Bachelor Thesis in Engineering Physics: Singularity Theorems

Marc Basquens
Director: Ramon Torres

2016

Contents

Abstract

In this work, the main singularity theorems are reviewed, and a detailed development of the required mathematical tools in their proofs is performed. Also, the plausibility and applicability of the assumptions of the theorems to realistic spacetimes and how the results might change if other assumptions were made are discussed. Finally, some examples to illustrate a range of cases regarding the fulfillment of the hypothesis of the theorems are provided and analyzed. The stress is put in how regular spacetimes succeed in avoiding the hypothesis in singularity theorems and which consequences this has on them.

Introduction

Since Einstein published his famous field equations of General Relativity in 1915, founding an essential yet not fully explored branch of physics, many efforts have been put in developing this theory. In just a few years, several solutions to the complicated equations were found, describing some interesting physical scenarios and even solving some problems such as the precession of the perihelion of Mercury, confirming the success of the new theory. However, it was not unnoticed that in most examples there were problematic regions, catastrophic events. In these 'places' there was a breakdown of not only General Relativity, but of all physics. Nevertheless, it seemed that the new theory itself was not only allowing for its own collapse, but also favouring it. These mysterious 'places' were called singularities, and in most cases, they seemed unavoidable.

However, the physicists that early developed General Relativity did not regard the singularities as a problem. Although there appeared infinite physical observable quantities at them, they just dismissed the singularities as mathematical artifacts due to the symmetry of the already known exact solutions of Einstein's equations or as unreachable effects. It would take a new generation of physicists to consider the problem more seriously. Although there are other tools for the study of singularities, Physics needed what we now call singularity theorems in order to clarify the nature of singularities, when would they appear and the problem of their abundance in GR. On the one hand, the physical importance of the topic was clear, since singularities are outside the limits of application of GR and it is crucial to keep under control when, where and under which conditions they may or must occur. But on the other hand, the study of singularity theorems was about to be a brand new field within General Relativity, than even Einstein had not predicted when he was developing his theory.

Indeed, in 1955 (ironically, a few days after Einstein's death) Raychaudhuri (and also Komar, independently) published what could be considered the first singularity theorem. This achievement motivated further investigation on the topic. Ten years later, in 1965, Penrose wanted to prove that singularities were not formed due to the assumption of spherical symmetry in the cosmological and astrophysical models by introducing several new ideas. As a result, he ended up proving his own singularity theorem. This could be considered the first 'modern' singularity theorem, in the sense that a new whole set of important concepts and developments were used in its making. This theorem inspired a lot of works, in particular those of Hawking who also published a singularity theorem in 1967. After some years of this revolution, several results were achieved, and in 1970, Penrose and Hawking collected them in a very strong theorem, which still is the main singularity theorem. After these years, there have been a lot of attempts to improve the theorems and to look for new results.

From a physical point of view, it is clear that the singularity theorems are relevant and necessary in order to make General Relativity a more complete theory and to understand what are and why do singularities appear. However, from a mathematical point of view they are also a milestone. All the tools necessary for it are a combination of several mathematical areas such as Differential Geometry or Topology. The ideas, techniques and mechanisms developed for the specific goal of the analysis of singularities can also be applied to more general scenarios and even modified with relatively little effort.

In this document, the probably most important singularity theorems by Raychaudhuri, Penrose and Hawking are reviewed, as well as the mathematical and physical foundations necessary to understand their meaning and conclusions. Moreover, some insights in their hypothesis and results are provided and why they are needed (or not) in order to ensure the existence of a singularity (and eventually, how to avoid it) is studied. This will be better illustrated with some final examples.

The approach used to develop the singularity theorems was fully classical: quantum gravity was not considered. Obviously, ignoring quantum effects makes the results 'wrong' in the sense that one expects them to considerably change the classical theorems. In fact, since the main results are not just valid for GR, but for a theory featuring a manifold, including quantum effects does not necessarily make the theorems unapplicable. Some comments will be made about that.

The structure of the document is as follows. In Chapter 1, the basic mathematical tools indispensable for the statements and proofs of the main singularity theorems are presented and developed and concepts such as energy conditions, congruences or trapped sets and introduced. In Chapter 2, three of the most important singularity theorems are stated and proved, along with some necessary previous results. In Chapter 3, the hypothesis used in the theorems, their plausibility and applicability are discussed, and how the conclusions change if some of them are modified or removed is explored. In Chapter 4, some examples of spacetimes are provided, paying special attention to the characteristics that make them singular or regular and comparing them to the hypothesis in the theorems.

The notation used in this work is quite standard, however a few remarks may be in order. Greek indices run from 0 to 3 (μ, ν, ...) while latin indices run from 1 to 3 (i, j, ...) and capital latin indices run from 2 to 3 (A, B, ...) unless it is explicitly stated. The units $8\pi G = c = 1$ will be used except explicitly stated, and Einstein's equations are written as:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = T_{\mu\nu}$$

where $g_{\mu\nu}$ is the metric, $R_{\mu\nu}$ is the Ricci tensor, R the curvature scalar and $T_{\mu\nu}$ the energy-momentum tensor. The tangent vectors with respect to the coordinates $\{x^\mu\}$ are noted by ∂_μ. The closure and boundary of a set ζ are $\bar{\zeta}$ and $\partial\zeta$ respectively and the set difference is denoted by $\setminus$. One forms will be denoted by bold symbols ($\boldsymbol{n}$), whereas vector will simply be written in standard style (v).

Chapter 1

Mathematical and physical basis

In this chapter the main mathematical and physical concepts involved in the development of the singularity theorems are presented. Throughout the chapter, reference [1] is strongly used.

1.1 Preliminaries

In General Relativity, the background where events take place is no longer a passive object. Instead, it is a main element in the theory and the behaviour of everything that is in the universe depends on its shape, which in turn depends on everything in it. As it is well known, this object is called spacetime and mathematically it is a manifold with a metric associated. However, physically, we can not accept any manifold or metric, which could have strange physics not corresponding with those of our world, since ultimately the goal of physics is to explain and predict our universe.

Thus, we will only consider 'well-behaved' spacetimes, in the sense of the following definition.

Definition 1.1. ([1] p. 704) A $\mathcal{C}^k$ *spacetime* is a pair (V_4, g), where V_4 is a paracompact, oriented, connected, Hausdorff, 4-dimensional differentiable manifold and g is a $\mathcal{C}^k$ Lorentzian metric, with signature $(-,+,+,+)$.

It may seem that we ask the spacetime a lot of topological properties, nevertheless, they are the ones that allow us to do some key mathematical steps that one would take for granted because of their naturalness. The metric must have this signature in order to recover the special relativity limit. We also require a differentiability grade to the metric in order to achieve as much smoothness as possible. However, this will not always be possible due to the properties of each particular metric, but we will at least demand $k \geq 2-$ (recall that a function f is $\mathcal{C}^{k-}$ if it is $\mathcal{C}^{k-1}$ and its $k-1$ derivatives are locally Lipschitz).

In order to navigate, identify and work with the points of the spacetime, one must introduce a coordinate system. Obviously, there exist not only one coordinate system but infinite, though some of them will have more convenient properties than others. It is worth to remark that the physics of each spacetime is fully contained in the metric and it does not depend on the coordinate system chosen, although sometimes could seem so.

Since the spacetime is a differentiable manifold, at each point $p \in V_4$ one can construct the *tangent space* to V_4: $T_pV_4 \cong \mathbb{R}^4$ with the metric $g|_p : T_pV_4 \times T_pV_4 \longrightarrow \mathbb{R}$. The elements in this vector space are called *tangent vectors*. Since the metric is Lorentzian, it is not definite positive and hence the tangent vectors can be classified as:

Definition 1.2. ([1] p. 704, [2] p. 38) A *vector* $v \in T_pV_4$ is:

- *timelike* if $g(v,v)|_p < 0$.
- *null* or lightlike if $g(v,v)|_p = 0$.
- *spacelike* if $g(v,v)|_p > 0$.
- *causal* if it is timelike or null.

Instead of $g(v,v)$ it is standard to use the Einstein summation convention: $g_{\mu\nu}v^\mu v^\nu = v^\mu v_\mu$, where v^α are the components of the vector: $v = v^\alpha \frac{\partial}{\partial x^\alpha} = v^\alpha \partial_\alpha$ in the basis of the tangent space $\{\partial_\mu\}$ defined by the coordinates $\{x^\mu\}$.

It must be remarked that, since the signature of the metric is $(-,+,+,+)$, there will be some directions in the tangent space with essentially a different nature than the others. These directions intend to represent time somehow and are associated with timelike vectors.

Besides, we can specify a sense of time by choosing an arbitrary timelike vector $t \in T_pV_4$ to be future-pointing at $p \in V_4$. By doing this, all non-zero timelike and null vectors $v \in T_pV_4$ are divided into future pointing if $v^\mu t_\mu < 0$ and past pointing if $v^\mu t_\mu > 0$. If the choice of the reference vector can be made in all the spacetime V_4 continuously, V_4 is called time-orientable. We will assume all the spacetimes to be time-orientable. This endows the tangent space at each point T_pV_4 with the same structure as Minkowski spacetime, namely, a two sheeted light cone separating past and future directions.

A very important object in General Relativity are *curves*. Curves are (possibly smooth) functions $\gamma : I \subseteq \mathbb{R} \longrightarrow V_4$ parametrized by some parameter, which is not unique. However, not all curves are of interest in physics, since they are to represent histories of particles through the universe. In this sense, we want them to have some relevance, given by the following definition.

Definition 1.3. ([1] p. 707) A curve $\gamma(\tau) \subset V_4$ parametrized by τ is *timelike* (resp. *null, spacelike, causal, future-directed, past-directed*) if its tangent vector $u = \frac{d\gamma}{d\tau}$ is timelike (resp. null, spacelike, causal, future-pointing, past-pointing) all along the curve.

Indeed, the curves we will focus on are the ones whose tangent vectors keep their behaviour. One could easily imagine a piecewise spacelike and timelike curve, but its physical interpretation would not be as clear, since by causality we want everything in our spacetime to move in a causal way. In order to study singularities, we will also be highly interested in knowing whether a curve starts or ends at some point or continues forever:

Definition 1.4. ([1] p. 714, [2] p. 184) Let $\gamma : I \longrightarrow V_4$ be a curve and $p \in V_4$ a point. p is a *right* (resp. left) *endpoint* of γ if $\forall U_p$ neighbourhood of p, $\exists u_0 \in I$ such that $\gamma(u) \in U_p$ $\forall u \geq u_0$ (resp. $u \leq u_0$). If a curve has no endpoints, it is *endless* or *inextendible*.

If the curve is causal and future-directed, we just call a right endpoint a future endpoint and a left endpoint a past endpoint. It is clear from the definition that the curve gets confined into a vicinity of the endpoint, independently of whether it has finite or infinite parameter.

Surfaces and Hypersurfaces

As it could not be otherwise, surfaces and hypersurfaces also have a prominent role in General Relativity. These are more complex than curves, since the extra dimensions allow for many new aspects to take into account.

Definition 1.5. ([1] p. 705, [2] p. 44) Let Σ be an orientable 3-dimensional manifold. A *hypersurface* is the image of a continuous piecewise $\mathcal{C}^3$ map $\Phi : \Sigma \longrightarrow V_4$.

By abuse of notation, we usually denote the hypersurface $\Phi(\Sigma)$ by simply Σ. Also, we will refer to the parametric form of the hypersurface as $x^\alpha = \Phi(u^i)$ where u^i are the 3 parameters parametrizing the hypersurface.

Given a hypersurface $\Sigma \subset V_4$ defined by its parametrization $\Phi^\mu(u^i)$, since it is a manifold by itself, we can construct its tangent space $T_p\Sigma \subset T_pV_4$ and use as a basis its natural basis given the parametrization

$$e_i = \frac{\partial \Phi^\mu}{\partial u^i} \frac{\partial}{\partial x^\mu}\bigg|_p .$$

These vectors are called tangent vectors of Σ.

We will be interested in the metric properties of hypersurfaces. These can be summarized by two different tensors, called the first and second fundamental forms.

Definition 1.6. ([1] p. 706, [2] p. 44) Let Σ be a hypersurface of V_4 parametrized by $\Phi^\mu(u^i)$. The *first fundamental form of* Σ is

$$\gamma_{ij} = g_{\mu\nu} \frac{\partial \Phi^\mu}{\partial u^i} \frac{\partial \Phi^\nu}{\partial u^j} = g_{\mu\nu} e_i^\mu e_j^\nu .$$

The first fundamental form is the tensor containing the scalar products of the tangent vectors of Σ and it can be regarded as the metric on Σ, so it is useful to perform metric calculations on Σ. However, if we want to know the way the hypersurface is embedded into V_4, this is not enough. For this, we have to introduce normal vectors.

Definition 1.7. ([1] p. 706, [2] p. 44) A non-zero one-form $\boldsymbol{n}$ defined on a hypersurface Σ with tangent vectors e_i^α is *normal* if $n_\mu e_i^\mu = 0$.

Obviously, the normal vector components will then be $n^\mu = g^{\mu\nu} n_\nu$.

Note that, for hypersurfaces, the normal one-form is defined up to a multiplicative factor. This is because, since a 3-dimensional basis of $T_p\Sigma$ is a subspace of the 4-dimensional T_pV_4 at each point p, and n_μ is orthogonal to each e_i^μ (and hence, to the whole $T_p\Sigma$), if another normal m_μ existed, by orthogonality, it must be parallel to n_μ.

Somehow, normal vectors enable us to describe a 3-dimensional object with a single vector.

Definition 1.8. ([1] p. 707, [2] p. 46) Let Σ be a hypersurface of V_4 parametrized by $\Phi^\mu(u^i)$ and with n_μ its normal form. The *second fundamental form of* n_μ is

$$K_{ij}(\boldsymbol{n}) = -n_\mu e_i^\nu \nabla_\nu e_j^\mu = e_j^\mu e_i^\nu \nabla_\nu n_\mu .$$

The second fundamental form is the tensor containing the projections of the covariant derivative of the normal vector onto the hypersurface.

As with the curves, hypersurfaces can have many shapes and behaviours, and mathematically, they are perfectly fine. However, the interesting ones are those whose behaviour does not change from point to point:

Definition 1.9. ([1] p. 706, [2] p. 44) A hypersurface Σ is *spacelike* (resp. null) if its normal form $\boldsymbol{n}$ is timelike (resp. null) everywhere on Σ.

As for surfaces, the definition and results are analogous with some changes:

Definition 1.10. ([1] p. 705) Let S be an orientable 2-dimensional manifold. A *surface* is the image of a continuous piecewise $\mathcal{C}^3$ map $\Phi : S \longrightarrow V_4$.

The definitions and formulas for surfaces are similar to those for hypersurfaces, but we have to take into account that, since they are 2-dimensional, a basis of their tangent space will have only 2 vectors, hence the first and fundamental forms will now be 2×2 tensors. Also, by the same reason, the normal space to the surface has dimension 2, so we can choose two linearly independent normal forms. If the surface is spacelike (this is, it has at least a timelike normal form), then we can choose the two normal forms to be null. Let them be $k^{\pm}_{\mu}$. Additionally, since the forms are null and cannot be normalised, there are two degrees of freedom. Once we fix one of them, we can normalize the other one with the normalization condition $k^{+}_{\mu} k^{-\mu} = -1$.

Geodesics

Among curves, geodesics are a particularly important group. In General Relativity, free particles with no other influence than themselves will follow the causal geodesics of the spacetime. Conversely, particles with acceleration will not describe geodesics, but other appropriate curves instead. Hence, geodesics can be thought as the curves without acceleration. The formal definition is:

Definition 1.11. ([1] p. 707, [3] p. 6, [4] p. 41) A *geodesic* is a curve $\gamma(u) \subset V_4$ parametrized by u of class $\mathcal{C}^2$ such that its tangent field $v^\mu = \frac{d}{du}\gamma^\alpha(u)$ satisfies $v^\mu \nabla_\mu v^\nu = A(u) v^\nu$ for some function $A(u)$ on the curve.

It is clear that if we solve for v^μ with initial conditions the equation of the condition on the definition we will find the geodesics of the spacetime in question. That equation is called the geodesic equation and it is a non-linear ODE system of order 2. However, it can be slightly simplified. By reparametrizing the curve, it is possible to make the right hand side of the geodesic equation vanish. Consider a new parameter $\tau = \tau(u)$ and call $\lambda = \frac{d\tau}{du}$ its jacobian. Then, by using the chain rule, $v(u) = \frac{d\gamma}{du} = \frac{d\gamma}{d\tau}\frac{d\tau}{du} = \tilde{v}(\tau)\lambda$, where the tilde stands to avoid confusion. Using this in the geodesic equation, we obtain $\lambda \tilde{v}^\alpha \nabla_\alpha (\tilde{v}^\beta \lambda) = A\lambda\tilde{v}^\beta$. The covariant derivative satisfies $\nabla_\alpha(\tilde{v}^\beta\lambda) = \lambda\nabla_\alpha\tilde{v}^\beta + \tilde{v}^\beta\nabla_\alpha\lambda$ and taking into account that $\nabla_\alpha\lambda = \partial_\alpha\lambda$ because it is a scalar, we end up with

$$\lambda^2 \tilde{v}^\alpha \nabla_\alpha \tilde{v}^\beta = \tilde{v}^\beta (A\lambda - \lambda\tilde{v}^\alpha\partial_\alpha\lambda).$$

We see we can make the right hand side vanish if it is fulfilled that $A\lambda = \lambda\tilde{v}^\alpha\partial_\alpha\lambda = \partial_v\lambda = \frac{d\lambda}{du}$. Hence, we can choose λ such that $\frac{d\lambda}{du} = A\lambda$, which leads to

$$\lambda = \lambda_0 e^{\int_{u_0}^{u} A(u)du},$$

taking into account the right initial conditions u_0 and the value of $\lambda_0 = \lambda(u_0)$ at this point. From this we can also see that, since $\frac{d\tau}{du} = \lambda$, then

$$\tau - \tau_0 = \lambda_0 \int_{u_0}^{u} e^{\int_{s_0}^{s} A(s)ds} du,$$

which due to the choice of initial conditions can be written as $\tau = af(u) + b$ where a, b are arbitrary. Such parameters are called *affine parameters* and they remain affine under the transformation $\tilde{\tau} = a\tau + b$, from this their name. All in all, this implies that the geodesic equation for affinely parametrized geodesics can be written as:

$$v^\mu \nabla_\mu v^\nu = 0. \tag{1.1}$$

Note that we are able to talk about timelike, spacelike or null geodesics since the norm of normalized tangent vectors will not change. We will always use the affine parameters when treating with geodesics not only because the problem of finding them becomes simpler, but because they have better properties. Namely, from their transformation law it is clear that if one affine parameter does not diverge, none of them will. However, a non-affine parameter could diverge while affine parameter do not. This will be important later because if a geodesic only attains finite values of its affine parameter, this is indicating a singularity. This fact motivates the following definition:

Definition 1.12. ([1] p. 708, [2] p. 33) A geodesic is *complete* if it is defined for all values of its affine parameter. A spacetime is *geodesically complete at* $p \in V_4$ if all geodesics emanating from p are complete. A spacetime is geodesically complete if it is complete $\forall p \in V_4$.

This concept will be crucial to the singularity theorems, since an incomplete geodesic means that a material particle travelling along it would suddenly disappear in a finite 'time', while a complete geodesic would allow the particle to travel forever. We will come back to this discussion later.

The geodesic equation 1.1 for a affinely parametrized geodesic $\gamma(\tau)$ can also be written in a totally equivalent way by developing the covariant derivative as:

$$\frac{d^2\gamma^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma}\Big|_{\gamma(\tau)} \frac{d\gamma^\nu}{d\tau}\frac{d\gamma^\sigma}{d\tau} = 0, \tag{1.2}$$

where $\Gamma^\mu_{\nu\sigma}$ are the Christoffel symbols, which are to be evaluated at the points of the geodesic. In this form, we are able to clearly see that the geodesic equation is a second order ODE system for $\gamma(\tau)$, hence giving as initial conditions a point $p = \gamma(0)$ and a tangent vector $v = \frac{d\gamma}{d\tau}|_{\tau=0}$ fully determines the solution in a certain interval of τ. We denote this set of solutions by $x^\mu = G(\tau; p, v)$, where $x^\mu = \gamma_{p,v}(\tau)$ is the point of the geodesic that starts at p with velocity v.

This observation provides a very convenient set of coordinates. Consider the exponential map $\exp : O \subseteq T_pV_4 \longrightarrow U \subseteq V_4$, defined by $\exp(v) = G(1; p, v)$ where $0 \in O$ and $p \in U$. Note that $\exp(\tau v) = G(1; p, \tau v) = G(\tau; p, v)$. Then we can naturally define:

Definition 1.13. ([1] p. 708, [2] p. 34) The *normal coordinates based at* p are $\{X^\mu\}$, defined by $x^\mu = G^\mu(1; p, X^\mu)$. A neighbourhood of p with the normal coordinates is called a normal neighbourhood. A *maximal normal neighbourhood* of p is denoted by $\mathcal{N}_p$.

Note that we can always take a convex subset from the set of points where normal coordinates hold and that outside $\mathcal{N}_p$, some geodesics can not be defined.

Normal coordinates are very convenient when working with geodesics emanating from a point p, since the expression of the points along the geodesics in the spacetime becomes simple. Namely, the coordinates of a geodesic with tangent vector v^α at p are simply $x^\alpha(t) = tv^\alpha$.

1.2 Energy conditions

In this section references [2] (Section 4.3) and [3] (Section 2.1) are used. Recalling the Einstein field equations (using natural units $c = 8\pi G = 1$),

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = T_{\mu\nu} \tag{1.3}$$

where $T_{\mu\nu}$ is the energy-momentum tensor, $R_{\mu\nu}$ the Ricci curvature tensor and R the Ricci curvature scalar, it is obvious that the energy density distribution in the spacetime directly affects its geometry. We impose the energy-momentum tensor to be symmetric $T^{\mu\nu} = T^{\nu\mu}$ and to satisfy the energy conservation condition $\nabla_\mu T^{\mu\nu} = 0$. These are standard conditions to the energy distribution in order to avoid degeneration. However, they are not very restrictive and a lot of possibilities are still allowed, many of which are unrealistic. The way in which we impose energy to behave in a convenient way are conditions on the energy-momentum tensor called energy conditions.

Although $T^{\nu\mu}$ is symmetric, its matrix representation $T^\nu_\mu = g_{\nu\rho} T^{\mu\rho}$ is not, hence it will not always diagonalize. Several cases arise then, but the most important is when the T^ν_μ tensor diagonalises, and we call it a type I energy-momentum tensor. The type I energy-momentum tensors are the most common in terms of energy density distributions that describe realistic spacetimes. In this case, there exists an orthonormal basis $\{e^\mu_A\}$ (we will exceptionally use capital latin letters running from 0 to 3 to differentiate each vector in the basis in the following development to avoid confusion) in which it can be written as:

$$T^{\mu\nu} = \rho e^\mu_0 e^\nu_0 + p_1 e^\mu_1 e^\nu_1 + p_2 e^\mu_2 e^\nu_2 + p_3 e^\mu_3 e^\nu_3 = T^{AA} e^\mu_A e^\nu_A,$$

with e^μ_0 timelike and e^μ_i spacelike. Therefore, this basis satisfies the relations

$$g_{\mu\nu} e^\mu_A e^\nu_B = \eta_{AB}, \qquad g^{\mu\nu} = \eta^{AB} e^\mu_A e^\nu_B, \tag{1.4}$$

where η_{AB} is the Minkowski metric. It is immediate that $T_{\mu\nu} e^\mu_A = T^{BB} \eta_{AB} e_{B\nu}$.

Note that this implies that an observer using this basis as a frame of reference will make local measurements as if the metric was flat. Also, this observer will measure:

$$T_{\mu\nu} e^\mu_A e^\nu_A = T_{AA}.$$

The physical interpretation of the obtained eigenvalues is straightforward: ρ represents the energy density as measured by an observer with tangent vector e^μ_0 and p_i represent the principal pressures in the three spacelike directions e^μ_i.

If the three principal pressures are equal, $p_i = p$, then the matter filling the universe is spatially isotropic. This corresponds to a perfect fluid if $e^\mu_0 = u^\mu$ is the velocity vector field of the fluid in the adapted frame of reference. By using the second relation in 1.4, we can rewrite its energy-momentum tensor as:

$$T^{\mu\nu} = \rho u^\mu u^\nu + p\left(e^\mu_1 e^\nu_1 + e^\mu_2 e^\nu_2 + e^\mu_3 e^\nu_3\right) = \rho u^\mu u^\nu + p\left(g^{\mu\nu} + u^\mu u^\nu\right) = \tag{1.5}$$

$$(\rho + p)\, u^\mu u^\nu + p g^{\mu\nu}. \tag{1.6}$$

Similarly, if two of the pressures are equal, $p_1 = p_r$, $p_2 = p_3 = p_T$, then there is a preferred axis in the fluid. Again, in the adapted frame of reference, the energy-momentum tensor can be written as:

$$T^{\mu\nu} = (\rho + p_T)\, u^\mu u^\nu + (p_r - p_T)\, e^\mu_1 e^\nu_1 + p_T g^{\mu\nu}.$$

We now proceed to introduce the relevant energy conditions used in singularity theory. In the following calculations, we use $p_0 = \rho$ to compactify the notation. Any timelike vector ($v^\alpha v_\alpha < 0$) can be written in the orthonormal coordinates in the form $v^\alpha = \gamma\left(e_0^\alpha + a e_1^\alpha + b e_2^\alpha + c e_3^\alpha\right)$ with $a^2 + b^2 + c^2 < 1$, where if we want v^α to be normalized, we need $\gamma = (1 - a^2 - b^2 - c^2)^{-\frac{1}{2}}$. Null vectors can be written as $k^\alpha = \gamma\left(e_0^\alpha + a e_1^\alpha + b e_2^\alpha + c e_3^\alpha\right)$ and must satisfy $a^2 + b^2 + c^2 = 1$. We will denote a, b, c by a^i and we will take $a^0 = 1$, hence the components of the vectors will be γa^μ. Note that the set of null vectors is the boundary of the set of timelike vectors, hence if a continuous function satisfies certain property for timelike vectors, by continuity, it will also hold for null vectors.

Weak energy condition (WEC)

The most basic energy condition is to require the energy density to be positive. The reasonability of this condition is classically undisputed, although quantum effects such as Casimir effect violate it.

Definition 1.14. ([1] p. 722, [2] p. 89, [3] p. 30) The *weak energy condition* is $T_{\alpha\beta} v^\alpha v^\beta \geq 0$, $\forall v^\alpha$ timelike.

By continuity of the energy-momentum tensor and the metric, the WEC also implies $T_{\alpha\beta} k^\alpha k^\beta \geq 0$ $\forall k^\alpha$ null.

