then, after integrating the term $\langle Y'\,,\, X' \rangle$ by parts and using a symmetry of the curvature, we get

$$H_\gamma(Y,X) = -\int_a^b \langle Y'' + R(Y,\gamma')\gamma' \,,\, X\rangle \,\mathrm{d}u - \\ - \left\langle \widetilde{\mathrm{II}}\big(Y(a),\gamma'(a)\big) - Y'(a)\,,\, X(a)\right\rangle .$$

Notice that, since $X(a)$ is tangent to P, we may replace $Y'(a)$ with $Y'(a)^\top$ in the inner product without altering the result. With that in mind, it is immediate from the formula above that, if Y is a P-Jacobi field, then $H_\gamma^\perp(Y,X) = 0$ for any $X \in T_\gamma^\perp\Omega$. Thus, $\mathfrak{J}(P,q) \subset \ker H_\gamma^\perp$. For the converse, we use the same technique as in Theorems 8 and 20: by exploiting the freedom in the choice of X, we first show that Y satisfies the Jacobi equation along γ, and then that

$$Y'(a)^\top - \widetilde{\mathrm{II}}\big(Y(a),\gamma'(a)\big) = 0\,;$$

together, these imply that Y is a P-Jacobi field. We get $\ker H_\gamma^\perp \subset \mathfrak{J}(P,q)$, which concludes the proof. □

In the next two subsections, we prove theorems which establish the connection between the ocurrence of focal points of P along γ and the positiveness of $H_\gamma^\perp$, first in the case of a cospacelike geodesic and then in the lightlike case.

2.5.1 The cospacelike case

Theorem 26. *Let P be a semi-Riemannian submanifold of M and $\gamma \in \Omega(P,q)$ a cospacelike P-normal geodesic. Then,*

1. *If γ contains no focal points of P, $H_\gamma^\perp$ is positive definite.*
2. *If q is the only focal point of P along γ, $H_\gamma^\perp$ is semidefinite, but not definite.*
3. *If there exists a focal point $\gamma(r)$ of P, with $a < r < b$, $H_\gamma^\perp$ is not semidefinite.*

Proof.

1. Suppose X and Y are P-Jacobi fields along γ. Because $Y(a)$ is tangent to P,

$$\langle X'(a)\,,\,Y(a)\rangle = \left\langle X'(a)^{\top}\,,\,Y(a)\right\rangle .$$

Now, the second condition for X to be a P-Jacobi field implies

$$\langle X'(a)\,,\,Y(a)\rangle = \left\langle \widetilde{\mathrm{II}}\big(X(a),\gamma'(a)\big)\,,\,Y(a)\right\rangle .$$

Finally, using the identity $\left\langle y\,,\,\widetilde{\mathrm{II}}(x,n)\right\rangle = -\langle \mathrm{II}(y,x)\,,\,n\rangle$,

$$\langle X'(a)\,,\,Y(a)\rangle = -\big\langle \mathrm{II}\left(X(a),Y(a)\right)\,,\,\gamma'(a)\big\rangle .$$

But, because of the symmetry of the second fundamental form, starting instead from $\langle Y'(a)\,,\,X(a)\rangle$, the result is the same. Thus,

$$\langle X'(a)\,,\,Y(a)\rangle = \langle X(a)\,,\,Y'(a)\rangle\,,$$

and, by Lemma 16, $\langle X'\,,\,Y\rangle = \langle X\,,\,Y'\rangle$ along all of γ.

Let $Y_1,\ldots,Y_k$ be a basis for the P-Jacobi fields which are perpendicular to γ. Then, for any $X \in T_\gamma^\perp\Omega$, there exist continuous and piecewise smooth functions $f^1,\ldots,f^k$ on $[a,b]$ which are uniquely defined and such that $X = f^iY_i$. By the observation above, $\left\langle Y_i'\,,\,Y_j\right\rangle = \left\langle Y_i\,,\,Y_j'\right\rangle$. Then, Lemma 17 applies, and

$$H_\gamma^\perp(X,X) = \int_a^b \langle A\,,\,A\rangle\,\mathrm{d}u + \langle X\,,\,B\rangle|_a^b - \big\langle \gamma'(a)\,,\,\mathrm{II}\left(X(a),X(a)\right)\big\rangle .$$

Now we show that the boundary terms cancel out. Using the fact that $X(b) = 0$ and the expression for B (see Lemma 17),

$$\langle X\,,\,B\rangle|_a^b = -\langle X(a)\,,\,B(a)\rangle = -f^i(a)\big\langle X(a)\,,\,Y_i'(a)\big\rangle .$$

Since $X(a)$ is tangent to P and the Y_i are all P-Jacobi fields,

$$\begin{aligned}\langle X\,,\,B\rangle|_a^b &= -f^i(a)\big\langle X(a)\,,\,Y_i'(a)^\top\big\rangle\\ &= -f^i(a)\Big\langle X(a)\,,\,\widetilde{\mathrm{II}}\big(Y_i(a),\gamma'(a)\big)\Big\rangle\\ &= -\Big\langle X(a)\,,\,\widetilde{\mathrm{II}}\big(X(a),\gamma'(a)\big)\Big\rangle\\ &= \big\langle \mathrm{II}\left(X(a),X(a)\right)\,,\,\gamma'(a)\big\rangle\,,\end{aligned}$$

cancelling out the last term.

Finally, since γ is cospacelike and each Y_i is orthogonal to γ, $\langle A\,,\,A\rangle \geq 0$; it follows that $H_\gamma^\perp(X,X) \geq 0$. Besides, if $H_\gamma^\perp(X,X) = 0$, then $\langle A\,,\,A\rangle$ must vanish identically, implying each f^i is constant. But because X vanishes at q, the f^i are all zero at b, and hence everywhere. Then, $X = 0$.

2. By Lemma 25, $H_\gamma^\perp$ has a nontrivial kernel, and is therefore not definite. Thus it remains to show that $H_\gamma^\perp(X,X) \geq 0$ for all nonvanishing $X \in T_\gamma\Omega$.

   Take one such $X \in T_\gamma^\perp\Omega$, and let $r_n = b - 1/n$. For each n large enough that r_n falls within the last smooth segment of X, define Y_n as the parallel transport of $X(r_n)$ along γ and

$$X_n(u) = \begin{cases} X(u)\,, & \text{for } u \in [a, r_n] \\ \left(1 - 2\dfrac{u - r_n}{b - r_n}\right)Y_n(u)\,, & \text{for } u \in \left(r_n, \frac{b+r_n}{2}\right) \\ 0\,, & \text{for } u \in \left[\frac{b+r_n}{2}, b\right] \end{cases}.$$

   Then, the sequence $\left\{H_\gamma^\perp(X_n,X_n)\right\}$ converges to $H_\gamma^\perp(X,X)$.[4] But, because there are no conjugate points to a on $(a, r_n]$, part 1 applies, and $H_\gamma^\perp(X_n,X_n) > 0$. Hence $H_\gamma^\perp(X,X) \geq 0$ and $H_\gamma^\perp$ is positive semidefinite.

3. By Proposition 13, using the fact that γ is cospacelike, we can always obtain $X \in T_\gamma^\perp\Omega$ such that $H_\gamma^\perp(X,X) > 0$. So it remains to show that the opposite inequality also occurs.

   By hypothesis, there exists a nontrivial P-Jacobi field Y on γ which vanishes

[4] See footnote 3 on page 30 and Appendix B.

at r. Let $X = Y$ on $[a, r)$ and $X = 0$ on $[r, b]$. Since Y is not identically zero, $X'(r^-) \neq 0$. Let $Z \in T_\gamma\Omega$ with $Z(a) = 0$ and $Z(r) = -X'(r^-)$, and let also $\delta > 0$.

$$H^\perp_\gamma(X + \delta Z, X + \delta Z) = H^\perp_\gamma(X, X) + 2\delta H^\perp_\gamma(X, Z) + \delta^2 H^\perp_\gamma(Z, Z)\,.$$

Because X is a piecewise Jacobi field, the first term vanishes. Meanwhile, because X vanishes identically on $[r, b]$, the second one reduces to

$$2\delta H^\perp_\gamma(X, Z) = 2\delta \int_a^r [\langle X' \,,\, Z'\rangle + \langle R(X, \gamma')Z \,,\, \gamma'\rangle]\, \mathrm{d}u\,.$$

Using $\langle X' \,,\, Z'\rangle = \langle X' \,,\, Z\rangle' - \langle X'' \,,\, Z\rangle$,

$$\begin{aligned} 2\delta H^\perp_\gamma(X, Z) &= 2\delta \int_a^r \left[\langle X' \,,\, Z\rangle' - \langle X'' + R(X, \gamma')\gamma' \,,\, Z\rangle\right] \mathrm{d}u \\ &= 2\delta\langle X'(r^-) \,,\, Z(r)\rangle \;=\; -2\delta\langle X'(r^-) \,,\, X'(r^-)\rangle \;<\; 0\,, \end{aligned}$$

the inequality stemming from the fact that γ is cospacelike. Then, for sufficiently small δ, $H^\perp_\gamma(X + \delta Z, X + \delta Z) < 0$. □

2.5.2 The lightlike case

Now we turn to the problem of locating focal points of a submanifold along a *lightlike* geodesic. The following result is an extension of Proposition 22; it shows that for a P-Jacobi field to be the variation vector field of a curve γ by *lightlike* geodesics perpendicular to P, it must satisfy one extra condition.

Proposition 27. *Let γ be a lightlike geodesic normal to P, a semi-Riemannian submanifold of M. A P-Jacobi field on γ is the variation vector field of a variation of γ by lightlike P-normal geodesics if, and only if, it is orthogonal to γ.*

Proof. Let Φ be such a variation. Since the longitudinal curves are all lightlike, $\langle U(u, v) \,,\, U(u, v)\rangle$—and hence also $E_\Phi(s)$—identically vanishes. Then, the first variation formula (2.7) implies that $\langle W', \gamma' \rangle \equiv 0$. By the Levi-Civita property,

$$\langle W', \gamma' \rangle = \langle W, \gamma' \rangle' .$$

But, since $W(a)$ is orthogonal to $\gamma'(a)$, the equation above implies that $W(u)$ remains perpendicular to $\gamma'(u)$ for all u.

For the converse, let Y be a P-Jacobi field orthogonal to γ. Then,

$$\langle Y', \gamma' \rangle = \langle Y, \gamma' \rangle' \equiv 0 ,$$

Specialising to $u = a$, we conclude that $Y'(a)^{\perp}$, the component of $Y'(a)$ orthogonal to P, is perpendicular to $\gamma'(a)$ (the tangent component does not contribute to the inner product above). $\gamma'(a)$ and $Y'(a)^{\perp}$ span a degenerate subspace of $T^{\perp}_{\gamma(a)}P$, and, since $\gamma'(a)$ itself is lightlike, we can infer that $Y'(a)^{\perp}$ canonically corresponds to a vector which is tangent to the lightcone in $T^{\perp}_{\gamma(a)}P$ at $\gamma'(a)$.[5] One can therefore construct a curve λ on the lightcone which has that vector as its initial velocity. Note that each point along that curve is a lightlike vector normal to P.

Now, take a curve σ on P such that $\sigma'(0) = Y(a)$. We construct a vector field Z along σ consisting of lightlike vectors normal to P in the following way: for each v, let $Z(v)$ be the normal parallel transport of $\lambda(v)$ along σ. Then,

$$Z'(0) = \widetilde{\mathrm{II}}\big(\sigma'(0), Z(0)\big) + Y'(a)^{\perp} .$$[6]

Now, we recognise in the previous formula $\sigma'(0) = Y(a)$ and $Z(0) = \gamma'(a)$; using the fact that Y is a P-Jacobi field,

$$Z'(0) = Y'(a)^{\top} + Y'(a)^{\perp} = Y'(a) .$$

[5]For a proof, see e.g. the section "Lorentz Causal Character" in [32], Chapter 5.

[6]To see this, note that we can define an auxiliary two-parameter vector field $\zeta(u, v)$ by first letting $\zeta(0, v) = \lambda(v)$ and then defining $\zeta(u, v)$ as the normal parallel transport of $\zeta(0, v)$ along σ up to parameter u. With this, $Z(v)$ becomes the composition of ζ with the curve $u = v$ in the (u, v)-parameter space, and $Z'(0)$ can be calculated in terms of the covariant derivatives of the curves $\zeta(0, v)$ and $\zeta(u, 0)$ at $v = 0$ and $u = 0$, respectively. The first term reduces to $Y'(a)^{\perp}$, because of the way the curve λ was constructed. Meanwhile, the second becomes $\widetilde{\mathrm{II}}\big(\sigma'(0), Z(0)\big)$, because $\zeta(u, 0)$ is a normal parallel transport of $\zeta(0, 0) = Z(0)$ along σ.

