General Relativity

Julien Larena

Département de physique

Université de Montpellier

September-December 2024

Forewords

These notes are intended as some supporting material for a General Relativity course in the CCP/Astro MSc program at the University of Montpellier. They are *not* meant to be entirely covered during the course as they are much more comprehensive than what we can study in 24 hours. Instead, I intend them to be a set of reference notes that students can use in their future endeavours. Some of the material included here will be covered in class as it is essential, while some applications will be selected depending, in part, from the students' interests and response to the course.

Chapter 2 is a summary of Special Relativity cast in a format and language that suits a smooth transition to General Relativity. It is *not* a Special Relativity course. In particular, it does not study physical situations in great details and does not cover relativistic electrodynamics. Rather, it concentrates on relativistic kinematics, to prepare the stage for the generalisations necessary to include gravitation in a relativistic theory. This means that it should contain only notions already encountered by most students, albeit in a language and form that might be unfamiliar to many.

Chapter 3 is the heart of these notes. First, it describes in some details the technical, mathematical tools of differentiable manifolds and calculus on such manifolds. Attempt has been made to keep this introduction 'rigorous' and, at the same time, easy to follow for physicists and focussed on the material directly related to the development of General Relativity. This has meant some compromise on the generality of discussions, in particular when it comes to the concept of affine connections. Gradually, the general material blends with the developments of General Relativity, taking the equivalence principle and special relativistic kinematics and dynamics as guiding principles. Finally, in its last part, this chapter explains how sources generate the gravitational field,

with a heuristic derivation of Einstein field equations. Chapters 2 and 3 contain a fair amount of mathematical statements. I have put in appendix A some necessary concepts that students might want to refer to when necessary. This is not a mathematics course and therefore, the stress will not be set on mathematics here. Classes will provide students with heuristic ways to continuation to navigate this material without command of higher mathematics. As we will see, a lot of the concept we need from differential geometry and the theory of manifolds are actually just usual concepts of calculus in $\mathbb{R}^n$ cast in a new language, more appropriate to the problems at hand. This is actually baked into what a manifold truly is, so that calculus on manifolds can be roughly summarised as standard calculus applied locally, with some tricks to glue the local calculii together. But I think it is important for more mathematically inclined students to be provided with some consistent story here, even if it is, of necessity, a truncated and simplified one.

The remaining 3 chapters explore the physics in configurations of the gravitational field applicable in particular contexts. Chapter 4 studies the properties of spacetime around isolated objects like stars and black holes. The discussion is limited to the Schwarzschild spacetime, although more realistic situations ought to be formulated in a Kerr setting. Unfortunately, this falls outside of what we can reasonably hope to study here. We start by obtaining the Schwarzschild solution from first principle. Then we study the trajectories of massive particles and photons around a spherical star, concentrating on the historical tests of General Relativity: deviation of light, gravitational redshift, advance of the perihelion of planets. Finally, we see how to extend the Schwarzschild geometry past the event horizon to construct the eternal Schwarzschild black hole.

Chapter 5 concentrates on the definition and study of gravitational waves. It starts with a general introduction to relativistic perturbation theory, a topic that will be central to the Cosmology course, in the second year of the master's programme. This discussion is fairly technical but students should make use of appendix B when necessary and concentrate here on the 'storyline' rather than the technical details. Then, this framework is applied to the free gravitational field in vacuum, i.e. gravitational plane waves. We also discuss what it means to detect gravitational waves by attempting to clarify what exactly happens when such a wave encounters matter. The chapter ends on a derivation of the quadrupole formula that explains how a weak, slowly varying source generates gravitational waves.

Finally, chapter 6 addresses the problem of cosmological solutions. In contrast with what usually appears in General Relativity courses, i.e. a study of various solutions to Einstein field equations

that are, in some loose or often historical sense, 'of cosmological interest', it is explicitly focussed on introducing, in the context of General Relativity, the first building block of our modern cosmological model. As such it differs from the previous chapters of these notes and aims at preparing the readers to the Cosmology course that is the continuation of this one in the second year of the programme.

There are countless books on Special and General Relativity. I can recommend here, a few of them in no particular order. For Special Relativity, [12] is a wonderful book, extremely well adapted to a General Relativity course, as it is written in the 'correct', modern language used here. For General Relativity, there is now a host of very good references, among which I particularly recommend [7, 13, 15, 19]. All of these have different, complementary takes on the topic. Such a diversity of viewpoints is important in a rich field such as General Relativity, which contains developments ranging from highly abstract and mathematical studies to applications is astrophysics. For the historical texts, [16] remains a remarkable reference with a lot of insight on many aspects of the theory. It is however, difficult to read as a course and should be used more as a pointed reference when one knows what one is looking for. [21] will please mathematically inclined readers and is very good for advanced topics. For very advanced topics, [14] is a compulsory reference.

I wish to thank Théo Paret and Lucas Maret, from the 2022/2023 MSc programme, for pointing out numerous typos in the original version those notes.

Notations and conventions

- Lorentzian metrics in 4 dimensions will be written in the $(-,+,+,+)$ signature. Physically, this means that positive spacetime intervals, $\mathrm{d}s^2$ will be spacelike and proper times will be $\mathrm{d}\tau^2 = -\mathrm{d}s^2$.
- 4-vectors will be denoted with boldface letters, capital or not, e.g.: $\boldsymbol{u}$, $\boldsymbol{X}$ etc.
- The same convention will apply to linear maps such as tensor fields but not to (scalar) functions.
- 3-vectors, i.e. the spatial part of 4-vectors will be denoted with an arrow, e.g.: $\vec{v}$ or $\vec{V}$.
- The notes are written in units with the speed of light equal to unity: $c = 1$. In some instances, we put the appropriate powers of c back into important formulæ. Students are encouraged to do that systematically using dimensional analysis. This is a very good exercise.

We list some useful numerical values that one may find useful when working on the content of these notes, especially on chapters 4, 5 and 6. Some of these values are approximate and the values adopted here will suffice to obtain results that are precise enough for our purposes.

1. *Fundamental Constants*:
   - Speed of light: $c = 299\,792\,458\, m \cdot s^{-1} \simeq 3 \times 10^8\ \mathrm{m} \cdot \mathrm{s}^{-1}$
   - Planck Mass: $M_p \simeq 1.67 \times 10^{-27}\ \mathrm{kg}$
   - Boltzmann constant: $k_B \simeq 1.38 \times 10^{-23}\ \mathrm{J} \cdot \mathrm{K}^{-1}$
   - Newton constant: $G \simeq 6.67 \times 10^{-11}\ \mathrm{N} \cdot \mathrm{m}^2 \cdot \mathrm{kg}^{-2}$.
2. *Conversion factors*:
   - $1\ \mathrm{AU} \simeq 1.5 \times 10^{11}\ \mathrm{m}$
   - $1\ \mathrm{eV} \simeq 1.6 \times 10^{-19}\ \mathrm{J}$
   - $1\ \mathrm{pc} \simeq 3 \times 10^{16}\ \mathrm{m}$
   - $1\ \mathrm{sterad} = 1\ \mathrm{rad}^2 = \left(\frac{180}{\pi}\right)^2\ \mathrm{deg}^2$
   - $1\ \mathrm{yr} \simeq 3.16 \times 10^7\ \mathrm{s}$.
3. *Sun's characteristics*:
   - $M_\odot \simeq 2 \times 10^{30}\ \mathrm{kg}$
   - $R_\odot \simeq 7 \times 10^8\ \mathrm{m}$

Contents

1 General Introduction

Contents

1.1 Why study General Relativity?

For a very long time after its formulation by Einstein in 1916, General Relativity remained a fringe subject in physical science. Although it was regarded as an elegant theory of gravitation that passed with remarkable success its first empirical tests, deviations from Newtonian physics in the observable physical world remained very small and largely unattainable by experiments and observations. Besides, its incompatibility with quantum mechanics, which was shaping the rest of physics, made it an awkward subject from a formal point of view. Until the late 1960s, there was no cosmological observations precise enough to test the relativistic model built in the 1930s to 1950s. Black holes where considered a mathematical curiosity and potentially a sign of failure for the theory and gravitational waves a natural prediction that was so faint that it could not be probed in any conceivable way. At the same time, particle and nuclear physics were developing hand-in-hand with remarkable experimental leaps. This context explains why most physicists turned to particle physics while General Relativity was mostly studied in mathematics departments. All this has changed dramatically over the last 40 years.

Cosmology has become a precision science which cannot be understood without General Relativity. Although the local dynamics of matter is everywhere quasi-Newtonian on cosmological scales, our understandings of the early Universe, the formation of structure and the dynamics of the largest scales cannot be understood in a Newtonian context. Black Holes have been observed, both small and (very) big, albeit indirectly, i.e. by the effect they have on surrounding matter. But this means that the most extreme objects predicted by General Relativity are now becoming part of astrophysics so that General Relativity is finding its way into astrophysics. And finally, gravitational waves have been detected and measured, vindicating yet again General Relativity, while opening a new window on the Universe; in a few decades, gravitational wave astrophysics will be part of the way we probe our Universe.

Clearly, it is now a wonderful time for young physicists to study General Relativity.

1.2 Structure of the course

This course takes a clear physical approach the General Relativity by choosing to present the theory through its manifestations in three iconic regimes:

- The gravitational field around stars and black holes; chapter 4. This is the occasion to develop

the tools to understand the standard tests of General Relativity: deviation of light by the Sun, advance of the perihelion of Mercury and gravitational spectral shift. Unfortunately, lack of time means that we cannot approach other important effects, e.g. the frame-dragging due to rotation or the technology of the GPS. Black holes are also studied in their simplest, unrealistic form i.e. without any rotation. It is an example where introducing rotation and going from the Schwarzschild to the Kerr metric requires significant technical and conceptual jumps. General Relativity is so complex that this is often the case: straying from the simplest, idealised situation might prove terribly difficult.

- Gravitational waves; chapter 5. These are the analogue in General Relativity of electromagnetic waves in electrodynamics. We will see that they are produced by quadrupolar motions at least and that they propagate into two polarisation modes. We will also characterise their physical effects, something that is often the occasion for some confusions that we will try to clarify.

- Cosmology and the dynamics of the Universe on very large scales; chapter 6. This chapter is a preparation for the advanced course on cosmology that is given in the second year of the MSc program. It introduces the homogeneous and isotropic description we use when dealing with the large scale dynamics of the Universe. It is the occasion to understand what distances are in relativistic cosmology and to understand the importance of past lightcones. It also allows us to introduce the matter-energy content of the Universe and discuss its thermal history.

In addition, we try not to completely sacrifice the mathematical elegance of the theory and we go to some length presenting this material in chapter 2 and in appendix B. This will be presented much more briefly in class than it is done here in the notes, the goal being to be ready to attack the physics chapter. Students are thus encouraged to read chapter 2 while focusing on what is discussed in class, and skipping the unnecessary digressions. However, the interested and/or mathematically inclined reader is welcome to pay more attention to the details and to ask for more explanation and resources when needed. Finally, the notes contain a chapter on Special Relativity, chapter 1. Once again, this chapter contains much more than what we will talk about in class. It is an attempt at reformulating Special Relativity in a way that makes it easy to jump to General Relativity and that is what interests us here. So it may differ in places from the way things were presented to you in L3, when the emphasis is usually on the electrodynamical side of things.

To work on this course and prepare for the exam, in addition to understanding the concepts and being able to explain them, you need to be able to reproduce some parts of this notes that will be indicated to you in class. Some calculations and developments will also be 'left to the students'. This is meant seriously and is examinable. Finally, you will also receive regular problem sheets with exercises and problems that will require you to apply what you've learnt and , hopefully, help you learn.

2

Special Relativity

Contents

2.1 Relativity in Newtonian physics

2.1.1 The concept of relativity through history

The concept of relativity has been central to our pictures of the natural world for a very long time, at least since the work of Aristotle (384 BCE-322 BCE) but certainly well before that, including in other traditions. It grew from attempts at formalising the concept of motion, when addressing the rather mundane question: "motion relative to what?" It should not be confused with relativism, which, in its various guises, is usually a statement about the ontology of our discourse rather than an attempt at formulating a theory of space and time, which is what will be at the centre of these notes. Of course, these questions are not independent, but their intertwining is subtle and relativism as to motion does not necessarily or simply correlate with ontological relativism (or any other derived form of it).

Mostly, we can distinguish two broad classes of attitudes in the description of motion. Either the concepts of space and time are absolute and motion must always, *in fine* be referred to these absolute yardsticks; or they are relational concepts bereft of an absolute reference. In the first class we find Plato, Kant and Newton; in the second we have Aristotle, Descartes and Leibniz but also, of course, Einstein. In this section, I would like to spend some time reviewing what are space, time and motion in Newtonian mechanics with a modern viewpoint on these concepts in order to contrast them with their formulation in special and general relativity. It would be interesting to expand on the concept of relativity of motion prior to the Newtonian revolution and this might be included here in a future iteration of these notes.

For now, let us begin our story with Newtonian mechanics.

2.1.2 Newtonian spacetime

The fundamental laws of Newtonian mechanics, Newton's three laws, are formulated in a very specific setting for spacetime, one which allows to use vector analysis and simple calculus and place these formalisms at the heart of mechanics. The central notions are *absolute* space and time as presented by Newton in his *Philosophiæ Naturalis Principia Mathematica* (1687) [17]:

Newtonian space and time

- **Absolute space**, in its own nature, without regard to anything external, remains always similar and immovable. Relative space is some movable dimension or measure of the absolute spaces; which our senses determine by its position to bodies: and which is vulgarly taken for immovable space [...] Absolute motion is the translation of a body from one absolute place into another: and relative motion, the translation from one relative place into another.

- **Absolute, true and mathematical time**, of itself, and from its own nature flows equably without regard to anything external, and by another name is called duration: relative, apparent and common time, is some sensible and external (whether accurate or unequable) measure of duration by the means of motion, which is commonly used instead of true time.

We will return later to the notion of relative space. What matters here is that Newton postulates the existence of absolute space and time, which remain identical to themselves *without regard to anything external*. Such entities, rooted in Newton's religious beliefs, were quite revolutionary at the time. In any case and in practise, what it means is that the theatre of nature unfolds in a fixed set-up that fits quite naturally in basic mathematical structures:

- **Space**, E, is Euclidean, i.e. that it is such that the shortest distance between two points is given along the straight line connecting those points, and the sum of the angles of a triangle always equals π. It is also infinite and without boundary. Once an origin has been chosen, vectors and couples of points can be identified and E can be represented by a vector space of dimension 3. The Euclidean nature of space means that it can be provided with a *scalar product* $\langle \cdot, \cdot \rangle$, which takes two vectors as inputs and returns a real number. As a scalar product, it has a few important properties:

  1. it is *bilinear*: $\forall (X, Y, Z) \in E^3$, $\forall (\lambda, \mu) \in \mathbb{R}^2$, $\langle \lambda X + \mu Y, Z \rangle = \lambda \langle X, Z \rangle + \mu \langle Y, Z \rangle$ and $\langle X, \lambda Y + \mu Z \rangle = \lambda \langle X, Y \rangle + \mu \langle X, Z \rangle$;
  2. it is *symmetric*: $\forall (X, Y) \in E^2$, $\langle X, Y \rangle = \langle Y, X \rangle$;
  3. it is *positive*: $\forall X \in E$, $\langle X, X \rangle \geq 0$;

4. it is *definite*: $\forall X \in E,\ \langle X, X\rangle = 0 \Rightarrow X = 0$.

We say that a vector $X \in E$ has length $\|X\| = \sqrt{\langle X, X\rangle}$ and that two non-zero vectors $(X, Y) \in E^2$ form an angle γ given by:

$$\cos\gamma = \frac{\langle X, Y\rangle}{\|X\|\|Y\|} \ . \tag{2.1}$$

Once given a basis of E, it is linearly isomorphic to $\mathbb{R}^3$. Let us call $\{e_1, e_2, e_3\}$ such a basis that, without loss of generality, we will assume orthonormal, that is we will assume:

$$\langle e_i, e_j\rangle = \delta_{ij} \ , \tag{2.2}$$

where δ_{ij} is the Kronecker symbol, equal to 1 if $i = j$ and 0 otherwise. Such a basis is often called *Cartesian*. Any vector $X \in E$ can then be uniquely represented by its *components*:

$$\forall X \in E, \exists! \left(X^1, X^2, X^3\right) \in \mathbb{R}^3,\ X = X^1 e_1 + X^2 e_2 + X^3 e_3 \ . \tag{2.3}$$

The length of a vector is then simply:

$$\|X\| = \sqrt{\left(X^1\right)^2 + \left(X^2\right)^2 + \left(X^3\right)^2} \ , \tag{2.4}$$

which is nothing but the Pythagorean theorem in 3 dimensions, and the scalar product of two vectors is:

$$\langle X, Y\rangle = X^1Y^1 + X^2Y^2 + X^3Y^3 = \sum_{i,j} \delta_{ij} X^i Y^j \ . \tag{2.5}$$

This can be written:

$$\|X\|^2 = \delta_{ij} X^i X^j \ , \tag{2.6}$$

using *Einstein's summation convention*, which consists in assuming that an index repeated as subscript and superscript in an expression is a dummy index that we have to sum over all its possible values. The object δ_{ij} is called the representation fo the Euclidean *metric* on E in the Cartesian basis $\{e_1, e_2, e_3\}$. when convenient, we will denote $\langle\cdot,\cdot\rangle$ as usual as the *dot product*:

$$\langle X, Y\rangle = X \cdot Y. \tag{2.7}$$

or using the Kroneker symbol when using indices:

$$\langle X, Y\rangle = \delta_{ij} X^i Y^j \ . \tag{2.8}$$

Moreover, we will use arrows to denote 3 dimensional vectors. Beware that Euclidean space can have a metric whose components are not simply δ_{ij} if we use a basis that is not Cartesian. Think of the spherical basis, for example. Let us now imagine two points $P \in E$ and $Q \in E$ that are infinitesimally close, i.e. such that $\overrightarrow{PQ} = \mathrm{d}x^i e_i$ with $|\mathrm{d}x^i| \ll 1$, then we can construct the quadratic quantity called the *line element* associated with the Euclidean metric:

$$\mathrm{d}s^2 = \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ , \tag{2.9}$$

such that $\mathrm{d}s$ is the infinitesimal length of $\overrightarrow{PQ}$. It will prove a very useful object in what follows.

- **Time** is simply a coordinate on the Euclidean line L: $t \in \mathbb{R}$. It is fixed by choosing an origin on the line and a basis vector. Equivalently, we need a single function $T : \mathbb{R} \to L$ such that $T(t) = P \in L$ which is continuous and bijective (hence necessarily monotonous).

In this framework, one can formulate Newton's three laws of mechanics, the principle of inertia, the second law that links acceleration and forces and the law of action and reaction.

2.1.3 Newtonian relativity

What I call here Newtonian relativity emerged slowly from a set of ideas to interpret a certain number of facts and principle usually subsumed in physics under the term of "Galilean invariance" as it was first formulated in a 'modern' way in Galileo's *Dialogue Concerning the Two Chief World Systems* (1632).

Newton, starting from Galileo's principles of equivalence and of inertia, postulated his famous law of inertia: if an object is not subject to any force, then the object will remain at rest (velocity $\vec{v} = \vec{0}$) or in constant, straight line motion (velocity $\vec{v}$ constant but non-zero). But there is something we left 'under the rug' here: the law of inertia mentions that objects will remain at rest or move with constant velocity, in absence of any force acting upon them. But what is to be understood as 'rest' and constant motion? These concepts are, after all, relative: one is at rest or in constant motion relatively to something else. Think about a car on a straight road, and a cow standing in a field next to the road: for the cow, certainly the poles along the field are at rest, and the car moves relative to them, but for the driver inside the car, all the parts of the car are at rest, while the cow and the poles move. In a nutshell, that means that the new mechanics formulated by Galileo and Newton

introduces a high degree of relativism; the notions of motion and rest have to be defined relative to *reference frames*. In which reference frames do objects not subjected to any force remain at rest or in uniform translation? These are known as *inertial frames*. Their nature can be enshrined in a:

Principle of relativity

Physical laws are identical when expressed in inertial frames.
In other words, identical physical experiments carried out in different inertial frames lead to identical results.

However, none of this tells us how to find or construct inertial frames. At the time of Newton, it had been known for a very long time (at least since the ancient Egyptians), that some stars visible in the sky did not apparently move with respect to our Sun. They were called the fixed stars. The idea of Newton was thus to define inertia with respect to his *absolute space* and to anchor this one by choosing the centre of the Sun as origin[1], and by picking three fixed stars. The three lines through the centre of the Sun passing through each of these three fixed stars thus defined a frame that could be used as a reference: the motion of objects with respect to this frame could determine whether or not the motions were inertial. Indeed, if the object is at rest or moving along a straight line at constant speed with respect to this absolute space, the motion can be said to be inertial. If it is not the case, then it is the sign that a force (or the combination of several forces) is acting on the object and makes its trajectory deviate from an inertial motion.
It is important to realize that, for Newton, absolute space was really absolute: the fixed stars were just practical means of identifying this space with a given physical reference frame. Nevertheless, for all practical purposes, this subtlety does not matter. The only relevant idea here is that inertia is only defined with respect to a reference frame.
Absolute space is an inertial frame by construction but what are the other ones? To define a reference frame, we pick up a point $O \in E$ that we call the origin. Through that point, we draw three infinite lines perpendicular to each other that we call the axes, D_1, D_2 and D_3. Any point M in E can then be described uniquely by a triplet of real numbers (x, y, z) corresponding to the distances of the orthogonal projections of M on each axis to the origin O. (x, y, z) are then called the coordinates of M in the frame $\mathcal{R}$. The set (O, D_1, D_2, D_3) is a (Cartesian) frame of reference. Let us call it

[1] Actually the centre of mass of the Solar System which differs slightly from the centre of the Sun.

$\mathcal{R}$. If we change the origin and consider a new point, $O^{'}$ as the origin, we can still define three perpendicular lines $D_1^{'}$, $D_2^{'}$ and $D_3^{'}$ through $O^{'}$ parallel to D_1, D_2 and D_3 respectively. Hence, we obtain a new frame of reference, $\mathcal{R}^{'}$. If the coordinates of $O^{'}$ in $\mathcal{R}$ are (a, b, c), and the coordinates of M in $\mathcal{R}$ are (x, y, z), it is easy to show that the coordinates of M in $\mathcal{R}^{'}$ are given by:

$$\begin{cases} x^{'} &= x - a \\ y^{'} &= y - b \\ z^{'} &= z - c \ . \end{cases} \tag{2.10}$$

Of course, we can change from $\mathcal{R}$ to $\mathcal{R}^{'}$ using the vector $\overrightarrow{OO^{'}} = (a, b, c)$ simply by writing the standard vectorial relation:

$$\forall M \in E \ , \overrightarrow{O^{'}M} = \overrightarrow{OM} - \overrightarrow{OO^{'}} \ . \tag{2.11}$$

This is a particular transformation called *translation*. If we consider a vector that is not a position vector, for example, a force $\vec{F}$ acting on a point $\vec{x}$, with a certain amplitude, the end point and the starting point of $\vec{F}$ will be translated by the same amount, so that the force itself remains unaffected by the translation.

One can also alter the axes of the frame while keeping the origin fixed. Consider the origin O. Instead of choosing D_1, D_2 and D_3 as axes, we could choose three other lines Δ_1, Δ_2 and Δ_3, perpendicular to each other. To go from D_1, D_2 and D_3 to Δ_1, Δ_2 and Δ_3, we need to apply a rotation, which is fully characterised by three real numbers, θ, ϕ and ξ, called the Euler angles. We have $(\theta, \xi) \in]-\pi, \pi[^2$, and $\phi \in [-\pi/2, \pi/2]$. This can be represented by a matrix $R(\theta, \phi, \xi)$ acting on vectors in E. If we denote by (x, y, z) the coordinates of a point $M \in E$ in the frame $\mathcal{R}$, and (x', y', z') the coordinates of M in the rotated frame $\mathcal{R}_\Delta$, we have:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R(\theta, \phi, \xi) \begin{pmatrix} x \\ y \\ z \end{pmatrix} , \tag{2.12}$$

where the matrix $R(\theta, \phi, \xi)$ is given by:

$$R(\theta, \phi, \xi) = \begin{bmatrix} \cos(\phi)\cos(\xi) & -\cos(\theta)\sin(\xi) + \sin(\theta)\sin(\phi)\cos(\xi) & \sin(\theta)\sin(\xi) + \cos(\theta)\sin(\phi)\cos(\xi) \\ \cos(\phi)\sin(\xi) & \cos(\theta)\cos(\xi) + \sin(\theta)\sin(\phi)\sin(\xi) & -\sin(\theta)\cos(\xi) + \cos(\theta)\sin(\phi)\sin(\xi) \\ -\sin(\phi) & \sin(\theta)\cos(\phi) & \cos(\theta)\cos(\phi) \end{bmatrix} . \tag{2.13}$$

Actually, the set of all matrices $R(\theta, \phi, \xi)$ for all the possible values of θ, ϕ and ξ, together with the standard matrix multiplication forms a group (the group of rotations in three dimensions), and the matrices are orthogonal:

$$\forall(\theta, \phi, \xi) \in]-\pi, \pi[\times[-\pi/2, \pi/2]\times]-\pi, \pi[, R(\theta, \phi, \xi)^T = R(\theta, \phi, \xi)^{-1} . \tag{2.14}$$

The principle of relativity states that Newtonian physics is invariant (i.e. to remain the same) when going from one inertial frame to another. Mathematically, one can show that the equations governing mechanics (Newton's second law) keep the same form if we translate and rotate the spatial coordinate system, *as long as the rotations are independent on time and translations uniform*. We already saw that by construction, objects free of forces were in uniform translation with respect to absolute space by definition. Thus, any of these objects and a set of three axes rotated with respect to the fixed stars define an inertial frame in which Newton;'s law can be applied.
The transformations between inertial frames are given the name of Galilean transformations.

Galilean transformation

Consider a reference frame $\mathcal{R} = (O, x, y, z, t)$ associated to an observer O. Let $\vec{v}$ be the constant velocity of an observer O' in motion with respect to O. Suppose that O and O' coincide at $t = 0$. Then, the position $(\vec{x}', t')$ of a point particle in the reference frame $\mathcal{R}' = (O', x', y', z', t')$ associated with O' can be deduced from position $(\vec{x}, t)$ of the same particle in $\mathcal{R}$ by the *Galilean transformation*:

$$\begin{cases} \vec{x}' &= R\vec{x} - t\vec{v} \\ t' &= t, \end{cases} , \tag{2.15}$$

where R is an arbitrary time-independent rotation.

If $\mathcal{R}$ is inertial, then so is $\mathcal{R}'$.

The last statement follows from the fact that Newton's second law is invariant under Galilean

transformations. Let us start with Newton's second law is R':

$$\vec{F}' = m\frac{d^2\vec{x}'}{dt'^2} = m\frac{d}{dt}\frac{d}{dt}(R\vec{x} - t\vec{v}) \tag{2.16}$$

$$= m\frac{d}{dt}\left(R\frac{d\vec{x}}{dt} - \vec{v}\right) \tag{2.17}$$

$$= mR\frac{d^2\vec{x}}{dt^2} = R\vec{F}\ . \tag{2.18}$$

Note that the set of Galilean transformations is actually a group:

Galilean group

The set of Galilean transformations $\mathcal{G}$, together with the standard composition of linear functions, form a group, called the Galilean group. An element of this group will be denoted $(R, \vec{v})$, where $R \in O(3)$, and $\vec{v} \in E$

Frames that are linked via Galilean transformations are called sometimes called preferred instead of *inertial* frames. They are very important, since they form the set of frames in which the laws of mechanics keep an invariant form. As is well-known, in non-inertial frames, inertial forces appear, which in Newtonian mechanics, are a trace of the motion of the frame relative to absolute space, as argued by Newton in his famous bucket experiment.

2.2 Enters electrodynamics

This picture of mechanics, now known as Newtonian mechanics, remains a beautiful achievement, certainly one of the most perfect scientific theory ever written, both for its aesthetic characteristics and its observational successes; after all, most of the phenomena that occur around us at human scales are accounted for in the framework of Newtonian mechanics with extremely high precision. Therefore, the downfall of this theory did not come from a failure of it to account for one or several observations and/or experiments; it came from the emergence of theoretical problems in the process of unification of mechanics with electromagnetism. Unification, even though it existed before, is an idea that has been central to physics since the beginning of the twentieth century and continue to be one of the main motivation behind the work of most theoretical and mathematical physicists. It is thus worth a few sentences in this course. Unification can roughly be described as the tentative to rid science from the emergence of separate theories describing seemingly separate natural phenomena.

Therefore, it relies on the idea that Nature is unique and should therefore obey a unified set of laws: subsuming the motions of stones on earth and of planets in the sky by postulating that they are all subject to a unique force, as Newton did, is clearly a process of unification. In a way unification as always been a guiding principle for scientists, but in the past, only very religious and/or mystical minds have made it a central ingredient of their approach to research; most scientists adopted a more pragmatic approach based on a principle of efficiency. The work of Einstein on relativity changed that. Suddenly, unification appeared as a very important guiding principle that could lead to deep changes in the way to do science: according to that idea, a disparity in the theoretical ways to treat separate physical phenomena is the sign that our theories must be amended to account for all the phenomena at a time, in a single coherent framework. The biggest success of this approach is clearly to be found in modern particle physics that accounts for the behaviour of all the matter around us with simply 12 elementary particles, and three different interactions: particle physicists have successfully unified the electromagnetic and weak interactions and work hard on integrating the strong nuclear force in this unification. The unification of gravity with other forces seems more out of reach for now, but string theorists, among others, have sought this dream for over four decades. This should show that a principle of unification, whatever its philosophical justifications might be for each individual adopting it, is indeed a powerful guideline for research. Special Relativity is the first unambiguous situation in which such a principle has been applied with success, and this should be an additional motivation to study its structure.

2.2.1 Maxwell's theory

In the late 1860s, J.C. Maxwell formulated a final version of his theory of electromagnetism phenomena. Electricity and magnetism had been studied for some time, but Maxwell's theory was the first complete theory to account for all the phenomenology of the time. In this setting, electricity and magnetism are unified in the electromagnetic field, consisting of two vectors, $\vec{E}$ and $\vec{B}$, such that, in vacuum:

$$\mathrm{div}\left(\vec{E}\right) = 0 \text{ (Gauss's law)} \qquad \vec{\mathrm{curl}}\left(\vec{E}\right) = -\frac{\partial \vec{B}}{\partial t} \text{ (Faraday's law)} \tag{2.19}$$

$$\mathrm{div}\left(\vec{B}\right) = 0 \qquad \vec{\mathrm{curl}}\left(\vec{B}\right) = \frac{1}{c^2}\frac{\partial \vec{E}}{\partial t} \text{ (Ampère's law).} \tag{2.20}$$

The operators divergence (div.) and curl ($\vec{\text{curl}}$) are defined, for a vector $\vec{f} = (f_x, f_y, f_z)$. In a Cartesian coordinate system (x, y, z) by:

$$\text{div}\vec{f} = \frac{\partial f_x}{\partial x} + \frac{\partial f_y}{\partial y} + \frac{\partial f_z}{\partial z} \tag{2.21}$$

$$\vec{\text{curl}}\vec{f} = \begin{pmatrix} \frac{\partial f_z}{\partial y} - \frac{\partial f_y}{\partial z} \\ \frac{\partial f_x}{\partial z} - \frac{\partial f_z}{\partial x} \\ \frac{\partial f_y}{\partial x} - \frac{\partial f_x}{\partial y} \end{pmatrix} . \tag{2.22}$$

These equations are known as Maxwell's equations in vacuum. They can be generalized to equations in a medium by introducing appropriate source terms. In vacuum, they lead to a system of two, independent but second order equations for each field:

$$\Delta\vec{E} - \frac{1}{c^2}\frac{\partial^2\vec{E}}{\partial t^2} = 0 \tag{2.23}$$

$$\Delta\vec{B} - \frac{1}{c^2}\frac{\partial^2\vec{B}}{\partial t^2} = 0 . \tag{2.24}$$

One recognizes wave equations, and the description of standard electromagnetic waves, the state of the electromagnetic field in vacuum (Δ is the usual vector Laplacian operator).

Finally, the electromagnetic force on a point particle of electric charge q and velocity $\vec{v}$, due to the electromagnetic field is given by Lorentz's law:

$$\vec{F} = q\left(\vec{E} + \vec{v} \wedge \vec{B}\right) . \tag{2.25}$$

Lorentz's law and Maxwell's equations are the only five laws necessary to describe the behaviour and the influence of the electromagnetic field. Maxwell's theory and Newton's mechanics together allowed to describe the whole of physics at the time. So where did the problem was?

2.2.2 Incompatibility with Galilean invariance

The problem was actually in the compatibility of both systems of laws. We have seen that Newton's mechanics is deeply rooted in its invariance under Galilean transformations: the laws of Newtonian mechanics keep identical forms in any two reference frames related by a Galilean transformation (i.e. the composition of a translation and a rotation). This isn't true of Maxwell's laws: their form changes between the two frames. Indeed, consider the equation for $\vec{E}$. Let $R = (0, t, \vec{x})$ and $\hat{R} = \left(0, \hat{t}, \vec{\hat{x}}\right)$ be

two frames such that $\hat{R}$ moves, relative to R with a constant velocity $\vec{v}$. Then, if we postulate that the law of invariance between frames is Galilean, we have that:

$$\vec{\hat{x}} = \vec{x} - t\vec{v} \tag{2.26}$$

$$\hat{t} = t\,, \tag{2.27}$$

where we assumed that the axes in both frames are aligned, for simplicity. Then, since it is a vector, the electromagnetic field is invariant when going from one frame to the other:

$$\vec{\hat{E}}\left(\hat{t}, \vec{\hat{x}}\right) = \vec{E}\left(t, \vec{x}\right)\ . \tag{2.28}$$

Finally, for any function $F : \mathbb{R}^4 \to \mathbb{R}$, noting as well $\hat{F}\left(\hat{t}, \vec{\hat{x}}\right) = F\left(t, \vec{x}\right)$, we can write the differential in two ways:

$$\mathrm{d}F = \frac{\partial \hat{F}}{\partial \hat{t}}\mathrm{d}\hat{t} + \sum_{i=1}^{3}\frac{\partial \hat{F}}{\partial \hat{x}^i}\mathrm{d}\hat{x}^i \tag{2.29}$$

$$= \frac{\partial F}{\partial t}\mathrm{d}t + \sum_{i=1}^{3}\frac{\partial F}{\partial x^i}\mathrm{d}x^i\ . \tag{2.30}$$

Then, using the Galilean laws of invariance, we get:

$$\frac{\partial^2}{\partial t^2}\cdot = \frac{\partial^2}{\partial \hat{t}^2}\cdot - 2\sum_{i=1}^{3} v^i \frac{\partial^2}{\partial t \partial \hat{x}^i}\cdot + \sum_{i,j=1}^{3} v^i v^j \frac{\partial^2}{\partial x^i \partial x^j}\cdot \tag{2.31}$$

$$\frac{\partial^2}{\partial {x^i}^2}\cdot = \frac{\partial^2}{\partial {\hat{x}^i}^2}\ . \tag{2.32}$$

We see that the wave equation in the frame $\hat{R}$ reads:

$$\Delta_{\vec{\hat{x}}}\vec{\hat{E}} - \frac{1}{c^2}\frac{\partial \vec{\hat{E}}}{\partial \hat{t}} = \left[2\sum_{i=1}^{3}\frac{v^i}{c^2}\frac{\partial^2}{\partial \hat{t}\partial \hat{x}^i}\cdot + \sum_{i,j=1}^{3}\frac{v^i v^j}{c^2}\frac{\partial^2}{\partial \hat{x}^i \partial \hat{x}^j}\cdot\right]\vec{\hat{E}} \neq \vec{0}. \tag{2.33}$$

It is therefore *not invariant under the group of Galilean transformations*. This means that, in principle, by carrying experiments on the electromagnetic field, two observers in relative uniform motion could tell which one is in which frame. In other words, electromagnetic phenomena would single out preferred observers among the inertial ones. That is in gross contradiction with the principle of inertial frames. Another, equivalent, way of formulating this problem is to consider the speed

of electromagnetic waves. In Maxwell's theory, it is c in any frame, since Maxwell's equations can be written with respect to any coordinate system. If it were a true Newtonian velocity, then it should follow the standard law of transformation of velocities. But it doesn't. And repeated experiments to check the effect have led to negative results: the speed of light remains the same in any reference frame. The fact that the laws of electromagnetism are not invariant under Galilean transformations thus introduced a big tension in the formulation of 19^{th} century theoretical physics: mechanical phenomena had, associated to them, inertial, preferred frames in which the laws of mechanics remained invariant, whereas electromagnetic phenomena could tell the difference between rest and uniform motion. Special Relativity was invented to circumvent this particular, un-æsthetic, situation. Einstein solved the problem by identifying a new law to go from one inertial frame to another one, a law that, in particular 'mixes' space and time coordinates. Since this is not a course on Special Relativity, we will simply remind the reader of the main properties and results of special relativity, in a language adapted to the jump to General Relativity.