Assuming a spacetime has a type I energy-momentum tensor and satisfies the WEC, then the energy density that an observer with tangent vector $v^\alpha = \gamma\left(e_0^\alpha + a e_1^\alpha + b e_2^\alpha + c e_3^\alpha\right)$ measures is:

$$0 \leq T_{\alpha\beta} v^\alpha v^\beta = T_{\alpha\beta} \gamma^2 a^A e_A^\alpha a^B e_B^\beta = \gamma^2 a^A a^B T_{\alpha\beta} e_A^\alpha e_B^\beta = \gamma^2 a^A a^B T^{AA} \eta_{AA} e_{A\beta} e_B^\beta = \\ = \gamma^2 a^A a^B T^{AA} \eta_{AA} \eta_{AB} = \gamma^2 (\rho + a^2 p_1 + b^2 p_2 + c^2 p_3).$$

In particular, taking an observer with tangent vector e_0^μ (this is, $a = b = c = 0$), the WEC implies

$$\rho \geq 0.$$

Similarly, if we choose an observer with tangent vector $e_0^\mu + a^i e_i^\mu$ (fixed i and $0 < a^i \leq 1$) we obtain the condition

$$0 \leq \rho + {a_i}^2 p_i \leq \rho + p_i.$$

This means that the WEC constrains the energy density in the spacetime to be positive. However, the principal pressures can still be negative, but bounded by $p_i \geq -\rho$.

Dominant energy condition (DEC)

A further, while still very reasonable condition, is that in addition to maintain the energy density seen by all the observers positive, the matter has to move in a causal way. $T^{\mu\nu} v_\mu$ is the matter momentum density measured by an observer with tangent vector v^α. Hence, this condition is:

Definition 1.15. ([1] p. 722, [2] p. 91, [3] p. 32) The *dominant energy condition* is $T_{\alpha\beta} v^\alpha v^\beta \geq 0$, $\forall v^\alpha$ timelike and that $T^{\alpha\beta} v_\beta$ must be causal.

We already have seen that the first of the conditions is equivalent to $\rho \geq 0$ and $\rho + p_i \geq 0$.

For $T^{\alpha\beta}v_\beta$ to be causal means that $T^{\mu\nu}v_\mu T_{\nu\gamma}v^\gamma \leq 0$. Hence, with a type I energy-momentum tensor:

$$0 \geq T^{AA}e^\mu_A e^\nu_A T^{BB} e_{B\nu} e_{B\gamma} a^C e_{C\mu} a^D e^\gamma_D = T^{AA}T^{BB}a^C a^D (e_{B\nu}e^\nu_A)(e^\mu_A e_{C\mu})(e_{B\gamma}e^\gamma_D) =$$
$$= T^{AA}T^{BB}a^C a^D \eta_{AB}\eta_{AC}\eta_{DB} = -\rho^2 + a^2 p_1^2 + b^2 p_2^2 + c^3 p_3^2,$$

where $a^2 + b^2 + c^2 \leq 1$. Taking $a = b = c = 0$ leads to $0 \geq -\rho^2$ which is the same $\rho \geq 0$ as before. Taking one of the $a_i \leq 1$ at a time while keeping the others at 0, leads to $0 \geq -\rho^2 + {a_i}^2 p_i^2$ which translates to $\rho^2 \geq a_i^2 p_i^2$. Since we are imposing the condition for $a_i \leq 1$ and we want to ensure its fulfilling for all a_i, we have to impose the least restrictive of them which is $a_i = 1$. Thus, the DEC implies:

$$\rho \geq 0, \quad \rho \geq |p_i|,$$

which can be rewritten by developing the absolute value as:

$$\rho \geq 0, \quad \rho + p_i \geq 0, \quad \rho - p_i \geq 0.$$

Strong energy condition (SEC)

The SEC requires $R_{\alpha\beta}v^\alpha v^\beta \geq 0$, $\forall v^\alpha$ timelike. However, unlike the previous conditions, this one does not have a direct, clear physical interpretation. Its importance comes from the Raychaudhuri equation 1.13 that we will present later, in which this quantity becomes important. The effect of imposing the SEC (through the Raychaudhuri equation) is to focus timelike congruences of curves, forcing them to move closer to each other, although the SEC alone is not sufficient for this to happen.

We would like to express this condition in terms of the energy-momentum tensor in order to be able to call it a proper energy condition. This can be done by using Einstein's equations 1.3. First take its trace by multiplying by $g^{\mu\nu}$:

$$R - \frac{1}{2}4R = T,$$

since $g_{\mu\nu}g^{\mu\nu} = \delta^\mu_\mu = 4$ in our 4-dimensional spacetime, $R = R^\mu_\mu$ and denoting $T = T^\mu_\mu$. Hence, $T = -R$. Finally, using this and equation 1.3 again in the SEC condition we obtain:

Definition 1.16. ([1] p. 722, [2] p. 95, [3] p. 31) The *strong energy condition* is $T_{\alpha\beta}v^\alpha v^\beta \geq \frac{1}{2}Tv^\nu v_\nu$ $\forall v^\alpha$ timelike.

The trace of a type I energy-momentum tensor can be easily calculated:

$$T = T^\mu_\mu = T^{\mu\nu}g_{\mu\nu} = \left(\rho e^\mu_0 e^\nu_0 + p_1 e^\mu_1 e^\nu_1 + p_2 e^\mu_2 e^\nu_2 + p_3 e^\mu_3 e^\nu_3\right) g_{\mu\nu} = p_\alpha \eta_{\alpha\alpha} = -\rho + \sum_i p_i.$$

Since we take v^μ normalized, $\frac{1}{2}Tv^\nu v_\nu = \frac{1}{2}(\rho - \sum_i p_i)$. Hence, the SEC for type I energy-momentum tensors translates into:

$$\rho\left(\gamma^2 - \frac{1}{2}\right) + \sum_i p_i \left({a^i}^2\gamma^2 + \frac{1}{2}\right) \geq 0.$$

Taking as a particular case the observer with tangent vector e_0^μ (this is, all of the $a^i = 0$, which means $\gamma = 1$) we obtain

$$\rho + \sum_i p_i \geq 0. \tag{1.7}$$

Taking the observer with tangent vector $e_0^\mu + a^i e_i^\mu$ (this is, one $a_i \neq 0$ which means $\gamma^2 = \frac{1}{1-a^{i2}}$), a simple calculation leads to

$$\rho + p_i \geq 0.$$

Null convergence condition (NCC)

The NCC is the analogous condition as SEC for null vectors.

Definition 1.17. ([1] p. 721, [2] p. 95 The *null convergence condition* is $R_{\alpha\beta}k^\alpha k^\beta \geq 0$, $\forall k^\alpha$ null.

From similar calculations as before, it can be seen that the NCC translates into $\rho + a^2 p_1 + b^2 p_2 + c^2 p_3 \geq 0$ with $a^2 + b^2 + c^2 = 1$.

As before, there is no direct physical motivation beyond the suggested by the null Raychaudhuri equation 1.15 of focusing null congruences of curves. Note that the SEC implies the NCC.

1.3 Congruences

In the spacetime manifold, there is an infinite number of possible curves to study. However, when considering some specific scenarios, we will not be interested in all of them. Often enough, we will want the curves to represent the evolution of fluid particles, or the possible paths from a particular point. In these cases, we will want to consider a set of curves such that fill a certain region without crossing each others after the initial point in order to represent correctly the mentioned above. This kind of sets of curves are called congruences.

Definition 1.18. ([1] p. 715, [3] p. 36) A *congruence of curves* in a domain $D \subseteq V_4$ is a family of curves Γ, such that $\forall p \in D$, $\exists! \gamma \in \Gamma$ such that $p \in \gamma$. A congruence is timelike (resp. null) if its tangent field is timelike (resp. null).

Then, it is clear that congruences are relevant for the study of the spacetime in the sense that we will not have to treat curves individually, since studying the geometry of the congruence as a whole will give us enough information.

We will develop the theory for timelike and null congruences separately because, although their developments and results are similar, there are some fundamental differences which are worth remarking. Hence, we will first review timelike congruences in detail and then we will do the same for null congruences, omitting the totally analogous parts with the previous case.

Timelike congruences

In this section references [2] (Section 4.1), [3] (Section 2.3) and [4] (Section 9.2) are used. Let u^μ be the tangent vector to a timelike congruence. We can always choose the parameter of the curves in such a way that the tangent vector of the congruence is unitary $u^\mu u_\mu = -1$. Since u^μ is timelike, we can choose three more spacelike vectors orthonormal to u^μ forming a basis of T_pV_4, or in other words, we can naturally separate the tangent space in $T_pV_4 = u(p) \oplus H_p$, with H_p a 3-dimensional spacelike subspace of T_pV_4. This subspace H_p, being orthogonal to u^μ, contains the spatial information of the congruence, which can be used to compute its geometric quantities. To do so, we have to construct the projector onto H_p, given by:

$$h_{\mu\nu} \equiv g_{\mu\nu} + u_\mu u_\nu \quad \Longleftrightarrow \quad h^\mu_\nu = \delta^\mu_\nu + u^\mu u_\nu.$$

In particular, the metric tensor is decomposed as $g_{\mu\nu} = -u_\mu u_\nu + h_{\mu\nu}$. It becomes clear that $h_{\mu\nu}$ will be the metric of a hypersurface orthogonal to u^μ at each point.

Some of its properties are immediate:

$$\begin{aligned} h^\mu_\nu h^\nu_\rho = \left(\delta^\mu_\nu + u^\mu u_\nu\right)\left(\delta^\nu_\rho + u^\nu u_\rho\right) = \delta^\mu_\nu \delta^\nu_\rho + \delta^\mu_\nu u^\nu u_\rho + u^\mu u_\nu \delta^\nu_\rho + u^\mu u_\nu u^\nu u_\rho = \\ = \delta^\mu_\rho + u^\mu u_\rho + u^\mu u_\rho - u^\mu u_\rho = h^\mu_\rho, \\ h^\mu_\mu = \delta^\mu_\mu + u^\mu u_\mu = 4 - 1 = 3, \quad h^\mu_\nu u^\nu = \left(\delta^\mu_\nu + u^\mu u_\nu\right) u^\nu = u^\mu - u^\mu = 0, \quad h_{\mu\nu} = h_{\nu\mu}, \end{aligned}$$

From this, it is clear that $h_{\mu\nu}$ is orthogonal to u^μ, so we can split any tensor parts parallel and orthogonal to u^μ. For instance, any 2-covariant tensor can be written as:

$$\begin{aligned} T_{\mu\nu} = T_{\sigma\rho}\delta^\sigma_\mu \delta^\rho_\nu = T_{\sigma\rho}\left(h^\sigma_\mu - u^\sigma u_\mu\right)\left(h^\rho_\nu - u^\rho u_\nu\right) = \\ = h^\sigma_\mu h^\rho_\nu T_{\sigma\rho} + u^\sigma u_\mu u^\rho u_\nu T_{\sigma\rho} - h^\sigma_\mu u^\rho u_\nu T_{\sigma\rho} - h^\rho_\nu u^\sigma u_\mu T_{\sigma\rho}. \end{aligned}$$

In particular, we can write the covariant derivative of u, which is the relevant quantity we want to compute to know the changes in the geometry of the congruence as:

$$\nabla_\nu u_\mu = h^\sigma_\mu h^\rho_\nu \nabla_\rho u_\sigma + u^\sigma u_\mu u^\rho u_\nu \nabla_\rho u_\sigma - h^\sigma_\mu u^\rho u_\nu \nabla_\rho u_\sigma - h^\rho_\nu u^\sigma u_\mu \nabla_\rho u_\sigma.$$

The acceleration of the congruence is $a^\mu \equiv u^\rho \nabla_\rho u^\mu$ (since it is the derivative of the velocity vector). We observe that this quantity is how much the congruence differs from being affinely parametrized geodesic (recall the geodesic equation 1.1), hence, if $a^\mu = 0$, u^μ is geodesic. Note that the acceleration is also orthogonal to u^μ since $a^\mu u_\mu = u^\rho u_\mu \nabla_\rho u^\mu = \frac{1}{2} u^\rho \nabla_\rho \left(u_\mu u^\mu\right) = \frac{1}{2} u^\rho \nabla_\rho (-1) = 0$.

Using this and the properties

$$\begin{aligned} h^\sigma_\mu a_\sigma = \left(\delta^\sigma_\mu + u^\sigma u_\mu\right) a_\sigma = a_\mu, \\ h^\rho_\nu u^\sigma u_\mu \nabla_\rho u_\sigma = \left(\delta^\rho_\nu + u^\rho u_\nu\right) u^\sigma u_\mu \nabla_\rho u_\sigma = u^\sigma u_\mu \nabla_\nu u_\sigma + u_\mu u^\sigma a_\sigma = 0, \end{aligned}$$

we get:

$$\nabla_\nu u_\mu = h^\sigma_\mu h^\rho_\nu \nabla_\rho u_\sigma + u_\mu u_\nu u^\sigma a_\sigma - h^\sigma_\mu u_\nu a_\sigma - h^\rho_\nu u^\sigma u_\mu \nabla_\rho u_\sigma = h^\sigma_\mu h^\rho_\nu \nabla_\rho u_\sigma - u_\nu a_\mu. \tag{1.8}$$

For the spatial part of the covariant derivative, it is useful to use a decomposition of the total deformation of the congruence in pure deformations which can be more easily understood. This is done by using:

$$h^\rho_\mu h^\sigma_\nu \nabla_\rho u_\sigma = \frac{\theta}{3} h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu},$$

where $\theta \equiv \nabla_\mu u^\mu$ is the expansion scalar, $\sigma_{\mu\nu} \equiv \frac{1}{2}h^\rho_\mu h^\sigma_\nu \left(\nabla_\sigma u_\rho + \nabla_\rho u_\sigma\right) - \frac{\theta}{3}h_{\mu\nu}$ is the shear tensor and $\omega_{\mu\nu} \equiv \frac{1}{2}h^\rho_\mu h^\sigma_\nu \left(\nabla_\sigma u_\rho - \nabla_\rho u_\sigma\right)$ is the rotation tensor. It is straightforward to see that $\sigma_{\mu\nu} = \sigma_{\nu\mu}$, $\omega_{\mu\nu} = -\omega_{\nu\mu}$, $\sigma^\mu_\mu = \omega^\mu_\mu = 0$ and $\sigma_{\mu\nu}u^\mu = \omega_{\mu\nu}u^\mu = 0$. Note that $\omega_{\mu\nu}$ and $\sigma_{\mu\nu}$ are traceless, so that the trace of the spatial part is concentrated in θ.

To understand why these quantities are useful, let us illustrate the 2-dimensional Newtonian case [3] (Section 2.2). For this, imagine a 2-dimensional deformable medium evolving in time as pictured in Figure 1.1. If the vector between two points in the plane is given by ξ^A, then for small distances we can write by approximating at first order

$$\frac{d\xi^A}{dt} \simeq K^A_B \xi^B,$$

where K^A_B is the tensor containing the medium dynamics at first order. Then, for small time intervals dt, then

$$\xi^A(t_0 + dt) = \xi^A(t_0) + K^A_B(t_0)\xi^A(t_0)dt.$$

To see the effect of the deformation tensors that concern us, let us consider a circle in the deformable plane defined by

$$\xi^A = \begin{pmatrix} r\cos\phi \\ r\sin\phi \end{pmatrix},$$

with $\phi \in [0, 2\pi)$ and let us see the resulting shape after a little time interval of being subject to each deformation tensor.

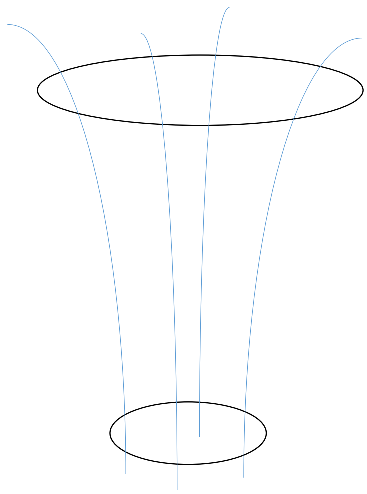

Figure 1.1: Deformation of a 2-dimensional circle in the deformable medium defined by the congruence (blue lines).

By symmetry of the shear tensor, antisymmetry of the rotation tensor and the fact that their traces vanish, in 2D the only possibility is to write :

$$\sigma^B_A = \begin{pmatrix} \sigma_+ & \sigma_\times \\ \sigma_\times & -\sigma_+ \end{pmatrix} \quad , \quad \omega^B_A = \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix},$$

where $A, B = 1, 2$. Through trivial calculations, one can see that the resulting shapes are:

$$\xi_\sigma^A = r \begin{pmatrix} (1 + dt\ \sigma_+) \cos\phi + dt\ \sigma_\times \sin\phi \\ dt\ \sigma_\times \cos\phi + (1 - dt\ \sigma_+) \sin\phi \end{pmatrix} \quad , \quad \xi_\omega^A = \begin{pmatrix} r\cos(\phi - \omega dt) \\ r\sin(\phi - \omega dt) \end{pmatrix}.$$

The radius at each angle of the figure after applying the shear tensor will be, up to first order in dt, $r(\phi) \simeq r(1 + dt\sigma_+ \cos 2\phi + dt\sigma_\times \sin 2\phi)$. As we can see, the effect of the shear tensor on the circle is to deform it into an ellipse with its major axis oriented at an angle determined by its parameters σ_+ and $\sigma_\times$. One can also check that the area of the new figure is still the same that of the circle, so overall, the effect is that of a shearing, and from this the name of the tensor. Similarly, it is trivial to see that the rotation tensor performs a rotation on the circle, also leaving invariant the area. The fact that none of the tensors has trace translates into the fact that they do not change the volume of the figure on which they are acting. All the trace of the spatial part of the covariant derivative is focused on θ, which will be the most important quantity.

To understand its effect, let us study the volume of shapes orthogonal to the congruence. Choose a particular curve γ in the congruence and a point $p \in \gamma$ such that $p = \gamma(\tau_p)$. Then construct a small 3-dimensional neighbourhood $\delta\Sigma(\tau_p)$ of p such that $\delta\Sigma(\tau_p)$ intersects γ orthogonally (and only γ, note that there is no reason to assume that the neighbouring curves intersects $\delta\Sigma(\tau_p)$ orthogonally), and such that through each $p' \in \delta\Sigma(\tau_p)$, there passes a different curve of the congruence γ' fulfilling $\gamma'(\tau_p) = p'$. We can always do this by reparametrizing the curves. This neighbourhood is three dimensional, and since each point in them is from a different geodesic, we can label the geodesics crossing it by some coordinates $\{y^i\}$. As we can do this for each τ, we can obtain a system of coordinates covering all the domain with $\{y^i, \tau\}$.

With this coordinates, since along each geodesic the only coordinate changing its value is τ, $\partial_\tau|_{y^i} = u^\mu$ is simply the tangent vector of the congruence. Let $e_i^\mu = \partial_{y^i}|_\tau$ the tangent vectors to $\delta\Sigma(\tau_p)$. Then, note that $u_\mu e_i^\mu = 0$ (again, only on γ) and that on each of the hypersurfaces, the first fundamental form is $h_{ij} = g_{\mu\nu} e_i^\mu e_j^\nu$, which implies $h_{ij} = h_{\mu\nu} e_i^\mu e_j^\nu$ on γ. The volume element on $\delta\Sigma(\tau_p)$ can be written as $\delta V = \sqrt{h}\ d^3y$ with $h = \det(h_{ij})$. Since y^i are constant along each curve, d^3y does not change when varying τ. Hence the variation of the volume element along the curves of the congruence is given by:

$$\frac{1}{\delta V}\frac{d}{d\tau}\delta V = \frac{1}{\sqrt{h}}\frac{d}{d\tau}\sqrt{h} = \frac{1}{2}h^{ij}\frac{dh_{ij}}{d\tau}.$$

Further developing the derivative of the metric by using that $e_i^\nu \nabla_\nu u^\mu = \partial_{y^i} u^\mu = \partial_{y^i}\partial_\tau x^\mu = \partial_\tau \partial_{y^i} x^\mu = \partial_\tau e_i^\mu = u^\nu \nabla_\nu e_i^\mu$, we get:

$$\frac{dh_{ij}}{d\tau} = \frac{d}{d\tau}\left(g_{\mu\nu} e_i^\mu e_j^\nu\right) = u^\alpha \nabla_\alpha \left(g_{\mu\nu} e_i^\mu e_j^\nu\right) = u^\alpha g_{\mu\nu} e_i^\mu \nabla_\alpha e_j^\nu + u^\alpha g_{\mu\nu} e_j^\nu \nabla_\alpha e_i^\mu =$$

$$= g_{\mu\nu} e_i^\mu e_j^\alpha \nabla_\alpha u^\nu + g_{\mu\nu} e_j^\nu e_i^\alpha \nabla_\alpha u^\mu = e_i^\mu e_j^\alpha \nabla_\alpha u_\mu + e_j^\nu e_i^\alpha \nabla_\alpha u_\nu = \left(\nabla_\mu u_\nu + \nabla_\nu u_\mu\right) e_i^\mu e_j^\nu.$$

Then:

$$\frac{1}{\delta V}\frac{d}{d\tau}\delta V = \frac{1}{2}h^{ij}\left(\nabla_\mu u_\nu + \nabla_\nu u_\mu\right) e_i^\mu e_j^\nu = \frac{1}{2}\left(\nabla_\mu u_\nu + \nabla_\nu u_\mu\right) g^{\mu\nu} = \nabla_\mu u^\mu = \theta.$$

Thus, θ is the variation of the volume element orthogonal to the congruence. Actually, the previous relation, since δV is a scalar, implies

$$\pounds_u V = \theta V. \tag{1.9}$$

Then, if f^i are three independent functions constant along each curve in the congruence, this is, $u^\mu \partial_\mu f^i = 0$, then df^i are orthogonal to u^μ and since they are independent, we can write the volume form as

$$V = v \; df^1 \wedge df^2 \wedge df^3.$$

Calculating the Lie derivative of this expression, we obtain:

$$\pounds_u V = \pounds_u v \; df^1 \wedge df^2 \wedge df^3 + v \; \pounds_u df^1 \wedge df^2 \wedge df^3 + v \; df^1 \wedge \pounds_u df^2 \wedge df^3 + v \; df^1 \wedge df^2 \wedge \pounds_u df^3.$$

Since $\pounds_u df^i = d\left(\pounds_u f^i\right) = d\left(u^\mu \partial_\mu f^i\right) = 0$, then on using 1.9 we simply have $\pounds_u v \; df^1 \wedge df^2 \wedge df^3 = \theta v \; df^1 \wedge df^2 \wedge df^3$ and thus, $\pounds_u v = u^\mu \partial_\mu v = \theta v$, or alternatively,

$$\theta = u^\mu \partial_\mu \left(\ln v\right). \tag{1.10}$$

To prove this statement, we have made use of a new coordinate system to make calculations easier. However, we have remarked that in general, the natural reference frame in those coordinates is not orthogonal (except on γ, where it is by construction). Since the most interesting frames of reference are the orthonormal ones, we are interested in knowing in which congruences it is possible to construct an adequate reference frame. This is given by:

Theorem 1.19. (Frobenius) ([1] p. 717, [3] p. 38) Let u^μ be the tangent vector to a timelike congruence. If the congruence is orthogonal to a hypersurface Σ, then $\omega_{\mu\nu} = 0$.

Proof. For the direct implication, assume u^μ is orthogonal to Σ, hence, it is proportional to the normal form, which in turn is the gradient of a scalar function. Hence, $u_\mu = -f \partial_\alpha \Phi$. To see that the congruence is irrotational, we have to see that $\omega_{\mu\nu} = 0$, which is equivalent to seeing that $\nabla_\nu u_\mu - \nabla_\mu u_\nu = 0$. The computation is straightforward:

$$\nabla_\nu u_\mu - \nabla_\mu u_\nu = -f \partial_\nu \partial_\mu \Phi - \partial_\nu f \partial_\mu \Phi + f \partial_\mu \partial_\nu \Phi + \partial_\mu f \partial_\nu \Phi = \frac{1}{f}\left(u_\mu \partial_\nu f - u_\nu \partial_\mu f\right).$$

Additionally, we know that $\omega_{\mu\nu} u^\nu = 0$ is orthogonal to the congruence. Using this condition, we obtain:

$$0 = \frac{1}{f}\left(u^\nu u_\mu \partial_\nu f - u^\nu u_\nu \partial_\mu f\right) = \frac{1}{f}\left(\partial_\mu f + u^\nu u_\mu \partial_\nu f\right).$$

Then, $\partial_\mu f = -u^\nu u_\mu \partial_\nu f$ and it is direct to see that:

$$\nabla_\nu u_\mu - \nabla_\mu u_\nu = \frac{1}{f}\left(u_\mu \partial_\nu f + u_\nu u^\nu u_\mu \partial_\nu f\right) = \frac{1}{f}\left(u_\mu \partial_\nu f - u_\mu \partial_\nu f\right) = 0.$$

Hence, $\omega_{\mu\nu} = 0$.

For the inverse implication, we assume that $\nabla_\nu u_\mu = \nabla_\mu u_\nu$ and we want to see that $\boldsymbol{u} = d\Phi$ for some function Φ. To see this, let us prove that $\boldsymbol{u} = u_\mu dx^\mu$ is a closed form.

$$d\boldsymbol{u} = d\left(u_\mu dx^\mu\right) = du_\mu \wedge dx^\mu = \left(\partial_\nu u_\mu dx^\nu\right) \wedge dx^m u = \partial_\nu u_\mu dx^\mu \wedge dx^\nu.$$

Since $\nabla_\nu u_\mu = \nabla_\mu u_\nu$, then it follows that $\partial_\nu u_\mu = \partial_\mu u_\nu$, and since $dx^\mu \wedge dx^\nu = -dx^\nu \wedge dx^\mu$, it is clear that $d\boldsymbol{u} = 0$. Thus, by Poincare's Lemma (see reference [5], Theorem 4.11), there exists a scalar function Φ such that $d\Phi = \boldsymbol{u}$, which means that $\boldsymbol{u}$ is proportional to the normal form of a certain hypersurface Σ, hence the congruence is orthogonal to Σ. □

Congruences emanating from a point have a similar property:

Lemma 1.20. ([1] p. 710) The timelike geodesic congruence emanating from a point p with affine parameter τ is orthogonal to the hypersurfaces $\Sigma_{\tau_0} = \{\tau = \tau_0\}$ within $\mathcal{N}_p$.