From this point, the construction proceeds exactly as in Proposition 22:

$$\Phi(u,v) = \exp\big((u-a)Z(v)\big)$$

defines a variation of γ by P-normal *lightlike* geodesics whose variational vector field which coincides with Y, since they are both Jacobi fields satisfying the same initial conditions at a. □

Theorem 28. *Let P be a spacelike submanifold of a Lorentz manifold M. If P has no focal points along a normal lightlike geodesic γ, then $H^{\perp}_{\gamma}$ is positive semidefinite. Furthermore, $H^{\perp}_{\gamma}(X,X) = 0$ implies that X is tangent to γ.*

Proof. Let $Y_1, \ldots, Y_k$ be a basis for the P-Jacobi fields perpendicular along γ, with $Y_1(u) = (u-a)\gamma'(u)$. Because there are no focal points along γ, $\{Y^1(u), \ldots, Y^{m-1}(u)\}$ remains a basis for $\gamma(u)^{\perp}$ for every u along γ, and any $X \in T^{\perp}_{\gamma}\Omega$ can be written as f^iY_i, for some set of piecewise smooth functions $f_1, \ldots, f_{m-1}$. Once again, by Lemmas 16 and 17,

$$\langle X', X'\rangle + \langle R(X,\gamma')X, \gamma'\rangle = \langle A, A\rangle + \langle X, B\rangle',$$

and $H^{\perp}_{\gamma}(X,X)$ reduces to $\int_a^b \langle A, A\rangle\,\mathrm{d}u$. Since A is orthogonal to γ and γ is lightlike, $\langle A, A\rangle \geq 0$, with equality holding if and only if A is tangent to γ. Therefore, $H^{\perp}_{\gamma}(X,X) \geq 0$. Equality holds if and only if $f^2, \ldots, f^{m-1}$ are all constant. But $X(q) = 0$ implies that $f^2, \ldots, f^{m-1}$ vanish at that point, and hence identically, so that $X = f^1Y_1$, which is tangent to γ. □

2.6 First consequences

We conclude this chapter by deducing alternative criteria for the presence of focal points along a timelike or lightlike geodesic, using Theorems 26 and 28, respectively.

Consider the following setup: let P be a spacelike hypersurface of a Lorentzian manifold M (so $\dim P = m-1$) and $\gamma : [a,b] \to M$ a timelike geodesic issuing orthogonally from P at $\gamma(a) = p$. Take an orthonormal basis $\{e_1, \ldots, e_{m-1}\}$

for T_pP, and let $E_1, \dots, E_{m-1}$ be the parallel transports of $e_1, \dots, e_{m-1}$ along γ, respectively. Then, for each $u \in [a, b]$, $\{E_1(u), \dots, E_{m-1}(u)\}$ forms an orthonormal basis for $\gamma'(u)^\perp$. If $f : [a, b] \to \mathbb{R}$ is a piecewise smooth function with $f(a) = 1$ and $f(b) = 0$, then, for each i, fE_i is an orthogonal vector field along γ which is tangent to P at a and vanishes at b. In other words, fE_i is in $T_\gamma^\perp\Omega$, and hence in the domain of $H_\gamma^\perp$, so that we can calculate

$$H_\gamma^\perp(fE_i, fE_i) = -f(a)^2\langle\gamma'(a)\,,\,\mathrm{II}(e_i, e_i)\rangle + \\ + \int_a^b \left[f'(u)^2\langle E_i\,,\,E_i\rangle + f(u)^2\langle R(E_i, \gamma')E_i\,,\,\gamma'\rangle\right] \mathrm{d}u\,.$$

Summing over all i,

$$\sum_{i=1}^{m-1} H_\gamma^\perp(fE_i, fE_i) = -\sum_{i=1}^{m-1}\langle\gamma'(a)\,,\,\mathrm{II}(e_i, e_i)\rangle + \\ + \int_a^b \left[(m-1)f'(u)^2 + f(u)^2\sum_{i=1}^{m-1}\langle R(E_i, \gamma')E_i\,,\,\gamma'\rangle\right] \mathrm{d}u\,.$$

The sum in the boundary term is proportional to the **mean curvature vector field** of P, which is defined as

$$H(p) = \frac{1}{m-1}\sum_{i=1}^{m-1}\mathrm{II}(e_i, e_i)\,.$$ [7]

For the sum within the integral, we have

$$\sum_{i=1}^{m-1}\langle R(E_i, \gamma')E_i\,,\,\gamma'\rangle = -\sum_{i=1}^{m-1}\langle R(E_i, \gamma')\gamma'\,,\,E_i\rangle\,.$$

[7] More generally, if P is any semi-Riemannian submanifold and the e_i form an orthonormal basis for T_pP, then

$$H(p) := \frac{1}{\dim P}\sum_{i=1}^{\dim P}\sigma_i\,\mathrm{II}(e_i, e_i)\,,$$

where $\sigma_i = \langle e_i\,,\,e_i\rangle$. It is regrettable that the mean curvature is typically denoted by the letter H, which is already in use in this text for the Hessian. However, this should not lead to any confusion, as the Hessian will always carry a subscript to indicate the curve on which it is defined.

We can define a vector field $E_0(u)$ along γ which completes $E_1(u), \ldots, E_{m-1}(u)$ to an orthonormal basis of $T_{\gamma(u)}M$, and then add and subtract $\langle R(E_0, \gamma')\gamma' \,,\, E_0\rangle$ in the formula above to obtain $\mathrm{Ric}(\gamma', \gamma')$:

$$\sum_{i=1}^{m-1} \langle R(E_i, \gamma')E_i \,,\, \gamma'\rangle = -\,\mathrm{Ric}(\gamma', \gamma') - \langle R(E_0, \gamma')\gamma' \,,\, E_0\rangle \,.$$

However, the field E_0 is collinear with γ', and hence the extra term vanishes, by the antisymmetry of R. We find

$$\sum_{i=1}^{m-1} H_\gamma^\perp(fE_i, fE_i) = -(m-1)\langle \gamma'(a) \,,\, H(p)\rangle + \\ + \int_a^b \left[(m-1)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\right] \mathrm{d}u$$

Now, suppose that for some function f obeying the previous assumptions,

$$\int_a^b \left[(m-1)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\right] \mathrm{d}u \le (m-1)\langle \gamma'(a) \,,\, H(p)\rangle \,. \tag{2.13}$$

Then,

$$\sum_{i=1}^{m-1} H_\gamma^\perp(fE_i, fE_i) \le 0 \,.$$

But, for this to hold, at least one of the terms in the sum on the left hand side has to be nonpositive, which, by Theorem 26, implies the existence of a focal point of P along γ in $(a, b]$. If the inequality in (2.13) is strict, the focal point lies within (a, b).

We shall call the function $k : NP \to \mathbb{R}$ given by

$$k(z) = \big\langle z \,,\, H\big(\pi(z)\big)\big\rangle = \frac{1}{m-1}\sum_{i=1}^{m-1} \langle z \,,\, \mathrm{II}(e_i, e_i)\rangle$$

the **convergence** of P (the projection $\pi(z)$ simply gives the base point of the normal vector z). The reason for this nomenclature is as follows. Let w be a unit tangent vector to P at the point p, $\gamma : [a, b] \to M$ a timelike geodesic issuing orthogonally from P at p, and consider a variation of γ by normal geodesics whose variation vector field is such that $W(a) = w$. Then, $\|W\| = \sqrt{\langle W \,,\, W\rangle}$ can be used as an estimate of the distance between γ and neighbouring geodesics in the variation. Its derivative along γ is

$$\|W\|' = \frac{\langle W' \,,\, W\rangle}{\sqrt{\langle W \,,\, W\rangle}}\,,$$

so that, at the starting point of the curve,

$$\|W\|'(a) = \frac{\langle W'(a) \,,\, w\rangle}{\sqrt{\langle w \,,\, w\rangle}} = \left\langle W'(a)^{\top} \,,\, w\right\rangle = \\ = \left\langle \widetilde{\mathrm{II}}\big(w, \gamma'(a)\big) \,,\, w\right\rangle = -\langle \mathrm{II}(w, w) \,,\, \gamma'(a)\rangle\,.$$

Therefore, the geodesics around γ have an initial tendency to converge towards γ if $\langle \mathrm{II}(w, w) \,,\, \gamma'(a)\rangle > 0$ and to diverge from it if $\langle \mathrm{II}(w, w) \,,\, \gamma'(a)\rangle < 0$. The convergence k is an averaged version of this quantity over all directions tangent to P.

With this definition, the argument we have presented in the preceding paragraphs can be restated as the following proposition:

Proposition 29. *Let P be a smooth spacelike hypersurface of M and $\gamma : [a, b] \to M$ a timelike geodesic issuing orthogonally from P. If there exists a piecewise smooth function $f : [a, b] \to \mathbb{R}$ such that $f(a) = 1$, $f(b) = 0$ and*

$$\int_a^b \left[(m-1)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\right] \mathrm{d}u \le (m-1)k\big(\gamma'(a)\big)\,, \qquad (2.14)$$

where k denotes the convergence of P, then there is a focal point of P along γ within $(a, b]$. If the strict inequality holds, that focal point lies in (a, b) instead.

The corollary below is an immediate consequence of this result:

Corollary 30. *Let P be a smooth spacelike hypersurface of M and $\gamma : [0, b] \to M$ a timelike geodesic issuing orthogonally from P. If:*

1. $\mathrm{Ric}(\gamma', \gamma') \geq 0$ *at every point along* γ*; and*

2. $k\big(\gamma'(0)\big) > 0$,

then, provided that $b \geq 1/k\big(\gamma'(0)\big)$*, there is a focal point of* P *along* γ *within* $\big(0, 1/k\big(\gamma'(0)\big)\big]$*. Note that condition 1 holds, in particular, if* (M, g) *satisfies the strong energy condition.*

Proof. Let us denote $k\big(\gamma'(0)\big)$ simply by k for short, and define $f(u) = 1 - ku$. Then, $f(0) = 1$, $f(1/k) = 0$ and $f'(u) = k$. Since $b \geq 1/k$, it is possible to restrict γ to $[0, 1/k]$ and calculate

$$\int\limits_0^{1/k} \big[(m-1)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\big]\,\mathrm{d}u = (m-1)k - \int\limits_0^{1/k} \mathrm{Ric}(\gamma', \gamma')f(u)^2\,\mathrm{d}u\,.$$

Because $\mathrm{Ric}(\gamma', \gamma')f(u)^2 \geq 0$ for all u, the result above is no greater than $(m-1)k$, and Proposition 29 guarantees the existence of the focal point. □

There is also an analogue of Proposition 29 for lightlike geodesics:

Proposition 31. *Let* P *be a spacelike submanifold of* M *with dimension* $m-2$*, and let* $\gamma : [a, b] \to M$ *be a lightlike geodesic issuing orthogonally from* P *at* $p = \gamma(a)$*. Suppose there exists a piecewise smooth function* $f : [a, b] \to \mathbb{R}$ *such that* $f(a) = 1$*,* $f(b) = 0$ *and*

$$\int\limits_a^b \big[(m-2)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\big]\,\mathrm{d}u \leq (m-2)k\big(\gamma'(a)\big)\,. \tag{2.15}$$

Then, there is a focal point of P *along* γ *in* $(a, b]$*. If the inequality is strict, the focal point is in* (a, b) *instead.*

Proof. The proof is very similar to the previous one. Let $e_1, \ldots, e_{m-2}$ be an orthonormal basis for T_pP and take parallel transports along γ to get $\{E_1, \ldots, E_{m-2}\}$.

For the given function f,

$$\sum_{i=1}^{m-2} H_\gamma^\perp(fE_i, fE_i) = -\left\langle \gamma'(a)\,,\, \sum_{i=1}^{m-2} \mathrm{II}(e_i, e_i) \right\rangle + \\ + \int_a^b \left[(m-2) f'(u)^2 + \sum_{i=1}^{m-2} \langle R(E_i, \gamma')E_i\,,\, \gamma' \rangle f(u)^2 \right] \mathrm{d}u\,.$$

The sum in the boundary term becomes $(m-2)H(p)$, as before. However, this time care must be taken when defining the basis on which to calculate the Ricci tensor. Let t be a unit timelike vector in T_pM orthogonal to P and T its parallel transport along γ. Then, define

$$S = T + \frac{1}{\langle T\,,\, \gamma' \rangle} \gamma'\,.$$

S is unit spacelike and is orthogonal to T and to each of the E_i; therefore, $\{T, S, E_1, \ldots, E_{m-2}\}$ is an orthonormal basis, which we can use to obtain:

$$\mathrm{Ric}(\gamma', \gamma') = -\langle R(T, \gamma')\gamma'\,,\, T \rangle + \langle R(S, \gamma')\gamma'\,,\, S \rangle + \sum_{i=1}^{m-2} \langle R(E_i, \gamma')\gamma'\,,\, E_i \rangle\,.$$

However,

$$\langle R(S, \gamma')\gamma'\,,\, S \rangle = \langle R(T, \gamma')\gamma'\,,\, T \rangle + \frac{1}{\langle T\,,\, \gamma' \rangle} \langle R(T, \gamma')\gamma'\,,\, \gamma' \rangle + \\ + \frac{1}{\langle T\,,\, \gamma' \rangle} \langle R(\gamma', \gamma')\gamma'\,,\, T \rangle + \frac{1}{\langle T\,,\, \gamma' \rangle^2} \langle R(\gamma', \gamma')\gamma'\,,\, \gamma' \rangle\,,$$

and, by the antisymmetry of R, all terms but the first vanish. Therefore,

$$\begin{aligned} \mathrm{Ric}(\gamma', \gamma') &= \sum_{i=1}^{m-2} \langle R(E_i, \gamma')\gamma'\,,\, E_i \rangle \\ &= -\sum_{i=1}^{m-2} \langle R(E_i, \gamma')E_i\,,\, \gamma' \rangle\,. \end{aligned}$$

Using this result, we find

$$\sum_{i=1}^{m-2} H_\gamma^\perp(fE_i, fE_i) = -(m-2)\langle\gamma'(a)\,,\,H(p)\rangle + \\ + \int_a^b \left[(m-2)f'(u)^2 - \mathrm{Ric}(\gamma',\gamma')f(u)^2\right] \mathrm{d}u\,.$$

Applying the hypothesis on f, we conclude that the sum on the left is nonpositive, which can only happen if at least one of the terms is also nonpositive. Thus, Theorem 28 guarantees the existence of the focal point. □

As in the timelike case, we have the following simple consequence:

Corollary 32. *Let P be an $(m-2)$-dimensional spacelike submanifold of M and $\gamma : [0,b] \longrightarrow M$ a lightlike geodesic issuing orthogonally from P. If:*

1. $\mathrm{Ric}(\gamma',\gamma') \geq 0$ *at every point along γ, and*
2. $k\big(\gamma'(0)\big) > 0$,

then, if $b \geq 1/k\big(\gamma'(0)\big)$, there is a focal point of P along γ in $\big(0, 1/k\big(\gamma'(0)\big)\big]$. Note that condition 1 is true, in particular, if the spacetime verifies the null energy condition.

