Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Dipartimento di Fisica e Astronomia
Corso di Laurea in Fisica

A Geometrical View on Space-Time Singularities

Relatore:
Prof. Alexandr Kamenchtchik

Presentata da:
Simone Coli

Anno Accademico 2022/2023

A Geometrical View on Space-Time Singularities

An Analysis on the Singularity Theorems

Simone Coli

Abstract

At the beginning of the development of General Relativity, from specific solutions of the Einstein field equations arose the existence of singular points inside the defined space-time, then called singularities. The existence of these objects seemed to be due to the presence of particular symmetries inside a solution. It can be shown that it is possible to prove that singularities are not the result of some peculiar solution. Such a proof is not trivial and requires the use of geometrical considerations. In this thesis, I shall introduce a series of results and definitions, allowing us to state and prove a series of theorems that predict the existence of singularities by showing that a particular space-time possesses some kind of incompleteness. These theorems are called the singularity theorems. In addition to that, I shall also include a series of considerations about the conditions that such theorems require, as well as a brief discussion about their consequences in physics.

Contents

Preface

This thesis has the goal of introducing the concept of space-time singularity from a strictly geometrical point of view, starting from the definition of the necessary mathematical tools. The majority of such tools, together with the notions and concepts used in the development of the theory had been inspired by the rather complete and exhaustive text "*The large scale structure of space-time*" by S.W. Hawking and G.F.R. Ellis, which had been recommended to me by Professor Alexandr Kamenchtchik, whom I would like to thank for putting up with me during the drafting of this document. In addition to such a book, I used some other minor references, which will be all listed at the end of the document. I would also like to thank Professor Roberto Casadio for the wonderful course *Basics of General Relativity* which had been very helpful to me in better understanding the basics of differential geometry as well as a few concepts of General Relativity.

After a brief introduction to the basic principles of the theory of General Relativity in Chapter 1, that is the theory of gravity from which we will derive a variety of useful results, we shall give a series of definitions and results necessary to the introduction and discuss the singularity theorems. In Chapter 2 I will introduce the weak and strong energy conditions as well as some consideration on the shape of the energy-momentum tensor. In the same chapter, I shall also introduce the concept of conjugate point, as well as the concept of arc length, both of which will turn out to be useful in Chapter 4. Chapter 3 will contain some considerations about causality, starting from the concept of orientability for space-time and then moving to some properties of the curves in an orientable space-time. I will then present the concept of achronal boundaries, Cauchy development, and global hyperbolicity, stating the causality conditions, defining the causal boundary of space-time, and concluding with the existence of geodesics and the definition of asymptotically simple spaces. The last chapter will finally talk about singularities, stating their definition and introducing the four singularity theorems provided with proofs and explanations about their conditions.

The reason I decided to choose this subject for my thesis has to be attributed to my interest in black holes from the beginning of my bachelor's degree. With the work I have done writing this document, I believe to have made one step forward toward a better understanding of the concepts behind this mysterious, but yet fascinating objects.

Chapter 1

Introduction to General Relativity

1.1 A mathematical model for space-time

The essence of a physical theory expressed in mathematical form is to identify, through mathematical concepts, physically measureable quantities.

Let us, therefore, begin considering for our model of space-time a pair $(\mathcal{M}, \mathbf{g})$ where $\mathcal{M}$ is a four-dimensional C^∞ Housdorf manifold and $\mathbf{g}$ is a metric with sg(+2) on $\mathcal{M}$. In general, two models $(\mathcal{M}, \mathbf{g})$ and $(\mathcal{M}', \mathbf{g}')$ will be taken as equivalent if they are isometric, that is if there exists a diffeomorphism $\theta : \mathcal{M} \to \mathcal{M}'$ which carries the metric $\mathbf{g}$ into $\mathbf{g}'$, *i.e.* $\theta_* \mathbf{g} = \mathbf{g}'$. This means that $\{(\mathcal{M}, \mathbf{g})\}$ is an equivalence class of pairs. For the sake of simplicity however, we shall consider only $(\mathcal{M}, \mathbf{g})$ as the representative member of such a class.

Choosing a manifold to represent space-time seems to be a natural option since it already contains the concept of continuity. Such an idea has been proven to work experimentally for distances down to 10^{-15} cm, using the scattering of Pions. Further confirmations at lower scales may be difficult, as it would require a particle of such energy that several other particles might be created during the measuring process, influencing the outcome of the observation. Therefore, a manifold could become an inappropriate model for space-time within lengths smaller than 10^{-35} cm, requiring the use of a theory which represents space-time using an alternative structure. However, such a breakdown would not be expected to affect General Relativity until the gravitational scale would become of that order. This would require extreme conditions, such as densities up to $10^{56}\,\mathrm{g\,cm^{-3}}$, which are completely beyond our present knowledge.

Nevertheless, by adopting a manifold model for space-time, and making some reasonable assumptions, it will be shown in Chapter 4 that some breakdowns of General Relativity must occur.

The previously defined metric, $\mathbf{g}$, allows non-zero vectors $\mathbf{X} \in \mathbf{T}_p$ in $p \in \mathcal{M}$ to be categorized depending on the result of their scalar product as: *space-like*, *time-like* or *null* according to whether $g(\mathbf{X}, \mathbf{X})$ is positive, negative or zero, respectively. The order of differentiability of $\mathbf{g}$, r, is such that allows the definition of the field equation.

Inside our manifold, $\mathcal{M}$, it is possible to define a number of fields, such as the electromagnetic, the neutrino's, *etc.*., which describes the matter content of space-time. These fields will obey relations and expression of tensorial nature on $\mathcal{M}$ in which spacial derivatives are defined to be covariant derivatives with respect to the symmetric connection given by the metric. This is because every connection on $\mathcal{M}$ different from the one defined by the metric, might be related to the latter through a tensor, which could be itself regarded as another field on $\mathcal{M}$. The theory one obtains depends on what fields one incorporates.

From now on, we shall denote the matter fields included in the theory with $\Psi_{(i)}^{a...b}{}_{c...d}$ where (i) is the index of the field. [Hawking and Ellis, 1973, pp. 56-59]

1.2 Principles

Now that the mathematical structures are defined, we shall deduct and discuss a series of principles which will lay down the basis for all further considerations.

> *(i) Space-time (which is the set of all events) is a four-dimensional manifold with a metric.*

> *(ii) The metric is measurable defining a coordinate system, such that the spacial distance between two nearby points is $|g(d\mathbf{x}, d\mathbf{x})|^{1/2}$, and the time distance between two events closely separated is $|-g(d\mathbf{x}, d\mathbf{x})|^{1/2}$.*

Nevertheless, the definition of the four-vector dot product $g(d\mathbf{x}, d\mathbf{x}) \doteq -(dx^0)^2 + (dx^1) + (dx^2)^2 + (dx^3)^2$, from Special Relativity, is not true everywhere in a general coordinate system. Still, it can be argued that such a frame exists locally. This assumption seems to suggest a curved manifold as a model of space-time, in which it is possible to find some particular coordinates, whose dot product has a similar form to the Minkowski dot product. Therefore:

> *(iii) The metric of space-time can be put in the canonical form $\eta_{\mu\nu}$ (Lorenz form) in any particular point by an appropriate choice of coordinates.*

Considering, now, the Newtonian approximation, which allows the existence of a locally inertial frame where particles will behave as in free fall, we obtain that, in a gravitational field, the acceleration of a particle is independent from its mass. The paths that these particles follow in this frame will take the shape of straight lines. This trajectory is, by definition, a geodesic in a curved manifold, therefore, a particle in free fall must move on a time-like geodesic through space-time. A more general way to express this statement would be the *strong equivalence principle*:

> *(iv) Any physical law which can be defined in tensorial notation in Special Relativity has the same form in a locally inertial frame of reference on a curve space-time.*

Given for example the conservation laws in Special Relativity:

$$(\rho U^{\alpha})_{,\alpha} = 0 \ , \tag{1.1}$$

where ρ is the density of particles and U^{α} is four-velocity, the strong equivalence principle implies that it can be regarded as a conservation law in a locally inertial frame of reference. In any coordinate system Eq. (1.1) becomes:

$$(\rho U^{\alpha})_{;\alpha} = 0 \ ,$$

in which the partial derivative had been substituted with the covariant derivative. Schutz [1985]

Let us now define the energy-momentum tensor $T^{\mu\nu}$ as a $(2,0)$ type tensor describing the distribution and dynamics of matter in space-time, whose (μ,ν) component given by the flux of μ^{th} of the four-momentum vector across a surface of constant coordinates x^{ν}. This serves in general relativity a similar role to the mass distribution in Newtonian physics. In particular, T^{00} is the flow of relativistic mass through time, defined as the *spatial mass density*; $T^{\mu 0}$ is the flow of the μ^{th} component of spatial momentum through time, defined as the *spacial μ–momentum density*; $T^{0\nu}$ is the flow of relativistic mass through a surface of constant coordinate x^{ν}, defined as *ν-flux of mass*; the rest of the components are the mechanical stresses. Tolish [2010] Since the components of momentum must be conserved, the conservation law takes the form:

$$(T^{\mu\nu})_{;\mu} = 0 \ . \tag{1.2}$$

Another way to define such a tensor is using the concept of action function. Given the Lagrangian of the matter field, in fact, we shall define the energy-momentum tensor as the action of the latter, that is:

$$T_{\mu\nu} = -\frac{\Delta g}{g}\mathcal{L}_m - 2\Delta\mathcal{L}_m,$$

in which, as we shall see in the following section, $\Delta \doteq \delta/\delta g_{\mu\nu}$

1.3 Einstein field equations

The principles deduced in the previous section give us a hint on the nature of the matter field as well as on the equation that this field has to obey. In particular, the two main conditions on the behavior of the latter, are the *strong equivalence principle*, generalized also to light signals, thus to *non-spacelike* curves, and the local conservation of the energy-momentum tensor.

It is possible to show that the shape of the equation that all such fields have to follow takes the form of Einstein's field equation:

$$(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}) + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \ , \tag{1.3}$$

that is a system of non-linear partial differential equation which expressed in the tensorial formalism takes a more compact form.

An alternative form uses the energy-momentum scalar instead of the Ricci scalar $8\pi T = R$ (such relation between the two scalars has been obtained diagonalizing the previous equation).

$$R^{\mu\nu} = 8\pi(T^{\mu\nu} - \frac{1}{2}Tg^{\mu\nu}) + \Lambda g^{\mu\nu} \ . \tag{1.4}$$

Proof. Let $I = I_c + I_m$ be the Einstein-Hilbert action function, in which I_m depends only on the matter component of the field, while I_c only on the curvature one. Let us define I_c such that:

$$I_c = \int_V \mathcal{L}_c d\nu = \int_V \mathcal{L}_c \sqrt{-g} d^4x \ , \tag{1.5}$$

such that $d\nu = 4!^{-1}\eta$ in which η is a 4-form. Being a scalar, we can take a guess at the Lagrangian in the simplest form possible, choosing to define it using the Ricci scalar R. With the purpose of taking also into account the expansion of the universe, we shall also include a scalar known as the cosmological constant Λ:

$$\mathcal{L}_c = \alpha(R - 2\Lambda) \ , \tag{1.6}$$

which gives to the action the following form:

$$I = \int_V (\alpha(R - 2\Lambda) + \mathcal{L}_m) d\nu \ , \tag{1.7}$$

where $\mathcal{L}_m$ is the Lagrangian of the matter field. Let us now apply the principle of least action on it. Using the notation Δ to indicate $\delta/\delta g_{\mu\nu}$ we specify that the variation is not applied on the parameters of the field, but rather on the components of the metric $g_{\mu\nu}$.

$$\delta I = \int_V \delta((\alpha(R - 2\Lambda) + \mathcal{L}_m)d\nu) \ . \tag{1.8}$$

The variation of the curvature Lagrangian is such that:

$$\begin{aligned}\delta((R - 2\Lambda)d\nu) = ((R - 2\Lambda)\ &\frac{1}{2}g_{\mu\nu}\delta g^{\mu\nu} + \\ &+ R_{\mu\nu}\delta g^{\mu\nu} + \delta R_{\mu\nu}g^{\mu\nu})d\nu \ ,\end{aligned}$$

in which the Ricci scalar is defined as the contraction of the Ricci tensor, through the metric ($R = g^{\mu\nu}R_{\mu\nu}$), and $\frac{1}{2}g_{\mu\nu}\delta g^{\mu\nu}$ is the variation of $d\nu$. That is:

$$\delta(d\nu) = \delta(\frac{1}{4!}\eta_{\nu\mu\lambda\sigma}),$$

where $\eta_{\nu\mu\lambda\sigma} = (-g)^{\frac{1}{2}} 4! \delta^1_{[\mu}\delta^2_{\nu}\delta^3_{\lambda}\delta^4_{\sigma]}$.

$$\delta(\frac{1}{4!}\eta_{\nu\mu\lambda\sigma}) = -\frac{\delta g}{2\sqrt{-g}}\delta^1_{[\mu}\delta^2_{\nu}\delta^3_{\lambda}\delta^4_{\sigma]} = \frac{g}{2\sqrt{-g}}g_{\mu}\nu\delta^1_{[\mu}\delta^2_{\nu}\delta^3_{\lambda}\delta^4_{\sigma]}$$

from which we obtain $\delta(d\nu) = \frac{1}{2}g_{\mu\nu}\delta g_{\mu\nu}d\nu$.

Now by the definition of the Ricci tensor, let us express it as the contraction of the Riemann tensor

$$\begin{aligned}R_{\mu\nu} \doteq R^{\alpha}_{\mu\alpha\nu} &= \\ &= \partial_\alpha\Gamma^\alpha_{\alpha\mu} - \partial_\nu\Gamma^\alpha_{\alpha\nu} + \Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\nu\beta}\Gamma^\beta_{\alpha\mu} \ ,\end{aligned}$$

therefore the variation of $R_{\mu\nu}$ gives

$$\begin{aligned}\delta R_{\mu\nu} = \partial_\alpha(\delta\Gamma^\alpha_{\alpha\mu}) - \partial_\nu(\delta\Gamma^\alpha_{\alpha\nu}) + \delta\Gamma^\alpha_{\alpha\beta}\Gamma^\beta_{\mu\nu} - \delta\Gamma^\alpha_{\nu\beta}\Gamma^\beta_{\alpha\mu} + \\ + \Gamma^\alpha_{\alpha\beta}\delta\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\nu\beta}\delta\Gamma^\beta_{\alpha\mu} \ .\end{aligned}$$

From the definition of covariant derivative of a (1,3) tensor, $\nabla_\alpha(\delta\Gamma^\alpha_{\mu\nu}) = \partial_\alpha(\delta\Gamma^\alpha_{\mu\nu}) + \Gamma^\beta_{\beta\alpha}\delta\Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\mu\beta}\delta\Gamma^\beta_{\nu\alpha} - \Gamma^\beta_{\nu\alpha}\delta\Gamma^\alpha_{\mu\beta}$ it follows that the previous expression can be simplified as:

$$\delta R_{\mu\nu} = (\delta\Gamma^\alpha_{\mu\nu})_{;\alpha} - (\delta\Gamma^\alpha_{\mu\nu})_{;\nu} \ ,$$

and the variation of the Rocci scalar becomes:

$$\begin{aligned}\delta R = R_{\mu\nu}\delta g^{\mu\nu} + \delta R_{\mu\nu}g^{\mu\nu} &= \\ &= R_{\mu\nu}\delta g^{\mu\nu} + \delta\Gamma^\alpha_{\mu\nu;\alpha}g^{\mu\nu} - \delta\Gamma^\alpha_{\mu\nu;\nu}g^{\mu\nu} \ .\end{aligned}$$

Renaming the index ν with α and vice versa in the last term of the previous equation, allows to have a homogeneous covariant derivative with respect to α:

$$\delta R = R_{\mu\nu}\delta g^{\mu\nu} + (\delta\Gamma^{\alpha}_{\mu\nu}g^{\mu\nu} - \delta\Gamma^{\nu}_{\mu\nu}g^{\mu\alpha})_{;\alpha} \ .$$

Now, by defining the tensorial object in the round parenthesis as V^{α} gives, we find the final form for the variation of Ricci's scalar:

$$\delta R = R_{\mu\nu}\delta g^{\mu\nu} + V^{\alpha}_{;\alpha} \ . \tag{1.9}$$

Because the integral of the action is over the whole space-time, the quantities $\delta\Gamma^{\alpha}_{\mu\nu}$ vanish. The reason behind this fact is the annulment assumption of the variation of gravitational field at the boundaries of our space-time, implying $\delta R = R_{\mu\nu}\delta g^{\mu\nu}$.

Furthermore, from linear algebra, we know that $\det(e^{A}) = e^{\mathrm{tr}(A)}$, which means that $\ln(\det(A)) = \mathrm{tr}(\ln(A))$, where A is an arbitrary matrix. By taking the variation of both sides we produce the following relationships:

$$\delta \ln(\det A) = \delta \mathrm{tr}(\ln(A)) \ ,$$

$$\frac{1}{\det(A)}\delta \det A = \mathrm{tr}(A^{-1}\delta A) \ ,$$

$$\delta \det A = \mathrm{tr}(A^{-1}\delta A) \ \det A \ .$$

Let now $A = g_{\mu\nu}$ and $A^{-1} = g^{\mu\nu}$, then

$$\delta g = \mathrm{tr}(g^{\mu\nu}\delta g_{\alpha\beta}) \ g = g^{\mu\nu}\delta g_{\mu\nu} g \ . \tag{1.10}$$

Merging Eq. (1.9) with Eq. (1.10)

$$\delta((R-2\Lambda)d\nu) = ((\frac{1}{2}R - \Lambda)g^{\mu\nu} + R^{\mu\nu})\delta g_{\mu\nu} dv \ . \tag{1.11}$$

Considering now the variation of the matter Lagrangian we get:

$$\delta(\mathcal{L}_m dv) = (\Delta\mathcal{L}_m + \frac{1}{2}g_{\mu\nu}\delta g^{\mu\nu}\mathcal{L}_m)dv \ .$$

Since the energy-momentum tensor had been defined as the action function of the matter Lagrangian, it follows that $T_{\mu\nu} = -\frac{\Delta g}{g}\mathcal{L}_m - 2\Delta\mathcal{L}_m$. Therefore, giving that $\Delta g = -g g_{\mu\nu}$ the previous expression becomes:

$$\delta(\mathcal{L}_m dv) = -T_{\mu\nu}\delta g_{\mu\nu} dv \ . \tag{1.12}$$

From the principle of least action the integrand of Eq. (1.8) vanishes for all values of $\delta g_{\mu\nu}$, since the volume cannot vanish,

$$\alpha(\frac{1}{2}R - \Lambda)g^{\mu\nu} + \alpha R^{\mu\nu} = \frac{1}{2}T^{\mu\nu} \ ,$$

$$R^{\mu\nu} + \frac{1}{2}Rg^{\mu\nu} - \Lambda g^{\mu\nu} = \frac{1}{2\alpha}T^{\mu\nu} \ .$$

Since we want to obtain the Newtonian laws of gravity in the approximation of a weak gravitational field the constant $\frac{1}{2\alpha}$ must be equal to $\frac{8\pi \mathrm{G}}{c^4}$, as shown in the next section.

$$R^{\mu\nu} + \frac{1}{2}Rg^{\mu\nu} - \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu} \ , \tag{1.13}$$

which is the Einstein field equations. Assuming now $G = 1$ as well as $c = 1$ gives the form previously mentioned. □

1.4 The Newtonian Limit

The constant 8π in Eq. (1.3) and Eq. (1.4) can be found by imposing that they must reduce to Newton's gravitational field in the case of a weak gravitational field. To make the derivation more general we shall use λ as our constant in front of the field equation.