Proof. In normal coordinates, we can write all the points in space by following the geodesics as $X^\mu = v^\mu \tau$. Plugging this into the geodesic equation 1.2, one obtains $\Gamma^\sigma_{\mu\nu}(\tau v) v^\mu v^\nu = 0$. From the known formula for the Christoffel symbols using the Levi-Civita connection, one trivially gets:

$$(2\partial_\mu g_{\nu\rho}(\tau v) - \partial_\rho g_{\mu\nu}(\tau v))\, v^\mu v^\nu = 0.$$

Since along geodesics the scalar products remain constant, in normal coordinates we can write

$$g_{\mu\nu}(\tau v) v^\mu v^\nu - g_{\mu\nu}(0) v^\mu v^\nu = 0.$$

Differentiating this with respect to v^ρ, one gets:

$$\begin{aligned} 0 =& \partial_\rho \left[g_{\mu\nu}(\tau v) v^\mu v^\nu - g_{\mu\nu}(0) v^\mu v^\nu \right] = \\ =& \partial_\rho g_{\mu\nu}(\tau v) v^\mu v^\nu + 2 g_{\mu\rho}(\tau v) v^\mu - \partial_\rho g_{\mu\nu}(0) v^\mu v^\nu - 2 g_{\mu\rho}(0) v^\mu = 2\frac{d}{d\tau}\left[g_{\rho\mu}(\tau v) \tau v^\mu - g_{\rho\mu}(0) \tau v^\mu \right]. \end{aligned}$$

From this we conclude that:

$$g_{\mu\nu}(X) X^\mu = g_{\mu\nu}(0) X^\mu. \tag{1.11}$$

In normal coordinates, the tangent vector of the normalized geodesic congruence emanating form p can be simply written as $u = \frac{X^\mu}{\sqrt{-g_{\mu\nu}(0) X^\mu X^\nu}} \partial_\mu$, and its corresponding one-form as

$$u = \frac{g_{\mu\nu}(X) X^\mu}{\sqrt{-g_{\mu\nu}(0) X^\mu X^\nu}} dX^\nu = \frac{g_{\mu\nu}(0) X^\mu}{\sqrt{-g_{\mu\nu}(0) X^\mu X^\nu}} dX^\nu = -d\left(\sqrt{-g_{\mu\nu}(0) X^\mu X^\nu} \right),$$

by using equation 1.11. Then, the congruence is orthogonal to the hypersurfaces

$$\tau = \sqrt{-g_{\mu\nu}(0) X^\mu X^\nu} = \mathit{constant}.$$

□

Since θ is the most relevant geometrical quantity in the congruence, we are interested in tracking the evolution of θ along the curves of the congruence. Its variation will be given by $\frac{D\theta}{d\tau} = u^\mu \nabla_\mu \theta = u^\mu \nabla_\mu \nabla_\nu u^\nu$ (keeping in mind that, since θ is a scalar, $\nabla_\mu \theta = \partial_\mu \theta$). Using the trace of the Ricci identity

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)\, u^\alpha = R^\alpha_{\rho\mu\nu} u^\rho, \tag{1.12}$$

contracted with u^μ one gets the relation

$$u^\mu (\nabla_\mu \nabla_\nu u^\nu - \nabla_\nu \nabla_\mu u^\nu) = -R_{\mu\nu} u^\mu u^\nu.$$

Furthermore, we can use that $\nabla_\nu (u^\mu \nabla_\mu u^\nu) = \nabla_\nu u^\mu \nabla_\mu u^\nu + u^\mu \nabla_\nu \nabla_\mu u^\nu$. Then, summing up, we can write:

$$u^\mu \partial_\mu \theta = u^\mu \nabla_\mu \theta = \nabla_\nu (u^\mu \nabla_\mu u^\nu) - \nabla_\nu u^\mu \nabla_\mu u^\nu - R_{\mu\nu} u^\mu u^\nu = \nabla_\nu a^\nu - \nabla_\nu u_\mu \nabla^\mu u^\nu - R_{\mu\nu} u^\mu u^\nu.$$

As we had seen before, the covariant derivative term can be written as $\nabla_\nu u_\mu = -u_\nu a_\mu + \frac{\theta}{3}h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}$. Since $a_\mu, h_{\mu\nu}, \sigma_{\mu\nu}, \omega_{\mu\nu}$ are orthogonal to u^μ, when developing the term $\nabla_\nu u_\mu \nabla^\mu u^\nu$, the only surviving contributions will be:

$$\nabla_\nu u_\mu \nabla^\mu u^\nu = \frac{\theta^2}{9} h_{\mu\nu} h^{\nu\mu} + \sigma_{\mu\nu}\sigma^{\nu\mu} + \omega_{\mu\nu}\omega^{\mu\nu}.$$

As seen before, $h_{\mu\nu}h^{\nu\mu} = h^\mu_\mu = 3, \sigma_{\mu\nu} = \sigma_{\nu\mu}, \omega_{\mu\nu} = -\omega_{\nu\mu}$, and putting all the pieces together, we finally get Raychaudhuri's equation ([1] p. 717, [3] p. 40):

$$u^\mu \partial_\mu \theta + \frac{\theta^2}{3} - \nabla_\mu a^\mu - \omega_{\mu\nu}\omega^{\mu\nu} + \sigma_{\mu\nu}\sigma^{\mu\nu} + R_{\mu\nu}u^\mu u^\nu = 0. \tag{1.13}$$

This equation is key to control the evolution of the congruence. In particular, it determines the behaviour of the volume element orthogonal to the congruence in equation 1.10. It rapidly becomes evident that if the volume element becomes 0 at some point, or equivalently $\theta \to -\infty$, the congruence as we have defined it will end there because the curves will meet. This motivates the following definition:

Definition 1.21. ([1] p. 718, [4] p. 223) A point $q \in V_4$ is *conjugate to a point* $p \in V_4$ if along a curve of the geodesic timelike congruence emanating from p it holds that $\lim_{x\to q} \theta(x) = -\infty$.

A point $q \in V_4$ is *focal to a spacelike hypersurface* $\Sigma \subset V_4$ if along a curve of the geodesic timelike congruence emanating orthogonally from Σ it holds that $\lim_{x\to q} \theta(x) = -\infty$.

Clearly, if $\theta < 0$, the geodesics in the congruence will get closer. However, in the extreme case of conjugate and focal points where $\theta \to -\infty$, the geodesics will be so highly concentrated that they will eventually meet at those points, as illustrated in Figure 1.2.

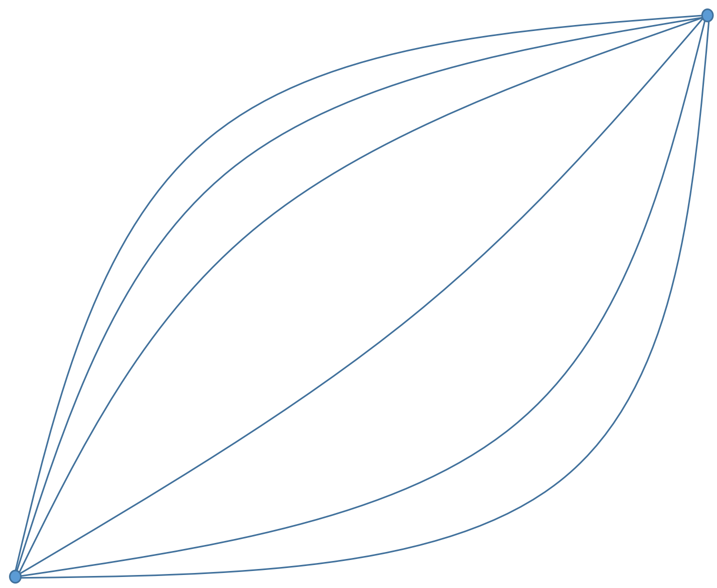

Figure 1.2: Two conjugate points. The geodesics in the congruence emanating from one point reconverge to its conjugate point.

As we will see later, one can assure the existence of conjugate or focal points if certain conditions are met. To that end, let us introduce now a new condition:

Definition 1.22. ([1] p. 721, [2] p. 101) The *timelike generic condition* holds if in each timelike geodesic with tangent vector v^μ, the condition $v_{[\alpha}R_{\rho]\mu\nu[\sigma}v_{\beta]}v^\mu v^\nu \neq 0$ is fulfilled at some point.

However, because of the antisymmetry properties of the Riemman tensor, one can rewrite the timelike generic condition easily as $R_{\rho\mu\nu\sigma}u^\mu u^\nu \neq 0$, and it can be proved that if the SEC is fulfilled strictly ($R_{\mu\nu}u^\mu u^\nu > 0$ for timelike u^μ), then the timelike generic condition holds. This condition, as its name already suggests, is a very generic condition that most of the spacetimes fulfill. Actually, its only purpose is to exclude from the singularity theorems a very specific selection of spacetimes. The condition is interpreted as the fact that every free particle has to interact with gravity in some way.

Null congruences

In this section reference [3] (Section 2.4) is used. We can develop the same ideas as before for null congruences, obtaining pretty similar results.

Let k^μ be the tangent vector to a null geodesic (not necessarily affinely parametrized) congruence, with $k^\mu k_\mu = 0$. In the timelike case, we took a basis of each T_pV_4 including the tangent vector to the timelike congruence. However, we can not do the same in the null case, because the metric has signature $(-,+,+,+)$, meaning that every vector in T_pV_4 must be able to be written as a linear combination of one timelike and three spacelike vectors. Since our available vector k^μ is null, we can not simply choose three more spacelike vectors to complete the basis, and in the same way we can not define a projector orthogonal to k^μ to track the variation in the geometry of the congruence by the same means as before.

The solution is to introduce a new null vector field l^μ linearly independent together with k^μ. Due to the 0 norm of k^μ and l^μ, we can also impose the normalization condition $l_\mu k^\mu = -1$. Note that now we are able to choose 2 linearly independent spacelike vectors to form a basis together with k^μ and l^μ.

Now, we can construct a projector truly orthogonal to k^μ defined by:

$$N_{\mu\nu} = g_{\mu\nu} + k_\mu l_\nu + k_\nu l_\mu \iff N^\mu_\nu = \delta^\mu_\nu + k_\nu l^\mu + k^\mu l_\nu,$$

and satisfying

$$\begin{gathered} N^\mu_\nu k^\nu = \left(\delta^\mu_\nu + k_\nu l^\mu + k^\mu l_\nu\right) k^\nu = k^\mu + l^\mu(k_\nu k^\nu) + k^\mu(l_\nu k^\nu) = k^\mu - k^\mu = 0, \\ N_{\mu\nu} = N_{\nu\mu}, \qquad N^\mu_\mu = \delta^\mu_\mu + k_\mu l^\mu + k^\mu l_\mu = 4 - 1 - 1 = 2, \qquad N^\mu_\nu l^\nu = 0, \end{gathered}$$

$$\begin{aligned} N^\mu_\nu N^\nu_\rho = &\left(\delta^\mu_\nu + k_\nu l^\mu + k^\mu l_\nu\right)\left(\delta^\nu_\rho + k_\rho l^\nu + k^\nu l_\rho\right) = \\ =&\delta^\mu_\rho + k_\rho l^\mu + k^\mu l_\rho + k_\rho l^\mu - l^\mu k_\rho + k^\mu l_\rho - k^\mu l_\rho = \delta^\mu_\rho + k_\rho l^\mu + k^\mu l_\rho = N^\mu_\rho. \end{aligned}$$

Since the spacelike subspace of the tangent space has now dimension 2, it seems logical that we can build an orthogonal surface to k^μ. We can not do it genuinely orthogonal to a hypersurface, because if k^μ is the normal vector of a hypersurface, since $k^\mu k_\mu = 0$, then it is also tangent to it. But we can find a surface such that the two null vectors k^μ and l^μ are orthogonal to it, and we can complete a basis of the tangent space with two more spacelike vectors tangent to the surface.

As we intend to prove the null version of the Raychaudhuri equation, the steps to follow are clear after the proof in the timelike case. We first separate the covariant derivative of k_μ, by using that $k^\mu \nabla_\nu k_\mu = 0$:

$$\begin{aligned}\nabla_\nu k_\mu &= \delta^\sigma_\mu \delta^\rho_\nu \nabla_\rho k_\sigma = \left(N^\sigma_\mu - k_\mu l^\sigma - k^\sigma l_\mu\right)\left(N^\rho_\nu - k_\nu l^\rho - k^\rho l_\nu\right)\nabla_\rho k_\sigma = \\ &= N^\sigma_\mu N^\rho_\nu \nabla_\rho k_\sigma - k_\mu l_\nu k^\rho l^\sigma \nabla_\rho k_\sigma - k_\mu k_\nu l^\rho l^\sigma \nabla_\rho k_\sigma - k_\nu l^\rho \nabla_\rho k_\mu - k^\rho l_\nu \nabla_\rho k_\mu - k_\mu l^\sigma \nabla_\nu k_\sigma = \\ &= N^\sigma_\mu N^\rho_\nu \nabla_\rho k_\sigma - k_\mu k_\nu l^\rho l^\sigma \nabla_\rho k_\sigma - k_\nu l^\rho \nabla_\rho k_\mu - k_\mu l^\sigma \nabla_\nu k_\sigma, \quad (1.14)\end{aligned}$$

since $k^\alpha \nabla_\alpha k^\beta = f k^\beta$.

From all these terms, the only genuinely orthogonal part to k^μ is $N^\sigma_\mu N^\rho_\nu \nabla_\rho k_\sigma$, because on being contracted with all $k^\mu, k^\nu, l^\mu, l^\nu$ it vanishes, unlike the other terms. Hence, as again, we can decompose the purely spatial part as:

$$N^\sigma_\mu N^\rho_\nu \nabla_\rho k_\sigma = \frac{1}{2}\vartheta N_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu},$$

with $\vartheta \equiv \nabla_\mu k^\mu + k^\mu l^\nu \nabla_\mu k_\nu$, obtained from the trace of equation 1.14, and

$$\begin{aligned}\sigma_{\mu\nu} &\equiv \frac{1}{2} N^\rho_\mu N^\sigma_\nu \left(\nabla_\sigma k_\rho + \nabla_\rho k_\sigma\right) - \frac{\vartheta}{2} N_{\mu\nu}, \\ \omega_{\mu\nu} &\equiv \frac{1}{2} N^\rho_\mu N^\sigma_\nu \left(\nabla_\sigma k_\rho - \nabla_\rho k_\sigma\right).\end{aligned}$$

Similarly, using the above expression for the covariant derivative of k^μ and the trace of the Ricci identity, we can derive the analogous to the Raychaudhuri equation in the null case ([3] p. 58):

$$k^\mu \partial_\mu \vartheta + \vartheta k^\mu l^\nu \nabla_\mu k_\nu + \frac{\vartheta^2}{2} + \sigma_{\mu\nu}\sigma^{\mu\nu} - \omega_{\mu\nu}\omega^{\mu\nu} + R_{\mu\nu} k^\mu k^\nu = 0. \quad (1.15)$$

The result is essentially the same as 1.13, with just a factor 2 as a difference, and hence, the behaviour of ϑ will be similar to the behaviour of θ for timelike congruences. Being this way, we can also define the concept of conjugate and focal points with null congruences as in the timelike case:

Definition 1.23. ([1] p. 725) A point $q \in V_4$ is *conjugate to a point* $p \in V_4$ if along a curve of the geodesic null congruence emanating from p it holds that $\lim_{x\to q} \vartheta(x) = -\infty$.

A point $q \in V_4$ is *focal to a spacelike surface* $S \subset V_4$ if along a curve of the geodesic null congruence emanating orthogonally from S it holds that $\lim_{x\to q} \vartheta(x) = -\infty$.

As said before, we want to know the conditions under which the existence of focal or conjugate points is assured. Again, we will need to define another condition:

Definition 1.24. ([1] p. 721, [2] p. 101) The *null generic condition* holds if in each null geodesic with tangent vector k^μ, the condition $k_{[\alpha} R_{\rho]\mu\nu[\sigma} k_{\beta]} k^\mu k^\nu \neq 0$ is fulfilled at some point. The *generic condition* holds if both the timelike and null generic conditions hold.

The interpretation of the null generic condition is no different from the timelike case, although we can not rewrite it in the same way. Though, there exists an easier way to check whether it is fulfilled:

Proposition 1.25. ([1] p. 723) Let k^μ be the tangent vector to a null geodesic. If the NCC ($R_{\mu\nu} k^\mu k^\nu > 0$) holds, the null generic condition is satisfied for this geodesic.

1.4 Maximal curves

Since we are in a Lorentzian space and vectors may have positive, negative or zero norm, the metric properties are not alike those of Riemannian spaces. In these, since all non-zero tangent vectors have positive norm, it makes sense to talk about curves of minimum length between two points. With a Lorentzian metric, this is not like this anymore, although we can still define a notion of length of curves in the standard way:

Definition 1.26. ([1] p. 725, [2] p. 105, [4] p. 233) The *length* of a piecewise differentiable curve $\gamma(u)$ with tangent vector $v(u)$ between $p = \gamma(u_1), q = \gamma(u_2) \in V_4$ is

$$L(p, q; \gamma) = \int_{u_1}^{u_2} \sqrt{|g_{\mu\nu} v^\mu v^\nu|} du. \tag{1.16}$$

The integral has to be computed over the differentiable segments, and it is well defined since it is invariant under reparametrizations of the curve, since if $\tilde{u}$ is the new parameter obeying $d\tilde{u} = \lambda\, du$, then:

$$L(p, q; \gamma) = \int_{\tilde{u}_1}^{\tilde{u}_2} \sqrt{|g_{\mu\nu} \lambda \tilde{v}^\mu \lambda \tilde{v}^\nu|}\, \frac{d\tilde{u}}{\lambda} = \int_{\tilde{u}_1}^{\tilde{u}_2} \sqrt{|g_{\mu\nu} \tilde{v}^\mu \tilde{v}^\nu|} d\tilde{u}.$$

It is obvious that $L = 0$ for null curves either geodesics or sequences of null geodesics segments, and for timelike curves, we have $L > 0$. Note that now, for curves joining two given points, the minimum possible arc length is $L = 0$, since any causal curve can be approximated by null segments in zigzag as shown in Figure 1.3.

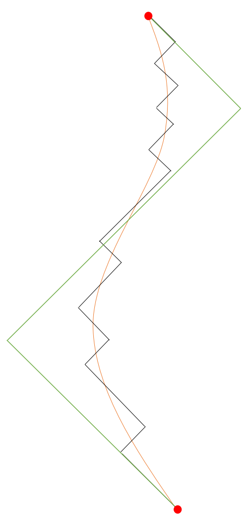

Figure 1.3: Any timelike curve (orange) can be approximated by curves formed of null geodesic segments in a zig zag (black and green).

However, there may be a maximum in the length of causal curves joining two points because a causal curve is relatively constrained when going from one point to another as its tangent vector must remain inside the light cone. This motivates the following definition:

Definition 1.27. ([1] p. 725) Let $p, q \in V_4$ be points and γ_0 a causal curve joining p and q. γ_0 is *maximal* if $\forall\gamma$ causal curve joining p and q holds $L(p, q; \gamma_0) \geq L(p, q; \gamma)$. γ_0 is *locally maximal* if $\exists\ U_{\gamma_0}$ neighbourhood of γ_0 such that $\forall\gamma \subseteq U_{\gamma_0}$ holds $L(p, q; \gamma_0) \geq L(p, q; \gamma)$.

1.5 Causality

In this section references [2] (Chapter 6) and [4] (Chapter 8) are used. The concept of causality in a general spacetime is not as simple as in $\mathbb{R}^4$ with Minkowski's metric, used in special relativity. In the latter, there is a two-sheeted light cone at each point, one sheet corresponding to the future, one to the past and timelike trajectories can only exist inside the cone, while null geodesics can only exist on the cone. Furthermore, this cone is always oriented the same way, so as a result, its causal structure is quite simple. However, when moving to a general spacetime, the topology can become much more complicated than that of a flat space. The light cones here are defined as:

Definition 1.28. ([1] p. 711, [4], p. 189) The *future light cone* in $p \in V_4$ is

$$\partial C_p^+ = \{\exp(v) \mid v \in \mathcal{N}_p \subset T_pV_4, v \text{ future directed, } g_{\mu\nu}v^\mu v^\nu = 0\}.$$

The *interior of the future light* cone is

$$C_p^+ = \{\exp(v) \mid v \in \mathcal{N}_p \subset T_pV_4, v \text{ future directed, } g_{\mu\nu}v^\mu v^\nu < 0\}.$$

Analogously with the past light cone.

One must notice that the light cones belong to the tangent spaces of each point. In the trivial Minkowski spacetime this happened too, but due to the trivial topology, one can identify the vectors in the tangent space with the points of the spacetime themselves. Now in general, the light cones change orientation from point to point and, in addition, some regions could be cut out of the spacetime, so to take this into account, we restrict the light cone to $\mathcal{N}_p \subset T_pV_4$, where we can assure that the geodesics are well behaved.

Proposition 1.29. ([1] p. 711) Let γ be a curve starting at p completely contained in $\mathcal{N}_p$. γ is a null geodesic if and only if $\gamma \subset \partial C_p^+$.

Proof. Let us use normal coordinates, in which the point p is the point 0. Let $\gamma^\mu(u)$ be the coordinates of a general curve in normal coordinates. Note that $\gamma^\mu(u)$ is geodesic if and only if $\gamma^\mu(u) = uv^\mu$. Let us denote the tangent vector of the curve $\gamma^\mu(u)$ as $v^\mu(u)$.

If γ is a null geodesic, then, $g_{\mu\nu}(0)\gamma^\mu(u)\gamma^\nu(u) = 0$ and $\gamma^\mu(u) = uv^\mu$. But this means that γ is contained into the light cone.

Conversely, if $\gamma \subset \partial C_p^+$, then $g_{\mu\nu}(0)\gamma^\mu(u)\gamma^\nu(u) = 0$. Differentiating this relation with respect to u, we find that $g_{\mu\nu}(0)v^\mu(u)\gamma^\nu(u) = 0$. Then, using equation 1.11, it follows that

$$g_{\mu\nu}(\gamma)\gamma^\mu(u)\gamma^\nu(u) = 0, \qquad g_{\mu\nu}(\gamma)v^\mu(u)\gamma^\nu(u) = 0.$$

Since v^μ is tangent to γ, which is null by the first equality, and v^μ and γ^μ are orthogonal by the second, then it must be that $v^\mu = f(u)\gamma^\mu$, hence v^μ is null too. Therefore, γ is a null geodesic. $\square$

Lemma 1.30. ([1] p. 726) The causal geodesics from p are maximal in $\mathcal{N}_p$.

Proof. Take $q \in \partial C_p^+ \cup C_p^+$. If $q \in \partial C_p^+$, by Proposition 1.29, the only causal curve joining p and q is the null geodesic and it is necessarily maximal.

If $q \in C_p^+$, by Lemma 1.20, the congruence of timelike geodesics emanating from p is orthogonal to the hypersurfaces $\tau = \sqrt{-g_{\mu\nu}(0)X^\mu X^\nu} = constant$ (using normal coordinates). We can define then a new coordinate system $\{y^\mu\} = \{\tau, y^i\}$ with $y^i = \frac{X^i}{\sqrt{-g_{\mu\nu}(0)X^\mu X^\nu}}$. By doing this, the line element has transformed into

$$ds^2 = -d\tau^2 + g_{ij}(y^\mu)dy^i dy^j,$$

where g_{ij} is positive definite.

Consider any curve from p to q parametrized by $\tau = u, y^i = \gamma^i(u)$. Then, its length 1.16 is

$$L(p, q; \gamma) = \int_0^{u_q} \sqrt{1 - g_{ij}(\gamma)\frac{d\gamma^i}{du}\frac{d\gamma^j}{du}} du.$$

Since g_{ij} is positive definite, $g_{ij}(\gamma)\frac{d\gamma^i}{du}\frac{d\gamma^j}{du} \geq 0$, so it is clear that the maximum length is attained when γ^i are constant, which are the timelike geodesics form p. □

Proposition 1.31. ([1] p. 729, [2] p. 115) Let γ be a causal curve from p to q. There is no neighbourhood of γ containing a timelike curve from p to $q \iff \gamma$ is a null geodesic segment from p to q with no conjugate point to p between p and q.

Proof. The inverse implication stands because since if γ is a null geodesic without conjugate points, then it is defined by Lemma 1.30. □

However, these sets are not as useful. In order to develop a causality theory, we will need to generalize them. Namely, the light cones can not be defined outside $\mathcal{N}_p$, which is inconvenient, since as long as just one of the geodesics can not be defined from some point on, $\mathcal{N}_p$ ends and this can easily happen when singularities are involved. Instead, the used sets are more general, while keeping the same idea:

Definition 1.32. ([1] p. 730, [2] p. 182-184, [4] p. 191) The *chronological future* of $p \in V_4$ is

$$I^+(p) = \{x \in V_4 | \text{ exists a future-directed timelike curve from } p \text{ to } x\}.$$

The *causal future* of p is

$$J^+(p) = \{x \in V_4 | \text{ exists a future-directed causal curve from } p \text{ to } x\}.$$

The *future horismos* of p is

$$E^+(p) = J^+(p) \setminus I^+(p).$$

For a set $\zeta \subseteq V_4$, its chronological future is $I^+(\zeta) = \bigcup_{p\in\zeta} I^+(p)$. Analogously for $J^+(\zeta), E^+(\zeta)$. The definitions with past instead of future are completely analogous and are denoted by a $-$ instead of a $+$.

The relation between these sets and the light cone is immediate. Choose a point p and consider the spacetime $(\mathcal{N}_p, g)$. Then, in this spactime, $I^+(p) = C_p^+$, $E^+(p) = \partial C_p^+$ and $J^+(p) = C_p^+ \cup \partial C_p^+$. However, outside the maximal normal neighbourhood, where geodesics do not behave as well as we wished, these equalities do not hold. This is the reason one uses in general the sets $I^+(\zeta), J^+(\zeta), E^+(\zeta)$ as this approach allows to treat the causality in a more general way than the light cones.

Note that the trivial curve from p to p, which we can treat as a null curve since it has 0 length, is also in $J^+(p)$, so $\zeta \subset J^+(\zeta)$. It is obvious by Proposition 1.31 that if $q \in E^+(p)$, then any causal curve connecting p and q must be a null geodesic segment, since no timelike curve can reach q.

Proposition 1.33. ([1] p. 731, [2] p. 182) $I^+(\zeta)$ is open and $I^+(\zeta) = I^+(\overline{\zeta})$

Proof. For the first part, let $p \in \zeta$, $q \in I^+(p)$ and γ be a timelike curve joining p and q. Consider $\mathcal{N}_q$ and its subset C_q^-, the past interior of the light cone, which is open by construction and choose a point $r \in C_q^- \cap \gamma$. Then, $q \in C_r^+$ and every point in C_r^+ can be reached by p by taking γ from p to r and then a suitable curve in C_r^+. Then, $q \in C_r^+ \subset I^+(p)$ the interior of the light cone is an open neighbourhood of q. Hence, $I^+(p)$ is open. Since the arbitrary union of open sets is open, it follows that $I^+(\zeta)$ is open.