Proof. Again, let k denote $k\big(\gamma'(0)\big)$ and define $f(u) = 1 - ku$. Then,

$$\int_0^{1/k} \left[(m-2)f'(u)^2 - \mathrm{Ric}(\gamma',\gamma')f(u)^2\right] \mathrm{d}u = (m-2)k - \int_0^{1/k} \mathrm{Ric}(\gamma',\gamma')f(u)^2\,\mathrm{d}u\,.$$

The integrand $\mathrm{Ric}(\gamma',\gamma')f(u)^2$ being nonnegative, the quantity above is no greater than $(m-2)k$, and the existence of a focal point follows from the previous proposition. □

Propositions 29 and 31 are never directly stated in O'Neill's textbook [32]; instead, they are only a step in the derivation of Corollaries 30 and 32 as consequences of Theorems 26 and 28. It was observed by Fewster and Kontou, in [31], that these

Propositions are themselves useful criteria for locating focal points along timelike and lightlike geodesics. Furthermore, as will be shown in greater detail in the next chapter, criteria of this form are particularly well-suited for attempts to relax the hypotheses of the singularity theorems of General Relativity to work with quantum energy conditions.

Chapter 3

Applications to singularity theorems

In this chapter, we present a review of the developments in the article [31], building upon the methods of Chapter 2 to arrive at the proofs of singularity theorems with weakened energy conditions. In Section 3.1, some auxiliary concepts and results are given; in Section 3.2, two points of comparison are established between the present setup and previous work on the subject; the introduction of the energy conditions used and the proofs of the singularity theorems themselves are the subject of Section 3.3; finally, in Section 3.4, we present our final remarks and conclusions.

3.1 Some auxiliary definitions and results

In this section, we present a few concepts appearing in the statements of the singularity theorems, some of which have been mentioned but not properly defined up until this point; most importantly, we define a globally hyperbolic spacetime and a future-converging submanifold. We also state two propositions which, when paired with our previous results on focal points, will give the proofs of the singularity theorems. The discussion will be brief and devoid of proofs, as the main goal here is simply the completeness of the text. For more in-depth treatments of these subjects, we direct the reader to the literature on Semi-Riemannian Geometry and General Relativity; see, e.g., [32, 34, 35]

3.1.1 Causality conditions

A causal geodesic $\gamma : [a, b] \to M$ is said to be **inextendible** if it cannot be extended to an interval containing $[a, b]$ whilst still remaining a causal geodesic. The concept of inextendibility is fundamental when dealing with the causal properties of spacetime, and is featured in the very definition of singularity:

Definition 33. A spacetime is said to be **singular** if it contains a causal geodesic which is inextendible and *incomplete*, in the sense that its affine parameter does not run from $-\infty$ to ∞.

Inextendibility is also in the basis of the concept of global hyperbolicity, which is the most relevant causality condition for our purposes:

Definition 34. A spacetime is **globally hyperbolic** if it contains a **Cauchy surface**, i.e., a smooth spacelike hypersurface which is intercepted by every inextendible causal geodesic exactly once.

Global hyperbolicity can be defined in several alternative equivalent ways, and implies that the spacetime verifies various other less stringent causality conditions. In order to better characterise globally hyperbolic spacetimes, we mention in passing some of these alternative definitions and causal properties:

- If a spacetime contains a Cauchy surface, it also contains infinitely many others, and they are all diffeomorphic. The spacetime can then be described as $\mathbb{R} \times S$, where S is any of its Cauchy surfaces.
- A globally hyperbolic spacetime possesses a global time function, i.e., a smooth function whose gradient is everywhere timelike. This property is known as **stable causality**. The existence of a global timelike vector field (call it T) also makes M a **time-orientable** spacetime. Then, each causal vector z tangent to M can be classified as **future-pointing** if $\langle z , T \rangle < 0$ or **past-pointing** if $\langle z , T \rangle > 0$. Causal curves are similarly classified as **future-** or **past-directed** according to whether their velocity vectors are future- or past-pointing.

- A globally hyperbolic spacetime also satisfies the **chronology condition**, i.e., it does not contain any closed timelike curves. In physical terms, time travel is impossible.

Global hyperbolicity and Cauchy surfaces are intimately related with the notion of *determinism* in classical physics. Indeed, any causal influence which reaches an event to the future of a Cauchy surface S must first pass through S. Therefore, with complete information on S, one should be able to describe the entire future of the Universe. As such, globally hyperbolic spaces are the natural setting for the initial value formulation of General Relativity, which seeks to rephrase the theory as the time evolution, via the Einstein equation, of initial data given on a Cauchy surface (say, the configuration of the Universe at the present time).

3.1.2 Future-converging submanifolds

Definition 35. Let P be a spacelike submanifold of a Lorentz manifold M with codimension greater than or equal to 2. We say that P is **future-converging** if its mean curvature vector field[1] H is timelike and past-pointing at each point.

Recalling the definition of the covergence of P (see page 45), it is clear that, if P is future-converging, then, for every future-pointing causal vector z orthogonal to P, $k(z) = \langle z\,, H\rangle > 0$. Then, according to the geometrical interpretation of the convergence previously discussed, all future-directed causal geodesics issuing orthogonally from P have an initial tendency to draw closer to their neighbours.

3.1.3 Causality and focal points

We begin this subsection by introducing some shorthand notation and conventions which will be used to work with the criteria established in Propositions 29 and 31 from here on out. When dealing with timelike geodesics, the hypotheses of the singularity theorems will always involve a certain Cauchy surface S, which will replace the generic spacelike submanifold P considered up until this point. Since S is a hypersurface, there is a single timelike direction orthogonal to S at

[1]Refer back to page 44 for the definition.

each point. We can therefore replace $(m-1)k\big(\gamma(a)\big)$, in the right hand side of Equation 2.14, by a scalar function K, obtained by constraining γ to be the unique future-directed unit-speed timelike geodesic issuing orthogonally from S at that point. With this, we can rewrite Equation 2.14 in the shorter form $J[f] \leq K$, where

$$J[f] := \int_a^b \big[(m-1)f'(u)^2 - \mathrm{Ric}(\gamma', \gamma')f(u)^2\big]\,\mathrm{d}u\,. \tag{3.1}$$

Similarly, in the lightlike case, we will replace P with an achronal, compact, smooth, future-converging submanifold T of dimension $m-2$. This allows us to fix a parametrisation for any future-directed lightlike geodesic γ issuing orthogonally from T, which cannot, in general, be done for lightlike curves in any canonical way. By hypothesis, the mean curvature vector field of T is past-pointing and timelike. Therefore, it can be written as $H = Kt/(m-2)$, where t is a *unit* past-pointing timelike vector field and K a positive function on T. We can then choose the unique affine parametrisation for γ such that $\langle \gamma'(a)\,,\, t\rangle = 1$. With this choice, the criterion (2.15) once again takes the form $J[f] \leq K$, with $J[f]$ defined as above, but with the $(m-1)$ factor replaced by $(m-2)$. For a lightlike geodesic γ parametrised in this way, we refer to the length of the parameter interval as the **T-length** of γ.

As was mentioned in Chapter 1, a crucial step in the proofs of singularity theorems is establishing that, if one assumes causal geodesic completeness, global hyperbolicity is at odds with the existence of focal points along every causal geodesic issuing normally from a certain spacelike submanifold. Then, the initial condition and the energy condition are used to show that, indeed, every such causal geodesic contains a focal point, which is a contradiction and forces one to discard the assumption of completeness.

Propositions 36 and 37, stated below without proof, establish this relationship between global hyperbolicity and focal points; the first one works within the setup of a Hawking-type singularity theorem, and the second in the Penrose case. These results are seldom stated thus isolated in the literature, and are usually simply a step in the proofs of the singularity theorems.

Proposition 36. *Let M be a globally hyperbolic spacetime and S a smooth and compact Cauchy surface in M. If every future-directed timelike geodesic which issues orthogonally from S and contains no focal points of S has proper length less than b, then every future-directed timelike curve issuing from S has proper length less than b (and M is therefore future timelike geodesically incomplete).*

Proposition 37. *Let M be a globally hyperbolic spacetime with non-compact Cauchy surface, and let T be a smooth $(m-2)$-dimensional spacelike submanifold of M which is achronal[2] and future-converging. If every future-directed lightlike geodesic issuing orthogonally from T with T-length at least b contains a focal point of T, then there exists an inextendible lightlike geodesic orthogonal to T with T-length less than b (and M is therefore future lightlike geodesically incomplete).*

3.2 Relations to previous work

Before proceeding to the proofs of the singularity theorems, we will follow two enlightening comments by Fewster and Kontou in [31] about the relations between their work and previous developments in the field. The first one regards the connection between the variational method used in their paper and the more traditional approach to singularity theorems, based on the Raychaudhuri equation. In the variational framework, the criterion for the existence of a focal point along a timelike geodesic is the existence of a function f satisfying the boundary conditions $f(a) = 1$, $f(b) = 0$ and the inequality (2.14). One way to obtain a sufficient condition for that to hold true is to search for the function f that *minimises* the functional J defined in Equation 3.1. If such a minimising function f exists, and it verifies the inequality $J[f] \leq K$, then Proposition 29 guarantees the existence of focal points. This is a standard variational problem, and the Euler-Lagrange equation for it is

$$(m-1)f'' + \mathrm{Ric}(\gamma', \gamma')f = 0\,,$$

[2] A set is achronal if it is intercepted by any timelike curve at most once.

subject to $f(a) = 1$ and $f(b) = 0$. Performing an integration by parts on the first term of J and applying the boundary conditions, we find

$$J[f] = -(m-1)f'(a) - \int_a^b [(m-1)f'' + f\,\mathrm{Ric}(\gamma',\gamma')]f\,\mathrm{d}u\,.$$

If f satisfies the Euler-Lagrange equation, then the integral term vanishes, and the criterion for the presence of a focal point along γ becomes

$$f'(a) \geq -\frac{K}{m-1}\,.$$

If one assumes f has no zeroes in $[a,b)$ [3], then it is possible to make the substitution $\theta = (m-1)f'/f$. One can then entirely rephrase the problem in terms of θ.

$$\theta' = (m-1)\left(\frac{f''}{f} - \frac{f'^2}{f^2}\right)$$

By the definition of θ and the differential equation for f,

$$\theta' = -\mathrm{Ric}(\gamma',\gamma') - \frac{\theta^2}{m-1}\,. \tag{3.2}$$

The boundary conditions become $\theta(a) = (m-1)f'(a)$ and $\theta \underset{u\to b^-}{\longrightarrow} -\infty$. Therefore, in terms of θ, the criterion for the existence of a focal point is the existence of a solution of (3.2) such that $\theta \underset{u\to b^-}{\longrightarrow} -\infty$ and $\theta(a) \geq -K/(m-1)$.

Notice the formal similarity between what was just done and the role played by the expansion θ in the standard approach to singularity theorems. There, the Raychaudhuri equation for the evolution of a geodesic congruence leads to the differential inequality

$$\theta' \leq -\mathrm{Ric}(\gamma',\gamma') - \frac{\theta^2}{m-1}$$

for θ. Then, using the Strong Energy Condition, it is possible to show that θ diverges to $-\infty$ in finite proper time, and that is precisely the criterion for the

[3] Which one can always do, by redefining b as the first root of f if necessary.

existence of a focal point along γ in that approach.

The second point made by Fewster and Kontou is that their results are an extension of the ones in the 2011 paper on singularity theorems by Fewster and Galloway [30]. This will be illustrated by showing that the energy conditions used in [30] *imply* that the criterion given in Proposition 29 is satisfied.