Let us consider a simple case in which the matter flow does not carry any force flow. That is a situation in which energy-momentum tensor's only non-zero component is T^{00}, which is the 0^{th} component of momentum in time. Such a flow through time is just the mass-energy density ρ, giving to the energy-momentum tensor the following form:

$$T = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \ .$$

Because of that, if we define the metric to be

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \qquad |h_{\mu\nu}| \ll 1 \ ,$$

so that we are in a weak gravitational field, the trace of $T_{\mu\nu}$ will be

$$T = T^{\mu}_{\mu} = \eta^{\mu\mu}T_{\mu\mu} = \eta^{00}T_{00} = -\rho \ .$$

Substituting this result in Eq. (1.4) as well as using T_{00} and g_{00} we get:

$$R_{00} = \frac{1}{2}\lambda\rho \ .$$

From the definition of Ricci tensor we can also write R_{00} as:

$$R_{00} = R^{\mu}_{0\mu0} = \partial_\mu \Gamma^{\mu}_{00} - \partial_0 \Gamma^{\mu}_{\mu0} + \Gamma^{\mu}_{\mu\xi}\Gamma^{\xi}_{00} - \Gamma^{\mu}_{0\xi}\Gamma^{\xi}_{\mu0} \ .$$

Since $\partial_0 \Gamma^{\mu}_{\mu0}$ is just the dime derivative, and we assumed a static field, it must vanish. Moreover, because Γ^{a}_{bc} is already of the first order in $g_{\mu\nu}$, then $\Gamma^{a}_{bc}\Gamma^{d}_{ef}$ has to be of the second order. We can then neglect them in a linear approximation of the metric, obtaining:

$$R_{00} = \partial_\mu \Gamma^\mu_{00} = \frac{1}{2}\lambda\rho \ .$$

From the definition of Christoffel's symbol, $\Gamma^\mu_{00} = -\frac{1}{2}\eta^{\mu\nu}\partial_\nu h_{00}$ it follows that

$$\eta^{\mu\nu}\partial_\mu\partial_\nu h_{00} = -\lambda\rho \ .$$

Given the form of $\eta^{\mu\nu}$ and the staticity of the field:

$$\partial_1\partial_1 h_{00} + \partial_2\partial_2 h_{00} + \partial_3\partial_3 h_{00} = -\lambda\rho \ ,$$

which is the definition of a Euclidean Laplacian

$$\nabla^2 h_{00} = -\lambda\rho \ . \tag{1.14}$$

Defining $h_{00} = -2\Phi$, we get that Eq. (1.14) is very similar to the Poisson equation for Newton's gravitational field

$$\nabla^2\Phi = 4\pi \mathrm{G}\rho \ .$$

To obtain the equivalence between the two expressions, λ must be equal to 8πG, showing that, in the weak gravity approximation, Einstein's theory of gravity indeed transforms into Newton's. Tolish [2010]

Chapter 2

Physical meaning of curvature

In this chapter, we shall discuss some mathematical results and structures necessary to introduce the concept of singularities.

2.1 Energy Conditions

Since the actual universe is made up of many fields, the form of the energy-momentum tensor is quite complex to compute. One has very little idea of the behavior of the matter field in the case of extreme density and pressure. Because of that, it might seem very hard to predict the existence of singularities from the right hand of the Einstein field equations 1.3. Nevertheless, if we are interested in studying the general features of space-time without considering the shape of the energy-momentum tensor itself, we shall strictly use geometric equations such as Raychaudhuri's:

$$\frac{d\theta}{ds} = -R_{\alpha\beta}V^{\alpha}V^{\beta} + 2\omega^2 - 2\sigma^2 - \frac{1}{3}\theta^2 + \dot{V}^{\alpha}_{;\alpha}, \tag{2.1}$$

$$\frac{d\hat{\theta}}{ds} = -R_{\alpha\beta}V^{\alpha}V^{\beta} + 2\omega^2 - 2\sigma^2 - \frac{1}{3}\theta^2, \tag{2.2}$$

where ω is the vorticity (which induce an expansion of the fluid), σ the shear (which induce a contraction of the fluid), θ the volumetric expansion, and V^{α} the time-like unit vector tangent to the flow, to obtain some interesting predictions. The first of the equations shown above is relative to a congruence of time-like curves, while the second to a congruence of null curves. By considering either a null-like vector field or one that is time-like ($\mathbf{T_p}$), both tangent to the geodesics belonging to the congruence with vanishing vorticity, and by assuming $\dot{\theta}$ negative, from the fact that gravity is always attractive, we derive a series of inequalities that will allow us to predict the existence of singularities without the knowledge of the exact form of the energy-momentum tensor. Martín-Moruno and Visser [2017]

The first of such inequalities will be the *weak energy condition*. It says that at each point $p \in \mathcal{M}$ the energy-momentum tensor for a time-like vector, and by continuity also for a null-like vector, $\mathbf{W} \in \mathbf{T_p}$ obeys $T_{\alpha\beta}W^\alpha W^\beta \geq 0$. Because of that, an observer whose world-line at p has a tangent vector $\mathbf{V}$ sees a positive local energy density $T_{\alpha\beta}V^\alpha V^\beta$. This assumption might be regarded as asking that, for any observer, the local energy distribution must be non-negative.

By expressing the energy-momentum tensor in p with respect to the orthonormal basis $\mathbf{E_1}, \mathbf{E_2}, \mathbf{E_3}, \mathbf{E_4}$ we hope to diagonalize it with the help of local Lorentz transformations. Such an undertaking is unfortunately not trivial. Thus, we need to find a different approach to solve this problem. One such method consists of finding four different partially diagonalized forms of such a tensor (assuming $\mathbf{E_4}$ time-like) and classifying them in terms of the space-like/light-like/time-like nature of their eigenvectors.

Type 1:

$$T^{\alpha\beta} = \begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & p_3 & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix}.$$

This is a *type 1* energy-momentum tensor, which is the general case when $\mathbf{E}_4$ is an eigenvector belonging to the eigenvalue μ, unique for $\mu \neq p_a$ (with $a = 1, 2, 3$). This represents the energy-density measured by an observer whose world-line has $\mathbf{E}_4$ as unit tangent vector at p. The other three eigenvalues p_a are the principal pressures in the three space-like directions $\mathbf{E}_\alpha$. This form of the energy-momentum tensor is valid for all the observed fields with non-zero rest mass and also for all the zero mass fields that are not contained in the case of type 2 tensors.

Type 2:

$$T^{\alpha\beta} = \begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & \nu - \kappa & \nu \\ 0 & 0 & \nu & \nu + \kappa \end{pmatrix}.$$

This is a more special case of the energy-momentum tensor, that is a *type 2* tensor, with a double null-like eigenvector ($\mathbf{E}_3 + \mathbf{E}_4 = k^\alpha = (0; 0; 1; 1)$), where $\nu = \pm 1$. The only observed occurrence of such a form is in the case of a zero rest mass field, representing the radiation traveling in the direction of $\mathbf{E}_3 + \mathbf{E}_4$. For the radiation field we will have p_1, p_2, and κ all vanishing identically.

Type 3:

$$T^{\alpha\beta} = \begin{pmatrix} p & 0 & 0 & 0 \\ 0 & -\nu & 1 & 1 \\ 0 & 1 & -\nu & 0 \\ 0 & 1 & 0 & \nu \end{pmatrix}.$$

In this special case the energy-momentum tensor has a triple null-like eigenvector ($\mathbf{E}_3 + \mathbf{E}_4 = k^\alpha = (0; 0; 1; 1)$)

Type 3:

$$T^{\alpha\beta} = \begin{pmatrix} p_1 & 0 & 0 & 0 \\ 0 & p_2 & 0 & 0 \\ 0 & 0 & -\kappa & \nu \\ 0 & 0 & \nu & 0 \end{pmatrix}.$$

This is the general case in which the energy-momentum tensor has no time-like or null-like eigenvector, with the condition $\kappa^2 < 4\nu^2$.

The weak energy condition for a type I will hold if

$$\mu \geq 0, \quad \mu + p_a \geq 0,$$

with $(a = 1, 2, 3)$, while for a type II tensor if

$$p_1 \geq 0, \quad p_2 \geq 0, \quad \kappa \geq 0, \quad \nu = +1.$$

For the last two types of energy-momentum tensor there are no observable fields which exhibit such a behavior.

The second inequality we shall consider is the *dominant energy condition* which states that for every time-like vector W_α, with $T_{\alpha\beta}W^\alpha W^\beta \geq 0$, the vector $T_{\alpha\beta}W^\alpha$ must be non-spacelike. Physically, this may be interpreted as saying that, for a local observer, the energy density must be non-negative and the local energy-flow vector is non-spacelike. Equivalently, that means having the energy component dominant with respect to the others in an orthonormal basis,

$$T^{00} \geq |T^{\alpha\beta}|.$$

Such a condition will hold for a type I tensor if

$$\mu \geq 0, \quad -\mu \leq p_a \leq \mu,$$

with $(a = 1, 2, 3)$, and for a type II if

$$\nu = +1, \quad \kappa \geq 0, \quad 0 \leq p_i \leq \kappa,$$

with $(i = 1, 2)$. This means that the dominant energy condition is just the weak energy condition, but with the additional requirement that the pressure must not exceed the

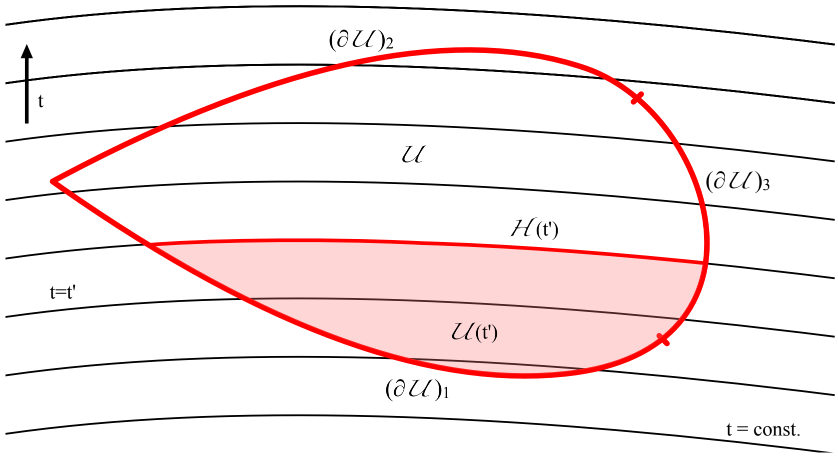

Figure 2.1: *$\mathcal{U}$ is a compact region of space-time with past and future non-spacelike boundaries $(\partial\mathcal{U})_1$, $(\partial\mathcal{U})_2$ and a time-like boundary $(\partial\mathcal{U})_3$*

energy density. Since the speed of sound is $dp_a/d\mu$, and it must be less than the speed of light (from the relativity postulates), using $c = 1$, the derivation of the above condition is trivial.

Let us now consider a C^2 function t whose gradient is time-like everywhere (such as in figure 2.1) and a region $\mathcal{U}$ in our space-time whose boundary $\partial\mathcal{U}$ consists of $(\partial\mathcal{U})_1$, whose normal form $\mathbf{n}$ is space-time and with $n_\alpha t_{;\alpha} g^{\alpha\beta}$ positive, $(\partial\mathcal{U})_2$, whose normal form $\mathbf{n}$ is space-time and with $n_\alpha t_{;\alpha} g^{\alpha\beta}$ negative, and $(\partial\mathcal{U})_3$. The sign of $\mathbf{n}$ is given by the requirement that the scalar product $\langle \mathbf{n}, \mathbf{X} \rangle$ must be positive for all vector $\mathbf{X}$ pointing out of $\mathcal{U}$. The symbol $\mathcal{U}(t')$ denotes the region of $\mathcal{U}$ for which $t < t'$, bounded at the top by $\mathcal{H}(t')$ that is the surface at $t = t'$.

It is now useful to introduce the following lemma for later use.

Lemma 2.1.1. *There is some positive constant value P such that for any tensor $S^{\alpha\beta}$ satisfying the dominant energy condition and vanishing on $(\partial\mathcal{U})_3$, the following inequality holds:*

$$\int_{\mathcal{H}(t)\cap\mathcal{U}} S^{\alpha\beta} t_{;\alpha} d\sigma_\beta \leq - \int_{(\partial\mathcal{U})_1} S^{\alpha\beta} t_{;\alpha} d\sigma_\beta + P \int^t \left(\int_{\mathcal{H}(t')\cap\mathcal{U}} S^{\alpha\beta} t_{;\alpha} d\sigma_\beta \right) dt' + \int^t \left(\int_{\mathcal{H}(t')\cap\mathcal{U}} S^{\alpha\beta}_{;\alpha} d\sigma_\beta \right) dt'.$$

As a direct consequence of this result we get the Conservation theorem:

Theorem 2.1.2 (Conservation theorem)**.** *If the energy-momentum tensor obeys the dominant energy condition, and is zero both on $(\partial\mathcal{U})_3$ and on $(\partial\mathcal{U})_1$ then it is zero everywhere.*

From this theorem we deduce that if the energy-momentum tensor vanishes on a set $\mathcal{S}$ then it also vanishes on the future Cauchy development $D^+(\mathcal{S})$, defined as the set of all points through which every past-directed, non-spacelike curve intersects $\mathcal{S}$.

By conjoining the weak and dominant energy condition it follows that the square mass distribution of the fluid must be non-negative, which means that it is not made of tachyons (special particles that are supposed to travel faster than light), and therefore the matter fluid can not travel faster than light.

For our purposes, the weak energy condition is important because it implies that matter always has a converging effect on congruences of null geodesics. Without vorticity, equation (2.2) becomes:

$$\frac{d\hat{\theta}}{ds} = -R_{\alpha\beta}V^\alpha V^\beta - 2\sigma^2 - \frac{1}{3}\theta^2,$$

meaning that $\hat{\theta}$ will monotonically decrease along a null geodesic when $R_{\alpha\beta}V^\alpha V^\beta \geq 0$. We shall call this the *null convergence condition*, and imply the weak energy condition in Einstein's field equation (1.3), independently of Λ.

On the other hand, for eq. (2.1) the expansion of θ for a time-like congruence with zero vorticity will monotonically decrease along a geodesic if $R_{\alpha\beta}V^\alpha V^\beta \geq 0$. This takes the name of *time-like convergence condition.* By the Einstein equation (1.4) the condition is valid if the energy-momentum tensor obeys the following inequality:

$$T_{\alpha\beta}V^\alpha V^\beta \geq V^\alpha V^\beta g_{\alpha\beta}\left(\frac{1}{2}T - \frac{1}{8\pi}\Lambda\right),$$

which for a type I tensor holds if

$$\mu + p_a \geq 0, \quad \mu + \sum_\alpha p_\alpha - \frac{1}{4\pi}\Lambda \geq 0,$$

and for a type II if

$$\nu = +1, \quad \kappa \geq 0, \quad p_1 \geq 0, \quad p_2 \geq 0, \quad p_1 + p_2 + \frac{1}{4\pi}\Lambda \geq 0.$$

The energy-momentum tensor will be said to satisfy the *strong energy condition* if it obeys the above inequalities when $\Lambda = 0$, that is a stricter, physically reasonable, requirement. For the general case, type I, such a requirement will be violated only by a negative energy density or a large negative pressure.

2.2 Conjugate points

The components of the vector **Z**, that represents the separation between a curve $\gamma(s)$ and neighboring curves in a congruence of time-like geodesics, satisfy what is called the Jacobi equation:

$$\frac{d^2Z^a}{ds^2} = -R_{a4b4}Z^b, \tag{2.3}$$

this is a differential equation whose solutions are called *Jacobi fields* along $\gamma(s)$. A Jacobi field might be regarded as representing the separation between a neighborhood of geodesics passing through a point q. Because of the second order nature of the equation, in every point of $\gamma(s)$ there are six independent Jacobi fields along such a curve. At a point q in which $\gamma(s)$ vanishes, the number of independent Jacobi fields reduce to three, expressed as:

$$Z^a(s) = A_{ab}(s)\frac{d}{ds}Z^b|_q.$$

In which A is a 3 by 3 matrix such that:

$$A_{ab}(s)|_q = 0,$$

$$\frac{d^2}{ds^2}A_{ab}(s) = -R_{a4b4}A_{cd}(s). \tag{2.4}$$

In such a context we can define the vorticity, shear and volumetric expansion as:

$$\omega_{ab} = A^{-1}_{c[b}\frac{d}{ds}A_{a]c}, \tag{2.5}$$

$$\sigma_{ab} = A^{-1}_{c(b}\frac{d}{ds}A_{a)c} - \frac{1}{3}\delta_{ab}\theta, \tag{2.6}$$

$$\theta = \det{(A)}^{-1}\frac{d}{ds}\det{(A)}. \tag{2.7}$$

Such that

$$A_{ca}\omega_{cd}A_{db} = \frac{1}{2}\left(A_{ca}\frac{d}{ds}A_{cb} - A_{cb}\frac{d}{ds}A_{ca}\right),$$

will be constant along $\gamma(s)$, and will vanish at q where A_{ab} is zero. Thus, ω_{ab} will vanish wherever A_{ab} is non-singular.

We shall now define a *conjugate point* p *of* q *along* $\gamma(s)$ as a point of $\gamma(s)$ for which there exists a Jacobi fields along such a curve, not identically zero, vanishing both on q and p. p may be regarded as a point in which the infinitesimal neighboring geodesics through q intersect. The Jacobi fields along $\gamma(s)$ which vanishes at q is described by A_{ab}. Thus, a point p is conjugate to q if, and only if, A_{ab} is singular at p. Since θ is expressed by eq. (2.7) and A_{ab} obeys eq. (2.4) in which $-R_{a4b4}$ is finite, then $d(\det A)/ds$ will be finite. Therefore, a point p will be conjugate to q along $\gamma(s)$ if θ there becomes infinite.

Proposition 2.2.1. *If at some point $\gamma(s_1)$ ($s_1 > 0$), the expansion θ has a negative value $\theta_1 < 0$ and if $R_{\alpha\beta}V^\alpha V^\beta \geq 0$ everywhere, then there will be a point conjugate to q along $\gamma(s)$ between $\gamma(s_1)$ and $\gamma(s_1 + (3/-\theta_1))$ showing that $\gamma(s)$ can be extended to this parameter value (which might be impossible if the space-time is geodetically incomplete).*

This means that if the time-like convergence condition holds and if the neighboring geodesics from q start converging on $\gamma(s)$, then some infinitesimal neighboring geodesic will intersect the curve, proving that it is extendible to that value of s.

Proposition 2.2.2. *If $R_{\alpha\beta}V^\alpha V^\beta \geq 0$ and if at some point $p = \gamma(s_1)$ the tidal force $R_{\alpha\beta\gamma\mu}V^\gamma V^\mu$ is non-zero, then there will be two values s_0, s_2 such that $\gamma(s_0) = q$ and $\gamma(s_2) = r$ will be conjugate of p along $\gamma(s)$, showing that such a curve can be extended to these values of s.*

Physically one would expect that, for real solutions, time-like geodesics will encounter some matter or some gravitational radiation along their path and therefore contain some points for which $R_{\alpha\beta\gamma\mu}V^\gamma V^\mu$ is non-zero. It would be reasonable to assume, then, that such a solution contains a pair of conjugate points, allowing to extend the solution in both directions.