As for the second part, the inclusion $I^+(\zeta) \subseteq I^+(\overline{\zeta})$ is obvious. For the converse, take a point $p \in I^+(\overline{\zeta})$. Then there is a point $q \in \overline{\zeta}$ that fulfills $q \in I^-(p)$. Since $I^-(p)$ is open, there is a neighbourhood of q totally contained in $I^-(p)$, but since $q \in \overline{\zeta}$, this neighbourhood must also contain some point $r \in \zeta$. Then, $r \in I^-(q)$ hence $p \in I^+(r) \subseteq I^+(\zeta)$. □

However, $J^+(\zeta)$ is not always closed, because of the possibility of missing points in the manifold. Note that in the example taking as the spacetime $(\mathcal{N}_p, g)$, $J^+(\zeta)$ is closed because of the simple properties of $\mathcal{N}_p$.

As it seems evident, the chronological and causal future give a lot of information about the spacetime as well as about their base set. We provide now some further useful classification of sets:

Definition 1.34. ([1] p. 731, [2] p. 186) An *achronal* set $\zeta \subseteq V_4$ is a set such that $I^+(\zeta) \cap \zeta = \varnothing$. An *acausal* set $\zeta \subseteq V_4$ is a set such that $J^+(\zeta) \cap \zeta = \varnothing$.

Note that if $I^-(\zeta) \cap \zeta = \varnothing$, ζ is achronal as well. It is clear that in an achronal set, two of its point can not be connected by any timelike curve, though they still could be joined by some null curves, whereas in acausal sets, points can not be connected by causal curves. Hence, achronal and acausal sets are in some way patches of the spacetime in a certain 'instant of time'.

Definition 1.35. ([1] p. 731, [2] p. 186) A *future set* $\zeta \subseteq V_4$ is a set such that $I^+(\zeta) \subseteq \zeta$.

It follows from the definition that a future set has to be necessarily 'big' in order to contain its chronological future. Note that for every ζ, $I^+(\zeta)$ is a future set.

The boundary of $I^+(\zeta)$, $\partial I^+(\zeta) = \partial J^+(\zeta)$, of any set ζ is achronal, because if a point $p \in \partial I^+(\zeta)$ could be reached by a timelike curve from a point in $\partial I^+(\zeta)$, then $p \in I^+(\zeta)$ and since $I^+(\zeta)$ is open, p can not belong to it and its boundary simultaneously. This has a special name:

Definition 1.36. ([1] p. 732, [2] p. 187) A *proper achronal boundary* is the boundary of a future set.

Proposition 1.37. ([1] p. 733, [2] p. 187) Any point in a proper achronal boundary is either contained in an acausal set, contained in a null geodesic segment or it is a past or future endpoint of the null geodesic segments.

In order to endow the spacetime with a sense of causality, this is, to respect the 'flow of time', one must impose further conditions, because until now, nothing forbids a causal curve to be in a closed loop or to influence its own past. Historically, the development of causality conditions led to a cascade of more and more restrictive conditions, since none of the previous solved the problem totally. Here we present some of the relevant causality conditions:

Definition 1.38. ([1] p. 734, [2] p. 189, 192, [4] p. 196) A spacetime (V_4, g) satisfies

- the *chronology condition* at $p \in V_4$ if $p \notin I^+(p)$.
- the *strong causality condition* at $p \in V_4$ if there are arbitrarily small neighbourhoods of p, $\mathcal{U}_p$, such that for all future directed causal curves γ, $\gamma \cap \mathcal{U}_p$ is not a disconnected set.

The chronology condition simply forces that an event p can not influence itself in the future. If this happened, past could be 'changed' since we could follow a trajectory and return to the same point for a second time. However, it could happen that even not passing through the same exact event p, a curve could influence a part of $I^+(p)$ by passing close to p. This is forbidden by the strong causality condition, since it imposes that a curve can not come back near to one of its points.

Still, both of the above conditions fail to avoid paradoxes, because they still allow for n points affecting the future of each other simultaneously, which clearly is a causal violation. Hence, it is needed another approach to avoid paradoxes. A way to ensure that, plainly speaking, 'future can not affect past' is given by the following idea:

Definition 1.39. ([1] p. 734, [2] p. 198) A spacetime (V_4, g) satisfies the *stable causality condition* if there exists a function f such that df is timelike everywhere.

Assume the condition holds and that $-df = -\partial_\mu f dx^\mu$ is future pointing. Let k^μ be a future-directed causal curve. Then, since both are future pointing, $-k^\mu \partial_\mu f < 0$. This means that $\frac{Df}{d\tau} > 0$, so f increases along every future directed causal curve. Furthermore, then the hypersurfaces $\{f = constant\}$ are spacelike because $-df$ is timelike. Hence, each causal curve can only intersect each of these hypersurfaces once. For this reason, f is called a 'time function', and it foliates the spacetime in 'instants of time', avoiding causality violations. Note that if the stable causality condition holds, then the chronology condition holds ([1] p. 739, [4] p. 199). However, not all problems are solved since this does not impose anything on the topology of the spacetime and some strange events can still happen.

Finally, the most important causality condition is a generalization of the strong causality condition:

Definition 1.40. ([1] p. 734, [2] p. 206) A spacetime (V_4, g) is *globally hyperbolic* if it satisfies the strong causality condition and $J^+(p) \cap J^-(q)$ is compact $\forall p, q \in V_4$.

Although this is a causality condition, its fulfilling implies many good properties of the spacetime which are extremely useful. We will discuss this later.

Now that the causality conditions are well established, we can take into account that in General Relativity information can not travel faster than light. We will see next that given a region Σ satisfying certain conditions, there will be a certain region in which everything will be entirely determined by Σ. This idea can be regarded as a kind of initial conditions from which all the information in some region can be predicted given that we know the appropriate initial information. To formalize this idea, we must introduce new sets:

Definition 1.41. ([1] p. 741, [2], p. 201, 202, [4] p. 200) The *future Cauchy domain of dependence* of a set ζ is

$$D^+(\zeta) = \{x \in V_4 \mid \forall \gamma \text{ past directed endless causal curve from } x, \gamma \cap \zeta \neq \varnothing\}.$$

The *future Cauchy horizon* of a set ζ is

$$H^+(\zeta) = \overline{D^+(\zeta)} \setminus I^- \left[D^+(\zeta)\right].$$

Analogously for past Cauchy domain and horizon. The full Cauchy domain of dependence is $D(\zeta) = D(\zeta)^+ \cup D(\zeta)^-$. The full Cauchy horizon is $H(\zeta) = H(\zeta)^+ \cup H(\zeta)^-$.

Their physical intuition is quite clear. If matter and energy propagate causally, every point in $D^+(\zeta)$ is exclusively influenced by matter and energy in ζ. Also, information in every point in $D^-(\zeta)$ can be predicted by knowing the initial information on ζ. Finally, $D(\zeta)$ is the set of events of which we can predict all the information by just its knowledge at ζ whereas $H(\zeta)$ is the 'limit' of what can be predicted. It seems clear that these definitions are only relevant for achronal sets, because for a set ζ with points that can be joined by timelike curves, the non-achronal subsets of ζ, $\tilde{\zeta}$, will have $D^+(\tilde{\zeta}) = \tilde{\zeta}$, hence we can not predict any extra information.

It is immediate that for ζ a closed achronal set, $\zeta \subseteq D^+(\zeta) \subseteq J^+(\zeta)$, $\partial D^+(\zeta) = H^+(\zeta) \cup \zeta$ and $H(\zeta) = \partial D(\zeta)$.

Proposition 1.42. ([1] p. 741) For ζ achronal, $H^+(\zeta)$ is achronal.

Proof. From the fact that $H^+(\zeta) \subset \overline{D^+(\zeta)}$ it follows that:

$$I^-[H^+(\zeta)] \subset I^-[\overline{D^+(\zeta)}] = I^-[D^+(\zeta)] \subset V_4 - H^+(\zeta),$$

where in the last step it has been used that ζ is achronal and that $H^+(\zeta) \subset \partial D^+(\zeta)$. Thus, $I^-[H^+(\zeta)] \cap H^+(\zeta) = \varnothing$ and $H^+(\zeta)$ is achronal. □

Since we are most interested in achronal sets, it seems logical to want to find its 'boundary' in the sense that we would like to know where this set ends 'spatially speaking', since it is already evident that this will be closely related to $H^+(\zeta)$. To develop this formally, we need to introduce that sense of 'only spatial boundary'. We call to this concept *edge*:

Definition 1.43. ([1] p. 742, [2] p. 202, [4] p. 200) The *edge* of a closed, achornal set ζ is $\text{edge}(\zeta) = \{x \in \overline{\zeta} \mid \forall U_x \text{ neighbourhood of } x,\ \exists\ p \in C_x^-, q \in C_x^+, \gamma \text{ future directed timelike curve from } p \text{ to } q \text{ such that } p, q \in U_x, \gamma \subset U_x \text{ and } \gamma \cap \zeta = \varnothing\}$.

Although the definition is quite technical, the basic idea is that of 'only spatial boundary'. A way to find easily the points that belong to the edge is by means of the formula $\text{edge}(\zeta) = (\bar{\zeta} \setminus \zeta) \cup \{\text{points where } \zeta \text{ is not a continuous 3-manifold}\}$, which can be proved. Intuitively, one might imagine (as in Figure 1.4) that since the points in $H^+(\zeta)$ are the limits of the predictable events, the points in edge(ζ) might have a tight connection to this set. This is true, and it is given by the following proposition:

Proposition 1.44. ([1] p. 742, [2] p. 203, [4] p. 203) Let ζ be a closed achronal set. Every $p \in H^+(\zeta)$ lies on a null geodesic contained entirely in $H^+(\zeta)$ which either is past inextendible or has a past endpoint on edge(ζ).

Proof. Available in the references. □

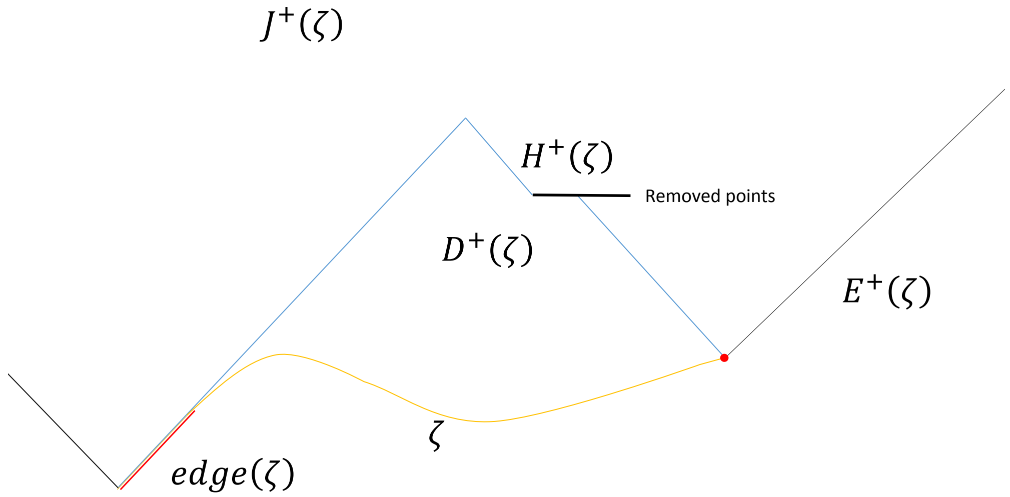

Figure 1.4: Relevant sets in causality in a general scenario. The points in edge(ζ) are in red. Note that points in $H^+(\zeta)$ are geodesics that start in edge(ζ) or have no past endpoint (since the removed points do not belong to the spacetime).

In terms of spacelike sets containing enough information to predict events in other regions, the most important are Cauchy hypersurfaces:

Definition 1.45. ([1] p. 742, [2] p. 205, [4] p. 201) A *partial Cauchy hypersurface* is an edgeless closed acausal set. A *global Cauchy hypersurface* (or simply *Cauchy hypersurface*) is an edgeless closed acausal set Σ such that $D(\Sigma) = V_4$.

It is clear that a partial and global Cauchy hypersurface are actually hypersurfaces, since they have no edge and hence they must be 3-dimensional manifolds. Provided the full Cauchy domain of dependence of a Cauchy hypersurface is all the spacetime, the entire past and future are determined by knowing all the information in it and moreover, since it is achronal, one can think about it as a representation of an 'instant of time' of the universe. However, to predict the future of the whole universe, one would have to know all the data of a spacelike hypersurface, which is impossible to do.

The difference between partial and global Cauchy hypersurfaces is subtle. Note that a Cauchy hypersurface has $H^+(\Sigma) = \varnothing$, while a partial Cauchy hypersurface has non-empty Cauchy horizon. Furthermore, every causal curve intersects a Cauchy hypersurface exactly once, while it intersects a partial Cauchy hypersurface at most once.

To illustrate the partial Cauchy hypersurfaces concept, consider the spacetime $(\mathbb{R}^4, \eta_{\mu\nu})$ with the Minkowski metric and consider a two-sheeted hyperboloid $t^2 - x^2 - y^2 - z^2 = 1$ (depicted in Figure 1.5). This hypersurface has no edge, since it extends to infinity. The light cones from the origin ($t^2 - x^2 - y^2 - z^2 = 0$) totally enclose the hyperboloid and the lower part of the cone is the Cauchy horizon of the lower sheet of the hyperboloid, and analogously with the upper parts. We clearly see that the null geodesics $\gamma(\tau) = (\tau, \pm\tau, 0, 0)$ do not cross the hyperboloid so each sheet is only a partial Cauchy hypersurface.

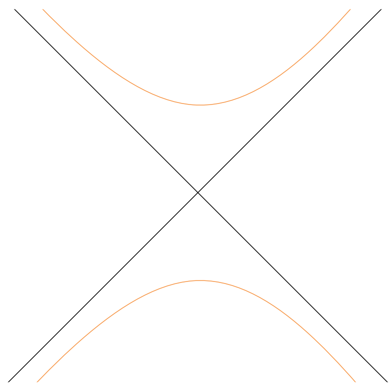

Figure 1.5: Light cone at the origin (black) and a two-sheeted hyperboloid (orange). Each sheet of the hyperboloid is a partial Cauchy hypersurface.

The hypersurfaces generated by the 'time function' in a stably causal spacetime are also in general partial Cauchy hypersurfaces. The fact that every causal curve intersects only once every Cauchy hypersurface is a very powerful indication that we could do a similar construction in this case:

Corollary 1.46. ([1] p. 747) If V_4 contains a Cauchy hypersurface Σ, then $V_4 \overset{f}{\cong} \mathbb{R} \times \Sigma$ and $\forall c \in \mathbb{R}, f^{-1}(\{c\} \times \Sigma)$ is a Cauchy hypersurface.

Proof. Choose a timelike congruence with tangent vector u^μ and let $\Pi : V_4 \to \Sigma$ be the map that takes each point through the curve in the congruence passing through it until it reaches Σ (which happens exactly once). Then the homeomorphism is given by $f : V_4 \to \mathbb{R} \times \Sigma$ with $f(x) = (\tau, \Pi(x))$, where τ is the parameter in the curve from $\Pi(x)$ corresponding to x, and the inverse is $f^{-1}(\tau, x) = y$, where y is the point at parameter τ from the curve emanating from Σ at x. This evidently is a bijection and both f and f^{-1} are continuous. $\square$

We have mentioned before that global hyperbolicity is not just a causality condition. This is because of the following property:

Proposition 1.47. ([1] p. 746, [4] p. 207) V_4 is globally hyperbolic $\iff$ V_4 contains a Cauchy hypersurface.

Hence global hyperbolicity implies all the good properties above, namely, we can naturally define a 'time', we could predict the whole future and past of the universe provided we know the appropriate information. Moreover, by Corollary 1.46, the spacetime is foliated and we have a time function, so the spacetime is also causally stable, and hence, the chronology condition also holds (also, [1] p. 740, [4] p. 205).

1.6 Trapped sets

In highly curved spacetimes, and in particular near singularities, it is likely that it appears a region so influenced by the curvature that even light can not escape from it. Mathematically, the tool we use to study these situations are trapped sets. Historically, the first concept to be developed were trapped surfaces, a very important particular case of trapped sets.

Let S be a spacelike surface so that we can choose two null normal fields to it, $k^{\pm\mu}$. The traces of its null fundamental forms are $K^{\pm} = K^{\pm A}_{\ A} = \gamma^{AB}K^{\pm}_{AB}$. Let us introduce a new parameter $\kappa = K^{+}K^{-}$. The surfaces can be classified as:

Definition 1.48. ([1] p. 775, 776) A spacelike surface S is:

- *trapped* if $\kappa > 0$ everywhere on S.
- *marginally trapped* if $\kappa = 0$ everywhere on S.
- *absolutely non-trapped* if $\kappa < 0$ everywhere on S.
- *untrapped* otherwise.

Moreover, S is future trapped if each of the $K^{\pm} < 0$ and past trapped if $K^{\pm} > 0$.

The interpretation of this definition is simple, one just has to realize that

$$K^{\pm} = \gamma^{AB}K^{\pm}_{AB} = N^{\mu\nu}\nabla_{\nu}k_{\mu} = \vartheta^{\pm},$$

are the expansions of the two null geodesic congruences tangent to $k^{\pm}$. Hence, a surface will be trapped if the two geodesic congruences orthogonal to it both converge or diverge.

However, a general surface will be just a 'patch', and if we consider an object and several surfaces around it, the scalar κ could give different results for each of them. Moreover, it is easy to build trapped surfaces in any spacetime, so to extract the useful information we use instead closed trapped surfaces:

Definition 1.49. ([1] p. 778) A *closed trapped surface* is a compact without boundary trapped surface.

A clear example of a closed trapped surface is a 2-sphere inside of a spherically symmetric black hole. Let us develop this example. A 2-sphere is a spacelike 2-dimensional sphere. The two families of null geodesics here are the outgoing, which travel outwards the sphere and ingoing, which travel inwards (see Figure 1.6). In absence of any gravitational influence, the outgoing congruence forms spheres when considering the surfaces at constant affine parameter that grow bigger as time passes, while the ingoing congruence forms smaller spheres. However, when we consider the same 2-sphere inside a black hole (a strong gravitational field), then both the outgoing and ingoing geodesics will form smaller spheres than the original one, since the geodesics that went outwards originally now bend and are forced to go inwards as well. Hence, every null geodesic starting at the surface will converge to the center and $\kappa > 0$. This is an example of closed trapped surface.

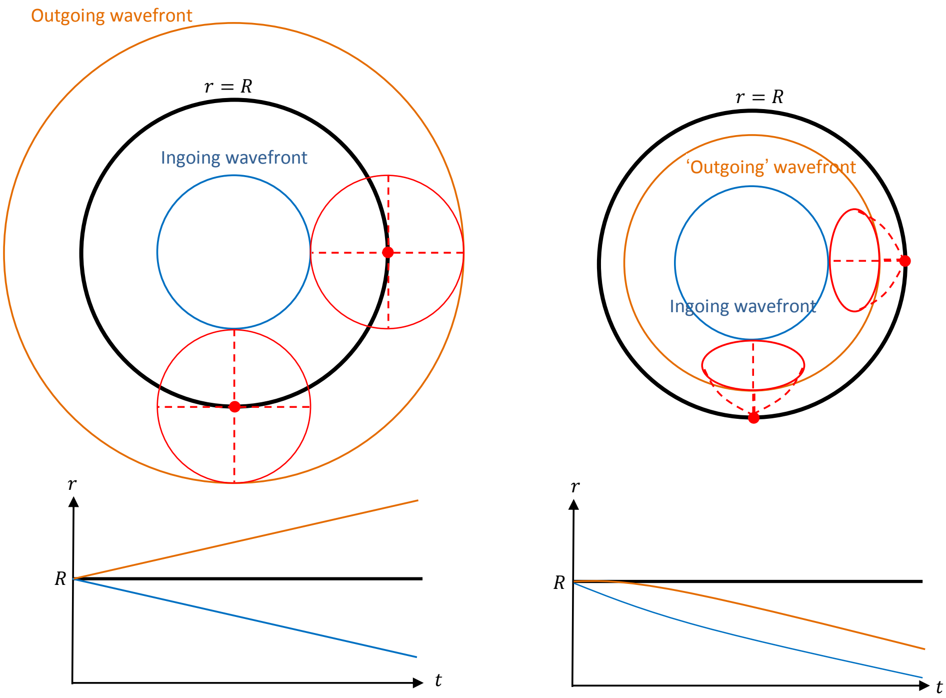

Figure 1.6: 2-spheres of areal radius $r = R$ emitting light rays inside a black hole (right) and outside (left). The signals emitted by each point in the 2-sphere (in red) on the left form two wave fronts, one expanding and the other contracting, while the wave fronts created by the signals emitted by 2-sphere on the right both contract. The 2-sphere on the right is a closed trapped surface.

In general, closed trapped surfaces may have any shape, although they must be homeomorphic to a sphere. An easy way to determine whether a generic surface is trapped or not is as described in [6]. Defining the mean curvature vector of a surface as $H^\mu = -K^- k^{+\mu} - K^+ k^{-\mu}$, one can trivially rewrite $\kappa = -\frac{1}{2} H_\mu H^\mu$. There exists a coordinate system $\{x^\mu\} = \{x^a, x^A\}$ where $a, b = 0, 1$ and $A, B = 2, 3$, in which a surface S can be parametrized as $\Phi(u^2, u^3) = (X^0, X^1, u^2, u^3)$ locally. Then, the null normals are of the form $k^\pm = k^\pm_b dx^b$. Assume the metric is written as:

$$ds^2 = g_{ab} dx^a dx^b + 2 g_{aA} dx^a dx^A + g_{AB} dx^A dx^B.$$

The second fundamental form in these conditions simply becomes:

$$K^\pm_{AB} = \delta^\mu_A \delta^\nu_B \nabla_\mu k^\pm_\nu = \delta^\mu_A \delta^\nu_B \left(\partial_\mu k^\pm_\nu - k^\pm_\lambda \Gamma^\lambda_{\mu\nu} \right) = -k^\pm_b \Gamma^b_{AB}.$$

Then, by defining $e^U = \sqrt{\det g_{AB}}$ and $g_a = g_{aA} dx^A$, it is straightforward to compute:

$$H_\mu = \delta_\mu^a \left(\partial_a U - \text{div} g_a \right),$$

and hence from this one can easily compute κ using $\kappa = -\frac{1}{2} g^{ab} H_a H_b$.

This provides a simple way to determine if a closed surface is trapped or not using the metric and the mean curvature vector. In particular, in spherical symmetry with general coordinates (x^0, x^1, θ, ϕ), the metric can be written as:

$$ds^2 = g_{ab} dx^a dx^b + R^2(x^a) d\Omega^2.$$

The relevant quantities are $U = \ln(R^2 \sin\theta)$ and $g_a = 0$, so

$$H_a = 2 \frac{\partial_a R}{R}.$$

Then, κ is easily calculated as:

$$\kappa = -\frac{2}{R^2}\, g^{ab}\, \partial_a R\, \partial_b R.$$

From the example and the intuition, one can already see that closed trapped surfaces have the property that the null geodesics starting at them will tend to get closer to each others. If we want to generalize the idea to more general sets, we can follow this idea. Like this, we define:

Definition 1.50. ([1] p. 780) A *future trapped set* is a non-empty achronal set ζ such that $E^+(\zeta)$ is compact.

Note that closed trapped surfaces are not necessarily trapped sets because they may not be achronal. Moreover, as seen before, if the spacetime is not null geodesically complete, $E^+(S)$ can be non-compact and hence S would not be a trapped set neither.

Chapter 2

Singularities and singularity theorems

2.1 Singularities

In this section reference [2] (Section 8.1) and [4] (Section 9.1) are used. The concept of singularity is well understood intuitively as a 'region' of the spacetime in which something is wrong or missing. One can rapidly think that points where some component of the metric or some curvature related quantity diverges are such 'bad' points. However, the complexity of some manifolds allows for many other phenomena, so this is not enough to cover all possibilities.

Consider an incomplete causal geodesic, and consider an observer travelling through it. Since the affine parameter is bounded by the incompleteness of the geodesic, the observer travelling along it would simply disappear at finite time. This indicates that something is missing in the spacetime.

If some physical quantity diverges at a point, we could be tempted to simply not consider it. However, this does not solve the problem of incomplete geodesics approaching that point. Furthermore, the divergence problem is also not solved because it becomes arbitrarily big around the catastrophic point. Actually, that point does not even belong to the manifold.

The problem of singularities at this point is severe, because we have to study some regions which 'do not exist' at all, since they are not places. Hence, we first have to find a way to describe such entities. Historically, the process of reaching a suitable description of the singularity theory was sheer trial and error, arising a new counterexample for each new proposal. However, this seems to have reached to an end with the following approach.

Definition 2.1. ([1] p. 759) An *envelopment of the spacetime* (V_4, g) is an embedding $\Phi : V_4 \longrightarrow \hat{V}_4$, with $\hat{V}_4$ a connected manifold. The *boundary of* $\Phi(V_4)$ is $\hat{\partial} V_4 \subset \hat{V}_4$. An *extension of* (V_4, g) is a spacetime $(\hat{V}_4, \hat{g})$ where $\hat{g} = \left(\Phi^{-1}\right)^{\star} g$. The extension is $\mathcal{C}^k$ regular if $\hat{g}$ is $\mathcal{C}^k$ at $\hat{\partial} V_4$. Otherwise, the extension is singular.

In the definition of extension appears the pullback of g by Φ^{-1}, which is calculated as: $\hat{g}(u, v) = \left(\Phi^{-1}\right)^{\star}(g)\ (u, v) = g(\Phi^{-1}(u), \Phi^{-1}(v))$, with $u, v \in \Phi(V_4)$.

With this formalism, we are able to treat those points that were not part of the spacetime initially. Note that we can extend a spacetime in many ways, some of them may be singular, some may be $\mathcal{C}^k$ and some even $\mathcal{C}^\infty$. This is indeed a problem because usually the extension must be fabricated taking into account some physical considerations, and this is not always straightforward. As for curves, note that now for some extensions, incomplete geodesics can become now complete. However, this may not happen for every extension. Hence, it would be excellent to be able to know in which cases any curve can be completed.

It seems clear that geodesic incompleteness indicates singularities, because this means that a free falling particle can simply disappear (or appear) from the spacetime. But not only geodesics may have this problem. Curves in general may also be 'incomplete' and this is again 'wrong' because observers can travel through them as well. In fact, even if the geodesics are complete, this still does not mean that all curves are complete in a spacetime.

We have already studied geodesics and seen that they can be parametrized by an affine parameter. However, general curves do not have any affine parameter associated. To define the sense of completeness of a general curve, we must introduce a new type of parameter:

Definition 2.2. ([1] p. 763, [2] p. 259) Let $\gamma(u)$ be a $\mathcal{C}^1$ curve parametrized by u passing through $\gamma(u_p) = p$, with tangent vector v. The *generalized affine parameter* is $\tau = \int_{u_p}^{u} \sqrt{\delta_{\mu\nu} v^\mu v^\nu}\, du$.