Let $\gamma : [0, +\infty) \to M$ be a cospacelike geodesic issuing orthogonally from a semi-Riemannian submanifold P, and fix a function $g \in \mathcal{C}_0^\infty(\mathbb{R})$ such that $g(u) = 1$ for all $u \in [0, 1]$ and g is nonincreasing for $u \geq 0$. Suppose there exists $c > 0$ such that

$$\frac{c}{2} + \liminf_{\tau \to +\infty} \left(- \int_0^\tau g^2(u/\tau) \mathrm{e}^{-2cu/(m-1)} \operatorname{Ric}(\gamma', \gamma') \, \mathrm{d}u \right) < K \, . \tag{3.3}$$

This is the form of the energy conditions used in [30]. Now, the task at hand is to use the information above to exhibit $b > 0$ and $f : [0, b] \to \mathbb{R}$ such that our criterion for focal points (2.14) is satisfied on $\gamma|_{[0,b]}$. The inequality above can be rephrased as

$$\liminf_{\tau \to +\infty} F(\tau) < K - \frac{c}{2} \, ,$$

where $F(\tau)$ denotes the quantity within parentheses in (3.3). If

$$\liminf_{\tau \to +\infty} F(\tau) \geq K - \frac{c}{2} \, ,$$

then, for every $\varepsilon > 0$, there exists $\tau_0 > 0$ such that $F(\tau) > K - \frac{c}{2} - \varepsilon$ for all $\tau > \tau_0$. Therefore, (3.3), which is the negation of this statement, implies the existence of $\varepsilon_0 > 0$ such that, for any $\tau > 0$, one can find $\tau' > \tau$ such that

$$F(\tau') \leq K - \frac{c}{2} - \varepsilon_0 \, . \tag{3.4}$$

Now, consider the family of functions

$$f_\tau(u) = g(u/\tau) \mathrm{e}^{-cu/(m-1)} \, ,$$

defined for each $\tau > 0$. We shall show that

$$\lim_{\tau\to+\infty} \int_0^\tau (m-1) f_\tau'(u)^2 \,\mathrm{d}u = \frac{c}{2} \,.$$

We have

$$f_\tau'(u) = \left(\frac{1}{\tau} g'(u/\tau) - \frac{c}{m-1} g(u/\tau)\right) \mathrm{e}^{-cu/(m-1)} \,,$$

and therefore

$$f_\tau'(u)^2 = \left(\frac{1}{\tau^2} g'(u/\tau)^2 - \frac{2c}{\tau(m-1)} g(u/\tau) g'(u/\tau) + \frac{c^2}{(m-1)^2} g^2(u/\tau)\right) \mathrm{e}^{-2cu/(m-1)} \,.$$

Now, because $0 \le u/\tau \le 1$ when u runs from 0 to τ and $g \equiv 1$ and $g' \equiv 0$ on $[0, 1]$,

$$\int_0^\tau (m-1) f_\tau'(u)^2 \,\mathrm{d}u = \frac{c^2}{m-1} \int_0^\tau \mathrm{e}^{-2cu/(m-1)} \,\mathrm{d}u = \frac{c}{2}\left(1 - \mathrm{e}^{-2c\tau/(m-1)}\right) .$$

Thus, given $\varepsilon > 0$, one can find $\tau_0 > 0$ such that

$$\int_0^\tau (m-1) f_\tau'(t)^2 \,\mathrm{d}t \le \frac{c}{2} + \varepsilon \,, \tag{3.5}$$

for all $\tau > \tau_0$. In particular, choose τ_0 such that this is true for the ε_0 in (3.4). Then, choose any $b > \tau_0$ such that (3.4) holds. Adding up (3.4) and (3.5) with $\tau' = \tau = b$

$$\int_0^b \left[(m-1)\left(f_b'\right)^2 - f_b^2 \operatorname{Ric}(\gamma', \gamma')\right] \mathrm{d}u \le K \,,$$

as intended.

3.3 Proving singularity theorems

This section is dedicated to the proofs of the singularity theorems with weakened energy conditions. In Subsection 3.3.1, we introduce the energy conditions to be assumed, and discuss the general strategy of the proof. Subsection 3.3.2 defines a family of test functions which will be used to work with both the energy condition and the criteria for the detection of focal points; the theorems are finally stated and proved in Subsections 3.3.3 and 3.3.4.

3.3.1 The energy condition

We will impose the following energy condition: for any causal curve $\gamma : I \to M$ and any test function $f \in \mathcal{C}^\infty(I)$,

$$\int_I \rho(u) f^2(u)\,\mathrm{d}u \geq -|||f|||^2\,, \tag{3.6}$$

where $\rho(u)$ denotes the energy density $\mathrm{Ric}\left(\gamma'(u)\,,\,\gamma'(u)\right)$ and the seminorm $|||\cdot|||$ is defined by

$$|||f|||^2 = Q_0(\gamma)\|f\|_2^2 + Q_N(\gamma)\left\|f^{(N)}\right\|_2^2. \tag{3.7}$$

In the expression above, $\|\cdot\|_2$ is the usual L^2-norm, and Q_0 and Q_N are positive constants (which may depend on the curve γ). Note that, for the constants Q_0 and Q_N to be well-defined, it is first necessary to fix a parametrisation for γ. This is done in the ways described at the end of Section 3.1 according to whether we are dealing with timelike or lightlike geodesics.

The form of this energy condition is inspired by quantum energy inequalities: notice how integrating ρ against the square of the test function f is reminiscent of how expectation values are calculated in quantum field theory, i.e., by smeared products. Also, note how this restriction does not prevent ρ from taking on negative values, but does limit violations of positivity by imposing a lower bound for certain time averages.

Equation 3.6 should make it clear why Propositions 29 and 31 are particularly well-suited to work with quantum-inspired energy inequalities: notice how the criterion for the existence of a focal point along γ and the energy condition are both stated in terms of a "test" function f. The method for obtaining singularity theorems with weakened energy hypotheses exploits this fact as follows. First, a suitable class of functions f is fixed. Then, the energy condition (3.6) is substituted into the criterion for the existence of focal points (2.14), where it controls the second term in the integral. The resulting inequality is a constraint on the convergence K, which, if satisfied, guarantees the existence of a focal point along the curve γ. This can then be used as an initial condition for a singularity theorem; if it holds pointwise across the entire spacelike submanifold under consideration, and the energy condition is also verified, then the existence of a singularity follows from Propositions 36 or 37.

There is, however, one difficulty which must be circumvented before the method above can be applied. The function f which appears in the criterion (2.14) is subject to the boundary conditions $f(0) = 1$ and $f(b) = 0$. Meanwhile, the test functions appearing in the energy condition are of compact support, and the energy integral has to be taken over the entirety of that support. Therefore, one cannot simply pick the same function f in both and integrate over the interval $[0, b]$. Fewster and Kontou present two different methods for choosing the function f so that the information provided by the energy condition can be used in conjunction with the criterion for focal points; these two methods lead to slightly different singularity theorems which are each best suited for different situations. After introducing a convenient family of test functions, we shall discuss each of these methods.

3.3.2 A class of test functions

Before proceeding with the definition of the test functions that will be used in this text, we should mention that they are different from the ones that appear in [31]. Fewster and Kontou construct their test functions by gluing identically vanishing segments to a regularised incomplete gamma function of a given order M (which is a polynomial of degree $2M - 1$). Therefore, the resulting function is not smooth, but only $2M - 1$ times continuously differentiable, and the energy condition is not immediately applicable. However, by using an approximation argument, the authors are able to establish that, when one chooses the order M suitably with respect to N (as appearing in the definition of the seminorm (3.7)), the energy inequality extends to the class of functions they use. Here, we shall attempt to bypass the need for this argument by defining test functions which are actually smooth.

To this end, we present a family of smooth, compactly supported functions on the real line whose L^2-norms (as well as those of their derivatives) can be neatly expressed in terms of the Euler gamma function. Throughout, we will often make use of the following data on trigonometric functions:

$$\frac{\mathrm{d}}{\mathrm{d}x}\sec x = \tan x \sec x \qquad \frac{\mathrm{d}}{\mathrm{d}x}\tan x = \sec^2 x$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\tan^2 x = \frac{\mathrm{d}}{\mathrm{d}x}\sec^2 x = 2\tan x \sec^2 x$$

For some real polynomial P and some integer $k \geq 1$, consider the function given by

$$g(x) = \sec^k x \, P(\tan x) \exp\!\left(-\frac{\tan^2 x}{2}\right), \tag{3.8}$$

on $(-\pi/2\,,\,\pi/2)$, and vanishing elsewhere. The exponential factor controls the divergences of the tangent and the secant near $\pm\pi/2$, thus making the connection with the identically vanishing parts of g smooth. This can be seen, for example, by applying L'Hôspital's rule. From here on out, we shall only concern ourselves with the behaviour of g within $(-\pi/2, \pi/2)$.

Suppose that the nth derivative of the function g has the form (3.8), for a given polynomial P_n. We will show that $g^{(n+1)}$ can be written in the same form, and thus obtain a recurrence formula for P_{n+1}. Indeed,

$$\begin{aligned} g^{(n+1)}(x) = k \sec^k x \tan x \, P_n(\tan x) \exp\left(-\frac{\tan^2 x}{2}\right) + \\ + \sec^{k+2} x \, P_n'(\tan x) \exp\left(-\frac{\tan^2 x}{2}\right) - \\ - \sec^{k+2} x \, \tan x P_n(\tan x) \exp\left(-\frac{\tan^2 x}{2}\right) . \end{aligned}$$

Then, using the trigonometric identity $\sec^2 x = \tan^2 x + 1$, we find

$$\begin{aligned} g^{(n+1)}(x) = \sec^k x \Big[\left(k - 1 - \tan^2 x\right) \tan x \, P_n(\tan x) + \\ + \left(1 + \tan^2 x\right) P_n'(\tan x) \Big] \exp\left(-\frac{\tan^2 x}{2}\right) , \end{aligned}$$

so that the intended result is proven, and the recurrence relation we sought turns out to be

$$P_{n+1}(t) = \left(t^2 + 1\right) P_n'(t) + \left[(k-1)t - t^3\right] P_n(t) .$$

We note the following important property of the formula above: P_{n+1} has the reverse parity of P_n. This is evident for the second term, where P_n is multiplied by an odd factor. For the first term, it follows from the fact that P_n' is odd if P_n is even and even if P_n is odd, and the multiplication by the even factor $\left(t^2 + 1\right)$ does not affect the parity.

Next, we will show how to express the L^2-norms of g and its derivatives in terms of the gamma function, whose definition we recall:

$$\Gamma(s) = \int_0^\infty t^{s-1} \mathrm{e}^{-t} \, \mathrm{d}t .$$

For simplicity, we shall assume that P_0 is either even or odd, so that the squares P_n^2 will be even for all n. Then, the squares of g and its derivatives will be even functions, and we will be able to replace the integration interval $(-\pi/2, \pi/2)$ with $(0, \pi/2)$.

$$
\begin{aligned}
\left\|g^{(n)}\right\|^2 &= \int_{-\pi/2}^{\pi/2} \sec^{2k} x\, P_n^2(\tan x) \exp\left(-\tan^2 x\right) \mathrm{d}x \\
&= 2 \int_0^{\pi/2} \frac{\tan x \sec^2 x}{\tan x} \sec^{2(k-1)} x\, P_n^2(\tan x) \exp\left(-\tan^2 x\right) \mathrm{d}x \\
&= 2 \int_0^{\pi/2} \frac{\tan x \sec^2 x}{\tan x} (\tan^2 x + 1)^{k-1} P_n^2(\tan x) \exp\left(-\tan^2 x\right) \mathrm{d}x
\end{aligned}
$$

Note how the factor $(\tan^2 x + 1)^{k-1} P_n^2(\tan x)$ is still an even polynomial in $\tan x$. Now, we make the substitution $t = \tan^2 x$ (so that $\frac{\mathrm{d}t}{\mathrm{d}x} = 2 \tan x \sec^2 x$), to obtain

$$
\left\|g^{(n)}\right\|^2 = \int_0^{\infty} t^{-1/2} (t+1)^{k-1} P_n^2\left(t^{1/2}\right) \mathrm{e}^{-t}\, \mathrm{d}t\,.
$$

This result can be immediately connected to the gamma function. For example, if the polynomial $(t+1)^{k-1} P_n^2\left(t^{1/2}\right)$ is given by $\sum_{i=0}^{N} a_i t^i$, then

$$
\left\|g^{(n)}\right\|^2 = \sum_{i=0}^{N} a_i \int_0^{\infty} t^{i-1/2} \mathrm{e}^{-t}\, \mathrm{d}t = \sum_{i=0}^{N} a_i \Gamma\left(i + \frac{1}{2}\right).
$$

Note how the gamma function is always evaluated at positive values, so that all the terms are well-defined. In fact, there is an analytic formula for the value of Γ at positive half-integer values of the argument:

$$
\Gamma\left(i + \frac{1}{2}\right) = \frac{(2i)!}{4^i i!} \sqrt{\pi} \qquad (\forall i \in \mathbb{N})\,.
$$

It will be convenient to translate and rescale g to obtain a function supported in $[0,1]$. Let $h(x) = g\big(\pi(x - 1/2)\big)$. Then, $h^{(n)}(x) = \pi^n g^{(n)}\big(\pi(x - 1/2)\big)$, and

$$
\left\|h^{(n)}\right\|^2 = \pi^{2n} \int_{-\infty}^{+\infty} g^{(n)}\big(\pi(x - 1/2)\big)^2 \mathrm{d}x = \pi^{2n-1} \left\|g^{(n)}\right\|^2.
$$

To work with the criterion for focal points, we would like a function which "stepped down" smoothly from 1 to 0. For convenience, we will instead define a fuction ψ which "steps up" smoothly from 0 to 1, and later reflect it as necessary. This is achieved by letting

$$\psi(u) = \frac{\int_0^u h(x)\,\mathrm{d}x}{\int_0^1 h(x)\,\mathrm{d}x} .$$

Then, for every $n \geq 1$, $\psi^{(n)} = \mu^{\frac{1}{2}} h^{(n-1)}$, where

$$\mu = \frac{1}{\left(\int_0^1 h(x)\,\mathrm{d}x\right)^2} .$$

The L^2-norms of all derivatives of ψ can then be expressed in terms of the known data for g:

$$\left\|\psi^{(n)}\right\|^2 = \mu\pi^{2n-3}\left\|g^{(n-1)}\right\| ;$$

$\|\psi\|^2$ itself is the only one which, in general, needs to be determined numerically.