We shall now consider not only a curve passing through a single point, but a whole congruence of time-like geodesics normal to a space-like three-surface $\mathscr{H}$, that is an imbedded three-dimensional submanifold defined by $f = 0$ where f is a function C^2, and such that $g^{\alpha\beta}f_{;\alpha}f_{;\beta} < 0$ for that value of f. Let $\mathbf{N}$ be the unit normal vector to $\mathscr{H}$ such that $N^\alpha = \left(-g^{\beta\gamma}f_{;\beta}f_{;\gamma}\right)^{-1/2}g^{\alpha\mu}f_{;\mu}$ and χ be the second fundamental tensor of $\mathscr{H}$, defined as $\chi_{\alpha\beta} = h^\gamma_\alpha h^\mu_\beta N_{\gamma\mu}$ symmetric by definition. Here $h_{\alpha\beta}$ is the first fundamental tensor for $\mathscr{H}$, or induced metric, and is defined as $h_{\alpha\beta} = g_{\alpha\beta} + N_\alpha N_\beta$. The congruence of time-like geodesics orthogonal to $\mathscr{H}$ consists of time-like geodesics whose unit tangent vector $\mathbf{V}$ is equal to $\mathbf{N}$ at $\mathscr{H}$, that is:

$$V_{\alpha;\beta} = \chi_{\alpha\beta}.$$

The separation between an infinitesimal neighboring geodesics normal to the surface and the normal geodesic $\gamma(s)$ is defined using the vector $\mathbf{Z}$, and will obey Eq. (2.3) as well as the initial condition at $q \in \mathscr{H}$:

$$\frac{d}{ds}Z^a = \chi_{ab}Z^b.$$

Similarly, it is possible to express the Jacobi fields along $\gamma(s)$ satisfying the above condition as:

$$Z^a(s) = A_{ab}(s)Z^b|_q,$$

in which A_{ab} obeys (2.4) at q, and it is a unit matrix such that:

$$\frac{d}{ds}A_{ab} = \chi_{ac}A_{cb}$$

A point p is said to be conjugate to $\mathscr{H}$ along $\gamma(s)$ if there is a non-zero Jacobi field along such a curve which satisfies the previous initial conditions at q and vanishes at p. Therefore, according to the case of a single point, p is conjugate to $\mathscr{H}$ if and only if A_{ab} is singular at p or equivalently θ goes to infinity, with initial value $\chi_{ab}g^{ab}$, since ω_{ab} is zero at q.

Proposition 2.2.3. *If $R_{\aleph\beta}V^{\alpha}V^{\beta} \geq 0$and $\chi_{\alpha\beta}g^{\alpha\beta} < 0$, there will be a point conjugate to $\mathscr{H}$ along $\gamma(s)$ within a distance $3/(-\chi_{\alpha\beta}g^{\alpha\beta})$ from $\mathscr{H}$, showing that such a curve is extendible.*

Considering the solution of the differential equation:

$$\frac{d^2}{dv^2}z^m = -R_{m4n4}Z^n \quad (m, n = 1, 2),$$

along a null geodesic $\gamma(v)$, we shall call it the Jacobi fields along such a curve. In the same way Z^m might be regarded as the components of $\mathbf{Z}$ in S_q (with $q \in \mathscr{H}$), expressed with respect to the basis $\mathbf{E}_1$ and $\mathbf{E}_2$, representing the vector connecting neighboring null geodesics passing through q. We shall say that p is a conjugate point to q, along a null geodesic, if there is a non-zero Jacobi filed along it which vanishes at q and p. The interpretation we derive from this point is the same we get for time-like geodesics.

Using the $\hat{A}_{mn}$ 2×2 matrix we find similar relations to the ones found in the case of time-like geodesics, as well as a similar proposition to proposition 2.2.1 and proposition 2.2.2:

Proposition 2.2.4. *If $R_{\alpha\beta}V^{\alpha}V^{\beta} \geq 0$ everywhere and if at some point $\gamma(v_1)$ the expansion $\hat{\theta}$ has a negative value $\hat{\theta}_1 < 0$, then there will be a point conjugate to q along $\gamma(v)$ between $\gamma(v_1)$ and $\gamma(v_1 + (2/-\hat{\theta}_1))$ showing that it is possible to extend the curve that far.*

Proposition 2.2.5. *If $R_{\alpha\beta}V^{\alpha}V^{\beta} \geq 0$ everywhere and if at $p = \gamma(v_1)$, $K^{\gamma}K^{\mu}K_{[\alpha}R_{\beta]\gamma\mu[\nu}K_{\rho]}$ is non-zero, then there will be two values v_0 and v_2 such that $q = \gamma(v_0)$ and $r = \gamma(v_2)$ will be conjugate along $\gamma(v)$ showing that it can be extented to those values of the parameter.*

This condition will be satisfied for null geodesics passing through matter showing that matter is not pure radiation, moving in the direction of the geodesic tangent vector $\mathbf{K}$. In the empty space it will be satisfied if the geodesic contains some point where the Weyl tensor is non-zero and $\mathbf{K}$ does not lie in one of the directions at that point for which $K^{\gamma}K^{\mu}K_{[\alpha}R_{\beta]\gamma\mu[\nu}K_{\rho]} = 0$. It seems physically reasonable to assume that both time-like and null geodesics will contain a point at which $K^{\gamma}K^{\mu}K_{[\alpha}R_{\beta]\gamma\mu[\nu}K_{\rho]}$ is not zero. This condition is called the *generic condition.*

To expand this concept to null geodesics orthogonal to a space-like two-surface $\mathscr{S}$, that is an imbedded two-dimensional submanifold defined locally by $f_1 = 0$, $f_2 = 0$, where f_1 and f_2 are C^2 functions such that at that values $f_{1;\alpha}$ and $f_{2;\alpha}$ do not vanish, are not parallel, and satisfy

$$(f_{1;\alpha} + \mu f_{2;\alpha})(fa_{1;\beta} + \mu f_{2;\beta})g^{\alpha\beta} = 0,$$

for two distinct values of μ. Then any vector lying in the two-surface is necessarily space-like. The two null vectors normal to $\mathscr{S}$, N_1^α and N_2^α, are proportional to $(f_{1;\beta}+\mu_1 f_{2;\beta})g^{\alpha\beta}$ and $(fa_{1;\beta} + \mu_2 f_{2;\beta})g^{\alpha\beta}$ respectively, and such that:

$$N_1^\alpha N_2^\beta g_{\alpha\beta} = -1.$$

To complete the pseudo orthonormal basis we shall introduce two space-like unit vectors Y_1^α and Y_2^α so that they are orthonormal to each other and to N_1^α, N_2^α. The two fundamental tensors on $\mathscr{S}$ can be now defined as:

$${}_n\chi_{\alpha\beta} = -N_{n\gamma;\mu}(Y_1^\gamma Y_{1\alpha} + Y_2^\gamma Y_{2_\alpha})(Y_1^\mu Y_{1\mu} + Y_2^\mu Y_{2\beta})$$

where $n = 1, 2$ and the two tensors ${}_1\chi_{\alpha\beta}$ and ${}_2\chi_{\alpha\beta}$ are symmetric.

There will be two distinct families of null geodesics normal to $\mathscr{S}$ corresponding to the two null normal vectors N_1^α and N_2^α. From this consideration, we shall fix our basis defining $\mathbf{E}_1 = \mathbf{Y}_1$, $\mathbf{E}_2 = \mathbf{Y}_2$, $\mathbf{E}_3 = \mathbf{N}_1$,$\mathbf{E}_4 = \mathbf{N}_2$. This projection into S_q of $\mathbf{Z}$ will satisfy the initial condition

$$\frac{d}{dv}Z^m = \chi_{2mn}Z^n$$

at q on $\gamma(v)$, and the differential equation

$$\frac{D^2}{dv^2}Z^\alpha = -R^\alpha_{\beta\gamma\mu}Z^\gamma K^\beta K^\mu \tag{2.8}$$

where D/dv represents the covariant derivative. Analogously to the case with time-like geodesics, having zero vorticity gives us an initial value for θ equal to $\chi_{2\aleph\beta}g^{\alpha\beta}$. Proposition 2.2.3 becomes:

Proposition 2.2.6. *If* $R_{\alpha\beta}K^\alpha K\beta \geq 0$ *everywhere and* $\chi_{2\alpha\beta}g^{\alpha\beta}$ *is negative, then there will be a point conjugate to* $\mathscr{S}$ *along* $\gamma(v)$ *within an affine distance of* $2/(-\chi_{2\aleph\beta}g^{\alpha\beta})$ *from the surface itself.*

To tie everything together, the definition of conjugate point implies the existence of self-intersections in families of geodesics.

2.3 Variation of arc-length

In this section, we shall consider time-like and non-spacelike curves which will be piece-wise C^3, but may have points in which their tangent vector is not continuous. In such cases we require that the two tangent vectors at that point

$$\left.\frac{\partial}{\partial t}\right|_{-}, \left.\frac{\partial}{\partial t}\right|_{+},$$

will satisfy

$$g\left(\left.\frac{\partial}{\partial t}\right|_{-}, \left.\frac{\partial}{\partial t}\right|_{+}\right) = -1,$$

that is, they point within the same half of the null cone (defined by the metric at that point).

Proposition 2.3.1. *Let $\mathscr{U}$ be a convex normal coordinate neighborhood about q. Thus, the point which can be reached from q by time-like (non-spacelike) curves in $\mathscr{U}$ are those of the form $\exp_q(\mathbf{X})$, with $\mathbf{X} \in \mathbf{T_q}$ and such that $g(\mathbf{X},\mathbf{X}) < 0$ ($g(\mathbf{X},\mathbf{X}) \leq 0$) (where we assume that the exponential map to be restricted to the neighborhood diffeomorphic to $\mathscr{U}$ by $\exp_q$ and centered in $\mathbf{T_q}$)*

Hence the null geodesics from q form a submanifold that works as a boundary for $\mathscr{U}$ reachable from q by a time-like or a non-spacelike curve. This proposition is rather important for the concept of causality, but we will not go into the detail of the proof. [Hawking and Ellis, 1973, p. 103]

Corollary 2.3.0.1. *If $p \in \mathscr{U}$ can be reached from q by a non-spacelike curve, but not from a time-like curve, than p lies on a null geodesic from q.*

Physically it means that it can only be reached by an object going at the speed of light.

The length of a non-spacelike curve from q to p is

$$L(\gamma, q, p) = \int_q^p \left[-g\left(\left.\frac{\partial}{\partial t}\right|_{-}, \left.\frac{\partial}{\partial t}\right|_{+}\right)\right]^{1/2},$$

over the differentiable section of the curve $\gamma(t)$.

In general, given a positive define metric, it is possible to find the shortest path between two points. Such a curve does not exist in the case of a Lorenz metric in which any curve can be deformed into a null curve having a length of zero. However, in certain cases there will be a non-spacelike curve that is the longest path between two points or between a point and a space-like three-surface. We shall consider first the case of two points, leaving the discussion on the sufficient condition to a further section.

Proposition 2.3.2. *Let q and p lie on a convex normal neighborhood $\mathscr{U}$. If q and p can be joined by a non-spacelike curve in $\mathscr{U}$, then the longest such curve is the unique non-spacelike geodesic in $\mathscr{U}$ from the first point to the second. Additionally, defining $\rho(q,p)$ as a length of the path, if it exists, and zero if it does not, $\rho(q,p)$ is a continuous function on $\mathscr{U} \times \mathscr{U}$.*

Let us now consider the case in which q and p are not necessarily contained in a convex normal neighborhood $\mathscr{U}$. By taking a small variation, that is a C^{1-} map $\alpha : (-\epsilon, \epsilon) \times [0, t_p] \to \mathscr{M}$, we shall derive the necessary conditions for a $\gamma(t)$ curve to be the longest path between q and p. A variation α of $\gamma(t)$ is such that:

(1) $\alpha(0, t) = \gamma(t)$;

(2) there is a subdivision $0 = t_1 < t_2 < \cdots < t_n = t_p$ of $[0, t_p]$ such that α is C^3 on each set $(-\epsilon, \epsilon) \times [t_i, t_i + 1]$;

(3) $\alpha(u, 0) = q,\ \alpha(u, t_p) = p$;

(4) for each constant $u,\ \alpha(u, t)$ is a time-like curve.

We can call the vector $\mathbf{Z} = (\partial/\partial u)_\alpha|_{u=0}$ *variation vector*. Vice versa, given a continuous piecewise C^2 vector field $\mathbf{Z}$ along $\gamma(t)$ vanishing at q and p, we may define a variation α for which $\mathbf{Z}$ will be the variation vector:

$$\alpha(u, t) = \exp_r(u\mathbf{Z}|_\mathbf{r}),$$

where $u \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$ and $r = \gamma(t)$.

Lemma 2.3.1. *The variation of the length from q to p under α is given by*

$$\left.\frac{\partial L}{\partial u}\right|_{u=0} = \sum_{i=1}^{n-1} \int_{t_i}^{t_{i+1}} g\left(\frac{\partial}{\partial u}, \left\{f^{-1}\frac{D}{\partial t}\frac{\partial}{\partial t} - f^{-2}\left(\frac{\partial f}{\partial t}\right)\frac{\partial}{\partial t}\right\}\right) dt$$
$$+ \sum_{i=2}^{n-1} g\left(\frac{\partial}{\partial u}, \left[f^{-1}\frac{\partial}{\partial t}\right]\right),$$

where $f^2 = g(\partial/\partial t, \partial/\partial t)$ is the magnitude of the tangent vector and $[f^{-1}\partial/\partial t]$ is the discontinuity at one of the singular points of the curve.

By choosing s as the parameter t of the arc-length, one can simplify the integral obtaining $g(\partial/\partial t, \partial/\partial t) = -1$. Now denoting V the unit tangent vector $\partial/\partial s$ the integral becomes:

$$\left.\frac{\partial L}{\partial u}\right|_{u=0} = \sum_{i=1}^{n-1} \int_{t_i}^{t_i+1} g(\mathbf{Z}, \dot{\mathbf{V}}) ds + \sum_{i=2}^{n-1} g(\mathbf{Z}, [\mathbf{V}])$$

where $\dot{\mathbf{V}} = D\mathbf{V}/\partial s$ is the acceleration. From these considerations, it is possible to see that for $\gamma(t)$ to be the longest path between two given points, it is a necessary condition that it should be an unbroken geodesic, otherwise one could choose a variation yielding a longer curve.

Considering now the case of a point joined with a time-like curve $\gamma(t)$ from a space-like three-surface $\mathscr{H}$ a variation α of $\gamma(t)$ is defined as being done before with the exception that condition (3) is replaced by:

(3) $\alpha(u,0)$ lies on $\mathscr{H}$ and $\alpha(u,t_p)=p$

Thus the variation vector $\mathbf{Z}=\partial/\partial u$ at $\mathscr{H}$ lies on $\mathscr{H}$.

Lemma 2.3.2. *In such a case the variation of the length of the path will be:*

$$\left.\frac{\partial L}{\partial u}\right|_{u=0}=\sum_{i=1}^{n-1}\int_{t_i}^{t_i+1} g(\dot{\mathbf{V}},\mathbf{Z})ds+\sum_{i=2}^{n-1} g(\mathbf{Z},[\mathbf{V}])+g(\mathbf{Z},[\mathbf{V}])|_{s=0}$$

The necessary condition for $\gamma(t)$ to be the longest path from $\mathscr{H}$ to the point p is that the curve must be an unbroken geodesic orthogonal to the surface itself.

Knowing that under a variation the first derivative of the length of a time-like geodesics is zero, we proceed to calculate the second derivative. That is to obtain a further understanding of the behavior of such a curve. Let us define a two-parameter variation as a C^1 map $\alpha:(-\epsilon_1,\epsilon)\times(-\epsilon_2,\epsilon_2)\times[0,t_p]\to\mathscr{M}$ of a geodesic from q to p such that:

(1) $\alpha(0,0,t)=\gamma(t)$;

(2) there is a subdivision $0=t_1<t_2<\cdots<t_n=t_p$ of $[0,t_p]$ such that α is C^3 on each set $(-\epsilon_1,\epsilon_1)\times(-\epsilon_2,\epsilon_2)\times[t_i,t_i+1]$;

(3) $\alpha(u_1,u_2,0)=q,\ \alpha(u_1,u_2,t_p)=p$;

(4) for each constant $u_1,\ u_2,\ \alpha(u_1,u_2,t)$ is a time-like curve.

Let us now define the two variation vectors as $\mathbf{Z_1}=(\partial/\partial u_1)|_{u_{1,2}=0}$ and $\mathbf{Z_2}=(\partial/\partial u_2)|_{u_{1,2}=0}$ and vice-versa, similar to the case of one variation, given two piecewise C^2 vector fields $\mathbf{Z_1}$ and $\mathbf{Z_2}$ along $\gamma(t)$ it is possible to define a variation such that the two vector fields would be the variation vectors:

$$\alpha(u_1,u_2,t)=\exp_r(u_1\mathbf{Z_1}+u_2\mathbf{Z_2})$$

where $r=\gamma(t)$.

Lemma 2.3.3. *Under a two-parameter variation the shape of the second derivative of the geodesic curve length will be:*

$$\left.\frac{\partial^2 L}{\partial u_2\partial u_1}\right|_{u_1,2=0}=\sum_{i=1}^{n-1}\int_{t_i}^{t_i+1} g\left(\mathbf{Z_1},\left\{\frac{D^2}{\partial s^2}(\mathbf{Z_2}+g(\mathbf{V},\mathbf{Z_2})\mathbf{V})-\mathbf{R}(\mathbf{V},\mathbf{Z_2})\mathbf{V}\right\}\right)ds$$
$$+\sum_{i=2}^{n-1} g\left(\mathbf{Z_1},\left[\frac{D}{\partial s}(\mathbf{Z_2}+g(\mathbf{V},\mathbf{Z_2})\mathbf{V})\right]\right).$$

Even if it is not obvious at first glance, from the definition of this expression we know it is symmetric in the two variation vector fields $\mathbf{Z_1}$ and $\mathbf{Z_2}$. It solely depends on the projection of $\mathbf{Z_1}$ and $\mathbf{Z_2}$ into the space orthogonal to $\mathbf{V}$. Thus, in our consideration we shall use only variations α whose variation vectors are orthogonal to $\mathbf{V}$. By defining a vectorial space of infinite dimension $\mathbf{T}_\gamma$ corresponding of all continuous, piecewise C^2 vector fields along $\gamma(t)$ orthogonal to $\mathbf{V}$ vanishing at q and p, we get that $\partial^2 L/\partial u_2 \partial u_1$ is a symmetric map of $\mathbf{T}_\gamma \times \mathbf{T}_\gamma$ to $\mathbb{R}$. It can be regarded as a symmetric tensor over $\mathbf{T}_\gamma$, written as:

$$\mathbf{L}(\mathbf{Z_1}, \mathbf{Z_2}) = \left. \frac{\partial^2 L}{\partial u_2 \partial u_1} \right|_{u_{1,2}=0}$$

It is also possible to calculate the second derivative with respect to the variation of the length of a geodesic curve $\gamma(t)$ from $\mathscr{H}$ to p, normal to the surface itself.

Lemma 2.3.4. *The second derivative of the length of $\gamma(t)$ from $\mathscr{H}$ to p is:*

$$\left. \frac{\partial^2 L}{\partial u_2 \partial u_1} \right|_{u_1,2=0} = \sum_{i=1}^{n-1} \int_{t_i}^{t_i+1} g\left(\mathbf{Z_1}, \left\{ \frac{D^2}{\partial s^2}(\mathbf{Z_2} - \mathbf{R}(\mathbf{V}, \mathbf{Z_2})\mathbf{V} \right\} \right) ds$$
$$+ \sum_{i=2}^{n-1} g\left(\mathbf{Z_1}, \left[\frac{D}{\partial s}\mathbf{Z_2} \right] \right) + g\left(\mathbf{Z_1}, \frac{\mathbf{D}}{\partial \mathbf{s}}\mathbf{Z_2}\right)\Big|_{\mathscr{H}} - \chi(\mathbf{Z_1}, \mathbf{Z_2})|_{\mathscr{H}},$$

where $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are orthogonal to $\mathbf{V}$ and $\chi(\mathbf{Z_1}, \mathbf{Z_2})$ is the second fundamental tensor of $\mathscr{H}$.