2.3 Minkowski spacetime

2.3.1 Constructing Minkowski spacetime

The basic building block of the theory is the concept of *event*: it is a physical occurrence that has no spatial extension and no duration in time. The permanence of a material particle, in this framework is then simply the existence of a continuous sequence of events called the *worldline* of the particle. We will denote by $\mathcal{M}$ the (abstract) set of all the events, called *Minkowski spacetime* and in the following, we will give a certain number of physical principles that will be used in the remainder of the text to provide a mathematical structure to $\mathcal{M}$.
The correct way to begin is to postulate that events are observables. In particular, we will select a particular class of observers that we will call *admissible*:

Admissible observers

To each admissible observer, on can attach a 3-dimensional, right-handed, Cartesian spatial coordinate system based on an agreed unit of space relative to which photons propagate rectilinearly in any direction.

The key point here is the *isotropy* of the light propagation. Notice that the rectilinearity of the

propagation is relative to a Cartesian coordinate system: in a rotating coordinate system, light might not follow straight lines.
This first definition gives a notion of space around each observer. We can now introduce a notion of time:

Local time

Each admissible observer is given an ideal standard clock based on an agreed unit of time according to which one can provide a quantitative temporal order to the events along the observer's worldline.

Here, the main point is that temporal order is only given *along the observer's worldline*: observer's cannot, yet, decide of the temporal order of events that are spatially separated from their own position. In order for them to do that, we will need a procedure to allow observers to compare their respective clocks; this is often called *synchronisation*. It turns out that this is a real problem, and that most of the effects of Special Relativity come from there. The idea to establish a way to synchronise clocks is to find a way for observers to communicate the results of reading their respective clocks. Light signals are most reliable communication signals because of the following experimental result:

Constancy of the speed of light

For an arbitrary admissible observer, the speed of light in vacuum as determined by the Fizeau procedure is independent of when the experiment is performed, the arrangement of the apparatus, the frequency of the signal and has the same numerical value c for all such observers.

The Fizeau procedure is a specific way of measuring the speed of light. It is described in any good physics book on Special Relativity. For the purpose of this course, we will only retain the fact that there exists a phenomenon that is characterised by its constant speed in any admissible frame of reference. As a matter of fact, in the remainder of this course, unless necessary for numerical evaluations, we will choose units of space and time such that $c = 1$ (geometrised units). In its own Cartesian spatial frame, an (admissible) observer determines time by synchronizing clocks using the following procedure:

- It has its own clock at the centre O of its frame.

- At each point P of its spatial coordinate system, it places a clock identical to the one at the origin.
- At a given time $t \in \mathbb{R}$ as read at O, it emits a spherical light signal.
- As the wave front encounters P, the clock placed at P is set at the time $t + \|\overrightarrow{OP}\|$ and set ticking, where $\|\overrightarrow{OP}\|$ is the Euclidean distance between O and P in the 3-dimensional space.

Let us consider an observer O with its coordinate frame $\left(x^0, x^1, x^2, x^3\right)$, where x^0 is the time as measured by the observer using his clock (procedure above), and $\left(x^1, x^2, x^3\right)$ are the Cartesian, spatial coordinates. Let $\hat{O}$ with coordinate frame $\left(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3\right)$ be another observer. Consider an event $\mathcal{E}$. It has coordinates in both frames. How are they related? In other words, what can we say about the mapping:

$$\mathcal{F} \quad : \quad \left\{ \begin{array}{ccc} \mathbb{R}^4 & \to & \mathbb{R}^4 \\ \left(x^0, x^1, x^2, x^3\right) & \mapsto & \mathcal{F}\left(x^0, x^1, x^2, x^3\right) = \left(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3\right) \end{array} \right. ? \tag{2.34}$$

First it must be *bijective*, so that one can go unambiguously from any admissible observer to the other one. Moreover, we will require an additional *causality* condition:

Causality condition

Any two (admissible) observers agree on the temporal order of any two events on the worldline of a photon. In other words, if two events along the worldline of a photons have coordinates $\left(x^0, x^1, x^2, x^3\right)$ and $\left(y^0, y^1, y^2, y^3\right)$ for O and $\left(\hat{x}^0, \hat{x}^1, \hat{x}^2, \hat{x}^3\right)$ and $\left(\hat{y}^0, \hat{y}^1, \hat{y}^2, \hat{y}^3\right)$ for $\hat{O}$, then $y^0 - x^0$ and $\hat{y}^0 - \hat{x}^0$ have the same sign.

We have not assumed which sign it should be, just that it should remain invariant by a transformation from one admissible coordinate system to another one. This means that $\mathcal{F}$ preserves order in the temporal coordinate.

Since photons propagate rectilinearly with constant speed 1, according to the principles stated above, two events on the worldline of a photon have coordinates with respect to O which satisfy:

$$\forall i \in \{1, 2, 3\},\ y^i - x^i = v^i \left(y^0 - x^0\right) , \tag{2.35}$$

for some constants v^i such that $\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 = 1$. This results in the following equation:

$$\left(y^1 - x^1\right)^2 + \left(y^2 - x^2\right)^2 + \left(y^3 - x^3\right)^2 - \left(y^0 - x^0\right)^2 = 0 . \tag{2.36}$$

This is the equation of a *cone* in $\mathbb{R}^4$ with vertex at (x^0, x^1, x^2, x^3). Of course, this short calculation must be valid in any admissible frame of reference, so that $\mathcal{F}$ must preserve the cone defined by Eq.(2.36), and map it into the cone:

$$\left(\hat{y}^1 - \hat{x}^1\right)^2 + \left(\hat{y}^2 - \hat{x}^2\right)^2 + \left(\hat{y}^3 - \hat{x}^3\right)^2 - \left(\hat{y}^0 - \hat{x}^0\right)^2 = 0. \tag{2.37}$$

Remarkably, this is all that is needed in order to fully characterise all the possible transformations $\mathcal{F}$, and the geometric structure of Special Relativity.

2.3.2 Metric structure on Minkowski spacetime

The few principles listed above and the conical structure of the set of photon paths are enough to formalise the structure of spacetime in Special Relativity.

Minkowski spacetime

Mikowski spacetime $\mathcal{M}$ is a set of points called *events*. It can be given the structure of a *4-dimensional vector space over* $\mathbb{R}$ whose vectors are the directed pairs of events[a]. On this vector space, is defined an *inner product* $\boldsymbol{\eta}$ *of index 1*, i.e. with signature $(-1, 1, 1, 1)$.
and $\boldsymbol{\eta}$ is usually called a *Lorentzian inner product* on $\mathcal{M}$.
There exists an (η-)orthonormal basis $\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)}\}$ of $\mathcal{M}$ such that, for any $(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{M}^2$, with $\boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)}$ and $\boldsymbol{w} = w^\mu \boldsymbol{e}_{(\mu)}$:

$$\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) = -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3 \ . \tag{2.38}$$

[a] Strictly speaking, this would give us an affine space. It would be a vector space only after identifying arrows with the same direction and length. We will not worry about this subtleties here.

The basis$\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)}\}$ is to be regarded as a frame of reference, and the coordinates of the vectors, $\boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)}$ are then identified with the time (v^0) and spatial position (v^1, v^2, v^3) of the events corresponding to the vector v, according to the principles above, i.e. they have to be attached to an observer carrying the frame of reference. This identification will become clear as we proceed further. Given the orthonormal basis $\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)}\}$, we know that, by definition:

$$\forall (\mu, \nu) \in \{0, 1, 2, 3\}^2, \boldsymbol{\eta}\left(\boldsymbol{e}_{(\mu)}, \boldsymbol{e}_{(\nu)}\right) = \eta_{\mu\nu} \ , \tag{2.39}$$

where:

$$\eta_{\mu\nu} = \begin{cases} -1 & \text{if } \mu = \nu = 0; \\ 1 & \text{if } \mu = \nu \neq 0 \\ 0 & \text{otherwise .} \end{cases} \tag{2.40}$$

The $\eta_{\mu\nu}$'s can be seen as the components of the $(0,2)$ tensor associated with $\boldsymbol{\eta}$, in the basis $\{\boldsymbol{e}_{(\mu)}\}$, and we have:

$$\forall(\boldsymbol{v},\boldsymbol{w}) \in \mathcal{M}, \boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)}, \boldsymbol{w} = w^\mu \boldsymbol{e}_{(\mu)} \Rightarrow \boldsymbol{\eta}(\boldsymbol{v},\boldsymbol{w}) = \eta_{\mu\nu} v^\mu w^\nu \ . \tag{2.41}$$

Moreover, for any $\boldsymbol{v} \in \mathcal{M}$, the function $\boldsymbol{\eta}(\boldsymbol{v},\cdot) : \mathcal{M} \to \mathbb{R}$ belongs to the dual of $\mathcal{M}$, $\mathcal{M}^*$. In particular, for any $\mu \in \{0,1,2,3\}$, we have that $\boldsymbol{e}^{(\mu)} = \boldsymbol{\eta}(\boldsymbol{e}_{(\mu)},.) \in \mathcal{M}^*$, and, by definition of the orthonormal basis $\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{2})}, \boldsymbol{e}_{(\mathbf{3})}\}$, this leads to:

$$\forall(\mu,\nu) \in \{0,1,2,3\}^2,\ \boldsymbol{e}^{(\mu)}\left(\boldsymbol{e}_{(\nu)}\right) = \begin{cases} -1 & \text{if } \mu = \nu = 0 \\ 1 & \text{if } \mu = \nu \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.42}$$

Therefore, if we define $\boldsymbol{\omega}^{(\mathbf{0})} = -\boldsymbol{\eta}(\boldsymbol{e}_{(\mathbf{0})},\cdot)$, and $\forall i \in \{1,2,3\}, \boldsymbol{\omega}^{(\boldsymbol{i})} = \boldsymbol{\eta}(\boldsymbol{e}_{(\boldsymbol{i})},\cdot)$, then $\{\boldsymbol{\omega}^{(\mathbf{0})}, \boldsymbol{\omega}^{(\mathbf{1})}, \boldsymbol{\omega}^{(\mathbf{2})}, \boldsymbol{\omega}^{(\mathbf{3})}\}$ is exactly the dual basis of $\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{2})}, \boldsymbol{e}_{(\mathbf{3})}\}$. In that case, we can write:

$$\boldsymbol{\eta} = \eta_{\mu\nu} \boldsymbol{\omega}^{(\mu)} \otimes \boldsymbol{\omega}^{(\nu)} \ . \tag{2.43}$$

Since $\mathcal{M}^*$ is in bijection with $\mathcal{M}$, the structure present on $\mathcal{M}$ via the inner product $\boldsymbol{\eta}$ is inherited by the dual $\mathcal{M}^*$: there is an inner product on the space of linear functions that is exactly like the one on the space $\mathcal{M}$. Actually, physicists give it the same name, $\boldsymbol{\eta}$ in our case. But it is now seen as a $(2,0)$ tensor acting on linear functions with components $\eta^{\mu\nu}$ on the basis $\left\{\boldsymbol{e}_{(\mu)} \otimes \boldsymbol{e}_{(\nu)}\right\}$ of doubly contravariant tensors: $\forall(\boldsymbol{a},\boldsymbol{b}) \in \mathcal{M}^* \times \mathcal{M}^*$, $\boldsymbol{a} = a_\mu \boldsymbol{\omega}^{(\mu)}$, $\boldsymbol{b} = b_\mu \boldsymbol{\omega}^{(\mu)}$, $\boldsymbol{\eta}(\boldsymbol{a},\boldsymbol{b}) = \eta^{\mu\nu} a_\mu b_\nu$. The components $\eta^{\mu\nu}$ are exactly the same as the components $\eta_{\mu\nu}$. All this justifies the 'lowering' and 'raising' of indices practised by physicists: to any vector $\boldsymbol{v} \in \mathcal{M}$ with components v^μ's such that $\boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)}$, we can associate a linear function, let us say $\tilde{\boldsymbol{v}} \in \mathcal{M}^*$ with components v_μ's on the dual basis: $\tilde{\boldsymbol{v}} = v_\mu \boldsymbol{\omega}^{(\mu)}$, and one has that the 'covariant components' of the vector are given, in terms of the 'contravariant ones' by: $v_\mu = \eta_{\mu\nu} v^\nu$.

Very often, the coordinates (x^0, x^1, x^2, x^3) associated to an orthonormal basis are renamed (t, x, y, z). Let us consider an infinitesimal displacement in Minkowski spacetime, defined by the vector:

$$\mathbf{d}\boldsymbol{p} = \mathrm{d}x^\mu \boldsymbol{e}_{(\mu)} = \mathrm{d}t \boldsymbol{e}_{(\mathbf{0})} + \mathrm{d}x \boldsymbol{e}_{(\mathbf{1})} + \mathrm{d}y \boldsymbol{e}_{(\mathbf{2})} + \mathrm{d}z \boldsymbol{e}_{(\mathbf{3})} \ . \tag{2.44}$$

We can define the infinitesimal 'length' of this vector as the *spacetime interval* ds such that:

$$\begin{aligned} \mathrm{d}s^2 =& \boldsymbol{\eta}\left(\boldsymbol{dp}, \boldsymbol{dp}\right) && (2.45)\\ =& -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ . && (2.46) \end{aligned}$$

2.3.3 Non-orthonormal bases

So far, we have defined orthonormal basis $\left\{\boldsymbol{e}_{(\mu)}\right\}$ canonically associated with Cartesian coordinates $\{x^\mu\} = \{t, x, y, z\}$, but there are other sets of coordinates one can use, e.g. spherical coordinates $\{t, r, \theta, \phi\}$. In this coordinate system, we have:

$$\begin{cases} x =r\sin\theta\cos\phi & (2.47)\\ y =r\sin\theta\sin\phi & (2.48)\\ z =r\cos\theta \ . & (2.49) \end{cases}$$

The canonical vectors associated with these coordinates $\left\{\boldsymbol{e}_{(\mathbf{0})}, \hat{\boldsymbol{e}}_{(\mathbf{1})}, \hat{\boldsymbol{e}}_{(\mathbf{2})}, \hat{\boldsymbol{e}}_{(\mathbf{3})}\right\}$ are such that an infinitesimal displacement reads:

$$\mathbf{d}\boldsymbol{p} = \mathrm{d}t\boldsymbol{e}_{\mathbf{0}} + \mathrm{d}x\boldsymbol{e}_{\mathbf{1}} + \mathrm{d}y\boldsymbol{e}_{\mathbf{2}} + \mathrm{d}z\boldsymbol{e}_{(\mathbf{3})} = \mathrm{d}t\boldsymbol{e}_{\mathbf{0}} + \mathrm{d}r\hat{\boldsymbol{e}}_{\mathbf{1}} + \mathrm{d}\theta\hat{\boldsymbol{e}}_{\mathbf{2}} + \mathrm{d}\phi\hat{\boldsymbol{e}}_{(\mathbf{3})} \ . \tag{2.50}$$

Therefore, using the relationship between coordinates, we get the interval:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2 \ , \tag{2.51}$$

Thus, although the basis remains orthogonal, we have that:

$$\begin{cases} \boldsymbol{\eta}\left(\hat{\boldsymbol{e}}_{(\mathbf{2})}, \hat{\boldsymbol{e}}_{(\mathbf{2})}\right) =r^2 \neq 0 & (2.52)\\ \boldsymbol{\eta}\left(\hat{\boldsymbol{e}}_{(\mathbf{3})}, \hat{\boldsymbol{e}}_{(\mathbf{3})}\right) =r^2\sin^2\theta \neq 0 \ , & (2.53) \end{cases}$$

so the coordinate basis associated to spherical coordinates is not orthonormal. It can be turned into an orthonormal basis by normalising each vector, giving the usual spherical basis, but this is not a coordinate basis; see subsection 2.5.1. Of course, we can also construct non-orthogonal bases in which the metric is not diagonal any more.

2.3.4 Classes of vectors in Minkowski spacetime

In the following, we will denote by $\boldsymbol{q}$ the quadratic form associated with $\boldsymbol{\eta}$ to simplify notations: $\forall \boldsymbol{v} \in \mathcal{M}$, $\boldsymbol{q}(\boldsymbol{v}) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) \in \mathbb{R}$. The fact that $\boldsymbol{\eta}$ is not positive definite implies that there are non-zero vectors $\boldsymbol{n} \in \mathcal{M}$ such that $\boldsymbol{q}(\boldsymbol{n}) = 0$.

Lightlike vectors

$\boldsymbol{n} \in \mathcal{M}$ is called a non-zero *null*, or *lightlike* vector iff $\boldsymbol{n} \neq 0$ and $\boldsymbol{q}(\boldsymbol{n}) = \boldsymbol{\eta}(\boldsymbol{n}, \boldsymbol{n}) = 0$.

Lightlike vectors are often called *null*. For example, the vector $\boldsymbol{v} = \boldsymbol{e}_{(0)} + \boldsymbol{e}_{(1)}$ is null:

$$\boldsymbol{q}(\boldsymbol{v}) = \boldsymbol{q}\left(\boldsymbol{e}_{(0)}\right) + 2\boldsymbol{\eta}\left(\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}\right) + \boldsymbol{q}(\boldsymbol{e}_{(1)}) = -1 + 0 + 1 = 0. \tag{2.54}$$

Let us recall an important result of the standard dot product in $\mathbb{R}^3$:

Cauchy-Schwartz inequality

Consider the vector space $\mathbb{R}^3$ and the standard dot product on $\mathbb{R}^3$. Let $\vec{u}$ and $\vec{v}$ be two non-zero vectors of $\mathbb{R}^3$. Then:

$$(\vec{u} \cdot \vec{v})^2 \leq (\vec{u} \cdot \vec{u})(\vec{v} \cdot \vec{v}), \tag{2.55}$$

where the equality holds iff $\vec{u}$ and $\vec{v}$ are linearly dependent.

The proof goes as follows. Consider:

$$\vec{z} = \vec{u} - \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} \, . \tag{2.56}$$

Then:

$$\vec{z} \cdot \vec{v} = 0 \, , \tag{2.57}$$

and therefore, by writing:

$$\vec{u} = \frac{\vec{u} \cdot \vec{v}}{\vec{v} \cdot \vec{v}} \vec{v} + \vec{z} \, , \tag{2.58}$$

we get:

$$\vec{u} \cdot \vec{u} = \frac{(\vec{u}. \cdot \vec{v})^2}{\vec{v} \cdot \vec{v}} + \vec{z} \cdot \vec{z} \geq \frac{(\vec{u} \cdot \vec{v})^2}{\vec{v} \cdot \vec{v}} \, . \tag{2.59}$$

Multiplying by $\vec{v}\cdot\vec{v}$, we get the result. Besides, if $\exists t \in \mathbb{R}, \vec{u} = t\vec{v}$, we clearly have: $(\vec{u}\cdot\vec{v})^2 = t^2\vec{v}\cdot\vec{v} = (\vec{u}\cdot\vec{u})(\vec{v}\cdot\vec{v})$. Conversely, if we have equality, then, we get:

$$\vec{u}\cdot\left(\frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}\right) = \vec{u}\cdot\vec{u}, \tag{2.60}$$

and therefore $\vec{u} = \frac{\vec{u}\cdot\vec{v}}{\vec{v}\cdot\vec{v}}\vec{v}$, because the dot product is non-degenerate.

Let us now consider two lightlike vectors $\boldsymbol{v} \in \mathcal{M}$ and $\boldsymbol{w} \in \mathcal{M}$ that are also η-orthogonal: $\boldsymbol{\eta}(\boldsymbol{v},\boldsymbol{w}) = 0$. Then:

$$\left(v^0\right)^2 = \left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 \tag{2.61}$$

$$\left(w^0\right)^2 = \left(w^1\right)^2 + \left(w^2\right)^2 + \left(w^3\right)^2 \ . \tag{2.62}$$

This means that:

$$\left(v^0\right)^2\left(w^0\right)^2 = \left(\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2\right)\left(\left(w^1\right)^2 + \left(w^2\right)^2 + \left(w^3\right)^2\right) \ . \tag{2.63}$$

On the other hand, because $\boldsymbol{\eta}(\boldsymbol{v},\boldsymbol{w}) = 0$, we have:

$$\left(v^0w^0\right)^2 = \left(v^0\right)^2\left(w^0\right)^2 = \left(v^1w^1 + v^2w^2 + v^3w^3\right)^2 \ . \tag{2.64}$$

This proves that the Cauchy-Schwartz inequality in the subspace $\mathbb{R}^3$ orthogonal to $\boldsymbol{e}_{(\mathbf{0})}$ and spanned by $\left\{\boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{2})}, \boldsymbol{e}_{(\mathbf{3})}\right\}$ is saturated, and therefore: $\exists t \in \mathbb{R},\ v^1\boldsymbol{e}_{(\mathbf{1})}+v^2\boldsymbol{e}_{(\mathbf{2})}+v^3\boldsymbol{e}_{(\mathbf{3})} = t\left(w^1\boldsymbol{e}_{(\mathbf{1})} + w^2\boldsymbol{e}_{(\mathbf{2})} + w^3\boldsymbol{e}_{(\mathbf{3})}\right)$, or, equivalently, $\forall i \in \{1,2,3\}, v^i = tw^i$.
Finally, that leads to:

$$\begin{aligned} v^0w^0 &= t\left(\left(w^1\right)^2 + \left(w^2\right)^2 + \left(w^3\right)^2\right) \\ &= t\left(w^0\right)^2 \ , \end{aligned} \tag{2.65}$$

so that we have: $v^0 = tw^0$. This shows that, necessarily, $\boldsymbol{v} = t\boldsymbol{w}$ for some $t \in \mathbb{R}$. Thus, we see that *orthogonal lightlike vectors are also parallel*: if $\boldsymbol{v} \in \mathcal{M}$ and $\boldsymbol{w} \in \mathcal{M}$ are non-zero null vectors, then $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal $\Leftrightarrow \exists t \in \mathbb{R}, \boldsymbol{v} = t\boldsymbol{w}$.

What is the link between this inner product and Special Relativity? Events in spacetime are points but since $\mathcal{M}$ is a vector space, once an observer has been chosen in the form of an origin O

for the frame $\{\boldsymbol{e}_{(\mu)}\}$, two events x and x_0 define a vector $\boldsymbol{v} = \boldsymbol{Ox} - \boldsymbol{Ox_0} \in \mathcal{M}$. In the following, we will often identify points and vectors starting at the origin. We will even omit to mention the origin when necessary and write $\boldsymbol{x} = \boldsymbol{Ox}$. If we write: $\boldsymbol{Ox} = x^\mu \boldsymbol{e}_{(\mu)}$ and $\boldsymbol{Ox_0} = x_0^\mu \boldsymbol{e}_\mu$. then $\boldsymbol{v}$ is lightlike if we have:

$$\boldsymbol{q}(\boldsymbol{v}) = -\left(x^0 - x_0^0\right)^2 + \left(x^1 - x_0^1\right)^2 + \left(x^2 - x_0^2\right)^2 + \left(x^3 - x_0^3\right)^2 = 0. \tag{2.66}$$

We know this equation. It is the condition for two events to be on the worldline of the same photon, and it is the equation of a cone in $\mathbb{R}^4$. We also saw that this implied that this condition must be *invariant* when changing the inertial frame. In our context, that means that the inner product of two vectors must be invariant by changing orthonormal bases. We will see the importance of that later. For the moment, we will therefore define the *nullcone* (or *lightcone*) at $\boldsymbol{x_0}$ in $\mathcal{M}$ by:

$$C\left(\boldsymbol{x_0}\right) = \left\{\boldsymbol{x} \in \mathcal{M}, \boldsymbol{q}\left(\boldsymbol{x} - \boldsymbol{x_0}\right) = 0\right\} \ . \tag{2.67}$$

Physically, it consists of all the events in $\mathcal{M}$ that can be connected to $\boldsymbol{x_0}$ via a light ray. For any to such events $\boldsymbol{x}$ and $\boldsymbol{x_0}$ we can therefore further define a *light ray*, or *null worldline*:

$$R_{\boldsymbol{x_0},\boldsymbol{x}} = \left\{\boldsymbol{v} \in \mathcal{M}, \exists t \in \mathbb{R}, \boldsymbol{v} = \boldsymbol{x_0} + t\left(\boldsymbol{x} - \boldsymbol{x_0}\right)\right\} \ . \tag{2.68}$$

We see that, clearly, $R_{\boldsymbol{x_0},\boldsymbol{x}} = R_{\boldsymbol{x},\boldsymbol{x_0}}$. Moreover, we also clearly have that $C\left(\boldsymbol{x_0}\right)$ is the (infinite) union of all the light rays through $\boldsymbol{x_0}$. So far, we have centred our attention on null vectors, i.e. vectors $\boldsymbol{v} \in \mathcal{M}$ for which $\boldsymbol{q}(\boldsymbol{v}) = 0$. We can also define two other type of vectors:

Timelike and spacelike vectors

Let $\boldsymbol{v} \in \mathcal{M}$. We say that:

(i) $\boldsymbol{v}$ is *timelike* iff $\boldsymbol{q}(\boldsymbol{v}) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) < 0$;

(ii) $\boldsymbol{v}$ is *spacelike* iff $\boldsymbol{q}(\boldsymbol{v}) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) > 0$.

Note that, for $\boldsymbol{v} \in \mathcal{M}$, if $\boldsymbol{v}$ is timelike, we have that: $\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 < \left(v^0\right)^2$, this corresponds to vectors *inside* the lightcone. Physically, that means that the distance (in $\mathbb{R}^3$) covered along the vector is less than the distance covered by a light ray in the same time lapse.
In the same way, for a spacelike vector, we have: $\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 > \left(v^0\right)^2$, and this corresponds to vectors *outside* the lightcone. In that case, the distance (in $\mathbb{R}^3$) covered along the vector is more

than the distance covered by a light ray in the same period of time.

Now, we have to show the following important result:

Time components of timelike and lightlike vectors

Let $(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{M}^2$, both non-zero, with $\boldsymbol{v}$ timelike, and $\boldsymbol{w}$ timelike or lightlike. Let $\left\{\boldsymbol{e}_{(\mu)}\right\}$ be an orthonormal basis of $\mathcal{M}$, such that: $\boldsymbol{v} = v^{\mu}\boldsymbol{e}_{(\mu)}$ and $\boldsymbol{w} = w^{\mu}\boldsymbol{e}_{(\mu)}$. Then:

(i) $v^0 w^0 > 0$, and we have $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) < 0$, or

(ii) $v^0 w^0 < 0$, and we have $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) > 0$.

Indeed, by definition:

$$\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) = -\left(v^0\right)^2 + \left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 \quad < \quad 0 \tag{2.69}$$

$$\boldsymbol{\eta}(\boldsymbol{w}, \boldsymbol{w}) = -\left(w^0\right)^2 + \left(w^1\right)^2 + \left(w^2\right)^2 + \left(w^3\right)^2 \quad \leq \quad 0\,. \tag{2.70}$$

Hence:

$$\left(v^0 w^0\right)^2 > \left(\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2\right)\left(\left(w^1\right)^2 + \left(w^2\right)^2 + \left(w^3\right)^2\right)\,. \tag{2.71}$$

Or, using the Cauchy-Schwartz inequality in $\mathbb{R}^3$:

$$\left(v^0 w^0\right)^2 > \left(v^1 w^1 + v^2 w^2 + v^3 w^3\right)^2\,, \tag{2.72}$$

which implies that:

$$\left|v^0 w^0\right| > \left|v^1 w^1 + v^2 w^2 + v^3 w^3\right|\,. \tag{2.73}$$

Therefore, we clearly have: $v^0 w^0 \neq 0$, and $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) \neq 0$.

If $v^0 w^0 > 0$, then:

$$v^0 w^0 = \left|v^0 w^0\right| > \left|v^1 w^1 + v^2 w^2 + v^3 w^3\right| \geq v^1 w^1 + v^2 w^2 + v^3 w^3\,. \tag{2.74}$$

This leads to $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) = -v^0 w^0 + v^1 w^1 + v^2 w^2 + v^3 w^3 < 0$. Similarly, if $v^0 w^0 < 0$, one finds that: $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) > 0$.

We have a simple corollary to this result: if a non-zero vector of $\mathcal{M}$ is orthogonal to a timelike vector, then it must be spacelike.

This result tells us that if two vectors, one timelike, $\boldsymbol{v}$, and the other timelike or null, $\boldsymbol{w}$, point in the same direction along the 'time axis', then we have $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) < 0$. Then we can define an equivalence relation $\sim$ as follow:

- We call $\tau = \{\boldsymbol{v} \in \mathcal{M}, \boldsymbol{q}(\boldsymbol{v}) < 0\}$ the set of all the timelike vectors of $\mathcal{M}$.
- Let $(\boldsymbol{v}, \boldsymbol{w}) \in \tau^2$, Then: $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) < 0 \Rightarrow \boldsymbol{v} \sim \boldsymbol{w}$.

You can check, as an exercise that $\sim$ is an equivalence relation on τ, i.e. that $\sim$ is:

- **reflexive**: $\forall \boldsymbol{v} \in \tau, \boldsymbol{v} \sim \boldsymbol{v}$;
- **symmetric**: $\forall (\boldsymbol{v}, \boldsymbol{w}) \in \tau^2, \boldsymbol{v} \sim \boldsymbol{w} \Rightarrow \boldsymbol{w} \sim \boldsymbol{v}$;
- **transitive**: $\forall (\boldsymbol{v}, \boldsymbol{w}, \boldsymbol{x}) \in \tau^3, (\boldsymbol{v} \sim \boldsymbol{w}$ and $\boldsymbol{w} \sim \boldsymbol{x}) \Rightarrow \boldsymbol{v} \sim \boldsymbol{x}$.

Moreover, this equivalence relation has exactly **two classes**, called τ^+ and τ^-, where $\tau^+ = \left\{\boldsymbol{v} \in \tau, v^0 > 0\right\}$ and $\tau^- = \left\{\boldsymbol{v} \in \tau, v^0 < 0\right\}$. We clearly have: $\tau = \tau^+ \cup \tau^-$. Elements of τ^+ all have the same time-orientation, and points toward positive values of the time coordinate. They are thus called *future-directed*. Elements of τ^- point toward negative values of the time coordinate and are therefore called *past-directed*. For each event $\boldsymbol{x_0} \in \mathcal{M}$, we can define the *time cone*, $C_T(\boldsymbol{x_0})$, the *future time cone*, $C_T^+(\boldsymbol{x_0})$ and the the *past time cone*, $C_T^-(\boldsymbol{x_0})$ by:

$$C_T(\boldsymbol{x_0}) = \{\boldsymbol{x} \in \mathcal{M}, \boldsymbol{q}(\boldsymbol{x} - \boldsymbol{x_0}) < 0\} \tag{2.75}$$

$$C_T^+(\boldsymbol{x_0}) = \left\{\boldsymbol{x} \in \mathcal{M}, \boldsymbol{x} - \boldsymbol{x_0} \in \tau^+\right\} \tag{2.76}$$

$$C_T^-(\boldsymbol{x_0}) = \{\boldsymbol{x} \in \mathcal{M}, \boldsymbol{x} - \boldsymbol{x_0} \in \tau^-\} \ . \tag{2.77}$$

Clearly, $C_T(\boldsymbol{x_0})$ is the interior of the null cone $C(\boldsymbol{x_0})$. It is made of two disjoint parts, $C_T^+(\boldsymbol{x_0})$ and $C_T^-(\boldsymbol{x_0})$ representing the timelike future and past of $\boldsymbol{x_0}$, respectively.
Now, we would like to extend these notions of future and past to lighlike vectors as well. Let us pick a non-zero lightlike vector $\boldsymbol{n} \in C(\boldsymbol{x_0})$. We have that:

$$\forall \boldsymbol{v} \in \tau^+, \boldsymbol{\eta}(\boldsymbol{n}, \boldsymbol{v}) > 0 \text{ or } \boldsymbol{\eta}(\boldsymbol{n}, \boldsymbol{v}) < 0 \ . \tag{2.78}$$

Indeed, let us suppose that we have two timelike vectors $(\boldsymbol{v_1}, \boldsymbol{v_2}) \in \tau^+$ such that:

$$\boldsymbol{\eta}(\boldsymbol{v_1}, \boldsymbol{n}) < 0 \text{ and } \boldsymbol{\eta}(\boldsymbol{v_2}, \boldsymbol{n}) > 0 \ . \tag{2.79}$$

Then, we have seen that, necessarily, $v_1^0 n^0 > 0$ and $v_2^0 n^0 < 0$. But, we also have that $v_1 \sim v_2$, and therefore, $v_1^0 v_2^0 > 0$. This trivially leads to a contradiction. Therefore, we can say that a lightlike vector $\boldsymbol{n}$ is *future-directed* iff $\forall \boldsymbol{v} \in \tau^+, \boldsymbol{\eta}(\boldsymbol{n}, \boldsymbol{v}) < 0$ and *past-directed* iff $\forall \boldsymbol{v} \in \tau^+, \boldsymbol{\eta}(\boldsymbol{n}, \boldsymbol{v}) > 0$.

One can then prove that two non-zero lightlike vectors $\boldsymbol{n_1}$ and $\boldsymbol{n_2}$ have the same time orientation iff $n_1^0 n_2^0 > 0$. This allows us to define the future and past null cones at $\boldsymbol{x_0} \in \mathcal{M}$ as the sets:

$$C^+(\boldsymbol{x_0}) = \{\boldsymbol{x} \in C(x_0), \boldsymbol{x} - \boldsymbol{x_0} \text{ future-directed}\} \tag{2.80}$$

$$C^-(\boldsymbol{x_0}) = \{\boldsymbol{x} \in C(\boldsymbol{x_0}), \boldsymbol{x} - \boldsymbol{x_0} \text{ past-directed}\}. \tag{2.81}$$

These sets define the *causal structure of spacetime* around x_0. The cone itself is the set of event connected to x_0 by light rays as we have seen, We will understand its interior as the set of events connected to x_0 by massive particles a bit later. The overall geometry formalised here is summarised and depicted on Fig. 2.1.

2.4 Lorentz transformations

In Special Relativity, the principle of relativity that we quoted in Newtonian physics still holds and inertial frames have not changed their nature: they are still frames in which an object free of forces is at rest on in uniform motion. Since the laws of physics must not change when we go from one inertial frame to another, the geometry of spacetime must not be altered by a transformation between inertial frame. In Special Relativity, these transformations are known as *Lorentz transformations*. Because we work on a vector space, we will also look for transformations between frames that are *linear*. Therefore, Lorentz transformations are linear mappings of $\mathcal{M}$ onto itself that *preserve* the metric (or inner product) $\boldsymbol{\eta}$, i.e. they are the $\boldsymbol{\eta}$-isometries of Minkowski spacetime. In particular, this implies that they preserve the causal structure described in the previous section and depicted in Fig. 2.1. From the passive viewpoint, they are the transformations that allow one to go from one inertial frame to another in Special Relativity, as we will see later.

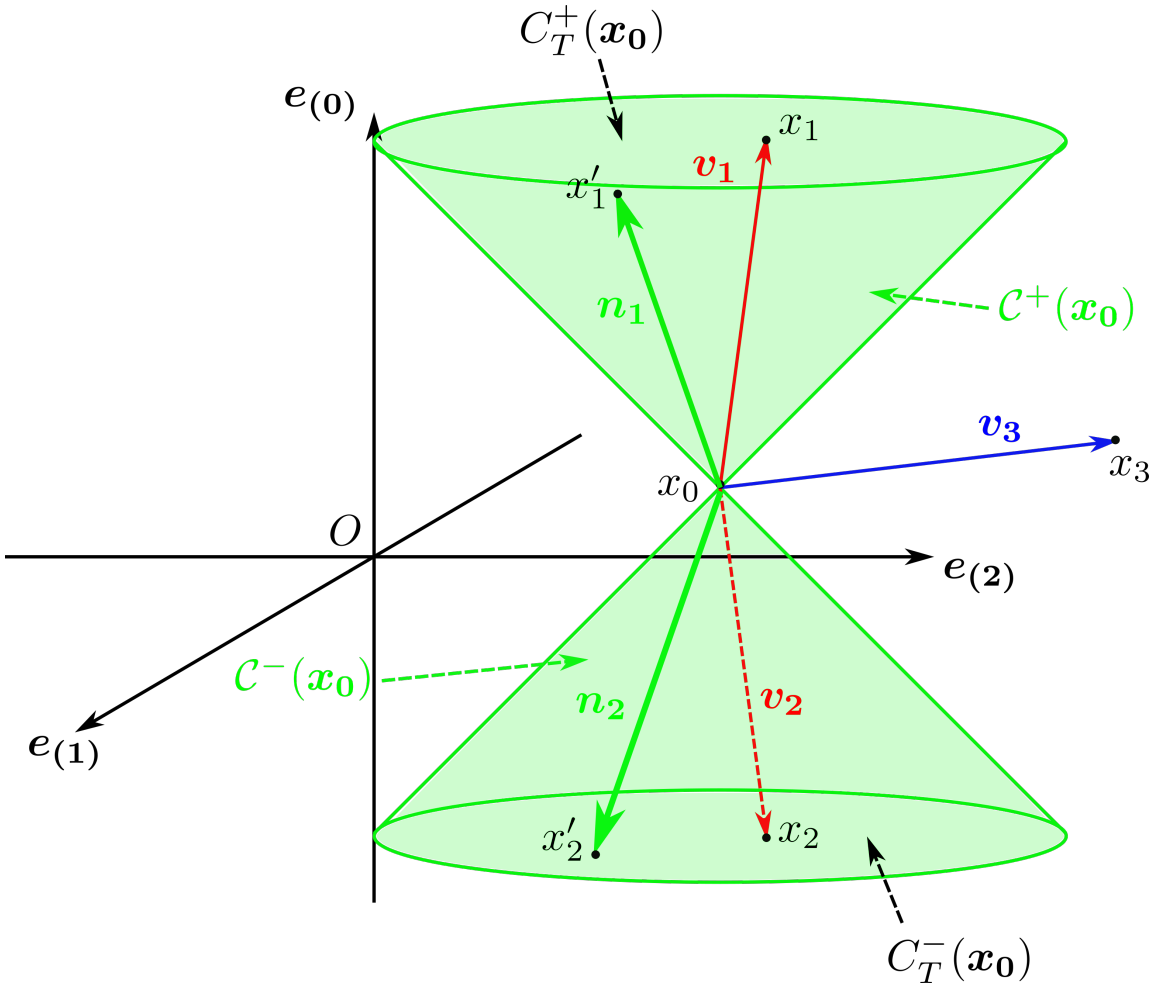


Figure 2.1: Minkowski spacetime in the orthogonal reference frame $\{\boldsymbol{e}_{(\mu)}\}$ of an observer O. One spatial dimension has been suppressed. The causal structure around an event x_0 has been represented by its lightcone in green. It represents the events connected to x_0 by light rays. x_1' is in the future of $x_0$since it is connected to it by a future-directed null vector, $\boldsymbol{v_1'}$. On the other hand, x_2' is in the past of x_0 and the null vector $\boldsymbol{v_2'}$ is past-directed. Timelike vectors are also separated into future directed (like $\boldsymbol{v_1}$) and past-directed (like $\boldsymbol{v_2}$) connecting x_0 to events that lie in its causal, timelike future (like x_1) or past (like x_2). Events like x_3 are connected to x_0 by spacelike vectors, like $\boldsymbol{v_3}$ and are not causally connected to x_0.