Finally, we define three auxiliary constants A, B and C related to the function ψ which will greatly simplify certain expressions in what follows:

$$\begin{aligned} A &= \|\psi\|^2 \\ B &= \|\psi'\|^2 \\ C &= \left\|\psi^{(N)}\right\|^2 . \end{aligned}$$

3.3.3 Strategy I: The SEC initially holds

The first method used to make the functions appearing in the energy condition compatible with the ones appearing in the criterion for focal points is to impose that the strong energy condition holds along γ for a certain interval after it departs from the Cauchy surface S. More precisely, we assume the existence of two constants

$b_0 \in (0, b)$ and $\rho_0 \geq 0$ such that

$$\rho(u) \geq \rho_0 \qquad (\forall u \in [0, b_0]) \,.$$

Let

$$f(u) = \begin{cases} 1\,, & \text{for } u \in [0, b_0] \\ \psi\left(\dfrac{b-u}{b-b_0}\right), & \text{for } u \in (b_0, b] \end{cases} \,,$$

where ψ is the test function previously defined. We also define an auxiliary function

$$\varphi(u) = \begin{cases} \psi\left(\dfrac{u}{b_0}\right), & \text{for } u \in [0, b_0) \\ 1\,, & \text{for } u \in [b_0, b] \end{cases} \,;$$

see Figure 3.1. Then, unlike f and φ, the product φf obeys the required boundary conditions at 0 and b. This means that we can apply the energy condition to φf. We will rewrite f^2 in terms of φf using

$$f^2 = (\varphi f)^2 + (1 - \varphi^2) f^2 = (\varphi f)^2 + (1 - \varphi^2)\,,$$

where the first equation can be written for any pair of functions and the second comes from the fact that $f \equiv 1$ on the support of $(1 - \varphi)^2$, whereas f differs from 1 only at points where $(1 - \varphi^2)$ vanishes.

Taking all of these observations into account,

$$\begin{aligned} \int_0^b \rho(u) f^2(u)\,\mathrm{d}u &= \int_0^b \rho(u)\left[(\varphi f)^2(u) + \left(1 - \varphi^2(u)\right)\right]\mathrm{d}u \\ &\geq -Q_0(\gamma)\|\varphi f\|^2 - Q_N(\gamma)\left\|(\varphi f)^{(N)}\right\|^2 + \rho_0 \int_0^{b_0} \left(1 - \varphi^2(u)\right)\mathrm{d}u \,. \end{aligned}$$

Now, because $\varphi f = \varphi$ on $[0, b_0]$ and $\varphi f = f$ on $[b_0, b]$, the squared norms in the expression above split, and can be rewritten in terms of the data for our family of

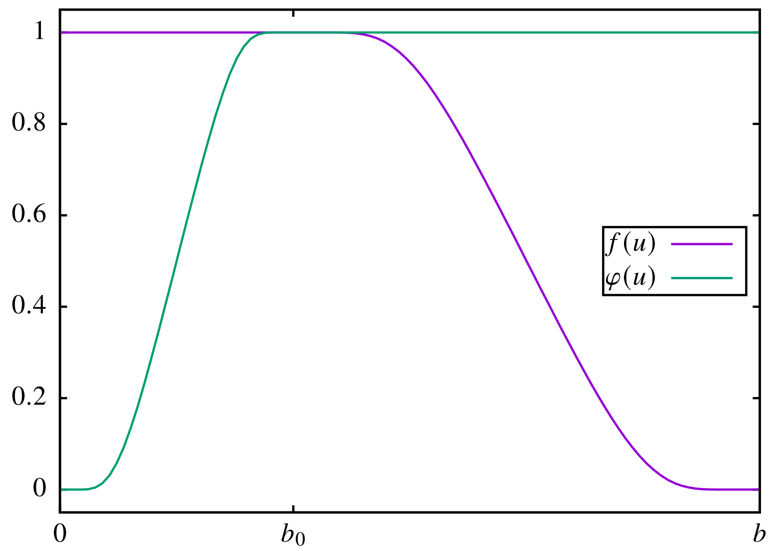

Figure 3.1: Example graphs showing the behaviour of the functions $f(u)$ and $\varphi(u)$.

test functions:

$$\int_0^b \rho(u) f^2(u)\,\mathrm{d}u \geq \rho_0 b_0 (1-A) - Q_0(\gamma) b A - \frac{Q_N(\gamma)C}{b_0^{2N-1}} - \frac{Q_N(\gamma)C}{(b-b_0)^{2N-1}}$$

The expression above gives us a bound on the second term of $J[f]$. The first one is easy to compute using that $f' \equiv 0$ on $[0, b_0]$ and applying the data on our test functions on the remainder of the interval:

$$\int_0^b (m-1) f'(u)^2\,\mathrm{d}u = (m-1)\int_{b_0}^b \psi'\left(\frac{b-u}{b-b_0}\right)^2 \mathrm{d}u = \frac{(m-1)B}{b-b_0}\,.$$

Finally, summing the two contributions, we find a bound for $J[f]$:

$$J[f] \leq -\rho_0 b_0 (1-A) + Q_0(\gamma) b A + \frac{(m-1)B}{b-b_0} + \frac{Q_N(\gamma)C}{b_0^{2N-1}} + \frac{Q_N(\gamma)C}{(b-b_0)^{2N-1}}\,.$$

If the right hand side of the above equation is less than or equal to K, then Proposition 29 guarantees the existence of a focal point of S along γ. Along with

Proposition 36, this allows us to state the following singularity theorem:

Theorem 38. *Let (M, g) be a globally hyperbolic spacetime of dimension $m > 2$ and let S be a compact smooth spacelike Cauchy surface for M. Suppose that, for some $b > 0$, there exist an integer $N \geq 1$ and constants Q_0 and Q_N such that:*

1. *For each unit speed timelike geodesic γ of proper length b issuing orthogonally from S,*

$$\int_\gamma \rho(u) f^2(u)\,\mathrm{d}u \geq -Q_0(\gamma)\|f\|^2 - Q_N(\gamma)\left\|f^{(N)}\right\|^2,$$

where $0 \leq Q_0(\gamma) \leq Q_0$, $0 \leq Q_N(\gamma) \leq Q_N$, with $\|\cdot\|$ denoting the usual L^2-norm and $\rho(u) = \mathrm{Ric}\left(\gamma'(u), \gamma'(u)\right)$.

2. *There exist $\rho_0 \geq 0$ and b_0 in $(0, b)$ such that, for each geodesic γ as above and each $u \in [0, b_0]$, $\rho(u) \geq \rho_0$ (i.e., the strong energy condition holds on $[0, b_0]$).*

3. *K, the convergence of S, satisfies*

$$K \geq \min\left\{\frac{m-1}{b_0}, \nu\right\},$$

where

$$\nu = -\rho_0 b_0(1-A) + Q_0 bA + \frac{(m-1)B}{b-b_0} + \frac{Q_N C}{b_0^{2N-1}} + \frac{Q_N C}{(b-b_0)^{2N-1}}.$$

Then, no future-directed timelike curve issuing from S has proper length greater than b, and M is future timelike geodesically incomplete.

If we instead apply our criterion to look for *lightlike* focal points of a compact, smooth, $(m-2)$-dimensional spacelike submanifold of M, the reasoning proceeds exactly as above, except for the substitution of factors of $m-1$ by $m-2$ and of the strong energy condition by the null energy condition. After including suitable causality hypotheses and applying Proposition 37, the following generalisation of Penrose's theorem is derived:

Theorem 39. *Let M be a globally hyperbolic spacetime with a noncompact Cauchy surface and let T be a smooth $(m-2)$-dimensional spacelike submanifold of M which is achronal and future-converging. Suppose that, for some $b > 0$, there exist an integer $N \geq 1$ and constants Q_0 and Q_N such that:*

1. *For every lightlike geodesic γ issuing orthogonally from T and with T-length b,*

$$\int_\gamma \rho(u) f^2(u)\,\mathrm{d}u \geq -Q_0(\gamma)\|f\|^2 - Q_N(\gamma)\left\|f^{(N)}\right\|^2,$$

where $0 \leq Q_0(\gamma) \leq Q_0$ and $0 \leq Q_N(\gamma) \leq Q_N$.

2. *There exist $\rho_0 \geq 0$ and $b_0 \in (0, b)$ such that $\rho(u) \geq \rho_0$ for each geodesic γ as above and every $u \in [0, b_0]$ (i.e., the null energy condition holds on $[0, b_0]$).*

3. *K, the convergence of T, satisfies*

$$K \geq \min\left\{\frac{m-2}{b_0}, \nu\right\},$$

with

$$\nu = -\rho_0 b_0(1-A) + Q_0 bA + \frac{(m-2)B}{b-b_0} + \frac{Q_N C}{b_0^{2N-1}} + \frac{Q_N C}{(b-b_0)^{2N-1}}.$$

Then, there exists an inextendible lightlike geodesic issuing orthogonally from T with T-length less than b, and M is future lightlike geodesically incomplete.

3.3.4 Strategy II: "Quantum Interest"

Take a unit speed timelike geodesic γ issuing orthogonally from S and extend it towards the past, obtaining, say, $\gamma : [-b_0, b] \to M$, with $b_0 > 0$ and $\gamma(0) \in S$.

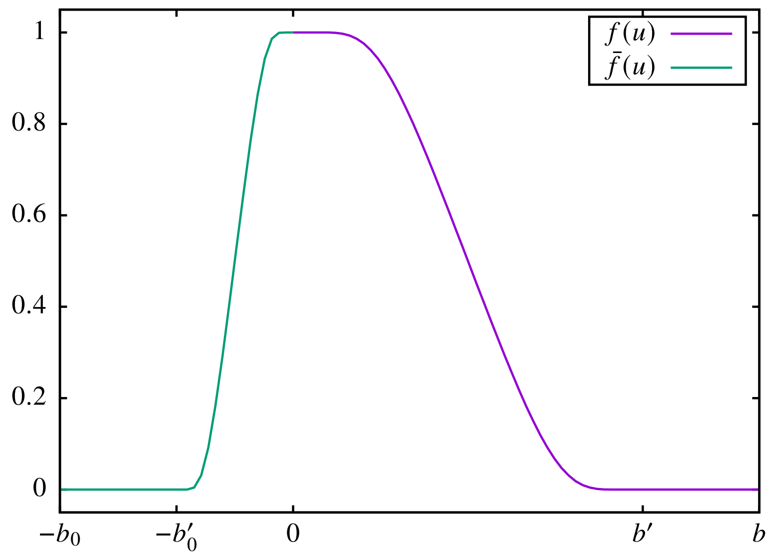

Figure 3.2: Example plots of f and $\bar{f}$. Note how the two coincide on the interval $[0, b]$.

Then, define

$$f(u) = \begin{cases} \psi\left(\dfrac{b' - u}{b'}\right), & \text{for } u \in [0, b') \\ 0\,, & \text{for } u \in [b', b] \end{cases} ,$$

for $b' \in (0, b)$. Let also

$$\bar{f}(u) = \begin{cases} 0\,, & \text{for } u \in [-b_0, -b_0'] \\ \psi\left(\dfrac{u + b_0'}{b_0'}\right), & \text{for } u \in (-b_0', 0) \\ f(u)\,, & \text{for } u \in [0, b] \end{cases} ,$$

where $b_0' \in (0, b_0)$. Figure 3.2 gives a graphical representation of the behaviour of f anf $\bar{f}$. Later on, we will use the free parameters b' and b_0' to optimise our bounds.

Now, we write the second term in $J[f]$ as

$$\int_0^b \rho(u) f^2(u)\,\mathrm{d}u = \int_{-b_0}^b \rho(u)\bar{f}^2(u)\,\mathrm{d}u - \int_{-b_0}^0 \rho(u)\bar{f}^2(u)\,\mathrm{d}u\,.$$

The energy condition can be applied to $\bar{f}$. To deal with the second term, we take a constant ρ_0 such that $\rho(u) \leq \rho_0$ for all $u \in [-b_0, 0]$. Then,

$$\int_0^b \rho(u) f^2(u)\,\mathrm{d}u \geq -Q_0(\gamma)\|\bar{f}\|^2 - Q_N(\gamma)\left\|\bar{f}^{(N)}\right\|^2 - \rho_0 \int_{-b_0}^0 \bar{f}^2(u)\,\mathrm{d}u\,.$$

The norms and the integral on the right hand side can all be written in terms of the data for our test functions; we find

$$\int_0^b \rho(u) f^2(u)\,\mathrm{d}u \geq -Q_0(\gamma)b'A - Q_0(\gamma)b_0'A - \frac{Q_N(\gamma)C}{b'^{2N-1}} - \frac{Q_N(\gamma)C}{{b_0'}^{2N-1}} - \rho_0 b_0' A\,.$$

Inserting into the formula for $J[f]$ and using the value calculated for the term involving f' in the previous subsection, we obtain

$$J[f] \leq \frac{(m-1)B}{b'} + Q_0(\gamma)b'A + Q_0(\gamma)b_0'A + \frac{Q_N(\gamma)C}{b'^{2N-1}} + \frac{Q_N(\gamma)C}{{b_0'}^{2N-1}} + \rho_0 b_0' A\,.$$

Exploiting the freedom of the parameters b' and b_0', we can guarantee the existence of a focal point of S within $[0, b]$ if

$$K \geq L + L_0\,,$$

where

$$L = \min\left\{Q_0(\gamma)b'A + \frac{(m-1)B}{b'} + \frac{Q_N(\gamma)C}{b'^{2N-1}} \;:\; b' \in (0, b)\right\} \tag{3.9}$$

$$L_0 = \min\left\{[Q_0(\gamma) + \rho_0]b_0'A + \frac{Q_N(\gamma)C}{{b_0'}^{2N-1}} \;:\; b_0' \in (0, b_0)\right\}. \tag{3.10}$$

Note how all terms depending on b' are grouped within L and those depending on b'_0 are in L_0. This leads to the following Hawking-type singularity theorem:

Theorem 40. *Let (M, g) be a smooth globally hyperbolic Lorentzian manifold of dimension $m > 2$ and S be a smooth, compact spacelike Cauchy surface for M. Suppose that, for some pair of positive numbers b and b_0, there exist an integer $N \geq 1$ and constants $Q_0, Q_N \geq 0$ such that:*

1. *Every unit speed future-directed timelike geodesic of proper length b issuing orthogonally from S can be extended to a geodesic of the form $\gamma : [-b_0, b] \to M$, along which*

$$\int_\gamma \rho(u) f^2(u)\, \mathrm{d}u \geq -Q_0(\gamma)\|f\|^2 - Q_N(\gamma)\left\|f^{(N)}\right\|^2$$

   holds for each test function $f \in \mathcal{C}_0^\infty([-b_0, b])$, where $0 \leq Q_0(\gamma) \leq Q_0$ and $0 \leq Q_N(\gamma) \leq Q_N$.

