UNIVERSITAT DE BARCELONA

Facultat de Matemàtiques
i Informàtica

GRAU DE MATEMÀTIQUES

Treball final de grau

General Theory of Relativity: Mathematical Elements and Hawking's Singularity Theorem

Autor: Sergi Novell Masot

Director: Dr. Ricardo García López
Realitzat a: Departament
de Matemàtiques i Informàtica

Barcelona, 16 de juny de 2020

Contents

Abstract

In this work, we study Riemannian and pseudo-Riemannian manifolds and their main properties. From them, we examine the special and general theories of relativity, and see how they arise from modelling space-time as special kinds of pseudo-Riemannian manifolds, the Lorentzian manifolds. Within this theory, we are able to give a rigorous formulation of the fundamental properties of cosmology and the Schwarzschild space-time.

We also wish to relate the behaviour of geodesics in a manifold with the intrinsic structure of the manifold. This results in the formulation of the Hopf-Rinow theorem in the case of Riemannian manifolds, and the Hawking singularity theorem, in Lorentzian manifolds.

2010 Mathematics Subject Classification. 58A05, 53Z05

Acknowledgements

I want to first express my deepest gratitude to my advisor, Ricardo García, for his one-year-long guidance, during which he has given me all the help and orientation I needed to progress on this work. Even during these tough and uncertain times we are going through, he has been always available to answer my doubts and, in general, to make this path much more enjoyable and interesting.

I am also truly thankful to my family, especially my parents, for supporting and encouraging me during all these years at University.

"The views of space and time which I wish to lay before you have sprung from the soil of experimental physics, and therein lies their strength. They are radical. Henceforth, space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality."

—Hermann Minkowski (1908)

Introduction

The theory of relativity remains as one of the greatest achievements of humankind. It is a complete reformulation of the classic kinematics and gravitational laws, that goes beyond them and ultimately challenges our understanding of reality. While being a physical theory with undisputed experimental validation, its formulation and results hold a striking beauty within them.

The general theory of relativity is rooted on the area of differential geometry, borrowing a lot of mathematical concepts from there and ultimately modelling space and time by bundling them as pseudo-Riemannian manifolds. Thus, in order to comprehend the meaning and formulation of general relativity, it is compulsory to understand a good deal of the underlying geometric objects and their functioning. Our intention in this work is twofold: first, to understand and present the mathematical elements in which general relativity is based; and secondly, to prove and compare, with the help of these mathematical concepts, the powerful Hopf-Rinow and Hawking theorems.

There is a myriad of literature which deals with general relativity, so in this work we intend not only to explain the important results, but to engage the reader in this wondrous endeavour, by illustrating the motivation behind the concepts that will show up. With the purpose of seeing how the theory unfolds naturally from its mathematical description, we present in this work the whole geometric background, that eventually leads to the basic results and applications of general relativity. After that, the final milestone of this work is to present a thorough deduction of the Hawking singularity theorem, and to see its relationship with the Hopf-Rinow theorem, its Riemannian version. The work is structured as follows.

In Chapter 1, we review the fundamental concepts and properties of smooth manifolds, which are the basic structures from which differential geometry is constructed. We will be following [Lee03].

Chapter 2 is the core of the geometric study that is done in this work. We start by adding an extra structure to smooth manifolds, the Riemannian metric, which allows us to relate different points of any smooth manifold to an extent unimaginable in the context of only smooth manifolds. Key concepts which we describe in detail are the ones of affine connections, covariant derivatives, geodesics and curvature, to name the main ones. They will be of invaluable help in the next chapter. Finally, we study the Hopf-Rinow theorem, whose translation into Lorentzian manifolds, as we said, will be seen to be the Hawking theorem. Here, we will mainly follow [GoNa14].

In Chapter 3, we introduce some physical facts and present the crucial concept of Lorentzian manifolds. Them both, in combination with our newly acquired knowledge of differential geometry, will allow us to derive the special and general theories of relativity, ending with the fascinating Einstein field equations.

After that, we can apply this whole framework to two highly relevant physical settings, namely the formulation of cosmology in the Robertson-Walker metric and the Schwarzschild spherical case. This is done in Chapter 4.

Finally, in Chapter 5, we derive Hawking's singularity theorem step by step, in a process that will give insight on the key properties of geodesics in Lorentzian manifolds. Again, in this chapter we will follow [GoNa14].

Chapter 1

Some background in smooth manifolds

If we wish to study the general theory of relativity, we must be really familiar with the nature of the mathematical building blocks it is constructed with. Thus, we consider essential to delve into the nature of the spaces in which general relativity is formulated: smooth manifolds.

First, we shall define what a **topological manifold** M is. It is a topological space which is Hausdorff, second countable and locally Euclidean. The first two properties are not usually of great importance in the study of smooth manifolds. However, the third property is key, and it means that for all $p \in M$ there exist open sets $U \subset M$, $U' \subset \mathbb{R}^n$ such that $p \in U \subset M$ and there exists an homeomorphism $\varphi : U \to U'$. We may say then that the dimension of M is n, and that any pair (U, φ) is a **coordinate chart** on M.

Smooth manifolds are born from topological manifolds, where some extra structure is defined in order to control how the various coordinate charts relate in the points where they intersect. Given two charts $(U, \varphi), (V, \psi)$ for which $U \cap V \neq \emptyset$, we define the **transition map** as $\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V)$, which lives in $\mathbb{R}^n$, so the basic analysis definitions hold. It is said that two charts with non-zero intersection are **smoothly compatible** if the transition map is a diffeomorphism.

Now, if we say that an **atlas** is a set of smoothly compatible charts that covers M, then a **maximal** smooth atlas will be one which is not contained into a strictly larger smooth atlas. It can be proved that any smooth atlas determines a unique maximal smooth atlas, so we can define a **smooth manifold** as a topological manifold which is endowed with a smooth atlas.

Given the smooth manifolds M and M', a map $F : M \to M'$ is said to be a **smooth map** if for all $p \in M$ there exist smooth charts $(U, \varphi), (V, \psi)$ such that $p \in U \subset M$ and $F(U) \subset V \subset M'$ and the map $\psi \circ F \circ \varphi^{-1}$ is smooth as a map between subspaces of $\mathbb{R}^n$. Thus, analogously to the case in $\mathbb{R}^n$, a map $F : M \to M'$ is said to be a **diffeomorphism** if it is a smooth map and it has an inverse map which is also smooth.

Now, a **smooth covering map** π between two connected and locally path connected smooth manifolds M and M' is a surjective smooth map for which, for all $p \in M$, there exists a connected neighborhood U of p such that for every arc-connected component C of $\pi^{-1}(U)$, $\pi|_C$ is a diffeomorphism. An interesting result is that any map which is a proper[1] local diffeomorphism is a smooth covering map.

In order to study the more advanced objects that are the building blocks of general relativity (covector fields, and, more generally, covariant tensors), we shall first understand how vectors and vector fields act on smooth manifolds. At any point $p \in M$, a **tangent vector** at p is defined to be a **derivation** at p, where a derivation is a linear map $\mathcal{X} : \mathcal{C}^\infty(M) \to \mathbb{R}$ that follows the product rule: for all $f, g \in \mathcal{C}^\infty(M)$, $\mathcal{X}(fg) = f(p)\mathcal{X}g + g(p)\mathcal{X}f$. The **tangent space** to M at p, T_pM, is the set of all derivations at p. This seemingly non-intuitive definition for tangent vectors turns out to be really useful, for it is seen that they adopt an expression that is quite simple and, indeed, intuitive.

Let us introduce an object that will play an important role in the understanding of tangent vectors and everything afterwards. Given a smooth map between smooth manifolds $F : M \to N$, the **differential** of F will be the map $d_pF : T_pM \to T_{F(p)}N$ such that for any $\mathcal{X} \in T_pM$, for all $f, g \in \mathcal{C}^\infty(M)$, we have that $d_pF(\mathcal{X})|_f = \mathcal{X}(f \circ F)$. It is well-defined and it is seen that, given any neighborhood U of a point p, the differential of the inclusion $d_pi : T_pU \hookrightarrow T_pM$ is an isomorphism.

This means that tangent spaces can be identified to tangent spaces in $\mathbb{R}^n$ by means of differentials applied to smooth charts. Thus, a tangent vector $\mathcal{X} \in T_pM$ can be expressed as a combination of some basis vectors $\left(\frac{\partial}{\partial x^1}|_p, \ldots, \frac{\partial}{\partial x^n}|_p\right)$ which are in turn related to the chosen local coordinates $(x^1, \ldots, x^n)$. Using this expression for tangent vectors, it is found that the differential of a map F can be explicitly computed on a given chart (U, φ) by obtaining the Jacobian matrix, in its usual form as the derivative in $\mathbb{R}^n$, in terms of the coordinate basis vectors.

[1] A map is said to be **proper** if the preimage of every compact set is compact.

With this definition in mind, one may want to operate not only with the tangent space to a given point, but with all tangent spaces together, or even generalize this to a broader class of vector spaces (which includes tangent vectors as well as other possible definitions of vector spaces attached to a given point). With this aim, a **vector bundle** of rank k over M is defined to be a smooth manifold E equipped with a smooth projection map $\pi : E \to M$ such that for each $p \in M$, $\pi^{-1}(p)$ is a real vector space of dimension k. Also, for every p there is a neighborhood U of p such that $\pi^{-1}(U)$ is diffeomorphic to $U \times \mathbb{R}^k$, and $\pi|_{\pi^{-1}(U)}$ is the standard projection $p_1 : U \times \mathbb{R}^k \to U$. Moreover, a **smooth section** of E is a smooth map $\sigma : M \to E$ such that $\pi \circ \sigma = Id|_M$.

It can be proved that the disjoint union of all the tangent spaces T_pM is indeed a smooth vector bundle, the so-called **tangent bundle** TM, where the projection map is the one that sends each vector to the point it is tangent to. Then, a **smooth vector field** on M is a smooth section X of the projection $\pi : TM \to M$, and it smoothly assigns to each $p \in M$ a vector $X_p \in T_pM$. Vector fields act on smooth functions in the following way: For any $f \in \mathcal{C}^\infty(M)$, a smooth vector field X will induce the real-valued function Xf, defined by $Xf(p) = X_pf$. It can be seen that this operation follows the product rule, and that the set of all vector fields over M is a vector space, $\mathcal{T}(M)$. Finally, for any smooth map $F : M \to N$ and any smooth vector field X over M, the **differential** of X is the differential of every vector of the field. However, it does not necessarily yield a vector field over N, let alone a smooth one. After this fact, it is said that two smooth fields X and Y (over M and N respectively) are **F-related** if the differential of X by F is Y.

Given the algebraic nature of the vector spaces T_pM, we can consider their dual spaces, T_p^*M, which are the **cotangent spaces** at p. This will result in another smooth vector bundle, the so-called **cotangent bundle** T^*M, which, as a set, is the disjoint union of all cotangent spaces T_p^*M, where the projection map is also the natural projection as with vector bundles. A **smooth covector field**, also called **1-form**, is a smooth section of the projection map $\pi : T^*M \to M$, that assigns to each $p \in M$ a covector $\omega_p : T_pM \to \mathbb{R}$.

Taking coordinates and defining some basis of T_pM, the basis for the cotangent space T_p^*M will be the collection $(\lambda_1|_p, \ldots, \lambda_n|_p)$ of the dual counterparts of the tangent space basis. Let us define the **differential covector field** of a function $f \in \mathcal{C}^\infty(M)$ as the one acting on every $X_p \in T_pM$ as $df_p(X_p) = X_pf$. It is seen that, for any coordinate chart (U, x^i), the correspondent cotangent basis vectors will be the differentials of the coordinate functions: $\lambda_p^i = dx_p^i$, for $i = 1, \ldots, n$.

An important definition is the one of the **pullback** F^* of a smooth map $F : M \to N$, which is the dual map of the differential defined before. By the nature of dual maps, it will locally have the form $F^* : T^*_{F(p)}N \to T^*_pM$, where for each $\omega_p \in T^*_{F(p)}N$ its image $F^*\omega$ will act on vectors $\mathcal{X} \in T_pM$ as $(F^*\omega)(\mathcal{X}) = \omega(d_pF(X))$. Contrary to the rather elusive nature of differentials when it is time to express them on coordinates, the pullback of a map $F : M \to N$ will be easily calculated as a function of the coordinates (y^j) of N as[2] $F^*\omega = (\omega_j \circ F)d(y^j \circ F)$. Moreover, an important property of pullbacks is that they are well-defined on 1-forms, i.e. the pullback of a 1-form is also a 1-form, which is not generally the case for differentials applied to vector fields.

Finally, the concept of line integrals, which is understood intuitively in $\mathbb{R}^n$, can now be generalized to smooth manifolds. Given a smooth curve segment[3] γ, where $\gamma : I = [a, b] \to M$, and a smooth covector field ω, the **line integral** of ω over γ is defined as $\int_\gamma \omega = \int_{[a,b]} \gamma^*\omega$. The properties of line integrals over manifolds are quite similar to the ones over $\mathbb{R}^n$, like the fact that for the differential *df* of a smooth function its line integral over any smooth curve will be the difference of the values of f at the endpoints of the curve.

Also in the same way as in $\mathbb{R}^n$, a smooth covector field is said to be **conservative** if its line integral over any closed smooth curve is zero. Another useful characteristics of smooth covector fields are exactness and closedness. A smooth covector field ω is *exact* if it is the differential of some smooth function f, and it is *closed* if its derivative over any smooth coordinates (x^i) fulfils the equation $\frac{\partial \omega_j}{\partial x^i} = \frac{\partial \omega_i}{\partial x^j}$ for all i, j. It is seen that the properties of being conservative and exact are equivalent, and they relate to the closedness just locally. Explicitly, a closed covector field will be a field that is exact in some neighborhood of every point $p \in M$.

The utility of the differential does not stop at the results we have summed up here. If $F : M \to N$ is a smooth map, let us define the **rank** of F at p as the rank, given by linear algebra, of the associated differential. A crucial result, that stems from its counterpart in $\mathbb{R}^n$, is the **rank theorem**, which says that if F has constant rank k on all points, it can take a simple form $F\left(x^1, \dots, x^k, x^{k+1}, \dots, x^n\right) = \left(x^1, \dots, x^k, 0, \dots 0\right)$ after choosing the right coordinates. This, together with the following definitions, will be the basis after which the concept of submanifold can be constructed. We define a smooth map F to be an **immersion** or a **submersion** if the differential

[2] It is expressed in the so-called **Einstein summation convention**, which sums along every two repeated indices that are one up and one down (e.g. $x^iE_i \coloneqq \sum_i x^iE_i$). We will be using this notation throughout this work.

[3] Everything we mention for smooth curve segments can be extended to piecewise smooth curve segments, which are smooth in all but a finite set of points.

d_pF is, respectively, injective or surjective at each point $p \in M$. An immersion that is an homeomorphism onto its image is called a **smooth embedding**.

Now, we define a subset $S \subset M$ to be an **embedded submanifold of dimension k** if the inclusion map $i : S \hookrightarrow M$ is a smooth embedding of rank k. Since this is a rather complex hypothesis to prove, there is a result that simplifies the identification of submanifolds, the **regular level set theorem**. It says that, for any smooth map $F : M \to N$ and given a point $c \in N$, for every set $S \coloneqq F^{-1}(c) \subset M$ such that the differential of F at p is surjective at all points of $F^{-1}(c)$, the set S is an embedded submanifold of M of dimension $k = \dim(M) - \dim(N)$.

The notions for manifolds seen so far are of use in embedded submanifolds too: the restriction of the domain and range of a smooth map to an embedded submanifold is also a smooth map; the tangent space to a point is a vector subspace of the tangent space, $T_pS \subset T_pM$, of dimension k; and if a smooth vector field in M is tangent to T_pS at all points, its restriction to S is a smooth vector field over S (for covector fields it is even more immediate, since they always restrict to a covector field over S).

With everything we briefly described so far, we can define a mathematical object that will be present from now on. A **covariant k-tensor** on a vector space V is a multilinear function $T : V \times \cdots \times V \to \mathbb{R}$, where multilinear means that T is linear on every one of its arguments. Tensors can be thought as a generalization of covectors, and, given a basis $(\varepsilon^1, \ldots, \varepsilon^n)$ of V^*, we can express every tensor as a linear combination $T = T_{i_1,\ldots,i_k}\varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k}$, where $1 \leq i_1, \ldots, i_k \leq n$.