2.4.1 Characterisation of the Lorentz group

First, we need to characterise Lorentz transformations in a few equivalent ways.

Isometries of Minkowski spacetime

Let $\boldsymbol{L} : \mathcal{M} \to \mathcal{M}$ be a linear mapping. Then, the following propositions are equivalent:

(i) $\boldsymbol{L}$ preserves the inner product $\boldsymbol{\eta}$, i.e.:

$$\forall (\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{M}^2, \boldsymbol{\eta}\left(\boldsymbol{L}(\boldsymbol{v}), \boldsymbol{L}(\boldsymbol{w})\right) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) \ . \tag{2.82}$$

(ii) $\boldsymbol{L}$ preserves the quadratic form $\boldsymbol{q}$ associated with $\boldsymbol{\eta}$:

$$\forall \boldsymbol{v} \in \mathcal{M}, \ \boldsymbol{q}\left(\boldsymbol{L}(\boldsymbol{v})\right) = \boldsymbol{q}(\boldsymbol{v}) \ . \tag{2.83}$$

(iii) For any $\left\{\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)}\right\}$ orthonormal basis of $\mathcal{M}$, there exists another orthonormal basis of $\mathcal{M}$, $\left\{\boldsymbol{\epsilon}_{(0)}, \boldsymbol{\epsilon}_{(1)}, \boldsymbol{\epsilon}_{(2)}, \boldsymbol{\epsilon}_{(3)}\right\}$ such that:

$$\forall \mu \in \{0, 1, 2, 3\}, \boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right) = \boldsymbol{\epsilon}_{(\mu)} \ . \tag{2.84}$$

We will prove each equivalence separately.

(i)⇒(ii) Let $(\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{M}^2$. Let us suppose that we have: $\boldsymbol{\eta}\left(\boldsymbol{L}(\boldsymbol{v}), \boldsymbol{L}(\boldsymbol{w})\right) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w})$ for all $\boldsymbol{v}$ and $\boldsymbol{w}$. Then, of course, because $\boldsymbol{q}(\boldsymbol{v}) = \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v})$, we have: $\boldsymbol{q}\left(\boldsymbol{L}(\boldsymbol{v})\right) = \boldsymbol{q}(\boldsymbol{v})$.

(ii)⇒(i) Here, we use the fact that:

$$\forall (\boldsymbol{v}, \boldsymbol{w}) \in \mathcal{M}^2, \ \boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) = \frac{1}{2}\left[\boldsymbol{q}(\boldsymbol{v} + \boldsymbol{w}) - \boldsymbol{q}(\boldsymbol{v}) - \boldsymbol{q}(\boldsymbol{w})\right] \ . \tag{2.85}$$

The invariance of $\boldsymbol{q}$ then implies the invariance of $\boldsymbol{\eta}$.

(i)⇒(iii) We know that:

$$\forall (\mu, \nu) \in \{0, 1, 2, 3\}^2, \boldsymbol{\eta}\left(\boldsymbol{e}_{(\mu)}, \boldsymbol{e}_{(\nu)}\right) = \eta_{\mu\nu} \ . \tag{2.86}$$

Now:

$$\boldsymbol{\eta}\left(\boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right), \boldsymbol{L}\left(\boldsymbol{e}_{(\nu)}\right)\right) = \boldsymbol{\eta}\left(\boldsymbol{e}_{(\mu)}, \boldsymbol{e}_{(\nu)}\right) = \eta_{\mu\nu} \ . \tag{2.87}$$

This proves that that set $\left\{\boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right)\right\}_{\mu\in\{0,1,2,3\}}$ is orthonormal. We just have to prove that it is linearly independent to prove that it is an orthonormal basis.
$\left\{\boldsymbol{e}_{(\mu)}\right\}$ being a basis, we have that:

$$v^{\mu}\boldsymbol{e}_{(\mu)} = 0 \Rightarrow \forall\mu \in \{0,1,2,3\}, v^{\mu} = 0 \ . \tag{2.88}$$

Thus:

$$v^{\mu}\boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right) = 0 \Rightarrow \boldsymbol{L}\left(v^{\mu}\boldsymbol{e}_{(\mu)}\right) = 0 \tag{2.89}$$

$$\Rightarrow v^{\mu}\boldsymbol{e}_{(\mu)} = 0 \ \text{(injectivity of } \boldsymbol{L}) \tag{2.90}$$

$$\Rightarrow \forall\mu \in \{0,1,2,3\},\ v^{\mu} = 0 \ \text{(linear independence of } \left\{\boldsymbol{e}_{(\mu)}\right\}). \tag{2.91}$$

The set $\left\{\boldsymbol{L}(\boldsymbol{e}_{(\mu)})\right\} = \left\{\boldsymbol{\epsilon}_{(\mu)}\right\}$ is therefore linearly independent.

(iii)⇒(i) We have:

$$\forall(\boldsymbol{v},\boldsymbol{w}) \in \mathcal{M}^2,\ \boldsymbol{\eta}\left(\boldsymbol{L}(\boldsymbol{v}),\boldsymbol{L}(\boldsymbol{w})\right) = \boldsymbol{\eta}\left(v^{\mu}\boldsymbol{\epsilon}_{(\mu)}, w^{\nu}\boldsymbol{\epsilon}_{(\nu)}\right) = \boldsymbol{\eta}\left(\boldsymbol{\epsilon}_{(\mu)},\boldsymbol{\epsilon}_{(\nu)}\right)v^{\mu}w^{\nu}\,, \tag{2.92}$$

which is exactly $\boldsymbol{\eta}\left(\boldsymbol{L}(\boldsymbol{v}),\boldsymbol{L}(\boldsymbol{w})\right) = \boldsymbol{\eta}(\boldsymbol{v},\boldsymbol{w})$.

Consider two orthonormal bases of $\mathcal{M}$, $\left\{\boldsymbol{e}_{(\mu)}\right\}$ and $\left\{\hat{\boldsymbol{e}}_{(\mu)}\right\}$ and a linear mapping $\boldsymbol{L}:\ \mathcal{M}\ \rightarrow\ \mathcal{M}$ such that $\forall\mu \in \{0,1,2,3\},\ \boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right) = \hat{\boldsymbol{e}}_{(\mu)}$. We know that there exists constants $\Lambda^{\mu}{}_{\nu} \in \mathbb{R}$ for any $(\mu,\nu) \in \{0,1,2,3\}^2$, such that:

$$\forall\nu \in \{0,1,2,3\}, \boldsymbol{e}_{(\nu)} = \Lambda^{\mu}{}_{\nu}\hat{\boldsymbol{e}}_{(\mu)} \ . \tag{2.93}$$

These are, by definition, the components of the linear mapping $\boldsymbol{L}^{-1}$ in the basis $\left\{\hat{\boldsymbol{e}}_{(\mu)}\right\}$[2]. If the coordinates of a vector $\boldsymbol{v} \in \mathcal{M}$ are such that: $\boldsymbol{v} = v^{\mu}\boldsymbol{e}_{(\mu)} = \hat{v}^{\mu}\hat{\boldsymbol{e}}_{(\mu)}$, then we have the law of transformation for vector components under Lorentz transformations:

$$\boxed{\forall\mu \in \{0,1,2,3\},\ \hat{v}^{\nu} = \Lambda^{\nu}{}_{\mu}v^{\mu}\ ,} \tag{2.94}$$

[2]The fact that we study the matrix representation of $\boldsymbol{L}^{-1}$ instead of $\boldsymbol{L}$ is of course irrelevant and one could have studied the matrix representation of $\boldsymbol{L}$ instead, but we stick here to the usual notations that one finds in standard physics and applied mathematics books. Note that, strictly speaking we have not even proved the existence of an inverse for a mapping $\boldsymbol{L}$ but it should be clear that such an inverse must exist if the orthonormal bases are to be treated on an equal ground.

Using the orthonormality of $\{\boldsymbol{e}_{(\mu)}\}$, $\forall(\mu,\nu)\in\{0,1,2,3\}^2$, $\boldsymbol{\eta}\left(\boldsymbol{e}_{(\mu)},\boldsymbol{e}_{(\nu)}\right)=\eta_{\mu\nu}$, we have:

$$\boldsymbol{\eta}\left(\Lambda^{\rho}{}_{\mu}\hat{\boldsymbol{e}}_{(\rho)},\Lambda^{\lambda}{}_{\nu}\hat{\boldsymbol{e}}_{(\nu)}\right)=\eta_{\mu\nu}\ , \tag{2.95}$$

which leads to:

$$\Lambda^{\rho}_{\mu}\Lambda^{\lambda}{}_{\nu}\boldsymbol{\eta}\left(\hat{\boldsymbol{e}}_{(\rho)},\hat{\boldsymbol{e}}_{(\lambda)}\right)=\eta_{\mu\nu}\ . \tag{2.96}$$

Using the orthonormality of $\{\hat{\boldsymbol{e}}_{(\mu)}\}$, we thus have:

$$\eta_{\rho\lambda}\Lambda^{\rho}{}_{\mu}\Lambda^{\lambda}{}_{\nu}=\eta_{\mu\nu}\ . \tag{2.97}$$

This relation must be satisfied by any linear mapping that preserves the inner product $\boldsymbol{\eta}$. Therefore, it defines such mappings. If we define the matrix:

$$\eta=\begin{bmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\ , \tag{2.98}$$

then Eq.(2.97) takes the simple form:

$$\boxed{\Lambda^{T}\eta\Lambda=\eta\ ,} \tag{2.99}$$

where the superscript T denotes the usual matrix transposition. This can be taken as an *operational definition of Lorentz transformations*.

The matrix η has components $\eta_{\mu\nu}$ or $\eta^{\mu\nu}$, depending on the summation required. This is reminiscent of the equivalence between $\mathcal{M}$ and $\mathcal{M}^*$ introduced by the inner product. For example, we see that $\eta^{-1}=\eta$ becomes simply: $\eta^{\mu\rho}\eta_{\nu\rho}=\delta^{\mu}_{\nu}$.

Note that we could have used the matrix $\mathcal{L}$ associated with the transformation $\boldsymbol{L}$ instead of $\boldsymbol{L}^{-1}$. It obeys the same relation:

$$\eta_{\rho\lambda}\mathcal{L}^{\rho}{}_{\mu}\mathcal{L}^{\lambda}{}_{\nu}=\eta_{\mu\nu}\ , \tag{2.100}$$

and we have: $\mathcal{L}=\Lambda^{-1}$. These matrices are related by the relation:

$$\mathcal{L}=\Lambda^{-1}=\eta\Lambda^{t}\eta\ , \tag{2.101}$$

where we have used the fact that $\eta^t=\eta^{-1}=\eta$. Usually, the components of $\mathcal{L}$ are denoted $\Lambda_{\mu}{}^{\nu}$.

Lorentz transformations

Any linear mapping $\boldsymbol{L} : \mathcal{M} \to \mathcal{M}$ which preserves the inner product $\boldsymbol{\eta}$ is such that:

$$\eta_{\rho\lambda} L^{\rho}{}_{\mu} L^{\lambda}{}_{\nu} = \eta_{\mu\nu} \ . \tag{2.102}$$

It is called a *general, homogeneous Lorentz transformation*.

This set actually forms a group called the *general (homogeneous) Lorentz group* and denoted $\mathcal{L}_{GH}$.

In fact, the Lorentz group we usually study in physics is the one we looked at first here, and it is a special representation of $\mathcal{L}_{GH}$:

Lorentz transformations: Matrix representation

Let $\left\{\boldsymbol{e}_{(\mu)}\right\}$ be an orthonormal basis of $\mathcal{M}$. Given $\boldsymbol{L} \in \mathcal{L}_{GH}$ a Lorentz transformation, we have a second orthonormal basis $\left\{\hat{\boldsymbol{e}}_{(\mu)}\right\}$ such that: $\forall \mu \in \{0,1,2,3\}$, $\boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right) = \hat{\boldsymbol{e}}_{(\mu)}$. We define an associated matrix Λ with components $\Lambda^{\mu}{}_{\nu}$ with respect to the basis $\left\{\boldsymbol{e}_{(\mu)}\right\}$ such that $\forall \mu \in \{0,1,2,3\}$, $\boldsymbol{e}_{(\mu)} = \Lambda^{\nu}{}_{\mu} \hat{\boldsymbol{e}}_{(\nu)}$; $\Lambda^{\nu}{}_{\mu}$ is actually the matrix associated with the Lorentz transformation $\mathcal{L}^{-1}$.
It verifies:

$$\Lambda^{\rho}{}_{\mu} \Lambda^{\lambda}{}_{\nu} \eta_{\rho\lambda} = \eta_{\mu\nu} \ . \tag{2.103}$$

These matrices have 16 components.
The set L_{GH} of all these matrices, given a specific basis $\left\{\boldsymbol{e}_{(\mu)}\right\}$, forms a *group*, called the *general homogeneous Lorentz group*. It has the same name as the group of transformations, even though, strictly speaking, it is just a representation of it. In the rest of these lecture notes, the Lorentz group will denote this set of matrices, with a fixed basis $\left\{\boldsymbol{e}_{(\mu)}\right\}$, rather than the set of transformations themselves. The results would be unchanged, but a bit more complicated to obtain.

The Lorentz group L_{GH} as defined here is the group of *passive* Lorentz transformations, i.e. of transformations that leave vectors invariant but change the orthonormal basis of $\mathcal{M}$. The linear mapping $\boldsymbol{L}$ whose inverse is represented by a matrix of L_{GH}, on the contrary, is an *active* transformation that keeps the coordinates of vectors fixed in both basis. Indeed, consider two orthonormal

bases, $\{\boldsymbol{e}_{(\mu)}\}$ and $\{\hat{\boldsymbol{e}}_{(\mu)}\}$ and an active Lorentz transformation $\boldsymbol{L} : \mathcal{M} \to \mathcal{M}$ such that:

$$\forall \mu \in \{0,1,2,3\},\ \boldsymbol{L}\left(\boldsymbol{e}_{(\mu)}\right) = \hat{\boldsymbol{e}}_{(\mu)}\ , \tag{2.104}$$

i.e.:

$$\boldsymbol{e}_{(\mu)} = \Lambda^{\nu}{}_{\mu}\hat{\boldsymbol{e}}_{(\nu)}\ . \tag{2.105}$$

where we denote by Λ the matrix associated to $\boldsymbol{L}$ in the basis $\{\hat{\boldsymbol{e}}_{(\mu)}\}$. Then, for any vector $\boldsymbol{v} \in \mathcal{M}$, we can write:

$$\boldsymbol{v} = v^{\mu}\boldsymbol{e}_{\mu} = \hat{v}^{\nu}\hat{\boldsymbol{e}}_{(\nu)} = \hat{v}^{\nu}\Lambda_{\nu}{}^{\mu}\boldsymbol{e}_{(\mu)}\ , \tag{2.106}$$

and indeed, keeping $\boldsymbol{v}$ fixed and changing $\{\boldsymbol{e}_{(\mu)}\}$ into $\{\hat{\boldsymbol{e}}_{(\nu)}\}$ via Λ, we get the old coordinates of $\boldsymbol{v}$, in $\{\boldsymbol{e}_{(\mu)}\}$, in terms of the new ones, in $\{\hat{\boldsymbol{e}}_{(\mu)}\}$:

$$\forall \mu \in \{0,1,2,3\},\ v^{\mu} = \Lambda_{\nu}{}^{\mu}\hat{v}^{\nu}\ . \tag{2.107}$$

On the other hand, we have that:

$$\boldsymbol{L}(\boldsymbol{v}) = \boldsymbol{L}\left(\hat{v}^{\mu}\hat{\boldsymbol{e}}_{(\mu)}\right) = \hat{v}^{\mu}\boldsymbol{L}\left(\hat{\boldsymbol{e}}_{(\mu)}\right) = \hat{v}^{\mu}\Lambda_{\mu}{}^{\nu}\boldsymbol{L}\left(\boldsymbol{e}_{(\nu)}\right) = \hat{v}^{\mu}\Lambda_{\mu}{}^{\nu}\hat{\boldsymbol{e}}_{(\nu)} = v^{\nu}\hat{\boldsymbol{e}}_{(\nu)}\ . \tag{2.108}$$

Therefore, the new vector $\boldsymbol{L}(\boldsymbol{v})$ has the same coordinates in $\{\hat{\boldsymbol{e}}_{(\mu)}\}$ than the old vector $\boldsymbol{v}$ in $\{\boldsymbol{e}_{(\mu)}\}$. These two types of transformations are completely equivalent; they only correspond to two different viewpoints.

Consider now a general Lorentz transformation $\Lambda^{\mu}{}_{\nu}$. It must verify:

$$\Lambda^{\rho}{}_{\mu}\Lambda^{\lambda}_{\nu}\eta_{\rho\lambda} = \eta_{\mu\nu}\ . \tag{2.109}$$

Putting $\mu = \nu = 0$ in this expression, we find that:

$$\left(\Lambda^{0}{}_{0}\right)^{2} = 1 + \left(\Lambda^{1}{}_{0}\right)^{2} + \left(\Lambda^{2}{}_{0}\right)^{2} + \left(\Lambda^{3}{}_{0}\right)^{2}\ . \tag{2.110}$$

In particular, we must have:

$$\left(\Lambda^{0}{}_{0}\right)^{2} \geq 1, \tag{2.111}$$

so that $\Lambda^{0}{}_{0} \geq 1$ or $\Lambda^{0}{}_{0} \leq -1$. A Λ with $\Lambda^{0}{}_{0} \geq 1$ is said to be *orthochronous* while it is said to be *non-orthochronous* if $\Lambda^{0}{}_{0} \leq -1$.

These names are justified by the following result. Let Λ be a Lorentz transformation and $\{\boldsymbol{e}_{(\mu)}\}$ an orthonormal basis of $\mathcal{M}$. Then, the following propositions are equivalent:

(i) Λ is orthochronous;

(ii) Λ preserves the time orientation of null vectors, i.e., for any $\boldsymbol{v} \in \mathcal{M}$ with $\boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)}$ such that $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) = 0$, we have that: v^0 and $\Lambda^0{}_\mu v^\mu$ have the same sign;

(iii) Λ preserves the time orientation of timelike vectors.

Indeed, consider a non-zero vector $\boldsymbol{v} = v^\mu \boldsymbol{e}_{(\mu)} \in \mathcal{M}$ that is either null or timelike. Using the Cauchy-Schwartz inequality of $\mathbb{R}^3$, we can write that:

$$\left(\Lambda^0{}_1 v^1 + \Lambda^0{}_2 v^2 + \Lambda^0{}_3 v^3\right)^2 \leq \left(\left(\Lambda^0{}_1\right)^2 + \left(\Lambda^0{}_2\right)^2 + \left(\Lambda^0{}_3\right)^2\right)\left(\left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2\right) . \quad (2.112)$$

Now, we have seen that $\Lambda^\rho{}_\mu \Lambda^\lambda{}_\nu \eta_{\rho\lambda} = \eta_{\mu\nu}$, which can be rewritten:

$$\Lambda^\mu{}_\rho \Lambda^\nu{}_\lambda \eta^{\rho\lambda} = \eta^{\mu\nu} . \quad (2.113)$$

This is a simple consequence of the fact that this expression is equivalent to $\Lambda\eta\Lambda^T = \eta$. Therefore, we have:

$$-\left(\Lambda^0{}_0\right)^2 + \left(\Lambda^0{}_1\right)^2 + \left(\Lambda^0{}_2\right)^2 + \left(\Lambda^0{}_3\right)^2 = -1 . \quad (2.114)$$

This implies that: $\left(\Lambda^0{}_0\right)^2 > \left(\Lambda^0{}_1\right)^2 + \left(\Lambda^0{}_2\right)^2 + \left(\Lambda^0{}_3\right)^2$. Now, since $\boldsymbol{v} \neq 0$ and it is timelike or null:

$$\left(v^0\right)^2 \geq \left(v^1\right)^2 + \left(v^2\right)^2 + \left(v^3\right)^2 , \quad (2.115)$$

so that, using the inequalities, we get:

$$\left(\Lambda^0{}_0 v^0\right)^2 > \left(\Lambda^0{}_1 v^1 + \Lambda^0{}_2 v^2 + \Lambda^0{}_3 v^3\right)^2 . \quad (2.116)$$

Let $\boldsymbol{w} \in \mathcal{M}$ such that $\boldsymbol{w} = \Lambda^0{}_0 \boldsymbol{e}_{(0)} + \Lambda^0{}_1 \boldsymbol{e}_{(1)} + \Lambda^0{}_2 \boldsymbol{e}_{(2)} + \Lambda^0{}_3 \boldsymbol{e}_{(3)}$. We know, because:

$$-\left(\Lambda^0{}_0\right)^2 + \left(\Lambda^0{}_1\right)^2 + \left(\Lambda^0{}_2\right)^2 + \left(\Lambda^0{}_3\right)^2 = -1 , \quad (2.117)$$

that $\boldsymbol{w}$ is timelike. In addition,

$$\left(\Lambda^0{}_0 v^0\right)^2 > \left(\Lambda^0{}_1 v^1 + \Lambda^0{}_2 v^2 + \Lambda^0{}_3 v^3\right)^2 \quad (2.118)$$

can be rewritten:

$$\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) \Lambda^0{}_\mu v^\mu < 0 , \quad (2.119)$$

which shows that $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w})$ and $\Lambda^0{}_\mu v^\mu$ have opposite signs.
Now, suppose that $w^0 = \Lambda^0{}_0 \geq 1$. Then, if $v^0 > 0$, we have $v^0 w^0 > 0$, and therefore, we know that $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) < 0$. But, then, necessarily, $\Lambda^0{}_\mu v^\mu > 0$. On the other hand, if $v^0 < 0$, we have $v^0 w^0 < 0$. which implies $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{w}) > 0$, so that $\Lambda^0{}_\mu v^\mu < 0$. Thus, we see that v^0 and $\Lambda^0{}_\mu v^\mu$ have the same sign. Following the same line of reasoning, if $\Lambda^0{}_0 \leq -1$, we can show that v^0 and $\Lambda^0{}_\mu v^\mu$ have opposite sign.

This shows that non-orthochronous Lorentz transformation have the unpleasant property of reversing the time orientation of all timelike and non-zero null vectors. This means that they transform 'forward clocks' into 'backward clocks'. Physically, this is not very attractive, and that is the reason why we choose to restrict the Lorentz group further. In the following, the Lorentz group will be the sub-group of $\mathcal{L}_{GH}$ made of *orthochronous transformations only*.
Finally, we need to introduce a further restriction. Consider the matrix relation:

$$\Lambda^T \eta \Lambda = \eta \ . \tag{2.120}$$

By taking its determinant and remembering that $\det \Lambda^T = \det \Lambda$, we get, straightforwardly:

$$(\det \Lambda)^2 = 1 \ , \tag{2.121}$$

and therefore, $\det \Lambda = \pm 1$. A Lorentz transformation Λ is said to be *proper* iff $\det \Lambda = 1$. Otherwise, it is *improper*. If we restrict our attention to orthochronous transformations, an improper orthochronous transformation is simply the composition of a proper orthochronous one with a transformation that changes the orientation of the spatial basis, from right-handed to left-handed and *vice-versa* (Can you prove it?). Such a change is quite arbitrary and does not contain any physical meaning whatsoever. Therefore, we choose to consider only orthochronous proper transformations, and we restrict ourselves to *admissible* orthonormal bases such that:

(i) $\boldsymbol{e}_{(\mathbf{0})}$ is timelike and future oriented;

(ii) $\left\{\boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{2})}, \boldsymbol{e}_{(\mathbf{3})}\right\}$ is right-handed.

Then, the set $\mathcal{L}$ of orthochronous proper Lorentz transformation forms a subgroup of $\mathcal{L}_{GH}$. We will call it the *Lorentz group* in the remainder of this course.
This Lorentz group contains a very important subgroup, $\mathcal{R}$, consisting of all the matrices of the

form:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & O & \\ 0 & & & \end{bmatrix}, \tag{2.122}$$

where O is an orthogonal matrix of determinant 1.

$\mathcal{R}$ is called the *rotation subgroup* of $\mathcal{L}$, and its elements are called *rotations*. For $\Lambda \in \mathcal{L}$, the following propositions are equivalent:

(i) Λ is a rotation;

(ii) ${\Lambda^1}_0 = {\Lambda^2}_0 = {\Lambda^3}_0 = 0$;

(iii) ${\Lambda^0}_1 = {\Lambda^0}_2 = {\Lambda^0}_3 = 0$;

(iv) ${\Lambda^0}_0 = 1$.

In the literature, $\mathcal{L}$ is often denoted $L_+^\uparrow$. This comes from the classification of Lorentz transformations hinted at above. Indeed, depending on the signs of $\det\Lambda$ and of whether ${\Lambda^0}_0 \geq 1$ or ${\Lambda^0}_0 \leq -1$, we have four different subsets of $\mathcal{L}_{GH}$, as illustrated in Table 2.1. Of course

	${\Lambda^0}_0 \geq 1$	${\Lambda^0}_0 \leq 1$
$\det\Lambda = 1$	$\mathcal{L} = L_+^\uparrow$ Proper orthochronous	$L_+^\downarrow$ Proper non-orthochronous
$\det\Lambda = -1$	$L_-^\uparrow$ Improper orthochronous	$L_-^\downarrow$ Improper non-orthochronous

Table 2.1: The four different subsets of $\mathcal{L}_{GH}$.

$\mathcal{L}_{GH} = L_+^\uparrow \cup L_-^\uparrow \cup L_+^\downarrow \cup L_-^\downarrow$, and these subsets are disjoint from each other.
Consider the following three mappings:

$$T = \eta\ ,\ P = -\eta \text{ and } Y = TP = PT = -I_4.$$

Clearly, $T \in L_-^\downarrow$, $P \in L_-^\uparrow$ and $Y \in L_+^\downarrow$, showing that all the subsets previously defined are non-empty. Geometrically, T corresponds to a *time-reversal*, P to an inversion of the orientation of space (called

parity), and Y is the combination of both transformations, i.e. an inversion of spacetime. Note that, among these four sets, only $L_+^\uparrow$ forms a group, since the other three are not stable by matrix multiplications: $T^2 = P^2 = Y^2 = I_4 \in L_+^\uparrow$.
Consider $\Lambda \in L_+^\uparrow = \mathcal{L}$. Each of the transformations T, P and Y induces a mapping of $\mathcal{L}$ into one of the other subsets of $\mathcal{L}_{GH}$ via the composition of linear applications. Indeed, for example, we have:

$$\det\ (T\Lambda) = \det T \det \Lambda = -1 \tag{2.123}$$

$$(T\Lambda)^0_0 = T^0_i \Lambda^i_0 = -\Lambda^0_0 \leq -1\ . \tag{2.124}$$

Therefore, $T\Lambda \in L_-^\downarrow$. Conversely, for any $U \in L_-^\downarrow$, $TU \in L_+^\uparrow$. Hence, the time reversal, T induces a mapping between $L_+^\uparrow$ and $L_-^\downarrow$; it is easily checked that this mapping is bijective. Similarly, P indices a bijective mapping between $L_+^\uparrow$ and $L_-^\uparrow$ and Y induces a bijective mapping between $L_+^\uparrow$ and $L_+^\downarrow$. Therefore, the study of the entire Lorentz group $\mathcal{L}_{GH}$ can be reduced to the study of $\mathcal{L} = L_+^\uparrow$: it is the only one which has the nice group structure, and the elements of the other parts of $\mathcal{L}_{GH}$ can be deduced from the ones of $L_+^\uparrow$. Nevertheless, one must not forget the other parts of the group: the time reversal and parity symmetries, as well as their composition do not play any role in classical physics where all the phenomena are invariant under these symmetries, but they are essential in Quantum Mechanics since some aspects of the theory turn out to be affected by the change in time orientation and/or space orientation.

2.4.2 Back to physics: interpretation of the components of a Lorentz transformation

Elements of $\mathcal{L}$ have $4 \times 4 = 16$ components. Nevertheless, because of Eq. (2.97) they are not all independent, and the orthochronous and proper characters also restrict the number of free parameters. Some of the remaining components have interesting physical interpretations that we will try to investigate now.
Let us start with two admissible bases $\left\{\boldsymbol{e}_{(\mu)}\right\}$ and $\left\{\hat{\boldsymbol{e}}_{(\nu)}\right\}$ corresponding, physically, to two frames of reference, F_1 and F_2. Consider the Lorentz transformation Λ such that: $\forall \mu \in \{0,1,2,3\}, \hat{\boldsymbol{e}}_{(\mu)} = \Lambda^\nu{}_\mu \boldsymbol{e}_{(\nu)}$; it corresponds to the change of frame $F_2 \rightarrow F_1$, in which the components of a vector $\boldsymbol{v}$ change as:

$$v^\nu = \Lambda^\nu{}_\mu \hat{v}^\mu\ . \tag{2.125}$$

Be careful that the role of the hatted and non-hatted coordinates are reversed compared to the previous section. First, let us consider a given worldline on which two events, x and $x + \delta x$ are at

rest in F_2. This means that the spatial point represented by x and $x + \delta x$ is a single spatial point, at rest with respect to F_2. This could be the worldline of a particle at rest in F_2. Write, in each basis: $\boldsymbol{\delta x} = \delta x^{\mu} \boldsymbol{e}_{(\mu)} = \delta \hat{x}^{\nu} \hat{\boldsymbol{e}}_{(\nu)}$, the separation vector between the two events. The fact that the two events are at rest in F_2 simply means that: $\delta \hat{x}^1 = \delta \hat{x}^2 = \delta \hat{x}^3 = 0$. $\delta \hat{x}^0$ then is the time separation between the two events, as measured in F_2. We have that:

$$\delta x^{\nu} = \Lambda^{\nu}{}_{0} \delta \hat{x}^0 \ . \tag{2.126}$$

Therefore:

$$\forall i \in \{1, 2, 3\}, \frac{\delta x^i}{\delta x^0} = \frac{\Lambda^i{}_0}{\Lambda^0{}_0} \ . \tag{2.127}$$

These ratios are constant and independent on the particular point at rest in F_2 that we consider. Physically, they correspond to the components of the *standard* $3-$*velocity of* F_2 *with respect to* F_1:

$$\vec{u} = u^1 \boldsymbol{e}_{(1)} + u^2 \boldsymbol{e}_{(2)} + u^3 \boldsymbol{e}_{(3)} \ , \tag{2.128}$$

with:

$$\forall i \in \{1, 2, 3\}, u^i = \frac{\Lambda^i{}_0}{\Lambda^0{}_0} \ . \tag{2.129}$$

Conversely, if we consider two events at rest in F_1, we find that the 3-velocity of F_1 w.r.t. F_2 is given by:

$$\vec{\hat{u}} = \hat{u}^1 \hat{\boldsymbol{e}}_{(1)} + \hat{u}^2 \hat{\boldsymbol{e}}_{(2)} + \hat{u}^3 \hat{\boldsymbol{e}}_{(3)} \ , \tag{2.130}$$

where:

$$\forall i \in \{1, 2, 3\}, \hat{u}^i = \frac{\left(\Lambda^{-1}\right)^i{}_0}{\left(\Lambda^{-1}\right)^0{}_0} = -\frac{\Lambda^0{}_i}{\Lambda^0{}_0} \ . \tag{2.131}$$

To carry on a bit further, let us observe that:

$$\sum_{i=1}^{3} \left(u^i\right)^2 = \sum_{i=1}^{3} \left(\hat{u}^i\right)^2 \tag{2.132}$$

$$= \frac{\left(\Lambda^0{}_0\right)^2 - 1}{\left(\Lambda^0{}_0\right)^2} \ . \tag{2.133}$$

This shows that: $\|\vec{u}\| = \left\|\vec{\hat{u}}\right\| = \beta$, i.e. that the magnitude of the 3-velocity of one frame relative to the other is the same in both cases, and is equal to[3]:

[3]Remember that we restrict our analysis to orthochronous transformations, for which $\Lambda^0{}_0 \geq 1$.

$$\beta = \sqrt{1 - \frac{1}{\left(\Lambda^0{}_0\right)^2}} \, . \tag{2.134}$$

Equivalently:

$$\Lambda^0{}_0 = \frac{1}{\sqrt{1-\beta^2}} = \gamma \, , \tag{2.135}$$

is known as the *Lorentz factor* of the transformation. Note that we have: $0 \leq \beta^2 < 1$. Since $\Lambda^0{}_0 = 1$ iff Λ is a rotation, we have that, for a rotation, $\beta = 0$, as expected physically. If we concentrate on Lorentz transformations Λ that are not rotations, $\beta \neq 0$, and we can write:

$$\vec{u} = \beta \vec{d} \, , \tag{2.136}$$

with $\vec{d} = d^i \boldsymbol{e}_{(i)}$ and $\forall i \in \{1,2,3\}$, $d^i = u^i/\beta$. Since $\beta = \|\vec{u}\|$, $\vec{d}$ is the *direction 3-vector* of F_2 relative to F_1, and its components, d^i, are the direction cosines of the line along which an observer in F_1 sees the fixed events in F_2 moving. Similarly, we can write:

$$\vec{\hat{u}} = \beta \vec{\hat{d}} \, , \tag{2.137}$$

with $\vec{\hat{d}} = \hat{d}^i \boldsymbol{e}_{(i)}$ and $\forall i \in \{1,2,3\}$, $\hat{d}^i = \hat{u}^i/\beta$. The $\hat{d}^i$'s are the direction cosines of the line along which an observer in F_2 sees the fixed events in F_1 moving. Using all these relations, we find that:

$$\forall i \in \{1,2,3\}, \left\{ \begin{array}{ll} \Lambda^i{}_0 & = \frac{\beta}{\sqrt{1-\beta^2}} d^i = \beta\gamma d^i \\ \Lambda^0{}_i & = -\frac{\beta}{\sqrt{1-\beta^2}} \hat{d}^i = -\beta\gamma \hat{d}^i \, . \end{array} \right. \tag{2.138}$$

We have fixed 7 components of the Lorentz transformation by using physically measurable quantities. This allows us to get some insight on the physics, already. Indeed, since for an event at rest in F_2, we have: $\delta x^\mu = \Lambda^\mu{}_0 \delta \hat{x}^0$, we see that the time interval between the two events in F_1, δx^0, is given, in terms of the time interval in F_2, $\delta \hat{x}^0$, by:

$$\delta x^0 = \frac{1}{\sqrt{1-\beta^2}} \delta \hat{x}^0 = \gamma \delta \hat{x}^0 \, . \tag{2.139}$$

Therefore, $\delta x^0 > \delta \hat{x}^0$, because $1/\sqrt{1-\beta^2} > 1$, which means that, considered in F_1, there is a *time dilation* between the two events, as compared to the same events considered in F_2: for an observer in F_2, the clocks in F_1 are running slow. Please note that in the limit $\beta \to 1$, the effect becomes

infinite: this is reminiscent of the fact that the speed of light (chosen equal to 1 in these lecture notes) is an unattainable limit for massive particles (such as clocks) in the theory.
Now that we have seen what happens to two events at rest in F_2, we can analyse the complementary situation of two events that are *simultaneous* in F_2, such that $\delta\hat{x}^0 = 0$ and a priori, $\delta\hat{x}^i \neq 0$ for $i \in \{1, 2, 3\}$ (they are not at the same spatial location). Then:

$$\delta x^0 = \Lambda^0{}_i \delta\hat{x}^i = -\beta\gamma \left(\hat{d}^1 \delta\hat{x}^1 + \hat{d}^2 \delta\hat{x}^2 + \hat{d}^3 \delta\hat{x}^3 \right) . \tag{2.140}$$

This means that, when the Lorentz transformation is not a rotation ($\beta \neq 0$), *the time difference between two events* in F_1 will not be zero in general: the two events will not be considered simultaneous in F_1. This illustrates the *relativity of simultaneity*. The two events will be seen as simultaneous iff:

$$\hat{d}^1 \delta\hat{x}^1 + \hat{d}^2 \delta\hat{x}^2 + \hat{d}^3 \delta\hat{x}^3 = 0 , \tag{2.141}$$

which means iff the *line joining the two events is perpendicular to the direction of relative motion between F_1 and F_2*. It is not hard to see that the previous equation characterises a plane in F_2, known as the plane of simultaneity of x.