Definition 2.3. ([1] p. 763, [2] p. 259) A $\mathcal{C}^1$ endless curve from $p \in V_4$ is *complete* if the generalized affine parameter is defined $\tau \in [0, \infty)$. A spacetime is *b-complete (bundle complete) at p* if all $\mathcal{C}^1$ curves emanating form p are complete. A spacetime is *b-complete* if it is so $\forall p \in V_4$.

The generalized affine parameter depends on the initial point p and the basis we choose, so one could ask if b-completeness is well defined. However, this is not a problem because even if the value is different, if one generalized affine parameter only reaches finite values, all of them do, and this is the only thing we want to know regarding completeness.

Hence it seems that b-completeness is what we need to get rid of singularities, because:

Proposition 2.4. ([1] p. 763, [2] p. 260) A b-complete spacetime has no regular exetnsion.

Proof. Suppose V_4 has a $\mathcal{C}^1$ regular extension given by the envelopment Φ. Then, there would exist a non-empty $\hat{\partial} V_4$ where the extended metric is $\mathcal{C}^1$. Since Φ is an isometry between V_4 and $\Phi(V_4)$, it leaves the generalized affinely parametrized curves from V_4 invariant in $\Phi(V_4)$. But the curves in $\Phi(V_4)$ approaching $\hat{\partial} V_4$ would have bounded generalized affine parameters, whereas the curves in V_4 are complete and thus their generalized affine parameter is unbounded, hence Φ can not exist and there is no regular extension. □

Since a b-complete spacetime has no regular extension, the only possible extensions are singular, and since it is b-complete, those extensions would only add points at infinity, which are unreachable and hence are not be considered singularities. Instead, we want singularities to be points where there is a curve that ends in an irregular way. Finally, we can define a singularity:

Definition 2.5. ([1] p. 763) A *singularity of* (V_4, g) relative to a singular extension $(\hat{V}_4, \hat{g})$ is the endpoint in $\hat{V}_4$ of a curve incomplete within (V_4, g).

This definition depends on the extension we choose, as expected because singularities do not belong to the spacetime, but to the extended one. This means that a 'point' may be singular or regular depending on the extension. If this happens, we call this singularity *removable*, because we can choose the regular extension and the problem disappears, although this may not always be desirable for physical reasons. If a singularity can not be removed with any extension, then it is called *essential*.

The different types of essential singularities are classified according to how 'bad' the curvature of the spacetime behaves when approaching the singular point [7]. This is measured using the Riemann curvature tensor, $R_{\mu\nu\sigma\rho}$, which fully determines the curvature and has 20 linearly independent components in 4 dimensions. However, its components are not tensors themselves and depend on the particular reference system chosen. The proper way to invariantly fully describe curvature is to construct scalars (hence, invariants) from $R_{\mu\nu\sigma\rho}$ and $g_{\mu\nu}$ by contractions, called scalar curvature invariants. In 4 dimensions, there are up to 14 algebraically independent curvature invariants.

Definition 2.6. ([1] p. 765, [2] p. 260) Essential singularities can be classified as:

- $\mathcal{C}^k$ *quasi-regular singularities* if all the components of the k-th covariant derivative of the Riemman tensor are locally bounded when approaching the singularity along any incomplete curve using a parallelly propagated orthonormal basis.
- $\mathcal{C}^k$ *non-scalar curvature singularities* if they are not $\mathcal{C}^k$ quasi-regular but all the scalar curvature invariants remain well-behaved when approaching the singularity.
- $\mathcal{C}^k$ *scalar curvature singularities* if they are not $\mathcal{C}^k$ quasi-regular and at least a scalar curvature invariant diverges when approaching the singularity.

Non-quasi regular singularities are called *matter singularities* if the problem arises with some component of the Ricci tensor.

It is clear that scalar curvature singularities are the most 'severe', since a curvature invariant, which is a intrinsic property of the spacetime, diverges. In general, the problem of proving the existence of a singularity is simple, given that we have several methods to do so. Namely, we have the singularity theorems.

2.2 Singularity theorems

In the first steps of General Relativity, the fact that many of the spacetimes considered by physicists had some singularity did not concern them in excess. However, later on it was realized that singularities were a big issue and the fact that they appeared so often was not satisfactory. This was one of the motivations for the development of the singularity theorems, to check that singularities were not mathematical artifacts arising from symmetries in the already known solutions. Besides, singularity theorems now serve to detect under which conditions we must expect a singularity, which is also useful when constructing spacetimes with no singularities at all.

We will present here three of the most important singularity theorems. Nevertheless, before that we will need to develop further some concepts introduced in the previous chapter.

The first singularity theorem to be proved was based on Raychaudhuri's equation. This theorem does not actually need most of the concepts discussed before because it was published prior to their creation, and certainly is different in many aspects to the 'modern' theorems which use the new Penrose's ideas. Still, besides its historical importance, this theorem is strong because of the strength of its result.

Theorem 2.7. (Raychaudhuri-Komar): ([1] p. 787) Let the spacetime be filled with a perfect fluid with velocity vector u^μ such that the congruence it generates is geodesic and irrotational. If the strong energy condition SEC holds, and the expansion of the congruence $\theta_0 > 0$ is positive at some instant, then, there is a matter singularity in the finite past along every integral curve of u^μ.

Proof. We are assuming $\omega_{\mu\nu} = 0$, $a^\mu = 0$ and $R_{\mu\nu}u^\mu u^\nu \geq 0$. The Raychaudhuri equation 1.13 in these conditions reads:

$$u^\mu \partial_\mu \theta + \frac{\theta^2}{3} = -\sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}u^\mu u^\nu \leq 0,$$

since the SEC holds and the other term is a square. Since the congruence is irrotational and has no acceleration, by Frobenius Theorem 1.19, $u = -d\tau$ for some function τ. This means that the hypersurfaces $\tau = C$ are orthogonal to the congruence. Let us denote by Σ_0 the hypersurface corresponding with $\tau = \tau_0$. We can build a coordinate system by taking a point $p \in \Sigma_0$ and choosing coordinates $\{x^i\}$ on it. Then, $\{\tau, x^i\}$ is a coordinate system that foliates the spacetime. In this coordinates, the covariant derivative along u, becomes $u^\mu \nabla_\mu f = \frac{df}{d\tau}$. Thus, in this system we can write, using the relation 1.10:

$$\frac{d\theta}{d\tau} + \frac{\theta^2}{3} = \frac{3}{V^{\frac{1}{3}}} \frac{d^2 V^{\frac{1}{3}}}{d\tau^2} \leq 0.$$

Integrating the differential equation, we obtain:

$$V^{\frac{1}{3}} \leq V_0^{\frac{1}{3}} \left[1 + \frac{\theta_0}{3}(\tau - \tau_0) \right],$$

where V_0, θ_0 are these quantities at τ_0. Hence we see that when we approach $\tau \to \tau_0 - \frac{3}{\theta_0} < \tau_0$ from the future, $V^{\frac{1}{3}} \to 0$. This already indicates that something will be wrong at this point. To check whether it is a matter singularity, let us do the explicit calculations. The energy-momentum tensor for a perfect fluid is $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p g^{\mu\nu} = \rho u^\mu u^\nu + p h^{\mu\nu}$. Imposing the conservation equation, $\nabla_\mu T^{\mu\nu} = 0$, we get:

$$u^\mu u^\nu \nabla_\mu \rho + \rho u^\mu \nabla_\mu u^\nu + \rho u^\nu \nabla_\mu u^\nu + \rho u^\nu \nabla_\mu u^\mu + h^{\mu\nu} \nabla_\mu p + p \nabla_\mu h^{\mu\nu} = 0.$$

Using the definition of the expansion $\theta = \nabla_\mu u^\mu$, and separating the terms in parallel and orthogonal to u^μ, which must be linearly independent, we obtain that:

$$u^\mu \nabla_\mu \rho + (\rho + p)\theta = 0, \qquad (\rho + p)\, a^\mu + h^{\mu\nu} \nabla_\mu p = 0. \tag{2.1}$$

In particular, denoting $a = V^{\frac{1}{3}}$, we can rewrite $\theta = \frac{1}{a}\frac{da}{d\tau}$ and using the chain rule on $\frac{d\rho}{d\tau}$, the first equation in our coordinate system takes the form:

$$\frac{d\rho}{da} + 3(\rho + p)\frac{1}{a} = 0 \quad \Longrightarrow \quad a^2 \frac{d\rho}{da} + 3(\rho + p)a = 0 \quad \Longrightarrow \quad \frac{d(\rho a^2)}{da} + a(\rho + 3p) = 0,$$

which leads to:

$$\rho = \frac{1}{a^2}\left(\rho_0 a_0^2 - \int_{a_0}^{a} a(\rho + 3p)da \right) = \frac{1}{a^2}\left(\rho_0 a_0^2 + \int_{a}^{a_0} a(\rho + 3p)da \right) \geq \frac{\rho_0 a_0^2}{a^2},$$

if $a < a_0$, where we have used that a is always a positive quantity, and that the SEC implies $\rho + 3p \geq 0$, hence the integral is positive. Thus, as $a \to 0$, $\rho \to \infty$. Since the momentum-energy tensor diverges, by Einstein's equations, the Ricci tensor diverges too. Thus, the singularity is a matter singularity. □

Although its conclusion is very strong, the theorem asks for the spacetime to be filled with a geodesic, irrotational perfect fluid. In particular, look at the second equation in 2.1. By imposing a geodesic congruence, $a^\mu = 0$, the pressure gradient is forced to vanish. All in all, these restrictions are very particular and of low applicability. Thus, we would like to have with less restrictive hypothesis, which can apply to most of the spacetimes. This is why physicists continued working on the topic.

Before stating the other theorems, it is time to prove some results that will be fundamental for the proofs. Let us start proving some properties about focusing of curves and its relation to curve incompleteness. This will be very important in the theorem's proofs because they are based on ensuring that at least a geodesic is incomplete. To that end, we are interested in conjugate and focal points, the points where curves reconverge and meet again. The important results about the appearance of conjugate and focal points are the ones that follow.

Proposition 2.8. ([1] p. 720, [2] p. 97, [4] p. 226) Let $p \in V_4$ be a point (resp. Σ be a hypersurface) and consider the timelike geodesic congruence emanating from it (resp. orthogonally to Σ). If $\theta_r < 0$ at some point r and $R_{\alpha\beta}v^\alpha v^\beta \geq 0$ SEC holds along the timelike geodesic γ through r, then $\exists q \in \gamma$ conjugate point to p (resp. focal to Σ) at a finite proper time $\tau < \frac{3}{|\theta_r|}$ after r if γ can be extended so far.

Proof. The considered congruence u^μ is irrotational in both cases: where it emanates from p and orthogonally from Σ. In the former, by Lemma 1.20, the congruence is orthogonal to an hypersurface of the form $\{\tau = \tau_0\}$, where τ is the affine parameter of u^μ. Hence, by Theorem 1.19 the congruence is irrotational. In the latter case, we just directly apply Theorem 1.19 and obtain that in both cases, the congruence is irrotational.

Hence, $\omega_{\mu\nu} = 0$ and $a^\mu = 0$ since u^μ is geodesic, so the Raychaudhuri equation 1.13 simply reads:

$$u^\mu \partial_\mu \theta + \frac{\theta^2}{3} + \sigma_{\mu\nu}\sigma^{\mu\nu} + R_{\mu\nu}u^\mu u^\nu = 0.$$

Using the fact that $\sigma_{\mu\nu}\sigma^{\mu\nu}$ is always positive and that the SEC holds we can write it as:

$$\frac{d\theta}{d\tau} + \frac{\theta^2}{3} = -\sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}u^\mu u^\nu \leq 0.$$

We can integrate this differential inequality, starting from the point r with value τ_r of the affine parameter of its geodesic where the expansion $\theta_r < 0$ is negative:

$$\int_{\theta_r}^{\theta} \frac{d\theta}{\theta^2} \leq -\int_{\tau_r}^{\tau} \frac{d\tau}{3}.$$

From this we obtain that:

$$\frac{1}{\theta} \geq \frac{1}{\theta_r} + \frac{\tau - \tau_r}{3}.$$

Since the right hand side is negative in the region $\tau - \tau_r \leq \frac{3}{-\theta_r}$ where we are working, and the left hand side is negative because θ was at $\tau = \tau_r$ and its derivative is negative, the same expression can be written as:

$$\frac{1}{\frac{1}{\theta_r} + \frac{\tau - \tau_r}{3}} \geq \theta.$$

Hence, $\theta \to -\infty$ at some value $\tau \leq \tau_r + \frac{3}{|\theta_r|}$ (because of the inequality), if the geodesic can be extended up to that affine parameter value. □

Proposition 2.9. ([1] p. 721, [2] p. 98, [4] p. 227) Assume the SEC holds along a future-directed timelike geodesic γ and that the timelike generic condition holds at $p = \gamma(\tau_p)$. Then γ has a pair of conjugate point if it can be extended so far.

Proof. Consider the future-directed timelike geodesic congruences emanating from each point $q = \gamma(\tau_q)$ with $\tau_q < \tau_p$. By Lemma 1.20 and Theorem 1.19, these congruences are irrotational. Let us analyze the quantity $\theta_q(\tau_p)$ for each of the congruences. If $\theta_q(\tau_p) \leq 0$, then, since the SEC holds, the congruences are geodesic and irrotational, from Raychauduri's equation 1.13 it follows that:

$$\frac{d\theta}{d\tau} = -\frac{\theta^2}{3} - \sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}u^\mu u^\nu \leq 0, \tag{2.2}$$

so that $\theta_q(\tau) \leq 0$ for $\tau > \tau_p$. Now, if $\theta_q(\tau) < 0$ at some point $\tau > \tau_p$, by Proposition 2.8, there exists a conjugate point r to q with $\tau_r > \tau_p$ and the statement holds. It remains to study the case where θ vanishes everywhere after τ_p. However, this is impossible because if it happened, then $\theta = \frac{d\theta}{d\tau} = 0$, and Raychaudhuri's equation would become:

$$0 = \sigma_{\mu\nu}\sigma^{\mu\nu} + R_{\mu\nu}u^\mu u^\nu, \tag{2.3}$$

and since both terms are non-negative, the only possibility left is $\sigma_{\mu\nu} = R_{\mu\nu}u^\mu u^\nu = 0$. Then, by equation 1.8 it follows that $\nabla_\nu u_\mu = 0$. Following the same steps than in Section 1.3, from equation 1.12, it is easy to see that this implies $R_{\rho\mu\sigma\nu}u^\mu u^\nu = 0$. But by assumption, this can not happen at p.

If $\theta_q(\tau_p) > 0$, the same construction does not work. Instead, consider a point $r = \gamma(\tau_r)$, with τ_r bigger than the value of the affine parameter of all the conjugate points considered in the previous case. Take the past directed timelike congruence emanating from r. If it has a conjugate point between r and p, then the result holds. Otherwise, if the past directed expansion of the congruence at p is positive, $\tilde{\theta}_q(\tau_p) > 0$ (and hence, the future-directed expansion is negative), then again by Proposition 2.8, the congruence would have to have a conjugate point before τ_r by construction, but this can not be since it reaches τ_r. Thus, $\tilde{\theta}_q(\tau_p) < 0$, and using Proposition 2.8 with the past-directed congruence form r, then there is a conjugate point to r at a finite proper time from p.

□

Proposition 2.10. ([1] p. 725, [2] p. 101, [4] p. 231) Let $p \in V_4$ be a point (resp. S be a spacelike surface) and consider the null geodesic congruence emanating from it (resp. orthogonally from S). If $\vartheta_r < 0$ at some point r and the NCC, $R_{\alpha\beta}k^\alpha k^\beta \geq 0$ holds along the null geodesic γ through r, then $\exists q \in \gamma$ conjugate point to p (resp. focal to S) at a finite proper time $\tau < \frac{2}{|\vartheta_r|}$ after r if γ can be extended so far.

Proof. Totally analogous to the proof of Proposition 2.8 using Raychaudhuri's equation 1.15 for null congruences and using the suitable reference system for null congruences. □

Proposition 2.11. ([1] p. 725, [2] p. 101, [4] p. 232) Assume the NCC and the null generic condition hold along a future-directed null geodesic γ. Then there exist $q, r \in \gamma$ conjugate points if γ can be extended so far.

Proof. Totally analogous to the proof of Proposition 2.9, using Raychaudhuri's equation 1.15.

□

It seems that the theory tends to make curves get closer with ease, namely, if SEC (or NCC) is fulfilled, the only thing that it takes for a geodesic to have a pair of conjugate points is to satisfy the (timelike or null) generic condition at one point (Propositions 2.9 and 2.11). Similarly with Propositions 2.8 and 2.10, which also ask for SEC (or NCC) and some extra condition. This, up to certain degree, explains why singularities are so abundant, although the focusing of curves alone does not imply the existence of a singularity. The only problem is that the SEC and NCC, as explained before, do not respond to any physical motivation, instead they are the conditions needed precisely for the convergences of congruences through Raychaudhuri's equation. In this sense, it is worth to discuss whether these conditions make enough sense for these propositions (and by extension, the theorems) be useful. We will do this later.

Furthermore, it is important to note that these propositions grant the existence of conjugate or focal points provided that the geodesics can be extended until them. This means that if the geodesics are incomplete before them for some reason, then the conjugate points will not exist. This will be key to the singularity theorems.

The next topic needing more coverage is that of trapped sets, since they are also an essential part of the singularity theorems.

Proposition 2.12. ([1] p. 780) Let S be a closed future trapped surface and assume the NCC holds. Then, either $E^+(S)$ is compact or the spacetime is null geodesically incomplete to the future, or both.

Proof. If the spacetime is null geodesically incomplete to the future, then the result holds. Assume the spacetime is future null geodesically complete. Since S is a future trapped surface, the expansions $\vartheta^\pm$ of both null geodesic congruences emanating orthogonally to S are negative on S. Since S is compact for it is a closed trapped surface, the maximum and the minimum of $\vartheta^\pm$ is attained on S. Let ϑ_M be the maximum value of both $\vartheta^\pm$ on S.

Since the NCC holds, by Proposition 2.10, since the spacetime is null geodesicall complete, there is a focal point to S at a finite affine parameter $\tau_M \leq -\frac{2}{\vartheta_M}$. Let K be the set containing all the null geodesics in both orthogonal congruences from S up to τ_M included. By construction, it is clear that K is compact and $E^+(S) \subseteq K$, so to see that $E^+(S)$ is compact we only need to see that $E^+(S)$ is closed.

Let $\{p_n\} \subset E^+(S)$ be a sequence of points and p its limit. To prove that it is closed we must see that $p \in E^+(S)$. By construction, $E^+(S) \subseteq K \subset J^+(S)$. K is closed, hence, $p \in K \subset J^+(S)$. Since $J^+(S) = E^+(S) \cup I^+(S)$, it is enough to see that $p \notin I^+(S)$. If $p \in I^+(S)$, since it is open, there would exist a neighbourhood $p \in \mathcal{U}_p \subset I^+(S)$ such that some $p_n \in I^+(S)$, but this is impossible since $p_n \in E^+(S)$. Therefore $p \in E^+(S)$, hence it is closed, and thus compact. □

Proposition 2.13. ([1] p. 781) Let $p \in V_4$ and assume the NCC holds. If the expansion of the future directed null geodesic congruence emanating from p becomes negative along every curve of the congruence, then, either $E^+(p)$ is compact or the spacetime is null geodesically incomplete, or both.

Proof. If the spacetime is null geodesically incomplete, then the result holds. Assume the spacetime is null geodesically complete. Since $\vartheta < 0$ at some point in each null geodesic from p and the NCC holds, by geodesic completeness, using Proposition 2.10 we know there will be a conjugate point to p along each geodesic before a finite value of the affine parameter.

Then, we can define the compact set K in the same way as in Proposition 2.12 and the rest is analogous to its proof. $\square$

Proposition 2.14. ([1] p. 785) Let S be a closed future trapped surface and assume the NCC and the strong causality condition hold. Then, either $E^+(S) \cap S$ is a trapped set or the spacetime is null geodesically incomplete, or both.

Proof. If the spacetime is null geodesically incomplete, then the result holds. Assume the spacetime is null geodesically complete. Since the NCC holds and the spacetime is null geodesically complete, by Proposition 2.12, $E^+(S)$ is compact. S is compact by definition, hence $E^+(S) \cap S$ is also compact and non-empty. We want to prove that $E^+[E^+(S) \cap S]$ is compact. To do so, we will compute the set directly. Let us first find $I^+[E^+(S) \cap S]$ and $J^+[E^+(S) \cap S]$.

We want to find $I^+[E^+(S) \cap S]$. Cover S with convex normal neighbourhoods. Since the strong causality conditions holds, we can choose the neighbourhoods such that their intersection with every causal curve is a connected set, and since S is spacelike, the intersection of S with the neighbourhoods can be made to be achronal. Since S is compact, we can extract a finite subcover $\{\mathcal{U}_i\}_{i=1\dots n}$. Let $q \in I^+(S)$. Let $p_1 \in \mathcal{U}_1$ such that $q \in I^+(p_1)$. Since $S \subset J^+(S)$, $S \cap J^+(S) = S$. In the case that $p_1 \in S \cap E^+(S)$, it follows that $q \in I^+[S \cap E^+(S)]$. Otherwise, since $p_1 \in S$ but $p_1 \notin E^+(S)$, then $p_1 \notin I^+(S)$ and there must exist $p_2 \in \mathcal{U}_2$ such that $p_1 \in I^+(p_2)$. Since we have taken $S \cap \mathcal{U}_1$ achronal, $p_2 \notin \mathcal{U}_1$, hence $p_2 \in \mathcal{U}_2 \setminus \mathcal{U}_1$. In the case that $p_2 \in S \cap E^+(S)$, it follows from $p_1 \in I^+(p_2)$ and $q \in I^+(p_1)$ that $q \in I^+[S \cap E^+(S)]$. Otherwise, we can repeat the process and find further points $p_k \in \mathcal{U}_k \setminus \left(\bigcup_{i=1}^{k} \mathcal{U}_i\right)$ until one of them satisfies $p_k \in S \cap E^+(S)$. Since the subcover is finite, and $S \cap E^+(S)$ is not empty, this must finish so there must be some $p_j \in S \cap E^+(S)$. Hence, $q \in I^+[S \cap E^+(S)]$ which implies $I^+(S) \subset I^+[S \cap E^+(S)]$. Since $S \cap E^+(S) \subset S$, then $I^+[S \cap E^+(S)] \subset I^+(S)$. Therefore, $I^+[E^+(S) \cap S] = I^+(S)$.

We now want to find $J^+[E^+(S) \cap S]$. Let $r \in J^+(S) = I^+(S) \cup E^+(S)$. Let us examine both cases. If $r \in I^+(S)$, then

$$r \in I^+(S) = I^+[E^+(S) \cap S] \subset J^+[E^+(S) \cap S].$$

If $r \in E^+(S)$, then there is a point $p \in S$ such that $r \in E^+(p)$, but $p \notin I^+(S)$ because otherwise $r \in I^+(S)$. Thus, $p \in S \setminus I^+(S)$ so $p \in E^+(S) \cap S$ and

$$r \in E^+[E^+(S) \cap S] \subset J^+[E^+(S) \cap S].$$

In both cases we obtained the same result, hence, if $r \in J^+(S)$, then $r \in J^+[E^+(S) \cap S]$ and $J^+(S) \subset J^+[E^+(S) \cap S]$. Since $E^+(S) \cap S \subset S$, trivially $J^+[E^+(S) \cap S] \subset J^+(S)$. Therefore, $J^+[E^+(S) \cap S] = J^+(S)$.

Therefore,

$$E^+[E^+(S) \cap S] = J^+[E^+(S) \cap S] \setminus I^+[E^+(S) \cap S] = J^+(S) \setminus I^+(S) = E^+(S),$$

which is compact by Proposition 2.12 and null geodesic completeness, so $E^+(S) \cap S$ is a trapped set. $\square$

Note that by Proposition 2.12 if NCC holds and the spacetime is null geodesically complete, then a closed trapped surface is a trapped set if it is achronal. But even if it is not, by

Proposition 2.14, there is always a subset of it that is a trapped set (provided the spacetime is null geodesically complete). Points are also trapped sets if the NCC holds and the null geodesic congruence emanating from them reconverge if, again, the spacetime is null geodesically complete. Note that if the spacetime is null geodesically incomplete, the main results in the propositions may still hold, because for that we only need that the null geodesics emanating from the surface or the point are complete, whereas the spacetime is incomplete as long as any geodesic is incomplete. This is why the statements include the option of the spacetime being null geodesically incomplete and the main result to hold.

So it seems that if the spacetime is null geodesically complete, the existence of trapped sets is guaranteed as long as few restrictive conditions are satisfied. On the other hand, if the spacetime is null geodesically incomplete, then it is likely to have a singularity. Since trapped sets have somehow a 'compact future', this also seems to indicate that something is wrong and possibly, a singularity. So again, it seems that the theory favours the appearance of singularities. As we will see next, trapped sets have a main role in the singularity theorems.

We also need some extra results about causality conditions. Note that all of them are related and are not independent. Here we state some properties:

Proposition 2.15. ([1] p. 780) Assume the chronology condition holds. If every null endless geodesic has a pair of conjugate points, then strong causality condition holds.

Proof. Suppose the strong causality condition does not hold at p. Take the maximal normal neighbourhood $\mathcal{N}_p$ and choose a sequence of nested neighbourhoods of p, $\{\mathcal{U}_n\}$ converging to p. Then, for each $\mathcal{U}_n$ there exists a causal curve γ_n starting in it, leaving and then returning to it. It can be seen that the limit curve γ of γ_n is a causal curve which starts are p and passes through p again. If γ was timelike, this would violate the chronology condition because $p \in I^+(p)$, hence it is impossible. Then γ must be a null geodesic. But by assumption, it has a pair of conjugate points and by Proposition 1.31, there exist two points of γ which can be joined through a timelike curve λ. But then, with pieces of the appropriate λ and γ_n, we can build a closed timelike curve, which again violates the chronology condition. Hence, this is not possible and the strong causality condition must hold. □

Corollary 2.16. ([1] p. 737) Assume the NCC, the chronology and the generic condition hold. Then, the spacetime either satisfies the strong causality condition or it is null geodesically incomplete.

Proof. Since the NCC and the null generic condition hold, by Proposition 2.11, either every null geodesic has a pair of conjugate points or the spacetime is null geodesically incomplete. But in the former, we can use Proposition 2.15 since all its assumptions are fulfilled, and then, the strong causality condition holds. Therefore, either the strong causality condition holds or the spacetime is null geodesically incomplete. □

Proposition 2.17. ([1], [2]) Let ζ be a closed achronal set. Every $p \in E^+(\zeta)$ either lies on a null geodesic with a past endpoint on edge(ζ) or $p \in \zeta$.