Hence, we can define the **bundle of covariant k-tensors** T^kM on a smooth manifold M as the vector bundle formed by covariant tensors that act on the vector spaces $T_pM \times \cdots \times T_pM$. **Smooth k-tensor fields**, therefore, are smooth sections of the natural projection $\pi : T^kM \to M$, that assign to each $p \in M$ a k-tensor over its tangent space T_pM. Pullbacks on smooth covariant tensor fields behave similarly as with covector fields, and given a map $F : M \to N$, the tensor **pullback** is $F^* : T^k(T_{F(p)}N) \to T^k(T_pM)$, which acts on covariant tensor fields S at any given point as $(F^*S)(\mathcal{X}_1, \ldots, \mathcal{X}_k) = S(dF(\mathcal{X}_1), \ldots, dF(\mathcal{X}_k))$.

After introducing all this necessary background in smooth manifolds, we are now in position to start our analysis of the spaces in which general relativity is formulated: **Riemannian** and **pseudo-Riemannian manifolds**.

Chapter 2

Riemannian Manifolds

2.1 Basic Notions

One may wonder that, if every point has its own tangent space which is not necessarily equivalent to tangent spaces at other points, there is no way in a smooth manifold to compare vectors at different points. Even the idea of measuring the distance between two points is foreign to smooth manifolds as far as we have seen, because the distance is usually defined in terms of the length of a given path. But, which path should we choose between two arbitrary points? How is length defined if the tangent space to a curve varies from point to point? To answer these questions, we need to add a new layer of structure over smooth manifolds, the concept of **metric**.

Definition 2.1. *A* **Riemannian metric** g *on a smooth manifold* M *is a smooth 2-tensor field on* M *which, given any* $p \in M$ *and any pair*[1] $v, w \in T_pM$*, has the following properties:*

(i) It is symmetric: $g(v, w)_p = g(w, v)_p$*;*

(ii) It is positive definite: $g(v, v)_p \geq 0$*, and* $g(v, v)_p = 0$ *if and only if* $v = 0$*.*

A smooth manifold M endowed with a Riemannian metric g is thus labelled as the **Riemannian manifold** (M, g). For the sake of convenience, we will also refer to g at some point p as $\langle \cdot, \cdot \rangle_p$, with $\langle v, w \rangle_p \coloneqq g(v, w)_p$. Now, we may extend the notion of diffeomorphism so that it transforms Riemann manifolds in such a way that their metrics are preserved.

[1] Having departed from the more abstract notions of tangent vectors as derivations $\mathcal{X}$, from now on we will label tangent vectors as usual: v, w, and so on.

Definition 2.2. *An* **isometry** *between two Riemannian manifolds* $(M, g_M), (N, g_N)$ *is a diffeomorphism* $F : M \to N$ *with the property that the pullback of the tensor* g_N *is* g_M*. Explicitly:* $(F^* g_N)|_{F(p)}(v, w) = g_M|_p(v, w)$ *for all* $p \in M$*,* $v, w \in T_pM$*.*

Two Riemannian manifolds for which there exists an isometry that connects them are called **isometric** Riemannian manifolds. Using the tensor formalism, we can write a Riemannian metric g in local coordinates as $g = g_{ij} dx^i \otimes dx^j$, using the Einstein convention.

Let us now define some basic concepts, which start to give answers to the questions we posed in the beginning of this chapter.

Definition 2.3. *The* **length** *of a vector* $v \in T_pM$ *is* $|v|_p \coloneqq \langle v, v \rangle_p^{1/2}$*.*

Definition 2.4. *The* **angle** θ *between two non-zero vectors* $v, w \in T_pM$ *is the angle correspondent to* $\cos\theta = \frac{\langle v,w \rangle_p}{|v|_p |w|_p}$*.*

Definition 2.5. *If we define the tangent vector to a smooth curve* $\gamma : I \to M$ *at a point* t_0 *as* $\dot{\gamma}(t_0) = d_{t_0}\gamma(\frac{d}{dt})$*, the* **length** *of the curve* γ *is* $L_g(\gamma) = \int_a^b |\dot{\gamma}(t)|_g dt$*.*

It can be seen that the length of any curve is independent on its parametrization, so it is a quantity that is intrinsic of the curve. Knowing that, and before we start working on connections, we can give an idea of what the distance between two points is, in a Riemannian manifold.

Definition 2.6. *Given two points* $p, q \in M$*, the* **distance** *between* p *and* q *is defined as* $d_g(p, q) = \inf\{L_g(\gamma) : \gamma$ *smooth curve from* p *to* $q\}$*.*

A remarkable result is that, equipped with this distance, every connected Riemannian manifold is a metric space.

As we drive towards an understanding of general relativity, it is the time now to define a key concept, the one of **pseudo-Riemannian manifolds**. They are smooth manifolds endowed with a smooth symmetric 2-tensor field g which is nondegenerate at every point. This last feature means that if $g(v, w)_p = 0$ for every vector w, then v must be 0, which is a weaker assumption than the one of being positive definite. As we see, Riemannian and pseudo-Riemannian manifolds have a quite similar definition, so they will have some analogous properties.

With this in mind, and before departing from the Riemannian metrics in our future journey through the pseudo-Riemannian Lorentzian metrics, we will remain a bit more within the scope of Riemannian manifolds in order to define and study there some structures that we will deal with later.

2.2 Affine Connections

The definition of distance between two points that we have seen, the infimum length between all curves that join the points, although formally correct, it is still not complete for us. For example, in a general case, we do not still have appropriate tools in order to identify such path of infimum distance, and we cannot compare tangent vectors at different points on M yet. For this purpose, we introduce the concept of **affine connections**.

Definition 2.7. *Given a smooth manifold M, an* **affine connection** *will be a map $\nabla : \mathcal{T}(M) \times \mathcal{T}(M) \to \mathcal{T}(M)$ such that, for all $X, Y, Z \in \mathcal{T}(M)$ and $f, g \in \mathcal{C}^\infty(M)$, the following conditions are fulfilled:*

(i) $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$;

(ii) $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$;

(iii) $\nabla_X (fY) = (Xf)Y + f\nabla_X Y$.

A technical concept which comes in useful in the analysis of Riemannian manifolds is the one of **Christoffel symbols**. In a given coordinate chart (U, x^i), they are the smooth functions $\Gamma^i_{jk} : U \to \mathbb{R}$, which are related to the affine connections of the basis vectors in the following way:

$$\nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} = \sum_{i=1}^{n} \Gamma^i_{jk} \frac{\partial}{\partial x^i}. \tag{2.1}$$

Using the conditions given in (2.7), it can be proved that any connection ∇ will have the following expression, in a given coordinate chart:

$$\nabla_X Y = \sum_{i=1}^{n} \left(XY^i + \sum_{j,k=1}^{n} \Gamma^i_{jk} X^j Y^k \right) \frac{\partial}{\partial x^i}, \tag{2.2}$$

for every two vector fields $X = X^i \frac{\partial}{\partial x^i}$, $Y = Y^i \frac{\partial}{\partial x^i}$ of $\mathcal{T}(M)$. This means, in particular, that the connection is defined locally, i.e. that its value at a point p only depends on the behaviour of X and Y in an arbitrarily small neighborhood of p.

This apparently abstract concept of affine connections will allow us to study the degree in which a vector field is aligned with a given curve, which is closely related to the analysis of paths between points in the manifold. Let us see:

Definition 2.8. *For any smooth curve $\gamma : I \to M$, a* **vector field along** *γ is a smooth map $X : I \to TM$ for which $X(t) \in T_{\gamma(t)}M$ for all $t \in I$.*

If $\dot{\gamma}(t) \neq 0$, and knowing that $\dot{\gamma}(t)$ is trivially a vector field along γ, the **covariant derivative of V along** γ is defined as

$$\frac{DX}{dt}(t) = \nabla_{\dot{\gamma}(t)} X. \tag{2.3}$$

This allows us to define **parallel transport**, and from there the crucial concept of **geodesics**.

Definition 2.9. *A vector field X along γ is* **parallel along** γ *if $\frac{DX}{dt}(t) = 0$ for all $t \in I$. A curve c whose tangent field $\dot{c}$ is parallel along c is called a* **geodesic**.

To distinguish the curves which are geodesics in a given Riemannian manifold M, we will label as c a geodesic of M, and a general curve as γ. Now, by the expression of the affine connection in terms of Christoffel symbols in (2.2), we see that a vector field X locally given by $X(t) = (X^1(t), ..., X^n(t))$ will be parallel along γ if there is a covering of $\gamma(I)$ by coordinate charts such that, on any of them,

$$\dot{X}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk} \dot{x}^j X^k = 0, \tag{2.4}$$

for $i = 1, ..., n$, where we note $\dot{X}^i = \frac{\partial X^i}{\partial t}$. Similarly, a geodesic that is locally given by $c(t) = (x^1(t), ..., x^n(t))$ will fulfil the equations

$$\ddot{x}^i + \sum_{j,k=1}^{n} \Gamma^i_{jk} \dot{x}^j \dot{x}^k = 0. \tag{2.5}$$

Thus, the study of geodesics and parallelism of vectors along a curve is also a study of these ODEs. The theory of differential equations tells us that there is a unique solution for these two sets of equations, given their initial conditions. That is to say, given a starting pair $(p, v) \in TM$, there is a unique geodesic c_v starting at p with initial tangent vector equal to v. In the same manner, given a curve γ with $\gamma(0) = p$, there is also a unique vector field X parallel along γ for which $X(0) = v$.

We now introduce a special case of affine connection, which has some interesting features: the **Levi-Civita connection**. Before that, we shall define two more concepts regarding connections.

Definition 2.10. *A connection ∇ on M is* **symmetric** *if, for all $X, Y \in \mathcal{T}(M)$,*

$$\nabla_X Y - \nabla_Y X = [X, Y], \tag{2.6}$$

which, by the definition of Christoffel symbols, is equivalent to

$$\Gamma^i_{jk} = \Gamma^i_{kj}, \tag{2.7}$$

for all $i, j, k \in \{1, \ldots, n\}$.

Definition 2.11. *Given a Riemannian manifold $(M, \langle\cdot,\cdot\rangle)$, a connection ∇ will be* **compatible** *with $\langle\cdot,\cdot\rangle$ if*

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle, \tag{2.8}$$

for all $X, Y, Z \in \mathcal{T}(M)$.

Above, $X\langle Y, Z\rangle$ refers to the product of the vector field X over the smooth map $\langle\cdot,\cdot\rangle$ applied to the function $\langle Y, Z\rangle$.

Definition 2.12. *We define a* **Levi-Civita connection** *as being a connection ∇ on $(M, \langle\cdot,\cdot\rangle)$ which is symmetric and compatible with $\langle\cdot,\cdot\rangle$.*

This definition, however, does not tell us neither if this connection can be found on an arbitrary manifold, nor which expression would it take. This is addressed in the following important theorem.

Theorem 2.13. (Fundamental Theorem of Riemannian Geometry) *For any Riemannian manifold (M, g), the Levi-Civita connection exists and is unique. The corresponding Christoffel symbols, after taking coordinates (x^i), are*

$$\Gamma^i_{jk} = \frac{1}{2}\sum_{l=1}^{n} g^{il}\left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l}\right), \tag{2.9}$$

with $g^{il} = (g_{il})^{-1}$.

Proof. We define the connection between $X, Y \in \mathcal{T}(M)$ as the one satisfying the **Koszul formula** for every $Z \in \mathcal{T}(M)$:

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle =& X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle - \\ &- \langle [X, Z], Y\rangle - \langle [Y, Z], X\rangle + \langle [X, Y], Z\rangle. \end{aligned} \tag{2.10}$$

This is well-defined, and satisfies the connection axioms defined in (2.7), so it is a connection. It is easily seen that the Koszul formula is constructed in such a way that the connection is symmetric and compatible with g. Now, to observe the form of the Christoffel symbols that is derived from the Koszul formula, we see that

$$\begin{aligned} 2\left\langle \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k}, \frac{\partial}{\partial x^l}\right\rangle &= \frac{\partial}{\partial x^j} g_{kl} + \frac{\partial}{\partial x^k} g_{jl} - \frac{\partial}{\partial x^l} g_{jk} \Leftrightarrow \\ \Leftrightarrow \left\langle \sum_{i=1}^{n} \Gamma^i_{jk} \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^l}\right\rangle &= \frac{1}{2}\left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l}\right) \Leftrightarrow \\ \Leftrightarrow \sum_{i=1}^{n} g_{il}\Gamma^i_{jk} &= \frac{1}{2}\left(\frac{\partial g_{kl}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^l}\right), \end{aligned}$$

which gives us the expression (2.9). □

2.3 The Exponential Map

Now we have seen the first properties of connections, but we are still on a rather abstract level. In order to obtain more tangible results, operating with the so-called **exponential map** is really convenient.

Let ∇ be an affine connection defined in a Riemannian manifold (M, g). Given a pair $(p, v) \in TM$, we know that there exists a unique geodesic $c_v : I \to M$ which is defined in a neighborhood $I_v = (-\epsilon, \epsilon)_v$ of 0, for which $c_v(0) = p$ and $\dot{c}_v(0) = v$. Given some $a \in \mathbb{R}$, we consider the interval $J_v = \left\{ \frac{t}{a} : t \in I_v \right\}$, and define the curve $\gamma_v : J_v \to M$ as $\gamma_v(t) = c_v(at)$.

Since a is a constant number, then $\dot{\gamma}_v(t) = a\dot{c}_v(at)$ and γ_v is also a geodesic. Thus, $\gamma_v(t) \coloneqq c_v(at)$ is the geodesic correspondent to the initial pair $(p, av) \in TM$, so $c_v(at) = c_{av}(t)$. Thanks to this "linear" property of geodesics, we can define the exponential map.

Definition 2.14. *Given some set $U \subset T_pM$ such that $1 \in I_v$ for all $v \in U$, the* **exponential map** *of U is defined as $\exp_p : U \to M$, with $\exp_p(v) = c_v(1)$.*

As we can see in Figure 2.1, the exponential map brings a vector $v \in T_pM$ to the point through which the geodesic c_v is passing at $t = 1$. The following result is easy to see.

Proposition 2.15. *For any $p \in M$, there exists an open neighborhood $U \subset T_pM$ of $\vec{0}$ such that the exponential map of U is a diffeomorphism onto its image, $\exp_p(U) \coloneqq V$. The image V is said to be a* **normal neighborhood** *of p.*

Proof. It is seen in [Pe16] p.172-173 (Lemma 5.2.6 and Theorem 5.2.3, the latter being a result from the theory of ODEs) that there is some open neighborhood $W \subset T_pM$ of $\vec{0}$ where the exponential map is well-defined and smooth. Now, if we see that its differential at the origin is bijective, by the inverse function theorem it follows that there is some open neighborhood U of $\vec{0}$ such that $U \subset W$ and the exponential map is a diffeomorphism in U.

To see this, we use the fact that the differential of a map F can be found as the derivative of the curve $F(\gamma(t))$ at $t = 0$, where γ is any smooth curve such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$. In our case,

$$d\exp_p(v)|_{\vec{0}} = \frac{d}{dt}\exp_p(tv)|_{t=0} = \frac{d}{dt}c_v(t)|_{t=0} = v,$$

so the differential is the identity, and hence the exponential map is indeed a local diffeomorphism at $\vec{0}$. □

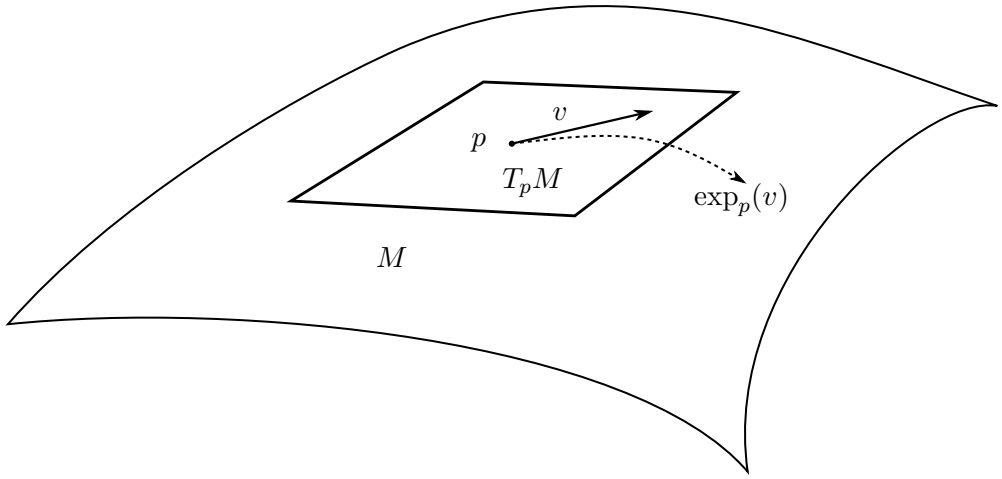

Figure 2.1: Visualization of the action of the exponential map over some $v \in T_pM$.