We see that the Lorentz group is vast. In particular, because it contains all the spatial rotations, the spatial axes $\left\{ \boldsymbol{e}_{(1)}, \boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)} \right\}$ and $\left\{ \hat{\boldsymbol{e}}_{(1)}, \hat{\boldsymbol{e}}_{(2)}, \hat{\boldsymbol{e}}_{(3)} \right\}$ can have any relative position as long as the orientation is preserved. This is reassuring because it is in line with the isotropy of space that we would like for physics, but it cloaks a lot of simple relations that appear much clearer if we restrain our study to a smaller subset of Lorentz transformations. Specifically, we will concentrate of bases that are such that their spatial axes have a definite, simple relationship: for any $i \in \{1, 2, 3\}$, we will align $\boldsymbol{e}_{(i)}$ and $\hat{\boldsymbol{e}}_{(i)}$. This is what we called a translation in the Galilean context. To start this process, let us first suppose that: $d^1 = 1 = -\hat{d}^1$ and $d^2 = d^3 = 0 = \hat{d}^2 = \hat{d}^3$. Then the direction vector is $\vec{d} = \boldsymbol{e}_{(1)} = -\vec{\hat{d}}$. That means that F_2 moves, relative to F_1 along the axis $\boldsymbol{e}_{(1)}$, with a velocity in the direction of the positive values of x^1. The form of the Lorentz transformation is then:

$$\Lambda = \begin{bmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ 0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ 0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{bmatrix} . \tag{2.142}$$

Further, using the condition Eq. (2.97), we get:

$$\Lambda = \begin{bmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ 0 & 0 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{bmatrix}, \tag{2.143}$$

and the matrix:

$$O = \begin{bmatrix} \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_2 & \Lambda^3{}_3 \end{bmatrix}. \tag{2.144}$$

is an orthogonal transformation of $\mathbb{R}^2$ with determinant 1: it is a rotation of $\mathbb{R}^2$. Its effect is to rotate the axes $\boldsymbol{e}_{(2)}$ and $\boldsymbol{e}_{(3)}$, keeping $\boldsymbol{e}_{(1)}$ fixed. We will therefore define the *standard configuration* to be the one in which $\hat{\boldsymbol{e}}_{(2)} = \boldsymbol{e}_{(2)}$ and $\hat{\boldsymbol{e}}_{(3)} = \boldsymbol{e}_{(3)}$, i.e. with this rotation of $\mathbb{R}^2$ being simply the identity map:

$$\Lambda = \begin{bmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{2.145}$$

The transformation of coordinates of an event $\boldsymbol{x} = x^\mu \boldsymbol{e}_{(\mu)} = \hat{x}^\mu \hat{\boldsymbol{e}}_{(\mu)}$, with $x^\mu = \Lambda^\mu{}_\nu \hat{x}^\nu$ is therefore simply:

$$\begin{cases} x^0 = \gamma\hat{x}^0 + \beta\gamma\hat{x}^1 & (2.146) \\ x^1 = \gamma\hat{x}^1 + \beta\gamma\hat{x}^0 & (2.147) \\ x^2 = \hat{x}^2 & (2.148) \\ x^3 = \hat{x}^3 . & (2.149) \end{cases}$$

The inverse transformation is then given by:

$$\Lambda^{-1} = \begin{bmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.150}$$

so that we have also:

$$\begin{cases} \hat{x}^0 = \gamma x^0 - \beta\gamma x^1 & (2.151) \\ \hat{x}^1 = \gamma x^1 - \beta\gamma x^0 & (2.152) \\ \hat{x}^2 = x^2 & (2.153) \\ \hat{x}^3 = x^3 \ . & (2.154) \end{cases}$$

Such transformations are called *special Lorentz transformations*. Strictly speaking, the velocity β is positive or zero, but because a special Lorentz transformation and its inverse only differ by a sign in front of β, it is customary to allow $\beta \in]-1,1[$. Then, by choosing $\beta > 0$ when the motion is positive along the $\boldsymbol{e}_{(1)}$ axis and $\beta < 0$ when it occurs in the negative direction, we can write any special Lorentz transformation as:

$$\forall \beta \in]-1,1[,\ \Lambda(\beta) = \begin{bmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.155}$$

with $\gamma(\beta) = 1/\sqrt{1-\beta^2}$. Usually, such a matrix $\Lambda(\beta)$ is called a *boost in the* $\boldsymbol{e}_{(1)}$*-direction*. Then the set of special Lorentz transformations is a subgroup of $\mathcal{L}$. The composition of two boosts in the $\boldsymbol{e}_{(1)}$-direction, $\Lambda(\beta_1)$ and $\Lambda(\beta_2)$ is a boost $\Lambda(\beta)$, with:

$$\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2} \ . \tag{2.156}$$

One should note that the composition of boosts along different directions is generally not a boost in any specific direction. The composition of boosts has a simple physical interpretation. Consider three frames, F_1, F_2 and F_3 related by boosts along the $\boldsymbol{e}_{(1)}$ direction. If the speed of F_2 relative to F_1 is β_1, and the speed of F_3 relative to F_2 is β_2, then the speed of F_3 relative to F_1 is *not* $\beta_1 + \beta_2$, as one would have expected from Galilean invariance, but:

$$\boxed{\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\beta_2} \ .} \tag{2.157}$$

This law is the *relativistic addition of velocities*. It is actually a law of *non-additivity of the velocities*. One can notice that if β_1 and β_2 have the same sign, i.e. if the motions of F_2 relative to

F_1 and F_3 relative to F_2 happens in the same direction, then β is always *smaller* than $\beta_1 + \beta_2$. Also, one sees that when $\beta_1 \to 1$ and $\beta_2 \to 1$, $\beta \to 1$: the speed of light indeed acts as a limit speed. The non-additivity of velocities is quite an inconvenient fact. We would like to define a 'velocity parameter', θ, that is additive when we compose special Lorentz transformations. Let us therefore suppose that we have two special Lorentz transformations with speeds β_1 and β_2 along the $\boldsymbol{e}_{(1)}$ direction. We associate θ_1 and θ_2 respectively to each transformation, and we must have relations of the form: $\beta_1 = f(\theta_1)$, $\beta_2 = f(\theta_2)$, where $f : \mathbb{R} \to \mathbb{R}$ is a function. Now, $\beta = (\beta_1 + \beta_2)/(1 + \beta_1\beta_2)$, and we must have an associated θ with $\beta = f(\theta)$ and $\theta = \theta_1 + \theta_2$ (additivity). This means that the function f must verify:

$$f(\theta_1 + \theta_2) = \frac{f(\theta_1) + f(\theta_2)}{1 + f(\theta_1) f(\theta_2)} . \tag{2.158}$$

This functional equation as at least one solution: $f = \tanh$. Therefore, we can choose our parameter θ to be: $\theta = \operatorname{atanh}(\beta)$. Because the function atanh is bijective from $]-1, 1[$ onto $\mathbb{R}$, the speed of light, $\beta = \pm 1$ corresponds to $\theta = \pm\infty$.

Using this *velocity parameter*, θ, the *hyperbolic form* of a special Lorentz transformation is, for any $\theta \in \mathbb{R}$:

$$\Lambda(\theta) = \begin{bmatrix} \cosh(\theta) & -\sinh(\theta) & 0 & 0 \\ -\sinh(\theta) & \cosh(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{2.159}$$

Using this form of the special Lorentz transformations, we can get a very important result, namely the fact that any proper, orthochronous Lorentz transformation can be written as the composition of two rotations and a boost:

Decomposition of a Lorentz transformation

Let $\Lambda \in \mathcal{L}$. Then, there exists a real number θ and two rotations R_1 and R_2 in $\mathcal{R}$ such that:

$$\Lambda = R_1 \Lambda(\theta) R_2 . \tag{2.160}$$

2.4.3 Spacetime diagrams

Thanks to the decomposition (2.160), we know that by applying the correct spatial rotations, we can always bring a physical situation to its description in terms of boosts only; that means that the

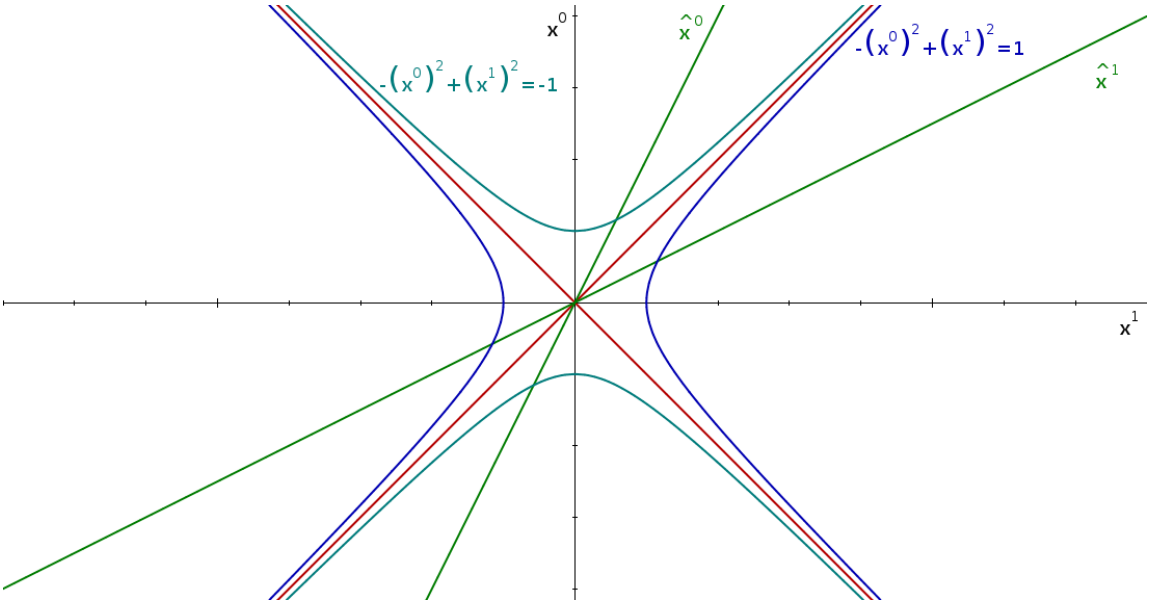


Figure 2.2: A spacetime diagram. The lightcone is represented by the two red lines, $x^0 = \pm x^1$.

physical content of special relativity is almost entirely contained in the properties of special Lorentz transformations. This is very convenient, since when considering these transformations, two spatial coordinates are left unchanged. Therefore, we can easily picture the effects of a boost on a piece of paper by simply suppressing the two dimensions that are not affected by the boost. Then, we have a 2-dimensional space spanned by $\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}\}$ that we can represent on paper. This leads to the construction of *spacetime diagrams*; see Fig. 2.2. The procedure is as follow:

- We draw two perpendicular axes along two unit vectors $\boldsymbol{e}_{(\mathbf{0})}$ and $\boldsymbol{e}_{(\mathbf{1})}$ that we label x^0 and x^1. Note that the perpendicularity is just convenient, it does not, in principle, correspond to the orthogonality in $\mathcal{M}$. The labels x^0 and x^1 are then the coordinates of the event x in the frame $\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}\}$. The intersection of the two axes is the origin of the frame, O.
- When there is a boost with parameter β (or equivalently, $\theta = \operatorname{atanh}(\beta)$) relating $\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}\}$ to another orthonormal basis $\{\hat{\boldsymbol{e}}_{(\mathbf{0})}, \hat{\boldsymbol{e}}_{(\mathbf{1})}\}$, in which an event $\boldsymbol{x} = \hat{x}^\mu \hat{\boldsymbol{e}}_{(\boldsymbol{\mu})}$, the axis labelled $\hat{x}^0$ is to be understood as the set corresponding to events with $\hat{x}^1 = 0$, i.e., $x^1 = \beta x^0$, with $\beta \in]-1, 1[$; in other words, this is a line passing through O with a slope $1/\beta$. Similarly, the axis labelled $\hat{x}^1$ is taken to be at $\hat{x}^0 = 0$, and therefore, $x^0 = \beta x^1$. This is therefore the line

through O with slope β. The lightcone corresponds to the two lines $x^0 = \pm x^1$.

The set of unit timelike (*resp.* spacelike) vectors corresponds then to the branches of hyperbolæ $-\left(x^0\right)^1 + \left(x^1\right)^2 = -1$ (*resp.* $-\left(x^0\right)^1 + \left(x^1\right)^2 = 1$). Since a boost leave the quadratic form $\boldsymbol{q}$ invariant, these hyperbolæ also correspond to the 'hyperbolæ' $-\left(\hat{x}^0\right)^1 + \left(\hat{x}^1\right)^2 = -1$ and $-\left(\hat{x}^0\right)^1 + \left(\hat{x}^1\right)^2 = 1$. This shows that by plotting the axes $\hat{x}^0$ and $\hat{x}^1$ as lines, we have distorted figures in the $\left(\hat{x}^0, \hat{x}^1\right)$ coordinate system. Indeed $-\left(\hat{x}^0\right)^1 + \left(\hat{x}^1\right)^2 = \pm 1$ should intersect the axes $\hat{x}^0$ and $\hat{x}^1$ at $\hat{x}^1 = \pm 1$ and $\hat{x}^0 = \pm 1$ respectively. But they don't on the graph; rather, because we have:

$$x^0 = -\beta\gamma\hat{x}^1 + \gamma\hat{x}^0 \tag{2.161}$$

$$x^1 = \gamma\hat{x}^1 - \beta\gamma\hat{x}^0 \,, \tag{2.162}$$

the points of intersection between the hyperbola $-\left(x^0\right)^1 + \left(x^1\right)^2 = -1$ and the axis $\hat{x}^0$, which should be at $\hat{x}^0 = \pm 1$, are actually, in the un-hatted coordinate system, at: $x^0 = \pm\gamma$ and $x^1 = -(\pm)\beta\gamma$, i.e., at a distance from the origin O given by: $\gamma\sqrt{1+\beta^2}$. A similar factor applies on the other hyperbola. That means that the 'true' coordinates in the $\left(\hat{x}^0, \hat{x}^1\right)$ basis can be obtained by projecting parallely along each axis and applying the scaling factor $\gamma\sqrt{1+\beta^2}$.

The 'lines of simultaneity' in F_2 appear with a slope, parallel to the axis $\hat{x}^1$ on this graph. We see immediately that they do not correspond to the line of simultaneity in F_1, which are horizontal. That illustrates clearly the relativity of simultaneity. Of course, since we have suppressed 2 dimensions here, in fact these lines of simultaneity are 3-D spaces, and they intersect, not at a point, like on the diagram, but on a plane: two observers in relative inertial motion agree on the simultaneity of events in a single plane: we have proven that previously.

Finally, let us analyse a striking consequence of this relativity of simultaneity: the 'contraction of length'. Consider two reference frames F_1 and F_2 with orthonormal bases $\left\{\boldsymbol{e}_{(\mu)}\right\}$ and $\left\{\hat{\boldsymbol{e}}_{(\mu)}\right\}$ respectively, in relative inertial motion and whose spatial axes are in standard configuration, so that the Lorentz transformation between them is a boost of velocity β: $\forall\mu \in \{0,1,2,3\}, \boldsymbol{e}_{(\mu)} = \Lambda\left(\beta\right)^{\nu}{}_{\mu}\hat{\boldsymbol{e}}_{(\nu)}$. The situation is represented on a spacetime diagram on Fig. 2.3. Let us consider a rigid rod at rest in F_2, and put along the $\hat{x}^1$ axis, whose end points are located at $\hat{x}^1 = 0$ and $\hat{x}^1 = 1$. The length of the rod has measured in F_2 is therefore equal to 1. On the spacetime diagram, at any time $\hat{x}^0$ in F_2, it corresponds to the line segment $[A, B]$, parallel to the axis $\hat{x}^1$: the worldlines of each extremity of the rod are the lines $\hat{x}^1 = 0$ and $\hat{x}^1 = 1$, respectively. On the other hand, in F_1, the length of the rod is the Euclidean length of the line segment joining these two worldlines at the *same time coordinate*

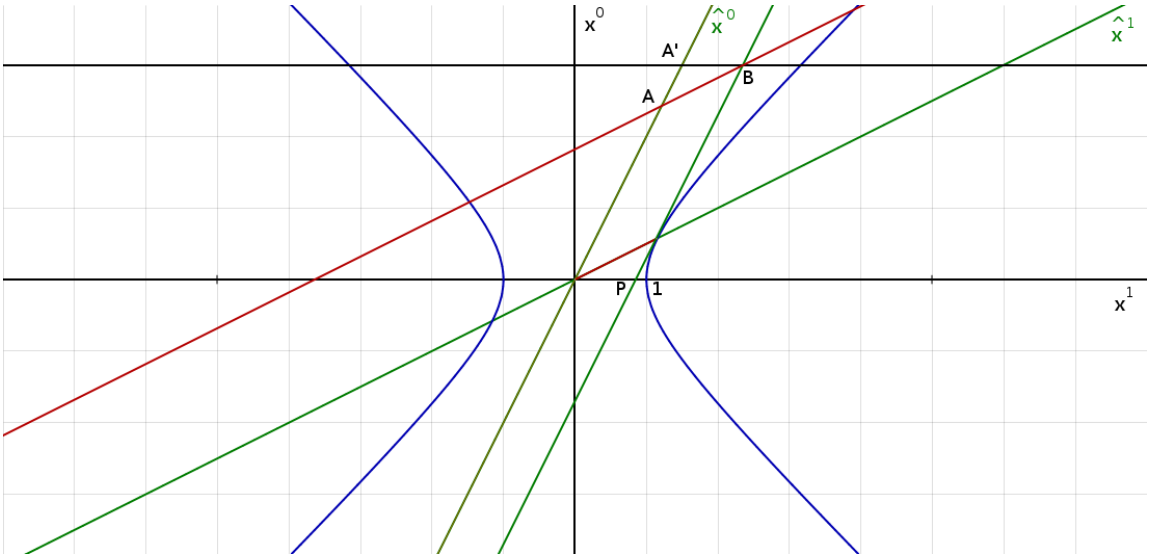


Figure 2.3: A spacetime diagram for the contraction of length. See description in the text.

in F_1, x^0, which corresponds to an horizontal line segment like $[A', B]$: this is the value of the x^1 coordinate of P on the diagram. By construction, this is less than 1 and therefore, the rod appears shorter in F_1 than in F_2. We can make that more precise via a calculation. Let us consider a rod lying at rest along the $\hat{x}^1$-axis of F_2 between $\hat{x}^1_0$ and $\hat{x}^1_1$, with $\hat{x}^1_0 < \hat{x}^1_1$. The length of that rod in F_2 is therefore $\Delta\hat{x}^1 = \hat{x}^1_1 - \hat{x}^1_0$. Consider first the left-hand end point of the rod. Its coordinates in F_2 are $(\hat{x}^0, \hat{x}^1_0, 0, 0)$ with $\hat{x}^0 \in \mathbb{R}$. Similarly, the right-hand end point has coordinates $(\hat{x}^0, \hat{x}^1_1, 0, 0)$ with the same $\hat{x}^0 \in \mathbb{R}$ (simultaneity in F_2).

In F_1, the length of the rod will be determined by considering the position of its endpoints simultaneously, that is at the same value of x^0 (and not $\hat{x}^0$, as in F_2). But, for any fixed x^0, the Lorentz transformation reads:

$$\begin{cases} \hat{x}^1_0 = \gamma x^1_0 - \beta\gamma x^0 & (2.163) \\ \hat{x}^1_1 = \gamma x^1_1 - \beta\gamma x^0 \ . & (2.164) \end{cases}$$

Hence, we see that the length of the rod in F_1, Δx^1 is linked to this length in F_2 by: $\Delta\hat{x}^1 = \gamma\Delta x^1$. Since $\gamma > 1$, that means that the length in F_1 appears *smaller* than in F_2: this is the *relativistic contraction of lengths*. Note that all these relativistic effects, contraction of lengths, dilation of time etc., are entirely symmetrical.

2.5 Particles in Minkowski spacetime

2.5.1 Curves in spacetime

In this section, we are going to analyse the description of particles in Special Relativity. We will start with photons and other massless particles and then turn to massive particles. First we need to introduce the concept of parametrised curve:

Parametrised curve

A parametrised curve is a map $c : I \subseteq \mathbb{R} \to \mathcal{M}$ which to any value of a real parameter λ in the interval I associates an event $c(\lambda)$ in $\mathcal{M}$.
The image $c(I) = C$ is the (geometric) curve and does not depend on the specific parameter λ used to describe it.

Given a parametrised curve $c(\lambda)$, we can define its tangent vector as follows:

$$\boldsymbol{X(\lambda)} = \lim_{\delta\lambda\to 0} \frac{c(\lambda+\delta\lambda) - c(\lambda)}{\delta\lambda} , \tag{2.165}$$

i.e. as the limit of the separation vector between events infinitely closed along the curve.

Let us consider a specific reference frame $\{\boldsymbol{e}_{(\mu)}\}$. An event $x(\lambda)$ along a curve C has components (coordinates as a point) $x^\mu(\lambda)$ and another event $x(\lambda+\delta\lambda)$ has components (coordinates as a point) $x^\mu(\lambda+\delta\lambda)$ so that, at first order in $\delta\lambda$:

$$x^\mu\left(\lambda+\delta\lambda\right) = x^\lambda + \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\delta\lambda \ . \tag{2.166}$$

Therefore, at the limit, the tangent vector has components:

$$X^\mu(\lambda) = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}(\lambda) \ . \tag{2.167}$$

Tangent vectors have a very useful interpretation in terms of *directional derivatives* along curves that will be very useful in General Relativity. Let us consider a map $f : \mathcal{M} \to \mathbb{R}$; then, we get, at first order in $\delta\lambda$:

$$f\left(x^\mu\left(\lambda+\delta\lambda\right)\right) = f\left(x^\mu(\lambda)\right) + \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\frac{\partial f}{\partial x^\nu}(\lambda)\delta\lambda \ . \tag{2.168}$$

Therefore, we see that we can write:

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda}(\lambda) = X^\mu(\lambda)\frac{\partial f}{\partial x^\mu}(\lambda) \ , \tag{2.169}$$

Therefore, we can formally write:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} = X^\mu \frac{\partial}{\partial x^\mu} \ . \tag{2.170}$$

This looks exactly like $\boldsymbol{X} = X^\mu \boldsymbol{e}_{(\mu)}$ provided we identify the basis vectors $\boldsymbol{e}_{(\mu)}$ with the partial derivative operators $\frac{\partial}{\partial x^\mu}$. This leads us to:

Tangent vectors as derivative operators

The tangent vector to a curve parametrised by $\lambda \in \mathbb{R}$ is a *differential operator* acting on functions $f : \mathcal{M} \to \mathbb{R}$ as:

$$X(f) = \frac{\mathrm{d}f}{\mathrm{d}\lambda} \ . \tag{2.171}$$

Given a frame $\{\boldsymbol{e}_{(\mu)}\}$ and the associated coordinates, x^μ, we have:

$$\boldsymbol{e}_{(\mu)} = \frac{\partial}{\partial x^\mu} \ , \tag{2.172}$$

and:

$$\boldsymbol{X} = X^\mu \frac{\partial}{\partial x^\mu} \ . \tag{2.173}$$

We will say that a curve in $\mathcal{M}$ is *timelike* iff its tangent vector $\boldsymbol{v}$ is timelike, $\boldsymbol{\eta}(\boldsymbol{v}, \boldsymbol{v}) < 0$, at every event along the curve. Similarly, we will say that a curve in $\mathcal{M}$ is *lightlike* or *null* iff its tangent vector $\boldsymbol{k}$ is lightlike, $\boldsymbol{\eta}(\boldsymbol{k}, \boldsymbol{k}) = 0$, at every event along the curve. Timelike and lightlike curves are often called *worldlines* of the associated particles travelling along them.

Non-coordinate bases

As we have seen in subsection 2.3.3, we can construct non-orthonormal coordinate bases such as the one for spherical coordinates, $\{\hat{x}^\mu\} = \{t, r, \theta, \phi\}$ which we now know that we can write:

$$\begin{cases} \boldsymbol{e}_{(0)} = \dfrac{\partial}{\partial t}, \ \hat{\boldsymbol{e}}_{(1)} = \dfrac{\partial}{\partial r} & (2.174) \\ \hat{\boldsymbol{e}}_{(2)} = \dfrac{\partial}{\partial \theta}, \ \hat{\boldsymbol{e}}_{(3)} = \dfrac{\partial}{\partial \phi} \ . & (2.175) \end{cases}$$

Clearly, since, for any function f, we have:

$$\frac{\partial f}{\partial \hat{x}^i} = \frac{\partial x^j}{\partial \hat{x}^i}\frac{\partial f}{\partial x^j} \ , \tag{2.176}$$

we can express the new basis in terms of the old one:

$$\begin{cases} \hat{\boldsymbol{e}}_{(1)} = \dfrac{\partial}{\partial r} = \sin\theta\cos\phi\dfrac{\partial}{\partial x} + \sin\theta\sin\phi\dfrac{\partial}{\partial y} + \cos\theta\dfrac{\partial}{\partial z} & (2.177) \\ \hat{\boldsymbol{e}}_{(2)} = \dfrac{\partial}{\partial \theta} = r\cos\theta\cos\phi\dfrac{\partial}{\partial x} + r\cos\theta\sin\phi\dfrac{\partial}{\partial y} - r\sin\theta\dfrac{\partial}{\partial z} & (2.178) \\ \hat{\boldsymbol{e}}_{(3)} = \dfrac{\partial}{\partial \phi} = -r\sin\theta\sin\phi\dfrac{\partial}{\partial x} + r\sin\theta\cos\phi\dfrac{\partial}{\partial y} \ . & (2.179) \end{cases}$$

Again we see immediately that the vectors are not unit vectors. We can of course define an orthonormal basis:

$$\begin{cases} \tilde{\boldsymbol{e}}_{(1)} = \hat{\boldsymbol{e}}_{(1)} & (2.180) \\ \tilde{\boldsymbol{e}}_{(2)} = \dfrac{1}{r}\hat{\boldsymbol{e}}_{(2)} & (2.181) \\ \tilde{\boldsymbol{e}}_{(3)} = \dfrac{1}{r\sin\theta}\hat{\boldsymbol{e}}_{(3)} \ , & (2.182) \end{cases}$$

but there is no coordinate system $\{\hat{x}^i\}$ such that $\hat{\boldsymbol{e}}_{(i)} = \frac{\partial}{\partial \hat{x}^i}$. Therefore, this orthonormal basis (the usual moving basis of spherical coordinates) is *not* a coordinate basis. An easy way to see that is to realise that partial derivatives ought to commute on smooth functions. Clearly, these $\tilde{\boldsymbol{e}}_{(i)}$ do not commute.

2.5.2 Massless particles

Massless particles such as photons move along lightlike curves. Let us pick up such a curve with parameter λ and the associated tangent vector $\boldsymbol{k}$ such that, in a frame $\{\boldsymbol{e}_{(\mu)}\}$, with $\boldsymbol{k} = k^\mu \boldsymbol{e}_{(\mu)}$:

$$\boldsymbol{\eta}(\boldsymbol{k},\boldsymbol{k}) = 0 \implies k^0 = \pm\|\vec{k}\| \ , \tag{2.183}$$

where we defined $\vec{k} = k^1\boldsymbol{e}_{(1)} + k^2\boldsymbol{e}_{(2)} + k^3\boldsymbol{e}_{(3)}$ the 3-vector corresponding to the direction of propagation of the photon at λ. $\boldsymbol{k}$ is called the *4-momentum* of the photon.

In classical field theory, the propagation of light is actually described by electromagnetism, which is a theory of waves. However, we are interested in the geometric optics limit of this theory, for which

the wavelengths of the waves are much smaller than other typical scales involved in the problems and for which it can be shown that the vector $\boldsymbol{k}$ is actually the *wavevector*, so that $k_\mu = \frac{\partial \Phi}{\partial x^\mu}$ is the variation of the wave's phase. In that case, if we choose $\boldsymbol{k}$, it can be shown that (see [6] for details):

$$k^0 = \pm\hbar\omega = \pm E \ , \tag{2.184}$$

where $E(\lambda)$ is the energy of the photon measured by the observer attached to the frame $\left\{\boldsymbol{e}_{(\mu)}\right\}$ and $\pm$ stands for future and past directed vectors respectively. Thus, we see that: $\|\vec{k}\| = E$, so we can write:

$$\boldsymbol{k} = E\left[\pm\boldsymbol{e}_{(\mathbf{0})} + \vec{\hat{k}}\right] \ , \tag{2.185}$$

where $\vec{\hat{k}}$ is the instantaneous direction of propagation of the photon and is orthogonal to $\boldsymbol{e}_{(0)}$. It is thus a unit spacelike vector. $\vec{k} = E\vec{\hat{k}}$ is the *3-momentum of the photon*. If the photons propagate freely (which in the geometric optics limit means that they do not encounter mirrors or a dioptre), according to the principle of inertia, they propagate in straight lines, i.e. that $\vec{\hat{k}}(\lambda)$ is a constant and, then E is also a constant.

To illustrate how Lorentz boosts act on photons consider a source S emitting photons of energy E isotropically in all directions in its rest frame (reference frame in which it is at rest) $\left\{\boldsymbol{e}_{(\mu)}\right\}$ associated with coordinates x^μ. The photons have 4-momentum $\boldsymbol{k} = E\left[\pm\boldsymbol{e}_{(\mathbf{0})} + \vec{\hat{k}}\right]$. An observer O moves at constant speed β along the x^1-axis and carries its own rest frame $\left\{\tilde{\boldsymbol{e}}_{(\mu)}\right\}$. The components of the 4-momentum of the photons transform under a Lorentz boost as[4]:

$$k^\mu = \Lambda^\mu{}_\nu \tilde{k}^\nu \ , \tag{2.186}$$

so that we get:

$$\begin{cases} E = k^0 = \gamma\left[\tilde{k}^0 - \beta\tilde{k}^1\right] = \gamma\tilde{E}\left[1 - \beta\cos\tilde{\alpha}\right] & (2.187) \\ E\cos\alpha = k^1 = \gamma\left[\tilde{k}^1 - \beta\tilde{k}^0\right] = \gamma\tilde{E}\left[\cos\tilde{\alpha} - \beta\right] & (2.188) \\ k^2 = \tilde{k}^2 & (2.189) \\ k^3 = \tilde{k}^3 \ , & (2.190) \end{cases}$$

where we have introduced the angle α between the direction of propagation of the photons and the x^1 axis in the frame of the source, the energy of the photons measured by O in its rest frame, $\tilde{E}$ and

[4]Remember that components of vectors change with the inverse matrix, compared to the basis vectors.

the angle $\tilde{\alpha}$ between the direction of propagation of the photons and the x^1-axis as measured by O. Using the pulsation $\omega = E/\hbar$ instead of the energy, we get immediately:

$$\begin{cases} \tilde{\omega} = \dfrac{\omega\sqrt{1-\beta^2}}{1-\beta\cos\tilde{\alpha}} & (2.191) \\ \cos\tilde{\alpha} = \dfrac{\cos\alpha+\beta}{1+\beta\cos\alpha} \ . & (2.192) \end{cases}$$

Eq. (2.191) shows that the frequency of photons observed by O is shifted with respect to its value in the source frame. This is the *relativistic Doppler shift*. For small velocities, $\beta \ll 1$, we have:

$$\tilde{\omega} \simeq \omega\,(1+\beta\cos\tilde{\alpha}) \ . \tag{2.193}$$

Photons emitted in the same direction that the source is moving ($\tilde{\alpha} = 0$) are *blueshifted* by an amount $\Delta\tilde{\omega} = \beta\omega$, while those emitted in the opposite direction ($\tilde{\alpha} = \pi$) are *redshifted* by $\Delta\tilde{\omega} = -\beta\omega$. Note that, contrary to the usual Doppler effect, photons emitted in the direction transverse to the motion of the source ($\tilde{\alpha} = \pi/2$) are redshifted; for small velocities, this transverse shift is of higher order though as it is given by:

$$\Delta\tilde{\omega}_\perp \simeq -\frac{1}{2}\beta^2\omega \ . \tag{2.194}$$

Eq. (2.192) describes the *aberration effect*: photons emitted in the source rest frame within a cone of opening angle $\alpha < \pi/2$ are seen by the observer to form a cone with opening angle $\alpha' < \alpha$: the beam is collimated in the direction of motion of the source relative to the observer. The effect is plotted on Fig. 2.4 for a few values of the velocity β. We see that, at the limit of ultrarelativistic motion, $\beta \to 1$, the beam is completely closed.

2.5.3 Massive particles

Massive particles move along *timelike worldlines*, i.e. curves in spacetime that are everywhere timelike. Let us consider such a particle of mass m and whose worldline, parametrised by a parameter λ, has a tangent vector $\boldsymbol{U}$, with $\eta(\boldsymbol{U},\boldsymbol{U}) < 0$. In a given frame $\{\boldsymbol{e}_{(\mu)}\}$ with coordinates $\{x^\mu\} = \{t, x^1, x^2, x^3\}$ where we used $t = x^0$, the curve is given parametrically by $x^\mu(\lambda)$ and we have $U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$. At each event $x \in \mathcal{M}$ along the curve, the tangent vector $\boldsymbol{U}$ points inside the local lightcone. For timelike curves, there is a favoured choice of parameter, namely, the *proper time* τ measured by the observer attached to the massive particle, and given by:

$$\mathrm{d}\tau^2 = -\mathrm{d}s^2 = \mathrm{d}t^2 - \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{2.195}$$

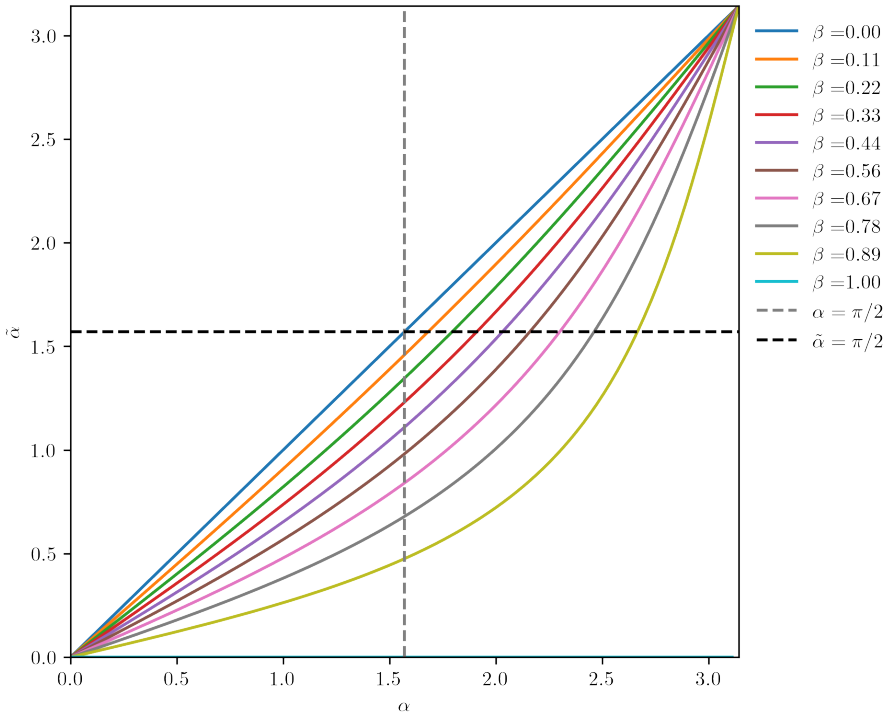


Figure 2.4: The relativistic aberration effect changes the observed emission angle α' with respect to the one in the source frame.