Proof. Let $p \in E^+(\zeta) = J^+(\zeta) \setminus I^+(\zeta) \supseteq \zeta$. If $p \in \zeta$, the result holds. Otherwise, as there exists a causal, but non-timelike curve γ starting at some $q \in \zeta$ to p, by Proposition 1.31, then γ is a null geodesic segment without conjugate points between q and p. It is only left to see that $q \in$ edge(ζ). Suppose $q \notin$ edge(ζ). Then, by definition of edge, there exist $r \in C_q^+$, $s \in C_q^-$ such that there is not any timelike curve λ joining s and r such that $\lambda \cap \zeta = \varnothing$.

Take a timelike curve $\tilde{\gamma}$ from s to r passing through q. Then, $q \in I^+(s)$ and since $p \in E^+(q)$, then $p \in J^+(s)$. But by Proposition 1.29, $p \notin E^+(s)$. Hence, $p \in I^+(s)$. However, since $q \notin \text{edge}(\zeta)$, we can find a timelike curve from s to p such that it intersects ζ. But this means that $p \in I^+(\zeta)$, which is impossible by assumption. Hence, $q \in \text{edge}(\zeta)$. □

Now we are in a condition to prove the modern and most important singularity theorems. After all the previous work, their proofs will be relatively simple. Their importance is due to their generality, since as we will see now, their hypothesis just ask for a few objects to exist or conditions to hold.

Theorem 2.18. (Penrose): ([1] p. 789, [2] p. 263, [4] p. 239) Assume the null convergence condition holds. If there exists a non-compact Cauchy hypersurface Σ and a closed trapped surface S, then the spacetime is null geodesically incomplete.

Proof. Suppose the spacetime is null geodesically complete. Since S is a closed trapped surface (assume it is future trapped, past trapped case leads to an analogous result), the NCC holds and the spacetime is null geodesically complete, by Proposition 2.12, $E^+(S)$ is compact. Note also that $E^+(S) = \partial J^+(S)$ is achronal because $J^+(S)$ is a future set

Let u^μ be the tangent vector to a timelike congruence and let Σ be a non-compact Cauchy hypersurface. Every curve of the congruence will intersect Σ exactly once, and due to the achronality of $E^+(S)$, every curve of the congruence will intersect it at most once. In these conditions, we can define the map $f : E^+(S) \to \Sigma$ which transports the points in $E^+(S)$ to Σ through the corresponding curves of the congruence.

Denoting $T = f\left(E^+(S)\right)$ the image of $E^+(S)$ by this map, $f_H : E^+(S) \to T$ is a homeomorphism, since it is clearly a bijection and it continuous with continuous inverse by continuity of the congruence u^μ [4]. As $E^+(S)$ is compact, so is T by the homeomorphism and hence, T is closed as a subset of Σ. $E^+(S)$ is a 3-dimensional submanifold of V_4 because it is an achronal proper boundary, so each point $p \in E^+(S)$ has a 3-dimensional neighbourhood homeomorphic to an open ball in $\mathbb{R}^3$ totally contained in $E^+(S)$, and since f_H is an homeomorphism, this also happens in T, which means T is an open subset of Σ. But as Σ is connected and $T = f\left(E^+(S)\right)$ is non-empty because the spacetime is null geodesically complete, and T is open and closed, then it must happen that $T = \Sigma$. However, T is compact while Σ is non-compact by assumption, so this is not possible by the homeomorphism and $E^+(S)$ can not be compact. Hence, the assumption we made at the beginning is false hence the spacetime must be null geodesically incomplete. □

After this theorem was published, a lot of research on the topic was performed. The culmination of all the subsequent works was the following theorem, which collects most of the results obtained until then:

Theorem 2.19. (Hawking-Penrose): ([1] p. 792, [2] p. 266, [4] p. 240) Assume the chronology, the generic and the strong energy conditions hold. If there exists at least one of the following:

(i) a compact achronal set Σ without edge,

(ii) a closed trapped surface S,

(iii) a point p such that the null geodesic families emanating from p reconverge, then, the spacetime is causal geodesically incomplete.

Proof. To prove the theorem, we need to use a lemma:

Lemma 2.20. (Hawking-Penrose): ([1] p. 791, [2] p. 267) The three following coditions can not hold simultaneously in a spacetime:

1. every endless causal geodesic has a pair of conjugate points,

2. the chronology condition is satisfied,

3. there is a trapped set.

Proof. The proof of this lemma is very technical and rather lengthy so we will not reproduce it here, but it can be found in the references. □

Now let us prove the theorem. Suppose the spacetime is causal geodesically complete. Since the SEC (and by continuity also NCC) and the generic condition hold and we have supposed causal geodesic completeness, by Propositions 2.9 and 2.11 every causal geodesic has a pair of conjugate points. Also, by assumption, the chronology condition holds. Then, by Lemma 2.20, there can not be a trapped set in the spacetime.

For (i), we will prove that any compact achronal set Σ without edge is a trapped set. Indeed, since Σ has no edge, by Proposition 2.17 there are no possible points of $E^{\pm}(\Sigma)$ other than those in Σ itself so we simply have $E^+(\Sigma) = E^-(\Sigma) = \Sigma$, which is compact by assumption, and hence Σ is a trapped set.

For (ii), since we are assuming that the spacetime is null geodesically complete, the NCC the chronology and the generic condition hold, by Corollary 2.16 the strong causality condition holds. Hence we can use Proposition 2.14 and we conclude that either $E^+(S) \cap S$ is a trapped set because the spacetime is null geodesically complete.

For (iii), we have to note that if the null congruences emanating from p reconverge, then $\vartheta < 0$ at some point in every curve, so by Proposition 2.13, and since the spacetime is null geodesically complete, $E^+(p)$ is compact and hence p is a trapped set.

If any of the three conditions is satisfied, then there exists a trapped set. But we have already seen that by Lemma 2.20, such set can not exist. Hence, the assumption we made is wrong and the spacetime is causal geodesically incomplete. □

The power of the theorems is so that after they were proved, physicists asked themselves whether reasonable astrophysical and cosmological models with no singularities did actually exist. Although they do, it is hard to build them by avoiding some of the hypothesis in the theorems. The next chapter is dedicated to the study of the assumptions of the theorems along with their conclusions, so we will discuss this in it.

Chapter 3

Assumptions and consequences of the singularity theorems

In this chapter, Chapter 6 of [1] is used. In the previous section, the most relevant singularity theorems have been shown. However, there are many other singularity theorems, each one with its own particularity, although all of them share the same structure: if certain type of conditions (that depend on each theorem, of course) are satisfied, then the spacetime is causal geodesically incomplete. One can collect this fact in the following outline:

Singularity theorems pattern. If the spacetime satisfies:

(i) an energy condition,

(ii) a causality condition, and

(iii) a boundary or initial condition

then it contains at least an incomplete causal geodesic.

It seems clear that, in order to make any sense, the conditions required by the theorems should be feasible and plausible. This is because the theorems were initially devised to explain the abundance of singularities in the solutions of Einstein's equations that were interesting in some way. In order to have stronger and more accurate theorems, we would like to have the most general possible conditions.

On the other hand, from the point of view of building regular spacetimes, if the hypothesis of the theorems are too general, it will be hard to find regular spacetimes with the desirable properties. In this sense, we would want the hypothesis to be weak to easily find suitable cosmological and astrophysical models.

All in all, the theorems are relatively strong, and at first it was not evident that regular cosmological models beyond the FLRW family or regular astrophysical models such as black holes could exist. Now we have some examples of them, built by avoiding the conditions in the singularity theorems.

Let us analyze each of the types of hypothesis in the theorems separately.

Energy and generic conditions

The motivation to include the energy and generic conditions is to focus the congruences of curves. Although generic conditions do not appear in the pattern above, they favour the existence of conjugate and focal points, which are the important element for the theorems.

As it is obvious from Raychaudhuri's equations 1.13 and 1.15, if the SEC and the NCC are fulfilled, respectively, the convergence of timelike and null curves is easier because they contribute to the negativity of $\frac{d\theta}{d\tau}$ and $\frac{d\vartheta}{d\tau}$, respectively. However, note that the rotation tensor and the acceleration terms can still act in the opposite direction. So, even if the energy conditions are satisfied, this alone does not imply the focusing of congruences.

The standard ways to ascertain focusing of curves are Propositions 2.8, 2.9, 2.10 and 2.11. These propositions rely on Raychaudhuri's equation under the appropriate situations, and Propositions 2.9 and 2.11 also use the timelike and null generic condition to reinforce the focusing effect.

Generic conditions are present in the hypothesis to exclude some type of spacetimes with special symmetries from the proofs of the theorems. As it is evident from the example in Section 4.1.1, generic conditions are indispensable for the theorems. The only problem is that in those few spacetimes excluded by the generic conditions, we can not apply the singularity theorems. One could try to develop simpler singularity theorems without generic conditions for the special cases not covered by the general ones. Unfortunately, simpler versions will not work, as proved by the many counter examples available.

With regard to the energy conditions demanded by the theorems listed in this work, Theorem 2.7 asks the SEC, Theorem 2.18 asks NCC and Theorem 2.19 ask the SEC and the generic condition. Nevertheless, there exist more refined versions which only require averaged conditions [8], [9], hence even if in some point the conditions does not hold, the theorem may still work if the spacetime is adequate.

However, the main problem with the required energy conditions in the theorems is their physical motivation. Indeed, both SEC and NCC do not have a clear physical interpretation, as they were only inspired by the term $R_{\mu\nu}u^{\mu}u^{\nu}$ in Raychaudhuri's equation 1.13. And actually, there are important physical examples that violate the SEC, such as the Higgs scalar field. Moreover, it is also remarkable that if one considers quantum effects, there appear several examples of phemonema violating even WEC. Some examples of this are Casimir effect, Hawking radiation or cosmological inflation. Thus it seems that quantum effects help avoiding the fulfillment of energy conditions and this could be a way to circumvent the singularity theorems.

Causality conditions

Causality conditions want to guarantee that causal paradoxes do not happen. For example, if a particle affects its own past, causality would break down. However, some of the strongest causality conditions also ensure the existence of maximal geodesics.

With regard to the causality conditions in the theorems listed before, Theorem 2.19 assumes the strong causality condition, Theorem 2.18 assumes global hyperbolicity through the existence of a Cauchy hypersurface and Theorem 2.7 does not explicitly assume any causality condition (although in the proof a partial Cauchy hypersurface is constructed from the congruence and it is performed within its Cauchy development, where global hyperbolicity holds). These conditions force timelike curves not to be closed and keep $E^{\pm}(\zeta)$ non-empty.

These kind of conditions are hardly questionable physically, as from our observations, it seems logic to impose an inviolable 'arrow of time'. However, some work can be done even without them. For example, from Theorem 5.7 of [1], the chronology conditions does not need to hold at all the spacetime. It is enough that it is fulfilled in a region causally disconnected from the 'violating' region. Moreover, in certain situations (like in [2], Chapter 8, Section 2, Theorem 4), even relaxing causality conditions does not help in avoiding the singularities.

Differentiability

A subtle but essential condition for the theorems is the metric g to be $\mathcal{C}^2$. This assumption is used in many points of the development of not only the theorems themselves, but also the mathematical machinery behind them. Namely, the normal coordinates of Definition 1.13 work well with $\mathcal{C}^2$ differentiability. Although they can also be defined with $\mathcal{C}^{2-}$, in this case, the change to normal coordinates is only continuous and not differentiable, hence differentiability of quantities in normal coordinates or dependence of the geodesics on the initial conditions are affected. Similarly, results on maximal curves, trapped sets, existence of conjugate points, etc. rely on the $\mathcal{C}^2$ differentiability.

This may seem just a technical condition, but in fact, there are some important cases of spacetimes whose metrics are not $\mathcal{C}^2$, for example, the matching of a interior with the exterior of a star, which is just $\mathcal{C}^{2-}$. However, proving the theorems with lower differentiability degrees is a really hard task, although it is widely accepted that the results of the singularity theorems will also hold with $\mathcal{C}^{2-}$ metrics. Hence, relaxing differentiability of the spacetimes to $\mathcal{C}^{2-}$ would probably not help at avoiding singularities. Relaxing even more the differentiability degree to $\mathcal{C}^1$ makes the singularity theorems even more inapplicable. Hence, in these conditions we can find any type of scenarios with or without singularities.

Boundary or initial conditions

These conditions are the analogous of boundary or initial conditions in other classical problems, in the sense that they determine the events in appropriate regions. In General Relativity, as we have seen, this may not be as simple as assigning a value to a function. Instead, surfaces, (Cauchy) hypersurfaces and trapped sets play a fundamental role in the determination of events in spacetimes.

In the theorems stated in Chapter 2, the boundary or initial conditions required by the theorems are: in Theorem 2.7, the expansion of the congruence must be positive at some hypersurface orthogonal to the congruence, in Theorem 2.18, the existence of a closed trapped surface and the non-compactness of a Cauchy hypersurface and in Theorem 2.19, a triple condition in which each of the cases leads to a trapped set. Note that in this theorem, condition (i) can only happen in spatially closed universes since an edgeless set has to expand to the limits of the spacetime, and it must be compact, hence finite.

As we have reviewed before, there are some ways (or special cases) in which energy and causality conditions and even differentiability degree can be relaxed while keeping the result of the singularity theorems. However, this can not be done with the boundary or initial conditions. They are *essential* hypothesis in the theorems and can not be removed or relaxed. Boundary or initial conditions are important because they usually assume or lead to the existence of a trapped set (or other conditions with an equivalent effect). As we have seen, this desirable effect is for null geodesics to converge and get confined. At first sight, one could think that this already is equivalent to a singularity. However, any of the energy, causal or boundary conditions alone is obviously not enough in order to ensure a singularity (this is precisely why we have the singularity theorems). Hence, although boundary conditions are the most relevant, they must be put together with other conditions properly in order to get the expected result.

Since they are the central assumptions, and since energy and causality conditions have been widely tested and studied empirically, the reasonability of boundary or initial conditions should be checked.

Conclusions of the theorems

It must be noted that the theorems just predict the incompleteness of at least one geodesic. This result is actually weak compared to what the theorem's names suggest, as it does not give much information about the characteristics of an eventual singularity (with the exception of Theorem 2.7, a very particular case in which we know all the information). Namely, we do not know if the singularity is removable or essential, as well as its location or severity.

With the information of the existence of an incomplete geodesic, what we can do is to look for extensions. However, this presents a problem, since a priori we do not have any physical motivation to choose a particular extension over the infinite number of alternatives we have. If we can choose a regular extension, it seems logical to do so. The problem is that we do not know if the extended spacetime will also satisfy the assumptions of the theorems, so maybe it has a singular extension at the end. It is worth to remark that until we choose a particular extension, we do not have any information about the kind of singularity we are dealing with, in case it exists. In the other hand, if there is a singularity, the theorems do not give any information about where it is or its severity, so again we have to select a particular extension to be able to say anything, and we will have to determine its properties on our own.

If we compare Theorem 2.7, with Theorems 2.18 and 2.19, one could say that, in the former, both the assumptions and the conclusions are very strong whereas in the other two, the assumptions are reasonably general and the conclusions rather vague. Although this is true, it does not mean that the usefulness of the theorems is small. Singularity theorems are still a big achievement because of their generality and historical, physical and mathematical importance.

Chapter 4

Examples

4.1 FLRW

For this section, references [10] (Section 2.4), [11] (Section 8.2), [2] (Section 5.3), [1] p. 751 are used. The *Friedman-Lemaître-Robertson-Walker* spacetimes are the homogeneous and isotropic solutions of Einstein's equations 1.3. These spacetimes are classified according to the sign of their spatial curvature, k, which can take the values $k = 0, 1, -1$, and are called flat, closed and open respectively. Their metric is given by:

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right], \tag{4.1}$$

where $r \in [0, \infty)$ for $k = -1, 0$ and $r \in [0, 1)$ for $k = 1$, $\theta \in [0, \pi)$, $\phi \in [0, 2\pi)$ as usual, $t \in I$ open interval and $a(t)$ is a positive function called *scale factor*. Note that the closed model is spatially closed, while the open and flat models are spatially infinite.

The flat FLRW spacetimes are nowadays accepted as *the* cosmological models for our universe. Hence, it is interesting to study their singularities. Let us study their properties for all values of k.

As it is evident from the metric 4.1, the energy-momentum tensor of the spacetime will diagonalize. In the preferred orthonormal reference frame of an observer with velocity ∂_t, $e_A = \{\partial_t, \frac{\sqrt{1-kr^2}}{a(t)}\partial_r, \frac{1}{r \; a(t)}\partial_\theta, \frac{1}{r \; \sin\theta \; a(t)}\partial_\phi\}$ $(A = 0, 1, 2, 3)$, it is easy to compute using $T_{\mu\nu} e^\mu_A e^\nu_A$, that the energy-momentum tensor takes the form of a perfect fluid, with

$$\begin{aligned} \rho &= 3\frac{a'(t)^2 + k}{a(t)^2}, \\ p &= -\frac{a'(t)^2 + 2a(t)a''(t) + k}{a(t)^2}. \end{aligned} \tag{4.2}$$

It seems clear that if $a(t_0) = 0$ for some t_0, then the metric becomes degenerate. Let us study what happens in such a case. Assume a Taylor expansion of $a(t)$ around t_0 such as $a(t_0 + t) = a_1 t + a_2 t^2 + O(t^3)$. Then, it follows that

$$\lim_{t\to 0} \rho = \lim_{t\to 0} 3\frac{(a_1 + 2a_2 t)^2 + k}{(a_1 t + a_2 t^2)^2} = 3 \lim_{t\to 0} \frac{a_1^2 + 4a_1 a_2 t + 4a_2 t^2 + k}{a_1^2 t^2 + 2a_1 a_2 t^3 + a_2 t^4}.$$

This limit diverges unless $k = -1$ and $a_1 = 1, a_2 = 0$. In any other case, then we energy-momentum tensor diverges and we have a matter singularity when approaching $t = t_0$. If the singularity is at our past, we call it *Big Bang* and if it is at out future, we call it *Big Crunch*.

The expansion of the fluid is $\theta = 3\frac{a'(t)}{a(t)}$ and its acceleration, shear and vorticity vanish: $a_\mu = 0$, $\omega_{\mu\nu} = \sigma_{\mu\nu} = 0$. Hence, the fluid follows geodesic and irrotational trajectories. These are the hypothesis of Theorem 2.7. However, for it to be applicable, $\theta > 0$ at some point and SEC must be fulfilled. It is clear that $\theta > 0 \iff a'(t) > 0$ at some value of t. Let us check the energy conditions.

Obviously, in a perfect fluid, the WEC is fulfilled if $\rho \geq 0$ and $\rho + p \geq 0$, the DEC is fulfilled if $\rho \geq 0$, $\rho + p \geq 0$ and $\rho - p \geq 0$ and the SEC is fulfilled if $\rho + p \geq 0$ and $\rho + 3p \geq 0$. Hence, a straightforward calculations leads to the conclusion that WEC holds if:

$$a'(t)^2 \geq a(t)a''(t) - k \quad \text{and, for } k = -1, \text{ also } a'(t) \geq 1,$$

the DEC holds if the WEC holds and additionally

$$\left(a(t)^2\right)'' + 2a'(t)^2 + 4k \geq 0,$$

and the SEC holds if

$$a'(t)^2 \geq a(t)a''(t) - k \text{ and } a''(t) \leq 0.$$

All in all, for Raychaudhuri-Komar Theorem 2.7 to apply, we need $a'(t_0) > 0$ for some t_0, $a'(t)^2 \geq a(t)a''(t) - k$ and $a''(t) \leq 0$. If these conditions hold, then as seen from the proof of Theorem 2.7, $\rho \to \infty$, which by expression 4.2 implies $a(t) \to 0$, and we have a Big Bang (or Big Crunch) in the model.

As for the other theorems in Chapter 2, it is better to study them in particular cases rather than in the general case to extract more meaningful information from them.

4.1.1 Einstein cosmology

The lack of observational evidence for the expansion of the universe at the beginnings of the 20th century motivated Einstein to find a completely static exact solution of his new theory in 1917. His model is a particular case of FLRW with $k = 1$, $a(t) = a_E$ constant and metric

$$ds^2 = -dt^2 + a_E^2 \left[\frac{dr^2}{1 - r^2} + r^2 \left(d\theta^2 + \sin^2\theta \; d\phi^2\right)\right].$$

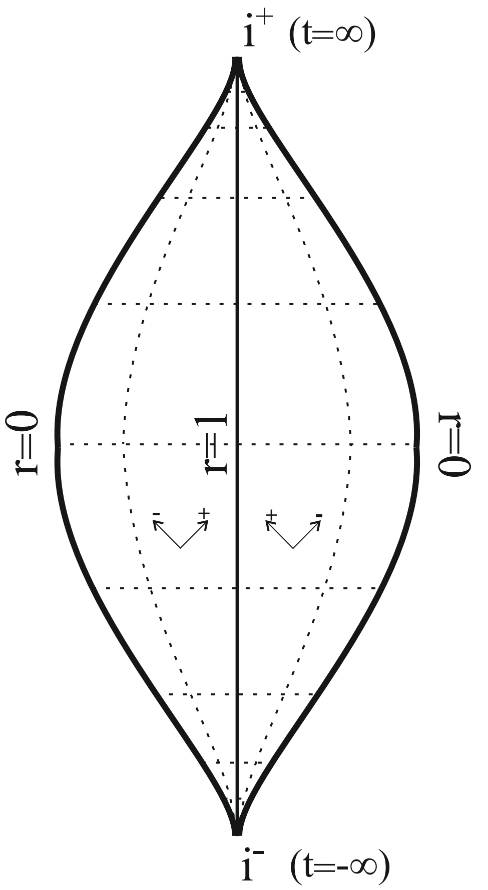

Figure 4.1: Penrose diagram for the static Einstein universe [10]. The horizontal dashed lines are the curves $t = constant$, while the curved dashed line going from top to bottom are the curves $r = constant$.

It is clear that this spacetime is static since its metric does not depend on the timelike coordinate t. Since $a'(t) = a''(t) = 0$, it is also clear that

$$\rho = \frac{3}{a_E^2},$$
$$p = \frac{-1}{a_E^2},$$

and that the WEC, DEC and SEC are satisfied everywhere and $\theta = 0$.

This spacetime is regular and inextendible as seen from the Penrose diagram in Figure 4.1.

Hence, although the congruence is geodesic and irrotational and SEC is fulfilled, Theorem 2.7 does not apply because the model is static ($\theta = 0$).

As for Penrose's Theorem 2.18, let us try to search for closed trapped surfaces. Since they must be compact, they must have a maximum value of r at some point p. At that point, the normal vector to the surface must be a linear combination of ∂_t and ∂_r. Using this, we obtain $\kappa = 2\frac{r^2-1}{a_E^2 r^2}$ at p. Since $0 < r < 1$, this quantity is never positive, so the closed surfaces are never trapped. Hence, we can not use Penrose's Theorem 2.18.

As for Hawking-Penrose Theorem 2.19, consider the hypersurfaces $\{t = t_0\}$. This set can not have any edge points because any timelike curve joining a point in its future with one in its past necessarily intersects it. Hence it is edgeless.

Posing the geodesic equations for this spacetime for constant θ, ϕ:

$$r'' + \frac{r}{1-r^2}(r')^2 = 0,$$
$$t'' = 0,$$

and the null condition

$$(t')^2 + \frac{a_E^2}{1-r^2}(r')^2 = 0,$$

it is easy to see that the curves

$$\gamma(\tau) = \left(\tau + t_0, \pm \sin\left(\frac{\tau}{a} + \arcsin r_0\right), \theta_0, \phi_0\right), \tag{4.3}$$

are radial null geodesics. This suggests that every null geodesic must cross the hypersurface $\{t = t_0\}$ as it can be checked from the Penrose diagram in Figure 4.1 and hence, it is a Cauchy hypersurface.

Also, since $r \in [0, 1)$ is finite, $\{t = t_0\}$ are compact, so it is an achronal edgeless compact hypersurface. Additionally, from equation 4.3, it is evident that there are reconverging null geodesics. However, although the SEC, the chronology condition and the initial conditions hold, Hawking-Penrose Theorem 2.19 can not be applied because the timelike generic condition is violated. For instance, consider the geodesic with tangent vector $u = \partial_t$. Since the Riemann tensor for this spacetime has no term proportional to $dx^\rho \otimes dt \otimes dt \otimes dx^\sigma$, then $R_{\rho\mu\nu\sigma}u^\mu u^\nu = 0$ and the timelike generic condition is violated.

Hence, the fact that this spacetime is static, spatially closed and possesses special symmetries forbids the existence of a singularity.

4.1.2 Linear state equation models

In this subsection reference [11] (Section 8.3) is used. Consider a perfect fluid in a flat FLRW spacetime satisfying the state equation

$$p = \omega\rho, \tag{4.4}$$

with ω constant. This model is of interest because there are several particular cases of this form with specific cosmological physical meaning. One of them is that with $\omega = 0$ and state equation $p = 0$, which corresponds to an universe filled with matter (dust) that exerts negligible pressure. Another one is that with $\omega = \frac{1}{3}$ and state equation $p = \frac{1}{3}\rho$, which represents radiation particles, which exert a radiation pressure as collected by the state equation. The most useful model is:

Empty (De Sitter cosmology)

This is the special case with $\omega = -1$ and hence $p = -\rho$. In this case, the ideal fluid energy-momentum tensor 1.5 can simply be written as $T^{\mu\nu} = pg^{\mu\nu}$, so we can rewrite Einstein's equations in the form:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \rho g_{\mu\nu} = 0.$$

Hence we can interpret this universe as being empty but having a cosmological constant $\Lambda = \rho$, which accounts for dark energy. This model is relevant because it approximates the early and late stages of our universe, in which the influence of dark energy is much more dominant than matter.

The conservation of energy equation gives us

$$0 = \nabla_\mu T^\mu_t = -\partial_t \rho - 3\frac{\partial_t a}{a}(\rho + p).$$

In adapted coordinates, where ∂_t is the velocity of the fluid we simply denote the derivative of f with respect to t as $\dot{f}$, and using equation 4.4 we obtain:

$$\frac{\dot{\rho}}{\rho} = -3(1+\omega)\frac{\dot{a}}{a}.$$

Using this along with the first Friedman equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\rho}{3},$$

arising from the tt component of Einstein's equations, we can trivially solve for $\rho(t)$ and $a(t)$. For $\omega \neq -1$, integrating we obtain

$$\begin{aligned} \rho(t) &= \frac{4}{\left(2(\sqrt{\rho_0})^{-1} + \sqrt{3}(1+\omega)(t-t_0)\right)^2}, \\ a(t) &= a_0 t^{\frac{2}{3(1+\omega)}}, \end{aligned}$$

and for $\omega = -1$ the equation for ρ reads $\dot{\rho} = 0$, hence the solution is:

$$\begin{aligned} \rho(t) &= \rho_0, \\ a(t) &= e^{\sqrt{\frac{\rho_0}{3}}t}. \end{aligned}$$

For $\omega \neq -1$ we note that there is a value of t for which ρ becomes infinite. Choosing that value as the origin of t, the expressions for ρ, p simplify to:

$$\begin{aligned} \rho &= \frac{4}{3t^2(1+\omega)^2}, \\ p &= \frac{4\omega}{3t^2(1+\omega)^2}, \end{aligned}$$

and for $\omega = -1$, conveniently defining $H = \sqrt{\frac{\rho_0}{3}}$, we can write:

$$\begin{aligned} \rho &= 3H^2, \\ p &= -3H^2. \end{aligned}$$

Hence, it is direct to compute that this model fulfills the WEC if $\omega \geq -1$, the DEC if $-1 \leq \omega \leq 1$ and the SEC if $\omega \geq -\frac{1}{3}$.