Example 2.16. The vector field E on $T_pM \setminus \{0\}$ defined by $E_v = \frac{v}{\|v\|}$ for all non-zero vectors of T_pM will have a differential by the exponential as

$$(d_v \exp_p)(E_v) = \frac{d}{dt} \exp_p \left(v + \frac{tv}{\|v\|} \right) \Big|_{t=0} = \frac{\dot{c}_v(1)}{\|v\|}. \tag{2.11}$$

We see that the differential of every vector of the field is the unit tangent vector to its geodesic at $t = 1$ (as $\|\dot{c}_v(t)\| = \|v\|$ for all $t \in I$). This will be useful, for it is a way of parametrizing the tangent field along any geodesic c_v.

Now, we will see that the exponential map allows us to translate the concept of open balls in $\mathbb{R}^n$ to Riemannian manifolds, where the geodesics c_v obtained with the exponential map correspond to the radii of the ball.

Definition 2.17. *Given an $\varepsilon > 0$ for which $\overline{B_\varepsilon(0)} \subset U$, the* **normal ball** *$B_\varepsilon(p)$ centered at p with radius ε is the image by the exponential map of $B_\varepsilon(0) \subset T_pM$, i.e. $B_\varepsilon(p) = \exp(B_\varepsilon(0))$. Analogously, the* **normal sphere** *is $S_\varepsilon(p) = \exp(\partial B_\varepsilon(0))$.*

The first step to see that the geodesics c_v are the equivalent of radii of a ball in $\mathbb{R}^n$ is the following proposition:

Proposition 2.18. *Given any normal sphere S_ε of p, the geodesics c_v starting at p are orthogonal to $S_\varepsilon(p)$.*

Proof. To prove the claim, we see that the vector fields $\dot{c}_v$ are orthogonal to any normal sphere, or what is the same, that $d \exp_p(E)$ is orthogonal to normal spheres.

We parametrize T_pM as $\tilde{\varphi}(r, \theta^i) \coloneqq r\varphi\left(\theta^1, \ldots, \theta^n\right)$, where φ is the standard parametrization of the sphere, r is the radial coordinate and (θ^i) are the angular coordinates of the sphere. We see that the field E is parallel to the radial coordinate and that $\|\frac{\partial}{\partial\theta^i}\| = \|\frac{\partial\tilde{\varphi}}{\partial\theta^i}\| = r\|\frac{\partial\varphi}{\partial\theta^i}\|$, so the angular coordinates vanish in the limit $r \to 0$.

Hence, the field $X \coloneqq d\exp_p(E)$ is also tangent to the radial directions $\frac{\partial}{\partial r}$, and the differentials Y_i of the vectors $\frac{\partial}{\partial\theta^i}$ are the angular coordinates of the normal spheres, which also approach 0 with $r \to 0$.

By the naturality of Lie brackets ([Lee03], p.92), we have $[X, Y_i] = d\exp_p\left(\left[\frac{\partial}{\partial r}, \frac{\partial}{\partial\theta^i}\right]\right)$, which is 0 by construction. By the properties of affine connections, and because the norm of X is constant and equal to 1, it is seen that $\langle X, Y_i\rangle$ is constant over any geodesic c_v:

$$X\langle X, Y_i\rangle = \langle X, \nabla_{Y_i}X\rangle = \frac{1}{2}Y_i\langle X, X\rangle = 0.$$

Finally, since Y_i vanish when $r \to 0$, this means that

$$\langle X, Y_i\rangle(\exp_p v) = \lim_{t\to 0}\left(\langle X, Y_i\rangle(\exp_p(tv)\right) = 0.$$

so we have indeed that X is orthogonal to the sphere $S_\varepsilon(p)$ for every ε. □

There is another property that the radius of a sphere $S_p(\varepsilon)$ has to fulfil, which is being the curve that minimizes the distance from p to any $q \in S_p(\varepsilon)$. This is precisely the object of the following proposition.

Proposition 2.19. *For any smooth curve $\gamma : [0, 1] \to M$ such that $\gamma(0) = p$ and $\gamma(1) \in S_\varepsilon(p)$, we will always have that $L(\gamma) \geq \varepsilon$. Furthermore, $L(\gamma) = \varepsilon$ if and only if γ is a reparametrization of a geodesic c_v.*

Proof. Safely imposing that $\gamma(t) = p$ only at $t = 0$, and that $\gamma([0, 1]) \subset B_\varepsilon(p)$, we can parametrize γ as the differential of the vector $r(t)n(t)$. Analogously as before, $r(t)$ and $n(t)$ are, respectively, the radial and angular components of the correspondent vector in T_pM. By the product rule,

$$\dot{\gamma}(t) = d\exp_p(\dot{r}n(t) + r(t)\dot{n}(t)),$$

where, since $n(t)$ are angular components of the sphere, we have that $n(t)$ is normal to $S_{r(t)}$ and $\dot{n}(t)$ is tangent to $S_{r(t)}$. With that, we can express $\dot{\gamma}(t)$ as a function of the radial and angular fields $X \coloneqq d\exp_p(\frac{\partial}{\partial r})$ and $Y \coloneqq r(t)d\exp_p(\dot{n}(t))$:

$$\dot{\gamma}(t) = \dot{r}(t)X_{\gamma(t)} + Y(t).$$

Since these fields are orthogonal by construction, the length function is easily seen to satisfy

$$L(\gamma) = \int_0^1 \langle \dot{r}(t)X_{\gamma(t)} + Y(t), \dot{r}(t)X_{\gamma(t)} + Y(t)\rangle^{\frac{1}{2}} dt$$
$$= \int_0^1 \left(\dot{r}(t)^2 + \|Y(t)\|^2\right)^{\frac{1}{2}} dt \geq \int_0^1 \dot{r}(t)dt = r(1) - r(0) = \varepsilon,$$

where the equality is obtained only when $Y(t) = \dot{n}(t) = 0$ for all $t \in [0,1]$, and $r(t)$ is a monotone function of t, i.e. when γ is a geodesic. □

This allows us to introduce a particularly interesting parametrization: the so-called **normal coordinates**. They can be defined on any normal neighborhood of a point p as $\varphi : U \subset \mathbb{R}^n \to M$, where φ acts on any n-tuple of $\mathbb{R}^n$ like

$$\varphi(x^1, \ldots, x^n) = \exp_p(x^1 v_1 + \cdots + x^n v_n).$$

These coordinates have some conceptually important properties, as:

(i) The Christoffel symbols are zero at p, $\Gamma^i_{jk}(p) = 0$. That means, for example, that the coordinate representation of the geodesic c_v for a given $v \in T_pM$, $v = (v^1, \ldots, v^n)$, is simply $(tv^1, \ldots, tv^n)$.

(ii) Given an orthonormal basis of T_pM, the Riemannian metric tensor at p is the identity, i.e $g_{ij}|_p = \delta_{ij}$.

These coordinates are constructed with the exponential map, which is in turn closely related to the behaviour of geodesics. Because of it, the parametrization of curves in this way is a measure of the length of the curve itself, so we can say that normal coordinates are the generalization to Riemannian manifolds of the arc length parametrization.

Now that we have seen how to find a neighborhood in which the geodesics are indeed the measure of distances, we wish to extend this to the whole manifold (M, g). We will prove the intuitive fact that the curves with minimum length between any two points, and hence the measure of distance between them, are given by geodesics. Before that, we will define a property that normal neighborhoods can have, which is the one of being **totally normal**.

Definition 2.20. *A* **totally normal neighborhood** $V \subset M$ *of* p *is a normal neighborhood such that there exists some* $\varepsilon > 0$ *for which* $V \subset B_\varepsilon(q)$ *for all* $q \in V$.

Lemma 2.21. *For every $p \in M$, there exists a totally normal neighborhood V such that $p \in V$.*

Proof. For every $(p, v) = (x^1, \dots, x^n, v^1, \dots, v^n) \in TM$, the associated geodesic starting at (p, v) can also be expressed as, for $i = 1, \dots, n$,

$$\begin{cases} \dot{x}^i = v^i \\ \dot{v}^i = -\sum_{j,k=1}^n \Gamma^i_{jk} \dot{x}^j \dot{x}^k \end{cases}$$

We can then express the solution of this equation as the vector field $X = ((\dot{x})^i, (\dot{v})^i)$,

$$X = \sum_{i=0}^n \dot{x}^i + \sum_{i=0}^n \dot{v}^i = \sum_{i=0}^n v^i - \sum_{i,j,k=0}^n \Gamma^i_{jk} \dot{x}^j \dot{x}^k.$$

This is the so-called **geodesic flow**, which in general is not globally defined. However, we can always define (see [GoNa14], p.30) a local restriction to $W \times I$, where $W \subset TM$ is a neighborhood of p for which all vectors fulfil $\|v\| < \varepsilon$, in which the local field is indeed well-defined. The number $\varepsilon > 0$ can be arbitrarily small, so the open interval $I \subset \mathbb{R}$ can be arbitrarily large.

We define a map $G : W \to M \times M$ as $G(q, v) \coloneqq (q, \exp_q(v))$, lowering ε if necessary. Analogously as in Proposition 2.15 we see that G is a diffeomorphism locally, so, shrinking W if necessary, we can assume that G is a diffeomorphism onto its image. Let V be an open neighborhood of p such that $V \times V \subset G(W)$. We see that, for any point $q \in V$, we have that $\{q\} \times \exp_q(B_\varepsilon(0))$ is exactly the subset of $G(W)$ for which the first component is q. Therefore, we finally obtain that $V \times V \subset G(W)$, so $V \subset \exp_q(B_\varepsilon(0))$ and hence V is a totally normal neighborhood of the point p. □

Having proved this technical lemma, we can finally observe that, as we were foreseeing during this chapter, geodesics are the paths between points that minimize the length function.

Theorem 2.22. *If $c : I \to M$ is a piecewise smooth curve between any two points $p, q \in M$ such that its length is minimal among all curves that connect p and q, then c is a reparametrization of a geodesic.*

Proof. The curve c can be divided in smooth segments for which the endpoints belong to a shared totally normal neighborhood U. By Proposition 2.19, in every one of these neighborhoods $c|_U$ will be a reparametrization of a geodesic, which extends to the fact that the whole curve c is indeed a reparametrization of a geodesic. □

Along this chapter, we have seen how closely related are the concepts of geodesics and distance. This is brought to another level with the crucial **Hopf-Rinow theorem**, which deserves to be a section on its own.

2.4 Hopf-Rinow Theorem

As we have been advancing in the knowledge of the nature of geodesics, every new result has pointed us to the fact that geodesics in (M, g) are the curves that determine the distance between any two points. Willing to relate geodesics with the structure of the metric space induced by g, we will look into the Hopf-Rinow theorem. Before, let us write some preliminary definitions.

Definition 2.23. *A Riemannian manifold (M, g) is* **geodesically complete** *if the exponential map is defined in all of T_pM, for every $p \in M$.*

We said at the beginning of the chapter that any connected Riemannian manifold is a metric space, endowed with the distance defined in 2.6. An additional property that can hold in metric spaces is the following.

Definition 2.24. *A metric space (M, d) is said to be* **complete** *if every Cauchy sequence is convergent. Remember that by definition a* **Cauchy sequence** *$(x_n)_{n\in\mathbb{N}}$ satisfies that, for any $\varepsilon > 0$, there exists some $n_0 \in \mathbb{N}$ such that $d(x_n, x_m) < \varepsilon$ for every $n, m > n_0$.*

One may suspect, since we have used the same word for both geodesics and metric spaces, that for a Riemannian manifold it is equivalent to be **complete** in terms of geodesics and as a metric space. This is addressed in the Hopf-Rinow theorem.

We have seen that geodesics between two points, if they exist, are the curves minimizing distance. We shall see in the following proposition, that will be central in the proof of the Hopf-Rinow theorem, that the property of being geodesically complete allows us to affirm that such geodesics always exist.

Proposition 2.25. *If (M, g) is a connected Riemannian manifold which is geodesically complete, it holds that for any two points $p, q \in M$ there exists a geodesic joining them with length equal to $d(p, q)$.*

Proof. We denote the distance between p and q as $d(p, q) =: r > 0$. Let us consider some $\varepsilon \in (0, r)$ for which $S_\varepsilon(p)$ is a normal sphere centered at p. Since $S_\varepsilon(p)$ is a compact submanifold of M and the distance function $x \mapsto d(x, q)$ is continuous, it will have a minimum value $x_0 \in S_\varepsilon(p)$. This x_0 corresponds to some vector $v \in T_pM$, and we can express it as $x_0 = \exp_p(\varepsilon v)$. Let us see that this is actually the wanted geodesic between p and q, i.e. that $q = \exp_p(rv) = c_v(r)$.

If we define the set A as

$$A = \{\, t \in [0, r] \;:\; d(c_v(t), q) = r - t \,\}, \tag{2.12}$$

then A is non-empty, since $0 \in A$, and closed (for the map $\alpha : t \mapsto d(c_v(t), q)$ is continuous and $A = \alpha^{-1}([0, r])$). To prove the proposition, we have to show that $r \in A$.

If $r \notin A$, then there would be some maximum $\tilde{t} < r$ of A, with $\tilde{x} \coloneqq c_v(\tilde{t})$. We can find some normal sphere centered in $\tilde{x}$, with radius $\delta \in (0, r)$. As before, there exists some point x_1 in this $S_\delta(\tilde{x})$ that minimizes the function $x \mapsto d(x, q)$. Let us see that the geodesic between $\tilde{x}$ and x_1 is the "continuation" of our geodesic c_v, i.e. $x_1 = c_v(\tilde{t} + \delta)$, and hence $\tilde{t}$ was not the maximum of A.

As $\tilde{x} \in A$, we have that $d(\tilde{x}, q) = r - \tilde{t}$. Since $S_\delta(\tilde{x})$ is a normal sphere, we also have that $d(\tilde{x}, q) = \delta + \min_{S_\delta(\tilde{x})} d(x, q) = \delta + d(x_1, q)$, so

$$d(x_1, q) = r - \tilde{t} - \delta. \tag{2.13}$$

We have that $d(p, q) - d(x_1, q) = r - (r - \tilde{t} - \delta) = \tilde{t} + \delta$, and as (M, d) fulfils the triangle inequality, then

$$d(p, q) \leq d(p, x_1) + d(x_1, q) \implies d(p, x_1) \geq \tilde{t} + \delta. \tag{2.14}$$

The piecewise smooth curve constructed as $c_v(t)$ until $\tilde{t}$, and then as the geodesic between $\tilde{x}$ and x_1, has length equal to $\tilde{t} + \delta$. Then, since it minimizes the distance between p and x_1, it is a geodesic, with $x_1 = c_v(\tilde{t} + \delta)$. Finally, we have that

$$d(c_v(\tilde{t} + \delta), q) = r - (\tilde{t} + \delta), \tag{2.15}$$

so $\tilde{t} + \delta \in A$, and it is proved that $\tilde{t}$ is not a maximal element of A. □

We are now in position to prove the theorem. Let us see:

Theorem 2.26. (Hopf-Rinow Theorem) *For a Riemannian manifold (M, g), the following statements are equivalent:*

(i) M is geodesically complete.

(ii) The induced metric space (M, d) is complete.