One can check straightforwardly that it is invariant by a change of inertial frames: it is a true scalar and only depends on the particle's motion, not on the frame in which it is evaluated. It is related to the time coordinate by:

$$\mathrm{d}\tau = \mathrm{d}t\sqrt{1 - \|\vec{v}\|^2} = \gamma^{-1}\mathrm{d}t \ , \tag{2.196}$$

where we introduced the *3-velocity* of the particle in the coordinate frame:

$$\vec{v} = \frac{\mathrm{d}x^i}{\mathrm{d}t}\boldsymbol{e}_{(i)} \ , \tag{2.197}$$

and the (time dependent) Lorentz factor of the particle with respect to the frame:

$$\gamma = \frac{1}{\sqrt{1 - \|\vec{v}\|^2}} \ . \tag{2.198}$$

The *4-velocity* the particle is then defined as:

$$\boldsymbol{u} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}\boldsymbol{e}_{(\mu)} \tag{2.199}$$

$$= \gamma \frac{\mathrm{d}x^\mu}{\mathrm{d}t}\boldsymbol{e}_{(\mu)} \tag{2.200}$$

$$= \gamma \left[\boldsymbol{e}_{(\mathbf{0})} + \vec{v}\right] \ . \tag{2.201}$$

One sees immediately that:

$$\boldsymbol{\eta}\left(\boldsymbol{u}, \boldsymbol{u}\right) = \eta_{\mu\nu} u^\mu u^\nu = u_\mu u^\mu = -1 \ . \tag{2.202}$$

Note that, according to our previous discussion, for any function f, we can always write:

$$\frac{\mathrm{d}f}{\mathrm{d}\tau} = \boldsymbol{u}(f) = u^\mu \frac{\partial f}{\partial x^\mu} \ . \tag{2.203}$$

Fig. 2.5 shows a typical timelike worldline with a spacelike dimension suppressed.

We can also define the particle's *4-momentum*:

$$\boldsymbol{p} = m\boldsymbol{u} \tag{2.204}$$

$$= m\gamma \boldsymbol{e}_{(\mathbf{0})} + \vec{p} \tag{2.205}$$

$$= E\boldsymbol{e}_{(\mathbf{0})} + \vec{p} \ , \tag{2.206}$$

with E the *energy* of the particle in the given frame, and $\vec{p} = m\gamma\vec{v}$ the *3-momentum* in the same frame. Note that:

$$\boldsymbol{\eta}\left(\boldsymbol{p}, \boldsymbol{p}\right) = -m^2 \ , \tag{2.207}$$

so that we get:

$$E = \sqrt{m^2 + \|\vec{p}\|^2} \ . \tag{2.208}$$

If the particle is not submitted to a net force, according to the principle of relativity, it should move along a straight line:

$$\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{0} \ . \tag{2.209}$$

Clearly, this is satisfied if:

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\tau} = 0 \ , \tag{2.210}$$

or equivalently:

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\tau} = 0 \ . \tag{2.211}$$

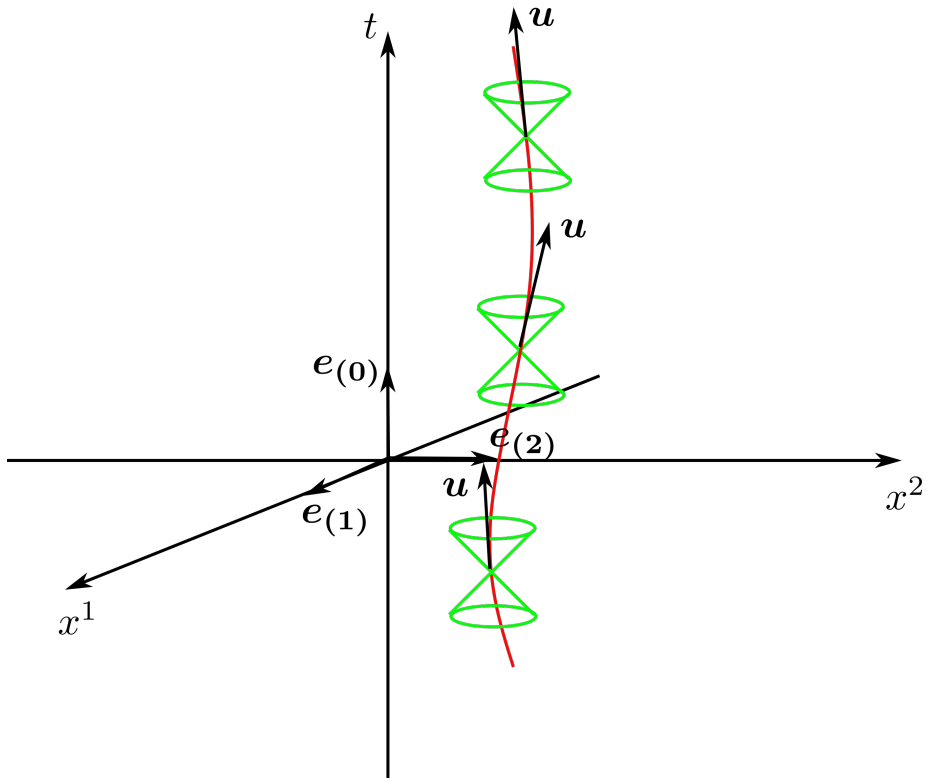


Figure 2.5: A timelike worldline (red) with 4-velocity $\boldsymbol{u}$. Some local lightcones along the worldline are represented in green.

Finally, we can define the *4-acceleration*:

$$\boldsymbol{A} = \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\tau} \,, \tag{2.212}$$

which, in components, becomes:

$$\boldsymbol{A} = \gamma \frac{\mathrm{d}\gamma}{\mathrm{d}t} \left(\boldsymbol{e}_{(\mathbf{0})} + \vec{v}\right) + \gamma^2 \vec{\mathfrak{a}} \tag{2.213}$$

$$= \frac{\mathrm{d}\gamma}{\mathrm{d}t} \boldsymbol{u} + \gamma^2 \vec{\mathfrak{a}} \,, \tag{2.214}$$

where we have defined the *3-acceleration*:

$$\vec{\mathfrak{a}} = \frac{\mathrm{d}v^i}{\mathrm{d}t} \boldsymbol{e}_{(i)} = \frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} \boldsymbol{e}_{(i)} \,. \tag{2.215}$$

Let us note that Eq. (2.202) implies that:

$$\boldsymbol{\eta}\left(\boldsymbol{A}, \boldsymbol{u}\right) = 0 \,, \tag{2.216}$$

which is equivalent to:

$$\frac{\mathrm{d}\gamma}{\mathrm{d}t} = \gamma^3 \vec{\mathfrak{a}} \cdot \vec{v} \,, \tag{2.217}$$

and implies that the 4-acceleration is always a spacelike vector.

What would be the equivalent to Newton's second law? We need a law that transforms correctly when going from one inertial frame to another, which means that it must only involve 4-vectors and proper scalars (invariant under a change of of inertial frame). We can write:

$$\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}\tau} = \boldsymbol{f} \,, \tag{2.218}$$

where $\boldsymbol{f}$ is the *4-force*. Note that Eq. (2.218) sums up 4 scalar equations, but they are not all independent because of Eq.(2.216) which adds a scalar constraint on the components of the 4-force:

$$\boldsymbol{\eta}\left(\boldsymbol{f}, \boldsymbol{u}\right) = 0 \,. \tag{2.219}$$

In order to get Newton's law:

$$\frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \vec{F} \,, \tag{2.220}$$

where $\vec{F} = F^i \boldsymbol{e}_{(i)}$ is the usual force, we need:

$$\boldsymbol{f} = f^0 \boldsymbol{e}_{(0)} + \gamma \vec{F} \,. \tag{2.221}$$

Then, using Eq. (2.219), we get:

$$f^0 = \gamma \vec{F} \cdot \vec{v} \,. \tag{2.222}$$

Thus:

$$\boldsymbol{f} = \gamma \left(\vec{F} \cdot \vec{v}\right) + \gamma \vec{F} \,, \tag{2.223}$$

and Eq. (2.218) is equivalent to:

$$\begin{cases} \dfrac{\mathrm{d}E}{\mathrm{d}t} = \vec{F} \cdot \vec{v} & (2.224) \\ \dfrac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \vec{F} \,. & (2.225) \end{cases}$$

The first equation is simply the equation giving the rate of variation of energy in terms of the power of the force and the second equation is Newton's law (be careful that $\vec{p}$ contains a Lorentz factor[5]).

[5]Since $\vec{p} = m\gamma\vec{v}$, one finds in many old textbooks (and a few modern ones too, unfortunately), a notion of relativistic varying mass $m\gamma$. This is misguided. The mass of a particle is a scalar invariant. Technically, together with the spin,

A variational principle for the free motion of particles

We can obtain the free motion of a massive particle from a variational principle by imposing that the worldline of a free particle travelling between two events separated by a timelike interval extremises the proper time taken by the particle to connect them.

Indeed, let us work in a given inertial frame and consider two events A and B and all possible timelike worldlines connecting them. Along each curve, we will have a proper time elapsed:

$$\tau_{AB} = \int_A^B \mathrm{d}\tau = \int_A^B \left[\mathrm{d}t^2 - \delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right]^{1/2} . \tag{2.226}$$

We cannot use τ as a parameter along the curves since this is the parameter we want to extremise on. Let us therefore introduce another parameter σ with $x^\mu(\sigma)$ the parametric equation of the worldline and such that $\sigma = 0$ corresponds to A and $\sigma = 1$ for B. Then:

$$\tau_{AB} = \int_0^1 \mathrm{d}\sigma \left[\left(\frac{\mathrm{d}t}{\mathrm{d}\sigma}\right)^2 - \delta_{ij}\frac{\mathrm{d}x^i}{\mathrm{d}\sigma}\frac{\mathrm{d}x^j}{\mathrm{d}\sigma}\right]^{1/2} . \tag{2.227}$$

Along a path with $t + \delta t$ and $x^i + \delta x^i$, we get:

$$\tau_{AB} + \delta\tau = \int_0^1 \mathrm{d}\sigma \left[\left(\frac{\mathrm{d}t + \delta t}{\mathrm{d}\sigma}\right)^2 - \delta_{ij}\frac{\mathrm{d}\left(x^i + \delta x^i\right)}{\mathrm{d}\sigma}\frac{\mathrm{d}\left(x^j + \delta x^j\right)}{\mathrm{d}\sigma}\right]^{1/2} . \tag{2.228}$$

Expanding at first order, integrating by parts and setting $\delta\tau = 0$ for every possible variation, we get, after a bit of work, that:

$$\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} = 0 , \tag{2.229}$$

which is exactly Eq. (2.210), as claimed.

This variational principle to determine the equations of force-free motion will be very important in General Relativity.

---

it labels irreducible representations of the Poincaré group. It has absolutely no reason to depend on the instantaneous velocity of the particle in a frame. In this course, the mass is m, period. the Lorentz factor enters the definition of the relativistic momentum when expressed in terms of the coordinate velocity $\vec{v}$, in the same way it enters the relativistic velocity, because we use t as a parameter along worldlines instead of τ.

2.6 Electrodynamics: classical field theory

As mentioned in Section 2.2, special relativity was built to try and reconcile Maxwell's formulation of electrodynamics with mechanics, in other words, the description of electromagnetic fields and interactions with the framework of Newtonian mechanics. How does electrodynamics look like in the framework of special relativity? This section will be a good preparation to the relativistic treatment of gravitation that will be the main focus of these notes. Indeed, as a classical relativistic, linear theory of a vector field, it is the simpler prototype of the classical, relativistic, non-linear theory of a tensor field that is general relativity.

2.6.1 Maxwell's equations

Let us start with Maxwell's equations written in some coordinate system $\{x^\mu\}$ associated to an admissible observer, in presence of some charge density $\rho\,(x^\mu)$ and some charge current $\vec{j}\,(x^\mu) = j^k \boldsymbol{e}_{(k)}$. The electric field $\vec{E} = E^i \boldsymbol{e}_{(i)}$ and the magnetic field $\vec{B} = B^i \boldsymbol{e}_{(i)}$ obey two sets of equations, the source-less constraints:

$$\begin{cases} \vec{\text{curl}}\vec{E} + \partial_t \vec{B} = \vec{0} \;\Leftrightarrow\; \epsilon^{ijk}\partial_j E_k + \partial_t B^i = 0 & (2.230) \\ \text{div}\,\vec{B} = 0 \;\Leftrightarrow\; \partial_i B^i = 0\,, & (2.231) \end{cases}$$

and the sourced equations:

$$\begin{cases} \text{div}\,\vec{E} = \epsilon_0^{-1}\rho \;\Leftrightarrow\; \partial_i E^i = \epsilon_0^1 \rho & (2.232) \\ \vec{\text{curl}}\vec{B} - \mu_0\epsilon_0\partial_t\vec{E} = \mu_0\vec{j} \;\Leftrightarrow\; \epsilon^{ijk}\partial_j B_k - \mu_0\epsilon_0\partial_t E^i = \mu_0 j^i\,. & (2.233) \end{cases}$$

The constants ϵ_0 and μ_0 are called the vacuum permittivity and the vacuum permeability respectively. They quantify how electric and magnetic fields react to the presence of charges and currents. We used the convenient notation: $\partial_\mu = \frac{\partial}{\partial x^\mu}$, and we introduced the totally antisymmetric tensor ϵ_{ijk}, which changes sign under the permutation of any two indices, with $\epsilon_{123} = 1$. A simple calculation shows that:

$$\begin{cases} \mu_0\epsilon_0\partial_t^2\vec{E} - \Delta\vec{E} = -\epsilon_0^{-1}\vec{\nabla}\rho - \mu_0\partial_t\vec{j} & (2.234) \\ \mu_0\epsilon_0\partial_t^2\vec{B} - \Delta\vec{B} = \mu_0\vec{\text{curl}}\vec{j}\,. & (2.235) \end{cases}$$

Thus $\vec{E}$ and $\vec{B}$ follow wave equations with a speed $(\mu_0\epsilon_0)^{-2}$, and we can conclude that:

$$\mu_0\epsilon_0 = c^{-2} \ . \tag{2.236}$$

From Eq. (2.231), we can directly see that the field $\vec{B}$ is divergence-less and is thus a pure curl; there exists a vector potential $\vec{A}$ such that:

$$\vec{B} = \vec{\text{curl}}\vec{A} \ \Leftrightarrow \ B^i = \epsilon^{ijk}\partial_j A_k \ . \tag{2.237}$$

Hence, we can rewrite Eq. (2.230):

$$\epsilon^{ijk}\partial_j \left[E_k + \partial_t A_k\right] = 0 \ , \tag{2.238}$$

so that the field $\vec{E} + \partial_t\vec{A}$ is curl-free and is thus a pure gradient. There exists a scalar potential ϕ such that:

$$\vec{E} + \partial_t\vec{A} = -\vec{\nabla}\phi \ \Leftrightarrow \ E_k = -\partial_k\phi - \partial_t A_k \ . \tag{2.239}$$

2.6.2 Covariant formulation of Maxwell's equations

Let us now try and make this theory manifestly compatible with special relativity? In this section, we keep $c \neq 1$ and, in order to retain all the formalism developed so far, this amounts to writing $x^0 = ct$.

First, let us introduce the 4-potential $\boldsymbol{A} = A^\mu \boldsymbol{e}_{(\mu)}$, also called the *electromagnetic field*:

$$A^\mu = \frac{\phi}{c}\delta^\mu{}_0 + A^k\delta^\mu{}_k \ , \tag{2.240}$$

as well as the *Maxwell tensor* $\boldsymbol{F} = F_{\mu\nu}\boldsymbol{\omega}^{(\mu)} \otimes \boldsymbol{\omega}^{(\nu)}$, with:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = 2\partial_{[\mu}A_{\nu]} \ . \tag{2.241}$$

Clearly, $\boldsymbol{F}$ is antisymmetric and is introduced because Eqs. (2.237)-(2.239) imply that:

$$\begin{cases} F_{0i} = -\dfrac{E_i}{c} & (2.242) \\ F_{12} = B_3 \ , \ F_{13} = -B_2 & (2.243) \\ F_{23} = B_1 \ . & (2.244) \end{cases}$$

Equivalently, for $F^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}F_{\rho\sigma}$, the components of the tensor $\boldsymbol{F}^* = F^{\mu\nu}\boldsymbol{e}(\boldsymbol{\mu}) \otimes \boldsymbol{e}_{(\nu)}$ dual to Maxwell's tensor $\boldsymbol{F}$, we get:

$$\begin{cases} F^{0i} = \eta^{00}\eta^{ij}F_{0j} = \dfrac{E_i}{c} & (2.245) \\ F^{12} = B_3 \ , \ F^{13} = -B_2 & (2.246) \\ F^{23} = B_1 \ . & (2.247) \end{cases}$$

If we introduce the *charge 4-current*, $\boldsymbol{j} = j^\mu \boldsymbol{e}_{(\mu)}$:

$$j^\mu = -c\rho\delta^{\mu 0} + j^i \delta_i{}^\mu \ , \tag{2.248}$$

then Maxwell's equations can be unified into two simple 4-dimensional equations:

$$\begin{cases} \partial_\mu F_{\nu\rho} + \partial_\rho F_{\mu\nu} + \partial_\nu F_{\rho\mu} = 0 & (2.249) \\ \partial_\mu F^{\mu\nu} = -\mu_0 j^\mu \ . & (2.250) \end{cases}$$

Note that, by antisymmetry of $\boldsymbol{F}$: $\partial_\mu\partial_\nu F^{\mu\nu} = 0$, so that Eq. (2.250) implies conservation of the charge 4-current:

$$\partial_\mu j^\mu = 0 \ . \tag{2.251}$$

This form of Maxwell's equations is called *covariant* because it is invariant under Lorentz transformations: all admissible observers will write these equations in the same way in their respective reference frame. In that sense, we recover that Maxwell's electrodynamics is invariant under the symmetries of special relativity. Indeed, let us consider a Lorentz transformation connecting the coordinates x^μ used above to another set of admissible coordinates $\tilde{x}^\mu$:

$$\tilde{x}^\mu = \Lambda_\nu{}^\mu x^\nu \ . \tag{2.252}$$

Since $\partial_\mu = \boldsymbol{e}_{(\mu)} = \Lambda^\nu{}_\mu \tilde{\boldsymbol{e}}_{(\nu)} = \Lambda^\nu{}_\mu \tilde{\partial}_\nu$ (with a slightly ridiculous but clear notation), we get for the dual basis:

$$\boldsymbol{\omega}^{(\nu)} = \Lambda_\mu{}^\nu \tilde{\boldsymbol{\omega}}^\mu \ . \tag{2.253}$$

Hence, some short calculations lead to the components on Maxwell's tensor in the new coordinate system:

$$\tilde{F}_{\mu\nu} = \Lambda_\mu{}^\rho \Lambda_\nu{}^\sigma F_{\rho\sigma} \ . \tag{2.254}$$

Similarly, for the dual tensor $\boldsymbol{F}^*$:

$$\tilde{F}^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu_\sigma F^{\rho\sigma} . \tag{2.255}$$

Plugging these into Eqs (2.249)-(2.250) and remembering that Lorentz transformations are invertible with constant matrix elements, we get, in the new coordinate system:

$$\begin{cases} \tilde{\partial}_\mu \tilde{F}_{\nu\rho} + \tilde{\partial}_\rho \tilde{F}_{\mu\nu} + \tilde{\partial}_\nu \tilde{F}_{\rho\mu} = 0 & (2.256) \\ \tilde{\partial}_\mu \tilde{F}^{\mu\nu} = -\mu_0 \tilde{j}^\mu \ , & (2.257) \end{cases}$$

confirming their invariance under Lorentz transformations.

Note that in terms of the vector potential, Eq. (2.249) is trivially satisfied, while Eq. (2.250) gives the wave equation:

$$\Box A^\nu - \eta^{\nu\rho}\partial_\rho \left(\partial_\mu A^\mu\right) = -\mu_0 j^\nu \ . \tag{2.258}$$

As is well known, this theory of the electromagnetic field is actually redundant as it presents some gauge freedom. Indeed, given a vector 4-potential A^μ, any other 4-vector $\tilde{A}^\mu$ related to it via:

$$\hat{A}^\mu = A^\mu + \eta^{\mu\nu}\partial_\nu \psi \ , \tag{2.259}$$

for an arbitrary function $\psi : \mathcal{M} \to \mathbb{R}$, is itself a 4-potential satisfying Maxwell's equations, since:

$$\hat{F}_{\mu\nu} = \partial_\mu \hat{A}_\nu - \partial_\nu \hat{A}_\mu = F_{\mu\nu} \ . \tag{2.260}$$

This gauge invariance means that we have some freedom in choosing the vector 4-potential. Since it can always be changed by a scalar function, we can impose any scalar constraint on it. For example, we could choose the *Lorenz gauge*:

$$\partial_\mu A^\mu = 0 \ , \tag{2.261}$$

for which the wave equation (2.258) is standard:

$$\Box A^\nu = -\mu_0 j^\nu \ . \tag{2.262}$$

Finally, let us emphasize the importance of using the 4-potential as the fundamental relativistic object to describe the electromagnetic field. Since it is a true 4-vector, under a Lorentz transformation $\boldsymbol{e}_{(\mu)} = \Lambda^\nu{}_\mu \tilde{\boldsymbol{e}}_{(\nu)}$, its components become:

$$\tilde{A}^\mu = \Lambda^\mu{}_\nu A^\nu \ , \tag{2.263}$$

so that, if an admissible observer measures $\boldsymbol{A} = 0$, so will any other. One the other hand, if any one of them measures a non-zero 4-potential, others may measure different components but they

will all agree that it is non-zero (because Lorentz transformations are invertible). As is well known, this is not the case for the electric and magnetic field. Indeed, Maxwell's dual tensor components transform as in Eq. (2.255), so that the electric field measured in the $\tilde{x}^\mu$ coordinate system will be given by:

$$\tilde{E}_i = c\tilde{F}^{0i} = c\Lambda^0{}_\rho \Lambda^i{}_\sigma F^{\rho\sigma} \ , \tag{2.264}$$

and thus for a boost, given by Eq. (2.145):

$$\begin{cases} \tilde{E}_1 = E_1 & (2.265) \\ \tilde{E}_2 = \gamma \left(E_2 - \beta c B_3\right) & (2.266) \\ \tilde{E}_3 = \gamma \left(E_3 + \beta c B_2\right) \ . & (2.267) \end{cases}$$

Similarly, the magnetic field measured in the $\tilde{x}^\mu$ coordinate system is obtained from the $\tilde{F}^{ij}$ components:

$$\begin{cases} \tilde{B}_1 = B_1 & (2.268) \\ \tilde{B}_2 = \gamma \left(B_2 + \beta\gamma \dfrac{E_3}{c}\right) & (2.269) \\ \tilde{B}_3 = \gamma \left(B_3 - \beta\gamma \dfrac{E_2}{c}\right) \ . & (2.270) \end{cases}$$

Therefore, if the first observer sees no magnetic field, the second one generally will: magnetic and electric fields are not covariant objects. This is best seen if we consider the case of a point charge q. An observer at rest with respect to this charge will measure a spherically symmetric, static electric field centred on the charge, and no magnetic field. Another observer, moving with respect to the charge at constant velocity $\vec{v}$ will however measure both an electric and a magnetic field.

As for the point particle, it is possible to arrive at an action principle for the electromagnetic field. Since the field equations are linear in $F_{\mu\ nu}$, the action ought to be quadratic in this tensor. It must also contain a term that is linear in the charge current. The only possibility is thus:

$$S = -\frac{1}{4}\int \mathrm{d}^4x F^{\mu\nu}F_{\mu\nu} + \int A_\mu j^\mu d^4x \ . \tag{2.271}$$

Varying this action with respect to A_μ and integrating by part, one recovers Eq. (2.250).

2.7 Accelerated frames

2.7.1 Local rest frame

So far, we have used inertial frames to study the motion of particles. However, as we have seen, massive particles subjected to external forces will not generally remain at constant speed in a given inertial frame:

$$\boldsymbol{A} = \frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}\tau} \neq 0 \ . \tag{2.272}$$

An observer O attached to such a worldline will not be inertial, but we might still be interested in knowing how she sees the world, i.e. how she makes measurements. This means that we would like to know how to construct its orthonormal rest frame $\left\{\hat{\boldsymbol{e}}_{(\mathbf{0})}, \hat{\boldsymbol{e}}_{(\mathbf{1})}, \hat{\boldsymbol{e}}_{(\mathbf{2})}, \hat{\boldsymbol{e}}_{(\mathbf{3})}\right\}$. For the timelike direction, we must use its 4-velocity:

$$\hat{\boldsymbol{e}}_{(\mathbf{0})} = \boldsymbol{u} \ . \tag{2.273}$$

This ensures that the time coordinate in this frame is the proper time τ as measured by the observer O along her worldline, i.e. that the observer is indeed at rest in this frame. For the spacelike part of the frame, the observer can pick any 3 orthonormal spacelike vectors $\hat{\boldsymbol{e}}_{(\boldsymbol{i})}$, as long as they satisfy:

$$\forall i \in \{1,2,3\},\ \boldsymbol{\eta}\left(\boldsymbol{u}, \hat{\boldsymbol{e}}_{(\boldsymbol{i})}\right) = 0 \ , \tag{2.274}$$

so they remain orthogonal to the timelike direction at every point along the observer's worldline.

Clearly, $\left\{\hat{\boldsymbol{e}}_{(\mathbf{0})}, \hat{\boldsymbol{e}}_{(\mathbf{1})}, \hat{\boldsymbol{e}}_{(\mathbf{2})}, \hat{\boldsymbol{e}}_{(\mathbf{3})}\right\}$ is not an inertial frame so it is not related to $\left\{\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{2})}, \boldsymbol{e}_{(\mathbf{3})}\right\}$ via a Lorentz transformation. Generically however, at fixed τ, we can write that the original inertial frame is given in terms of the non-inertial one at $x^\mu(\tau)$ via:

$$\boldsymbol{e}_{(\mu)} = \Lambda^{\nu}{}_{\mu}(\tau)\hat{\boldsymbol{e}}_{(\nu)} \ , \tag{2.275}$$

where $\Lambda^{\nu}{}_{\mu}(\tau)$ are the components of the map between frames at the specific point $x^\mu(\tau)$ along the accelerated observer's worldline. It is a Lorentz transformation at fixed τ but varies from point to point. Imposing that $\hat{\boldsymbol{e}}_{(\mathbf{0})} = \boldsymbol{u}$ implies:

$$\Lambda^{\mu}{}_{0}(\tau) = u^\mu(\tau) = \gamma(\tau)\left[\delta^\mu{}_0 + v^j\delta^\mu_j\right] \ , \tag{2.276}$$

where $\gamma(\tau) = \mathrm{d}t/\mathrm{d}\tau$ is the instantaneaous Lorentz factor of O with respect to the inertial frame. The other components of the transformation are fixed by determining how the spacelike vectors

change from one event to another along the worldline of the observer. Since the frames at τ and $\tau + \mathrm{d}\tau$ are both locally inertial, they must be related by a Lorentz transformation that maps $\hat{\boldsymbol{e}}_{(\mathbf{0})}(\tau)$ into $\hat{\boldsymbol{e}}_{(\mathbf{0})}(\tau + \delta\tau)$. Such an infinitesimal Lorentz transformation is fully charaterised by a matrix $\hat{\Lambda}^{\nu}{}_{\mu}(\tau)$ such that:

$$\hat{\Lambda}^{\rho}{}_{\mu}\hat{\Lambda}^{\sigma}{}_{\nu}\eta_{\rho\sigma} = \eta_{\mu\nu} \ . \tag{2.277}$$

By developing at first order in $\mathrm{d}\tau$:

$$\hat{\Lambda}^{\nu}{}_{\mu}(\tau) = \delta^{\nu}_{\mu} + \hat{\Omega}^{\nu}{}_{\mu}\mathrm{d}\tau \ , \tag{2.278}$$

we find that:

$$\hat{\Omega}^{\nu\mu} = -\hat{\Omega}^{\mu\nu} \ . \tag{2.279}$$

Thus, the infinitesimal Lorentz transformation is fully characterised by an antisymmetric rank two tensor:

$$\boldsymbol{\Omega} = \hat{\Omega}^{\mu\nu}\hat{\boldsymbol{e}}_{(\mu)} \otimes \hat{\boldsymbol{e}}_{(\nu)}. \tag{2.280}$$

This tensor is arbitrary since it is a prescription the observer gives herself to transport her coordinate system along her trajectory in spacetime. If we demand that spatial vectors of the frames do not rotate (mix with each other), so that the transformation only affects the plane spanned by $\boldsymbol{u}$ and $\boldsymbol{A}$:

$$\begin{cases} \dfrac{\mathrm{d}\hat{\boldsymbol{e}}_{(\mathbf{0})}}{\mathrm{d}\tau} = \boldsymbol{A} & (2.281) \\ \dfrac{\mathrm{d}\hat{\boldsymbol{e}}_{(\boldsymbol{i})}}{\mathrm{d}\tau} = \alpha_i \boldsymbol{u} \ , & (2.282) \end{cases}$$

for some constant α_i, we see that the only possibility for $\hat{\Omega}^{\mu\nu}$ is:

$$\begin{cases} \hat{\Omega}^{0i} = -\hat{\Omega}^{i0} = \alpha_i = \hat{A}^i & (2.283) \\ \hat{\Omega}^{00} = \hat{\Omega}^{ij} = 0 \ . & (2.284) \end{cases}$$

In an arbitrary coordinate system, this gives:

$$\Omega^{\mu\nu} = u^{\mu}A^{\nu} - A^{\mu}u^{\nu} \ , \tag{2.285}$$

i.e.:

$$\boldsymbol{\Omega} = \boldsymbol{u} \otimes \boldsymbol{A} - \boldsymbol{A} \otimes \boldsymbol{u} \ . \tag{2.286}$$

The basis $\{\hat{\boldsymbol{e}}_{(\mu)}\}$ thus defined is called a *Fermi-Walker* transported basis. The associated coordinates, $\{\hat{x}^\mu\}$ are called Fermi coordinates. Physically, it amounts to fixing spatial directions using gyroscopes so as to cancel the Coriolis forces in the local frame while keeping the other non-inertial forces: the spatial frames are in relative translation but not in relative rotation.

Let us now consider a particle $\mathcal{P}$, of mass m and proper time T, in motion under the influence of some forces. Let us call its trajectory $X^\mu(T)$ in the inertial frame In the inertial frame. Then, its 4-velocity has components $U^\mu = \frac{\mathrm{d}X^\mu}{\mathrm{d}T}$ in that frame and since Newton's law apply, we get:

$$m\frac{\mathrm{d}U^\mu}{\mathrm{d}T} = f^\mu \ , \tag{2.287}$$

where $\boldsymbol{f} = f^\mu \boldsymbol{e}_{(\mu)}$ is the 4-force acting on $\mathcal{P}$. If we now examine the motion of $\mathcal{P}$ as seen by the accelerated observer $\mathcal{O}$, we must write:

$$\begin{aligned} m\frac{\mathrm{d}\boldsymbol{U}}{\mathrm{d}T} =& m\frac{\mathrm{d}\hat{U}^\mu}{\mathrm{d}T}\hat{\boldsymbol{e}}_{(\mu)} + m\hat{U}^\mu\frac{\mathrm{d}\hat{\boldsymbol{e}}_\mu}{\mathrm{d}T} && (2.288)\\ =& \left[m\frac{\mathrm{d}\hat{U}^\nu}{\mathrm{d}T} + m\frac{\mathrm{d}\tau}{\mathrm{d}T}\hat{\Omega}^\nu{}_\mu\hat{U}^\mu\right]\hat{\boldsymbol{e}}_{(\nu)} && (2.289)\\ =& \hat{f}^\mu\hat{\boldsymbol{e}}_{(\nu)} \ . && (2.290) \end{aligned}$$

Therefore, the dynamics obeys:

$$m\frac{\mathrm{d}\hat{U}^\nu}{\mathrm{d}T} = \hat{f}^\nu + \hat{f}^\nu_{\text{inertial}} \ , \tag{2.291}$$

where:

$$\hat{f}^\nu_{\text{inertial}} = -m\hat{\Gamma}\hat{\Omega}^\nu{}_\mu\hat{U}^\mu \tag{2.292}$$

represents the inertial forces in the Fermi-Walker accelerated frame. Here $\Gamma = \frac{\mathrm{d}\tau}{\mathrm{d}T}$ is the instantaneous Lorentz factor between the inertial frame of $\mathcal{O}$ and the one of $\mathcal{P}$. Clearly, as expected, Newton's law is not invariant when going from an admissible observer to a non-admissible, accelerating one. The lack of invariance is manifested by the appearance of non-intertial forces, as is the case in Newtonian physics when going from Galilean to non-Galilean frames.

Note that Maxwell's equations (2.249)-(2.250) are also affected by going to an accelerated frame. For example, Eq. (2.250) becomes:

$$\tilde{\partial}_\mu\tilde{F}^{\mu\nu} + \frac{1}{u^\alpha}\frac{\mathrm{d}\Lambda_\beta{}^\lambda}{\mathrm{d}\tau}\left[\Lambda^\nu{}_\lambda\tilde{F}^{\alpha\beta} + \Lambda^\alpha{}_\lambda\tilde{F}^{\beta\nu}\right] = -\mu_0\tilde{j}^\nu \ . \tag{2.293}$$

Therefore, the laws of electrodynamics would have to be modified to remain covariant in general frames.

2.7.2 Example: Rindler observers

To illustrate this, let us imagine that an observer O travels in an inertial frame $\left\{\boldsymbol{e}_{(\mathbf{0})}\right\}$ with coordinates (t, x, y, z), along the x-axis, and with the norm of the 4-acceleration constant equal to g. Thus, we have:

$$\boldsymbol{\eta}\left(\boldsymbol{A}, \boldsymbol{A}\right) = g^2 \,, \tag{2.294}$$

and using the normalisation of the 4-velocity and Eq. (2.216), we get:

$$\begin{cases} -\left(u^0\right)^2 + \left(u^1\right)^2 = -1 & (2.295) \\ -u^0 A^0 + u^1 A^1 = 0 & (2.296) \\ -\left(A^0\right)^2 + \left(A^1\right)^2 = g^2 \,. & (2.297) \end{cases}$$

This implies that:

$$\begin{cases} A^0 = \pm\, g u^1 & (2.298) \\ A^1 = \pm\, g u^0 \,. & (2.299) \end{cases}$$

This leads to:

$$\frac{\mathrm{d}^2 u^1}{\mathrm{d}\tau^2} = g^2 u^1 \,, \tag{2.300}$$

Solving this equation and plugging the solution back in $A^0 = \pm g u^1$, we get the solution:

$$\begin{cases} u^0(\tau) = \cosh(g\tau) & (2.301) \\ u^1(\tau) = \sinh(g\tau) \,. & (2.302) \end{cases}$$

Choosing $t(0) = 0$ and $x(0) = x_0$, we get:

$$\begin{cases} t(\tau) = \dfrac{1}{g}\sinh(g\tau) & (2.303) \\ x(\tau) = x_0 + \dfrac{1}{g}\cosh(g\tau) \,. & (2.304) \end{cases}$$

The worldline is a branch of the hyperbola given by:

$$(x - x_0)^2 - t^2 = \frac{1}{g^2} \,. \tag{2.305}$$

For small velocities, $u^1 \simeq g\tau \ll 1$ and we recover the Newtonian parabola. The trajectory is represented on a spacetime diagram in the inertial frame on Fig. 2.6. Notice that observer's velocity asymptotically tends to $c = 1$ both in the past and future. This is a particle coming from infinity at a speed close to the speed of light, approaching the origin before turning back and going to infinity at a speed approaching the speed of light.

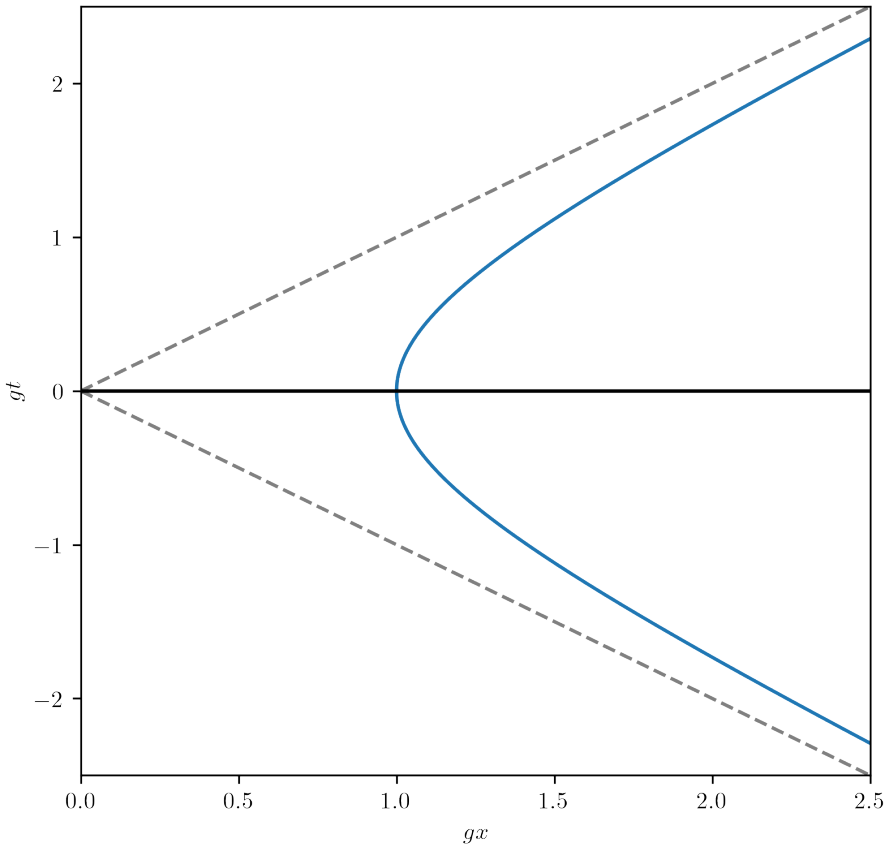


Figure 2.6: Spacetime diagram in the inertial frame of the motion of an observer with constant 4-acceleration. The dashed grey lines are the lightcone of the observer at the origin of the inertial frame.