2. *There exists a finite upper bound ρ_0 such that $\rho(u) \leq \rho_0$ for every geodesic γ as above and every $u \in [-b_0, 0]$.*

3. *K, the convergence of S, verifies*

$$K \geq L + L_0$$

   at every point of S, with L and L_0 given by (3.9) and (3.10).

Then, no future-directed curve issuing from S has proper length greater than b and M is future timelike geodesically incomplete.

Note how decreasing ρ_0 makes L_0 also decrease, and hence, relaxes the requirement on the initial convergence K. This is the reason why we refer to this strategy as "quantum interest": violations of the SEC before S make it so that the energy tends to be positive after S, and this makes the focussing of geodesics more intense and the appearance of singularities more likely.

Again, applying the same reasoning to the setup of the Penrose theorem yields the following generalisation:

Theorem 41. *Let M be a globally hyperbolic spacetime with noncompact Cauchy surface and T be a smooth $(m-2)$-dimensional spacelike submanifold of M which is achronal and future-converging. Suppose that for some pair of positive numbers $b, b_0 > 0$ there exist an integer $N \geq 1$ and positive constants Q_0 and Q_N such that:*

1. *Every future directed lightlike geodesic γ issuing orthogonally from T and with T-length b can be extended to $[-b_0, b]$; besides, along the extended curve,*

$$\int_\gamma \rho(u) f^2(u)\,\mathrm{d}u \geq -Q_0(\gamma)\|f\|^2 - Q_N(\gamma)\left\|f^{(N)}\right\|^2,$$

   for every test function $f \in \mathcal{C}_0^\infty([-b_0, b])$, where $0 \leq Q_0(\gamma) \leq Q_0$ and $0 \leq Q_N(\gamma) \leq Q_N$.

2. *There exists a finite upper bound ρ_0 such that, along each extended geodesic as above, $\rho(u) \leq \rho_0$ for each $u \in [-b_0, 0]$ (i.e., on the portion of γ that lies to the past of T).*

3. *At each point of T, the convergence K verifies $K \geq L + L_0$, where*

$$L = \min\left\{Q_0(\gamma) b' A + \frac{(m-2)B}{b'} + \frac{Q_N(\gamma)C}{b'^{2N-1}} \;:\; b' \in (0, b)\right\}$$

$$L_0 = \min\left\{[Q_0(\gamma) + \rho_0] b_0' A + \frac{Q_N(\gamma)C}{{b_0'}^{2N-1}} \;:\; b_0' \in (0, b_0)\right\}.$$

Then, there exists an inextendible future directed lightlike geodesic issuing orthogonally from T with T-length less than b, and M is future lightlike geodesically incomplete.

3.4 Final remarks and conclusion

With these results in hand, the most immediate application is to try and derive singularity theorems for fields which are known to respect a certain energy inequality. For example, the final sections of both [30] and [31] are concerned with the Klein-Gordon field, as well as a more recent paper by Fewster, Kontou and Brown [36], which does the same on a curved background. The question of whether these theorems are useful for other types of classical and quantum fields is an interesting one for future work in the area.

One other line of inquiry which can be pursued is that of optimising the choice of the family of test functions one uses. Note that both the choice made here and the one in [31] are perfectly arbitrary, motivated by operational convenience. It is likely that the optimal choice is dependent on the particular setting, and, while the resulting family might be much more technically demanding to work with, it might lead to stronger singularity theorems (i.e., requiring less initial convergence).

Finally, we reiterate that the results discussed here are not proof that singularity theorems hold in quantum theory. As was already mentioned at the start, one can only expect results of this kind if the gravitational field itself is treated at the quantum level. These theorems are, however, the best indication of the answer that can be obtained to date, considering their potential application to the semiclassical theory.

Appendix A

Some mathematical prerequisites

In this appendix, we have attempted to collect a number of definitions and results from Semi-Riemannian Geometry which are not standard in general relativity textbooks but are essential to the understanding of this text. These prerequisites are grouped under two sections, one on Jacobi fields and one on Semi-Riemannian submanifolds. Our treatment leaves out some geometrical concepts which, despite being used throughout the text and not being standard topics in the Physics literature, are only required in proofs, and not essential for the understanding of the results themselves. For these more advanced topics, such as the exponential map, we refer the reader to the specialised mathematical literature (e.g., [32, 34]).

A.1 Jacobi fields

Definition 42. Let Y be a vector field on a geodesic γ. Y is said to be a **Jacobi Field** on γ if it satisfies the Jacobi Equation:

$$Y'' + R(Y, \gamma')\gamma' = 0\,.$$

Theorem 43. *Let $\gamma : [a, b] \to M$ be a geodesic with $\gamma(a) = p$, and let $v, w \in T_pM$. Then, there exists a unique Jacobi field Y on γ such that $Y(a) = v$ and $Y'(a) = w$.*

Proof. Take a parallel frame $E_1, \ldots, E_m$ on γ and define component functions for

Y via $Y = Y^i E_i$ and components for v and w via $v = v^i E_i(a)$ and $w = w^i E_i(a)$. Because γ is a geodesic, the components of γ' on this parallel frame are all constants: $\gamma' = c^i E_i$. Then, the Jacobi equation for Y can be rewritten as

$$0 = Y'' + R(Y, \gamma')\gamma' = \left[\left(Y^i\right)'' + R_{jkl}{}^i Y^j c^k c^l \right] E_i \,,$$

which is equivalent to

$$\left(Y^i\right)'' + R_{jkl}{}^i Y^j c^k c^l = 0 \qquad (1 \le i \le m) \,.$$

This system is also subject to the initial conditions

$$Y^i(a) = v^i \quad , \quad \left(Y^i\right)'(a) = w^i \qquad (1 \le i \le m) \,.$$

By standard results in ordinary linear differential equations, because the components of the curvature tensor are smooth, this system has a uniquely defined, smooth solution for Y on the domain of γ. □

The Jacobi equation is linear, and hence the set of Jacobi fields on γ is a real vector space. Since, by the Lemma above, each of these fields is uniquely determined by the initial conditions for Y and Y' at a point along γ, the dimension of this vector space is $2m$.

We say that a vector field V along a curve $\gamma : [a, b] \to M$ is **tangent** if $V = f\gamma'$ for some smooth function f on $[a, b]$. V is **perpendicular** to γ if $\langle V \,, \gamma' \rangle = 0$ at each point.

If $\gamma'(t)$ is not lightlike, $V(t)$ has a unique decomposition of the form $V^{\perp}(t) + V^{\top}(t)$, where $V^{\perp}(t)$ is orthogonal to $\gamma'(t)$ and $V^{\top}(t)$ is collinear with $\gamma'(t)$. When $\gamma'(t)$ is lightlike, the tangential direction is itself orthogonal, and there is no canonical way of splitting $V(t)$ into perpendicular and non-perpendicular components.

If γ is a geodesic, $V \perp \gamma$ implies $V' \perp \gamma$, since

$$\langle V' \,, \gamma' \rangle = \langle V \,, \gamma' \rangle' - \langle V \,, \gamma'' \rangle = 0 \,.$$

If $V = f\gamma'$, then

$$V' = f'\gamma' + f\gamma'' = f'\gamma',$$

implying that V' is also a tangent vector field. Therefore, along non-lightlike geodesics, V' (and also all higher order covariant derivatives of V) splits into tangential and perpendicular components and

$$(V')^{\top} = \left(V^{\top}\right)' \quad \text{and} \quad (V')^{\perp} = \left(V^{\perp}\right)'.$$

Proposition 44. *Let Y be a vector field on a geodesic γ.*

1. *If Y is tangent, the following are equivalent:*
   - *(i) Y is a Jacobi field.*
   - *(ii) $Y'' = 0$.*
   - *(iii) $Y(t) = (\alpha t + \beta)\gamma'(t)$, with $\alpha, \beta \in \mathbb{R}$.*
2. *If Y is a Jacobi field, the following are equivalent:*
   - *(i) Y is perpendicular to γ.*
   - *(ii) There exist distinct points a and b in the domain of γ such that $Y(a) \perp \gamma'(a)$ and $Y(b) \perp \gamma'(b)$.*
   - *(iii) There exists a in the domain of γ such that $Y(a)$ and $Y'(a)$ are orthogonal to $\gamma'(a)$.*
3. *If γ is not lightlike, the following are equivalent:*
   - *(i) Y is a Jacobi field.*
   - *(ii) $Y^{\top}$ and $Y^{\perp}$ are Jacobi fields.*

Proof.

1. The antisymmetry of the curvature operator implies that *(ii)* is just the form the Jacobi equation takes when Y is tangent, so it is immediately equivalent to *(i)*. Besides, if $Y = f\gamma'$, $Y'' = 0$ if and only if $f'' = 0$, since γ is a geodesic. Thus, *(ii)* is also equivalent to *(iii)*.

2. If we calculate $\langle Y\,,\,\gamma'\rangle''$ and apply the Jacobi equation and the fact that γ is a geodesic, we get

$$\langle Y\,,\,\gamma'\rangle'' = -\langle R(Y,\gamma')\gamma'\,,\,\gamma'\rangle = 0\,,$$

by elementary properties of the curvature. Therefore, $\langle Y\,,\,\gamma'\rangle$ is an affine function $\alpha t + \beta$, and its derivative $\langle Y'\,,\,\gamma'\rangle$ is the constant α; it results that all three assertions in the statement are equivalent to $\alpha = \beta = 0$.

3. Decomposing Y into tangential and orthogonal components,

$$Y'' + R(Y,\gamma')\gamma' = \left(Y^{\top}\right)'' + \left(Y^{\perp}\right)'' + R(Y^{\perp},\gamma')\gamma' \tag{A.1}$$

(the term $R(Y^{\top},\gamma')\gamma'$ vanishes by the antisymmetry of the curvature). Also by properties of the curvature,

$$\left\langle R(Y^{\perp},\gamma')\gamma'\,,\,\gamma'\right\rangle = \left\langle R(\gamma',\gamma')Y^{\perp}\,,\,\gamma'\right\rangle = 0\,.$$

Therefore, the first term in the right hand side of (A.1) is tangential to γ, whereas the final two are orthogonal; the sum can only be zero if these components both vanish, and so

$$Y'' + R(Y,\gamma')\gamma' = 0 \qquad \Longleftrightarrow \qquad \begin{cases} \left(Y^{\top}\right)'' = 0 \\ \left(Y^{\perp}\right)'' + R(Y^{\perp},\gamma')\gamma' = 0 \end{cases}. \qquad \square$$

Definition 45. Given a nonzero vector $v \in T_pM$, the **tidal force operator** $F_v : v^{\perp} \longrightarrow v^{\perp}$ is defined as

$$F_v(y) = -R(y,v)v\,.$$

Notice that, in terms of the tidal force, the Jacobi equation for a vector field Y

along γ can be written as $Y'' = F_{\gamma'}(Y)$.

Proposition 46. *F_v is a self-adjoint linear operator on $v^\perp$ and* $\operatorname{tr} F_v = -\operatorname{Ric}(v,v)$.

Proof. Self-adjointness follows from symmetries of the curvature:

$$\begin{aligned}\langle F_v(x)\,,\, y\rangle = \langle -R(x,v)v\,,\, y\rangle = \langle R(v,x)v\,,\, y\rangle = \\ = \langle R(v,y)v\,,\, x\rangle = \langle -R(y,v)v\,,\, x\rangle = \langle F_v(y)\,,\, x\rangle\,.\end{aligned}$$

If v is not lightlike, let $\{e_1,\ldots,e_{m-1}\}$ be an orthonormal basis for $v^\perp$. Then,

$$\operatorname{tr} F_v = \sum_{i=1}^{m-1} \sigma_i \langle -R(e_i,v)v\,,\, e_i\rangle = -\operatorname{Ric}(v,v)\,,$$

where $\sigma_i = \langle e_i\,,\, e_i\rangle = \pm 1$. If v is lightlike, its direction is itself contained in $v^\perp$, and thus the metric is degenerate in this subspace. Therefore, the basis has to be picked more carefully for the calculation of the trace. Take a lightlike vector w such that $\langle v\,,\, w\rangle = -1$ and let

$$e_1 = \frac{1}{\sqrt{2}}(v+w) \quad \text{and} \quad e_2 = \frac{1}{\sqrt{2}}(v-w)\,;$$

then, e_1 is unit timelike, e_2 unit spacelike, and they are mutually orthogonal. We can then complete to an orthonormal basis $\{e_1,\ldots,e_m\}$ of T_pM, and calculate

$$\operatorname{Ric}(v,v) = \sum_{i=1}^{m} \sigma_i \langle R(e_i,v)v\,,\, e_i\rangle\,.$$

Replacing e_1 and e_2 by their expressions in terms of v and w, the first two terms in the sum become

$$\sigma_1 \langle R(e_1,v)v\,,\, e_1\rangle = -\frac{1}{2}\langle R(w,v)v\,,\, w\rangle$$

and

$$\sigma_2 \langle R(e_2,v)v\,,\, e_2\rangle = \frac{1}{2}\langle R(w,v)v\,,\, w\rangle\,;$$

three other terms vanish in each of these equations by virtue of containing three or more factors of v. The two expressions above cancel out, and we conclude

$$\begin{aligned}\mathrm{Ric}(v,v) &= \sum_{i=3}^{m} \sigma_i \langle R(e_i,v)v\,,\,e_i\rangle \\ &= -\sum_{i=3}^{m} \sigma_i \langle F_v(e_i)\,,\,e_i\rangle - \langle F_v(v)\,,\,v\rangle \\ &= -\operatorname{tr} F_v\,,\end{aligned}$$

where, in the penultimate line, we are free to add the last term, since it is zero. □

A.2 Semi-Riemannian submanifolds

A.2.1 Preliminary concepts

Definition 47. A manifold N is a **submanifold** of M if the following conditions hold:

1. N is a topological subspace of M; and
2. the inclusion map $i : N \to M$ is smooth and its differential is injective at every point of N.