Proof. If M is geodesically complete, we will see that for any $p \in M$, every closed and bounded set $K \subset M$ containing p is compact, from where it is easy to see then that (M, d) is complete. First, if K is bounded, there is some $R > 0$ such that it fits inside $B_R(p) \coloneqq \{q \in M : d(p, q) < R\}$. By Proposition 2.25, any $q \in B_R(p)$ is connected to p by some geodesic, which will have length shorter than R. Hence, $B_R(p) \subset \exp_p\left(\overline{B_R(0)}\right)$, where $\exp_p\left(\overline{B_R(0)}\right)$ is compact for being a continuous image of the compact set $\overline{B_R(0)}$.

Finally, K is compact since it is a closed subset of a compact set. It follows that the closure of any Cauchy sequence in M is compact, for it being closed and bounded, so every Cauchy sequence has a convergent subsequence. Since this applies to any such sequence, we have that, indeed, every Cauchy sequence converges, so (M, d) is complete.

Conversely, let us assume that (M, d) is a complete metric space and let us consider some normalized geodesic c defined only for all $t < \tilde{t}$, where by normalized we mean that $||\dot{c}(t)|| \equiv 1$. By this property, given any sequence $\{t_n\}$, we will have that $d(c(t_n), c(t_m)) \leq |t_n - t_m|$. Therefore, if we take any sequence $\{t_n\}$ converging to $\tilde{t}$, the corresponding sequence $\{c(t_n)\}$ will be a Cauchy sequence, and hence it will converge to $p \coloneqq \lim_{t \to \tilde{t}} c(t)$. By the result seen in [One83], p.130 (8. Lemma), we just proved that c can be extended *ad infinitum*, hence proving the completeness of geodesics. □

This theorem is of great importance, and not only because it characterizes the relationship between geodesics and distances. We will see that the Hawking singularity theorem, which we aim to understand in this work, can be seen as the translation of the Hopf-Rinow theorem into pseudo-Riemannian manifolds.

As we mentioned before, although we wish to jump into pseudo-Riemannian manifolds, into which general relativity is formulated, we are considering first the case of Riemannian manifolds. We have several reasons to proceed in this way. First, since there are a lot of notions that are shared by them both, it is harmless (and very pedagogical!) to work first with Riemannian manifolds, and then make the according modifications for the pseudo-Riemannian case.

Also, there is the fact that Riemannian manifolds can be understood quite intuitively, and the results seem to be "logical" from our experience. For instance, the fact that the distance between two points corresponds to some minimizing geodesic (a straight line in the case of $\mathbb{R}^n$) seems not difficult to grasp. However, as we will see, this gets a bit trickier when we deal with pseudo-Riemannian metrics.

The first part of this chapter, dealing with fundamental definitions and connections, can safely remain untouched when we deal with pseudo-Riemannian metrics, for the positive definiteness is not a necessary condition there (non-degeneracy suffices). Nevertheless, everything concerning distance may have to be reformulated, or at worst, abandoned, in pseudo-Riemannian manifolds. Before getting there, though, we shall introduce a crucial concept in both types of manifolds, which is none other than the **curvature**.

2.5 Curvature

Curvature has been a subject of study since the origins of mankind, or at least since the origins of mathematics. From the Greeks until Gauss, there had been a consistent effort in order to parametrize the behaviour of curves and surfaces. But, despite it being an interesting field on its own, there was still something missing in order to incorporate the notion of curvature into general manifolds, not necessarily inhabiting real Euclidean spaces. As we have seen for other concepts, we need to introduce it as a theoretical construct, which, although it may seem too abstract at first, it will make sense as we advance.

Definition 2.27. *Given a manifold M and some connection ∇ defined on M, the* **curvature** R^∇ *is an operator that assigns to every two vector fields $X, Y \in \mathcal{T}(M)$ the map $R^\nabla_{X,Y} : \mathcal{T}(M) \to \mathcal{T}(M)$, defined as*

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

We may notate it equivalently as $R^\nabla_{X,Y}(Z)$ or $R^\nabla(X,Y)Z$, depending on the complexity of the expression for the fields X, Y, Z.

Now, this can be seen as a map that, in every point $p \in M$, takes three vectors $X_p, Y_p, Z_p \in T_pM$ and outputs another vector, namely $R^\nabla_{X_p,Y_p}(Z_p)$. Hence, if it is seen to be multilinear in X,Y and Z, the curvature will be a smooth $(1,3)$-tensor field. Indeed, it is easy to see, using the properties of connections, that, for any smooth fields and maps,

(i) $R^\nabla(fX_1 + gX_2, Y)Z = fR^\nabla(X_1, Y)Z + gR^\nabla(X_2, Y)Z$,

(ii) $R^\nabla(X, fY_1 + gY_2)Z = fR^\nabla(X, Y_1)Z + gR^\nabla(X, Y_2)Z$,

(iii) $R^\nabla(X,Y)(fZ_1 + gZ_2) = fR^\nabla(X,Y)Z_1 + gR^\nabla(X,Y)Z_2$.

This tensor field is called the **Riemann tensor**, and it will play a central role in general relativity. Let us see now how its tensor coordinates $R^{\nabla l}_{ijk}$ look like, where

$$R^\nabla = \sum_{i,j,k,l=1}^{n} R^{\nabla l}_{ijk} dx^i \otimes dx^j \otimes dx^k \otimes \frac{\partial}{\partial x^l}.$$

Since a tensor can be fully defined from its action on basis vectors, then the following result allows us to determine R^∇.

Proposition 2.28. *For some chart $(U, (x^i))$, if the Christoffel symbols of the connection ∇ are Γ^i_{jk}, then the Riemann tensor coordinates are*

$$R^{\nabla l}_{ijk} = \frac{\partial \Gamma^l_{jk}}{\partial x^i} - \frac{\partial \Gamma^l_{ik}}{\partial x^j} + \sum_{m=1}^{n} \Gamma^m_{jk}\Gamma^l_{im} - \sum_{m=1}^{n} \Gamma^m_{ik}\Gamma^l_{jm}. \tag{2.16}$$

Proof. Let us see how R^∇ looks applied to basis vectors $\frac{\partial}{\partial x^i}$, where $\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0$.

$$\begin{aligned} R^\nabla\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^k} &= \nabla_{\frac{\partial}{\partial x^i}} \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} - \nabla_{\frac{\partial}{\partial x^j}} \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^k} \\ &= \nabla_{\frac{\partial}{\partial x^i}} \left(\sum_{m=1}^{n} \Gamma^m_{jk} \frac{\partial}{\partial x^m}\right) - \nabla_{\frac{\partial}{\partial x^j}} \left(\sum_{m=1}^{n} \Gamma^m_{ik} \frac{\partial}{\partial x^m}\right) \end{aligned}$$

Operating and reordering terms, we get that it is equal to

$$\begin{aligned} &\sum_{m=1}^{n} \left(\frac{\partial}{\partial x^i}\Gamma^m_{jk} - \frac{\partial}{\partial x^j}\Gamma^m_{ik}\right)\frac{\partial}{\partial x^m} + \sum_{m,l=1}^{n} \left(\Gamma^m_{jk}\Gamma^l_{im} - \Gamma^m_{ik}\Gamma^l_{jm}\right)\frac{\partial}{\partial x^l} \\ =&\sum_{l=1}^{m} \left(\frac{\partial \Gamma^l_{jk}}{\partial x^i} - \frac{\partial \Gamma^l_{ik}}{\partial x^j} + \sum_{m=1}^{n} \left(\Gamma^m_{jk}\Gamma^l_{im} - \Gamma^m_{ik}\Gamma^l_{jm}\right)\right)\frac{\partial}{\partial x^l} \end{aligned}$$

which is the result we wanted to obtain. □

Note that we have not introduced the concept of metric into the curvature yet. Theorem 2.13 still holds for pseudo-Riemannian manifolds, so there is a unique Levi-Civita connection ∇ which is well-defined for any pseudo-Riemannian manifold. Hereafter, we consider some Riemannian or pseudo-Riemannian manifold (M, g) and the correspondent Levi-Civita connection ∇, together with our new acquaintance, its Riemann tensor $R \coloneqq R^\nabla$.

To explain qualitatively how the Riemann tensor works, let us consider parallel transport of a vector through a closed curve. Trivially, in Euclidean spaces the vector at the endpoints of the curve is the same, but it is not always like this for other manifolds. The curvature of the manifold alters the parallel transport of vectors, and the Riemann tensor is a way to measure this displacement and hence the curvature itself. This is why it is formulated in terms of the Levi-Civita connection, because it is the manner in which we analyse parallel transport of vectors.

However, working with a $(1, 3)$ tensor is rather cumbersome, for it is a mixed tensor and both our understanding of its behaviour and operating with it can get quite involved. This is the reason why we would rather **lower**[2] the first index so that we

[2]The operation of lowering or raising tensor indices, which is not trivial, can be done considering the so-called **musical isomorphisms** (see e.g. [Lee18], p.26).

are left with a $(0,4)$ Riemann tensor[3], which acts on smooth vector fields as

$$R(X,Y,Z,W) = g(R(X,Y)Z,W). \tag{2.17}$$

Looking at equation (2.16), and knowing that the Levi-Civita connection is symmetric, we guess that the Riemann tensor should have some symmetries over its arguments.

Proposition 2.29. *For any $X,Y,Z,W \in \mathcal{T}(M)$, the Riemann tensor R has the following properties:*

(i) $R(X,Y,Z,W) = -R(Y,X,Z,W)$,

(ii) $R(X,Y,Z,W) = -R(X,Y,W,Z)$,

(iii) $R(X,Y,Z,W) = R(Z,W,X,Y)$,

(iv) $R(X,Y,Z,W) + R(Y,Z,X,W) + R(Z,X,Y,W) = 0$ **(Bianchi identity)**.

Proof. Proving these properties is a matter of operating with equation (2.16), assuming that ∇ is symmetric and compatible with the metric. We omit the details, because they hold little relevance for our purpose in this work. For an explicit proof, see for example [GoNa14], p.125-127. □

A concept that will come in useful in order to define an appropriate metric for cosmology applications in Chapter 4, is the one of **sectional curvature**.

Definition 2.30. *For any 2-dimensional subspace $\Pi \subset T_pM$, for any two non-proportional vectors $X_p, Y_p \in \Pi$, their* **sectional curvature** *K is defined as*

$$K(X_p,Y_p) = -\frac{R(X_p,Y_p,X_p,Y_p)}{\|X_p\|^2\|Y_p\|^2 - \langle X_p,Y_p\rangle^2}.$$

This can be seen to be independent of the choice of coordinates and vectors, and equivalent to the Riemann tensor when considered over all possible subspaces. Now, although we have made some steps to make the Riemann tensor more feasible to work with, we can still simplify it more by the means of **contraction**, which we define loosely as follows.

Definition 2.31. *Given a tensor $T^{j_1,\dots,j_r}_{i_1,\dots,i_s}$ of rank (r,s), a* **metric contraction** *$T^{j_1,\dots,j_{r-1}}_{i_1,\dots,i_{s-1}}$ consists in choosing two indices $i_k, j_l =: m$ and summing over them.*

$$T^{j_1,\dots,j_{r-1}}_{i_1,\dots,i_{s-1}} = \sum_{m=0}^{n} T^{j_1,\dots,m,\dots,j_r}_{i_1,\dots,m,\dots,i_s} \tag{2.18}$$

[3]It is sometimes labelled as curvature tensor. For the sake of simplicity, we stick with the first denomination.

We shall note that this can be extended to the **covariant** and **contravariant** case, where we would respectively take two indices i_k, i_l and j_k, j_l. To give an easy example, the contraction of a $(2,0)$ tensor, i.e. a matrix, is its trace. Indeed, given a matrix (a^{ij}) with dimension n, the only possible contraction is the contravariant contraction $i, j =: m$, that results in $\mathrm{tr}(a) = \sum_{m=0}^{n} a^{mm}$.

It can be seen that this contraction operation does not change the tensor basic properties, and that the covariant derivative commutes with it (see [One83], p.83). We may wonder if we could use this to simplify the Riemann tensor to an extent that it still retains its main structure but is instead a $(0,2)$ tensor. As we will see later, this works, and the tensor is named after the inventor of tensor calculus, Gregorio Ricci-Curbastro.

Definition 2.32. *The* **Ricci curvature**, **Ric**, *is the following contraction of the Riemann tensor* R^i_{jlk}*:*

$$\mathrm{Ric}_{ij} = \sum_{k=1}^{n} R^k_{kij}$$

It can be easily seen, by applying the symmetry conditions of R given in Proposition 2.29, that the Ricci tensor is symmetric, so given any $X, Y \in \mathcal{M}$, we will have $\mathrm{Ric}_{ij}(X,Y) = \mathrm{Ric}_{ij}(Y,X)$.

Going even further, we could wonder what would happen if we took one more step, i.e. if contracting the Ricci tensor yields any significant quantity.

Definition 2.33. *We define the* **scalar curvature** S *as the trace of the Ricci tensor:*

$$S = \sum_{i=1}^{n} \mathrm{Ric}_{ii}$$

It turns out that, indeed, both the Ricci and scalar curvature are key objects in general relativity. Time will come when we remember them again, in the unveiling of Einstein field equations. Now that we have the main mathematical ingredients, let us introduce the physical context, to put ourselves in the situation in which Einstein was in 1905.

Chapter 3

The Theory of Relativity

3.1 Special Relativity

Einstein's theory of relativity was a pivotal accomplishment in our understanding of space, time, and the relation between them, which still puzzles us a century later. But, to begin with, what do we understand by space and time? Is space the scenario in which some processes occur and evolve through time? Can we still cling to some construction which we can label as "absolute space and time", some special point of view in which we formulate the fundamental laws of physics? Let us look first into how can we give sense mathematically to the physical concepts of space and time.

Definition 3.1. *A* **reference frame** S *is a choice of space and time coordinates. We call it an* **inertial reference frame** *if every free particle (one which is not subject to any force) either stays at rest or moves at constant speed with respect to this system.*

It follows that, given an inertial frame, another reference frame will be inertial if their relative velocity $\vec{v}$ is constant, i.e. if they are not accelerated with respect to each other.

Before Einstein, the established model was the one inherited from the classical ideas of Newton and Galileo. In it, each reference frame is the assignment of the real space $\mathbb{R}^3$, with some basis (e_1, e_2, e_3), to some trajectory along the time coordinate, $t \in \mathbb{R}$. All phenomena happening in some inertial frame can be translated to any other inertial frame by the means of **Galilean transformations**. In the simple case where the two frames S, S' are such that S' is going away from S in the x direction at positive speed v, the coordinates of any point in S' will relate to the ones in S by

$$x' = x - vt; \quad y' = y; \quad z' = z; \quad t' = t.$$

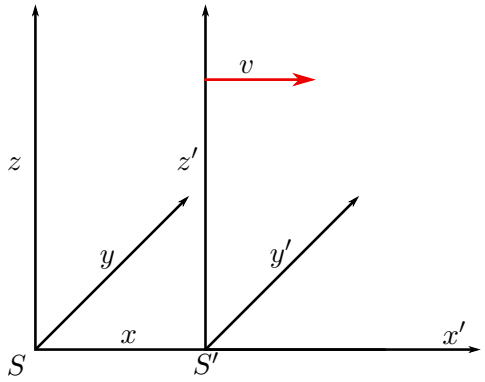

Figure 3.1: An inertial frame S' departing from S at constant speed $\vec{v} = (v, 0, 0)$.

Here, it is taken as an axiom that time is an absolute static quantity, which is the same on every frame. As is usually the case, this axiom does not come from some observational proof but more from a intuitive or philosophical statement taken as truth. The relationship between coordinates and relative velocity is also intuitively seen as true: If I am driving at 100 km/h and a car passes me at 120 km/h, then I see this car (and all of its reference frame) going away from me at 20 km/h, in the direction we are driving in.

This approach, however, has a huge pitfall. If we were to apply this Galilean transformation to a ray of light, then from different reference frames we could see light going away from us at different speeds. However, it was seen in Maxwell's theory of electromagnetism (and backed up experimentally in the Michelson-Morley experiment) that the speed of light is an absolute constant, c, which is independent of the reference frame. This deemed the classical viewpoint incomplete, or even plain incorrect, until Einstein came up with a solution. It is said that one man's trash is another man's treasure, so if this invariance of c brings problems, let us assume it as a hypothesis.