Now, to build the orthonormal rest frame of the observer, we set:

$$\hat{\boldsymbol{e}}_{\mathbf{(0)}} = \boldsymbol{u} = \cosh(g\tau)\boldsymbol{e}_{(0)} + \sinh(g\tau)\boldsymbol{e}_{\mathbf{(1)}} \ . \tag{2.306}$$

Then, we pick $\hat{\boldsymbol{e}}_{\mathbf{(2)}} = \boldsymbol{e}_{\mathbf{(2)}}$ and $\hat{\boldsymbol{e}}_{\mathbf{(3)}} = \boldsymbol{e}_{\mathbf{(3)}}$. This choice is completely arbitrary; any other pairs

of orthonormal spacelike vectors, even if they rotate with respect to the inertial frame as τ changes would have been acceptable. To fix $\hat{\boldsymbol{e}}_{\mathbf{(1)}}$, we require that it remains orthogonal to all the other ones. Clearly:

$$\hat{\boldsymbol{e}}_{(1)} = a(\tau)\boldsymbol{e}_{\mathbf{(0)}} + b(\tau)\boldsymbol{e}_{\mathbf{(1)}} \tag{2.307}$$

will work provided:

$$\begin{cases} -a(\tau)\cosh(g\tau) + b\sinh(g\tau) = 0 & \left[\boldsymbol{\eta}\left(\hat{\boldsymbol{e}}_{\mathbf{(1)}}, \hat{\boldsymbol{e}_{\mathbf{(0)}}}\right) = 0\right] \quad (2.308) \\ -a^2(\tau) + b^2(\tau) = 1 & \left[\boldsymbol{\eta}\left(\hat{\boldsymbol{e}}_{\mathbf{(1)}}, \hat{\boldsymbol{e}_{\mathbf{(1)}}}\right) = 1\right] \ . \quad (2.309) \end{cases}$$

We can thus set:

$$a(\tau) = \sinh(g\tau) \ \text{ and } b(\tau) = \cosh(g\tau) \ , \tag{2.310}$$

and we get:

$$\hat{\boldsymbol{e}}_{(1)} = \sinh(g\tau)\boldsymbol{e}_{\mathbf{(0)}} + \cosh(g\tau)\boldsymbol{e}_{\mathbf{(1)}} \tag{2.311}$$

$\left\{\hat{\boldsymbol{e}}_{(\boldsymbol{\mu})}\right\}$ constructed this way provides a local rest frame for the observer along her worldline. A cursory look at the various expressions give the frame rotation tensor:

$$\hat{\Omega}^{0i} = \mp g\delta^{i1} \ , \tag{2.312}$$

as well as the inertial forces present in that frame:

$$\begin{cases} \hat{f}^0_{\text{inertial}} = \pm\, m\hat{\Gamma} g\hat{U}^1 & (2.313) \\ \hat{f}^1_{\text{inertial}} = \pm\, m\hat{\Gamma} g\hat{U}^0 \ . & (2.314) \end{cases}$$

For example, a particle at rest in the accelerated frame experiences the inertial force along the $\hat{\boldsymbol{e}}_{\mathbf{(1)}}$ axis only:

$$\hat{f}^1_{\text{inertial,rest}} = \pm mg \ , \tag{2.315}$$

as expected.

Let us assume that the observer starts at $t = 0$ at the surface of the Earth and travels out with the acceleration $g = 10\ \text{m}\cdot\text{s}^{-2}$. After 40 years elapsed on its ship ($\tau = 40$ years, so that $g\tau/c \simeq 42$), 3×10^9 years will have elapsed in the inertial frame ($t = 3\times10^9$ years)! This is the twin paradox, which is not a paradox at all. After that much time, it will be infinitesimally close to the speed of light (in the inertial frame): $v \simeq \tanh(42)$.

To conclude, let us imagine that a physicist is at rest in the inertial frame at the origin and emits light isotropically with pulsation ω^* in the direction $\vec{k}^*$. The wavevector of the photons reaching the observer at proper time τ is thus:

$$\boldsymbol{k} = \hbar\omega^* \left[\boldsymbol{e}_{(\mathbf{0})} + \vec{k}^*(\tau) \right] , \tag{2.316}$$

with (straight line between emitter and observer in the inertial frame):

$$\vec{k}^*(\tau) = \boldsymbol{e}_{(\mathbf{1})} . \tag{2.317}$$

In its rest frame, the observer receiving the photons would project this wavevector on its attached basis to get its *observed pulsation* $\omega(\tau)$ and *observed direction of arrival* $\vec{\hat{k}}$:

$$\begin{cases} \hbar\omega(\tau) = -\boldsymbol{\eta}\left(\boldsymbol{k}, \hat{\boldsymbol{e}}_{(\mathbf{0})}\right) & (2.318) \\ \hbar\omega(\tau)\hat{k}^i = \delta^{ij}\boldsymbol{\eta}\left(\boldsymbol{k}, \hat{\boldsymbol{e}}_{(\boldsymbol{j})}\right) . & (2.319) \end{cases}$$

Here this gives simply:

$$\begin{cases} \omega(\tau) = \omega^* \mathrm{e}^{-g\tau} & (2.320) \\ \hat{k}^i = -\delta^{i1} . & (2.321) \end{cases}$$

Photons arrive to the observer in its direction of motion, with spectral shift corresponding to a blueshift when the observer approaches the source and a redshift when it goes away; see Fig 2.7.

2.8 Gravitation: the equivalence principle

So far, Special Relativity sets a generic framework to formulate the kinematics of particles and fields like the electromagnetic field in inertial frames. We have seen how to include accelerated observers in this framework but, for those observers, like in Newtonian mechanics, the laws of physics need to be amended by the introduction of *inertial forces*. As the classical exercise about zero gravity in the space station orbiting Earth at constant height shows, as long as a gravitational field is constant, it can always be compensated by applying a constant acceleration and defining a local accelerated frame in which gravitation is absent. This is a manifestation of the equivalence principle, the first step in establishing General Relativity by noticing the tight relationship between accelerations and gravitational fields.

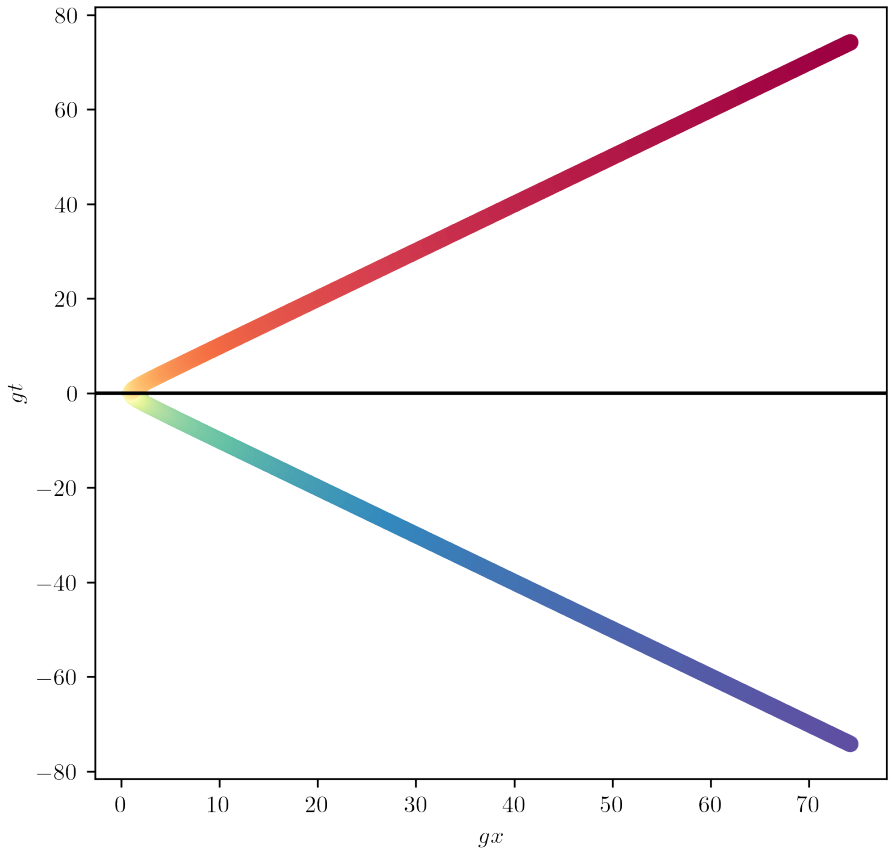


Figure 2.7: Spacetime diagram in the inertial frame of the motion of an observer with constant 4-acceleration. The color along the trajectory represent the spectral shift (in a log scale).

2.8.1 Einstein's equivalence principle

The notion of equivalence is intimately related to the universality of free fall for massive bodies. This universality was 'discovered' by Galileo and set as a fundamental principle by Newton in casting his law of gravitation. Let us consider a body of inertial mass m_i, in free fall in a gravitational field Φ, In Newtonian mechanics, in an inertial frame, its trajectory obeys:

$$m_i \vec{a} = -m_g \vec{\nabla} \Phi \,, \tag{2.322}$$

where m_g is the gravitational mass of the body. While the inertial mass tells us how the object opposes the impulse received by a given arbitrary force $\vec{F}$, the gravitational mass tells us how the

object reacts to a gravitational field; it is the *gravitational charge*. There is no *a priori* reason for these two numbers to coincide. However, it is an *empirical fact* that for any object, independently of its internal constituents, its shape and any other intrinsic properties, the two masses coincide:

$$m_i \simeq m_g \ . \tag{2.323}$$

Nowadays, this is one of the best empirically tested statement about the physical world. Using torsion balances to measure the differential of accelerations between two masses of different composition, one can form the Eötvös parameter:

$$\eta = 2\frac{a_2 - a_1}{a_2 + a_1} = 2\frac{m_{2,g}/m_{2,i} - m_{1,g}/m_{1,i}}{m_{2,g}/m_{2,i} + m_{1,g}/m_{1,i}} \ . \tag{2.324}$$

If the equivalence principle is valid, we must have $\eta = 0$. The most recent precision measurement of this parameter was made by the MICROSCOPE satellite and returned [20]:

$$\frac{m_{2,g}}{m_{2,i}} - \frac{m_{1,g}}{m_{1,i}} = [-1 \ \pm 9\,(\text{stat}) \ \pm 9\,(\text{sys})] \times 10^{-15} \tag{2.325}$$

between a pair of masses in titanium and platinum. This equivalence has now been extensively tested between different types of materials, but also for big objects such as the Moon and Earth in the field of the Sun and it is now also being tested for antimatter at CERN.
Given this level of accuracy it is thus reasonable to assume, with Newton and Einstein, that, at least up to the precision currently necessary to describe gravitational phenomena:

$$m_i = m_g \ . \tag{2.326}$$

As a direct consequence of this equivalence, we obtain:

Universality of free-fall

$$\vec{a}_i = -\vec{\nabla}\Phi \tag{2.327}$$

for any object i, irrespective of its mass, internal composition etc.

In other words, everything 'falls the same way in a given gravitational field.' This leads to:

Galilean equivalence principle

It is not possible to distinguish, locally, between a uniform gravitational field $\vec{g}$ and a uniform acceleration $\vec{a} = -\vec{g}$ in absence of gravitation.

The situation is best illustrated in the thought experiment of the lift; see Fig. 2.8. We are in an

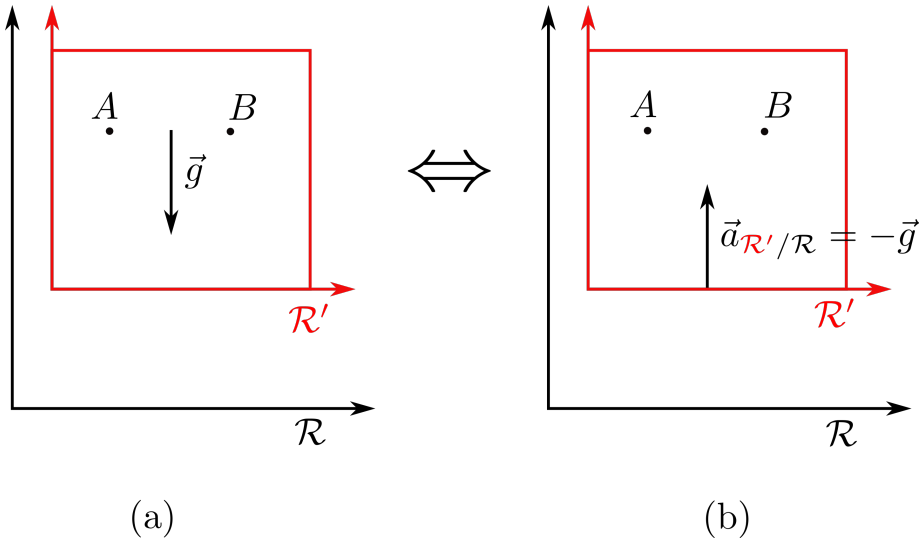


Figure 2.8: (a) The lift, materialised by the red box and to which we attach a reference frame $\mathcal{R}'$, is at rest in the inertial frame $\mathcal{R}$. In that inertial frame, a uniform gravitational field $\vec{g}$ is present and A and B are in free fall in $\mathcal{R}$ and in $\mathcal{R}'$. (b) The lift is now uniformly accelerated with acceleration $\vec{a} = -\vec{g}$ with respect to the inertial frame $\mathcal{R}$ in which gravitation has been turned off. A and B are now free in $\mathcal{R}'$ and therefore uniformly accelerated in the frame of the lift $\mathcal{R}'$. An observer in the lift cannot distinguish between the two situations.

inertial frame $\mathcal{R}$, in which we consider two massive objects A and B contained in a lift. An observer is attached to the lift in the form of a reference frame $\mathcal{R}'$. In the first instance, (a) on Fig. 2.8, a uniform gravitational field $\vec{g} = -\vec{\nabla}\Phi$ is present in $\mathcal{R}$ and $\mathcal{R}'$ is at rest with respect to $\mathcal{R}$ and is thus itself inertial. Then, Newton's law applies and:

$$\vec{a}_{A/\mathcal{R}'} = \vec{a}_{B/\mathcal{R}'} = \vec{g} \,, \tag{2.328}$$

so that both objects are in free fall in the lift.

In the other set-up, (b) on Fig. 2.8, there is no gravitational field or any other force acting on A

and B. They are at rest (or in uniform translation but it is irrelevant here) in $\mathcal{R}$. But now, the lift accelerates in $\mathcal{R}$, with an acceleration exactly equal to $-\vec{g}$, where $\vec{g}$ is the gravitational field present in setting (a). Then, once again:

$$\vec{a}_{A/\mathcal{R}'} = \vec{a}_{B/\mathcal{R}'} = -\vec{a}_{\mathcal{R}'/\mathcal{R}} = \vec{g} \ . \tag{2.329}$$

Therefore, from the point of view of the observer in the lift, there is absolutely nothing distinguishing the two situations, from a dynamical point of view. By conducting experiments with objects free of forces, she cannot say if she is at rest in a uniform gravitational field or accelerated in an environment free of gravitation. Note that this could be formulated in a slightly different way by saying that one can always cancel a gravitational field $\vec{g}$ applied in an inertial frame locally by choosing the local non-inertial frame with acceleration $\vec{g}$ with respect to the inertial frame. Then, the inertial forces generated by this acceleration exactly cancel the gravitational field, and in the accelerated frame, free falling particles actually appear free of force. This is zero-gravity. Why did we insist on the *local* restriction to the equivalence principle? Because it is absolutely crucial that the gravitational field be *uniform*. Indeed, if the field varies appreciably over the scales involved, e.g. the distance between A and B in the lift, then we cannot define a unique accelerated frame in which both masses accelerate the same way; see Fig. 2.9. In other words, acceleration can cancel the value of the gravitational field at a point but it cannot get rid of *tidal effects* at that point. This will have a striking implication in General Relativity. Einstein's decisive step was to generalise the equivalence principle to any experiment an observer could carry, not just free-fall ones:

Einstein's equivalence principle

No physics experiment can distinguish, locally, between a uniform gravitational field acting in the local inertial frame and a uniform acceleration with respect to that local inertial frame. The *local inertial frame* is the frame in which the laws of Special Relativity must hold.

In other words, the equivalence principle as understood by Einstein tells us that *locally*, spacetime ought to be Minkowski spacetime and the laws of physics ought to be written as the ones of Special Relativity. However, as the example in Fig. 2.9 illustrates, because of tidal effects, these local inertial frames will in general vary from point to point and one will have to find ways to "glue" them together if one wants a more global picture of spacetime. This is why the concept of *manifold* introduced in chapter 3 will become crucial in General Relativity.

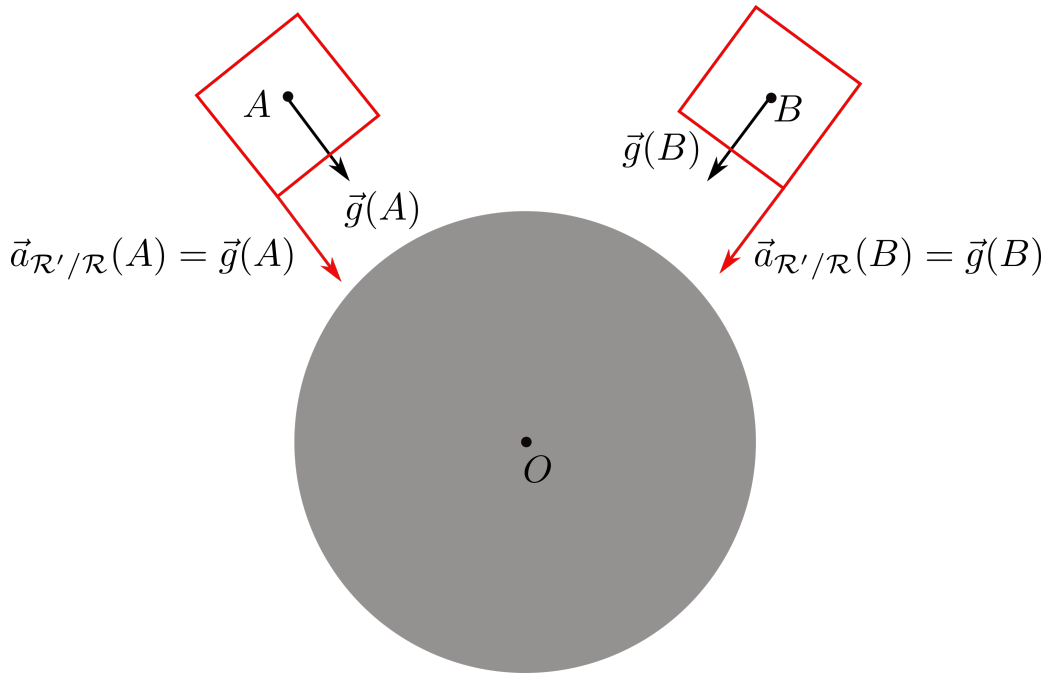


Figure 2.9: The gravitational field generated by a spherical distribution of mass is inhomogeneous in space. If two points A and B are separated by a distance large compared to the typical distance of variation of the field, then one cannot cancel the effects of the field by choosing a single appropriate accelerated frame. Rather, one needs to define accelerated frames around both points.

2.8.2 Gravitational redshift

We can already obtain a genuine physical effect from Einstein's equivalence principle, even in absence of a full theory: the gravitational spectral shift.

Consider two observers P and Q sitting respectively at the top and bottom of a tall building at the surface of the Earth, at rest in the terrestrial reference frame $\mathcal{R}_T$ assumed inertial. We will assume that the height of the building, h, is much smaller than the Earth radius, R_T, so that the gravitational field of the Earth can be considered uniform on the scales of the problem. Each observer has an atomic clock allowing them to measure their own proper time; see panel (a) on Fig. 2.10. At regular intervals of their proper time $\delta\tau_P$, P sends light pulses towards Q. Can we determine the interval of their own proper time $\delta\tau_Q$ at which observer Q receives the signals? According to Einstein's equivalence principle, this situation is equivalent to the one represented on panel (b) of Fig. 2.10:

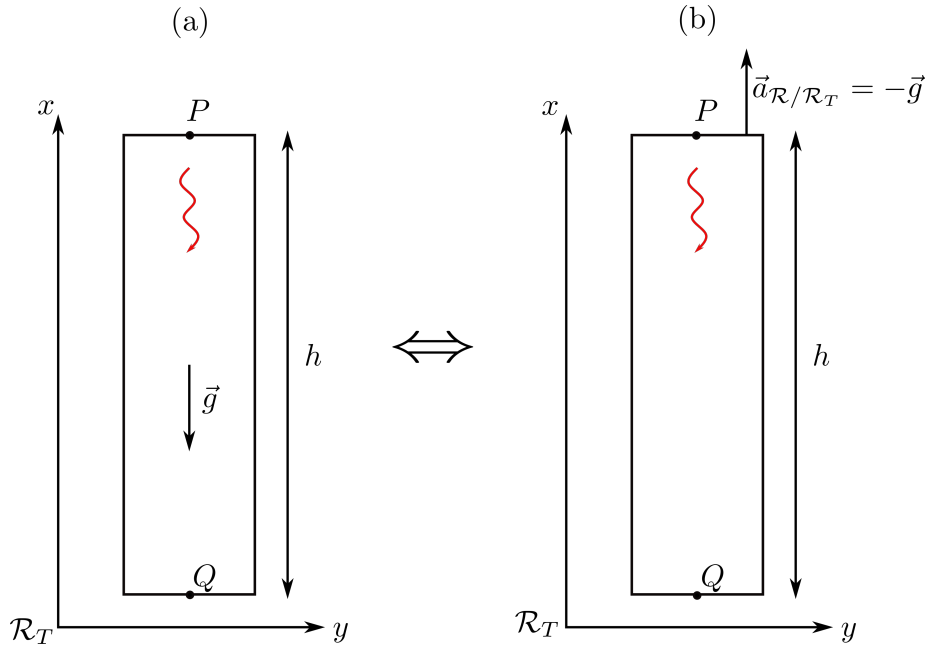


Figure 2.10: (a) P sends light signals towards Q, in a tower at rest in the terrestrial rest frame $\mathcal{R}_T$. The Earth's gravitational field $\vec{g}$ is uniform on the scales involved. (b) P and Q are embarked in a rocket moving with a uniform acceleration $-\vec{g}$ with respect to $\mathcal{R}_T$ in which gravitation has been turned off. According to the equivalence principle, both situations are equivalent.

the gravitational field is turned off and P and Q are at rest in a rocket moving with respect to $\mathcal{R}_T$ with a constant acceleration $-\vec{g}$. This means that the reference frame $\mathcal{R}$ of the rocket is not inertial. Note that Einstein's version of the equivalence principle is required here to asssume that *everything* can be translated in this equivalence, including the trajectory of light signals which do not follows from Newton's law. In any case, this means that we should be able to do calculations in configuration (b) and translate the results to configuration (a). In order to simplify calculations, we will assume that the speed of the rocket is much smaller than the speed of light, $v \ll 1$, and that the gravitational field/acceleration is also weak, $gh/c^2 = gh \ll 1$. This allows us to work at first order in these quantities and neglect Lorentz factors. In particular, we can write that proper times along the worldlines of P and Q are simply given by the time coordinate in $\mathcal{R}_T$. In $\mathcal{R}_T$, the trajectories of

P and Q are given by:

$$\begin{cases} x_P(t) = h + \frac{1}{2}gt^2 & (2.330) \\ x_Q(t) = \frac{1}{2}gt^2 \ . & (2.331) \end{cases}$$

They are depicted on Fig. 2.11. Along light rays, we can write $\Delta t = \pm \Delta x$, so that along the ray

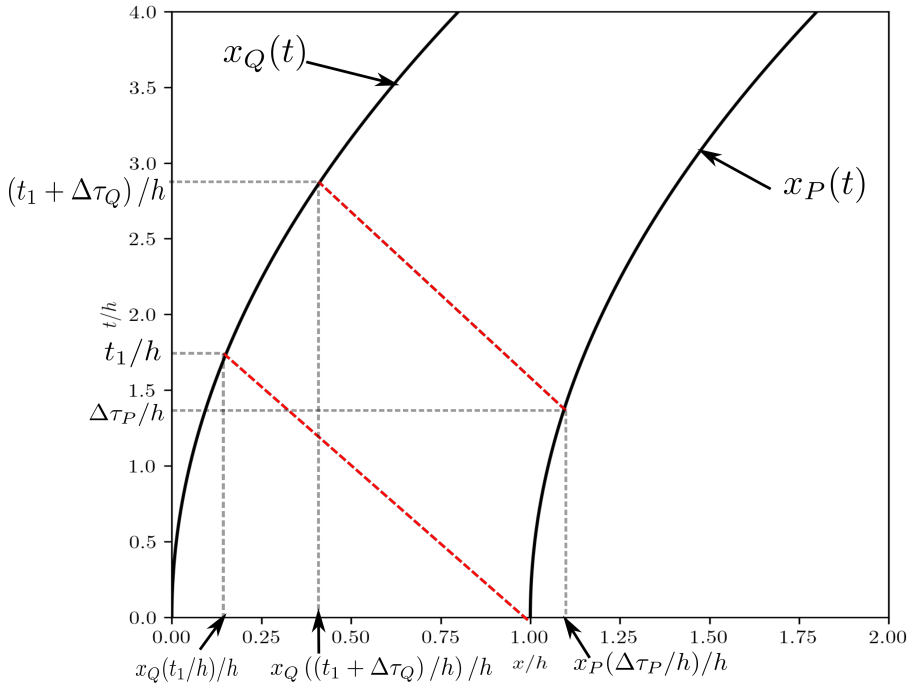


Figure 2.11: Trajectories of P and Q in $\mathcal{R}_T$ for $gh = 0.1$ for illustration purposes. The red dashed lines are the light signals emitted at $t = 0$ and $t = \delta\tau_P$ by P.

emitted at $t = 0$ and the one emitted at $t = \tau_P$ we have:

$$\begin{cases} x_P(0) - x_Q(t_1) = t_1 & (2.332) \\ x_P(\Delta\tau_P) - x_Q(t_1 + \delta\tau_Q) = t_1 + \Delta\tau_Q - \Delta\tau_P \ . & (2.333) \end{cases}$$

At first order in $\Delta\tau_P$ and $\Delta\tau_Q$, this gives:

$$\begin{cases} \frac{1}{2}gt_1^2 + t_1 - h = 0 & (2.334) \\ h - \frac{1}{2}gt_1^2 - gt_1\Delta\tau_Q \simeq t_1 + \Delta\tau_Q - \Delta\tau_P \ . & (2.335) \end{cases}$$

Solving for t_1 and keeping only leading terms in gh, we get:

$$t_1 \simeq h \ , \tag{2.336}$$

as expected, which then gives:

$$\Delta\tau_Q = \underbrace{(1 - gh)}_{<1} \Delta\tau_P \ . \tag{2.337}$$

Therefore, $\Delta\tau_Q < \Delta\tau_P$ and light signals are received at shorter intervals at Q than they are emitted at P.

According to the equivalence principle, this result must apply to the pair of observers at rest in the gravitational field of the earth. Noting that, in that case, the field Φ is given by:

$$\vec{\nabla}\Phi = \vec{g} = -g\vec{e}_x \ , \tag{2.338}$$

we get $\Phi\left(x_P\right) - \Phi(x_Q) = -gh$, so that choosing $\Phi\left(x_Q\right) = 0$ as the reference, we get $-gh = \Phi(h)$. Thus:

$$\Delta\tau_Q = (1 + \Phi(h))\, \Delta\tau_P \ . \tag{2.339}$$

The gravitational field affects local measurements of time: if Q knows the emission period by P, $\Delta\tau_P$, and measures their reception period, $\Delta\tau_Q$, they will conclude that they have measured fewer ticks of their own clock than their were on P's clock between the two emissions. If the two clocks had been synchronised initially, they will then conclude that time slows down at the bottom of the tower compared to its top. If $\Delta\tau_P$ and $\Delta\tau_Q$ are the periods of an electromagnetic wave as emitted by P and received by Q, then the result implies that the wave is received at Q with a frequency $\nu_Q > \nu_P$, i.e. that it is blueshifted. Of course, if the signal is sent from Q to P , by symmetry of light paths, the signal is redshifted. This was the first prediction Einstein made using his version of the equivalence principle, before having obtained the full theory of General Relativity. It was actually instrumental in his thoughts to arrive at this theory, although it was only experimentally checked many years later [18].

2.8.3 Incompatibility of gravitation and Special Relativity

However, this picture is problematic, as it leads to some inconsistency with Special Relativity. A way to see it is an argument originally due to Schild. Consider the set-up (a) in Fig. 2.10. P and Q are at rest in the inertial frame $\mathcal{R}_T$. Therefore, their worldlines are straight, vertical lines; see

Fig. 2.12 that shows a spacetime diagram drawns in $\mathcal{R}_T$. Light rays emitted by P towards Q ought to be straight lines with a slope of $-\pi/4$ (red dashed lines in Fig. 2.12). The problem is that the

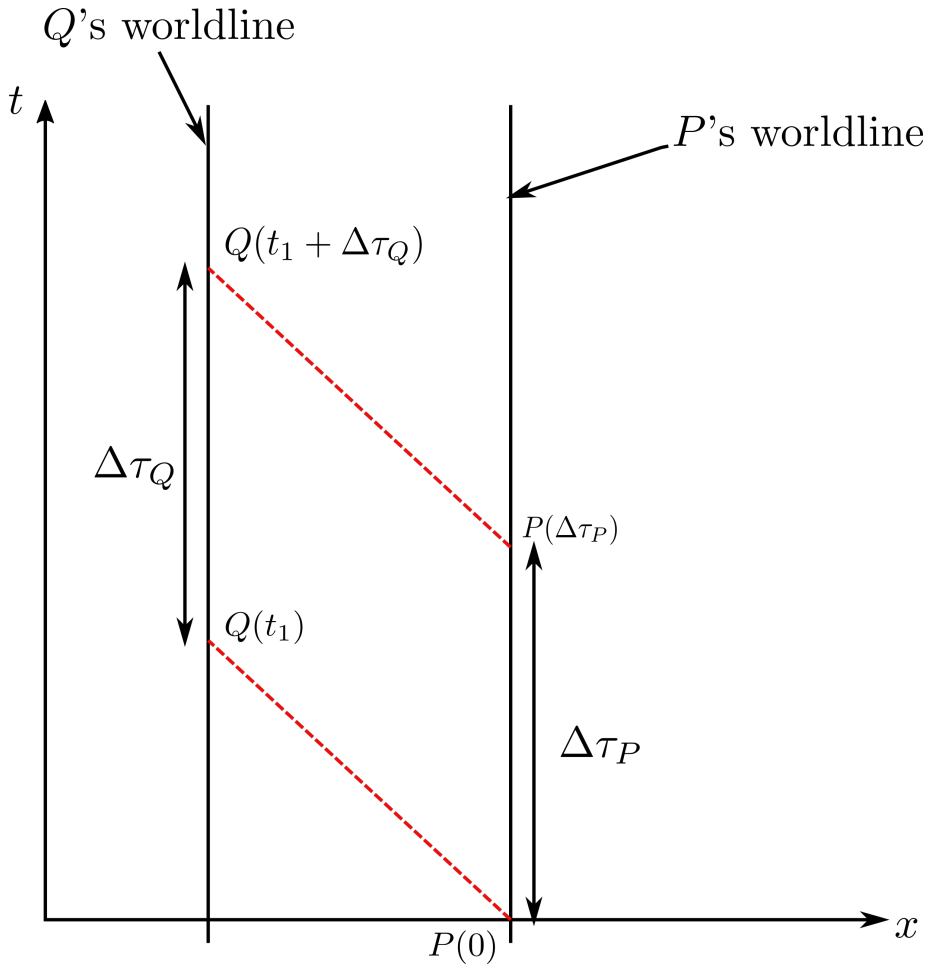


Figure 2.12: Spacetime diagram of situation (a) of Fig. 2.10, in $\mathcal{R}_T$.

parallelogram $P(0)P\left(\Delta\tau_P\right)Q\left(t_1+\Delta\tau_Q\right)Q(t_1)$ has opposite sides of different lengths: $\Delta\tau_P \neq \Delta\tau_Q$. This is completely impossible in Special Relativity and signals that *in presence of gravitational fields, it is impossible to extend an inertial frame over and arbitrarily extended regions of spacetime*.

3

The geometry of spacetime

Contents

3.1 Introduction

This chapter is the central part of these notes. It sets the general stage in which the applications of the next chapters will unfold. First, we define the concept of differentiable manifold and we formulate the concept of spacetime in this setting. We will see that the equivalence principle is naturally embedded in the manifold framework. Then, we define vectors, forms and tensors on manifolds and emphasize the link between vectors and derivatives. Via the introduction of a metric tensor, we start to do geometry.
Then, limiting our discussion to the 4-dimensional spacetime manifold of General Relativity from now on, we introduce parallel transport and geodesics and we show that geodesics are the worldlines of free-falling particles in spacetime. This leads us to consider where gravitational manifests itself and we see that it shows up in how neighbouring geodesics move relative to each other; this is the concept of geodesic deviation which will lead us to Riemann curvature.
Finally, we see how this Riemann curvature is linked to the sources of the gravitational field via the Einstein field equations.

The first section of this chapter contains some technical points that can be overlooked by the reader with a physicist's mind. Such a reader can concentrate on the big picture and assume that all the technical, especially topological properties that objects need to satisfy for definitions and results to make sense are satisfied in physics. Appendix A contains some definitions of many of these concepts for a reader more inclined to follow the mathematics. From the section on calculus, though, the concepts and methods presented are essential for the understanding of the course.

3.2 The concept of manifold

3.2.1 The basic ideas

Manifolds are an abstract generalisation to objects of arbitrary whole dimensions (this notion of dimension will be made more precise later) of intuitive notions about curves and surfaces. Generically, a curve in the three dimensional Euclidean space E^3 is a set of points that can be locally parametrised by a single real number t, via a triplet of real-valued functions $(x(t), y(t), z(t))$, and a surface can be locally parametrised by a pair of real numbers (u, v) via the parametric equations

$(x(u,v), y(u,v), z(u,v))$. This means that, by continuous deformations, a curve and a surface can respectively be locally mapped into $\mathbb{R}$ and $\mathbb{R}^2$: they are locally homeomorphic to $\mathbb{R}$ and $\mathbb{R}^2$[1]. A (real) manifold is simply a topological space that is locally homeomorphic to $\mathbb{R}^n$ for $n \in \mathbb{N}$. Here, the subtle concept is locality: the homeomorphism that allows to parametrise a curve or a surface in terms of coordinates usually depends on the point of the curve or the surface considered, it varies from point to point. Therefore, manifolds, although locally homeomorphic to $\mathbb{R}^n$, generally differ from it globally. If this is the case, one may have to introduce several local coordinate systems. Moreover, the same point of the manifold be described by different coordinate systems that are all acceptable. These coordinate systems are inessential; they are just convenient ways to describe locally the manifold itself. Nevertheless, in these lecture notes we will focus on manifolds for which the change of coordinates can be done smoothly; this will allow for the development of differential calculus on manifolds.

In order to grasp the main ideas, let us focus on one of the simplest (for our every day life intuition) manifold: the unit 2-sphere S^2 embedded in E^3. As an object embedded in $\mathbb{R}^3$, this is defined by:

$$S^2 = \left\{(x,y,z) \in \mathbb{R}^3,\, x^2 + y^2 + z^2 = 1\right\} \ . \tag{3.1}$$

Of course, one can also introduce spherical coordinates $(r, \theta, \phi) \in \mathbb{R}_+ \times [0,\pi] \times [0, 2\pi[$ in E^3. In these coordinates, the 2-sphere is simply characterised by $r = 1$; a very simple equation indeed. That means that points on the sphere can be labelled by two real numbers $(\theta, \phi) \in [0,\pi] \times [0, 2\pi[$, in agreement with the intuition that a sphere is two-dimensional. Can any point be labelled like that though? No! These coordinates cover the entire sphere except the poles ($\theta = 0$ and $\theta = \pi$), where ϕ is not even defined[2]! In essence, this is therefore a local mapping between the sphere and $\mathbb{R}^2$, because it excludes two points. Besides this problem, spherical coordinates on the sphere also suffer from another pitfall: imagine a point that describes a circle at $\theta = \text{cst}$, starting from $\phi = 0$. The value of ϕ for this point increases gradually until it approaches 2π, getting closer and closer to the position where it started. That means that in a neighbourhood of the great circle $\phi = 0$, the coordinate system is discontinuous. Since we want to develop calculus on objects like the sphere, and since calculus is essentially based on smoothness, this is a problem. To alleviate this problem,

[1] A map $f : X_1 \to X_2$ between two topological spaces X_1 and X_2 is an homeomorphism if it is continuous and its inverse $f^{-1} : X_2 \to X_1$ exists and is itself continuous. In these notes, all maps that need to be homeomorphic for definitions to work, in particular charts, will be homeomorphic.