The expansion of the fluid congruence is $\theta = \sqrt{3\rho}$, which is equivalent to $\theta = \frac{2}{t(1+\omega)}$ except for $\omega = -1$, for which we simply have $\theta = 3H$.

The hypothesis of Raychaudhuri's Theorem 2.7 are fulfilled for certain values of ω. The theorem requires the fluid congruence to be geodesic and irrotational, which is fulfilled by all FLRW spacetimes, and $\theta_0 > 0$ and the SEC to be satisfied. For this to happen, we have already seen that $\omega \geq -\frac{1}{3}$, in which case θ always reaches positive values. Hence, if $\omega \geq -\frac{1}{3}$, the universe has a matter singularity at a finite past and we are talking about a cosmological model with a Big Bang.

However, although Theorem 2.7 fails to apply when $\omega < -\frac{1}{3}$, there still is a matter singularity for $\omega > -1$ since $a(0) = 0$ in that range, making $t = 0$ singular as seen before.

Let us study further the De Sitter cosmology. Its metric can be written as:

$$ds^2 = -dt^2 + e^{2Ht}\left[dr^2 + r^2\left(d\theta^2 + \sin^2\theta\; d\phi^2\right)\right],$$

with $t > 0$.

As we have seen before, this spacetime satisfies the WEC and DEC but not the SEC.

Since the SEC does not hold, neither Raychaudhuri's Theorem 2.7 nor Hawking-Penrose Theorem 2.19 can be applied.

There is only Penrose's Theorem 2.18 left to check. In this particular case, the NCC reads $\rho + p \geq 0$, which is fulfilled (because $\rho + p = 0$). Also, the hypersurfaces $\{t = t_0\}$ are Cauchy hypersurfaces and non-compact. One can compute $\kappa = 2H^2 - \frac{2e^{-2Ht}}{r^2}$ at the maximum value of the coordinate r of the closed surfaces, so they are trapped when their range of $r \geq \frac{1}{H}e^{-HT}$. This means there exist closed trapped surfaces. Hence, this spacetime has at least an incomplete geodesic.

We can try to build an extension. By using the coordinate change [2] (p. 124):

$$\begin{aligned}
t &= \frac{1}{H}\ln\left(\cosh(H\hat{t})\cos\chi + \sinh(Ht)\right),\\
x &= \frac{1}{H}\frac{\cosh(H\hat{t})\sin\chi\cos\theta}{\cosh(H\hat{t})\cos\chi + \sinh(H\hat{t})},\\
y &= \frac{1}{H}\frac{\cosh(H\hat{t})\sin\chi\sin\theta\cos\phi}{\cosh(H\hat{t})\cos\chi + \sinh(H\hat{t})},\\
z &= \frac{1}{H}\frac{\cosh(H\hat{t})\sin\chi\sin\theta\sin\phi}{\cosh(H\hat{t})\cos\chi + \sinh(H\hat{t})},
\end{aligned}$$

one obtains the metric

$$ds^2 = -d\hat{t}^2 + \frac{\cosh^2(H\hat{t})}{H^2}\left(d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\theta d\phi\right)\right),$$

with $\hat{t} \in \mathbb{R}, \chi \in [0,\pi], \theta \in [0,\pi], \phi \in [0,2\pi]$. Note that the singularities at $\chi = 0,\pi$ and $\theta = 0,\pi$ are due the spherical coordinates only and are totally removable. Now, this spacetime is regular, as can be seen from the Penrose diagram in Figure 4.2.

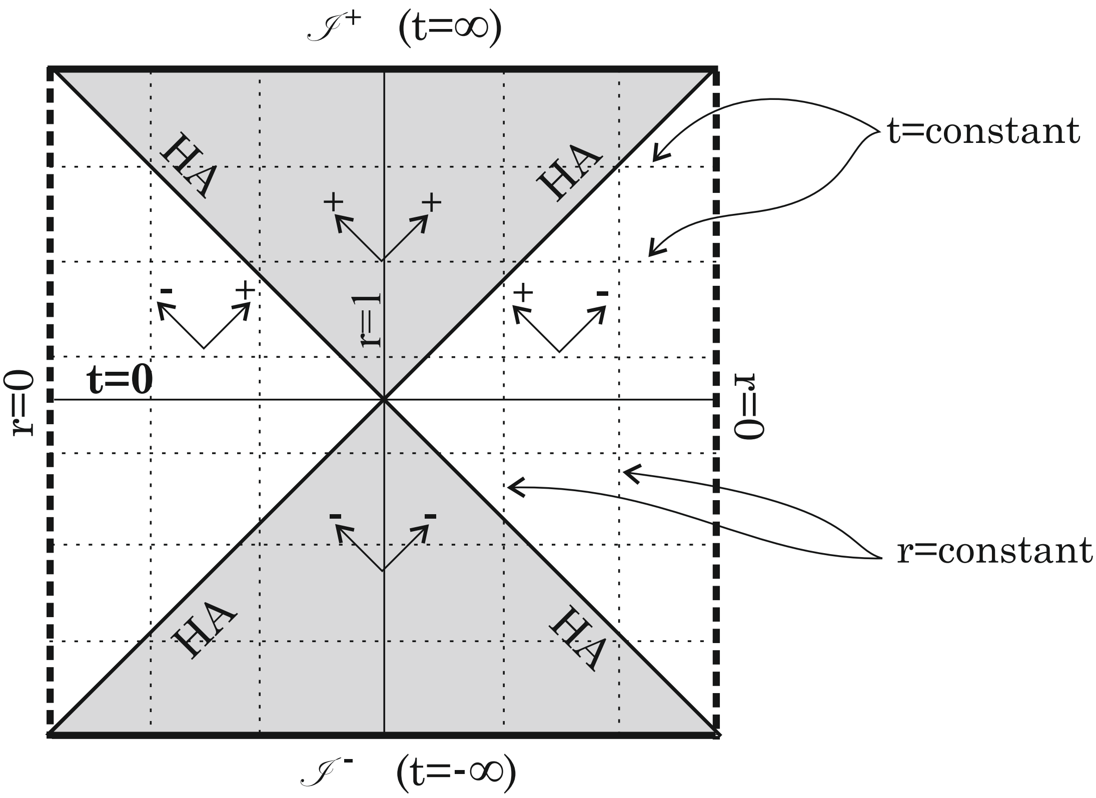

Figure 4.2: Penrose diagram for the De Sitter universe [10]. In the shaded regions there are closed trapped surfaces. The horizontal dashed lines represent the curves $t = constant$ and the vertical dashed lines represent $r = constant$.

Notice that this extended spacetime is spatially closed with $0 < r < 1$. To see this, use the variable change $r = \sin\chi$, so that $d\chi = \frac{1}{\sqrt{1-r^2}}dr$. Using this, the metric becomes:

$$ds^2 = -d\hat{t}^2 + \frac{\cosh^2(H\hat{t})}{H^2}\left(\frac{dr^2}{1-r^2} + r^2\left(d\theta^2 + \sin^2\theta d\phi\right)\right),$$

which is the metric 4.1 with $k = 1$, corresponding to a spatially closed universe. We will drop the hat of $\hat{t}$ from now on.

In the adapted reference frame, we find that the energy momentum tensor is of the type of a perfect fluid, again with:

$$\rho = 3H^2,$$
$$p = -3H^2.$$

Hence, the WEC and DEC are fulfilled, but the SEC is not. The expansion of the fluid congruence is $\theta = 3H\tanh(Ht)$ and its acceleration is $a_\mu = 0$.

Hence, although the fluid congruence is geodesic, and $\theta > 0$ for some t, we can not use Raychaudhuri's Theorem 2.7 because the SEC is violated.

The NCC reads $\rho + p \geq 0$, which is fulfilled because the equality is achieved. The spacetime is globally hyperbolic because there exist Cauchy hypersurfaces of the form $\{t = \textit{constant}\}$. However, since it is spatially closed, the Cauchy hypersurfaces can not be non-compact. Hence, Penrose's Theorem 2.18 can not be applied.

As for Hawking-Penrose Theorem 2.19, as we have already discussed, there exists a compact achronal set without edge (a Cauchy hypersurface). Looking at the existence of closed trapped surfaces, consider κ for a point which maximizes χ in such surfaces. Then, using the null normal vectors $k^+ = \partial_t - H\frac{\sqrt{1-r^2}}{\cosh(Ht)}\partial_r$, $k^- = \frac{1}{2}\partial_t - H\frac{\sqrt{1-r^2}}{2\cosh(Ht)}\partial_r$, it is direct to compute that

$$\kappa = \frac{H^2\left(r^2(1+\cosh(2Ht)) - 2\right)}{r^2\cosh^2(Ht)},$$

hence closed surfaces with all of their points fulfilling $\cosh(2Ht) \geq \frac{2-r^2}{r^2}$ will be trapped.

However, although the spacetime is globally hyperbolic, hence it also fulfills the chronology condition and it also contains closed trapped surfaces, since the SEC is violated, we can not apply Hawking-Penrose Theorem 2.19.

Hence it seems that the violation of the energy condition SEC along with the fact that this spacetime is spatially closed helps at avoiding singularities.

4.2 Senovilla cosmology

Before this cosmological model was published, the regularity of the known solutions was due to the unfulfillemnt of an energy, causal or generic condition, which allowed them to avoid the singularity theorems. This posed the question of whether a regular spacetime could satisfy the desirable conditions in order to be considered a cosmological model. In fact, even the desirable conditions for a spacetime to be considered an acceptable cosmological model have been under discussion as explained in [1]. It seems that an accepted definition for a theoretical classical cosmological model is any non-static spacetime filled with matter. However, since we do not completely know everything about our universe, this kind of definitions are speculative and strongly depend on our current knowledge, hence they are subject to changes if new physical discoveries are made. Nevertheless, for the moment, we will use this definition of a cosmological model.

The first such regular spacetime to be found was a manifold with the metric:

$$ds^2 = \cosh^4(at)\cosh^2(3a\rho)\left(-dt^2 + \frac{\sinh^2(3a\rho)}{\cosh^2(3a\rho)-1}d\rho^2\right) + \\ + 4\cosh^4(at)\frac{\cosh^2(3a\rho)-1}{36a^2\cosh^{\frac{2}{3}}(3a\rho)}d\phi^2 + \cosh^{-2}(at)\cosh^{-\frac{2}{3}}(3a\rho)dz^2,$$

with $a > 0$. The full analysis of why this spacetime is regular is in [12], where all the geodesics are computed and their completeness (and other properties) is checked. Namely, it is shown that every null geodesic meets each of the hypersurfaces $\{t = t_0\}$. Hence, this means that they are Cauchy hypersurfaces, and hence, the spacetime is globally hyperbolic. Thus, it is causally simple (and the chronology condition also holds) and t a time function (its gradient is timelike). Since every causal curve has to meet each $\{t = t_0\}$ and t is a time function, then every causal curve can be extended to arbitrary values of their affine parameter, so the spacetime is causal b-complete and hence, regular.

Taking the observer with velocity ∂_t and using the natural orthonormal basis suggested by the coordinates, one finds that the spacetime is filled with a perfect fluid with:

$$\rho = 15a^2\text{sech}^4(at)\text{sech}^4(3a\rho),$$
$$p = 5a^2\text{sech}^4(at)\text{sech}^4(3a\rho),$$

so it is obvious that the spacetime satisfies the WEC, DEC and SEC everywhere. Additionally, since the SEC is fulfilled strictly (and hence also the NCC), by Proposition 1.25, the generic condition holds.

The expansion of the fluid congruence is

$$\theta = 3a\ \text{sech}^2(at)\text{sech}(3a\rho)\tanh(at),$$

and its acceleration is

$$a_\mu = -3a\left(\cosh(6a\rho) - 1\right)\sinh^{-1}(6a\rho)\delta^\rho_\mu.$$

Even though this spacetime satisfies the SEC and $\theta > 0$, Theorem 2.7 can not be applied because its acceleration is non-zero, hence the fluid congruence that fills the spacetime is not geodesic.

As for Penrose Theorem 2.18, although the NCC holds and there exists a non-compact Cauchy hypersurface, there are not any closed trapped surfaces. To see this, take a compact without boundary spacelike surface S. Since it is compact, it must have a maximum value of the coordinate ρ at a certian point p. At p, the normal vector to S must be a linear combination of ∂_t and ∂_ρ (since at p the tangent vectors to S must be linear combinations of ∂_θ and ∂_ϕ only, because S is spacelike and ρ is maximum at p). Then, at p we have

$$\kappa = -\frac{1}{8}a^2\frac{(\cosh(a(t-6\rho)) + 5\cosh(at))(\cosh(a(6\rho+t)) + 5\cosh(at))}{\sinh^2(3a\rho)\cosh^4(3a\rho)\cosh^6(at)} < 0.$$

Hence, there does not exist any closed trapped surface.

As for Hawking-Penrose Theorem 2.19, something similar happens. The SEC, chronology and generic conditions are fulfilled. However, the initial condition is not. We have already seen that closed trapped surfaces can not exist. Light cones do not reconverge at any point because considering the null radial geodesics, one arrives at the system:

$$\dot{t} = |\dot{\rho}|, \quad \dot{\rho} = \cosh^{-4}(at)\cosh^{-2}(3a\rho).$$

It is clear that for future directed outgoing null radial geodesics, ρ it not bounded. Hence, through each point there is at least a null geodesic that diverges, which means that the light cones can not reconverge. Finally, compact achronal edgeless sets can not exist due to the spatial openness of the spacetime.

Thus, it is clear that this spacetimes satisfies all the desirable energy, causal and the generic conditions and that despite this, it is regular because the initial condition required by the theorems is not satisfied. Hence, there is still place for regular cosmological models fulfilling the conditions in which we are interested despite the existence of the singularity theorems. Unfortunately, this spacetime is not a realistic model for our universe due to its cylindrical symmetry, since we observe spatial isotropy and homogeneity.

4.3 Modification of Schwarzschild Metric

In this section, we will set the classical gravitation constant to $G_0 = 1$ in numerical computations (and will be explicit in the expressions), hence Einstein's equations become $G_{\mu\nu} = 8\pi G_0 T_{\mu\nu}$.

The very first exact solution of Einstein's equations was the Schwarzschild metric. This spacetime represents the empty exterior of a chargeless, non-radiative star. The original spacetime describes the region $r > 2M$, but it can be extended to cover the values $r < 2M$. This new region is the so called black hole region and it has a singularity at $r = 0$.

The following metric is a modification of the Schwarzschild metric, allowing for new physical effects:

$$ds^2 = -\left(1 - \frac{2M(r)}{r}\right) dt^2 + \left(1 - \frac{2M(r)}{r}\right)^{-1} dr^2 + r^2 d\Omega^2,$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta \, d\phi^2$ is the usual volume element on the 2-spheres.

Using the orthonormal frame of reference $\{\left(1 - \frac{2M(r)}{r}\right)^{-1/2} \partial_t, \left(1 - \frac{2M(r)}{r}\right)^{1/2} \partial_r, \frac{1}{r}\partial_\theta, \frac{1}{r\sin\theta}\partial_\phi\}$ for the region $r > 2M(r)$ and $\{\left(1 - \frac{2M(r)}{r}\right)^{1/2} \partial_r, \left(1 - \frac{2M(r)}{r}\right)^{-1/2} \partial_t, \frac{1}{r}\partial_\theta, \frac{1}{r\sin\theta}\partial_\phi\}$ for the region $r < 2M(r)$, the energy-momentum tensor becomes diagonal with a preferred axis, with:

$$\begin{aligned} \rho &= \frac{1}{4\pi G_0}\frac{M'(r)}{r^2}, \\ p_r &= -\frac{1}{4\pi G_0}\frac{M'(r)}{r^2}, \\ p_T &= -\frac{1}{8\pi G_0}\frac{M''(r)}{r}. \end{aligned} \tag{4.5}$$

It can be seen that with these coordinates, the geodesics approaching $r = 2M(r)$ are not complete, hence it is a singularity. However, one can introduce an extension whose coordinates are given by:

$$(u, r, \theta, \phi) = \left(t + r + 2M(r)\ln\left(\frac{r}{2M(r)} - 1\right), r, \theta, \phi\right),$$

which transforms the metric into:

$$ds^2 = -\left(1 - \frac{2M(r)}{r}\right) du^2 + 2dudr + r^2 d\Omega^2, \tag{4.6}$$

with $u \in (-\infty, \infty)$, $r > 0$. See Figure 4.10 for an example.

Now, $r = 2M(r)$ is a regular point, so that it is a removable singularity and we will not study it further.

The other suspicious point could be $r = 0$. To study it, it is usual to do a local analysis of the spacetime around it. Assume a Taylor expansion of $M(r) = Ar^n + O(r^{n+1})$, $n \geq 0$, $A \neq 0$. One has to keep in mind that the results obtained from this analysis only hold in a small neighbourhood of $r = 0$. To study this point, let us state a useful result:

Theorem 4.1. [13] Assume the metric of a spacetime can be written in the form:

$$ds^2 = -a(r)dt^2 + b(r)dr^2 + r^2 d\Omega^2,$$

with $a(r), b(r) \in C^2$. All the second order scalar curvature invariants of the spacetime are finite at $r = 0 \iff a(0) = b(0) = 1$ and $a'(0) = b'(0) = 0$.

Since our spacetime is spherically symmetric, we can use Theorem 4.1. In our case, $a(r) = (1 - 2Ar^{n-1})$, $b(r) = (1 - 2Ar^{n-1})^{-1}$ and hence,

$$a'(r) = -2A(n-1)r^{n-2},$$
$$b'(r) = \frac{2A(n-1)r^{n-2}}{(1 - 2Ar^{n-1})^2}.$$

Clearly, $a(0) = b(0) = 1$ for $n \geq 2$ and $a'(0) = b'(0) = 0$ is fulfilled if $n \geq 3$. Thus, the theorem tells us that for $n < 3$, the spacetime has a scalar curvature singularity at $r = 0$, and that for $n \geq 3$ it has not (however, there still is the possibility that there is a singularity of another kind).

Improved Schwarzschild metric

We now analyze a particular case of this metric arising when quantum corrections are taken into account [14]. The intention is to build a model of a regular black hole which takes into account quantum gravity effects. In this scenario, the gravitational constant becomes a function of r, which translates into :

$$M(r) = \frac{G_0 M r^3}{r^3 + \tilde{\omega} G_0 (r + \gamma G_0 M)},$$

where G_0 is the gravitational constant (since we are dealing with quantum models, it is useful to use natural units $c = \hbar = 1$) and quantum field theory predicts the values $\gamma = \frac{9}{2}$, $\tilde{\omega} = \frac{118}{15\pi}\hbar$. We will use these values for the numerical computations. This spacetime represents the quantum vacuum surrounding a massive spherically symmetric object. However, one can still use Einstein's equations to define effective energy density and pressures as in expression 4.5. M represents the total mass of the black hole (in units of the Planck mass m_P), as it can be checked by computing

$$\int_0^\infty \rho(r) 4\pi r^2 dr = M.$$

The behaviour of $M(r)$ is shown in Figure 4.3.

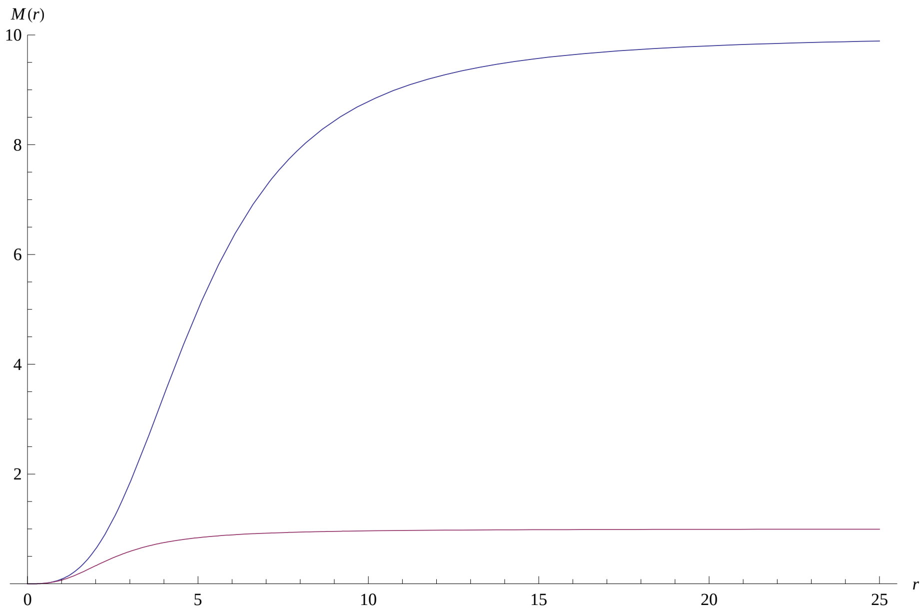

Figure 4.3: Shape of $M(r)$ using the values $M = 10$ for the upper curve and $M = 1$ for the lower curve (and the values $G_0 = 1$, $\gamma = \frac{9}{2}$ and $\tilde{\omega} = \frac{118}{15\pi}$). When $r \to 0$, $M(r) \to 0$ and as $r \to \infty$, $M(r) \to G_0 M$.

Its Taylor expansion at $r = 0$ is:

$$M(r) \simeq \frac{r^3}{\tilde{\omega}\gamma G_0} - \frac{r^4}{\tilde{\omega} M \gamma^2 G_0^2} + O(r^5),$$

hence, by the previous analysis, $r = 0$ is not a scalar singularity.

As for the energy conditions, the WEC is satisfied if $\rho \geq 0$, $\rho + p_r \geq 0$ and $\rho + p_T \geq 0$. In this case, the WEC is fulfilled in all points. The DEC is satisfied if the WEC holds and $\rho - p_r \geq 0$ and $\rho - p_T \geq 0$. This is satisfied in the region $r \leq R_{DEC}$, where R_{DEC} is the only positive solution (see Figures 4.4 and 4.5) of:

$$F_{DEC}(R_{DEC}) = R_{DEC}^4 + 3G_0 M\gamma R_{DEC}^3 - 3G_0\omega R_{DEC}^2 - 8G_0^2 M\gamma\omega R_{DEC} - 6G_0^3 M^2\gamma^2\omega = 0.$$

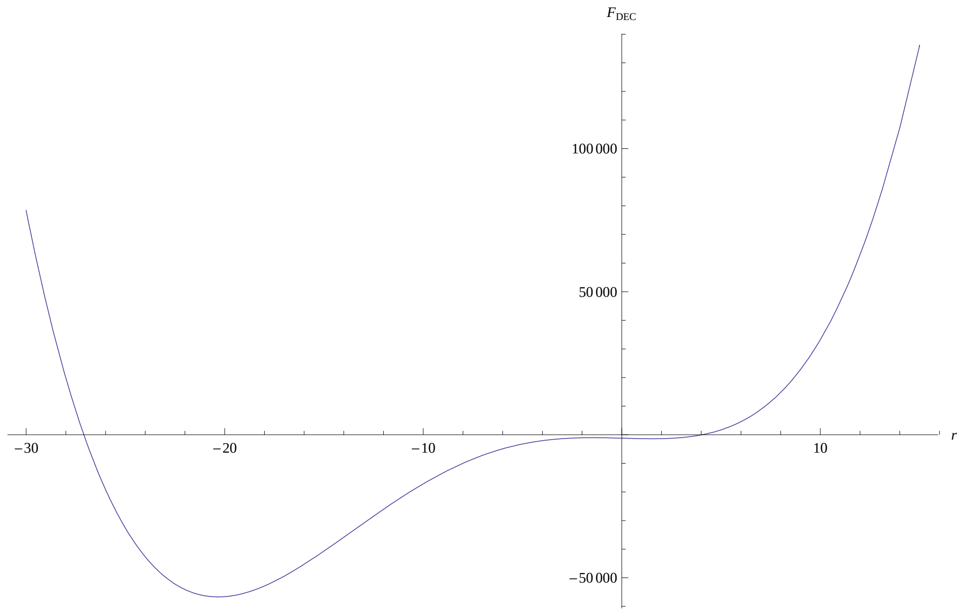

Figure 4.4: 4th order polynomial $F_{DEC}(r)$ for $M = 2$. It has two complex conjugate roots, one negative root and one positive root. The positive root is R_{DEC}. The qualitative result is independent of the specific value of M.

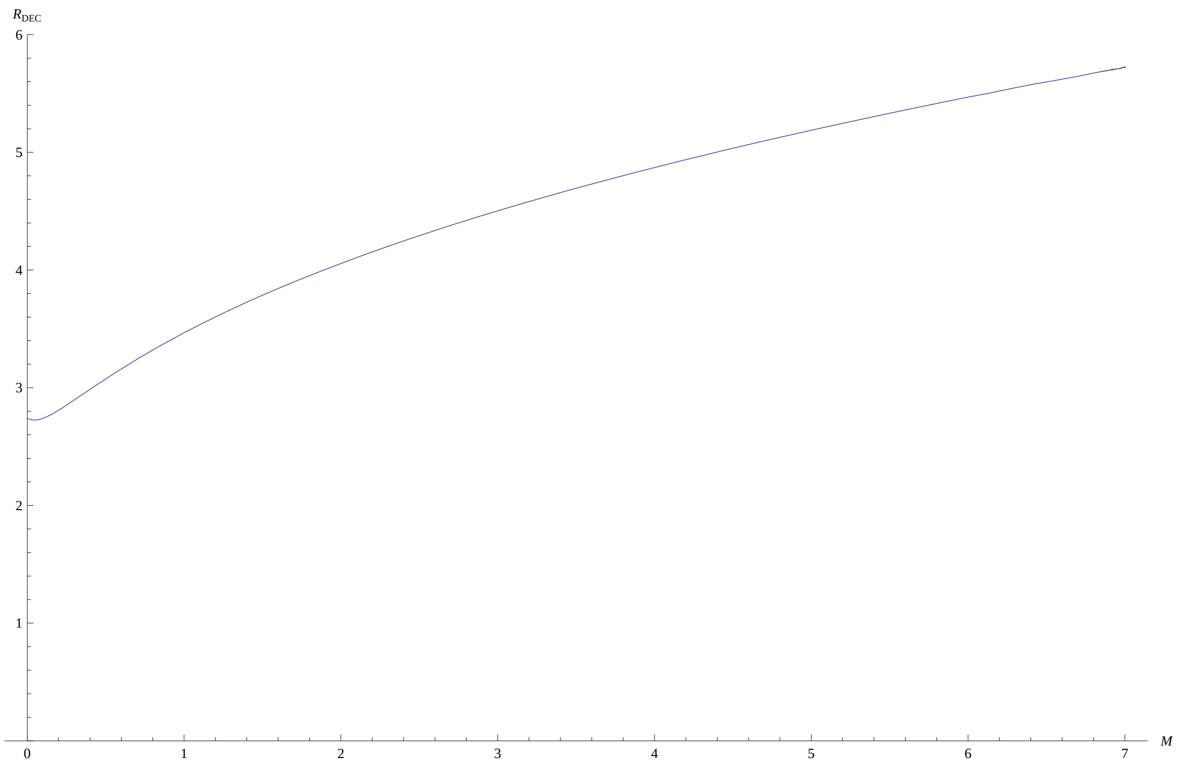

Figure 4.5: Evolution of the limit of the region where the DEC holds, R_{DEC} with M.