3.1.1 Lorentz Transformations

Einstein's theory of special relativity starts its reasoning with two postulates:

1. All the laws of physics are equivalent in all inertial reference frames.

2. The speed of light, c, is the same in every inertial reference frame.

We saw that Galilean transformations fail to satisfy Postulate 2, so the transformation laws between inertial reference frames have to come in the form of the **Lorentz transformations**. In these transformations, the quantities of space and time are

entangled in the quantity $X^\mu := (ct, x, y, z)$, which determines completely any trajectory in a given reference frame. Assuming the postulates, again for frames S, S' where S' moves at $\vec{v} = (v, 0, 0)$ respect[1] to S, it can be seen that the transformation law for a point X^μ into S' will be:

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \tag{3.1}$$

where we label $\beta = \frac{v}{c}$ and $\gamma = \frac{1}{\sqrt{1-\beta^2}}$. As a proof for this result, we can refer to the reasoning of Einstein himself in [Ein16], p.139.

As we see, the concepts of time and space are intertwined, and are unique in every inertial frame. If one does some simple calculations, it is seen that ct' and x' respectively grow and decrease with v, reaching asymptotic values when $v \to c$. These are well-documented consequences of the postulates, known as **time dilation** and **length contraction**. Note too that, by just assuming that c is constant in every inertial frame, we have seen that in fact the speed of light c is unsurpassable.

Given all of this, it is obvious that we need to substitute the Galilean conception of space and time as $\mathbb{R} \times \mathbb{R}^3$ by some alternative formulation, and this is exactly where pseudo-Riemannian metrics enter the conversation. For this, we should abandon the idea that the space in which we work is "intuitive", for we have seen that the most "intuitive" approach was not capable of handling the invariance of c.

Let us consider the space of the elements of the form $X^\mu = (ct, x, y, z)$, which is obviously $\mathbb{R}^4$ when considered purely as a smooth manifold, with the identification $T_{X^\mu}\mathbb{R}^4 \cong \mathbb{R}^4$ for every X^μ. We want to endow it with some metric which, as said in the first postulate, is invariant under change of frame, i.e. invariant under Lorentz transformations.

If one looks at the expression (3.1) and considers some arbitrarily small trajectory $(ct, x, y, z) \to (ct + \Delta(ct), x + \Delta x, y + \Delta y, z + \Delta z)$, then for any two inertial frames S, S' it holds that

$$-(\Delta ct')^2 + (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2 = -(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2. \tag{3.2}$$

This quantity, which is invariant over change of frame, is called the **action**, labelled as s^2.

[1] Of course it applies to any possible direction, not just the direction $\hat{x}$. We base our calculations in this chapter in that S' departs from S at $\vec{v} = v\hat{x}$.

The invariance of s^2 seen in (3.2) can be extended to differential displacements, where

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2. \tag{3.3}$$

This expression is familiar to us, for it being analogous to the definition of a metric tensor. In fact, if we define the **Minkowski metric** as the pseudo-Riemannian metric over $\mathbb{R}^4$ defined by equation (3.3), we have the desired metric. Using the so-called natural units, where we assign $c = 1$, then the matrix expression of the metric is, in the canonical standard basis of $\mathbb{R}^4$,

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where $\mu, \nu = 0, ..., 3$. We have not specified the point $p \in \mathbb{R}^4$ in whose tangent space we work, because it is identical in all of $\mathbb{R}^4$. Thus, given any two vectors $v, w \in T_p\mathbb{R}^4$ at some point p, their inner product by the Minkowski metric will be $\eta(v, w) = \eta_{\mu\nu}v^\mu w^\nu$. Applying the results seen so far to various physical situations where some object is travelling at speeds closer to c yields bewildering and counter-intuitive results[2] that are indeed correct, such is the power of this reformulation of space and time. Or, should we say, of space-time.

3.2 Lorentzian manifolds

Now we have a metric which is invariant under Lorentz transformations, and hence faithful to the postulates of special relativity. However, thus far we have been studying phenomena occurring in Riemannian manifolds, but we see that we are not in this case any more: we have to introduce the concept of **Lorentzian manifolds**.

Definition 3.2. *Given a pseudo-Riemannian metric tensor g_p in a manifold M, the* **signature** *$\nu = (p, n, z) \in \mathbb{N}^3$ of g_p is the array defined by:*

- $p = dim\{v \in T_pM : g(v, v) > 0\}$,
- $n = dim\{v \in T_pM : g(v, v) < 0\}$,
- $z = dim\{v \in T_pM : g(v, v) = 0\}$.

In terms of matrix representation, by simple linear algebra, it reduces to the number of positive, negative and zero eigenvalues of the matrix expression of g_p. It can be seen, too, that the signature is the same all over the manifold M.

[2]The interested reader may appreciate, for example, the thorough analysis of the so-called Twin paradox found in [Be12].

Example 3.3. The Minkowski metric has signature $\nu = (3, 1, 0)$, and we denote $\mathbb{R}^4$ equipped with the Minkowski metric as $\mathbb{R}^3_1$.

Definition 3.4. *A* **Lorentzian manifold** *is a pseudo-Riemannian manifold of dimension $n \geq 2$ such that its signature is $\nu = (n-1, 1, 0)$. A 4-dimensional Lorentzian manifold is called a* **space-time**.

As we mentioned earlier, the concepts and basic properties of connections, geodesics and curvature for Lorentzian manifolds (which are a specific case of pseudo-Riemannian manifolds) are in general the same as seen in Chapter 2. It is everything related to distance that falls apart under the light of pseudo-Riemannian metrics, because distance is a concept essentially defined for positive definite metrics.

For example, in the case of $\mathbb{R}^3_1$ it is obvious that we cannot find minimizing geodesics or define a metric structure under which distances are defined. We could see, actually, that the length of a path in $\mathbb{R}^3_1$ between two given events is *maximized* by the Euclidean straight line, which is exactly the opposite of what happens in the Riemannian case[3]. Talking about paths, since the Minkowski metric allows for negative-valued inner products, we shall introduce the following classification in order to compare different curves in $\mathbb{R}^3_1$.

Definition 3.5. *Any vector $v \in \mathbb{R}^3_1$ is defined as*

(i) **spacelike** *if $\eta(v, v) > 0$,*

(ii) **lightlike** *if $\eta(v, v) = 0$,*

(iii) **timelike** *if $\eta(v, v) < 0$.*

A curve $\gamma \subset \mathbb{R}^3_1$ will be **spacelike**, **lightlike** *or* **timelike** *if its tangent field $\dot{\gamma}$ is always spacelike, lightlike or timelike, respectively.*

The set of timelike vectors, which is clearly bounded by the set of lightlike vectors (named the **light cone**), has two connected components, as can be seen in Figure 3.2. Given that timelike vectors are inherent to the movement of regular bodies, this separation of components gives a way to discern the past from the future. To pick one of the components as **future-pointing** gives the Minkowski space a **time orientation**.

This gives rise to the essential concept of **causality**. It is said that there is a **causal relation** between two events A and B in space-time if there is some timelike or lightlike path γ connecting them. In this case, if this timelike curve γ starts at A,

[3] Actually, it can be proven a kind of "reversed" triangle inequality, where the length of the path is smaller if the spatial part of the path is *longer*. For a proof of this result, see e.g. [Man93]

it can be proved that A comes before B in every reference frame, and vice versa. On the other side, if there is some spacelike path connecting A and B we say that there is no causality between A and B. This means both that these two events cannot influence each other, and that there exist two inertial frames, where in one of them A precedes B, and in the other one it is the opposite case. In this way, the concept of *simultaneity* as it was known before is broken once and for all.

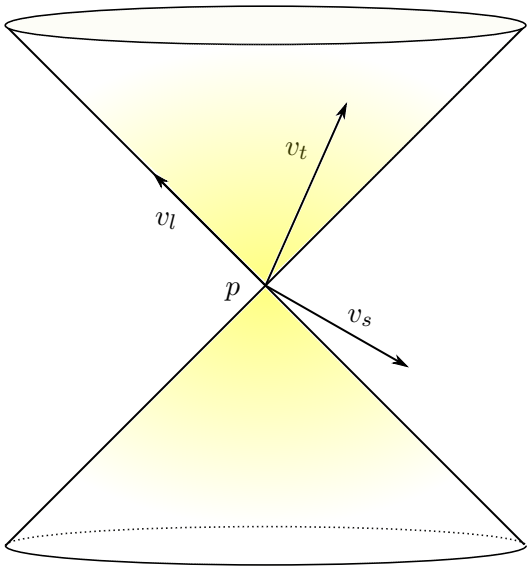

Figure 3.2: Different vectors in relation to the light cone of some point p. In this case, v_t, v_l and v_s are respectively timelike, lightlike and spacelike vectors. In particular, the light cone of a point encompasses all points of the past and future (given a certain time orientation) that have a causal relationship with it.

3.3 General Relativity

We have seen that, strange as it may be, movement in the absence of external force is determined by the Lorentz transformations. However, this is a rather particular description of nature, for the bodies in the universe are usually subject to forces of some kind. So, in order for a theory of gravity to be a proper substitution of the classical theories of motion and dynamics from Newton and Galileo, the action of force had to be considered. But, which force? We know that there are four main forces in nature: the strong and weak nuclear forces, the electromagnetic force and gravity.

General relativity, like Newtonian physics, is meant to model macroscopic phenomena, a term historically defined as those phenomena that can be distinguished by the naked eye. Since the action of nuclear forces is at the scales of atomic size, they are not considered in general relativity. Electromagnetism, differently, is a force that theoretically has infinite range, in the same ratio as gravity, $1/r^2$. But, since it is seen that at macroscopic scales bodies are usually neutral in terms of charge, and hence immune to electromagnetic force, this third force can be neglected as well. This is, of course, not the case for gravity.

3.3.1 The Equivalence Principle

How can we take gravity into account if acceleration yields non-inertial reference frames, which the relativity principle does not consider? Einstein made a crucial observation, in the form of the **equivalence principle**.

Prior to Einstein, there was a physical equivalence he was well aware of: the fact that inertial mass, the one that is accelerated by some force $F = m_i a$ is undistinguishable from gravitational mass, which is the ratio between the gravitational force and acceleration, $F_g = m_g g$.

Going further, in one of his most celebrated mental experiments[4], he affirmed that this equivalence actually extends to a and g, in the following way. An observer bound to a gravitational field would feel the same as another observer in vacuum which was suffering an acceleration $a = g$. Even more interestingly, an observer at rest in vacuum would feel the same as an observer under free fall, on a gravitational field. Then, the equivalence principle can be expressed as:

A reference frame in free fall is locally equivalent to an inertial reference frame.

This was the missing link, as we will see, by which the results of special relativity could be applied to accelerating frames, thus allowing a general theory of relativity to come to fruition.

3.3.2 Einstein Field Equations

In the same fashion as in special relativity, we want to consider space-time as a Lorentzian manifold M and model its behaviour. As said before, there are always some assumptions prior to any physical model, so let us impose the following two: the principle of equivalence and the hypothesis of **general covariance**.

[4] See his own considerations in [Ein07], section V.

This general covariance principle can be seen as an extension of the first postulate of special relativity. It states that physical laws are actually invariant over all reference frames, not only inertial ones. This also means that these laws can (and should) be expressed with mathematical objects that are not dependent on the frame, such as tensors.

These two principles suggest that we should also assume that the structure of M is such that

(i) The geodesics in space-time are the free-falling trajectories.

(ii) M locally reduces to the special relativistic case. In other words, the tangent spaces to each point are Minkowski spaces, $\mathbb{R}^3_1$.

This is explained in more detail in [Wa10], p.66-68. We should thus be able to formulate the basic rules of motion considering the geometry of M, by acknowledging how the structure of M affects trajectories of objects. This was the very intention of Einstein. For that purpose, the Riemann tensor was a good bet, for the behaviour of geodesics and intrinsic properties of M are encoded within it. For the sake of simplicity, we could consider instead the Ricci and scalar curvatures.

On the other side, the mass, or energy (which were deemed equivalent in special relativity), should also have some effect on the dynamics of gravitation, and the final result should reduce to the Newtonian case at conditions of low speed or weak force. This fact is sometimes added as an assumption too, and is quantitatively expressed in the **Poisson equation**

$$\Delta\varphi = 4\pi G\rho, \tag{3.4}$$

where ρ is the matter density and φ is the gravitational field, G being the Newtonian gravitational constant. As we wish to compare this distribution of energy with some tensorial combination that includes Ric and S, it has to be expressed as a tensor too. This gives rise to the appearance of the $(0,2)$ **stress-energy tensor**, $T_{\mu\nu}$.

Like in the case of the Minkowski metric, from now on we will work with natural units, which are $c = 1$ and $G = 1$. We will not expand on how the actual equations were obtained, but just sketch some possible procedure.

First, the stress-energy tensor, which models the distribution of energy as a fluid, satisfies that $\nabla(T_{\mu\nu}) = 0$. There is also a property of the curvature tensors, which is seen e.g. in [Lee18], p.209, by which $\nabla(\mathrm{Ric}_{\mu\nu} - \frac{1}{2}Sg_{\mu\nu}) = 0$, where g is the metric tensor of the Lorentzian manifold.

Thus, adding a constant term Λ to fine-tune the expression, we finally get the **Einstein field equations**:

$$\mathrm{Ric}_{\mu\nu} - \frac{1}{2} S g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{3.5}$$

where the 8π factor comes from the boundary condition given by the Poisson equation, and the constant Λ is called the **cosmological constant**. We see that this is an equation relating $(0, 2)$ tensors on a manifold of dimension 4, so it translates into a system of 16 partial differential equations.

In the Einstein field the following facts are made tangible: that the existence of energy results in the curvature of space-time, and that the gravitational force is encoded into how the trajectories of objects are affected by this curvature.

Thus, *the curvature of space-time is the medium through which the gravitational interaction takes place.*

Chapter 4

Notable applications

We have just seen that space-time is determined by a Lorentzian metric, and that for a given space-time to be feasible option to describe a real setting, its metric has to satisfy the Einstein field equations. However, the Einstein field equations, as beautiful as they might be, are really complex to solve analytically. Working with the curvature tensors and generalizing them to all the space in which we are working can be a strenuous task, if not impossible.

Nevertheless, there have been a handful of studied settings which are simple enough to make such analytic derivation of the resulting metric. We are going to review the two most important ones, the Schwarzschild solution and the arising of cosmology.

4.1 The Schwarzschild Solution

One of the cases for which Newtonian gravitation theory had a straight-forward solution was the simple case of a single sphere or point of mass M. In this case, the gravitational pull felt by a body of mass m in the vicinity of M would go in the direction between them, and have modulus

$$F = G\frac{M \cdot m}{r^2}, \tag{4.1}$$

where r is the modulus of the distance between them.

One could wonder if the analogous of such simple solution could be attained by the means of general relativity. This is what Schwarzschild did, just after Einstein's general relativity came out. Let us see.

The assumptions to find the field outside the sphere will be the following:

(i) The sphere is the unique source of gravitational force, and any other mass outside of it is neglected.

(ii) Spherical symmetry: The field created by the sphere, and thus the whole system, is invariant under rotations.

(iii) Time invariance: The system is static, i.e. there is no variation in the gravitational field over time.

(iv) The cosmological constant Λ is 0.