In both cases we shall say that a time-like geodesic $\gamma(t)$ from a point q to another p or from a surface $\mathscr{H}$ to a point p is *maximal* if $L(\mathbf{Z_1}, \mathbf{Z_2})$ is negative semi-definite. That is, if such a curve is not maximal there exists a small variation α which yields a longer curve.

Proposition 2.3.3. *A time-like geodesic $\gamma(t)$ from a point q to a point p is maximal if, and only if, there is no point conjugate to q along the curve in the interval (q, p).*

A similar result occurs when considering, instead of two points, a point and a surface. That is:

Proposition 2.3.4. *A time-like geodesic $\gamma(t)$ from $\mathscr{H}$ to a point p is maximal if and only if there is no point in the interval (q, p), conjugate to $\mathscr{H}$ along the curve.*

By considering the variations of non-spacelike curves $\gamma(t)$ from q to p, we should be interested in finding the circumstances under which it is possible to find a variation α of $\gamma(t)$ which makes the scalar product $g(\partial/\partial t, \partial/\partial t)$ negative everywhere, in other words, the variation yields a time-like curve from q to p. Thus, the variation α should be of the form:

$$\begin{aligned}\frac{\partial}{\partial u}\left(g\left(\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right)\right) =& 2g\left(\frac{D}{\partial u}\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right) = 2g\left(\frac{D}{\partial t}\frac{\partial}{\partial u},\frac{\partial}{\partial t}\right)\\ &= 2\frac{\partial}{\partial t}\left(g\left(\frac{\partial}{\partial u},\frac{\partial}{\partial t}\right)\right) - 2g\left(\frac{\partial}{\partial u},\frac{D}{\partial t}\frac{\partial}{\partial t}\right),\end{aligned}$$

which, in order to be a time-like curve, it has to be less than or equal to zero on the curve.

Proposition 2.3.5. *If p and q are joined by a non-spacelike curve $\gamma(t)$ which is not a null geodesic, then they can also be joined by a time-like curve.*

To summarize, if $\gamma(t)$ is a geodesic curve, then the parameter t may be taken as the affine parameter, on the other hand, if $\gamma(t)$ is not a geodesic curve, then it can be varied to give a time-like geodesic. It is therefore necessary, but not sufficient, that the variation vector $\partial/\partial u$ of a variation α yielding a time-like curve to be everywhere orthogonal to the tangent vector $\partial/\partial t$ of the curve, otherwise $(\partial/\partial t)g(\partial/\partial u, \partial/\partial t)$ would be positive somewhere on $\gamma(t)$. The first derivative of the metric with respect to such a variation will be zero, it is then necessary to examine the second derivative.

In the case of null geodesics, the two-parameter variation α will be defined similarly as before except for the restriction to the variation vectors, that are requested to be orthogonal to the tangent vector of the $\partial/\partial t$ geodesic.

Under such a variation the study of the behavior of L is not convenient since, for $g(\partial/\partial t, \partial/\partial t) = 0$, $(-g(\partial/\partial t, \partial/\partial t))^{1/2}$ is not differentiable. Instead, we consider

$$\Lambda \doteq -\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} g\left(\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right) dt.$$

A necessary, but not sufficient, condition for a variation α yielding a time-like curve $\gamma(t)$ is that Λ should become positive between q and p. By considering the second term of the Taylor expansion of the metric around a point, one has:

$$\begin{aligned}\frac{1}{2}\frac{\partial^2}{\partial u_2 \partial u_1}\left(g\left(\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right)\right) &= \frac{\partial^2}{\partial u_2 \partial t}\left(g\left(\frac{\partial}{\partial u_1},\frac{\partial}{\partial t}\right)\right) - \frac{\partial}{\partial u_2}\left(g\left(\frac{\partial}{\partial u_2},\frac{D}{\partial t}\frac{\partial}{\partial t}\right)\right) =\\ &= \frac{\partial^2}{\partial u_2 \partial t}\left(g\left(\frac{\partial}{\partial u_1},\frac{\partial}{\partial t}\right)\right) -\\ &- g\left(\frac{\partial}{\partial u_1},\left\{\frac{D^2}{\partial t^2}\frac{\partial}{\partial u_2} - \mathbf{R}\left(\frac{\partial}{\partial t},\frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t}\right\}\right),\end{aligned}$$

which means that

$$\begin{aligned}\frac{1}{2}\left.\frac{\partial^2}{\partial u_2 \partial u_1}\right|_{u_{1,2}=0} &= \sum \int g\left(\frac{\partial}{\partial u_1}, \left\{\frac{D^2}{\partial t^2}\frac{\partial}{\partial u_2} - \mathbf{R}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial u_2}\right)\frac{\partial}{\partial t}\right\}\right) dt \\ &+ \sum g\left(\frac{\partial}{\partial u}, \left[\frac{D}{\partial t}\frac{\partial}{\partial u_2}\right]\right),\end{aligned} \tag{2.9}$$

It follows that the variation of Λ vanishes for a variation vector proportional to the tangent vector $\partial/\partial t$ since, because of the anti-symmetry of the Riemann tensor, we have that $\mathbf{R}(\partial/\partial t, \partial/\partial t)(\partial/\partial t) = 0$. From the previous relation for the variation of the length of a time-like curve, one can reduce it to a time-like geodesic by considering only the projection of the variation vector into the space S_q at each q. Thus, introducing a pseudo-orthonormal basis $\mathbf{E_1}, \mathbf{E_2}, \mathbf{E_3}, \mathbf{E_4}$ with $\mathbf{E_4} = \partial/\partial t$, along $\gamma(t)$ the variation will depend only on the components Z^m of the variation vector.

Proposition 2.3.6. *If there is no conjugate point to q in the interval (q,p), then the value of $d^2\Lambda/du^2|_{u=0}$ will be negative for all variations α of $\gamma(t)$ whose variation is orthogonal to the tangent vector to the curve, and not zero everywhere, nor proportional to the tangent vector. That is, there is no small variation of $\gamma(t)$ which gives a time-like curve between q and p.*

On the other hand:

Proposition 2.3.7. *If there is a point conjugate to q, r, in the set (q,p) then there will be a variation α which yields a time-like curve from q to p.*

In the case of a path from a space-like two-surface to a point, the previous proposition translate to the following two.

Proposition 2.3.8. *If $\gamma(t)$ is a null geodesic orthogonal to a space-like two surface $\mathscr{S}$, from the surface to a point p, and if there is no point in the set $[\mathscr{S}, p]$ conjugate to $\mathscr{S}$, then no small variation of γ can give a time-like curve from $\mathscr{S}$ to p.*

Proposition 2.3.9. *If there is a point conjugate to $\mathscr{S}$, r, in $(\mathscr{S}, p)$ then there will be a variation α which yields a time-like curve from $\mathscr{S}$ to p.*

Chapter 3

Causality

3.1 Orientability

We shall now introduce a new concept that is not the consequence of any direct geometric considerations. In our region of space-time, there is a well-defined arrow of time that arise directly from the second law of thermodynamics in a quasi-isolated system. Physically, it would appear reasonable to assume the existence of a local thermodynamic arrow of time defined continuously on all the points of the manifold. The only requirement is that it should be possible to define a continuous division of non-spacelike vectors in two classes, which we will arbitrarily label as past- and future-directed. In this case, we will say that the space-time is *time-oriented*. It is important to point out that not all the solutions to the Einstein field equations give a space-time with such a feature (such as in the case of the de Sitter universe). However, it is possible to remove this problem by simply not defining the point at which this condition does not hold. That is if a space-time $(\mathcal{M}, \mathbf{g})$ is not time-orientable, then it has a double covering space $(\widetilde{\mathcal{M}}, \mathbf{g})$ which is orientable. We shall define $\widetilde{\mathcal{M}}$ as the set of all pairs (p, α), where $p \in \mathcal{M}$ and α is one of the two possible orientation of time at p. Such a structure with the addition of the projection map $\pi : (p, \alpha) \to p$ we get that in fact $\widetilde{\mathcal{M}}$ is a double covering of $\mathcal{M}$. In general, we will have that either $\widetilde{\mathcal{M}}$ consists of two distinct components, and therefore $(\mathcal{M}, \mathbf{g})$ is time-orientable, or $\widetilde{\mathcal{M}}$ is connected and therefore $(\mathcal{M}, \mathbf{g})$ is not orientable, but $(\widetilde{\mathcal{M}}, \mathbf{g})$ is. From now on we shall consider either $(\mathcal{M}, \mathbf{g})$ time-orientable or its covering space. Proving the existence of singularities in this space implies their existence also in $(\mathcal{M}, \mathbf{g})$.

3.2 Causal curves

Assuming space-time to be time-orientable we shall consider two sets $\mathscr{S}$ and $\mathscr{U}$, defining the concept of *chronological future* $I^+(\mathscr{S}, \mathscr{U})$ *of* $\mathscr{S}$ *related to* $\mathscr{U}$ as the set of all points in $\mathscr{U}$ which can be reached from $\mathscr{S}$ by future directed time-like curves in $\mathscr{U}$. In the whole manifold the set $I^+(\mathscr{S}, \mathcal{M})$ will be denoted simply as $I^+(\mathscr{S})$ since given $p \in \mathcal{M}$ if it can be reached by a future-directed time-like curve from $\mathscr{S}$ then there is a small neighborhood of p which can be reached by such curves.

By replacing 'future' with 'past' and + by - in the previous definition, from the duality nature of the orientability, one obtains the definition of the $I^-(\mathscr{S}, \mathcal{M})$ set.

The *causal future of $\mathscr{S}$ related to $\mathscr{U}$* is denoted by $J^+(\mathscr{S}, \mathscr{U})$, and it will be the union of $\mathscr{S} \cap \mathscr{U}$ and the set of all the points in $\mathscr{U}$ reachable from $\mathscr{S}$ by a future-directed non-spacelike curve in $\mathscr{U}$. From section 2.3 we know that a non-spacelike curve, which is not a null geodesic can be deformed into a time-like curve between two points. Thus, if $\mathscr{U}$ is an open set and $p, q, r \in \mathscr{U}$ then either:

$$q \in J^+(p, \mathscr{U}), r \in I^+(q, \mathscr{U}),$$

or

$$q \in I^+(p, \mathscr{U}), r \in J^+(q, \mathscr{U}),$$

which implies that $r \in I^+(p, \mathscr{U})$. That is $\overline{I^+}(p, \mathscr{U}) = \overline{J^+}(p, \mathscr{U})$ and $\dot{I}^+(p, \mathscr{U}) = \dot{J}^+(p, \mathscr{U})$, where for a general set $\mathscr{H}$, $\overline{\mathscr{H}}$ denotes the closure of the original set and, $\dot{\mathscr{H}}$ denotes the boundary, such that:

$$\dot{\mathscr{H}} = \overline{\mathscr{H}} \cap \overline{(\mathcal{M} - \mathscr{H})}.$$

As in the previous case, $J^+(\mathscr{S}, \mathcal{M})$ will be written as $J^+(\mathscr{S})$, and it represents the region of space-time that can be causally affected by events in $\mathscr{S}$. Such a set may not be closed even if $\mathscr{S}$ is made of a single point.

In general, it is possible to construct a space-time, given a series of causal properties using the following method: one starts by considering some space-times, then cuts out any closed set and, if desired, pastes it together appropriately. The result of this procedure is still a manifold with a Lorentz metric, thus still a space-time, even though it might seem rather incomplete since some points had been cut out. Such an incompleteness can be resolved as said before, by applying an appropriate transformation that moves the cut-out points to infinity.

We shall define the *future horizons of $\mathscr{S}$ relative to $\mathscr{U}$* as $E^+(\mathscr{S}, \mathscr{U})$ that is $J^+(\mathscr{S}, \mathscr{U}) - I^+(\mathscr{S}, \mathscr{U})$, and similarly as before $E^+(\mathscr{S}, \mathcal{M})$ can be written as $E^+(\mathscr{S})$. In particular, if $\mathscr{U}$ is an open set then, by proposition 2.3.5, all the points in $E^+(\mathscr{S}, \mathscr{U})$ must lie on a future-directed null-geodesic from $\mathscr{S}$, at the same time if $\mathscr{U}$ is a convex normal neighborhood around p, then, from proposition 2.3.1, $E^+(p, \mathscr{U})$ consists of the future-directed null-geodesics in $\mathscr{U}$ from p and forms the boundary of both $I^+(p, \mathscr{U})$ and $J^+(p, \mathscr{U})$.

For the sake of simplicity, from now on we shall extend the definition of time-like and non-spacelike curves from differentiable piecewise to continuous differentiable. Although we will still say that it is non-spacelike if locally every two points of the curve can be joined by a piecewise differentiable non-spacelike curve. Given a continuous curve $\gamma : F \to \mathcal{M}$, where F is a continuous interval of $\mathbb{R}$, it will be future-directed and non-spacelike if, for every element of F, there exists a neighborhood $G \subset F$ of $t \in F$ and a convex normal neighborhood $\mathscr{U}$ of γ in $\mathcal{M}$ such that for any $t_1 \in G$, $\gamma(t_1) \in J^-(\gamma(t), \mathscr{U}) - \gamma(t)$ if $t_1 < t$ and $t_1 \in G$, $\gamma(t_1) \in J^+(\gamma(t), \mathscr{U}) - \gamma(t)$ if $t < t_1$.

On the other hand, we shall say that γ is future-directed and time-like if the same condition holds for I.

From now on, with time-like or non-spacelike curve, we will refer to a continuous curve. Two curves are considered to be equivalent if it is possible to find one from the reparametrization of the other.

That said, before giving an important result coming from these considerations, let us define as *future endpoint* p of a future-directed non-spacelike curve $\gamma : F \to \mathcal{M}$ a point such that for every neighborhood $\mathscr{V}$ of p there is a $t \in F$ such that $\gamma(t_1) \in \mathscr{V}$ for every $t_1 \in F$ with $t_1 \geq t$. Let now a non-spacelike curve be *future-inextendible* (*future-inextendible in $\mathscr{S}$*) if it has no future endpoint (no endpoint in $\mathscr{S}$). A point p will be a *limit point* of an infinite sequence of non-spacelike curve λ_n if every neighborhood of p intersects an infinite number of the λ_n, while a non-spacelike curve λ will be said to be a *limit curve* of the sequence λ_n if there is a sub-sequence λ'_n such that, for every $p \in \lambda_n$, λ'_n, converges to p.

Lemma 3.2.1. *Let $\mathscr{S}$ be an open set and λ_n an infinite sequence of non-spacelike curves in $\mathscr{S}$ which are future-inextendible in $\mathscr{S}$. If $p \in \mathscr{S}$ is a limit point of λ_n, then through that point there is a non-spacelike curve λ which is future-inextendible in $\mathscr{S}$ and which is a limit curve for λ_n.*

3.3 Achronal boundaries

By proposition 2.3.1 we know that, in a convex normal neighborhood $\mathscr{U}$, the boundary of $I^+(p, \mathscr{U})$ or $J^+(p, \mathscr{U})$ derives from the future-directed null geodesics arising from p. Let now a set $\mathscr{S}$ be *achronal* if $I^+(\mathscr{S}) \cap \mathscr{S}$ is empty, that is if there are no two points of $\mathscr{S}$ with a time-like separation. Meanwhile, a set $\mathscr{S}$ is said to be a *future set* if $\mathscr{S} \supset I^+(\mathscr{S})$. A set such as $\mathcal{M} - \mathscr{S}$ can be defined as a past set.

Proposition 3.3.1. *If $\mathscr{S}$ is a future set, then $\dot{\mathscr{S}}$ (the boundary of thi set) is a closed, imbedded, achronal three-dimensional C^{1-} submanifold.*

The boundary of a set that has the properties defined in the previous proposition is said to be an *achronal boundary*. These sets can be divided into four distinct subsets $\dot{\mathscr{S}}_N$, $\dot{\mathscr{S}}_+$, $\dot{\mathscr{S}}_-$, $\dot{\mathscr{S}}_0$, such that for a point $q \in \dot{\mathscr{S}}$ there may exist two points $p, r \in \dot{\mathscr{S}}$ with $p \in E^-(q) - q$, $r \in E^+(q) - q$. The different subsets can be defined depending on the existence of the two points, according to the following scheme:

	$\exists p$	$\nexists p$
$\exists r$	$\dot{\mathscr{S}}_N$	$\dot{\mathscr{S}}_-$
$\nexists r$	$\dot{\mathscr{S}}_+$	$\dot{\mathscr{S}}_0$

Given a point q we shall say that if $q \in \dot{\mathscr{S}}_N$, than $r \in E^+(p)$ since $r \in J^+(p)$ and $r \notin I^+(p)$, meaning that there is a null segment in $\dot{\mathscr{S}}$ through q. If $q \in \dot{\mathscr{S}}_+$ (or $\dot{\mathscr{S}}_-$) then q is the future (or past) endpoint of the null geodesics in $\dot{\mathscr{S}}$. Finally, $\dot{\mathscr{S}}_0$ is space-like (more strictly, acausal).

A useful condition on such sets arises from the following lemma, introduced by Penrose:

Lemma 3.3.1. *Let $\mathscr{W}$ be a neighborhood of $q \in \dot{\mathscr{S}}$, and the boundary of a future set, then:*

(i) $I^+(q) \subset I^+(\mathscr{S} - \mathscr{W})$ implies $q \in \dot{\mathscr{S}}_N \cup \dot{\mathscr{S}}_+$;

(ii) $I^-(q) \subset I^+(\mathcal{M} - \mathscr{S} - \mathscr{W})$ implies $q \in \dot{\mathscr{S}}_N \cup \dot{\mathscr{S}}_-$;

As an example of application of this result we shall consider $\dot{J}^+(\mathscr{H}) = \dot{I}^+(\mathscr{H})$, that is the boundary of the closed set $\mathscr{H}$. We know, by a previous result, it is an achronal manifold, using the above lemma we now know that every point of $\dot{J}(\mathscr{H}) - \mathscr{H}$ belongs to $[\dot{J}^+(\mathscr{H})]_N$ or $[\dot{J}^+(\mathscr{H})]_+$. This allows us to assume that $\dot{J}(\mathscr{H}) - \mathscr{H}$ is generated by null geodesic segments which may have future endpoints in $\dot{J}^+(\mathscr{H}) - \mathscr{H}$. However, if they do have past endpoints, the only possibility is to find them on $\mathscr{H}$ itself. This example is related to what Penrose called the plane wave solution.

Let us now define a *causally simple* set as an open set $\mathscr{U}$ which for every compact set $\mathscr{H} \subset \mathscr{U}$,

$$\dot{J}^+(\mathscr{H}) \cap \mathscr{U} = E^+(\mathscr{H}) \cap \mathscr{U}\,,$$

and

$$\dot{J}^-(\mathscr{H}) \cap \mathscr{U} = E^-(\mathscr{H}) \cap \mathscr{U}\,.$$

That is equivalent to say that $\dot{J}^+(\mathscr{H})$ and $\dot{J}^-(\mathscr{H})$ are closed in $\mathscr{U}$.