If M and N are semi-Riemannian manifolds, N is said to be a **semi-Riemannian submanifold** of M if its metric coincides with the restriction of that of M.

Note that, in the definition of a semi-Riemannian submanifold, requiring that N be a semi-Riemannian manifold beforehand is meaningful. If one simply takes N as a submanifold of M and attempts to endow it with a metric tensor by restricting that of M, the result may be a degenerate metric, which does not make N a semi-Riemannian manifold. A trivial illustration of this fact is a light ray in Minkowski space.

The following proposition will be used in the proofs of some of the results that follow:

Proposition 48. *Let N be a submanifold of M. If a vector field $X \in \mathfrak{X}(M)$ is tangent to N (i.e., if $X(p) \in T_pN$ for each $p \in N$), then $X|_N \in \mathfrak{X}(N)$. Furthermore, if Y*

is another vector field tangent to N,

$$[X\,,\,Y]|_N = [\,X|_N\,,\,Y|_N]\,.$$

If N is a submanifold of M, vector fields on the inclusion map $i : N \longrightarrow M$ are called **M-vector fields** on N and denoted $\overline{\mathfrak{X}}(N)$. In particular, the restriction of any vector field on M to N is in $\overline{\mathfrak{X}}(N)$.

If N is now a *semi-Riemannian* submanifold of M, each T_pN is a nondegenerate subspace of T_pM, and hence

$$T_pM = T_pN \oplus \big(T_pN\big)^\perp\,,$$

where $\big(T_pN\big)^\perp$ is also nondegenerate. If n and m are the dimensions of N and M, $\dim\big(T_pN\big)^\perp = m - n$ is called the **codimension** of N in M. The semi-Riemannian index of $\big(T_pN\big)^\perp$ is called the **coindex** of N and

$$\operatorname{ind} M = \operatorname{ind} N + \operatorname{coind} N\,.$$

The elements of $\big(T_pN\big)^\perp$ are said to be **normal** to N, whereas those in T_pN are **tangent**. Besides, for every $p \in N$ and every $v \in T_pM$, there exists a unique decomposition

$$v = v^\top + v^\perp\,,$$

where $v^\top \in T_pN$ and $v^\perp \in \big(T_pN\big)^\perp$. If applied pointwise to a vector field $X \in \overline{\mathfrak{X}}(N)$, this decomposition yields two new vector fields: one everywhere tangent to N, denoted $X^\top$, and one everywhere normal to N, denoted $X^\perp$. The set of normal vector fields to N is denoted $\mathfrak{X}^\perp(N)$. The tangent and normal vector fields to N are both submodules of $\overline{\mathfrak{X}}(N)$, and

$$\overline{\mathfrak{X}}(N) = \mathfrak{X}(N) \oplus \mathfrak{X}^\perp(N)\,.$$

A.2.2 The induced connection

In this section, we will consider N a semi-Riemannian submanifold of M, and denote the Levi-Civita connections of N by ∇ and of M by $\overline{\nabla}$. If $X \in \mathfrak{X}(N)$ and $Y \in \overline{\mathfrak{X}}(N)$, it is not immediately possible to ascribe meaning to $\overline{\nabla}_X Y$, since X and Y are not vector fields defined on an open subset of M. However, for $p \in N$, X and Y can be extended to vector fields $\overline{X}$ and $\overline{Y}$ defined on a neighbourhood U of p in M. Then, one could tentatively define

$$\overline{\nabla}_X Y = \left(\overline{\nabla}_{\overline{X}} \overline{Y}\right)\Big|_{U \cap N} .$$

We will show below that this object is indeed well-defined, i.e., does not depend on the particular extensions of X and Y used. The result is also a vector field in $\overline{X}(N)$. This operation is called the **induced connection** on N.

Proposition 49. $\overline{\nabla}_X Y$ *is a well-defined smooth M-vector field on N.*

Proof. Being the restriction of a smooth vector field on M to N, the induced connection is a smooth M-vector field on N. All we need to show is that the choice of the extensions $\overline{X}$ and $\overline{Y}$ does not influence the final result. Let U be a coordinate neighbourhood of M which intercepts N, and let the extension of Y be given by $\overline{Y} = f^i \partial_i$ in that basis. Then,

$$\overline{\nabla}_{\overline{X}} \overline{Y} = \overline{X}\left[f^i\right] \partial_i + f^i \overline{\nabla}_{\overline{X}} \partial_i .$$

The value that the second term takes at $p \in U \cap N$ only depends on the pointwise values of $\overline{X}$ and the functions f^i. Therefore, its value is the same regardless of the extensions $\overline{X}$ and $\overline{Y}$ chosen. As for the first term, if $\overline{X}$ correctly extends X, then

$$\overline{X}\left[f^i\right]\Big|_{U \cap N} = X\left[f^i\big|_{U \cap N}\right] ,$$

which is also completely determined by the values of X and Y on N. □

The induced connection inherits a series of properties from the connection of M:

Proposition 50. *Let $\overline{\nabla}$ denote the induced connection on $N \subset M$, and let $X, Y \in \mathfrak{X}(N)$ and $Z, W \in \overline{\mathfrak{X}}(N)$.*

1. *$\overline{\nabla}_X Z$ is $\mathcal{C}^\infty(N)$-linear in X.*
2. *$\overline{\nabla}_X Z$ is $\mathbb{R}$-linear in Z.*
3. *$\overline{\nabla}_X fZ = X[f]Z + f\overline{\nabla}_X Z$, for all $f \in \mathcal{C}^\infty(N)$.*
4. *$\overline{\nabla}_X Y - \overline{\nabla}_Y X - [X\,,\, Y] = 0$.*
5. *$X[\langle Z\,,\, W\rangle] = \left\langle \overline{\nabla}_X Z\,,\, W\right\rangle - \left\langle Z\,,\, \overline{\nabla}_X W\right\rangle$.*

It is a crucially important fact that, when X and Y are both in $\mathfrak{X}(N)$, $\overline{\nabla}_X Y$ is not in general also tangent to N. In particular, $\overline{\nabla}$ does not reduce to the intrinsic connection of N.

Proposition 51. *If $X, Y \in \mathfrak{X}(N)$, then $\nabla_X Y = \left(\overline{\nabla}_X Y\right)^\top$.*

Proof. Take a third vector field $Z \in \mathfrak{X}(N)$ and let $\overline{X}$, $\overline{Y}$ and $\overline{Z}$ be extensions of X, Y and Z to M. Then, by the Koszul formula,

$$\begin{aligned} 2\left\langle \overline{\nabla}_{\overline{X}}\overline{Y}\,,\, \overline{Z}\right\rangle = {} & \overline{X}\left[\left\langle \overline{Y}\,,\, \overline{Z}\right\rangle\right] + \overline{Y}\left[\left\langle \overline{X}\,,\, \overline{Z}\right\rangle\right] - \overline{Z}\left[\left\langle \overline{X}\,,\, \overline{Y}\right\rangle\right] - \\ & - \left\langle \overline{X}\,,\, \left[\overline{Y}\,,\, \overline{Z}\right]\right\rangle - \left\langle \overline{Y}\,,\, \left[\overline{X}\,,\, \overline{Z}\right]\right\rangle + \left\langle \overline{Z}\,,\, \left[\overline{X}\,,\, \overline{Y}\right]\right\rangle. \end{aligned}$$

By definition, the left hand side restricts to $2\left\langle \overline{\nabla}_X Y\,,\, Z\right\rangle$ on N. Meanwhile, the right hand side reduces to the same expression in terms of X, Y and Z. By the Koszul formula for the Levi-Civita connection of N, that expression equals $2\langle \nabla_X Y\,,\, Z\rangle$. Therefore, we have

$$2\left\langle \overline{\nabla}_X Y\,,\, Z\right\rangle = 2\langle \nabla_X Y\,,\, Z\rangle\,,$$

for all $Z \in \mathfrak{X}(N)$. This implies $\left(\overline{\nabla}_X Y\right)^\top = \nabla_X Y$. □

The *normal* component of $\overline{\nabla}_X Y$ for X and Y tangent to N also has several interesting properties. It is called the **second fundamental form** or **shape tensor** of N and defined using the following expression:

$$\overline{\nabla}_X Y = \nabla_X Y + \mathrm{II}(X, Y)\,. \tag{A.2}$$

The proposition below establishes its most basic properties:

Proposition 52. *The function* $\mathrm{II} : \mathfrak{X}(N) \times \mathfrak{X}(N) \to \mathfrak{X}^{\perp}(N)$ *defined by* (A.2) *is* $\mathcal{C}^{\infty}(N)$*-bilinear and symmetric.*

Proof. The expression $\mathrm{II}(X, Y)$ is immediately $\mathcal{C}^{\infty}(N)$-linear in X, since it is just the normal component of $\overline{\nabla}_X Y$. We will now prove that $\mathrm{II}(X, Y)$ is symmetric, and then $\mathcal{C}^{\infty}(N)$-linearity in Y will follow. Indeed,

$$\mathrm{II}(X, Y) - \mathrm{II}(Y, X) = \left(\overline{\nabla}_X Y - \overline{\nabla}_Y X\right)^{\perp} = [X\,,\,Y]^{\perp} = 0\,,$$

where we have used item 4 from Proposition 50 and the fact that the Lie bracket of two vector fields in $\mathfrak{X}(N)$ is also in $\mathfrak{X}(N)$. ◻

Once again, it follows from properties of $\overline{\nabla}$ that $\mathrm{II}(X, Y)$ at a point $p \in N$ depends on X only through its pointwise value $X(p)$. By symmetry, the same holds for Y. Therefore, II smoothly assigns to each $p \in N$ a bilinear transformation $T_pN \times T_pN \to \left(T_pN\right)^{\perp}$.

Contraction of II yields a normal vector field on N, denoted by H and known as the **mean curvature vector field** of $N \subset M$:

$$H(p) = \frac{1}{n}\sum_{i=1}^{n} \sigma_i\, \mathrm{II}(e_i, e_i)\,,$$

where $\{e_1, \ldots, e_n\}$ is a basis for T_pN and $\sigma_i = \frac{1}{\langle e_i\,,\,e_i\rangle}$.

A.2.3 The normal connection

The **normal connection** of a sub-Riemannian manifold $N \subset M$ is defined in a similar way to the induced connection. However, instead of taking the tangential component of $\overline{\nabla}_X Y$ for X and Y two tangential vector fields on N, we now take the *normal* component of $\overline{\nabla}_X Z$ for X a tangential vector field and Z a normal vector field on N:

$$\nabla^{\perp}_X Z := \left(\overline{\nabla}_X Z\right)^{\perp} \qquad \left(\forall X \in \mathfrak{X}(N)\,,\ \forall Z \in \mathfrak{X}^{\perp}(N)\right).$$

The normal connection has the following elementary properties:

1. $\nabla^{\perp}_X Z$ is $\mathcal{C}^{\infty}(N)$-linear in X and $\mathbb{R}$-linear in Z.
2. $\nabla^{\perp}_X fZ = X[f]Z + f\nabla^{\perp}_X Z$, for all $f \in \mathcal{C}^{\infty}(N)$.
3. $X[\langle Z\,,\,W\rangle] = \left\langle\nabla^{\perp}_X Z\,,\,W\right\rangle + \left\langle Z\,,\,\nabla^{\perp}_X W\right\rangle$, for all $X \in \mathfrak{X}(N)$ and $Z, W \in \mathfrak{X}^{\perp}(N)$.

Consider a curve $\gamma : I \to N$ and let Z be a vector field along γ normal to N. We say that Z is **normal parallel** along γ if $\nabla^{\perp}_{\gamma'} Z$ vanishes identically. The concept of **normal parallel transport** is analogous to that of parallel transport for the usual connection, and is established by the following proposition:

Proposition 53. *Let $p \in N$, $\gamma : (-\varepsilon, \varepsilon) \to N$ a curve such that $\gamma(0) = p$ and $z \in \left(T_p N\right)^{\perp}$. Then, there exists a unique vector field Z normal parallel along γ such that $Z(0) = z$.*

This result follows from existence and uniqueness theorems for linear ordinary differential equations by noticing that the map

$$\nabla^{\perp}_{\gamma'} : I \times \mathfrak{X}^{\perp}(N) \longrightarrow \mathfrak{X}^{\perp}(N)$$

which appears in the constraint $\nabla^{\perp}_{\gamma'} Z = 0$ is linear in Z (note that it is not $\mathcal{C}^{\infty}(N)$-linear, but it *is* $\mathbb{R}$-linear).