Now, we can express any possible Lorentzian metric in spherical coordinates as

$$ds^2 = -A(r,t)dt^2 + B(r,t)dr^2 + C(r,t)r^2d\theta^2 + D(r,t)r^2sin^2\theta d\phi^2, \tag{4.2}$$

for some functions A, B, C and D dependent on r and t. Since the system is static, these functions will be just functions of r. Also, at a given radius r, the expression of ds^2 needs to be the same along all angular coordinates because of spherical symmetry, so $C(r) = D(r)$ for all r. We can set $C = D = 1$ without loss of generality, by adjusting A and B conveniently. Then, the metric is of the form

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2d\theta^2 + r^2sin^2\theta d\phi^2. \tag{4.3}$$

To solve the system, we "just" have to find the functions $A(r)$ and $B(r)$, using the Einstein field equations. By assumption (iii), the absence of mass outside the sphere implies that $T_{\mu\nu} = 0$, so the Einstein equations reduce to the so-called **vacuum Einstein equations**:

$$\mathrm{Ric}_{\mu\nu} - \frac{1}{2}Sg = 0. \tag{4.4}$$

Since we have the expression for the metric, calculating the Ricci and scalar curvatures it is a matter of finding the Christoffel symbols and using equation (2.16) to find R, and finally contracting R into Ric and S. To the delight of the reader, we spare the tedious details of such calculation, and give the expression for the tensors.

$$\begin{aligned}
\mathrm{Ric}_{00} &= -\frac{A''}{2B} + \frac{A'B'}{4B^2} + \frac{(A')^2}{4AB} - \frac{A'}{r\cdot B};\\
\mathrm{Ric}_{11} &= \frac{A''}{2A} - \frac{(A')^2}{4A^2} - \frac{A'B'}{4AB} - \frac{B'}{r\cdot B};\\
\mathrm{Ric}_{22} &= -\frac{r\cdot A'}{2AB} + \frac{1}{B} - \frac{r\cdot B'}{2B^2} - 1;\\
\mathrm{Ric}_{33} &= \sin^2\theta R_{22};\\
\mathrm{Ric}_{ij} &= 0, \text{ if } i \neq j,
\end{aligned}$$

where of course A and B are still functions of r, and by A' and A'' we mean the ordinary differentiation of the function respect to r. The scalar curvature will be

$$S = -\frac{A''}{AB} + \frac{(A')^2}{2A^2B} + \frac{A'B'}{2AB^2} - \frac{2A'}{r \cdot AB} + \frac{2B'}{r \cdot B^2} - \frac{2}{r^2 \cdot B} + \frac{2}{r^2}.$$

Putting Ric and S into the Einstein field equations and operating with the differential equations thus obtained (there are terms A', A'' in the mix) we finally get

$$A = K_1 \left(1 - \frac{K_2}{r}\right), \quad B = \frac{1}{1 - K_2/r},$$

where K_1 and K_2 are integration constants. The metric is thus

$$ds^2 = -K_1 \left(1 - \frac{K_2}{r}\right) dt^2 + \frac{1}{1 - K_2/r} dr^2 + r^2 d\theta^2 + r^2 sin^2\theta d\phi^2.$$

As gravity vanishes at infinite distance from the source, the limit when $r \to \infty$ has to be the Minkowski metric, the one corresponding to flat space, and hence $K_1 = 1$. Also, since Newtonian gravity must be a particular case for weak gravity fields, it is seen that $K_2 = 2M$ (in natural units), M being the mass of the sphere. Then,

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \frac{1}{1 - 2M/r} dr^2 + r^2 d\theta^2 + r^2 sin^2\theta d\phi^2. \tag{4.5}$$

As we have the metric of space-time, the geodesic structure can be found, and with it the velocity and acceleration of any free-falling body, the latter being[1]

$$a = \frac{d^2r}{d\tau^2} = \frac{M}{r^2} \left(1 - 2M/r\right)^{-\frac{1}{2}}. \tag{4.6}$$

After applying this solution to the Solar System, it was found that it was not only extremely precise, but also it accurately modelled the orbit of Mercury, which was not predicted well with Newton's theory. We now know that this was because Newton's theory is only an approximation for the case where the gravitational field is weak, which is not the case at the location of Mercury. This was the first huge success of general relativity, in a chain of never-ending valid predictions. It would be all good only if we had not noticed that some coefficients of the metric reach infinity values in two cases:

(i) $r = 0$: This one is somehow understandable, since it could mean that the metric is not well-defined at the center of the sphere.

(ii) $r = 2M$: This spherical shell is labelled as the **event horizon**, and $r_S = 2M$ is called the **Schwarzschild radius**.

[1] Assuming non-angular motions and $a(0) = 0$. It is expressed in units of *proper time* τ, where this proper time quantity models the flow of time considering relativistic effects.

We define the points at which the metric goes to infinity as **singular points**, or just **singularities**. By (ii), the submanifold $r_S = 2M$ divides the manifold in two parts, and prevents the metric to be defined along all the space. This is extremely interesting, for it has no obvious explanation in terms of classical or relativistic physics.

How would this look like, in a physical sense? Let us consider that the massive sphere is a single star, which is a good approximation since its mass completely overwhelms any other lesser (non-star) body nearby. From basic astrophysics, it is known that nearly all kinds of stellar bodies satisfy that their radius R is bigger than r_S, so this singularity cannot be found on the outside of any of them[2]. Otherwise, any stellar body for which r_S is bigger than its radius is then defined to be a **black hole**.

These mysterious objects, whose existence was proven five decades later, feature this supposed singularity at their Schwarzschild radius, which is now relevant for it falls outside of them. They are the remnants of a star that collapsed within itself and whose density is sufficiently high to permit that $R < 2M$. But, why are they called this way?

To see it qualitatively, if the event horizon existed in a Schwarzschild metric, the space would no longer be connected. Hence, if it were indeed a physical singularity, no geodesic could trespass this spherical shell, in any direction. However, there is physical evidence that black holes accrete matter and light, thus the space (except at $r = 0$) should be geodesically complete, regardless of this apparent singularity. It turns out that this singularity was produced by the choice of coordinates, and it was corrected later in the so-called **Kruskal extension**. Nevertheless, it is a surprising coincidence that the Schwarzschild radius is exactly the maximum radius at which light cannot escape the gravitational influence of the mass. By the maximality of c, it means in particular that *every body that crosses the event horizon can never go back again, including light.*

To recapitulate, although we have seemingly solved the Schwarzschild setting and found an explanation for the discontinuity of ds^2 at $r = 2M$, we still have this singularity at $r = 0$ chiming in. In order to understand the fundamental nature of singularities, we have to delve into what the Hawking theorem says to us, which will be exactly the object of next chapter. Before that, we still want to see another huge consequence of general relativity, in the following section.

[2]To see how ridiculously small is the Schwarzschild radius in "common" bodies, we have for example that for the Sun it is $r_S \approx 3$ km, and for the Earth it is $r_S \approx 1$ cm.

4.2 Cosmology

We have seen that just the Schwarzschild solution alone is already a replacement and refinement of the classical theory of gravitation. We may wonder, then, if we could apply relativity to less simple settings, so what about trying to model the Universe as a whole? In this case, we have to assume that the Universe satisfies two important hypotheses: the one of **perfect fluid** and the so-called **cosmological principle**. Let us quickly review them:

The **cosmological principle** is the combination of the assumption of isotropy (which is proved by observations), and the so-called *Copernican principle*, which roughly says that we are not in a privileged position inside the Universe. Combining both statements we claim that *at sufficiently large scales, the Universe is isotropic and homogeneous*, and so is the expression of the metric.

Also at sufficiently large scales (also named *cosmological scales*), we may assume the hypothesis that the distribution of energy in the Universe is given by the one of **perfect fluid**. This is the simplest possible model of fluid, which is isotropic, and there is neither transport of energy nor friction, and hence the stress-energy tensor takes a very simple form. Indeed, as it can bee seen in [Cep07], p.111, the tensor is diagonal, and has the following form:

$$T_{\mu\nu} = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}, \tag{4.7}$$

ρ being the energy density and p the pressure of the fluid.

As before, we need to characterize the metric in order to solve the Einstein equations. Now, although we may consider that the Universe varies over time, which is true, let us separate the time and space description, defining the metric as

$$ds^2 = -b(t)^2 dt^2 + a(t)^2 d\tilde{s}^2.$$

As before, we can take $b(t)^2 = 1$ without loss of generality. Here, $d\tilde{s}^2$ will be the spatial description of the metric, where space is a Riemannian 3-submanifold N, since the metric is positive definite in N. As N is isotropic, the Riemann tensor is defined by

$$R_{ijkl}(p) = K_p(g_{il}g_{jk} - g_{ik}g_{jl}) \tag{4.8}$$

(see a proof in [GoNa14], p.128).

Since the Universe is homogenous, if we change the origin of coordinates the curvature should not change, so we see that actually K_p is constant, i.e. the spatial submanifold N has constant curvature[3], which we denote by K. Since the concept of sectional curvature, defined in Chapter 2, has the same information as the Riemann tensor, this means that every possible sectional curvature all over N is constant and equal to K. By the Killing-Hopf theorem, which is detailed in [Lee18], p.348, this means that the manifold N is either isometric to $\mathbb{S}^3$, $\mathbb{R}^3$ or $\mathbb{H}^3$, for the cases when, respectively, $K > 0, K = 0$ and $K < 0$. Thus, we have three possible metrics for N, and we can find spherical coordinates (r, θ, ϕ) such that they look like

$$\begin{cases} d\tilde{s}^2 = dr^2 + \sin^2 r \left(d\theta^2 + sin^2\theta d\phi^2\right), & \text{if } K > 0; \\ d\tilde{s}^2 = dr^2 + r^2 \left(d\theta^2 + \sin^2 \theta d\phi^2\right), & \text{if } K = 0; \\ d\tilde{s}^2 = dr^2 + \sinh^2 r \left(d\theta^2 + \sin^2 \theta d\phi^2\right), & \text{if } K < 0. \end{cases} \tag{4.9}$$

This, in turn, after the due reparametrizations, can easily be expressed in a more compact form,

$$d\tilde{s}^2 = \frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + sin^2\theta d\phi^2\right), \tag{4.10}$$

where k is a parameter of just the sign of the curvature, that can be $k = 1, 0, -1$, and relates to K in that $K(t) = k/a(t)^2$. This is as far as we can get with our hypotheses, and now adding $d\tilde{s}^2$ to the whole metric, we obtain the so-called **Robertson-Walker metric**.

$$ds^2 = -dt^2 + a(t)^2 \left(\frac{dr^2}{1 - kr^2} + r^2 \left(d\theta^2 + sin^2\theta d\phi^2\right) \right). \tag{4.11}$$

In the same manner as with the Schwarzschild metric, we need to find the expression for the curvature tensors and solve the needed differential equations, this time over $a(t)$. Again, after a laborious calculation of the Christoffel symbols and then the Riemann tensor, we obtain the Ricci and scalar curvatures:

$$\begin{aligned} &\mathrm{Ric}_{00} = -3\frac{\ddot{a}}{\dot{a}}; \\ &\mathrm{Ric}_{ii} = \frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2} + \frac{2k}{a^2}, \quad \text{for } i = 1, 2, 3; \\ &S = -6\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{\ddot{a}}{a} + \frac{k}{a^2}\right], \end{aligned}$$

where as before the Ricci tensor is diagonal. Putting these expressions into the Einstein field equations, together with the stress-energy tensor seen in (4.7), we get

[3] We should not mistake this curvature, which concerns the spatial submanifold N of space-time, with the curvature of the tetra-dimensional space-time. For instance, the spatial curvature can be 0, but, as we have seen, space-time is always curved since there is matter in it. This is still a source of confusion nowadays, so it must be made clear.

the following two differential equations for $a(t)$, the **Friedmann equations**:

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi}{3}\left(\rho(t) + 3p(t)\right) + \frac{\Lambda}{3} \tag{4.12}$$

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi}{3}\rho(t) + \frac{\Lambda}{3} - \frac{k}{a(t)^2}. \tag{4.13}$$

Let us just admire for a moment the fact that the Einstein equations have allowed us to naturally parametrize the whole universe, from just the assumptions of perfect fluid and the cosmological principle. With the Friedmann equations, the science of cosmology was born.

If we want to determine what the exact spatial curvature of universe is, we have to get help from observational data. As an overly simplified explanation, it was observed that, looking the furthest away possible, any triangle of cosmic size satisfies that the sum of its angles is 180°. Hence, the space is actually flat, and $k = 0$.

However, we are still clueless about some issues. What is this function $a(t)$, and the constant Λ? If these equations are really determining a dynamical system, is the universe evolving over time? Also, why do we have an under-determined system, with three variables and just two equations? Let us answer the first two questions.

If we look at the Robertson-Walker metric, we see that $a(t)$ is a function that tells about the "size" of the metric, in that the distance between any two points will be proportional to the value of $a(t)^2$. Then, if we fix any two points, although we do not move the points, their relative distance will vary with time according to $a(t)$. The Friedmann equations tell us that in general $a(t)$ is not a constant function, so it is obtained that the metric, and hence the universe itself, expands or contracts with the passing of time.

Einstein, who was an advocate of a static universe, introduced in his equations a mathematically valid artefact, the cosmological constant Λ. By carefully adjusting this constant, a static universe can be obtained[4]. However, as it was later seen by observing the rate at which other galaxies depart from us, the Universe is expanding, so theoretically this constant was no longer required. But finally, ground-breaking observations at the end of the 20th century pointed out that the universe expansion was actually accelerating. It turns out that this constant Λ can be a measure of this acceleration, whose source is not understood and thus named *dark energy*.

[4] As we see in the first Friedmann equation, if we remove the cosmological constant term and consider that $\rho, p > 0$, which was the classical view, then we obtain a negative acceleration $\ddot{a}(t)$ and hence an always present tendency of the universe to contract.

Finally, we should mention that we have derived the Friedmann equations from the most general setting possible, but we have not yet introduced any actual assumption on the nature of the constituents of the universe. Then, it is logical that the system of equations is under-determined. In physics, one characterizes the properties of a substance through its **equation of state**, which usually relates its correspondent thermodynamic quantities: pressure, volume, density... It makes sense to introduce an equation of state of any component of the universe as

$$p = \omega\rho, \tag{4.14}$$

where ω is a constant that is characteristic of the substance. For example, for ordinary matter we have $\omega = 0$, whereas for dark energy we have $\omega = -1$.[5]

In this case, the system of equations is well-determined, and, after some experimental data that tells us about the evolution of the components of the universe through time, the evolution of the expansion of the universe can be obtained. There is attached below a (very) qualitative representation of this evolution.

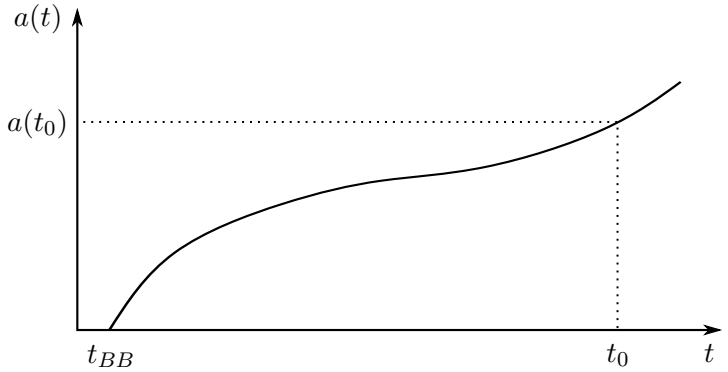

Figure 4.1: Evolution of $a(t)$, from the Big bang (BB) until present (t_0). A rigorous explanation and plot can be found e.g. in [Cep07] p.174.

We can see that if we go sufficiently far away to the past, about $13.8 \cdot 10^9$ years, $a(t)$ tends to reach zero, for which a quick look into the Friedmann equations sees that this would provoke a singularity. This hypothetical singularity is named **the Big Bang**, and it should mark the origin of the universe as we know it.

We have collected in this chapter some reasons to desire to understand the nature of singularities. Luckily, it is now time to do so.

[5]This seemingly unrealistic equation of state, where the density of energy does not decay when the universe expands, corresponds to the Λ term (and thus to dark energy) in the Friedmann equations, as seen e.g. in [Cep07], p.130.

Chapter 5

Hawking Singularity Theorem

Let us look now into the nature of singularities. We have seen that at certain points of space-time the expression for the metric may have an asymptotic discontinuity, and we want to characterize how does space-time behave in the vicinity of those points. Since singularity points limit the completeness of geodesics (i.e. geodesics cannot trespass those points), we want to obtain a way to relate this completeness of geodesics with intrinsic properties of the manifold.

This sounds familiar, for it is exactly what was done in the Hopf-Rinow theorem. However, this was a theorem for Riemannian manifolds, where it is used that any such manifold can be given a metric structure and we had appropriate means to calculate distances in terms of geodesics. This is not the case for Lorentzian manifolds, so we want to see how can we manage to obtain a corresponding result in this context.

5.1 Preliminary notions on causality

First, we need to make some more assumptions on the structure of space-time, so that we are working with manifolds that are realistic enough. We thus need to extend the notions given in Definition 3.5, which are the starting point to understand the relations of causality in space-time.