3.4 Causality conditions

In Chapter 1 principia *(i)* required that the causality should have held only locally, leaving the discussion on the global equations open. Thus, it may seem possible to

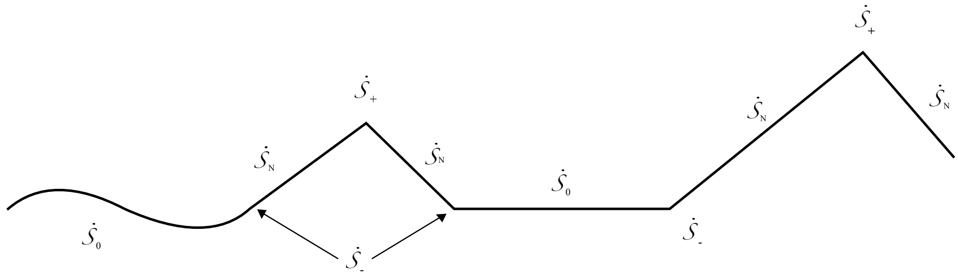

Figure 3.1: *The figure shows how an achronal boundary $\dot{\mathscr{S}}$ can be divided into the four sets just defined. Where $\dot{\mathscr{S}}_N$ is null-like, $\dot{\mathscr{S}}_0$ time-like, and $\dot{\mathscr{S}}_+$, $\dot{\mathscr{S}}_-$ are the two future and past endpoints, respectively, of a null geodesic in the set.*

construct closed time-like curves. However, the existence of such curves would lead to some logical paradoxes. For example, one could imagine that with a rocketship one could travel around such a curve, arriving before he had set off, and preventing himself from even starting the trip. To avoid the necessity of changing our philosophical definition of free will we shall introduce what is called the *chronology condition* which states that there are no closed time-like curves. However, there might be some points of space-time at which this condition do not hold, the set of such points will be the *chronology violating* of $\mathcal{M}$ with the following features:

Proposition 3.4.1 (Carter)**.** *The chronology violating set of $\mathcal{M}$ is a disjoint union of sets of the form $I^+(q) \cap I^-(q) \in \mathcal{M}$.*

Proposition 3.4.2. *If $\mathcal{M}$ is compact, the chronology violating set of $\mathcal{M}$ is non-empty.*

From this result, it would seem reasonable to assume that space-time is non-compact, in addition to the fact that any compact four-dimensional manifold, on which is defined a Lorentz metric, cannot be simply connected.

The *causality condition* holds, therefore, if there are no closed non-spacelike curves, which is similar to what we said in proposition 3.4.2.

Proposition 3.4.3. *The set of points at which the causality does not hold is the disjoint union of sets of the form $J^-(q) \cap J^+(q)$, $q \in \mathcal{M}$.*

If the causality condition is violated at $q \in \mathcal{M}$, but the chronology condition holds, then there must be a closed null geodesic curve γ through q. Let v be an affine parameter for the curve, and $\dots, v_{-1}, v_0, v_1, \dots$ be successive values of v at q, then it is possible to compare the tangent vector $\partial/\partial v|_{v=v_0}$ with $\partial/\partial v|_{v=v_1}$, parallel transporting $\partial/\partial v|_{v=v_0}$ round γ. Since they point in the same direction, then they must be proportional to each other: $\partial/\partial v|_{v=v_1} = a(\partial/\partial v)|_{v=v_0}$ where the factor a is defined by saying that, given the affine distance covered over the nth circuit of γ, $(v_{n+1} - v_n)$, it is equal to $a^{-n}(v_1 - v_0)$.

Thus, if $a > 1$, v never attains the value $(v_1 - v_0)(1 - a^{-1})^{-1}$ and therefore the curve is geodetically incomplete in the future direction, even though it is possible to go around it an infinite number of times. On the other hand, if $a < 1$, γ is incomplete in the past direction, while for $a = 1$ it is incomplete in both directions.

Moreover, from the following proposition we can obtain another meaning of the a factor.

Proposition 3.4.4. *If γ is a closed null geodesic curve, incomplete in the future direction, then there is a variation of γ which moves each point of γ toward the future and which yields a closed time-like curve.*

Proposition 3.4.5. *If:*

(a) $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ *for every null vector* $\mathbf{K}$*;*

(b) the generic condition holds (meaning that every null geodesic contain a point at which $K_{[\alpha}R_{\beta]\mu\nu[\rho}K_{\delta]}K^\mu K^\nu \neq 0$*);*

(c) the chronology condition holds on $\mathcal{M}$*,*

then the causality condition holds on $\mathcal{M}$*.*

Showing that for physically realistic solutions, the causality and chronology condition are equivalent.

In addition to ruling out closed non-spacelike curves, it would seem reasonable to exclude situations with non-spacelike curves which return arbitrarily close to the origin point, or to curves which return arbitrarily close to the origin point of the first. There seems to be more than a countably infinite hierarchy of such higher degree causality conditions depending on the number and order of the limiting processes involved. Our interest will remain on the first three of these conditions, giving the ultimate causality condition at the end.

The *future (past) distinguishing condition* holds at a point $p \in \mathcal{M}$ if every neighborhood of such point contains a neighborhood of p which no future (past) directed non-spacelike curves from p intersects more than once. That is, if $I^+(q) = I^+(p)$ $(I^-(q) = I^-(p))$ implies that $q = p$.

The *strong causality condition* is said to hold at p if every neighborhood of p contains a neighborhood of p in which no non-spacelike curve intersects more than once.

Proposition 3.4.6. *If the conditions (a) to (c), in proposition 3.4.5 hold and if we add a fourth condition (d) which states that* $\mathcal{M}$ *is null and geodesically complete, then the strong causality condition holds on* $\mathcal{M}$*.*

Corollary 3.4.0.1. *The past and future distinguishing condition would also hold on* $\mathcal{M}$ *since they are implied by the strong causality.*

A connected concept to the three higher degree causality condition is the phenomenon of imprisonment. A non-spacelike curve γ, that is future inextendible can behave in one of three ways:

(i) enter and remain within a compact set $\mathscr{S}$,

(ii) not remain within any compact set, but continually re-enter a compact set $\mathscr{S}$,

(iii) not remain within any compact set $\mathscr{S}$ and not re-enter any such set more than a finite amount of times.

In (iii) γ can be regarded as going off to the edge of space-time, that is either infinite or a singularity, while in the first and second case, we shall say that γ is *totally* or *partially future imprisoned* in $\mathscr{S}$, respectively.

Proposition 3.4.7. *If the strong causality condition holds on the compact set $\mathscr{S}$, then there can be no future-inextendible, non-spacelike curve totally or partially future imprisoned in $\mathscr{S}$.*

Proposition 3.4.8. *If the future or past distinguishing condition holds on a compact set $\mathscr{S}$, then there can be no future-inextendible, non-spacelike curve totally future imprisoned in $\mathscr{S}$.*

The causality relations on $(\mathcal{M}, \mathbf{g})$ may be used to generate a topology on $\mathcal{M}$ defined as *Alexandrov Topology* and such that a set is defined to be open if, and only if, it happens to be the union of one or more sets of the form $I^+(p) \cap I^-(q)$, with p, $q \in \mathcal{M}$. Since $I^+(p) \cap I^-(q)$ is an open set in the manifold topology, then any open set in the Alexandrov topology will be open in the manifold topology, but the opposite is not necessarily true.

However, if the strong causality condition holds on $\mathcal{M}$, then around any point $r \in \mathcal{M}$ one can find a local causality neighborhood $\mathscr{U}$. In these terms the Alexandrov topology of $(\mathscr{U}, \mathbf{g}|_{\mathscr{U}})$ may be regarded as space-time on its own, and clearly the same as the manifold topology of $\mathscr{U}$, thus the Alexandrov topology of $\mathcal{M}$ is the same as the manifold topology since $\mathcal{M}$ can be covered by a local causality neighborhood. This means that, if the strong causality condition holds, one can determine the topological structure of space-time by the observation of causal relationships.

Even by imposing the strong causality condition, it is still possible to have a space-time that is on the verge of violating the chronology condition in which the slightest variation of the metric can lead to closed time-like curves. Because General Relativity is presumably the classical limit of some, yet to-discover, quantum theory of space-time, in such a theory the Uncertainty principle would prevent the metric from having an exact value at every point. Thus, we need that any property of our space-time, in order to be physically significant, and must have some sort of stability, meaning it should be a property of the *nearby* space-time. To give a more precise meaning to *nearby* we need to define a topology on the set of all space-times, that is, all the non-compact four-dimensional manifolds and all the Lorenz metrics on them.

We shall leave the problem of uniting in one connected topology space manifold of different topologies, and focus only on putting a topology on the set of all C^r Lorenz metrics ($r \geq 1$). This can be done in many ways, depending on whether one requires a *nearby* metric to be nearby just its value (C^0 topology) or also to its kth derivative (C^k topology) and to whether one requires it to be nearby everywhere (open topology) or only on compact sets (compact open topology).

For our purposes, we will be interested in a C^0 open topology, which may be defined as the symmetric tensor spaces $T^0_{S2}(p)$ (type (0,2) tensor) at every point $p \in \mathcal{M}$ form a manifold $T^0_{S2}(\mathcal{M})$, that is the bundle of symmetric tensors type (0,2) over $\mathcal{M}$. A Lorentz metric $\mathbf{g}$ on $\mathcal{M}$ is the value of a map $\hat{g} : \mathcal{M} \to T^0_{S2}$ assigning to each point of the manifold an element of T^0_{S2}. Applying the projection π to such a map, we get $\pi \circ \hat{g} = 1$, where $\pi : T^0_{S2} \to \mathcal{M}$, which sends $x \in T^0_{S2}(p)$ to p. Letting now $\mathscr{U}$ be an open set in $T^0_{S2}(\mathcal{M})$ and $O(\mathscr{U})$ the set of all C^0 Lorentz metrics $\mathbf{g}$ such that $\hat{g}(\mathcal{M})$ is contained in $\mathscr{U}$, then the open sets in the C^0 open topology of the C^r Lorentz metrics on $\mathcal{M}$ are defined to be the union of one or more sets of the form $O(\mathscr{U})$.

If the space-time metric $\mathbf{g}$ has an open neighborhood in the C^0 open topology such that there are no closed time-like curves in any metric belonging to the neighborhood, then we shall say that the *stable causality condition* holds on $\mathcal{M}$. (This condition can be generalized to C^r topology, but not to compact ones, since in them each neighborhood of any metric contains closed time-like curves.) In other words, one can slightly expand the light cones at every point without introducing closed time-like curves.

Proposition 3.4.9. *The stable causality condition holds everywhere on $\mathcal{M}$ if, and only if, there is a function f on $\mathcal{M}$ whose gradient is everywhere time-like.*

The function f may be regarded as a sort of cosmic time, in the sense that it increases along every future-directed non-spacelike curve. This means that the space-like surfaces $\{f = \text{constant}\}$ may be thought of as non-unique simultaneity surfaces, whose compactness induces a diffeomorphism between them. This condition is not necessarily true if some of them are not compact.

3.5 Cauchy developments

In Newton's theory, there is an instantaneous action at a distance, and in order to predict future events one has to know the state of the entire universe at the present time, as well as assume some boundary conditions at infinity, such as a potential which goes to zero. On the other hand, in a relativistic theory, from postulate (i) in Chapter 1, we know that events at different points of space-time are causality connected only if there is a non-spacelike curve that goes from a point to the other. Thus, a knowledge of the appropriate data on a closed set $\mathscr{S}$ would determine events in a region $D^+(\mathscr{S})$ to the future of $\mathscr{S}$ called the *future Cauchy development* or *domain of dependence* of $\mathscr{S}$, and it is defined as the set of all points $p \in \mathcal{M}$ such that every past-inextendible non-spacelike curve through p intersects $\mathscr{S}$. The definition given by Penrose of the Cauchy development of $\mathscr{S}$, however, is slightly different from the previous and states that it is the set of all points $p \in \mathcal{M}$ such that every past-inextendible time-like curve through p intersects $\mathscr{S}$. We shall denote such a set as $\widetilde{D}^+(\mathscr{S})$.

Proposition 3.5.1. $\widetilde{D}^+(\mathscr{S}) = \overline{D^+}(\mathscr{S})$.

The future boundary of $D^+(\mathscr{S})$, $\overline{D^+}(\mathscr{S}) - I^-(D^+(\mathscr{S}))$, marks the limit of the region predictable from the knowledge of data on $\mathscr{S}$. This set will be the *future Cauchy horizon* of $\mathscr{S}$ and will be denoted as $H^+(\mathscr{S})$. In addition to that, it will intersect $\mathscr{S}$ if it is null or if it has an *edge*, that is the set of all points $p \in \mathscr{S}$ for an achronal set, such that in every neighborhood $\mathscr{U}$ of q there are points $p \in I^-(\shortparallel, \mathscr{U})$, and $r \in I^+(\shortparallel, \mathscr{U})$ which can be joined by a time-like curve in $\mathscr{U}$ not intersecting $\mathscr{S}$. By a similar argument to proposition 3.3.1, it follows that if edge($\mathscr{S}$) is empty for a non-empty achronal set $\mathscr{S}$, then it is a three-dimensional imbedded C^{1-} submanifold.

Proposition 3.5.2. *For a closed achronal set $\mathscr{S}$,*

$$edge(H^+(\mathscr{S})) = edge(\mathscr{S}).$$

Proposition 3.5.3. *Let $\mathscr{S}$ be a closed achronal set, then $H^+(\mathscr{S})$ is generated by null geodesic segments, which either have no past endopoints or have past endopoints at edge($\mathscr{S}$).*

Corollary 3.5.0.1. *If edge($\mathscr{S}$) vanishes, then $H^+(\mathscr{S})$, if non-empty, is an achronal three-dimensional imbedded C^{1-} manifold which is generated by null geodesic segments which have no past endopoints.*

We shall call an acasual set $\mathscr{S}$ with no edge, a *partial Cauchy surface* that is a space-like hypersurface intersected by a non-spacelike curve not more than once. Supposing there is a connected space-like hypersurface $\mathscr{S}$ (with no edges) which some non-spacelike curve γ intersects at point p_1 and p_2, then one could join p_1 and p_2 by a curve μ in $\mathscr{S}$ such that $\mu \cup \lambda$ would be a closed curve crossing $\mathscr{S}$ only once. This curve could not be continuously deformed to zero since such a deformation could add an even number of points to the number of times it crosses $\mathscr{S}$. Thus, $\mathcal{M}$ may not be simply connected, meaning that we could unwrap $\mathcal{M}$ by going to the simply connected universal covering manifold $\widetilde{\mathcal{M}}$, in which every connected component of the image of $\mathscr{S}$ is a space-like hypersurface and therefore a partial Cauchy surface in $\widetilde{\mathcal{M}}$. However, it is possible that using the universal covering manifold may unwrap $\mathcal{M}$ more than required, resulting in a non-compact partial Cauchy surface, even though $\mathscr{S}$ was compact.

For the purposes of our analysis, we want a covering manifold which unwraps $\mathcal{M}$ sufficiently so that each connected component of the image of $\mathscr{S}$ was a partial Cauchy surface, but so that every component remains homeomorphic to $\mathscr{S}$. This can be obtained by recalling that the universal covering manifold may be defined as the set of all pairs of the form $(p, [\lambda])$ with $p \in \mathcal{M}$, and $[\lambda]$ is an equivalent class of curves in $\mathcal{M}$, from q to p, and homotopic modulo q to p. In addition to that, we shall define the covering manifold $\mathcal{M}_H$ as the set of all pairs $(p, [\lambda])$ where now $[\lambda]$ is an equivalent class of curves from $\mathscr{S}$ to p, and it is the largest covering manifold such that each component of the image of $\mathscr{S}$ is homeomorphic to $\mathscr{S}$. On the other hand, we can define a different covering manifold $\mathcal{M}_G$ defined as the set of all pairs $(p, [\lambda])$ where $[\gamma]$ is an equivalent class of curves from q to p passing through $\mathscr{S}$ the same number of times, in which crossing in the future direction is considered positive while crossing in the past direction is considered to be negative. In these regards, we can define $\mathcal{M}_G$ as the smallest covering manifold in which each connected component of the image of $\mathscr{S}$ divides the manifold into two parts. In each case, the topology and differential structures of the covering manifold are fixed by the requirement that the projection mapping $(p, [\lambda])$ into p is locally a diffeomorphism.

Now, defining $D(\mathscr{S}) = D^+(\mathscr{S}) \cup D^-(\mathscr{S})$, we shall call a local Cauchy surface *global Cauchy surface* (simply *Cauchy surface*) if $D(\mathscr{S})$ is equal to $\mathcal{M}$. That is a space-like hypersurface that every non-spacelike curve intersects exactly once. An example of Cauchy surfaces is $\{x^4 = \text{constant}\}$ in the Minkowski space, defined as hyperboloids which are only partial Cauchy surfaces since the past and future null cones of the origin are Cauchy horizon for these surfaces.

If there is a Cauchy surface for $\mathcal{M}$, then one may predict the state of the universe at any time by knowing the relevant data on the surface. However, one could not know the data unless one was at the future of every point on the surface, which would be impossible in most cases. Thus, in General Relativity one's ability to predict the future is limited both by the difficulty of knowing data on the whole of a space-like surface and by the possibility that, even if one did, it would still be insufficient.

3.6 Global hyperbolicity

A concept very close to the Cauchy development is the global hyperbolicity, which is a property of a set $\mathscr{N}$ which arises if the strong causality condition holds on it and if, for any two points $p, q \in \mathscr{N}$, the set $J^+(p) \cap J^-(q)$ is compact and contains $\mathscr{N}$ itself. This can be thought of as saying that $J^+(p) \cap J^-(q)$ does not contain any points on the edge of space-time. The name of such a property comes from the fact that the wave equation for a $\delta-$function source at some point $p \in \mathcal{M}$ has a unique solution vanishing outside $\mathscr{N} - J^+(p, \mathscr{N})$.

We call *causally simple* a set $\mathscr{N}$ if, for every compact set $\mathscr{H}$ contained in $\mathscr{N}$, $J^+(p) \cap \mathscr{N}$ and $J^-(q) \cap \mathscr{N}$ are closed in $\mathscr{N}$.

Proposition 3.6.1. *An open globally hyperbolic set $\mathscr{N}$ is causally simple.*

Corollary 3.6.0.1. *If $\mathscr{H}_1$ and $\mathscr{H}_2$ are compact sets in $\mathscr{N}$, then $J^+(\mathscr{H}_1) \cap J^-(\mathscr{H}_2)$ is compact.*

A different way to define the global hyperbolicity starts by considering the following statements: for two points $p, q \in \mathcal{M}$ such that the strong causality condition holds on $J^+(p) \cap J^-(q)$, we define $C(p, q)$ to be the space of all (continuous) non-spacelike curves from p to q. Given two curves $\gamma(t)$ and $\lambda(u)$ they will represent the same point of $C(p, q)$ if one is the reparametrization of the other, *i.e.* if there is a continuous monotonic function $f(u)$ such that $\gamma(f(u)) = \lambda(u)$. The topology of such a space $(C(p, q))$ is defined by saying that a neighborhood of γ in $C(p, q)$ consists of all the curves in $C(p, q)$ whose points in $\mathcal{M}$ lie in a neighborhood $\mathscr{W}$ of the points of γ in $\mathcal{M}$. In these terms, the other definition of global hyperbolicity states that an open set $\mathscr{N}$ is globally hyperbolic if $C(p, q)$ is compact for all $p, q \in \mathscr{N}$.

Even though they are different from each other, these definitions are equivalent, as it is shown by the following results.

Proposition 3.6.2. *Let $\mathscr{N}$ be an open set on which holds the strong causality condition, and such that*

$$\mathscr{N} = J^-(\mathscr{N}) \cap J^+(\mathscr{N}),$$

then $\mathscr{N}$ is globally hyperbolic if, and only if, $C(p, q)$ is compact for all $p, q \in \mathscr{N}$.

The relationship between the global hyperbolicity and Cuchy development is given by the following results:

Proposition 3.6.3. *If $\mathscr{S}$ is a closed achronal set, then int$(D(\mathscr{S})) \doteq D(\mathscr{S}) - \dot{D}(\mathscr{S})$, if not empty, is globally hyperbolic.*

Even if we will not go over the details of the proof, we still introduce a series of important lemmas:

Lemma 3.6.1. *If $p \in D^+(\mathscr{D}) - H^+(\mathscr{S})$, then every past-inextendible non-spacelike curve through p intersects $I^-(\mathscr{S})$.*

Corollary 3.6.1.1. *If $p \in$ int$(D(\mathscr{S}))$ then every inextendible non-spacelike curve through p intersects $I^-(\mathscr{S})$ and $I^+(\mathscr{S})$.*

Lemma 3.6.2. *The strong causality condition holds on int$(D(\mathscr{S}))$.*

Proposition 3.6.4. *If $q \in$ int$(D(\mathscr{S}))$, then $J^+(\mathscr{S} \cap J^-)$ is compact or empty.*

To show that the global hyperbolicity holds on $D(\mathscr{S})$ and not only on its interior, we need to introduce some extra conditions.