Much like the second fundamental form $\mathrm{II}(X, Y)$ is defined as the normal part of the induced connection for two tangent vector fields X and Y, we define an object $\widetilde{\mathrm{II}}(X, Z)$ as the *tangent* component of $\overline{\nabla}_X Z$ for any X tangent and Z normal to N. $\widetilde{\mathrm{II}}$ is also $\mathcal{C}^{\infty}(N)$-bilinear and depends only pointwise on X and Z.

In summary, we have

$$\begin{aligned}
\overline{\nabla}_X Y &= \nabla_X Y + \mathrm{II}(X, Y) \\
\overline{\nabla}_X Z &= \widetilde{\mathrm{II}}(X, Z) + \nabla^{\perp}_X Z\,,
\end{aligned}$$

where, in each equation, the first term in the right hand side is tangent to N, and the second is normal. II and $\widetilde{\mathrm{II}}$ are also related by the following equation:

$$\langle \mathrm{II}(X,Y)\,,\,Z\rangle = -\left\langle X\,,\,\widetilde{\mathrm{II}}(Y,Z)\right\rangle,$$

for all $X, Y \in \mathfrak{X}(N)$ and all $Z \in \mathfrak{X}^{\perp}(N)$. Indeed, since $\langle Y\,,\,Z\rangle = 0$ everywhere,

$$\begin{aligned} X[\langle Y\,,\,Z\rangle] &= \left\langle \overline{\nabla}_X Y\,,\,Z\right\rangle + \left\langle Y\,,\,\overline{\nabla}_X Z\right\rangle \\ 0 &= \left\langle \left(\overline{\nabla}_X Y\right)^{\perp}\,,\,Z\right\rangle + \left\langle Y\,,\,\left(\overline{\nabla}_X Z\right)^{\top}\right\rangle \\ 0 &= \langle \mathrm{II}(X,Y)\,,\,Z\rangle + \left\langle Y\,,\,\widetilde{\mathrm{II}}(X,Z)\right\rangle \end{aligned}$$

(the form written previously then follows by symmetry of II).

Appendix B

Completion of the proofs of Theorems 18 and 26

Recall that in the proofs of Theorems 18 and 26 there was left pending the point of the convergence of the sequence $\left\{H_\gamma^\perp(X_n, X_n)\right\}$ to $H_\gamma^\perp(X, X)$, for a certain family of vector fields X_n on γ which converged pointwise to X. In this Appendix, we shall clear up that point, thus completing both proofs. For definiteness, we shall work within the context of Theorem 18, but the procedure is exactly the same for Theorem 26.

The following corollary of Taylor's theorem shall play a fundamental role in the proof:

Proposition 54. *Let $f : [a, b] \to \mathbb{R}$ be a smooth function and suppose $f^{(0)}(c) = \ldots = f^{(N)}(c) = 0$, for some $c \in [a, b]$. Then, for every $x \in [a, b]$,*

$$f(x) = (x - c)^{N+1} g(x) ,$$

where g is another smooth function on $[a, b]$.

Proof. Take the Taylor expansion of f around c to order N, which is legitimate since f is smooth:

$$f(x) = \sum_{n=0}^{N} \frac{f^{(n)}(c)(x - c)^n}{n!} + E_N(x) ,$$

where the error $E_N(x)$ is given by

$$E_N(x) = \frac{1}{N!} \int_c^x (x-t)^N f^{(N+1)}(t)\,\mathrm{d}t\,.$$ [1]

Note that, by the hypotheses, the Taylor polynomial vanishes, and we are left simply with $f(x) = E_N(x)$. By introducing the substitution $t = (x-c)s + c$ into the error formula, we find the alternate form

$$E_N(x) = \frac{(x-c)^{N+1}}{N!} \int_0^1 (1-s)^N f^{(N+1)}\big((x-c)s + c\big)\,\mathrm{d}s\,,$$

and, since f is smooth, the integral term is infinitely differentiable with respect to x. □

Recall that we had defined the sequence $\{X_n\}$ via

$$X_n(u) = \begin{cases} X(u)\,, & \text{for } u \in [a, r_n] \\ \left(1 - 2\dfrac{u - r_n}{b - r_n}\right) Y_n(u)\,, & \text{for } u \in \left(r_n, \frac{b+r_n}{2}\right) \\ 0\,, & \text{for } u \in \left[\frac{b+r_n}{2}, b\right] \end{cases}\,,$$

where $r_n = b - 1/n$, Y_n is the parallel transport of $X(r_n)$ along γ, and the sequence is taken to start from n large enough that all the r_n lie within the final smooth component of the (piecewise smooth) vector field X. Thus, we have, for the covariant derivative of X_n along γ,

$$X_n'(u) = \begin{cases} X'(u)\,, & \text{for } u \in [a, r_n) \\ -2nY_n(u)\,, & \text{for } u \in \left(r_n, \frac{b+r_n}{2}\right) \\ 0\,, & \text{for } u \in \left(\frac{b+r_n}{2}, b\right] \end{cases}\,.$$

[1]See, e.g., [37], chapter 7.

Note that X_n' is not defined at the breaks r_n and $\frac{b+r_n}{2}$. We can then calculate

$$H_\gamma^\perp(X_n, X_n) = \int_a^{r_n} [\langle X', X'\rangle + \langle R(X,\gamma')X, \gamma'\rangle] \, \mathrm{d}u + \\ + \int_{r_n}^{\frac{b+r_n}{2}} 4n^2 \langle Y_n, Y_n\rangle \, \mathrm{d}u + \\ + \int_{r_n}^{\frac{b+r_n}{2}} [1 - 2n(u - r_n)]^2 \langle R(Y_n,\gamma')Y_n, \gamma'\rangle \, \mathrm{d}u \, .$$

Since $r_n \to b$, it is clear that the first term converges to $H_\gamma^\perp(X, X)$. We now show that the other terms tend to 0. Because Y_n is parallel, the integrand in the second term is constant and equal to $4n^2\langle X(r_n), X(r_n)\rangle$. Besides, the fact that $X(b) = 0$ implies that the function $\langle X, X\rangle$ vanishes with derivative 0 at the same point:

$$\langle X, X\rangle'(b) = 2\langle X(b), X'(b)\rangle = 0 \, .$$

By Proposition 54 above, this observation implies that $\langle X, X\rangle$ can be written, in the last smooth component of X, as $(b - u)^2 g(u)$, where g is another smooth function. Therefore, in particular,

$$\langle X(r_n), X(r_n)\rangle \leq (b - r_n)^2 G \, ,$$

where G denotes the supremum of $|g|$ in the final smooth component of X, and hence

$$\left| \int_{r_n}^{\frac{b+r_n}{2}} 4n^2 \langle Y_n, Y_n\rangle \, \mathrm{d}u \right| \leq \frac{1}{2n} \cdot 4n^2 \cdot \frac{G}{n^2} \, ,$$

which goes to zero as n tends to infinity.

Now we turn to the third term. Using the definition of sectional curvature and the fact that Y_n is perpendicular to γ, we can write

$$\langle R(Y_n, \gamma')Y_n\,,\, \gamma'\rangle = c\langle Y_n\,,\, Y_n\rangle K(Y_n, \gamma')\,,$$

where $c = \langle \gamma'\,,\, \gamma'\rangle$. Once again, $\langle Y_n\,,\, Y_n\rangle$ is constant and its module bounded by G/n^2. The sectional curvature K is also bounded at each point along the curve, and hence we can find a positive real number M such that $|K(Y_n, \gamma')| < M$ for each $u \in \left[r_n, \frac{b+r_n}{2}\right]$ and each n. Therefore,

$$\left| \int_{r_n}^{\frac{b+r_n}{2}} [1 - 2n(u - r_n)]^2 \langle R(Y_n, \gamma')Y_n\,,\, \gamma'\rangle \,\mathrm{d}u \right| \leq \frac{MGc}{n^2} \left| \int_{r_n}^{\frac{b+r_n}{2}} [1 - 2n(u - r_n)]^2 \,\mathrm{d}u \right|.$$

A simple integration by substitution shows that the integral on the right hand side is $1/6n$, which allows us to conclude that this term also goes to zero as $n \longrightarrow +\infty$.

Bibliography

1. Penrose, R. Gravitational Collapse and Space-Time Singularities. *Phys. Rev. Lett.* **14,** 57–59 (3 1965).
2. Hawking, S. W. The occurrence of singularities in cosmology. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **294,** 511–521 (1966).
3. Hawking, S. W. The occurrence of singularities in cosmology. II. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **295,** 490–493 (1966).
4. Hawking, S. W. The Occurrence of Singularities in Cosmology. III. Causality and Singularities. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **300,** 187–201 (1967).
5. Hawking, S. W. & Penrose, R. The Singularities of Gravitational Collapse and Cosmology. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **314,** 529–548 (1970).
6. Senovilla, J. M. M. & Garfinkle, D. The 1965 Penrose singularity theorem. *Classical and Quantum Gravity* **32,** 124008 (2015).
7. Senovilla, J. M. M. Singularity Theorems and Their Consequences. *General Relativity and Gravitation* **30,** 701–848 (1998).
8. Fewster, C. J. *Lectures on quantum energy inequalities* 2012. arXiv: 1208.5399 [gr-qc].
9. Epstein, H., Glaser, V. & Jaffe, A. Nonpositivity of energy density in Quantized field theories. *Nuovo Cim.* **36,** 1016 (1965).

10. Ford, L. H. Quantum Coherence Effects and the Second Law of Thermodynamics. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **364,** 227–236 (1978).

11. Dawson, S. P. A quantum weak energy inequality for the Dirac field in two-dimensional flat spacetime. *Classical and Quantum Gravity* **23,** 287–293 (2005).

12. Dawson, S. P. & Fewster, C. J. An explicit quantum weak energy inequality for Dirac fields in curved spacetimes. *Classical and Quantum Gravity* **23,** 6659–6681 (2006).

13. Pfenning, M. J. Quantum inequalities for the electromagnetic field. *Physical Review D* **65** (2001).

14. Fewster, C. J. & Teo, E. Bounds on negative energy densities in static spacetimes. *Physical Review D* **59** (1999).

15. Fewster, C. J. A general worldline quantum inequality. *Classical and Quantum Gravity* **17,** 1897–1911 (2000).

16. Fewster, C. J. & Verch, R. A Quantum Weak Energy Inequality for Dirac Fields in Curved Spacetime. *Communications in Mathematical Physics* **225,** 331–359 (2002).

17. Fewster, C. J. & Mistry, B. Quantum weak energy inequalities for the Dirac field in flat spacetime. *Physical Review D* **68** (2003).

18. Fewster, C. J. & Pfenning, M. J. A quantum weak energy inequality for spin-one fields in curved space–time. *Journal of Mathematical Physics* **44,** 4480 (2003).

19. Fewster, C. J. & Hollands, S. Quantum Energy Inequalities In Two-Dimensional Conformal Field Theory. *Reviews in Mathematical Physics* **17,** 577–612 (2005).

20. Fewster, C. J. & Kontou, E.-A. Quantum strong energy inequalities. *Physical Review D* **99** (2019).

21. Fewster, C. J. & Osterbrink, L. W. Quantum energy inequalities for the non-minimally coupled scalar field. *Journal of Physics A: Mathematical and Theoretical* **41,** 025402 (2007).

22. Smith, C. J. An absolute quantum energy inequality for the Dirac field in curved spacetime. *Classical and Quantum Gravity* **24,** 4733–4750 (2007).

23. Yu, H. & Wu, P. Quantum inequalities for the free Rarita-Schwinger fields in flat spacetime. *Physical Review D* **69** (2004).

24. Fewster, C. J. in *Wormholes, Warp Drives and Energy Conditions* (ed Lobo, F. S. N.) 215–254 (Springer International Publishing, Cham, 2017).

25. Tipler, F. J. Energy conditions and spacetime singularities. *Phys. Rev. D* **17,** 2521–2528 (10 1978).

26. Tipler, F. J. General relativity and conjugate ordinary differential equations. *Journal of Differential Equations* **30,** 165–174 (1978).

27. Roman, T. A. Quantum stress-energy tensors and the weak energy condition. *Phys. Rev. D* **33,** 3526–3533 (12 1986).

28. Roman, T. A. On the "averaged weak energy condition" and Penrose's singularity theorem. *Phys. Rev. D* **37,** 546–548 (2 1988).

29. Wald, R. & Yurtsever, U. General proof of the averaged null energy condition for a massless scalar field in two-dimensional curved spacetime. *Phys. Rev. D* **44,** 403–416 (2 1991).

30. Fewster, C. J. & Galloway, G. J. Singularity theorems from weakened energy conditions. *Classical and Quantum Gravity* **28,** 125009 (2011).

31. Fewster, C. J. & Kontou, E.-A. A new derivation of singularity theorems with weakened energy hypotheses. *Classical and Quantum Gravity* **37,** 065010 (2020).

32. O'Neill, B. *Semi-Riemannian Geometry With Applications to Relativity* (Elsevier Science, 1983).

33. Misner, C., Thorne, K., Wheeler, J. & Kaiser, D. *Gravitation* (Princeton University Press, 2017).

34. Godinho, L. & Natário, J. *An Introduction to Riemannian Geometry: With Applications to Mechanics and Relativity* (Springer International Publishing, 2014).
35. Wald, R. *General Relativity* (University of Chicago Press, 1984).
36. Brown, P. J., Fewster, C. J. & Kontou, E.-A. A singularity theorem for Einstein–Klein–Gordon theory. *General Relativity and Gravitation* **50** (2018).
37. Apostol, T. M. *Calculus, Volume I: One-Variable Calculus, with an Introduction to Linear Algebra* 2nd ed. (John Wiley & Sons, 1967).