Definition 5.1. *A space-time M is* **time-orientable** *if there exists some smooth vector field $X \in \mathcal{T}(M)$ for which $\langle X, X\rangle < 0$ at all points $p \in M$. It is possible then to have a definite* **time orientation** *of the space-time by orienting all the tangent spaces (which are Minkowski spaces) either towards the future or the past.*

This allows us to rigorously quantify the arrow of time. Giving a space-time a time-orientation we can check if a curve points to the future, the past, or neither of them.

Definition 5.2. *A timelike trajectory $c : I \to M$ is* **future-directed** *if $\dot{c}(t)$ is future-pointing for all $t \in I$. For every $p \in M$, the set $I^-(p)$ of all points that can be linked to p by some timelike curve is denominated as the* **chronological past of** *p. Conversely, the set $I^+(p)$ of points that that can be connected with p by a timelike trajectory starting at p is the* **chronological future of** *p.*

If we allow the tangent vectors to the trajectory to be either timelike or lightlike (instead of them being all timelike), then c is said to be a **causal curve**, and the correspondent concepts to $I^-(p)$ and $I^+(p)$ will be the ones of the **causal past** $J^-(p)$ and **causal future** $J^+(p)$. Since we want to know when space-times are not geodesically complete (i.e. when geodesics cannot be extended any more), the following definitions are crucial:

Definition 5.3. *A smooth future-directed causal curve $c : (a, b) \to M$ will be* **future-inextendible** *if $\lim_{t\to b} c(t)$ does not exist. This gives also the definition for* **past-inextendible** *curves, which is the analogous case for past-directed curves. Given some set $S \subset M$, the* **future domain of dependence** *$D^+(S)$ is the set of points $p \in M$ for which all past-inextendible causal curves starting at p intersect S. With an analogous definition we also get the* **past domain of dependence** *$D^-(S)$.*

The domains of dependence are then a measure of the incompleteness of geodesics in a given region around any set of points S. They also allow us to define the following important concepts.

Definition 5.4. *A space-time M is* **stably causal** *if there exists a smooth function $t : M \to \mathbb{R}$, called* **time function***, such that its gradient is timelike. Let M be any such manifold. Then, we define a* **Cauchy hypersurface**[1] *as a level set $S_a \coloneqq t^{-1}(a)$ of some $a \in \mathbb{R}$ such that $D(S_a) \coloneqq D^+(S_a) \cup D^-(S_a) = M$. If all level sets of the time function t are Cauchy hypersurfaces, then M is said to be* **globally hyperbolic***.*

The property of a space-time being globally hyperbolic assures us that the causal relations are not unnatural and is therefore a sign of physically sensible properties. Hereafter, when we refer to any space-time M we will implicitly assume that it is stably causal and globally hyperbolic. We shall prove now the following technical proposition, that will be important later on.

Proposition 5.5. *Given a space-time M, a Cauchy hypersurface S and a point $p \in D^+(S)$, then $A \coloneqq D^+(S) \cap J^-(p)$ is a compact set of M.*

Proof. Since we have to derive topological properties, we define a basis for the topology of M that will be useful. A geodesically convex set $U \subset M$ (i.e. such that all

[1] It can easily be seen that they are 3-submanifolds of M, hence the hypersurface label.

points within U can be linked with some geodesic) will be a **simple neighborhood** if it is diffeomorphic to an open ball bounded by a compact submanifold of a bigger geodesically convex open set.

It can be seen that simple neighborhoods are a basis for the topology of M, and thus every open cover of M can be expressed in terms of some union of simple neighborhoods. We want to prove that any open cover of A by simple neighborhoods has some finite subcover. Let us assume it is false, for some cover $\{U_n\}_{n\in\mathbb{N}}$ (we can assume that this cover is countable because smooth manifolds satisfy the second countability axiom). We take a sequence $\{q_n\}_{n\in\mathbb{N}}$ such that $q_n \in A \cap U_n$ and $q_n \neq q_m$ if $n \neq m$. In this way, the sequence does not have accumulation points, and since by definition $\overline{U}_n$ is compact, in each U_n there is only a finite amount of such points.

We now define a sequence $\{p_n\}_{n\in\mathbb{N}}$ that will cause a contradiction and thus prove the claim. We start by setting $p_1 = p$, where p belongs to some U_1, and take some q_m such that $q_m \notin U_1$. By construction there is some future-directed causal curve c_1 from q_m to p_1 (because $A \subset J^-(p)$) that crosses the boundary ∂U_1 at some point r_{1m}. Since we have assumed that A is not compact, there are infinite points q_m and therefore infinite points $r_{1m} \in \partial U_1$. They accumulate at some point $p_2 \in \partial U_1$, by the compactness of $\overline{U}_1$, as pictured in Figure 5.1.

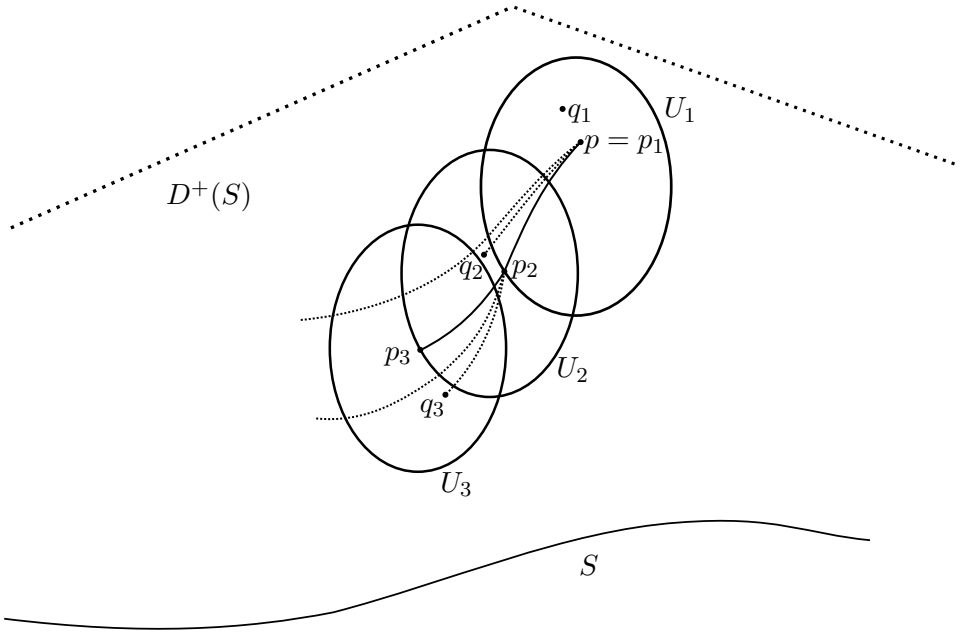

Figure 5.1: Construction of the sequence $\{p_n\}_{n\in\mathbb{N}}$.

This sequence of points r_{1m} corresponds to a sequence of causal geodesics $\{\gamma_{1m}\}$, which by the properties of simple neighborhoods also converges naturally to a causal geodesic γ_1 between p_1 and p_2. Then, $p_2 \in J^-(p)$. Since $p \in D^+(S)$, then $t(r_{1m}) \geq t(S)$, which implies $t(p_2) \geq t(S)$ and thus $p_2 \in D^+$, so $p_2 \in A$. Also, $p_2 \notin U_1$, so we now consider a simple neighborhood U_2 for which $p_2 \in U_2$.

Doing the exact same procedure we find p_3 as an accumulation point of the intersections between causal curves from p_2 to q_m and ∂U_2. Again, it is found that p_2 and p_3 are joined by a causal curve γ_2, and equally as before $p_3 \in A$. We can repeat this *ad infinitum*, since the cover $\{U_n\}$ has no finite subcover, and find a sequence $\{p_n\}$ such that all points belong to different neighborhoods U_i and all successive points are joined by a causal geodesic γ_n.

This piecewise smooth causal curve $\cup_i \gamma_i$ can be smoothed on each p_i so that we finally have a past-directed causal curve γ starting at p. We have seen recursively that for all points p_i we have $t(p_i) \geq t(S)$, and therefore $\gamma \cap S = \emptyset$. Also, the sequence $\{p_n\}$ has no limit, since there are no accumulation points by the construction in terms of simple neighborhoods. To sum it up, we have obtained a past-inextendible smooth causal curve starting at p that does not intersect S, reaching a contradiction with the definition of $D^+(S)$ and therefore proving that A is indeed compact. □

5.2 Singular space-times

A common justification of the singularities featured in the models we have studied was that they appeared because the model was an ideal and perfectly symmetric one, and that they should disappear once real-world effects were taken in account. We intend to prove in what follows that this is false, for they are instead intrinsic real properties of the physical setting. Labelling as **singular** all space-times that are not geodesically complete, we will analyse under which conditions a space-time happens to be singular. Let us apply some of the geometric concepts into our situation:

Given a space-time M and a Cauchy hypersurface $S \subset M$, we can extend the exponential map (which defines the geodesics) to all $p \in S$, in the following way. Given the unique future-pointing unit vector field n that is normal to S, and some set of geodesics $\{c_p\}_{p \in S}$ with initial tangent vectors n_p at every p, the exponential map of some open set $U \subset \mathbb{R} \times S$ is defined as $\exp(t, p) = c_p(t)$. An important definition is the one of **conjugate points** to S, which are all the points which are critical points of some exponential map $\exp_p = c_p$, for $p \in S$.

Since $S = t^{-1}(a)$ has constant time, the restriction of the metric g to S will be zero on all the components related to time: $g_{0i}|_S = 0$. We can take local coordinates (x_1, x_2, x_3) on S around some point p, and adding the time coordinate we will have full local coordinates (t, x_1, x_2, x_3) in some open neighborhood V around p. The neighborhood V will contain any point $q = \exp(t_0, p)$ which is not conjugate to S, so we can find the value of $\frac{\partial g_{0i}}{\partial t}$ along the geodesic $\exp_p$, for $i = 1, 2, 3$:

$$\begin{aligned}\frac{\partial g_{0i}}{\partial t} &= \frac{\partial}{\partial t}\left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial x^i}\right\rangle = \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial x^i}\right\rangle \\ &= \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial t}\right\rangle = \frac{1}{2}\frac{\partial}{\partial x^i}\left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right\rangle = 0,\end{aligned}$$

which follows naturally from the compatibility with $\langle\cdot,\cdot\rangle$ and the symmetry properties of ∇. Thus, in this coordinates, g_{0i} are the same as in S ($g_{0i} = 0$), and we have that S has not only constant time, but that it is also orthogonal to the time coordinate. This set of coordinates is called a **synchronized coordinate system**. If we set $\gamma_{ij} \coloneqq \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\rangle$, we have that the Christoffel symbols are

$$\begin{aligned}\Gamma^0_{00} &= \Gamma^i_{00} = 0, \\ \Gamma^i_{0j} &= \sum_{k=1}^{3} \gamma^{ik}\beta_{kj},\end{aligned}$$

with $\gamma^{ij} = (\gamma_{ij})^{-1}$ and $\beta_{ij} = \frac{1}{2}\frac{\partial \gamma_{ij}}{\partial t}$. Then, with this notation, the Ricci tensor has the following first component:

$$\mathrm{Ric}_{00} = -\frac{\partial}{\partial t}\left(\sum_{i,j=1}^{3} \gamma^{ij}\beta_{ij}\right) - \sum_{i,j,k,l=1}^{3} \gamma^{jk}\gamma^{il}\beta_{ki}\beta_{lj}. \tag{5.1}$$

Defining $\theta \coloneqq \sum_{i,j=1}^{3} \gamma^{ij}\beta_{ij}$ we obtain, by applying the matrix equality $(\log(\det A))' = \mathrm{tr}(A^{-1}A')$, that

$$\theta = \frac{1}{2}\mathrm{tr}\left((\gamma_{ij})^{-1}\frac{\partial}{\partial t}\gamma_{ij}\right) = \frac{1}{2}\frac{\partial}{\partial t}\log\gamma. \tag{5.2}$$

This new function, which is named the **expansion**, facilitates greatly our work, since it has a singularity for points where the synchronized coordinates are zero, and these are precisely the conjugate points to S.

Another hypothesis that will have to be assumed in order to deal with "realistic" space-times is the **strong energy condition**, which is defined as the property that every timelike vector field $V \in \mathcal{T}(M)$ satisfies that $\mathrm{Ric}(V_p, V_p) \geq 0$, for all $p \in M$. Assuming this, while using the tools we have just defined, allows us to walk our steps towards the Hawking theorem.

Proposition 5.6. *Given a space-time M that satisfies the strong energy condition, let $S \subset M$ be a Cauchy hypersurface. For every point $p \in S$ such that $\theta = \theta_0 < 0$, its associated geodesic c_p contains at least a point conjugate to S, whose distance from S is, at most, of $-\frac{3}{\theta_0}$ in the future.*

Proof. Applying the strong energy condition to (5.1) and (5.2), we get

$$\frac{\partial \theta}{\partial t} + \sum_{i,j,k,l=1}^{3} \gamma^{jk}\gamma^{il}\beta_{ki}\beta_{lj} \leq 0. \tag{5.3}$$

We are free to choose an orthonormal basis, by applying the Gram-Schmidt process to the x_i coordinates, so we can set $\gamma^{ij} = \delta_{ij}$. Also, using the algebraic matrix inequality $(\mathrm{tr}(A))^2 \leq n \cdot \mathrm{tr}(\mathrm{A^t A})$, we obtain that the second term of (5.3) satisfies

$$\sum_{i,j,k,l=1}^{3} \gamma^{jk}\gamma^{il}\beta_{ki}\beta_{lj} = \sum_{i,j=1}^{3} \beta_{ji}\beta_{ij} = \mathrm{tr}(\beta_{ij} \cdot (\beta_{ij})^t) \geq \frac{1}{3}\theta^2,$$

so then we have that $\frac{\partial \theta}{\partial t} + \frac{1}{3}\theta^2 \leq 0$, which brings us to $-\frac{\partial \theta}{\theta^2} \leq \frac{\partial t}{3}$. Integrating:

$$\frac{1}{\theta} \geq \frac{1}{\theta_0} + \frac{t}{3} \tag{5.4}$$

Hence, $1/\theta$ crosses zero at some $t \leq -\frac{3}{\theta_0}$, which signals a point conjugate to S. □

We have seen a result that tells us when a conjugate point is expected to be found, starting from some arbitrary point in S. It can be seen that in this case the geodesic c_p does not maximize distance[2], which is made tangible in the following proposition.

Proposition 5.7. *Given a space-time M and a Cauchy hypersurface S, we consider a point $p \in M$ and a timelike geodesic c which goes through p and is also orthogonal to S. Then, c will not be a maximizing geodesic between p and S if there is a point q between p and S which is conjugate to S.*

Proof. The proof is outlined in [GoNa14], p.302. Since q is conjugate to S, it is seen as an intersection point of geodesics that are orthogonal to S. We can thus find one such geodesic $\tilde{c}$ that has the same length from S to q as c. Then, considering some geodesically convex neighborhood V of q, and two points $r, s \in V$ such that $r \in \tilde{c}$ and s is in c between p and q, there is a geodesic c_V in V between r and s. The new curve formed by $\tilde{c}$, c_V and the upper part of c has strictly more length than c, since in a geodesically convex neighborhood the distance is maximized by geodesics[3]. □

[2]Note that, as we said in p.27, in Lorentzian manifolds the geodesics are not minimizing curves, but rather tend to be instead maximizing trajectories. In Minkowski space-time geodesics are always maximizing curves, but this is not necessarily true in more general space-times.

[3]It is because of the generalized twin paradox ([GoNa14], p.262), which roughly states that, among two trajectories with the same spatial endpoints, the maximal path is the one which has undergone less acceleration.

Although we guess that all these results are important on their own, we may not find yet any intuitive significance in them. With the introduction of the following important theorem, which is the critical argument in the proof of Hawking's theorem, we will understand their part in the picture.