Proposition 3.6.5. *If $\mathscr{S}$ is a closed achronal set such that $J^+(\mathscr{S}) \cap J^-(\mathscr{S})$ is both strongly causal and either*

(1) acausal, meaning that only $\mathscr{S}$ is casual, or

(2) compact,

then $D(\mathscr{S})$ is globally hyperbolic.

By proposition 3.6.3 we know that if for an open set $\mathscr{N}$ there existes a Cauchy surface, then it implies the global hyperbolicity of $\mathscr{N}$. And the following proposition shows that the opposite is also true.

Proposition 3.6.6. *If an open set $\mathscr{N}$ is globally hyperbolic, then, regarding it as a manifold, $\mathscr{N}$ is homeomorphic to $\mathbb{R} \times \mathscr{S}$ where $\mathscr{S}$ is a three-dimensional manifold, and for each $a \in \mathbb{R}$, $\{a\} \times \mathscr{S}$ is a Cauchy surface for $\mathscr{N}$.*

3.7 The existence of geodesics

The crucial importance of global hyperbolicity in Chapter 4 lies in the following result:

Proposition 3.7.1. *Let p and q lie in a globally hyperbolic set $\mathscr{N}$ with $q \in J^+(p)$, then there is a non-spacelike geodesics from p to q whose length is greater or equal to that of any other non-spacelike curve from p to q.*

Proof. We shall consider a general proof, introduced by Avez and Seifert, using the compactness of $C(p, q)$.

Let us consider $C(p, q)$ and a dense subset $C'(p, q)$ containing all the time-like C^1 curves from p to q. The length of one of these curves λ is defined as

$$L[\lambda] = \int_p^q (-g(\partial/\partial t, \partial/\partial t))^{1/2} dt,$$

where t is a C^1 parameter on λ. The function L is not continuous on $C'(p, q)$ since any neighborhood of λ contains a zigzag piecewise almost null curve of arbitrarily small length. The reason behind this lack of continuity is that we used a C^0 topology in which two curves are close to each other if their points in $\mathcal{M}$, but not necessarily their vectors, are close too. We could use a C^1 topology on $C'(p, q)$ to make L continuous, but this would generate problems of its own because in these cases $C'(p, q)$ will not be compact, in fact one has a compact space only when includes all the continuous non-spacelike curves.

To solve the problem of continuity we shall, instead, use a C^0 topology, and extend the definition of L to $C(p,q)$.

The signature of the metric, making a time-like curve wiggle, actually reduces its length, thus L is not lower semi-continuous, however:

Lemma 3.7.1. *L is upper semi-continuous in the C^0 on $C'(p,q)$.*

By considering a neighborhood $\mathscr{U}$ of a continuous non-spacelike curve λ in $\mathcal{M}$ and letting $l(\mathscr{U})$ be the least upper bound of the length of time-like curves in $\mathscr{U}$, we shall define $L[\gamma]$ as the greatest lower bound of $l(\mathscr{U})$ for all neighborhood $\mathscr{U}$ of λ in $\mathcal{M}$. That is the definition of the length of the curve λ, which works for all curves λ between p and q having a C^1 time-like curve in every neighborhood. That is for all the points in $C(p,q)$ lying in the closure of $C'(p,q)$. From what we have found in Chapter 2, a non-spacelike curve from p to q, which is not an unbroken null geodesic curve, can be variated to give a piecewise C^1 time-like curve from p to q, and by rounding off the corners of such curve, we get a C^1 time-like curve from p to q. Thus, points in $C(p,q) - \overline{C'(p,q)}$ are unbroken null geodesics, and we define their length to be zero.

This definition of L makes it an upper semi-continuous function on the compact space $\overline{C'(p,q)}$ (more precisely it makes it a continuous non-spacelike curve which satisfies a local Lipschitz condition, and it is differentiable almost everywhere). If such a space is empty, but $C(p,q)$ is not, p and q are joined by an unbroken null geodesic and there are no non-spacelike curves from p to q that are not unbroken null geodesics. On the other hand, if $\overline{C'(p,q)}$ is not empty, then it will contain some point at which L attains its maximum value, meaning that there is a curve γ from p to q whose length is greater or equal to the one of any other curve of this kind. By proposition 2.3.2, γ must be a geodesic curve, as otherwise, one could find points $x, y \in \gamma$ which lay in a convex normal coordinate neighborhood and could be joined by a geodesic segment of greater length than the partition of such a curve between x and y. □

Corollary 3.7.1.1. *If $\mathscr{S}$ is a C^2 partial Cauchy surface, then to each point $q \in D^+(\mathscr{S})$ there is a future-directed time-like geodesic curve orthogonal to $\mathscr{S}$ of length $d(\mathscr{S}, q)$ which does not contain any point conjugate to $\mathscr{S}$ between $\mathscr{S}$ and q.*

3.8 The causal boundary of space-time

In this section, we shall discuss a method (created by Geroch, Kronheimer, and Penrose) for attaching a boundary to space-time, whose construction depends only on the causal structure of $(\mathcal{M}, \mathbf{g})$. This means that it does not distinguish between boundary points at a finite distance (singular point) and boundary points at infinity.

We shall assume that $(\mathcal{M}, \mathbf{g})$ satisfies the strong causality condition, then any point p in $(\mathcal{M}, \mathbf{g})$ is uniquely determined by its chronological past $I^-(p)$ or future $I^+(p)$. By defining the chronological past as $\mathscr{W} \doteq I^-(p)$ we shall assume it to have the following properties:

(1) $\mathscr{W}$ is open;

(2) $\mathscr{W}$ is a past set $I^-(\mathscr{W}) \subset \mathscr{W}$;

(3) $\mathscr{W}$ cannot be expressed as the union of two proper subsets which have properties (1) and (2).

A set with these properties will be called *indecomposable past set*, IP in short. The *indecomposable future set*, IF in short, is defined by simply changing past with future in the previous three properties.

It is possible to divide IPs into two different classes: the proper IPs (PIPs) which are the past of points in $\mathcal{M}$, and the terminal IPs (TIPs) which are not the past of any point in $\mathcal{M}$. Now we can consider these TIPs as representing points of the causal boundary (*c-boundary*) of $(\mathcal{M}, \mathbf{g})$. Another way of seeing it is by considering the TIPs as the past of future-inextendible time-like curves, thus, one can regard the past $I^-(\gamma)$ of a future-directed curve γ as representing the future endpoint of γ on the c-boundary. Moreover, we shall say that another curve γ' has the same endpoint if, and only if, $I^-(\gamma) = I^-(\gamma')$.

Proposition 3.8.1. *A set $\mathscr{W}$ is a TIP if, and only if, there is a future-inextendible time-like curve γ such that $I^-(\gamma) = \mathscr{W}$.*

Let us denote with $\hat{\mathcal{M}}$ the set of all IPs of the space $(\mathcal{M}, \mathbf{g})$, then it represents the points of $\mathcal{M}$ plus a future c-boundary; similarly, let $\check{\mathcal{M}}$ be the set of all IFs of $(\mathcal{M}, \mathbf{g})$, representing $\mathcal{M}$ plus a past c-boundary. These sets can be considered as an extension of I, J, and E in the sense that, for each $\mathscr{U}\, \mathscr{V} \subset \hat{\mathcal{M}}$ we have

$$\begin{aligned}
&\mathscr{U} \in J^-(\mathscr{V}, \hat{\mathcal{M}}) \quad \text{if} \quad \mathscr{U} \subset \mathscr{V};\\
&\mathscr{U} \in I^-(\mathscr{V}, \hat{\mathcal{M}}) \quad \text{if} \quad \mathscr{U}\, I^-(q) \text{ for some point } q \in \mathscr{V};\\
&\mathscr{U} \in E^-(\mathscr{V}, \hat{\mathcal{M}}) \quad \text{if} \quad \mathscr{U} \in J^-(\mathscr{V}, \mathcal{M}) \text{ but not} \mathscr{U} \in I^-(\mathscr{V}, \hat{\mathcal{M}}).
\end{aligned}$$

Because of this structure, the IP-space $\hat{\mathcal{M}}$ is a causal space, and there exists a natural injective map $I^- : \mathcal{M} \to \hat{\mathcal{M}}$ which sends a point $p \in \mathcal{M}$ into $I^-(p) \in \hat{\mathcal{M}}$. Such a map is an isomorphism of the causal relation J^- as $p \in J^-(q)$ if, and only if, $I^-(q \in J^-(I^-(q), \hat{\mathcal{M}}))$. In addition to that the causality relation is preserved by I^-, but not in the opposite direction. Proceeding similarly it is possible to define the causal relations on $\check{\mathcal{M}}$.

What we want to do is use the two spaces $\hat{\mathcal{M}}$ and $\check{\mathcal{M}}$ to define a third space $\mathcal{M}^*$ which has the form $\mathcal{M} \cup \Delta$, where Δ is the c-boundary of $(\mathcal{M}, \mathbf{g})$. To do so, let us start by defining a space $\mathcal{M}^{\#} \doteq \hat{\mathcal{M}} \cup \check{\mathcal{M}}$, with each PIF identified with the corresponding

PIP. In other words, $M^{\#}$ corresponds to the points of $\mathcal{M}$ together with the TIPs and TIFs.

A basis for the topology of the topological space $\mathcal{M}$ is provided by sets of the form $I^{+}(p) \cap I^{-}(q)$. Unfortunately, one cannot use a similar approach to define a basis for the topology of $\mathcal{M}^{\#}$ as there may be some points that are not in the chronological past of any point of this set. However, it is possible to obtain a topology of $\mathcal{M}$ from a sub-basis consisting of sets of the form $I^{+}(p)$, $I^{-}(p)$, $\mathcal{M} - \overline{I^{+}}(p)$, and $\mathcal{M} - \overline{I^{-}}(p)$. Using this analogy Penrose has shown how one can define a topology on $\mathcal{M}^{\#}$. For any IF $\mathscr{A} \in \check{\mathcal{M}}$, one can define

$$\mathscr{A}^{\text{int}} \doteq \{\mathscr{V} : \mathscr{V} \in \hat{\mathcal{M}} \text{ and } \mathscr{V} \cap \mathscr{A} \neq \emptyset\},$$

and

$$\mathscr{A}^{\text{ext}} \doteq \{\mathscr{V} : \mathscr{V} \in \hat{\mathcal{M}} \text{ and } \mathscr{V} = I^{-}(\mathscr{W}) \Rightarrow I^{-}(\mathscr{W}) \not\subset \mathscr{A}\}.$$

For an IF $\mathscr{B} \in \hat{\mathcal{M}}$ the two sets $\mathscr{B}^{\text{int}}$ and $\mathscr{B}^{\text{ext}}$ are defined similarly, thus the open sets of $\mathcal{M}^{\#}$ are defined to be the union and finite intersection of sets of the form $\mathscr{A}^{\text{int}}$, $\mathscr{A}^{\text{ext}}$, $\mathscr{B}^{\text{int}}$, and $\mathscr{B}^{\text{ext}}$. In these manners, the sets $\mathscr{A}^{\text{int}}$ and $\mathscr{A}^{\text{ext}}$ are the analog of $I^{+}(q)$ and $I^{-}(q)$ in $\mathcal{M}^{\#}$. In particular if $\mathscr{A} = I^{+}(p)$ and $\mathscr{V} = I^{-}(q)$ then $\mathscr{V} \in \mathscr{A}^{\text{int}}$ if, and only if, $q \in I^{+}(p)$. However, by the definition, it is also possible to incorporate TIPs into $\mathscr{A}^{\text{int}}$. In the same way, the sets $\mathscr{A}^{\text{ext}}$ and $\mathscr{B}^{\text{ext}}$ are the analogues of $\mathcal{M} - \overline{I^{+}}(p)$ and $\mathcal{M} - \overline{I^{-}}(q)$.

Finally, one can obtain $\mathcal{M}^{*}$ identifying the smallest number of points in the space $\mathcal{M}^{\#}$ necessary to make it a Hausdorff space. More precisely $\mathcal{M}^{*}$ is the quotient space $\mathcal{M}^{\#}/R_h$, where R_h is the intersection of all equivalence relation $R \subset \mathcal{M}^{\#} \times \mathcal{M}^{\#}$ for which $\mathcal{M}^{\#}/R$ is Hausdorff. The space $\mathcal{M}^{*}$ has a topology induced form $\mathcal{M}^{\#}$ which agrees with the topology of $\mathcal{M}$ on the subset $\mathcal{M}$ of $\mathcal{M}^{*}$.

3.9 Asymptotically simple spaces

In order to study bounded physical systems such as stars, we need to investigate spaces which are defined to asymptotically flat, *i.e.* the metric approaches Minkowski's at large distances from the system.

Let us consider the validity of the strong causality condition, even if in the original Penrose definition it was not required. However, it will not generate a loss in generality in the situation we will consider, allowing to a simplification of the matter.

Let us say that a time- and space-orientable space $(\mathcal{M}, \mathbf{g})$ is said to be *asymptotically simple* if there exists a strong causal space $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ and an imbedding $\theta : \mathcal{M} \to \widetilde{\mathcal{M}}$ which imbeds $\mathcal{M}$ as manifold with smooth boundary $\partial\mathcal{M}$ in $\widetilde{\mathcal{M}}$ such that:

(1) there is a smooth (say C^8 at least) function Ω on $\widetilde{\mathcal{M}}$ such that on $\theta(\mathcal{M})$, Ω is positive and $\Omega^2 \mathbf{g} = \theta_*(\tilde{\mathbf{g}})$, meaning $\tilde{\mathbf{g}}$ is conformal to $\mathbf{g}$ on $\theta(\mathcal{M})$;

(2) on $\partial\mathcal{M}$, $\Omega = 0$ and $d\Omega \neq 0$;

(3) every null geodesic in $\mathcal{M}$ has two endpoints on $\partial\mathcal{M}$.

We shall define $\overline{\mathcal{M}} \doteq \mathcal{M} \cup \partial\mathcal{M}$, which is a rather more general definition than what we wanted, since it includes cosmological models. In order to restrict it to spaces which are asymptotically flat spaces we will say that a space $(\mathcal{M}, \mathbf{g})$ is *asymptotically empty and simple* if satisfies conditions (1), (2), (3), and:

(4) $R_{\alpha\beta} = 0$ on an open neighborhood of $\partial\mathcal{M}$ in $\overline{\mathcal{M}}$.

We can consider the boundary $\partial\mathcal{M}$ as being at infinity, in the sense that any affine parameter in the metric $\mathbf{g}$ on a null geodesic in $\mathcal{M}$, may attain unboundedly large values near $\partial\mathcal{M}$. That is because an affine parameter v of a metric $\mathbf{g}$ is related to the parameter $\tilde{v}$ of the metric $\tilde{\mathbf{g}}$ by the transofrmation $dv/d\tilde{v} = \Omega^{-2}$. Since $\Omega = 0$ at $\partial\mathcal{M}$ then $\int dv$ diverges.

Moreover, from condition (2) and (4) it follows that the boundary $\partial\mathcal{M}$ is a null hypersurface.

In the Minkowski space, $\partial\mathcal{M}$ contains the two null surfaces $\mathscr{I}^+$ and $\mathscr{I}^-$, each with a topology of the form $\mathbb{R}\times S^2$. To show that a similar structure exists in any asymptotically simple and empty space, let us start from the following considerations. Because $\partial\mathcal{M}$ is a null surface, we have that $\mathcal{M}$ lies locally on to its past or future, showing that $\partial\mathcal{M}$ must contain two diconnected components. The first is $\mathscr{I}^+$ on which null geodesics in $\mathcal{M}$ have their future endpoints, and the second is $\mathscr{I}^-$ on which they have their past endpoints.

Having said that, we shall now establish some important properties.

Lemma 3.9.1. *An asymptotically simple and empty space* $(\mathcal{M}, \mathbf{g})$ *is causally simple.*

Proposition 3.9.1. *An asymptotically simple and empty space* $(\mathcal{M}, \mathbf{g})$ *is globally hyperbolic.*

Lemma 3.9.2. *Let* $\mathscr{W}$ *be a compact set of an asymptotically empty and simple space* $(\mathcal{M}, \mathbf{g})$*, then every null geodesic generator of* $\mathscr{I}^+$ *intersects* $\dot{J}^+(\mathscr{W}, \overline{\mathcal{M}})$ *once, where the dot indicates the boundary in* $\overline{\mathcal{M}}$.

Corollary 3.9.2.1. $\mathscr{I}^+$ *is topologically* $\mathbb{R} \times (\dot{J}(\mathscr{W}, \overline{\mathcal{M}}) \cap \partial\mathcal{M})$.

Now it is possible to show that $\mathscr{I}^+$ (and $\mathscr{I}^-$) and $\mathcal{M}$ are topologically identical to the case of the Minkowski space.

Proposition 3.9.2. *In an asymptotically simple and empty space* $(\mathcal{M}, \mathbf{g})$*,* $\mathscr{I}^+$ *amd* $\mathscr{I}^-$ *are topologically equal to* $\mathbb{R} \times \mathbb{R}^2$*, and* $\mathcal{M}$ *is* $\mathbb{R}^4$.

Penrose has shown that this implies a vanishing Weyl tensor of the metric $\mathbf{g}$ on $\mathscr{I}^+$ and $\mathscr{I}^-$, which can be interpreted as saying that the components of the Weyl tensor go as different powers of the affine parameter on a null geodesic near $\mathscr{I}^+$ and $\mathscr{I}^-$. In addition to that he also introduced conservation laws for the energy-momentum tensor as measured from $\mathscr{I}^+$, in terms of integrals on $\mathscr{I}^+$.

We have that the null surfaces $\mathscr{I}^+$ and $\mathscr{I}^-$ form nearly all the c-boundary Δ of $(\mathcal{M}, \mathbf{g})$. This is because any point $p \in \mathscr{I}^+$ defines a TIP $I^-(p, \overline{\mathcal{M}}) \cap \mathcal{M}$. Considering now λ as an inextendible curve in $\mathcal{M}$, if it has a future endpoint at $p \in \mathscr{I}^+$ then TIP $I^-(\lambda)$ is the same as defined by p. If it does not have a future endpoint on $\mathscr{I}^+$ then $\mathcal{M} - I^-(\lambda)$ must be empty. TIPs, therefore, consist of one for each point of $\mathscr{I}^+$, and one extra TIP being $\mathcal{M}$ itself. Similar considerations are true for TIFs.

Now, it is clear that two TIPs and TIFs, corresponding to $\mathscr{I}^+$ and $\mathscr{I}^-$ are not non-Hausdorff separated. Thus, the c-boundary of any asymptotically simple and empty space $(\mathcal{M}, \mathbf{g})$ is the same as the one of the Minkowski space-time.

Asymptotically simple and empty spaces include Minkowski space and the asymptotically flat spaces containing bounded objects such as stars, which do not undergo gravitational collapse. However, there are some solutions that are not asymptotically simple and empty, since there are null geodesics which do not have endpoints. Nevertheless, these spaces do have asymptotically flat regions, similar to the asymptotically simple and empty ones. It is therefore possible to define a *weakly asymptotically simple and empty* space $(\mathcal{M}, \mathbf{g})$ as a space in which there is an asymptotically simple and empty space $(\mathcal{M}', \mathbf{g}')$ and a neighborhood $\mathscr{U}'$ of $\partial\mathcal{M}'$ in $\mathcal{M}'$ such that $\mathscr{U}' \cap \mathcal{M}'$ is isometric to an open set $\mathscr{U}$ of $\mathcal{M}$.