[2] Convince yourself of that by writing the transition from Cartesian coordinates to spherical coordinates

we could decide to let ϕ run in $\mathbb{R}_+$, but then the same point on the sphere would correspond to an infinite number of pairs of coordinates, $(\theta, \phi + 2k\pi)$, for $k \in \mathbb{Z}_+$.

Of course, there exists many ways (an infinite number actually) to label points on S^2. Another example are *stereographic coordinates* $(X, Y) \in \mathbb{R}^2$, projected from the North pole N, for example:

$$X = \frac{x}{1-z} \;\; ; \;\; Y = \frac{y}{1-z} \, , \tag{3.2}$$

with $z^2 = 1 - x^2 - y^2$. These coordinates are continuous and even smooth everywhere on $S^2 \backslash \{N\}$, but they are not defined at N. So again, they can only represent the sphere locally, in the sense that they cannot be extended to every point of the sphere. Therefore, one sees that, in order to label all the points of the sphere by a set of two real numbers, one necessarily has to use several maps from S^2 onto $\mathbb{R}^2$. This is a key idea that results from the locality condition mentioned earlier.

The arbitrariness in the coordinate system that allows one to describe the same point of a manifold like S^2 by different sets of coordinates is actually one of the reasons why manifolds are so important in physics, since it is very similar to the principle of relativity of physics that states that the behaviour of a physical system is independent from the frame used to describe it.

So, what is the key idea to remember? We have hinted at the fact that it is impossible to label all the points of S^2 with a single set of two real numbers such that both of the following conditions are satisfied simultaneously:

(1) 'Close' points on S^2 have 'close' values of their coordinates. Rigorously that means that any neighbourhood of any point of S^2 is mapped continuously to a connected open set of $\mathbb{R}^2$.

(2) Each point of S^2 is represented by a unique set of coordinates.

Nevertheless, we can certainly require that these conditions (1) and (2) be satisfied *locally*, i.e. that, around every point $p \in S^2$, we define an open set U_p that is mapped continuously onto an open set of $\mathbb{R}^2$. But then, what happens if two open sets U_p and U_q for $p \in S^2$ and $q \in S^2$, $p \neq q$ intersect? Then, in order to define a proper differential structure, we will impose a condition that requires that, in the intersection $U_p \cap U_q$, the coordinate systems attached to U_p and U_q are linked by a smooth transformation, ensuring that, on that part of the sphere, the two descriptions of the sphere in terms of coordinates are compatible and can be interchanged smoothly.

3.2.2 Differential manifolds

We define differential manifolds as follows:

Differentiable manifold

M is a differentiable manifold (also hereafter manifold) of dimension $n \in \mathbb{N}^*$ iff:

(1) M is a Hausdorff topological space;

(2) there exists a family of open sets $\{U_i\}$, with $i \in I$ (set of indices) which covers M, i.e. such that $\bigcup_{i\in I} U_i = M$;

(3) for any $i \in I$, there exists an homeomorphism $\phi_i : U_i \to U_i^{'} \subseteq \mathbb{R}^n$, $U_i^{'}$ being open;

(4) given U_i and U_j such that $U_i \cap U_j \neq \emptyset$, the map $\psi_{ij} = \phi_j \circ \phi_i^{-1}$ from $\phi_i\left(U_i \cap U_j\right)$ into $\phi_j\left(U_i \cap U_j\right)$ is infinitely differentiable.

The pair (U_i, ϕ_i) is called a *chart*, and the collection $\{(U_i, \phi_i)\}$ an *atlas*. An open set U_i is called a *coordinate neighbourhood* and ϕ_i a *coordinate function*, or coordinates in short. In $\mathbb{R}^n$, ϕ_i is represented by the set of functions $\left(x^1(p), ..., x^n(p)\right) \in \mathbb{R}^n$ for any $p \in U_i$. This n-uples is also called the coordinates of p relative to ϕ_i. This is all illustrated on Fig. 3.1. Anticipating a bit on our definition of curves on manifolds, we can also understand a local chart as the "knitting" of the manifold by "threads" (curves) of constant values of the local coordinates; see Fig. 3.2. This image is at the origin of the name "curvilinear coordinates" that we find in old textbooks.

Let us emphasise that a point $p \in M$ exists independently of any coordinate system used to described the manifold locally around p. The choice of coordinates is arbitrary and is usually determined by convenience in a specific development. Note that, out of convenience, we will sometimes use the incorrect notation $x = \left(x^1(p), ..., x^n(p)\right) \in \mathbb{R}^n$ to denote the point $p \in M$ when a coordinate system has been fixed and there is no ambiguity. Conditions (2) and (3) in the definition of a manifold ensure that M is locally Euclidean, i.e. in any local chart, M looks like an open set of $\mathbb{R}^n$. But, we do not require this to be true globally.

If two neighbourhoods U_i and U_j have a non-empty intersection, then each point of $U_i \cap U_j$ is assigned two coordinate systems, ϕ_i and ϕ_j. The fourth condition in the definition then ensures that the transition between one coordinate system and the other be smooth (infinitely differentiable). If $\phi_i(p) = \left(x^1(p), ..., x^n(p)\right) = x$ and $\phi_j(p) = \left(y^1(p), ..., y^n(p)\right) = y$, the transition function $\psi_{ij}(x) = y$ and we have explicitly that, for any $j \in \{1, 2, .., n\}$, $y^j(x)$ is a smooth function of

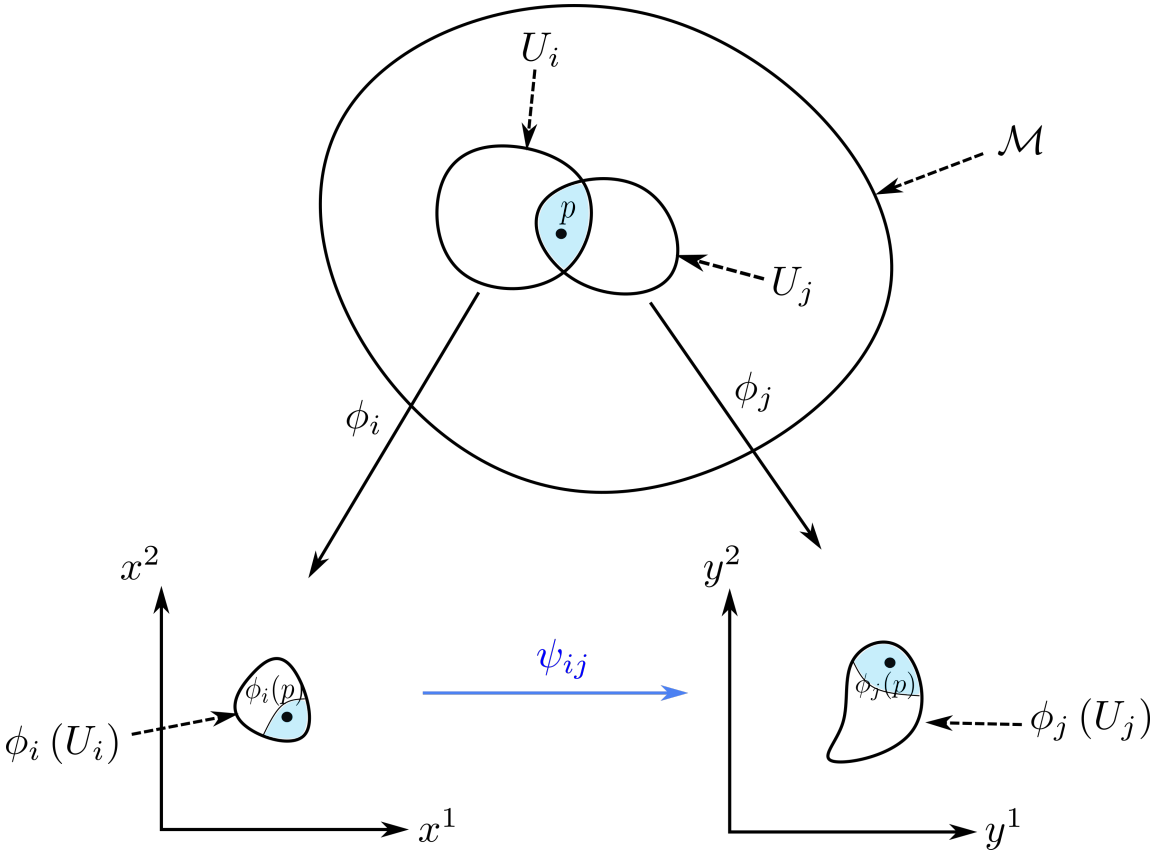


Figure 3.1: Schematic representation of local charts on a manifold.

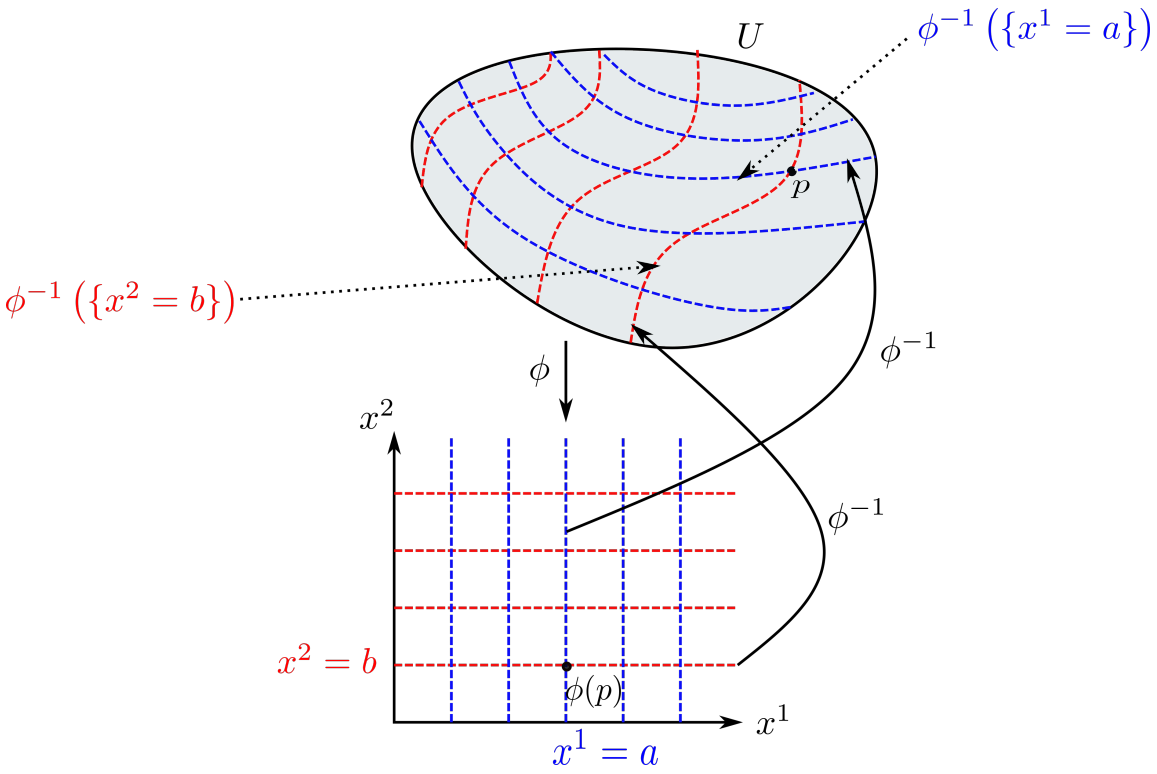


Figure 3.2: A local chart and the "knitting" on the manifold by "threads" at constant values of the coordinates. The piece of surface U is given local coordinates $\{x^1, x^2\}$ by the chart (U, ϕ). Each point p on the surface is at the intersection of two curves that are the pre-image of the line at constant x^1 (blue dashed) and constant x^2 (red dashed).

$\left(x^1(p), ..., x^n(p)\right)$, that is, each y^j is infinitely differentiable with respect to any x^i[3]. Of course, because both ϕ_i and ϕ_j are homeomorphism, ψ_{ij} is also an homeomorphism, and therefore ψ_{ij}^{-1} is continuous; this ensures that the change from x to y and the change from y to x are equally possible and infinitely differentiable.

Given two atlases $\{(U_i, \phi_i)\}$ and $\left\{(V_j, \psi_j)\right\}$ of the same manifold M, if their union is still an atlas of M, we say that the two atlases are *compatible*. Compatibility in that sense is an equivalence relation and its the equivalence classes are called *differentiable structures* on M. Whether or not a given manifold M has more than one differentiable structure is a very difficult question. For example, S^7 has 28 inequivalent differential structures! Even more strikingly, a space like $\mathbb{R}^4$ turns out to have a infinite number of differential structures!

We can now illustrate this definition by a few examples.

The vector space $\mathbb{R}^n$ for $n \in \mathbb{N}^*$ is the most trivial example of a manifold. In that case, it is enough to use a single chart covering the whole manifold, whose coordinate function is the identity. Of course, one can rely on more complicated charts, according to the principle that a manifold is independent on the chart used to describe it.

In one dimension, there are only two connected differential manifolds: a line (formally identical to $\mathbb{R}$; see diffeomorphism later) and the circle S^1. $\mathbb{R}$ is a trivial sub-case of the previous example, and we can concentrate on S^1. If we want to satisfy our axioms, we need several charts, otherwise we would have discontinuous maps at one point at least (remember our discussion on S^2), which would not be homeomorphisms. S^1 is defined via: $S^1 = \left\{(x, y) \in \mathbb{R}^2,\ x^2 + y^2 = 1\right\}$. Let:

$$\phi_1 : \begin{cases} U_1 = \left\{(x, \sqrt{1-x^2}), x \in]-1, 1[\right\} & \rightarrow &]-1, 1[\\ p = (x, \sqrt{1-x^2}) & \mapsto & x \end{cases}, \tag{3.3}$$

$$\phi_2 : \begin{cases} U_2 = \left\{(x, -\sqrt{1-x^2}), x \in]-1, 1[\right\} & \rightarrow &]-1, 1[\\ p = (x, -\sqrt{1-x^2}) & \mapsto & x \end{cases}, \tag{3.4}$$

$$\phi_3 : \begin{cases} U_3 = \left\{(\sqrt{1-y^2}, y), y \in]-1, 1[\right\} & \rightarrow &]-1, 1[\\ p = (\sqrt{1-y^2}, y) & \mapsto & y \end{cases}, \tag{3.5}$$

[3]The differentiability here is with respect to the usual partial differentiation of calculus on $\mathbb{R}^n$

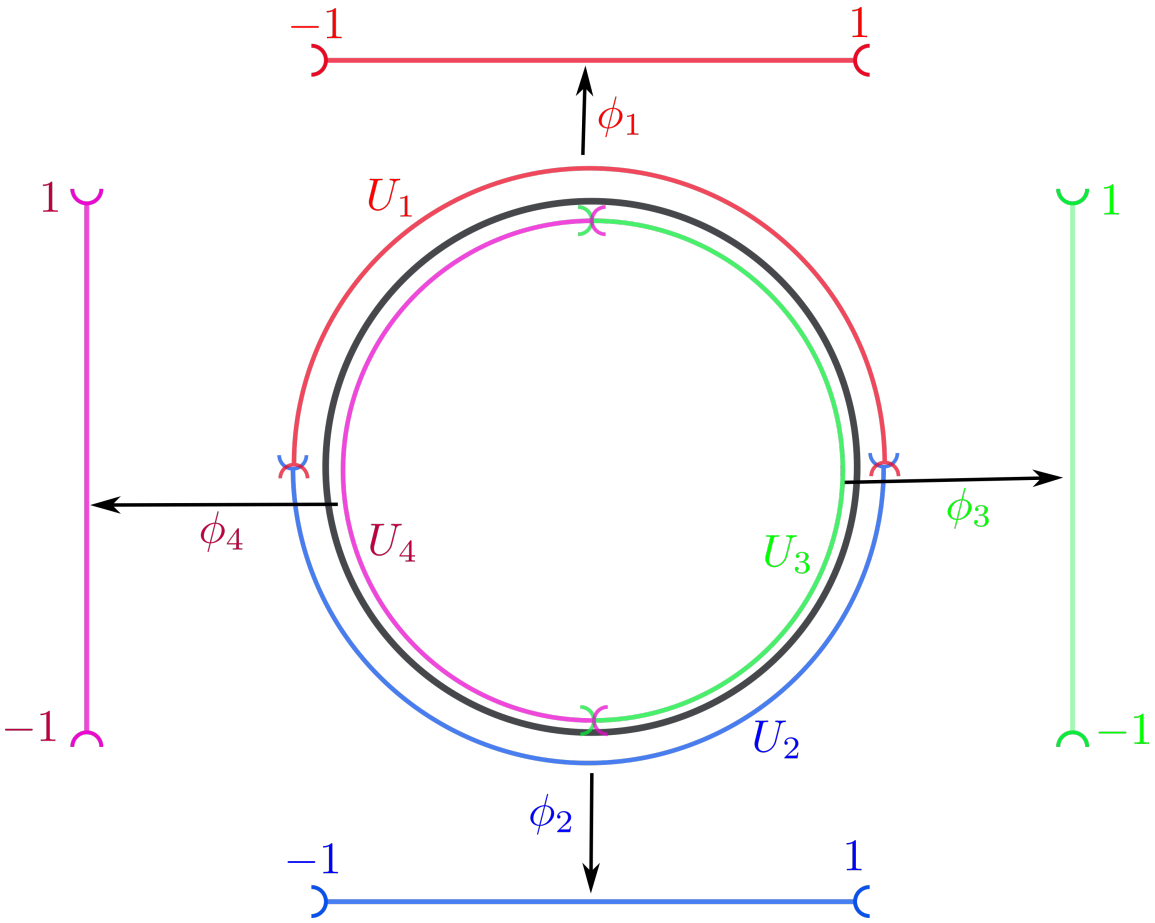


Figure 3.3: Schematic representation of the atlas (3.3)-(3.6).

and:

$$\phi_4 : \left\{ \begin{array}{ccc} U_4 = \left\{(-\sqrt{1-y^2}, y), y \in]-1,1[\right\} & \rightarrow &]-1,1[\\ p = (-\sqrt{1-y^2}, y) & \mapsto & y \end{array} \right. . \tag{3.6}$$

The atlas is depicted on Fig. 3.3.

Then, ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are continuous and invertible, and their inverses are also continuous. Moreover, all the transition functions are infinitely differentiable homeomorphisms. For example: U_1 and U_3 intersect and $\psi_{13} = \phi_3 \circ \phi_1^{-1} :]0,1[\rightarrow]0,1[$ reads:

$$\forall x \in]0,1[, \psi_{13}(x) = \phi_3\left((x, \sqrt{1-x^2})\right) = \sqrt{1-x^2}\,, \tag{3.7}$$

which is indeed infinitely differentiable. You can check the other transition functions as an exercise.

Another atlas on the circle

Can you find another atlas for S^1?

As another example, we can consider the n-dimensional sphere:

$$S^n = \left\{ \left(x^0, x^1, ..., x^n\right) \in \mathbb{R}^{n+1}, \sum_{i=0}^{n} \left(x^i\right)^2 = 1 \right\} . \tag{3.8}$$

We define $2(n+1)$ coordinate neighbourhoods U_i as follows: for any $i \in \{0, ..., n\}$,

$$U_{i+} = \left\{ \left(x^0, x^1, ..., x^n\right) \in S^2, x^i > 0 \right\} \tag{3.9}$$

$$U_{i-} = \left\{ \left(x^0, x^1, ..., x^n\right) \in S^2, x^i < 0 \right\} . \tag{3.10}$$

The corresponding coordinate maps, $\forall i \in \{0, 1, ..., n\}$, $\phi_{i+} : U_{i+} \longrightarrow \mathbb{R}^n$ and $\phi_{i-} : U_{i-} \longrightarrow \mathbb{R}^n$ are defined via:

$$\begin{aligned} \phi_{i+}\left((x^0, ..., x^n)\right) &= (x^0, ..., x^{i-1}, x^{i+1}, ..., x^n) \\ \phi_{i-}\left((x^0, ..., x^n)\right) &= (x^0, ..., x^{i-1}, x^{i+1}, ..., x^n). \end{aligned}$$

Geometrically, $\phi_{i\pm}$ are the projections of the hemispheres $U_{i\pm}$ onto the plane $x^i = 0$. Obtaining the transition functions to verify that they are smooth is a bit cumbersome, but it can easily be done in low dimensions (ex.: $n = 2$).

Stereographic atlas on S^2

Using stereographic projections, obtain an atlas for S^2. Generalise to S^n.

As a last example, consider the torus:

The torus

The torus is the set:

$$T^2 = \left\{ (x, y, z) \in \mathbb{R}^3, \left(\sqrt{x^2 + y^2} - 1 \right)^2 + z^2 = \frac{1}{4} \right\} . \tag{3.11}$$

Show that one can define $(\theta, \varphi) \in [0, 2\pi[^2$ such that:

$$(x, y, z) \in T^2 \Leftrightarrow x = (1 + \frac{1}{2}\cos\varphi)\cos\theta \, , \, y = x = (1 + \frac{1}{2}\cos\varphi)\sin\theta \, , \, z = \frac{1}{2}\sin\varphi. \tag{3.12}$$

Represent T^2. Often T^2 is presented as the Cartesian product $S^1 \times S^1$, why?
Why can't the induced map $T^2 \to (\theta, \varphi)$ be an atlas? Propose an atlas.

3.2.3 The spacetime manifold of General Relativity

These general considerations on manifolds are quire useful, especially because in General Relativity we often have to deal with parts of spacetime that are themselves manifolds, or with symmetry Lie groups, so knowing about manifolds in general is quite useful. But the central feature of the theory is that spacetime itself needs to be described as a generic 4 dimensional manifold equipped with a metric tensor. As we saw in the previous chapter, the equivalence principle teaches us that gravitation is geometry, and that it is a naturally *local* concept: we can always cancel a gravitational field locally by choosing an appropriately accelerated reference frame, i.e. an appropriate coordinate system in which physics is described by Special Relativity. In particular, this means that locally, spacetime is homeomorphic to Minkowski spacetime: $\mathbb{R}^4$ equipped with the metric $\boldsymbol{\eta}$; we will see that it is not sufficient and that an extra condition needs to be imposed on the metric derivatives. But, if we forget for a moment about the metric, this is exactly what a 4 dimensional manifold is: a set of points locally homeomorphic to $\mathbb{R}^4$. So here it is.

Spacetime of General Relativity

The spacetime of General Relativity is a 4 dimensional differentiable manifold $\mathcal{M}$. A point of $\mathcal{M}$ is called an *event*.

What of the metric? It turns out to be the important part since it encodes the gravitational field. But before we can make sense of it, we need to define vectors. Everything that will be defined

afterwards will make sense for general manifolds. However, from now on, unless otherwise stated, *we will restrict our presentation to the 4 dimensional manifold of General Relativity*. When the word manifold is used without any further details, it will be assumed to be differentiable and 4 dimensional.

Finally, as we have seen, a striking feature of manifolds is that locally, their properties do not depend on the coordinate systems chosen to cover them. Thus, if we formulate the laws of physics on a manifold, we should expect them to be *invariant* under any permissible coordinate change. This is known as *general covariance* and we will get back to it later.

Let us fix some notations. Given a point $p \in \mathcal{M}$ and an open set U around p, local charts (U, ϕ) will define local coordinate systems which will be denoted $\phi(p) = \left(x^0, x^1, x^2, x^3\right) = (x^\mu)$. In other words, spacetime indices will run from 0 to 3, like in the special relativistic case, and be referred to by Greek letters. When restraining the range to $\{1, 2, 3\}$, we will use Latin indices instead, usually from the second part of the alphabet.

3.3 Calculus on manifolds

In this section, we are going to introduce on manifolds the usual notions of calculus and geometry. The tactics will essentially consists in "localising" everything and, using charts to "bring down" objects in $\mathbb{R}^4$, where we can rely on our usual techniques. There is one conceptual exceptions though. As we have seen in the context of Special Relativity, it is possible to think of vectors in $\mathbb{R}^4$ as derivative operators, associated with derivatives along curves. This is certainly a bit cumbersome in $\mathbb{R}^4$. But on a generic manifold, this is the only possible way to make sense of the concept of vector! In a sense, calculus on manifolds ties together geometric concepts like vectors to analysis concepts like derivatives. As it turns out, this is a deep connection that does not stop with vectors (although we will only talk about these here).

3.3.1 Functions

Functions

A function on a manifold $\mathcal{M}$ is a smooth map $f : \mathcal{M} \to \mathbb{R}$, i.e. that to each point $p \in \mathcal{M}$, it associates a real number $f(p) \in \mathbb{R}$. In physics, it is often dubbed a *scalar field*.

It is the simplest type of map we can define on a manifold. You can think of it as, for example, the map that associates a temperature to a point at the surface of the Earth. Such an example is displayed on Fig. 3.4 in colour: each point is coloured according to its temperature and the function $T : S^2 \to \mathbb{R}$ maps points on the sphere to points on the real line, here represented by the colour bar at the bottom. If we introduce coordinate charts then such functions on manifolds can be viewed,

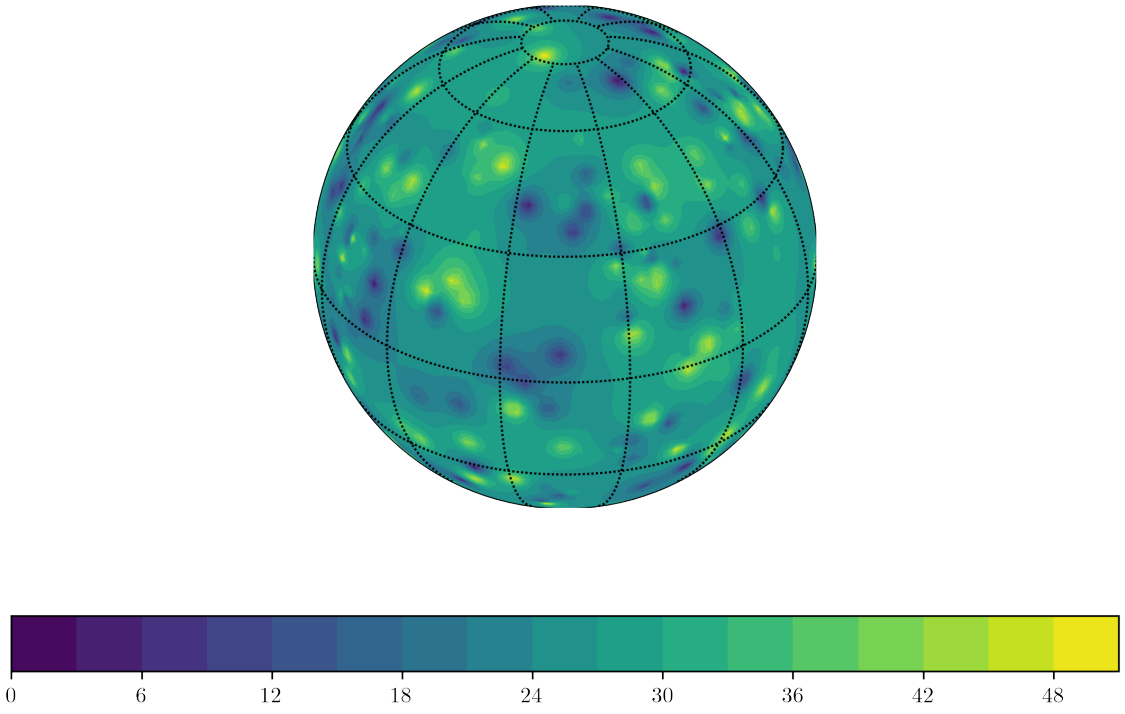


Figure 3.4: A function associating temperature to the surface of a sphere.

locally, as usual functions on $\mathbb{R}^4$. Let (U, φ) be a chart around $p \in \mathcal{M}$ and $f : \mathcal{M} \to \mathbb{R}$ a function. Then $f \circ \phi^{-1} : \mathbb{R}^4 \to \mathbb{R}$ is a usual function on $\mathbb{R}^4$. If we call x^μ the coordinates of p in the chart, then $f(p) = f\left(\phi^{-1}\left(x^\mu\right)\right) = f\left(x^0, x^1, x^2, x^3\right)$, where we used the usual physicists' abuse of notations to write the last equality. Thus $f \circ \phi^{-1}$ is just a regular function on $\mathbb{R}^4$ for which we can define partial derivatives etc. The adjective "smooth" in the definition above refers to the fact that in any local chart, $f \circ \phi^{-1}$ is smooth in the usual calculus sense (at least twice continuously differentiable for us, more usually just infinitely differentiable). This will be the case throughout from now on.
Often, when the chart is not ambiguous, physicists tend yo forget about ϕ^{-1} and simply write $f(p) =$

$f(x^\mu)$. This is confusing and horrible but permitted as long as we keep in mind what we really mean! We will denote by $\mathcal{F}(\mathcal{M})$ the set of all smooth functions on the manifold $\mathcal{M}$. One can check, as an exercise, that *functions are invariant under coordinate transformations*.

3.3.2 Curves

The objects we need to understand next are, in a way the inverse of functions: curves.

Curves on a manifold

A *parametrised curve* on a manifold $\mathcal{M}$ is a smooth map $c : I \subseteq \mathbb{R} \to \mathcal{M}$ which, to any real number $\lambda \in I$, called a *parameter*, associates a point on $\mathcal{M}$, $c(\lambda) = p$.
The image $c(I)$ is called the *unparametrised curve* or simply curve described by c.

The situation is represented on Fig. 3.5.

There is a bit of ambiguity here as we called the map c between $\mathbb{R}$ and $\mathcal{M}$ the parametrised curve while $c(I)$ is simply the curve or unparametrised curve. This is simply because one can always choose a different parameter $\sigma \in J \subseteq \mathbb{R}$ via a change of parameter, i.e. a bijective map $\psi : I \to J$ such that $\sigma = \psi(\lambda)$. Then, we obtain a new parametrised curve $c' = c \circ \psi^{-1}$ with $c'(J) = c(I)$: both maps describe the same unparametrised, geometrical curve on $\mathcal{M}$. This ability to reparametrise curves will be important later.
Given a local chart (U, ϕ) assumed to cover the whole curve[4] for each point on the curve, we have coordinates: $\phi\left[c(\lambda)\right] = x^\mu(\lambda)$ so that the map $\phi \circ c : I \to \mathbb{R}^4$ is the parametric equation of the curve.

3.3.3 Vectors

We are now ready to take the next step and define other geometrical objects such as vectors, covectors and tensors. We already know the notion of a vector in E^n as 'an arrow' pointing from one point to another, and, more generally, as the elements of a vector space. Of course, on manifolds, this

[4] If it were not the case, we could just study the part of the curve covered by the chart and move from chart to chart along the curve in the appropriate way. This would introduce some complication involving transition maps between charts but it would not affect the arguments here. As a matter of fact, this is a general remark in what follows: we will liberally use purely local arguments, assuming that the extrapolation to global ones is made of straightforward algebraic manipulations. This is not always true, but in this course, it will be.

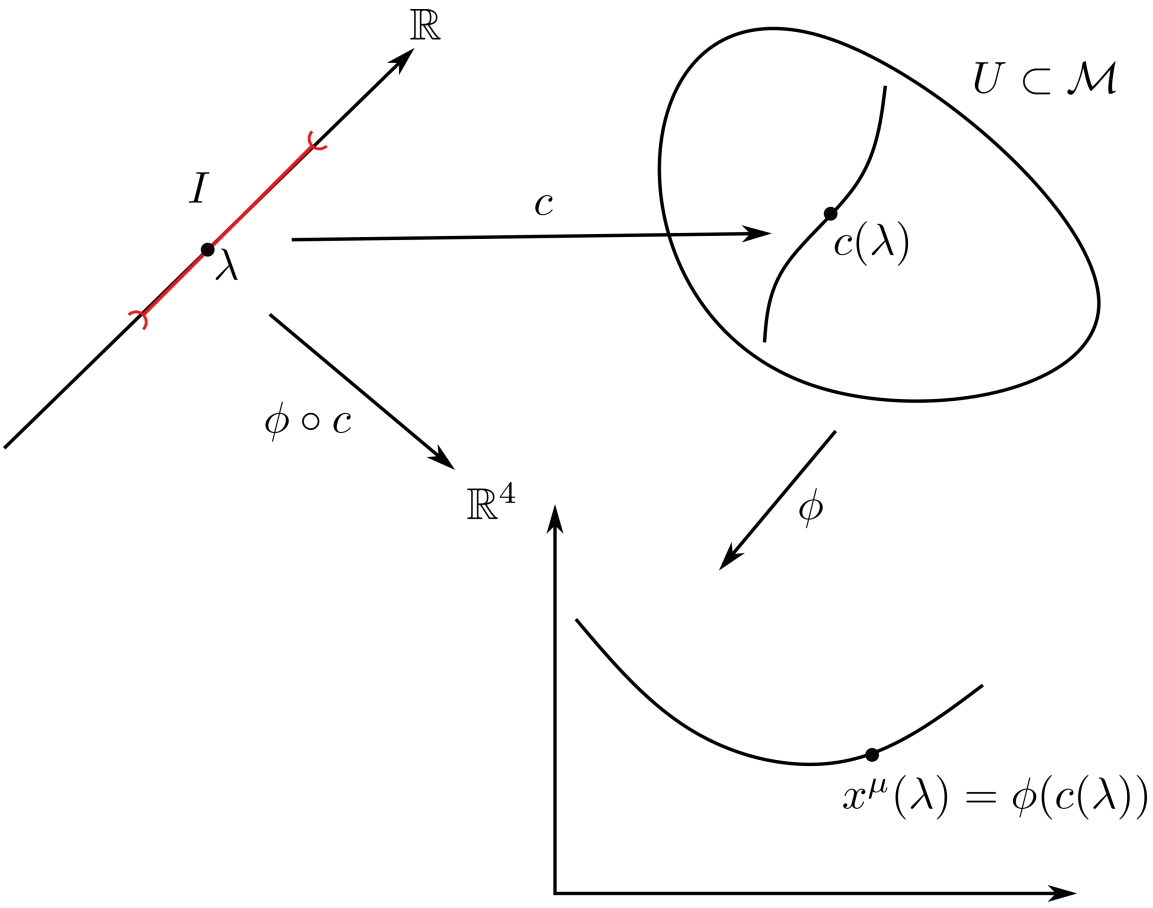


Figure 3.5: A curve c defined on the manifold $\mathcal{M}$.

is a bit trickier: how would one draw such a straight arrow on the surface of a sphere? One could select two points on the sphere and join them by an oriented arc; but which one? That would be the closest analogy, of course. But that is not how things are done in differential geometry, although the two descriptions might prove equivalent if analysed carefully. Locally, spacetime manifolds are homeomorphic to $\mathbb{R}^4$, which is a vector space, so that should provide the structure we want to define vectors. And indeed, we shall define a vector at a point p in a manifold as the tangent vector to a curve on $\mathcal{M}$ that passes through p.