Theorem 5.8. *Given a space-time M, a Cauchy hypersurface S and some point $p \in D^+(S)$, there exists a timelike curve between p and S which has maximal length and is a geodesic orthogonal to S.*

Proof. Let $T(S,p)$ be the set of timelike curves between p and S, which are closed subsets of $A = D^+(S) \cap J^-(p)$ and hence compact, since A has been showed to be compact in Proposition 5.5. It can be seen (e.g. in [Nab88], p.166) that the set $C(A)$ of all compact subsets of A is a compact metric space, with the so-called **Hausdorff metric** d_H. This metric, given an underlying metric d, acts on every two subsets of A in the following way:

$$d_H(K,L) = \inf\{\varepsilon > 0 \,:\, K \subset U_\varepsilon(L); L \subset U_\varepsilon(K)\},$$

where the ε-neighborhoods U_ε are defined as $U_\varepsilon(\Sigma) \coloneqq \{p \,:\, d(p,\Sigma) < \varepsilon\}$ (as pictured qualitatively in Figure 5.2). Since the set of lightlike vectors is the boundary for the set of timelike vectors, the set $\overline{T(S,p)}$ can be seen as the set of continuous causal curves between p and S. Now, we can assume without loss of generality that $t(S) = 0$, and then the length of a given geodesic c is simply

$$\tau(c) = \int_0^{t(p)} |\dot{c}(t)| dt. \tag{5.5}$$

We want to see that this length function is upper semicontinuous, i.e. that for every $\delta > \tau(c)$ there is a neighborhood V of c such that $\tau(\gamma) < \delta$ for all the curves $\gamma \subset V$. We can work with the arclength function u, which in the case of c is equivalent to τ. Let us see how will be the length of some curve γ in a small open set $U_\varepsilon(c)$.

Shrinking U_ε if necessary, we can define an arclength function in all of U_ε from the original $u(c)$, by imposing that the constant length lines are orthogonal to c (this is again better seen in Figure 5.2). In this way, the gradient of u reduces to $\dot{c}$ when applying it on c. This means that for γ it will be equivalent to say that $du(\dot{\gamma}) = 1$ in some point than to say that the gradient reduces to $\dot{\gamma}$ there, i.e. $\langle \dot{\gamma}, \mathrm{grad}(u)\rangle = 1$. By this, we can express the vector $\dot{\gamma}$ at any point as the sum of the tangent and orthogonal components to $\mathrm{grad}(u)$, where we label the orthogonal component as X :

$$\dot{\gamma} = \frac{1}{\langle \mathrm{grad}(u), \mathrm{grad}(u)\rangle}\mathrm{grad}(u) + X \implies |\dot{\gamma}| = \left| \frac{1}{\langle \mathrm{grad}(u), \mathrm{grad}(u)\rangle} + \langle X, X\rangle \right|^{\frac{1}{2}},$$

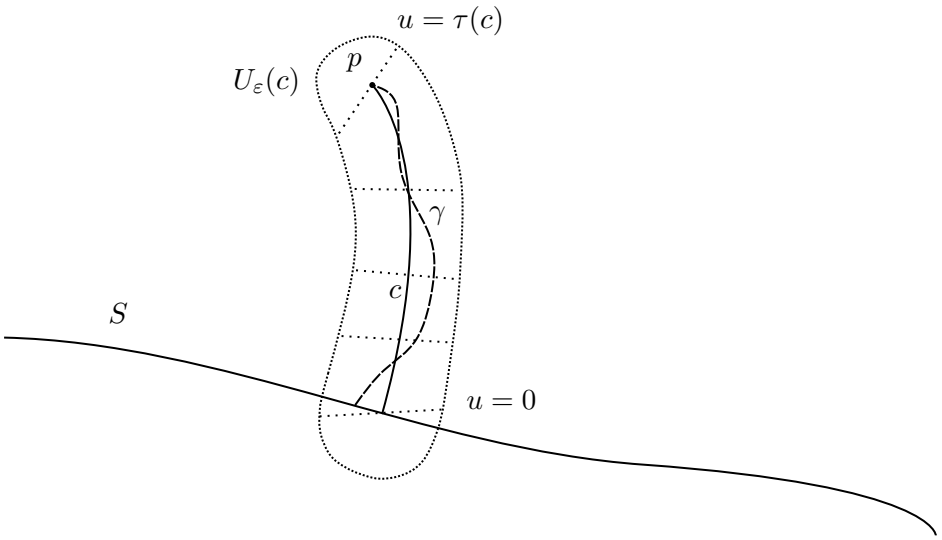

Figure 5.2: Sketch of the situation. In a neighborhood $U_\varepsilon(c)$ of c we consider curves γ and see that the length function is upper semicontinuous.

since the components are orthogonal. Now, as $\mathrm{grad}(u)$ is timelike, the product $\langle \mathrm{grad}(u), \mathrm{grad}(u)\rangle$ is negative, and is exactly -1 along c by definition. For any $\delta > 0$, we can find an $\varepsilon > 0$ such that U_ε allows for so little variation of γ from c that the following is satisfied:

$$-\frac{1}{\langle \mathrm{grad}(u), \mathrm{grad}(u)\rangle} < \left(1 + \frac{\delta}{2\tau(c)}\right)^2.$$

Then, taking into account the fact that γ can start in a different point than $c(0)$, which we label as γ_0, using the arclength parametrization we have

$$\tau(\gamma) = \int_{u(\gamma_0)}^{\tau(c)} |\dot{\gamma}| du, \tag{5.6}$$

so, by the previous inequality,

$$\begin{aligned}\tau(\gamma) &= \int_{u(\gamma_0)}^{\tau(c)} \left| -\frac{1}{\langle \mathrm{grad}(u), \mathrm{grad}(u)\rangle} - \langle X, X\rangle \right|^{\frac{1}{2}} \\ &< \int_{u(\gamma_0)}^{\tau(c)} \left(1 + \frac{\delta}{2\tau(c)}\right) du = \left(1 + \frac{\delta}{2\tau(c)}\right) \cdot (\tau(c) - u(\gamma_0)).\end{aligned}$$

Again, shrinking U_ε if needed, we can have that, for a given δ, u is sufficiently small so that $\tau(\gamma) < \tau(c) + \delta$, and thus τ is upper semicontinuous over c. This can naturally

be extended to the curves living in $\overline{T(S,p)}$, by taking the limit:

$$\tau(c) = \lim_{\epsilon \to 0} \sup\{\tau(\gamma) \ : \ \gamma \in B_\varepsilon(c) \cap T(S,p)\}. \tag{5.7}$$

If we add the fact that $\overline{T(S,p)}$ is compact, then the extended τ function must have a maximum for some element $\tilde{c} \in \overline{T(S,p)}$. For the moment, this just means that there is some sequence $\{c_n\}$ that tends to $\tilde{c}$, so we want to see that this maximal element corresponds actually to some real geodesic between S and p.

This can be done by dividing $\tilde{c}$ so that each step is in a geodesically convex neighborhood, and constructing a finite sequence of points $\{p_i\}$ along $\tilde{c}$ in which also each step is in such a well-behaved neighborhood. Given the fact that locally the time function is well-defined, the sequence $\{c_n\} \to \tilde{c}$ translates into the sequences of points $\{p_{i_n}\}$, that converge to the points p_i. The piecewise smooth curve $\tilde{\gamma}$ joining all points p_i from S until p is clearly a piecewise smooth geodesic, and by construction has length equal to $\tau(\tilde{c})$.

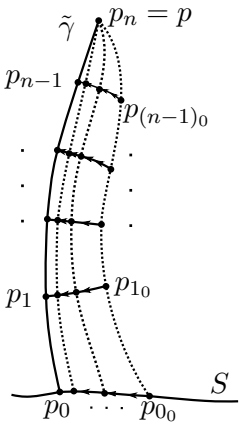

Figure 5.3: Sketch of the procedure to obtain the maximal geodesic $\tilde{\gamma}$. The curves of the sequence that tends to $\tilde{\gamma}$ are piecewise smooth, and $\tilde{\gamma}$ is smooth.

Since a piecewise smooth curve has to be smooth also on the division points in order to be maximal (again, by the generalized twin paradox), we have that $\tilde{\gamma}$ is a smooth curve, and therefore a timelike geodesic. Since it has maximal length, if it were not orthogonal to S at $\tilde{\gamma}(0)$, by the expression of synchronized coordinates at $\tilde{\gamma}(0)$ we could obtain an alternative curve orthogonal to S with larger length than $\tilde{\gamma}$. In conclusion, $\tilde{\gamma}$ is a maximizing geodesic between S and p which is orthogonal to S. □

We are ready to state and prove Hawking's theorem in a simple and natural way.

Theorem 5.9. (Hawking Singularity Theorem) *Let M be a space-time that fulfils the strong energy condition. If the expansion θ is such that $\theta \leq \theta_0 < 0$ on some Cauchy hypersurface S, then M is singular.*

Proof. Let us see that all future-directed timelike geodesics which are orthogonal to S are future-inextendible after $\tilde{\tau} = -\frac{3}{\theta_0}$. If this was not true, we could consider for instance some geodesic c defined along $[0, \tilde{\tau} + \varepsilon]$, with $\varepsilon > 0$.

If we now consider the point $p \coloneqq c(\tilde{\tau}+\varepsilon)$, by Theorem 5.8 we would have a geodesic of maximal length γ between S and p. By this maximizing property, this curve would satisfy $\tau(\gamma) \geq \tilde{\tau} + \varepsilon$, and thus have some conjugate point q to S, by Proposition 5.6. This means, by Proposition 5.7, that γ is not a maximizing geodesic anymore, therefore reaching contradiction. □

In a nutshell, this theorem gives a condition on the expansion quantity θ that results in unavoidable singularity points. Also, considering a perturbation on the space-time to account for "real-world effects", if θ is still bounded below 0 then the singularities do not vanish.

Let us see now the singularities that we encountered in the previous chapter, in the light of this new knowledge. Both the Schwarzschild [4] and cosmology solutions satisfy the strong energy condition, as a quick peek in the expression of the stress-energy tensor can reveal. Also, since the property of being globally hyperbolic means that the inextendible geodesics from every point in the considered space-time intersect every constant time hypersurface, they are seen to be globally hyperbolic, singularity points aside.

Indeed, both Schwarzschild and Robertson-Walker metrics are stably causal, given that in both of them the time coordinate gives a way to obtain a global smooth time function. Then, for the Schwarzschild case, the time invariance of the system implies that any "time snapshot" $S(t = a)$ (i.e. any Cauchy hypersurface) intersects any geodesic from any point in space-time that is time-oriented towards $S(t = a)$. In other words, any geodesic within $r < 2M$ will, at some point, pass through some point with $t = a$.

For the zero curvature Robertson-Walker metric (the one that is the current description of the universe and hence the one we will consider), it is clear by homogeneity and the flatness of its Cauchy hypersurfaces that it is also globally hyperbolic.

[4] Given the apparent singularity at $r = 2M$, we consider in what follows just the internal region $r < 2M$.

For the Schwarzschild setting, we want to look at points within the event horizon, $r < 2M$, since, as we said, our understanding of the metric gives an artificial singularity at $r = 2M$. In this region, we have that the metric can be expressed in another way by replacing the coordinate dr^2 by the now more appropriate $d\tau^2$, where

$$\tau = \int_r^{2M} \left(\frac{2M}{u} - 1\right)^{-\frac{1}{2}} du.$$

This coordinate will act as the temporal coordinate. Since $(\frac{2M}{r} - 1) > 0$ because in this case $r < 2M$, the temporal dt coordinate from the original Schwarzschild metric will become a spatial one, in substitution for the removed dr^2. They are a synchronized system of coordinates, and the metric is

$$g = -d\tau^2 + \left(\frac{2M}{r} - 1\right) dt^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2.$$

A straightforward calculation from this yields

$$\sum_{i,j=1}^{3} \beta_{ij} dx^i dx^j = \frac{dr}{d\tau}\left(-\frac{M}{r^2} dt^2 + r d\theta^2 + r \sin^2\theta d\varphi^2\right).$$

Taking into account the Barrow law, and after that the inverse function theorem, we have that

$$\frac{d\tau}{dr} = -\left(\frac{2M}{r} - 1\right)^{-\frac{1}{2}} \implies \frac{dr}{d\tau} = -\left(\frac{2M}{r} - 1\right)^{\frac{1}{2}}.$$

The expression of θ can now be obtained:

$$\theta = \left(\frac{2M}{r} - 1\right)^{-\frac{1}{2}} \left(\frac{2}{r} - \frac{3M}{r^2}\right), \tag{5.8}$$

which is negative in the region $r < \frac{3M}{2}$. This means that we have indeed a singularity which is inevitable after crossing this region, and which cannot be "corrected" by small perturbations. In short, the singularity is physical and real. This results in the astrophysical fact that once a star collapse crosses some threshold (which depends on its initial mass) then it necessarily ends in the formation of a black hole.

On the other side, the case of the flat Robertson-Walker metric for cosmology is way simpler, since we can directly see that $\beta_{ij} = \frac{\dot{a}}{a}\gamma_{ij}$, and then

$$\theta = \frac{3\dot{a}}{a}. \tag{5.9}$$

We know that the singularity should happen at $t \to 0$, so we want to look in the past-oriented direction. The fact that the Universe has been constantly expanding means that $\dot{a}$ is always negative looking into the past. This produces an always negative value for θ and thus, as before, the Big Bang singularity is confirmed and solid.

Conclusions

After all has been said and done, we see that we have been able to get a big picture on the mathematical nature of space-time. We have studied the main building blocks of differential geometry and have applied them to understand the model of space and time that is the general theory of relativity.

We have started from the fundamental tools and definitions from the field of smooth manifolds, and then we have devoted ourselves to the analysis of Riemannian geometry. One thing that we have noted is that in differential geometry, both for smooth and Riemannian manifolds, it is common to define concepts in a manner that seems rather abstract at first, but that makes more sense as the subject is developed. Since there are often various ways to approach a subject, the one that is more intuitive at first is not necessarily the wisest option.

When working on the special and general theory of relativity in this work, we have tried to maintain a balance: since we are interested in its mathematical aspects, we have tried to formulate it by always having in mind the underlying geometric theory, studied in the previous chapters. Nonetheless, we have tried at the same time to explain the most relevant physical properties, in order to see the extent of the revolution that was brought to physics by the appearance of differential geometry (and specially pseudo-Riemannian geometry).

These revolutionary physics are epitomized in the analysis of the Schwarzschild and the Robertson-Walker metrics. We have seen, on the one hand, how a geometric treatment really facilitates the calculations in some situations that were too complex otherwise. On the other hand, we have been able to contemplate how incredibly precise yet mind-blowing is general relativity when it is put into practice.

The final chapter has brought the difficulty to another level, with intricate mathematical reasonings that were both challenging and motivating, and in any case totally worthwhile. We have seen that, while in Riemannian manifolds the condition for geodesic completeness is related to the simple property of metric completeness, in Lorentzian manifolds there is a much more subtle and complex relationship. The Hawking singularity theorem is a big modern milestone in the history of mathematics and physics, which is a great continuation of general relativity and a wonderful way to end this work.

Bibliography

[Be12] L. Benguigui, *A tale of two twins*, arXiv:1212.4414 (2012).

[Cep07] J. Cepa, *Cosmología Física*, Ediciones Akal (2007).

[Ein16] A. Einstein, *Relativity: The Special and General Theory*, New York, H. Holt and company (1916).

[Ein07] A. Einstein, *On the relativity principle and the conclusions drawn from it* (English translation), Jahrbuch der Radioaktivität und Elektronik, 4, 411-462 (1907).

[GoNa14] L. Godinho, J. Natário, *An Introduction to Riemannian Geometry*, Springer (2014).

[Lee03] J. M. Lee, *Introduction to Smooth Manifolds*, Springer (2003).

[Lee18] J. M. Lee, *Introduction to Riemannian Manifolds*, Springer (2018).

[Man93] E. B. Manoukian, *On the reversal of the triangle inequality in Minkowski spacetime in relativity*, European Journal of Physics, Volume 14, Issue 1, pp. 43 (1993).

[Nab88] G. L. Naber, *Spacetime and Singularities: An Introduction*, Cambridge University Press (1988).

[One83] B. O'Neill, *Semi-Riemannian Geometry: With Applications to Relativity*, Academic Press (1983).

[Pe16] P. Petersen, *Riemannian Geometry*, Springer (2016).

[Wa10] R. M. Wald, *General Relativity*, University of Chicago Press (2010).