The SEC is satisfied if $\rho + p_r \geq 0$, $\rho + p_T \geq 0$ and $\rho + p_r + 2p_T \geq 0$. This holds in the region $r > R_{SEC}$, where R_{SEC} is the positive solution (see Figures 4.6 and 4.7) of:

$$3R_{SEC}^4 + 6G_0 M\gamma R_{SEC}^3 - G_0\omega R_{SEC}^2 - 3G_0^2 M\gamma\omega R_{SEC} - 3G_0^3 M^2\gamma^2\omega = 0.$$

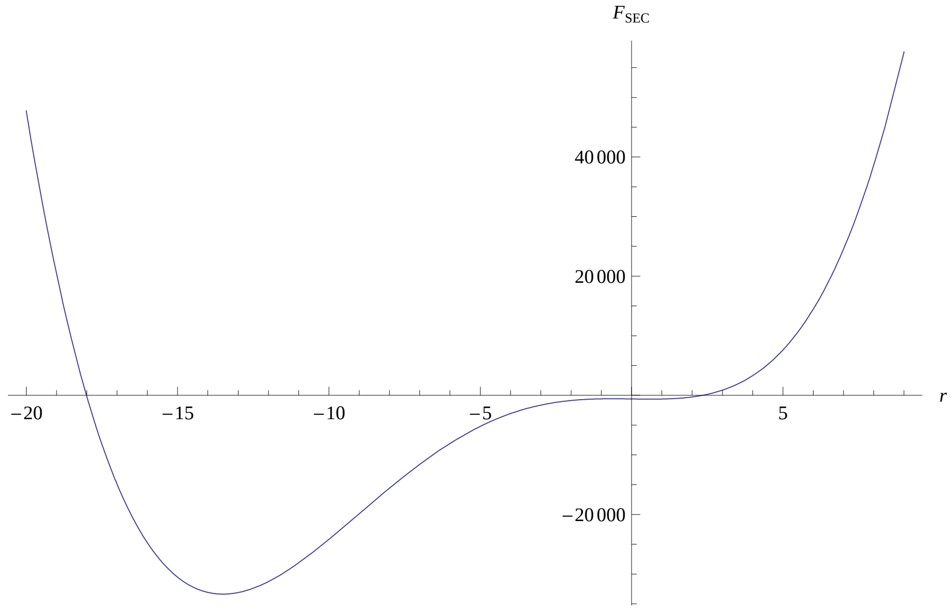

Figure 4.6: 4th order polynomial $F_{SEC}(r)$ for $M = 2$. It has two complex conjugate roots, one negative root and one positive root. The positive root is R_{SEC}. The qualitative result is independent of the specific value of M.

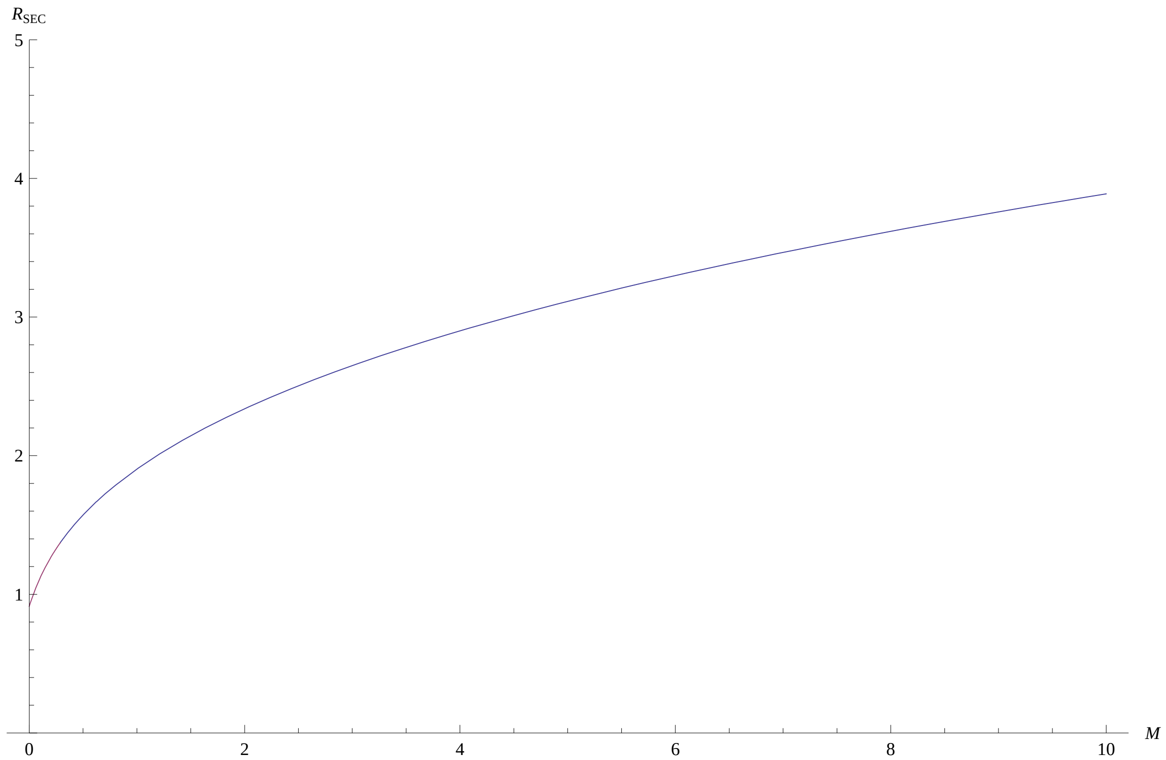

Figure 4.7: Evolution of the limit of the region where the SEC holds, R_{SEC} with M.

We now study the existence of closed trapped surfaces. Since they are compact, they must have a maximum (and a minimum) value of the coordinate r. At those points, the normal null vectors to the surface must be a linear combination of ∂_t and ∂_r. Taking the future pointing normal vectors $k^+ = \partial_t + \frac{r-2M(r)}{r}\partial_r$ and $k^- = \frac{r}{2(r-2M(r))}\partial_t - \frac{1}{2}\partial_r$ for the region $r > 2M(r)$ and $k^+ = \partial_t - \frac{r-2M(r)}{r}\partial_r$ and $k^- = \frac{r}{2(r-2M(r))}\partial_t + \frac{1}{2}\partial_r$ for the region $r < 2M(r)$ we obtain:

$$\kappa = -\frac{2\left(r-2M(r)\right)}{r^3} = -\frac{2}{r^2}\frac{r^3 - 2G_0 M r^2 + \omega G_0 r + G_0^2 M\omega\gamma}{r^3 + G_0\omega\left(r + G_0 M\gamma\right)}.$$

To study the behaviour of κ, we only need to track the sign of the third order polynomial $F(r) = r^3 - 2G_0Mr^2 + \omega G_0 r + G_0^2 M\omega\gamma$. First, we should find its number of roots. To do this, it is enough to study its discriminant:

$$\Delta = 32M^4\gamma\omega G_0^5 + 4M^2\omega^2 G_0^4 - 27M^2\gamma^2\omega^2 G_0^4 - 36M^2\gamma\omega^2 G_0^4 - 4\omega^3 G_0^3.$$

If $\Delta > 0$, the polynomial has 3 different real roots, if $\Delta < 0$, the polynomial has 1 real root and if $\Delta = 0$, all roots are real but there are at least two equal roots. These different behaviours are collected in Figure 4.8.

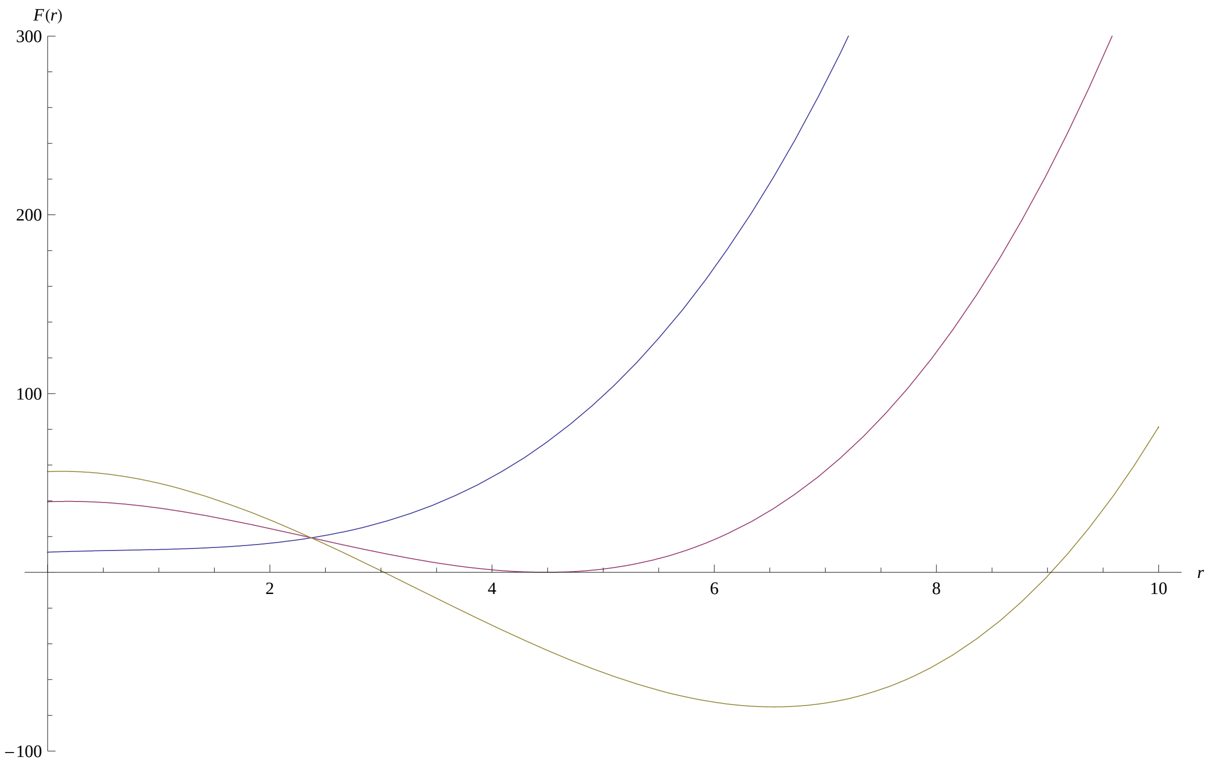

Figure 4.8: Possible shapes of $F(r)$ depending on the value of its discriminant. The values of M used are $M = 1$ for the case $\Delta < 0$ (blue curve), $M = M_{crit} \simeq 3.5$ for the case $\Delta = 0$ (purple curve) and $M = 5$ for the case $\Delta > 0$ (yellow curve).

Hence, depending on the value of M we have different scenarios. Its critial value, at which $\Delta = 0$, is:

$$M_{crit} = \sqrt{\frac{\omega}{64G\gamma}}\sqrt{36\gamma + 27\gamma^2 + \sqrt{2+\gamma}\left(2+9\gamma\right)^{\frac{3}{2}} - 4}.$$

For $M > M_{crit}$, $\Delta > 0$ and for $M < M_{crit}$, $\Delta < 0$. Hence, if $M < M_{crit}$, there is only 1 real root, which must be at a negative value of r since $F(0) = G_0^2 M\omega\gamma > 0$ and $\lim_{r\to-\infty} F(r) = -\infty$. Hence, $F(r) > 0$ for the relevant region $r > 0$ and $\kappa < 0$. Thus, for $M < M_{crit}$ there are not any closed trapped surfaces.

For $M > M_{crit}$, there is a region of r in which $F(r) < 0$. Let us call the two roots delimiting this region r_{min}, r_{max}, which are the two positive roots of $F(r)$. If $r_{min} < r < r_{max}$, then $\kappa > 0$. Hence, if the r coordinate of a compact closed surface is inside the limits $r_{min} < r < r_{max}$, it is a closed trapped surface.

This means that for heavy black holes ($M > M_{crit}$), there are two event horizons: r_{min} and r_{max}. Outside these horizons, $F(r) > 0$, and since we can write $g_{tt} = -\frac{r-2M(r)}{r} = -\frac{F(r)M(r)}{G_0 M r^3}$ and $M(r) > 0$, the vector ∂_t is timelike. In contrast, inside both horizons, $F(r) < 0$ and ∂_t is spacelike. This mimics the behaviour of the classical Schwarzschild black hole. Note that for the limit $M \to \infty$, $F(r) \simeq r^3 - 2G_0 M r^2 = r^2(r - 2G_0 M)$, which recovers the value of the horizon of the classical Schwarzschild black hole, $r = 2G_0 M$.

As M decreases and approaches M_{crit}, the two horizons get closer until at $M = M_{crit}$ they coincide in the double zero of $F(r)$. If we decrease M further, then the horizons disappear altogether. This is depicted in Figure 4.9.

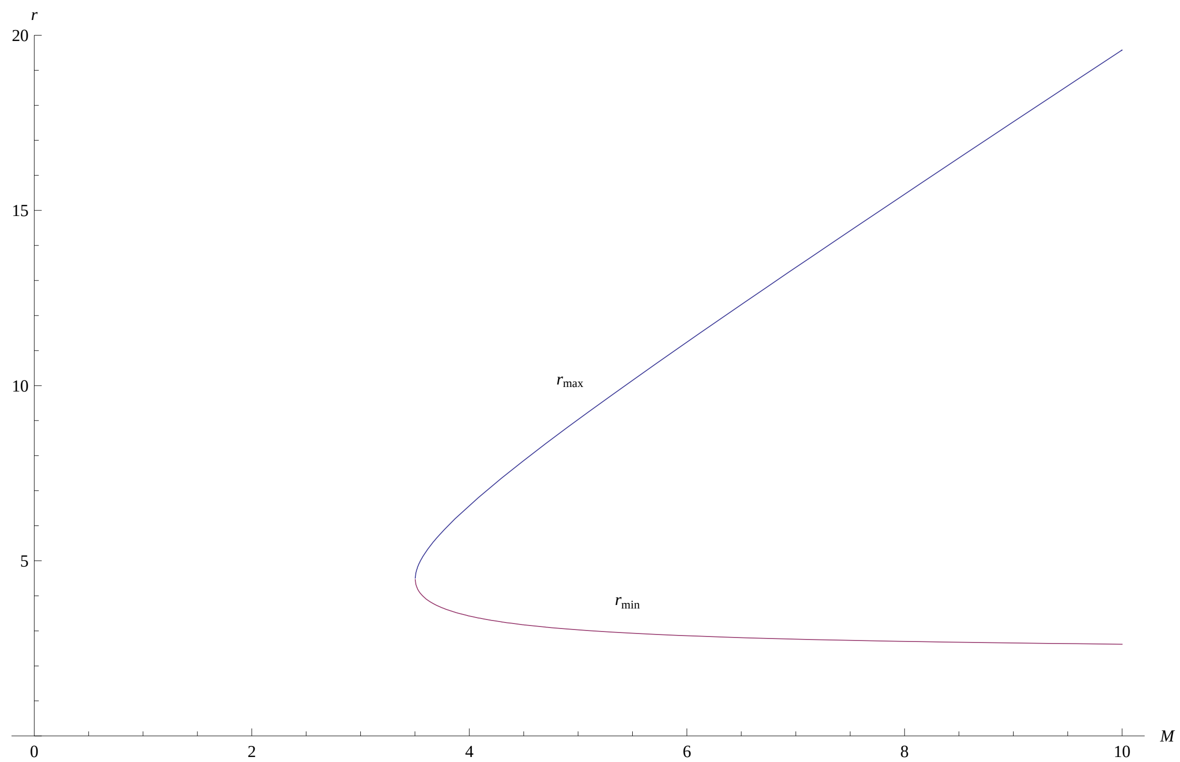

Figure 4.9: Evolution of the apparent horizons r_{min}, r_{max} as a function of the mass of the black hole. The horizons only appear for $M > M_{crit}$.

As for the singularity theorems in this case, Raychaudhuri's Theorem 2.7 can not be applied because this spacetime is not filled by a perfect fluid.

Hawking-Penrose Theorem 2.19 can not be applied because the SEC is only fulfilled in a certain region $r > R_{SEC}$.

Let us try with Penrose's Theorem 2.18. The NCC in this spacetime translates into

$\rho + p_T + a^2(p_r - p_T) \geq 0$ for $1 \geq a \geq 0$. Since $p_r - p_T = -(\rho + p_T) \leq 0$ as we have already seen, then $\rho + p_T + a^2(p_r - p_T) \geq \rho + p_T + p_r - p_T = \rho + p_r = 0$ and the NCC holds. Penrose's Theorem 2.18 also asks for a closed trapped surface, which happens only for $M > M_{crit}$ and for a non-compact Cauchy hypersurface. However, in this spacetime there are Cauchy hypersurfaces only for $M < M_{crit}$. It can be seen from the Penrose diagram for $M > M_{crit}$ in Figure 4.10 that the horizon $\{r = r_{min}\}$ is a Cauchy horizon for any edgeless acausal hypersurface in regions I and I', hence their Cauchy development can never be all the manifold.

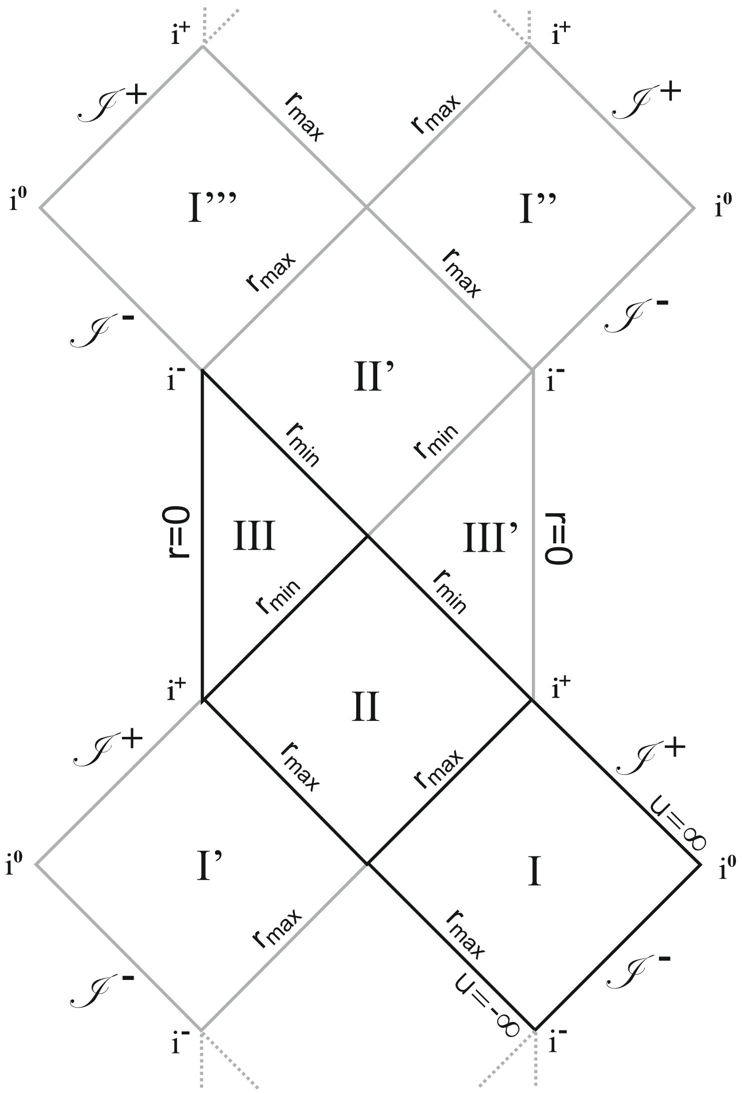

Figure 4.10: Penrose diagram of the maximal extension of the quantum regular black hole for $M > M_{crit}$ [10]. The regions labeled with the same number are isometric. The regions I are asymptotically flat. Regions II and III are part of the BH. In II, the 2-spheres are trapped. Note that $r = 0$ is not a singularity. There exist radial timelike geodesics crossing the event horizons r_{max} and r_{min} and then bouncing in region III and leaving the black hole [14] (Section G). The regions I, II and III (emphasised with bold lines) are the regions covered by the coordinates used in metric 4.6.

Hence it seems that the fact this spacetime is regular comes from the fact that the SEC does not hold in all its points and that its causal structure is non-trivial.

Conclusion

The main singularity theorems and their proofs have been reviewed by studying in detail the necessary tools and the main results. An insight into both the mathematical and physical views of the reasons and assumptions that motivated and led to the theorems has been made. Special attention has been payed to the feasibility and applicability of some of them while exploring what might happen when considering some variations in their hypothesis.

A variety of interesting regular and singular spacetimes have been studied. The features that make them fulfill the hypothesis of the theorems or, on the contrary, that allow them to avoid the key hypothesis have been identified.

Although some of the early spacetimes obtained from Einstein's equations were already singularity-free, other relevant solutions including both cosmological and astrophysical models were not. After the singularity theorems were proved, influenced by their strength, one could have though that regular cosmological models outside the FLRW family (since it is already difficult to find regular FLRW models) or regular astrophysical objects such as the black hole could not exist. However, we now know that the singularity theorems are not as restrictive as it was thought when they were first developed and that there are ways to avoid the singularities while preserving some desirable conditions. Hence, by having studied the singularities, we have been able to see that, although in certian situations General Relativity seems to favour their existence, there are still ways to avoid them. This has been done many times, using the own singularity theorems as a guide to build regular spacetimes. However, as we have seen, this will probably imply that in order to avoid the key points in the theorems, the resulting spacetimes tend to be rather artificial and hence little useful to model our universe. The natural question arising is whether we will be able to do so with realistic spacetimes that can describe our universe or astrophysical objects.

Although the results seen here are powerful and meaningful, we have also pointed out that they are far from ideal. As we have seen, the theorems just predict the incompleteness of at least one geodesic and we do not know if the singularity is removable or essential, as well as its location or severity. However, although this result is rather vague, singularity theorems are still a big achievement because of their generality and historical, physical and mathematical importance.

Hence an obvious direction for further development would be to improve the theorems, namely, improving in terms of differentiability of the metric, taking into consideration more plausible energy conditions or improving the rather vague conclusion of the theorems. However, as we have already mentioned, these refinements are very challenging mathematically.

Another important advancement would be to include quantum effects in the theory. Recall that all the results and developments in the first chapters were built with General Relativity in mind, but they are valid for any gravitational theory using a manifold as the spacetime. As we have seen in Section 4.3, quantum effects may change the causal structure of the spacetime or make it violate the energy conditions. This makes it harder for singularities to exist, allowing for regular black holes, for instance. If quantum theory is included in cosmological models, then theories describing early stages of the universe such as inflation arise. Inflation predicts high negative pressures, which violate the usual energy conditions. Hence this means that the classical theorems do not apply in this case, leaving the door open for regular models. Also, if quantum gravity is considered, it is possible to obtain regular cosmological models. It seems, then, that quantum effects are likely to solve many problems of classical General Relativity. In this sense, the next step would be to study different approaches to quantum gravity and to modify the singularity theorems accordingly to achieve a more complete theory.

Bibliography

[1] J. M. M. Senovilla, "Singularity theorems and their consequences", General Relativity and Gravitation **30**, 701–848 (1998).

[2] S. Hawking and G. Ellis, *The large scale structure of space-time* (Cambridge University Press, 1973).

[3] E. Poisson, *A relativist's toolkit* (Cambridge University Press, 2004).

[4] R. M. Wald, *General relativity* (The University of Chicago Press, 1984).

[5] M. Spivak, *Calculus on manifolds : a modern approach to classical theorems of advanced calculus*, Mathematics monograph series (Reading, Mass. Addison-Wesley, 1965).

[6] J. M. M. Senovilla, "Trapped surfaces, horizons, and exact solutions in higher dimensions", Class. Quant. Grav. **19**, L113 (2002), arXiv:`hep-th/0204005 [hep-th]`.

[7] S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, NY, 1972).

[8] T. A. Roman, "On the "averaged weak energy condition"and penrose's singularity theorem", Physical Review D **37**, 546 (1988).

[9] A. Borde, "Geodesic focusing, energy conditions and singularities", Classical and Quantum Gravity **4**, 343 (1987).

[10] R. Torres, "Construcción de modelos de estrellas y burbujas de vacío o radiativas, en el marco de la relatividad general: un planteamiento global", PhD thesis (UPC, Barcelona, Spain, 2001).

[11] S. M. Carroll, *Spacetime and geometry: An introduction to general relativity* (2004).

[12] F. J. Chinea, L. Fernández-Jambrina, and J. M. M. Senovilla, "Singularity-free spacetime", Phys. Rev. D **45**, 481–486 (1992).

[13] K. Lake and P. Musgrave, "The regularity of static spherically cylindrically and plane symmetric spacetimes at the origin", General Relativity and Gravitation **26**, 917–925 (1994).

[14] A. Bonanno and M. Reuter, "Renormalization group improved black hole space-times", Phys. Rev. **D62**, 043008 (2000), arXiv:`hep-th/0002196 [hep-th]`.

Glossary

$D^+(\zeta)$ Cauchy domain of dependence 25

$E^+(\zeta)$ Future horismos 22

$H^+(\zeta)$ Cauchy horizon 25

$I^+(\zeta)$ Chronological future 22

$J^+(\zeta)$ Causal future 22

$\mathcal{C}^{k-}$ **function** 3

$\mathcal{N}_p$ Maximal normal neighbourhood 7

acausal 23

achronal 23

affine parameter 7

b-completeness 32

Cauchy domain of dependence 25

Cauchy horizon 25

Cauchy hypersurface 26

causal 4

causal future 22

chronological future 22

chronology condition 24

closed trapped surface 28

congruence 11

conjugate points 17

curve 4