Chapter 4

Space-time Singularities

In this chapter, we shall use the results from Chapter 2 and 3 to establish some basic results about space-time singularities.

4.1 The definition of singularities

The definition of a space-time singularity is rather different from the one we are used to, for example, in electrodynamics. That is because, if one decides to define singularities as the points in which the metric tensor is either not defined or not differentiably suitable, then it would be possible to just remove these points and define the space-time with the remaining manifold, which would be singularity free according to this definition. In the context of the normal equations of physics, this definition would be reasonable, since these laws would not hold at such a point, making any measurement impossible. Because of that, in Chapter 1, we defined our space-time to be a pair $(\mathcal{M}, \mathbf{g})$ where $\mathbf{g}$ is a suitable differentiable Lorenzian metric, while we ensured that no regular points were omitted from $\mathcal{M}$ along with singular points.

The problem of defining whether space-time has singularities or not can be translated to study whether any singular point had been cut out. To show this we hope to find an incomplete space-time of some sort.

For a manifold $\mathcal{M}$ on which there is a positive defined metric $\mathbf{g}$ it is possible to define a function $\rho(x, y)$ which is the greatest lower bound of the length of curves from x to y. Such a function is a distance function, and can also be regarded as a metric in the topological sense. We shall say that $(\mathcal{M}, \mathbf{g})$ is *metrically complete* (*m-complete*) if every Cauchy sequence concerning ρ converges to a point in $\mathcal{M}$. Equivalently, $(\mathcal{M}, \mathbf{g})$ will be m-complete if every C^1 curve of finite length has an endpoint. It, therefore, follows that m-completeness implies *geodesic completeness* (*g-completeness*), meaning that every geodesic can be extended to arbitrary values of its affine parameter. It is possible to show that g-completeness and m-completeness are equivalent for a positively defined metric.

On the other hand, in a Lorentz metric, it is not possible to define a topological metric, leaving us only with the g-completeness from which it is possible to distinguish three different types, depending on the nature of the geodesic (time-like, space-like, or null-like). By removing a regular point from the space-time it becomes incomplete in all three ways. This, however, does not imply that a g-complete manifold in one way is also complete in the other two.

The incompleteness of a time-like geodesic has the immediate physical meaning of saying that a freely moving observer, or particle, would have no history after (or before) a finite interval of time. Even if the meaning is quite simple, what it implies is quite objectionable from a physical point of view, therefore we shall regard it as a consequence of a singular space-time. Even though the affine parameter of a null geodesic does not have the same physical meaning of the proper time in a time-like geodesic, we still consider a space-time that has an incomplete null-like geodesic to be singular. It is possible to regard this kind of curve as the history of a massless particle. On the other hand, since nothing moves on a space-like curve, it is quite hard to associate the incompleteness of a space-like geodesic with any physical significance. Thus, we shall assume that *time-like and null-like geodesic completeness is the minimum requirement to have a non-singular space-time.* Meaning that the incompleteness of such geodesics implies a singular space-time.

What is useful about taking time-like and null-like incompleteness as indicators of the presence of singularities is that on this basis one can establish a series of theorems about their occurrence. However, the class of incomplete geodesics in our space-time does not include all those we want to consider as singular in some other sense. For example, we have considered as singular a space-time in which a free-falling observer comes to an untimely end, moreover, it is also possible to regard as singular a space-time in which the geodesics are complete, but there is an inextendible time-like curve of bounded acceleration and finite length. This is an observer on a suitable rocketship and a finite amount of fuel which, after a finite interval of time, would no longer be represented as a point of space-time.

Our goal is to find some sort of generalization of the concept of the affine parameter to all C^1 curves which are geodesics or not. A possible solution to this problem is defining a different notion of completeness by requiring that every C^1 curve of finite length, measured by this parameter, has an endpoint.

The idea we are going to develop was first suggested by Ehrsman, that is: let $\lambda(t)$ be a C^1 curve through $p \in \mathcal{M}$, and let $\{\mathbf{E}_\nu\}$ be a basis for T_p, then we shall propagate $\{\mathbf{E}_\nu\}$ along λ to obtain a basis for $T_{\lambda(t)}$ for each value of the parameter t. Then the tangent vector $\mathbf{V} = (\partial/\partial t)_{\lambda(t)}$ can be expressed as $\mathbf{V} = V^\nu(t)\mathbf{E}_\nu$, and by defining a generalized affine parameter u on λ as

$$u = \int_p \left(\sum_\nu V^\nu V^\nu \right)^{1/2} dt.$$

Therefore, u depends only on the choice of p and of the basis $\{\mathbf{E}_\nu\}$ at that point. Given a different choice of the basis at p, $\{\mathbf{E}_{\nu'}\}$ it is always possible to find a non-singular matrix A^μ_ν such that:

$$\mathbf{E}_\nu = \sum_{\mu'} A^{\mu'}_\nu \mathbf{E}_{\mu'}$$

As $\{\mathbf{E}_{\nu'}\}$ and $\{\mathbf{E}_\nu\}$ are parallelly transported along λ such a relation is maintained with constant $A^{\mu'}_\nu$. Since this matrix is non-singular, there is a constant $C > 0$ such that

$$C \sum_\nu V^\nu V^\nu \leq \sum_{\nu'} V^{\nu'} V^{\nu'} \leq C^{-1} \sum_\nu V^\nu V^\nu.$$

Thus, the length of a curve λ is finite in the parameter u if, and only if, it is finite in the parameter u'. If λ is a geodesic, then u is an affine parameter, but such a parameter can be defined in any C^1 curve. A pair $(\mathcal{M}, \mathbf{g})$ is said to be *bundle complete* (*b-complete*) if there is an endpoint for every C^1 curve of finite length, as measured by the generalized affine parameter. If the length, in one of these parameters, is finite, then it will be finite in all of them, so that we will not lose anything by restricting the basis to be orthonormal. If the metric $\mathbf{g}$ is positively defined then the generalized affine parameter is an arc-length and therefore the b-completeness coincides with the m-completeness. Even if the metric is not positively defined it is still possible to define b-completeness. In addition to that, b-completeness does imply g-completeness, however, it is not possible to say the opposite.

We shall therefore say that space-time is *singularity-free* if it is b-complete.

From intuition one imagines a singularity to involve an unboundedly large growth of the curvature near a singular point. However, since we excluded the singular points from our definition of space-time, the concept of near and unboundedly large are not defined on them. What one can do is to say that, on a b-incomplete curve, points that are near the singularity correspond to a value of the generalized affine parameter that is near the upper bound. The definition of unboundedly large is quite harder to give since the size of the components of the curvature tensor depends on the basis on which it is measured. A possibility to overcome this difficulty is by considering some scalar polynomials in $g_{\alpha\beta}$, $\eta_{\alpha\beta\mu\nu}$, and $R_{\alpha\beta\mu\nu}$. A b-incomplete curve will correspond to a scalar polynomial curvature singularity (*s.p. curvature singularity*) if any of these scalar polynomials is unbounded on the incomplete curvature. However, in a Lorentz metric, these polynomials do not fully characterize the Riemann tensor since in plane-wave solutions the scalar polynomials are all zero, but the Riemann tensor does not vanish. Because of that, even though these polynomials remain small, the curvature might become very large in some sense.

Alternatively one might measure the components of the curvature tensor in a parallelly propagated basis along a curve. A b-incomplete curve will correspond to a curvature singularity with respect to a parallel translated basis (*p.p. curvature singularity*) if any of these components is unbounded on a curve. In particular, we have that an s.p. curvature singularity implies a p.p. curvature singularity.

4.2 Singularity theorems

In many solutions of the Einstein field equations, it is possible to find singularities that arise from the symmetry of the solution itself. Because of that, such results do not necessarily have any physical meaning. From this consideration, there were speculations that singularities were simply the result of some symmetries and did not appear in general solutions, since there were some solutions with space-like singularities which did not have the full number of functions expected in a general solution. This, however, is not the case, since, thanks to Belinskii, Kahaltnikov, and Lifshitz, it has been found that there are other classes of solutions that seem to have the full number of arbitrary functions and contain singularities. Their methods to predict the existence of singularities draw light on their structure. However, it is not clear whether the power series they used converges, neither one obtains a general condition on the inevitability of singularities. Nevertheless, we still shall use their results as a support to our view of singularities, that is the implication of infinite curvature in the general case.

To predict the occurrence of singularities by studying the incompleteness of space-time, we shall introduce a series of theorems, called singularity theorems, the first (modern version) of which was introduced by Penrose, to prove the occurrence of a singularity in the gravitational collapse of a dying star beyond the Schwarzschild radius. In this case, if the collapse is spherically symmetric we have a solution that can always be integrated explicitly, always obtaining a singularity. However, this is not trivial if we consider some irregularities or a small amount of angular momentum. Penrose showed that (differently from the Newtonian case) once the star had passed inside the Schwarzschild surface it would be impossible to get out. Moreover, from more general criteria, he proved that this is the case not only for a spherically symmetric collapse, but also for solutions that do not have any exact symmetry. Thus, there exists a *close trapped surface* $\mathscr{T}$, that is a C^2 closed space-like two-surface, such that the two families of null geodesics orthogonal to $\mathscr{T}$ converge at $\mathscr{T}$. This means that such a surface finds itself in a very strong gravitational field such that even the outgoing light beams are dragged back (this is what the convergence implies). Since nothing can travel faster than light the matter within $\mathscr{T}$ is trapped inside a succession of two-surfaces of smaller and smaller areas.

Theorem 4.2.1. *Space-time* $(\mathcal{M}, \mathbf{g})$ *cannot be null geodesically complete if:*

1. $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ *for all null vectors* K^α*;*

2. *there is a non-compact Cauchy surface* $\mathscr{H}$ *in* $\mathcal{M}$*;*

3. *there is a closed trapped surface* $\mathscr{T}$ *in* $\mathcal{M}$*.*

To prove this result we shall show that the boundary of the future of $\mathscr{T}$ would be compact if $\mathcal{M}$ were null geodesically complete, which allows us to show that it will be incomplete when $\mathscr{H}$ is not compact.

Proof. From proposition 3.6.3 we know that the existence of a Cauchy surface implies that $\mathcal{M}$ is globally hyperbolic, and therefore Cauchy is simple by proposition 3.6.1. This means that the boundary of $J^+(\mathscr{T})$ will be $E^+(\mathscr{T})$ and will be generated by null geodesic segments which have past endpoints on $\mathscr{T}$ and are orthogonal to $\mathscr{T}$. By assuming $\mathcal{M}$ to be geodesically complete, then by condition (1) and (3), and proposition 2.2.6, there would be a point conjugate to $\mathscr{T}$ along every future-directed null geodesic orthogonal to $\mathscr{T}$ within an affine distance of $2c^{-1}$, where c is the value of ${}_n\hat{\chi}_{\alpha\beta}g^{\alpha\beta}$ at the point of intersection between the null geodesic and $\mathscr{T}$. By proposition 2.3.9, all the points on such a null geodesic beyond the point conjugate to $\mathscr{T}$ would lie in $I^+(\mathscr{T})$. Thus, each segment of $\dot{J}(\mathscr{T})$ would have a future endpoint at or before the point conjugate to $\mathscr{T}$. At $\mathscr{T}$ one could assign continuously an affine parameter on each null geodesic orthogonal to the surface.

By considering the continuous map $\beta : \mathscr{T} \times [0, b] \times Q \to \mathcal{M}$, where Q is a discrete set containing 1 and 2, which takes a point $p \in \mathscr{T}$, with an affine distance v, on one of the two future-directed null geodesics through p orthogonal to $\mathscr{T}$, since $\mathscr{T}$ is compact, then there will be some minimum value c_0 of $-{}_1\hat{\chi}_{\alpha\beta}g^{\alpha\beta}$ and $-{}_2\hat{\chi}_{\alpha\beta}g^{\alpha\beta}$. Therefore, if $b_0 = 2c_0^{-1}$, then $\beta(\mathscr{T} \times [0, b_0] \times Q)$ would contain $\dot{J}^+(\mathscr{T})$. Because of that, $\dot{J}^+(\mathscr{T})$ is compact since it is the closed subset of a compact set. This is possible if the Cauchy surface $\mathscr{H}$ was compact, this way $\dot{J}(\mathscr{T})$ could meet up around the back and create a compact Cauchy surface homeomorphic to $\mathscr{H}$. However, if $\mathscr{H}$ is not compact, then there would be some problems. That is because $\mathcal{M}$ admits a past-directed C^1 time-like vector field and each integral curve of such a field would intersect $\mathscr{H}$ and $\dot{J}^+(\mathscr{T})$ at the most once. Now it is possible to define a continuous map $\alpha : \dot{J}^+(\mathscr{T}) \to \mathscr{H}$, such that if $\dot{J}^+(\mathscr{T})$ is compact then also its image $\alpha(\dot{J}^+(\mathscr{T}))$ would be compact and homeomorphic to $\dot{J}^+(\mathscr{T})$. However, since $\mathscr{H}$ is not compact, $\alpha(\dot{J}^+(\mathscr{T}))$ cannot contain the whole set $\mathscr{H}$, since, by proposition 3.3.1 $\dot{J}^+(\mathscr{T})$ and $\alpha(\dot{J}^+(\mathscr{T}))$ would be a three-dimensional no-bounded manifold, showing that the assumption of null geodesic completeness is not valid.

□

The first condition of the theorem ($R_{\alpha\beta} \geq 0$) had been discussed in Chapter 2, and it holds no matter what is the value of the cosmological constant Λ, meaning that the energy density must be non-negative for every observer. The second condition, on the other hand, is always satisfied in at least some part of the space-time. Finally, the third condition had been introduced only to show that $\alpha(\dot{J}(\mathscr{T}))$ could not be the whole of $\mathscr{H}$, which can be achieved by requiring that there exists a future-directed inextendible curve from $\mathscr{H}$ not intersecting $\dot{J}(\mathscr{T})$, meaning that the theorem would still hold even if $\mathscr{H}$ was compact.

The weakness of this theorem is the requirement of having $\mathscr{H}$ as a Cauchy surface, which was used firstly to show that $\mathcal{M}$ was causally simple, implying that the generator of $\dot{J}^+(\mathscr{T})$ had past endpoints on $\mathscr{T}$, and secondly to ensure that under a map α every point of $\dot{J}^+(\mathscr{T})$ was mapped into a point of $\mathscr{H}$. That is because there are some solutions of the Einstein field equations, such as the Reissner-Nordström solution, which do not have Cauchy surfaces.

The only thing that such a theorem tells us is that at the end of the collapse of a star, there will be either a singularity or a Cauchy horizon. Such a result is very important since in both cases our ability to predict the future breaks down, however, it does not answer the question about the occurrence of singularities in physically realistic solutions. Therefore, we need a theorem that will help solve this problem, without the assumption of the existence of Cauchy surfaces in the hypothesis. In addition to that, it must have the condition $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ for all time-like and null-like vectors, so that we make it reasonable in all the possible solutions. We also require the validity of the chronology condition, that is all the time-like geodesics are non closed. The condition of existence of a closed trapped surface is now just one of three possible conditions, therefore this new theorem is already applicable in a wide variety of situations. In particular one of the alternative conditions is that there should be a compact partial Cauchy surface, while the third asks for a point whose past (or future) light cone starts converging again.

Theorem 4.2.2 (Hawking and Penrose)**.** *Given a generic space-time* $(\mathcal{M}, \mathbf{g})$*, it will not be time-like and null-like geodesically complete if:*

(1) $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ *for every non-spacelike vector* K^α*;*

(2) The generic condition, introduced in Chapter 2, is satisfied, i.e. *every non-spacelike geodesic contains a point at which* $K_{[\alpha}R_{\beta]\mu\nu[\rho}K_{\gamma]}K^\mu K^\nu \neq 0$ *where* $\mathbf{K}$ *is the tangent vector to the geodesic;*

(3) The chronology condition holds on $\mathcal{M}$*,* i.e. *there are no closed time-like curves;*

(4) At least one of the following structure will exist:

(i) a compact achronal set without edge,

(ii) a closed trapped surface,

(iii) a point p such that on every past (or future) null geodesic from p the divergence $\hat{\theta}$ *of the null geodesics from the previous point becomes negative, meaning that the null geodesics are focused by the matter curvature and start to reconverge.*

Equivalently:

Theorem 4.2.3. *The following three conditions cannot hold all at once:*

(a) every inextendible non-spacelike geodesic, contains a pair of conjugate points;

(b) the chronology condition holds on $\mathcal{M}$*;*

(c) *there is an achronal set $\mathscr{S}$ such that $E^+(\mathscr{S})$ or $E^-(\mathscr{S})$ is compact (we shall say that such a set is respectively,* future trapped *or* past trapped*).*

If we prove the last version of the theorem true, the first will then follow, since if $\mathcal{M}$ is time-like and null-like geodesically complete, then condition (1) and (2) would imply (a) by propositions 2.2.2 and 2.2.5, while (3) is equivalent to (b), and condition (1), (4) would imply (c), since in case (i), $\mathscr{S}$ would be an achronal compact set without edge and such that

$$E^+(\mathscr{S}) = E^-(\mathscr{S}) = \mathscr{S}.$$

In the two cases (ii) and (iii), $\mathscr{S}$ would be the closed trapped surface and p the point at which the divergence becomes negative, respectively. By proposition 2.2.4, 2.2.6, 2.3.7, 2.3.9, $E^+(\mathscr{S})$ and $E^-(\mathscr{S})$ would be the component being the intersections of the closed sets $\dot{J}^+(\mathscr{S})$ and $\dot{J}^-(\mathscr{S})$ with compact sets consisting of all the null geodesics of some finite length from $\mathscr{S}$.

Lemma 4.2.4. *If $\mathscr{S}$ is a closed set and if the strong causality condition holds on $\overline{J^+}(\mathscr{S})$ then $H^+(\overline{E^+}(\mathscr{S}))$ is non-compact or empty.*

Corollary 4.2.4.1. *If $\mathscr{S}$ is a future trapped set, there is a future-inextendible time-like curve γ contained in $D^+(E^+(\mathscr{S}))$.*

Proof. Let us consider the compact set $\mathscr{F}$ defined to be $E^+(\mathscr{S}) \cap \overline{J^-}(\gamma)$. Now, since γ is contained into $\text{int}(I^+(E^+(\mathscr{S})))$, then $E^-(\mathscr{S})$ would consist of $\mathscr{F}$ and a portion of $\dot{J}^-(\gamma)$, and, because γ is future-inextendible, the null geodesic segments generating $\dot{J}^-(\gamma)$ could have no future endpoints. However, from statement (a) every inextendible non-spacelike geodesic contains a pair of conjugate points, which, by proposition 2.3.7, implies that the past-inextendible extension ν' of each generating segment ν of $\dot{J}^-(\gamma)$ would enter $I^-(\gamma)$. Thus, there will be a past endpoint for ν at or before the first point p of $\overline{\nu' \cap I^-(\gamma)}$, and since $I^+(\gamma)$ is an open set, then a neighborhood of p would contain some points in $I^+(\gamma)$ of some neighboring null-like geodesics. The affine distance of p from $\mathscr{F}$ will, therefore, be upper semi-continuous, and the past horizon $E^-(\mathscr{F})$ would be a compact set since it is the intersection of a closed set $\dot{J}(\gamma)$ and a compact set generated by null geodesic segments of a bounded affine length from $\mathscr{F}$. It would therefore follow from the previous lemma that there exists a past-inextendible time-like curve λ, in $D^-(E^-(\mathscr{F}))$, such that there would be an infinite sequence of its points a_n which satisfy the following criteria:

(I) $a_{n+1} \in I^-(a_n)$,

(II) no compact segment of λ contains more than a finite number of a_n.

Let b_n be a similar sequence on γ but with I^+ instead of I^- in (I) with $b_1 \in I^+(a_1)$.
As γ and λ are contained in the globally hyperbolic set $D(E^-(\mathscr{F}))$, there would be a non-spacelike geodesic μ_n of maximum length between each a_n and the corresponding

b_n. Each of them would intersect $E^+(\mathscr{S})$, thus there will be a $q \in E^+(\mathscr{S})$ that is a limit point of the $\mu_n \cap E^+(\mathscr{S})$ and a non-spacelike direction at q being a limit of the directions of the μ_n. Let μ'_n be a subsequence of μ_n such that $\mu'_n \cap E^+(\mathscr{S})$ converges to q and that the directions of μ'_n at $E^+(\mathscr{S})$ converges to the limit direction. Let, now, μ be the inextendible geodesic in the direction of the limit through q. By (a) we know there will be some conjugate points x and y on μ such that $y \in I^+(x)$, with respectively future and paste counterparts on μ, x' and y'. From proposition 2.3.3, there is some value $\epsilon > 0$ and a time-like curve α from x' to y' whose length is ϵ plus the length of μ from x' to y'.

Let $\mathscr{U}$ and $\mathscr{V}$ be two convex normal coordinate neighborhoods of, respectively, x' and y' both of which do not contain any curve of length $\frac{1}{4}\epsilon$, and let x^n, y^n be $\dot{\mathscr{U}} \cap \alpha$ and $\dot{\mathscr{V}} \cap \alpha$, respectively. If x_n and y_n are, respectively, points on μ'_n converging to x' and y', then, for a sufficiently large n, the length μ'_n from x'_n to y'_n will be less than $\frac{1}{4}\epsilon$ plus the length of μ from x' to y', and we will have x'_n and y'_n belonging in $I^-(x'', \mathscr{U})$ and $I^-(y'', \mathscr{V})$. Going along α from x'_n, through x'', to y'', and from y'' to y'_n we would have a longer non-spacelike curve than μ'_n from x'_n to y'_n. From property (II), however, a'_n must lie to the past of x'_n on μ'_n, while b'_n would lie to the future of y'_n on μ'_n, therefore μ'_n has to be the longest non-spacelike curve from x'_n to y'_n. Which establishes the desired condition. □

With this theorem, we have proved the existence of singularities under very general conditions. However, it does not give us any hint on whether a singularity would occur in the past or in the future. In case (ii) of the 4th condition, which asks for the existence of a compact space-like surface, one has no reason to believe it should be in the future rather than in the past. Nevertheless, in case (i) one would expect the singularity to be in the future, because of the existence of a closed trapped surface, while in case (iii), because this condition wants the past null-like cone to start reconverging, one would expect to find a singularity in the past.

By strengthening condition (iii), that is assuming all past-directed time-like, as well as null-like, geodesics from p to start reconverging toward a compact region in $J^-(p)$, it is possible to say that one would still expect a singularity to exist in the past.

Theorem 4.2.5 (Hawking)**.** *If:*

(1) $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ *for every non-spacelike vector* K^α*;*

(2) the strong causality condition holds on $(\mathcal{M}, \mathbf{g})$*;*

(3) there is some past-directed unit time-like vector $\mathbf{W}$*, at point* p*, and a positive constant* b *such that if* $\mathbf{V}$ *is the unit tangent vector to the past-directed time-like geodesic, through* p*, then on each such geodesic the expansion* $\theta \doteq V^\alpha_{;\alpha}$ *of these geodesics would become less than* $-3c/b$ *within a distance of* b/c *from* p*, where* $c \doteq -W^\alpha V_\alpha$*. This implies that there is a past incomplete non-spacelike geodesic through* p*.*

Proof. Let us begin by considering K^α to be the parallelly transported tangent vector to the past-directed non-spacelike geodesics through p, with normalization $K^\alpha W_\alpha = -1$. Then, for the time-like geodesic through p, we will have $K^\alpha = c^{-1}V^\alpha$, meaning that $K^\alpha_{;\alpha} = c^{-1}V^\alpha_{;\alpha}$. Since $K^\alpha_{;\alpha}$ is continuous on the non-spacelike geodesic, it will become less than $-3/b$ on the null geodesic through p within an affine distance of b.

If we now consider a pseudo-orthonormal tetrad $\mathbf{Y_{(1)}}$, $\mathbf{Y_{(2)}}$, $\mathbf{Y_{(3)}}$, and $\mathbf{Y_{(4)}}$ on a null geodesic, with $\mathbf{Y_{(1)}}$ and $\mathbf{Y_{(2)}}$ two space-like unit vectors and $\mathbf{Y_{(3)}}$, $\mathbf{Y_{(4)}}$ two null-like vectors such that $Y^\alpha_{(3)}Y_{(4)\alpha} = -1$ and $\mathbf{Y_{(4)}} = \mathbf{K}$, then we will have the expansion $\hat{\theta}$ of the null geodesics through p defined as:

$$\begin{aligned}\hat{\theta} &= K_{\alpha;\beta}(Y^\alpha_{(1)}Y^\beta_{(1)} + Y^\alpha_{(2)}Y^\beta_{(2)}) \\ &= K^\alpha_{;\alpha} + K_{\alpha;\beta}(Y^\alpha_{(3)}Y^\beta_{(4)} + Y^\alpha_{(4)}Y^\beta_{(3)}).\end{aligned}$$

The second term will vanish since K^α is parallelly transported, while the third can be expressed as $\frac{1}{2}(K_\alpha K^\alpha)_{;\beta}Y^\beta_{(3)}$ which is negative since $K_\alpha K^\alpha$ is zero on null geodesics and negative for time-like geodesics. Thus showing that $\hat{\theta}$ will get under $-3/b$ within an affine distance of b along a null geodesic through p. This means that if all the past-directed null geodesics are complete, then the horizon $E^-(p)$ will be compact, and any point $q \in J^-(E^-(p)) - E^-(p)$ would be in $I^-(p)$, and not in $J^+(E^-(p))$, since $E^-(p)$ is achronal. Therefore:

$$J^+(E^-(p)) \cap J^-(E^-(p)) = E^-(p),$$

will be compact, and, by proposition 3.6.5, $D^-(E^-(p))$, globally hyperbolic.

Considering now proposition 3.7.1, each point $r \in D^-(E^-(p))$ would be joined to p by a non-spacelike geodesic such that it will not contain any point conjugate to p between r and p. Thus, from proposition 2.2.1, $D^-(E^-(p))$ would be contained in $\exp_p(F)$ where F is the compact region of T_p which consists of all past-directed non-spacelike vectors K^α such that $K^\alpha W_\alpha \leq -2b$.

If all past-directed non-spacelike geodesics are complete, then we will have a compact $\exp_p(F)$, since $\exp_p(K^\alpha)$ would be defined for all $K^\alpha \in F$, and $\exp_p(F)$ is the image of a compact set under a continuous map. However, from the corollary of lemma 4.2.4 we understand that $D^-(E^-(p))$ contains a past-inextendible time-like curve, which by proposition 3.4.7 implies that it could not be totally imprisoned in the compact set $\exp_p(F)$. Therefore, the assumption that all past-directed non-spacelike geodesics are complete has to be false. □

The last two theorems are the most useful when it comes to studying singularities since it has been shown that their condition would apply in a variety of physical situations. However, what may happen is that what we found were not singularities, but closed time-like curves violating the causality condition. Such a case is much harder to deal with, and therefore we should just consider what happens to our space-time in the case of this violation. In particular, we want to know if the violation of the causality condition would prevent the occurrence of singularities. The following theorem will show that in certain situations this is not the case, meaning that a causality breakdown is not a way out.

Theorem 4.2.6 (Hawking)**.** *Space-time is not geodetically complete if:*

(1) $R_{\alpha\beta}K^\alpha K^\beta \geq 0$ *for every non-spacelike vector* K^α*;*

(2) there exists a compact space-like three-surface $\mathscr{S}$ *(without edge);*

(3) the unit vectors normal to $\mathscr{S}$ *are either everywhere converging or diverging on* $\mathscr{S}$*.*

Condition (2) may be regarded as asking for a spatially closed universe, while condition (3) is asking for either a contracting or expanding universe. As done before, to prove this theorem we shall work in $\hat{\mathcal{M}}$ which is a covering manifold containing each connected component of the image of $\mathscr{S}$, one of which will be denoted as $\hat{\mathscr{S}}$.

Proof. By condition (2) and (3) the contraction χ^α_α of the second fundamental form of $\hat{\mathscr{S}}$ has a negative upper bound on $\mathscr{S}$, thus if $\mathcal{M}$, and therefore $\hat{\mathcal{M}}$, is time-like geodesically complete, then there would be a point conjugate to $\hat{\mathscr{S}}$ on every future-directed geodesic orthogonal to the surface $\hat{\mathscr{S}}$ within a finite distance upper bound b from $\hat{\mathscr{S}}$, as stated in proposition 2.2.3. By the corollary to proposition 3.7.1, to every point $q \in D^+(\hat{\mathscr{S}})$, there is a future-directed geodesic orthogonal to $\hat{\mathscr{S}}$ which does not contain any point conjugate to $\hat{\mathscr{S}}$ between $\hat{\mathscr{S}}$ and q.

Let now $\beta : \hat{\mathscr{S}} \times [0, b] \to \hat{\mathcal{M}}$ to be a differentiable map which takes a point $p \in \hat{\mathscr{S}}$ up a future-directed geodesic through p by a distance $s \in [0, b]$. Then we will have a compact set $\beta(\hat{\mathscr{S}} \times [0, b])$ which contains $D^+(\hat{\mathscr{S}})$, implying $\overline{D^+}(\hat{\mathscr{S}})$ and $H^+(\hat{\mathscr{S}})$ to be compact. By assuming the validity of the strong causality condition, the contraction in (3) will follow from lemma 4.2.4. However, it is possible to show that one can still obtain a contraction even without the validity of such a condition.

Let us consider a point $q \in H^+(\hat{\mathscr{S}})$, because every past-directed non-spacelike curve from q to $\hat{\mathscr{S}}$ would consist of a null geodesic segment in $H^+(\hat{\mathscr{S}})$ and a non-spacelike curve in $D^+(\hat{\mathscr{S}})$, it follows that $d(\hat{\mathscr{S}}, q)$ would be less or equal to b. Thus, since d is lower semi-continuous, one could find an infinite sequence of points $r_n \in D^+(\hat{\mathscr{S}})$ converging to q such that $d(\hat{\mathscr{S}}, r_n)$ converges to $d(\hat{\mathscr{S}}, q)$. To each r_n there will be associated at least one element of $\beta^{-1}r_n$ of $\hat{\mathscr{S}} \times [0, b]$, and, because such a set is compact, then there would be an element (p, s) which is a limit point for $\beta^{-1}(r_n)$. By continuity, we will have $s = d(\hat{\mathscr{S}}, q)$ and $\beta(p, s) = q$, thus to every point $q \in H^+(\hat{\mathscr{S}})$ there will be a time-like geodesic of length $d(\hat{\mathscr{S}}, q)$ from $\hat{\mathscr{S}}$. Let us now consider a point $q_1 \in H^+(\hat{\mathscr{S}})$ lying to the past of q on the same null geodesic λ which generates $H^+(\hat{\mathscr{S}})$. Joining the geodesic of length $d(\hat{\mathscr{S}}, q_1)$ from $\hat{\mathscr{S}}$ to q_1, to the segment of λ between q_1 and q we shall obtain a non-spacelike curve of length $d(\hat{\mathscr{S}}, q_1)$ from $\hat{\mathscr{S}}$ to q, which could be varied to give a longer curve between these two endpoints. Therefore, $d(\hat{\mathscr{S}}, q)$ will strictly decrease along every past-directed generator of $H^+(\hat{\mathscr{S}})$ (with $q \in \hat{\mathscr{S}}$). From proposition 3.5.2 such a generator cannot have past endpoints leading to a contraction, since, as $d(\hat{\mathscr{S}}, q)$ is lower semicontinuous in q, it would have a minimum on the compact set $H^+(\hat{\mathscr{S}})$. □

The necessity of condition (2) arises from an example given by Penrose himself, in which he shows how in the Minkowski space $(\mathcal{M}, \eta)$ the non-compact surface $\mathscr{S} : (x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 = -1$ with $x^4 < 0$, is a Cauchy surface with $\chi^\alpha_\alpha = -3$ at all points. If one considers a surface similar to the previous, but such that $(x^1)^2 + (x^2)^2 + (x^3)^2 + (x^4)^2 < 0$ with $x^4 < 0$, it is possible to define a group of discrete isometries G such that $\mathscr{S}/G$ is compact. Moreover, from what is required by theorem 4.2.6, the space $(\mathcal{M}/G, \eta)$ would be time-like geodesically incomplete since one could not extract the identification under G to the whole space $\mathcal{M}$. In this case, the incompleteness singularity arises from bad global properties and is not accompanied by a curvature singularity.

Interestingly it is possible to change the condition of theorem 4.2.6 and maintain its validity. For example by replacing conditions (2) and (3) with:

(2') $\hat{\mathscr{S}}$ is a Cauchy surface for the covering manifold $\hat{\mathcal{M}}$;

(3') χ^α_α is bounded away from zero on $\mathscr{S}$.

Because in this case it is not possible to have a Cauchy horizon, all future-directed time-like geodesics from $\hat{\mathscr{S}}$ must have a length less than some finite upper bound.

In addition to that, Geroch had shown that keeping condition (2) while replacing conditions (1) and (3) by:

(1") $R_{\alpha\beta} K^\alpha K^\beta \geq 0$ for every non-spacelike vector, even if $R_{\alpha\beta} = 0$;

(3") there is a point $p \in \hat{\mathscr{S}}$ such that any inextendible non-spacelike curve intersection $\hat{\mathscr{S}}$ also intersects both $J^+(p)$ and $(J^-(p))$;

one gets that either the Cauchy development of $\hat{\mathscr{S}}$ is flat, or $\hat{\mathcal{M}}$ is time-like geodetically incomplete. Condition three can be regarded as saying that an observer at p can see every particle intersecting $\hat{\mathscr{S}}$ and vice-versa. Moreover, by considering a topological space $S(p)$, out of each of these surfaces, then, by conditions (2) and (3"), it must be compact.

Chapter 5

Conclusions

The structure of all the theorems introduced in the previous chapter, seem to follow a general pattern:

Theorem 5.0.1 (General Pattern)**.** *If:*

1. *an energy condition,*
2. *a causality condition, and*
3. *a boundary or initial condition*

holds, then our space-time will contain at least one incomplete causal geodesic.

However, this basic structure works only when adequate energy, causality, and boundary conditions are specified.

By looking at the proofs of these theorems, we observe that, from a geometric point of view, they are based either on the construction of at least one maximal geodesic, that is a geodesic with no conjugate or focal points, showing that such a geodesic cannot be complete due to the chosen energy condition, or on the assumption of g-incompleteness, leading to the existence of compact achronal boundaries, which, however, do not exist if space-time is spatially open. From a physical point of view, on the other hand, the concept behind each proof is different depending on the theorem. For theorem 4.2.1 and theorem 4.2.5, in particular, the idea behind them is the assumption of having a set bound to be trapped. There is, therefore, a certain region of space-time (let it be a surface, a slice, or a point) that has its future or its past initially contained within a compact and contracting spatial region. Meaning that all matter contained in such a region cannot escape from a spatially finite contracting zone. That is, of course, as long as gravity remains attractive and there is no way out back in time through the violation of causality. The particles in such a region would get closer together until either they collapse into a region too small for such amount of matter or radiation, thus creating a singularity, or they reach an untrapped region, allowing them to escape. In the last case, the future of such a particle will be all contained in $J(\mathscr{S}, \mathscr{U})$ that is within the achronal boundary $\mathscr{H}$, which is compact if it does not reach any singularity. Such a prospect is possible in a variety of physical solutions. However, if we consider the ideal case, in which all the particles are required to cross the Cauchy horizon $H^+(\mathscr{S})$ of the proper archonal boundary $\mathscr{H}$, then they will be freed from the catastrophic influence that gravity exerts on them. This eventuality cannot occur, since the Cauchy horizon $H^+(\mathscr{S})$ must be either non-compact or empty. Thus, some particles are able to travel indefinitely without having to leave $D^+(\mathscr{H})$, so that they either approach a singularity, or go out to infinity. In the last case, we have to take into account the event horizon of the curve which goes to infinity. Such a set is an acrhonal boundary as well, which, combined with an adequate subset of the $\mathscr{H}$, gives yet another proper achronal boundary, which, by the energy condition, is either compact, or reaches a singularity. If we extend this reasoning also to past-directed particles, meaning that there will be another particle that can reach infinity, then its combination with the previous case would give us access to every point in time, from the past to the future, remaining within a finite spatial region, avoiding the focusing effect. This is, however, not possible, since the arrow of time is thermodynamically inevitable, and gravity always remains attractive.

From this consideration, we concluded that for an adequate energy condition, General Relativity favors the existence of incomplete curves. This can also be seen by considering theorem 4.2.3, in which we stated that conditions (i) and (iii) cannot hold at the same time. Why it happens can be made more clear by studying some particular situations, however, a general proof is still unknown.

Even if in this chapter we used words such as particles or matter to discuss singularities, all our previous considerations had been made independently of these concepts. In particular, we have seen that at the early stages of our understanding of singularities, there was the idea that they were simply a property of some particular solutions, arising from particular symmetries. However, we showed that geodesic incompleteness can also occur in an empty space-time. Thus, the intuitive idea that singularities may have something to do with the existence of matter, or with its bad definition, is wrong. An example of that, which we will not fully cover, are space-times with simple pure gravitational waves and no matter whatsoever. The case of colliding plane waves allows no alternative but the destruction of such space-times by the existence of singularities. [Senovilla, 1998, pp. 796–798]

5.1 Conclusions of the singularity theorems

In conclusion, what arises from these theorems is the existence of at least one incomplete causal geodesic. However, they do not give us any other information, leaving the problem of the existence of extensions and singularities open.

In the case of g-incompleteness, we have seen that it is possible to remove singularities and make regular extensions possible, the problem now becomes which one to choose among the huge amount of possibilities. Moreover, there is no reason to assume that the conditions, or other equivalent ones, which hold on the unextended space-time, will remain valid also on its extension. Meaning that the energy, causality, or boundary conditions that led us to the existence of incomplete geodesic may not be satisfied in the extended space-time.

On the other hand, when g-incompleteness indicates an essential singularity, the problem of its character, severity, and location remains. For example, one may think that singularities would be found in the past and that they would all be of big-bang type (that is a singularity on a singular extension in which every past-directed causal curve, with no endpoints, approaches a singular set $\mathscr{S}$ at a finite affine parameter), but they are nothing of that sort. It has been shown that, in general, this is not the case, both for homogeneous and inhomogeneous space-times, meaning that there is no reason to presume that cosmological singularities have any definite property. [Senovilla, 1998, pp. 806–807]

Bibliography

Stephen W. Hawking and George F. Ellis. *The Large Scale Structure of Space–Time*. Cambridge University Press, 1973.

Bernard F. Schutz. *A first course in general relativity*. Cambridge University Press, 1985. pages 171–198.

Alexander Tolish. General relativity and the newtonian limit. pages 14–17. University of Chicago Math, 2010. URL https://api.semanticscholar.org/CorpusID:14050709.

Prado Martín-Moruno and Matt Visser. *Classical and Semi-classical Energy Conditions*, pages 193–213. Springer International Publishing, Cham, 2017. ISBN 978-3-319-55182-1. doi: 10.1007/978-3-319-55182-1_9. URL https://doi.org/10.1007/978-3-319-55182-1_9.

José M. M. Senovilla. Singularity theorems and their consequences. *General Relativity and Gravitation*, 30(5):701–848, May 1998. ISSN 1572-9532. doi: 10.1023/A:1018801101244. URL https://doi.org/10.1023/A:1018801101244.