Let $c : I \subseteq \mathbb{R} \to \mathcal{M}$ be a curve on the manifold $\mathcal{M}$, where $I =]a, b[$ is an open interval of $\mathbb{R}$. Let $p \in M$ be on the curve and suppose, for simplicity, that $p = c(0)$. By definition, we can use a given chart (U, ϕ) with $p \in U$ such that, locally, any point of $c(]a, b[)$ that is in U is mapped onto

$(x^0(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda))$ where $\lambda \in I$ is the parameter along the curve. This defines the map $\phi \circ c :]a, b[\to \mathbb{R}^4$. Now, for any smooth function $f : \mathcal{M} \to \mathbb{R}$ such that c is in the domain of f, we can define the function $F = f \circ \phi^{-1}$ from $\mathbb{R}^4$ into $\mathbb{R}$. The action of f on points of the curve $c(]a, b[)$ is given by the map $f \circ c : \mathbb{R} \to \mathbb{R}$ which is differentiable by construction. By writing $f \circ c = f \circ \phi^{-1} \circ \phi \circ c$ we see that:

$$\forall \lambda \in]a, b[,\ c(\lambda) \in U,\ f(c(\lambda)) = F\left(x^0(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda)\right) , \tag{3.13}$$

and therefore:

$$\left.\frac{\mathrm{d}f(c(\lambda))}{\mathrm{d}\lambda}\right|_{\lambda=0} = \sum_{\mu=0}^{3} \left.\frac{\partial F}{\partial x^\mu}\right|_{\left(x^0(0), x^1(0), x^2(0), x^3(0)\right)} \left.\frac{\mathrm{d}x^\mu(\lambda)}{\mathrm{d}\lambda}\right|_{\lambda=0} . \tag{3.14}$$

At this stage, it is convenient to introduce Einstein summation convention like we did in the chapter on Special Relativity, so that we may write the previous relation:

$$\left.\frac{\mathrm{d}f(c(\lambda))}{\mathrm{d}\lambda}\right|_{\lambda=0} = \left.\frac{\partial F}{\partial x^\mu}\right|_{\left(x^0(0), x^1(0), x^2(0), x^3(0)\right)} \left.\frac{\mathrm{d}x^\mu(t)}{\mathrm{d}t}\right|_{\lambda=0} \tag{3.15}$$

The right hand side of this formula defines a map:

$$\boldsymbol{\mathcal{X}} = \left.\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\right|_{\lambda=0} \left.\frac{\partial .}{\partial x^\mu}\right|_{\phi(c(0))} \tag{3.16}$$

that acts on functions $f \circ \phi^{-1}$ and returns a number. By 'lifting everything up' on the manifold, i.e. by defining the map $\Psi_\phi : \mathcal{F}(\mathcal{M}) \to \mathcal{F}(\mathbb{R}^n)$ such that $\Psi_\phi(f) = f \circ \phi^{-1}$, this defines a map $\boldsymbol{X_p} = \boldsymbol{\mathcal{X}} \circ \Psi_\phi : \mathcal{F}(\mathcal{M}) \to \mathbb{R}$ defined on the set of functions on $\mathcal{M}$ and returning a number:

$$\boldsymbol{X_p}(f) = \left.\frac{\mathrm{d}f(c(\lambda))}{\mathrm{d}\lambda}\right|_{\lambda=0}, \tag{3.17}$$

that is, the *directional derivative of f along the curve c at p*. In what follows we will not bother with the difference between $\boldsymbol{X_p}$ and $\boldsymbol{\mathcal{X}}$. Effectively, this amounts to identifying the tangent space (see below) and the copy of $\mathbb{R}^4$ used to define the local chart. It is not quite correct, but it is sufficient for our purpose. The map $\boldsymbol{X_p} : \mathcal{F}(\mathcal{M}) \to \mathbb{R}$ thus defined is called the *tangent vector* to the curve c at p associated with the parameter λ. The previous relation shows that the expression of the tangent vector in a given coordinate basis is $\boldsymbol{X_p} = \left.\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\right|_{\lambda=0} \left.\frac{\partial .}{\partial x^\mu}\right|_{\phi(c(0))}$. Denoting $X^\mu_p = \left.\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\right|_{\lambda=0}$, and abusing notations once more (replacing $\phi(p)$ by p in the partial derivatives), we get the expression of a vector in a given chart:

$$\boxed{\boldsymbol{X_p}[\cdot] = \left.\frac{\mathrm{d}\cdot}{\mathrm{d}\lambda}\right|_p = X_p^\mu \left.\frac{\partial\cdot}{\partial x^\mu}\right|_p .} \tag{3.18}$$

The X_p^μ's are the *components* of the vector $\boldsymbol{X_p}$ in the coordinate chart chosen.

As one can see, tangent vectors are *derivative operators acting locally on functions defined on the manifolds*. This is actually what we need to define vectors in general, irrespective to the curves they are tangent to:

Vectors in general

Let $\mathcal{M}$ be a manifold. Let $p \in M$. A *tangent vector* at p is a map: $\boldsymbol{X_p} : \mathcal{F}(M) \to \mathbb{R}$ such that:

(i) $\forall (f,g) \in \mathcal{F}(M)^2$, $\forall (\alpha,\beta) \in \mathbb{R}^2$, $\boldsymbol{X_p}\,[\alpha f + \beta g] = \alpha \boldsymbol{X_p}[f] + \beta \boldsymbol{X_p}[g]$;

(ii) $\forall (f,g) \in \mathcal{F}(M)^2$, $\boldsymbol{X_p}[fg] = \boldsymbol{X_p}[f]g(p) + f(p)\boldsymbol{X_p}[g]$ (Leibniz rule).

Note that in (ii), the operations between functions is a standard product, not a composition. These two conditions are the standard ones required to define differential operators. The set of tangent vectors at p is denoted $T_p(\mathcal{M})$. It is called the *tangent space* of $\mathcal{M}$ at p. If one supplements $T_p(\mathcal{M})$ by the addition of vectors and the multiplication of vectors by real numbers according to:

$$\forall f \in \mathcal{F}(\mathcal{M}) \quad , \quad \forall (\boldsymbol{X_p}, \boldsymbol{Y_p}) \in T_p(\mathcal{M}),\ (\boldsymbol{X_p} + \boldsymbol{Y_p})(f) = \boldsymbol{X_p}(f) + \boldsymbol{Y_p}(f) \tag{3.19}$$

$$\forall f \in \mathcal{F}(\mathcal{M}) \quad , \quad \forall \lambda \in \mathbb{R},\ \forall \boldsymbol{X_p} \in T_p(\mathcal{M}),\ (\lambda \boldsymbol{X_p})(f) = \lambda \boldsymbol{X_p}(f), \tag{3.20}$$

one simply constructs a vector space (show it), which, a posteriori, justifies the name of a vector for elements of $T_p(\mathcal{M})$. It is simple to see that vectors defined as tangent to curves indeed obey the definition of vectors we just gave (by linearity of the directional derivative). But, is the converse true? That means, can one write any tangent vector at p, $\boldsymbol{X_p}$, in coordinates as a derivative along a curve:

$$\boldsymbol{X_p}[\cdot] = X_p^\mu \left.\frac{\partial\cdot}{\partial x^\mu}\right|_p \ ? \tag{3.21}$$

The answer is yes, but it is quite tricky to prove so here we will admit it. This being the case, we see that $T_p(\mathcal{M})$ is a *vector space* of dimension 4. Indeed, the vectors $\frac{\partial}{\partial x^\mu}$ are themselves tangent

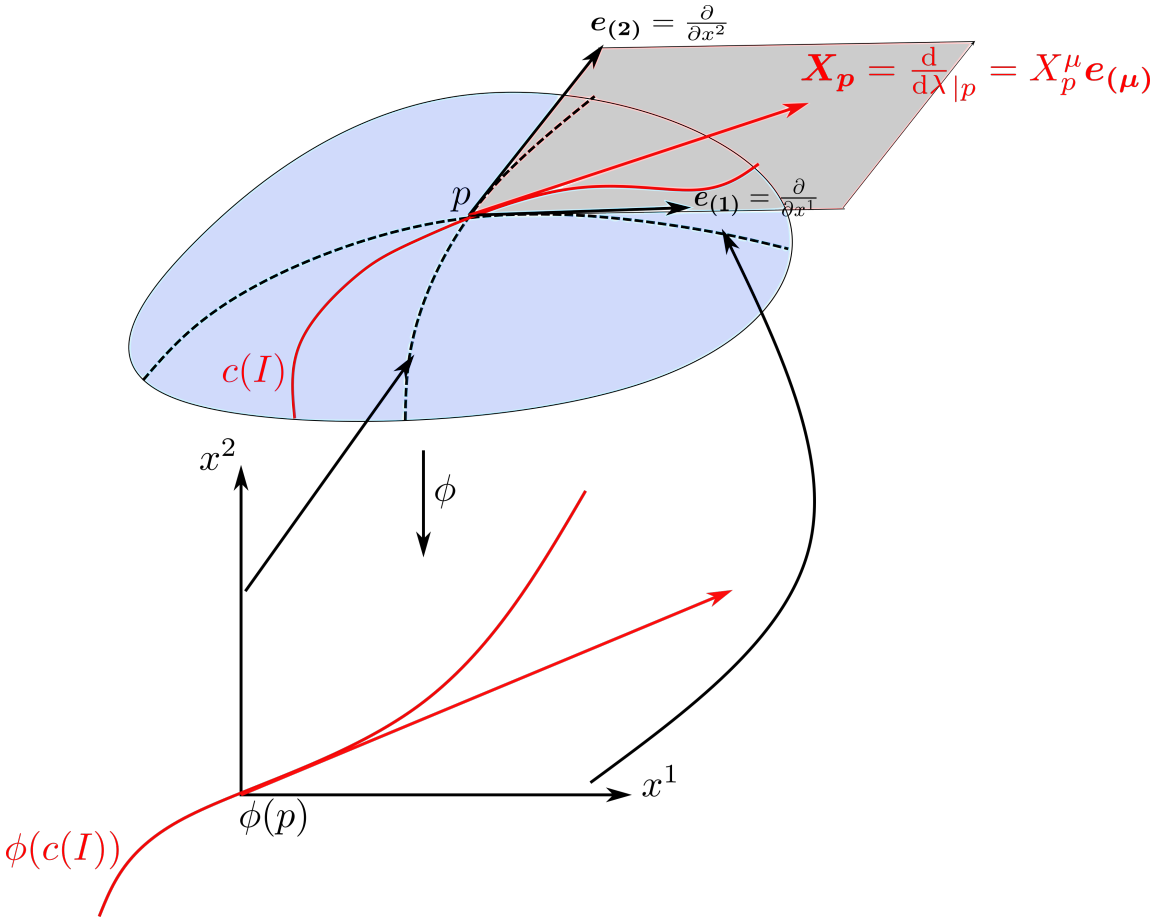


Figure 3.6: A curve parametrised by λ and its tangent vector at a point p in red. The tangent space is represented in grey and is identified to the copy of $\mathbb{R}^2$ used to define the chart. The subscript p has been omitted for the canonical basis to ease notations.

to the coordinate axes by definition so they are clearly linearly independent ($a^\mu \frac{\partial f}{\partial x^\mu} = 0$ for all f implies $a^\mu = 0$; just take $f = x^\rho$ varying ρ to show it) and they generate $T_p\mathcal{M}$ according to what we just said. Thus:

Canonical basis of T_pM

Once a local chart has been chosen, the vectors:

$$\boldsymbol{e}_{(\mu),\boldsymbol{p}} = \left.\frac{\partial}{\partial x^\mu}\right|_p , \tag{3.22}$$

for $\mu \in \{0,1,2,3\}$, form a basis of T_pM called the *canonical basis*.

Note that the partial derivatives have to be understood as keeping all the other coordinate fixed by definition:

$$\frac{\partial f}{\partial x^0} = \left.\frac{\partial f}{\partial x^0}\right|_{(x^1,x^2,x^3)} , \tag{3.23}$$

etc.

The situation is illustrated on Fig. 3.6 in the two dimensional case. An important property of vectors is the way their components are affected by a change of coordinates on the manifold:

Transformation of the components of vectors under coordinate changes

Let $\boldsymbol{X_p} \in T_p(M)$. Since T_pM is a vector space, given two local charts x^μ and $\tilde{x}^\mu$, it has two different bases, $\left\{\left.\frac{\partial}{\partial x^\mu}\right|_p\right\}$ and $\left\{\left.\frac{\partial}{\partial \tilde{x}^\mu}\right|_p\right\}$, and we have:

$$\boldsymbol{X_p} = X^\mu_p \left.\frac{\partial}{\partial x^\mu}\right|_p = \tilde{X}^\nu_p \left.\frac{\partial}{\partial \tilde{x}^\nu}\right|_p . \tag{3.24}$$

Then:

$$\forall \nu \in \{0,1,2,3\}, \tilde{X}^\nu_p = \frac{\partial \tilde{x}^\nu}{\partial x^\mu} X^\mu_p . \tag{3.25}$$

Indeed, consider the curve $c :]a,b[\to \mathcal{M}$ that has for tangent vector at p $\boldsymbol{X_p}$. We denote by $(x^0(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda)) = (\phi \circ c)(t\lambda)$, coordinates in the chart (U,ϕ) and $(\tilde{x}^0(\lambda), \tilde{x}^1(\lambda), \tilde{x}^2(\lambda), \tilde{x}^3(\lambda)) = (\varphi \circ c)(t)$, coordinates in the chart (V,φ) with $\phi(p) = (x^0(0), x^1(0), x^2(0), x^3(0))$ and $\varphi(p) = (\tilde{x}^0(0), \tilde{x}^1(0), \tilde{x}^2(0), \tilde{x}^3(0))$. Given any $f : \mathcal{M} \to \mathbb{R}$ defined along the curve locally around p, we have, by definition:

$$\boldsymbol{X_p}[f] = X^\mu_p \frac{\partial}{\partial x^\mu} \left(f \circ \phi^{-1}\right)(x^0, x^1, x^2, x^3)\Big|_{\phi(p)} . \tag{3.26}$$

Since we have seen that: $(f \circ \phi^{-1})(x^0, x^1, x^2, x^3) = (f \circ \varphi^{-1})(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)$, we get:

$$\boldsymbol{X_p}[f] = X_p^\mu \frac{\partial}{\partial x^\mu} \left(f \circ \varphi^{-1}\right)(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\Big|_{\varphi(p)} \tag{3.27}$$

$$= X_p^\mu \frac{\partial \tilde{x}^\nu}{\partial x^\mu} \frac{\partial}{\partial \tilde{x}^\nu} \left(f \circ \varphi^{-1}\right)(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\Big|_{\varphi(p)} . \tag{3.28}$$

But since in the (V, φ) chart, we have:

$$\boldsymbol{X_p}[f] = \tilde{X}_p^\nu \frac{\partial}{\partial \tilde{x}^\nu} \left(f \circ \varphi^{-1}\right)(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\Big|_{\varphi(p)} , \tag{3.29}$$

and these equalities have to be true for any f, the result follows:

$$\tilde{X}_p^\nu = X_p^\mu \frac{\partial \tilde{x}^\nu}{\partial x^\mu} . \tag{3.30}$$

The union of the tangent spaces at every point of a manifold $\mathcal{M}$ is called the *tangent bundle* of $\mathcal{M}$, and is denoted $T\mathcal{M}$:

$$T\mathcal{M} = \bigcup_{p \in \mathcal{M}} T_p\mathcal{M} . \tag{3.31}$$

If a vector of $T_p\mathcal{M}$ is assigned to each point p of the manifold $\mathcal{M}$ in a smooth way, the result is a *vector field*:

Vector field

A *vector field* on a manifold $\mathcal{M}$ is a map:

$$\boldsymbol{X} : \begin{cases} \mathcal{M} & \to & T\mathcal{M} \\ p & \mapsto & \boldsymbol{X_p} \in T_pM \end{cases} , \tag{3.32}$$

such that X is smooth.

In other words, $\boldsymbol{X}$ is a vector field on $\mathcal{M}$ iff $\boldsymbol{X}(f) \in \mathcal{F}(\mathcal{M})$ for any $f \in \mathcal{F}(\mathcal{M})$. Given an atlas, one also speaks of the components X^μ of a vector field by identifying them to the components of $\boldsymbol{X}(p) = \boldsymbol{X_p}$ in the local chart for every p. Therefore, the components of $\boldsymbol{X}$ are functions. In a given coordinate system, the vector fields $\boldsymbol{e_{(\mu)}} = \frac{\partial}{\partial x^\mu}$ are to be understood as the fields of partial derivatives at every point locally. Therefore, they are a basis of vector fields and we can write vector fields in a chart:

$$\boldsymbol{X} = X^\mu \boldsymbol{e}_{(\mu)} = X^\mu \frac{\partial}{\partial x^\mu} \,, \tag{3.33}$$

where the X^μ's are functions on the spacetime called *components* of $\boldsymbol{X}$ in the canonical basis associated with the chart.

The set of vector fields on a manifold $\mathcal{M}$ will be denoted $\mathcal{X}(\mathcal{M})$.

In the notes, we will work with vector fields rather than vectors and, to simplify notations, we will often omit the localisation ay p unless stated otherwise.

Finally, let us note that it is sometimes useful to define bases of the tangent spaces and of the vector fields that are not canonical, i.e. not associated with a coordinate system, $\{\hat{\boldsymbol{e}}_{(a)}\}$ for $a \in \{0,1,2,3\}$. Note that we use a Latin letter from the beginning of the alphabet to retain the possibility that these vectors are not associated with coordinates. These vectors (fields) define a coordinate system iff:

$$\forall (a,b) \in \{0,1,2,3\}^2,\ \left[\hat{\boldsymbol{e}}_{(a)}, \hat{\boldsymbol{e}}_{(b)}\right] = \hat{\boldsymbol{e}}_{(a)}\hat{\boldsymbol{e}}_{(b)} - \hat{\boldsymbol{e}}_{(b)}\hat{\boldsymbol{e}}_{(a)} = 0 \,, \tag{3.34}$$

where this needs to be understood using a function f, as:

$$\hat{\boldsymbol{e}}_{(a)}\left(\hat{\boldsymbol{e}}_{(b)}(f)\right) = \hat{\boldsymbol{e}}_{(b)}\left(\hat{\boldsymbol{e}}_{(a)}(f)\right) \,, \tag{3.35}$$

i.e. as saying that differentiation along the curve tangent to $\hat{\boldsymbol{e}}_{(b)}$ and then along the one tangent to $\hat{\boldsymbol{e}}_{\boldsymbol{b}}$ is the same thing as differentiation along the curve tangent to $\hat{\boldsymbol{e}}_{(\boldsymbol{a})}$ and then along the one tangent to $\hat{\boldsymbol{e}}_{\boldsymbol{b}}$. The direct implication is trivial but the converse is a bit trickier to prove; we will admit it here. The brackets $[\cdot,\cdot]$ are known as the *Lie brackets*.

3.3.4 Cotangent space

Since $T_p\mathcal{M}$ is a vector space, we can define its *dual*, consisting of the linear maps defined on $T_p\mathcal{M}$, sending it to $\mathbb{R}$; see appendix A:

One-forms

Let $\mathcal{M}$ be a manifold and $p \in M$. The cotangent space at p is the vector space, denoted $T_p^*\mathcal{M}$, of linear functions:

$$\boldsymbol{w_p} : \begin{cases} T_p\mathcal{M} & \to & \mathbb{R} \\ \boldsymbol{X_p} & \mapsto & \boldsymbol{w_p}(\boldsymbol{X_p}). \end{cases} \tag{3.36}$$

A vector of $T_p^*\mathcal{M}$ is called a *dual vector*, a *covector* or a *cotangent vector*, sometimes a *one-form* or even a *covariant vector* at p. The simplest example of a one-form is provided by the differential of a function at p, $\mathrm{d}f_p$. Letting $f : \mathcal{M} \to \mathbb{R}$, it is defined via:

$$\forall \boldsymbol{X_p} \in T_p\mathcal{M}, \mathrm{d}f_p(\boldsymbol{X_p}) = \boldsymbol{X_p}[f] = X_p^\mu \left.\frac{\partial f}{\partial x^\mu}\right|_p , \tag{3.37}$$

where, for simplicity, we have again replaced $f \circ \phi^{-1}$ by f, where ϕ is a local coordinate function. In terms of these local coordinates, remember that we have a coordinate basis $\{\boldsymbol{e}_{(\mu)}\} = \left\{\frac{\partial}{\partial x^\mu}\right\}$ for $T_p\mathcal{M}$ (up to identification with $T_p\mathbb{R}^n$ via the pushforward of ϕ). Therefore, we can define a canonical *dual basis* associated to $\{\boldsymbol{e}_{(\mu)}\}$, that is a basis of $T_p^*\mathcal{M}$ (up to identification with $T_p^*\mathbb{R}^n$ via the pullback of ϕ; see appendix B), usually denoted $\left\{\boldsymbol{\omega}^{(\mu)}\right\} = \{\mathrm{d}x^\mu\}$ in this context, and characterised by:

$$\mathrm{d}x^\mu \left(\frac{\partial}{\partial x^\nu}\right) = \delta^\mu{}_\nu \, . \tag{3.38}$$

In this basis, any differential has components $(df)_\mu$ given by:

$$\mathrm{d}f_p = \mathrm{d}f_\mu dx^\mu \, . \tag{3.39}$$

Therefore, we have:

$$\begin{aligned} \mathrm{d}f_p(\boldsymbol{X_p}) &= (\mathrm{d}f_\mu \mathrm{d}x^\mu)\left(X^\nu \frac{\partial}{\partial x^\nu}\right) = \mathrm{d}f_\mu X^\nu \mathrm{d}x^\mu \left(\frac{\partial}{\partial x^\nu}\right) \\ &= \mathrm{d}f_\mu X^\mu \, . \end{aligned} \tag{3.40}$$

Hence, by identifying with the definition:

$$\forall \mu \in \{0,1,2,3\}, \mathrm{d}f_\mu = \left.\frac{\partial f}{\partial x^\mu}\right|_p . \tag{3.41}$$

In general, any one-form ω_p is written, in a local coordinate system, $\boldsymbol{\omega_p} = \omega_\mu \mathrm{d}x^\mu$, and the action of this one-form on vectors at p is given, in coordinates, by:

$$\boldsymbol{\omega_p}(\boldsymbol{X_p}) = \omega_\mu X_p^\mu \ . \tag{3.42}$$

If, instead of a canonical basis in $T_p\mathcal{M}$ we use a non-coordinate basis $\left\{\boldsymbol{e_{(a)}}\right\}$, we can of course define its dual basis in $T^*\mathcal{M}$, $\left\{\boldsymbol{\omega}^{(a)}\right\}$ such that:

$$\boldsymbol{\omega}^{(\boldsymbol{a})}\left(\boldsymbol{e_{(b)}}\right) = \delta^a{}_b \ . \tag{3.43}$$

All the previous results carry forward trivially.
We can also define an inner product between one-forms and vectors at a point:

$$\langle .,. \rangle : \left\{ \begin{array}{ccc} T_p^*\mathcal{M} \times T_p\mathcal{M} & \to & \mathbb{R} \\ (\boldsymbol{\omega_p}, \boldsymbol{X_p}) & \mapsto & \left\langle \boldsymbol{\omega_p}, \boldsymbol{X_p} \right\rangle = \boldsymbol{\omega_p}(\boldsymbol{X_p}) \end{array} \right. . \tag{3.44}$$

If we now consider a change of coordinate on $\mathcal{M}$ at p, taking two charts (U, ϕ) and (V, φ) with $p \in U \cap V$, we must have, for a one-form $\boldsymbol{\omega} \in T_p^*M$: (forgetting the index p from now one to simplify notations):

$$\boldsymbol{\omega} = \omega_\mu \mathrm{d}x^\mu = \tilde{\omega}_\nu \mathrm{d}\tilde{x}^\nu \ , \tag{3.45}$$

where we have noted (x^μ) the coordinates in $\phi(U)$, and $(\tilde{x}^\nu)$ those in $\varphi(V)$, and similarly for the associated canonical dual bases. Then, we must have:

$$\boldsymbol{\omega}(\boldsymbol{X}) = \omega_\mu X^\mu = \tilde{\omega}_\nu \tilde{X}^\nu \ . \tag{3.46}$$

Since $X^\nu = \frac{\partial x^\nu}{\partial \tilde{x}^\mu} \tilde{X}^\mu$, this leads to:

$$\omega_\mu \frac{\partial x^\mu}{\partial \tilde{x}^\nu} \tilde{X}^\nu = \tilde{\omega}_\nu \tilde{X}^\nu \ , \tag{3.47}$$

and therefore, under coordinate changes, the components of one-forms transform as:

$$\boxed{\forall \mu \in \{0,1,2,3\},\ \tilde{\omega}_\mu = \omega_\nu \frac{\partial x^\nu}{\partial \tilde{x}^\mu} \ .} \tag{3.48}$$

The set $T^*\mathcal{M} = \bigcup_{p \in \mathcal{M}} T_p^*\mathcal{M}$ is the *cotangent bundle* of the manifold $\mathcal{M}$. Similarly to what happens for vectors, we can define one-form fields as follows:

One-form field

A field of one-forms (or field of covectors) is an application $\boldsymbol{\Omega} : \mathcal{M} \to T^*\mathcal{M}$ that associates a one-form $\boldsymbol{\omega} \in T_p^*\mathcal{M}$ to any point $p \in M$ smoothly. Then, each component of a one-form field, $\Omega_\nu(p)$ is a function.

The set of one-form fields on a manifold $\mathcal{M}$ will be denoted $\Omega(\mathcal{M})$.

3.3.5 Tensors

At any point p of a manifold $\mathcal{M}$, we now have two vector spaces, $T_p\mathcal{M}$ and its dual $T_p^*\mathcal{M}$, therefore, we can define tensors as usual on vector spaces; see appendix A:

Tensors

A *tensor* of order $(r, s) \in \mathbb{N}^2$ at $p \in \mathcal{M}$ is a multilinear map:

$$\boldsymbol{T} : \begin{cases} \underbrace{T_p^*\mathcal{M} \times ... \times T_p^*\mathcal{M}}_{r \text{ times}} \times \underbrace{T_p\mathcal{M} \times ... \times T_p\mathcal{M}}_{s \text{ times}} & \to \quad \mathbb{R} \\ (\boldsymbol{\omega^1}, ..., \boldsymbol{\omega^r}, \boldsymbol{X_1}, ..., \boldsymbol{X_s}) & \mapsto \quad \boldsymbol{T}(\boldsymbol{\omega^1}, ..., \boldsymbol{\omega^r}, \boldsymbol{X_1}, ..., \boldsymbol{X_s}) \end{cases} . \tag{3.49}$$

In the local coordinate basis and its dual, we get:

$$\boldsymbol{T} = T^{\mu_1 ... \mu_r}{}_{\nu_1 ... \nu_s} \frac{\partial}{\partial x^{\mu_1}} \otimes ... \otimes \frac{\partial}{\partial x^{\mu_r}} \otimes \mathrm{d}x^{\nu_1} \otimes ... \otimes \mathrm{d}x^{\nu_s} , \tag{3.50}$$

where the $T^{\mu_1 ... \mu_r}{}_{\nu_1 ... \nu_s}$ are numbers called the *components* of the tensor $\boldsymbol{T}$ in the coordinate basis. Then, for any r one-forms and s vectors:

$$\begin{aligned} \boldsymbol{T}(\boldsymbol{\omega^1}, ..., \boldsymbol{\omega^r}, \boldsymbol{X_1}, ..., \boldsymbol{X_s}) &= T^{\mu_1 ... \mu_r}{}_{\nu_1 ... \nu_s} \left(\frac{\partial}{\partial x^{\mu_1}} \otimes ... \otimes \frac{\partial}{\partial x^{\mu_r}} \otimes \mathrm{d}x^{\nu_1} \otimes ... \otimes \mathrm{d}x^{\nu_s} \right) \\ &\quad \left(\omega^1_\mu \mathrm{d}x^\mu, ..., \omega^r_\mu \mathrm{d}x^\mu, X_1^\nu \frac{\partial}{\partial x^\nu}, ..., X_s^\nu \frac{\partial}{\partial x^\nu} \right) \qquad (3.51) \\ &= T^{\mu_1 ... \mu_r}{}_{\nu_1 ... \nu_s} \left((\omega^1_\mu \mathrm{d}x^\mu)(\frac{\partial}{\partial x^{\mu_1}}) \times ... \times (\omega^r_k \mathrm{d}x^\mu)(\frac{\partial}{\partial x^{\mu_r}}) \right. \\ &\quad \left. \times \mathrm{d}x^{\nu_1}(X_1^\nu \frac{\partial}{\partial x^\nu}) \times ... \times \mathrm{d}x^{\nu_s}(X_s^\nu \frac{\partial}{\partial x^\nu}) \right) \qquad (3.52) \\ &= T^{\mu_1 ... \mu_r}{}_{\nu_1 ... \nu_s} \omega^1_{\mu_1} ... \omega^r_{\mu_r} X_1^{\nu_1} ... X_s^{\nu_s} . \qquad (3.53) \end{aligned}$$

Note how the Einstein summation convention is making our life easier! The set of tensors of order $(r, s) \in \mathbb{N}^2$ at $p \in \mathcal{M}$ is a vector space (can you show it?), denoted $T^r_{s,p}\mathcal{M}$. Again, we denote by $T^r_s\mathcal{M} = \bigcup_{p\in\mathcal{M}} T^r_{s,p}\mathcal{M}$ and we call it the *tensor bundle* of order (r, s). We can then define a *tensor field* of order (r, s) by the smooth application: $\boldsymbol{T} : M \to T^r_s\mathcal{M}$ such that $\boldsymbol{T(p)} \in T^r_{s,p}\mathcal{M}$. The set of tensor fields of type (r, s) on a manifold $\mathcal{M}$ will be denoted $\mathcal{T}^r_s(\mathcal{M})$.

Other structures as tensors

What kind of tensors are functions, vectors and one-forms?

Finally, we can give the law of transformations of the components of a tensor by a local change of chart:

Transformations of the components of tensors under coordinate changes

Let $\boldsymbol{T} \in T^r_{s,p}\mathcal{M}$ at $p \in \mathcal{M}$. Let (U, ϕ) and (V, φ) be two local charts at p with $p \in U \cap V$. Noting $(x^\mu) = \phi(p)$ and $(\tilde{x}^\nu) = \varphi(p)$ respectively, the components of the tensor $\boldsymbol{T}$ in the chart φ are given, in terms of the components in the chart ϕ by:

$$\tilde{T}^{\mu_1\ldots\mu_r}{}_{\nu_1\ldots\nu^s} = \frac{\partial\tilde{x}^{\mu_1}}{\partial x^{\rho_1}}\ldots\frac{\partial\tilde{x}^{\mu_r}}{\partial x^{\rho_r}}\frac{\partial x^{\sigma_1}}{\partial\tilde{x}^{\nu_1}}\ldots\frac{\partial x^{\sigma_s}}{\partial\tilde{x}^{\nu_s}}T^{\rho_1\ldots\rho_r}{}_{\sigma_1\ldots\sigma_s} \,. \tag{3.54}$$

The proof is similar to the ones given for vectors and one-forms and is left to the reader.
Often, especially in the physics literature, the indices of the components of a tensor that are 'upstairs' are called *contravariant*, because, in a local coordinate change, they transform the opposite way, compared to the basis vectors, whereas the indices 'downstairs' are called *covariant*, because they transform the same way as the basis vectors. The law of transformation of tensor components under a local change of chart is very important as, when given a multilinear function, it allows one to test if it is a tensor: 'if an object carrying indices transforms like a tensor, it is one, if not, it is not'. If such an object only transforms like a tensor under a subset of coordinate transformations, then it is called a tensor 'under these transformations'. Also note that, according to the law of transformation, if the components of a tensor are zero in a given chart, then they are zero in any chart, which means that the tensor itself is identically zero at the point of the manifold considered.

Finally, let us note a very useful property of tensors, illustrating it with a $(1, 1)$ tensor. In a given local basis on $T_p\mathcal{M}$, $\boldsymbol{e}_{(a)}$, and its dual basis $\boldsymbol{\omega}^{(a)}$ in $T^*_p\mathcal{M}$, the components of a tensor $\boldsymbol{T} \in T^1_{1,p}$ are

given by:

$$T^a{}_b = \boldsymbol{T}\left(\boldsymbol{\omega}^{(a)}, \boldsymbol{e}_{(b)}\right) \ . \tag{3.55}$$

The proof of this statement is a good exercise left to the reader.

3.4 The metric tensor

3.4.1 Definition

In General Relativity, there is a tensor that plays a very important role, as it carries the degrees of freedom associated with the gravitational field: the metric tensor. As we have seen, in Special Relativity, geometric quantities such as distances, time intervals, length of vectors and angles are determined via a bilinear map $\boldsymbol{\eta}$ defined on vectors, that we called the Minkowski metric tensor. Besides, the equivalence principle led us to state that locally, in a suitably chosen coordinate system, namely in a local inertial frame, the properties of spacetime must reduce to those of Special Relativity. This means that on our spacetime manifold $\mathcal{M}$, we must have a $(0,2)$ tensor field acting on vectors in tangent spaces which, locally, can have the same components as $\boldsymbol{\eta}$ in a suitably chosen coordinate system. This tensor is the *metric tensor*.

The interest of this new structure is threefold. First, it will allow one to define the scalar product of two vectors in the same tangent space, thus enabling one to talk about length of vectors and angles between vectors at the same point in spacetime. Second, it will also allow us to fix very nicely a 'natural' way to compare vectors and tensors at different points of the manifold. Finally, it will define a natural inner product at a given point of the manifold, allowing one to obtain the standard identification between T_pM and its dual, i.e. leading to a very convenient identification between vectors and one-forms.

Metric tensor

Let $\mathcal{M}$ be the spacetime manifold. A *pseudo-Riemannian metric* $\boldsymbol{g}$ on $\mathcal{M}$ (or simply a *metric tensor* $\boldsymbol{g}$ on $\mathcal{M}$) is a tensor field of type $(0,2)$ (bilinear form) which satisfies the following properties at any point $p \in \mathcal{M}$:

(i) **Symmetry**: $\forall (\boldsymbol{U}, \boldsymbol{V}) \in \mathcal{X}(\mathcal{M})$, $\boldsymbol{g}(\boldsymbol{U}, \boldsymbol{V}) = \boldsymbol{g}(\boldsymbol{V}, \boldsymbol{U})$;

(ii) **Non-degeneracy**: if for $\boldsymbol{V} \in \mathcal{X}(\mathcal{M})$, $\forall \boldsymbol{U} \in \mathcal{X}(\mathcal{M})$, $\boldsymbol{g}(\boldsymbol{U}, \boldsymbol{V}) = 0$, then $\boldsymbol{V} = 0$.

Therefore, locally, a pseudo-Riemannian metric is a symmetric non-degenerate bilinear form. Note that a pseudo-Riemannian metric is not necessarily positive-definite: there might exist vectors $\boldsymbol{V} \neq 0$ such that $\boldsymbol{g}(\boldsymbol{V}, \boldsymbol{V}) = 0$, something we are familiar with from Special Relativity.
In a local chart, we can write[5]:

$$\boldsymbol{g} = g_{\mu\nu} \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu \tag{3.56}$$

Since the $g_{\mu\nu}$'s form a real symmetric matrix, the associated eigenvalues are real. Since $\boldsymbol{g}$ is pseudo-Riemannian, some eigenvalues can be negative[6]. It turns out that the number of negative eigenvalues is called the *index* of the metric and is an intrinsic property. If this number is equal to one, one speaks of a *Lorentzian* metric, like in the case of the metric of spacetime in Special Relativity. This is the case we are interested in. It is always possible to choose a local coordinate system such that, locally, the metric may be written as a diagonal matrix with ± 1 as eigenvalues. This means that, locally, one can always find a coordinate system such that $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$ (Minkowski metric). Actually, to encode the equivalence principle, we will need a little bit more:

Local inertial frame

At any event $C \in \mathcal{M}$, we can find local coordinates X^μ such that:

$$\boldsymbol{g}_{|C} = \eta_{\mu\nu} \mathrm{d}X^\mu \otimes \mathrm{d}X^\nu \ , \tag{3.57}$$

and:

$$\forall (\mu, \nu, \rho) \in \{0, 1, 2, 3\}^3, \ \frac{\partial g_{\mu\nu}}{\partial X^\rho}(C) = 0 \ . \tag{3.58}$$

The associated frame is called a *local inertial frame* at C. In this frame, the laws of physics are those of Special Relativity.

We will come back to inertial frames later in greater details. Special Relativity relied on the existence of a specific class of frames, called inertial frames, which were related to each other via specific transformations, the Lorentz transformations. In General Relativity, inertial frames will play a role, essentially by the mere fact that they exist. But by the very nature of a manifold, all coordinate systems are treated on a equal footing. This is *general covariance*. It means that all the

[5]Note that we will mostly work in coordinate frames. However, everything can be rewritten in non-coordinate frames by the appropriate change of notations.

[6]Note that in any case, the eigenvalues must be non-zero because the matrix has to be non-degenerate, and invertible.

laws of physics will have to be written in a form that remains the same in *every allowed coordinate system*. We will get back to this later. For now, we can just check that under a generic coordinate transformation, $x^\mu \mapsto \tilde{x}^\mu$ the components of the metric tensor transform as:

$$\tilde{g}_{\mu\nu}(\tilde{x}) = \frac{\partial x^\rho}{\partial \tilde{x}^\mu}\frac{\partial x^\sigma}{\partial \tilde{x}^\nu} g_{\rho\sigma}(x) \ . \tag{3.59}$$

3.4.2 Classification of vectors

In the same way as in Special Relativity, vectors at p can then be given a 'length' via the scalar product interpretation of the metric. Let $\boldsymbol{X}$ be a vector field. Its length function is given by:

$$L^2(p) = \boldsymbol{g}\left(\boldsymbol{X}, \boldsymbol{X}\right) = g_{\mu\nu} X^\mu X^\nu \ . \tag{3.60}$$

Because L^2 is a function, its values are invariant under a coordinate transformation therefore, we can evaluate them in the local inertial frame at each point and conclude that $T_p\mathcal{M}$ has *the same causal structure as Minkowski spacetime*.

Types of vectors at a point

Let $p \in \mathcal{M}$ and $X \in T_p\mathcal{M}$ be a vector at p. Then:

- If $\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{X}) < 0$, then $\boldsymbol{X}$ is *timelike*;
- If $\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{X}) = 0$ and $\boldsymbol{X} \neq 0$, then $\boldsymbol{X}$ is *lightlike* or *null*;
- If $\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{X}) > 0$, then $\boldsymbol{X}$ is *spacelike*.

A curve tangent to $\boldsymbol{X}$ at p is either timelike, lightlike or spacelike at p depending on the case; see Fig. 3.7.
Note that we will say that two non-zero vectors are orthogonal iff $\boldsymbol{g}\left(\boldsymbol{u}, \boldsymbol{v}\right) = 0$.

Remember that in Special Relativity we could pick a time orientation by picking a unit timelike vector. Then, this time orientation was preserved under orthochronous Lorentz transformations. In General Relativity, a time orientation can also be specified but only locally. If we want it to extend to finite regions of spacetime, it needs to involve a timelike vector field, i.e. a vector field that is

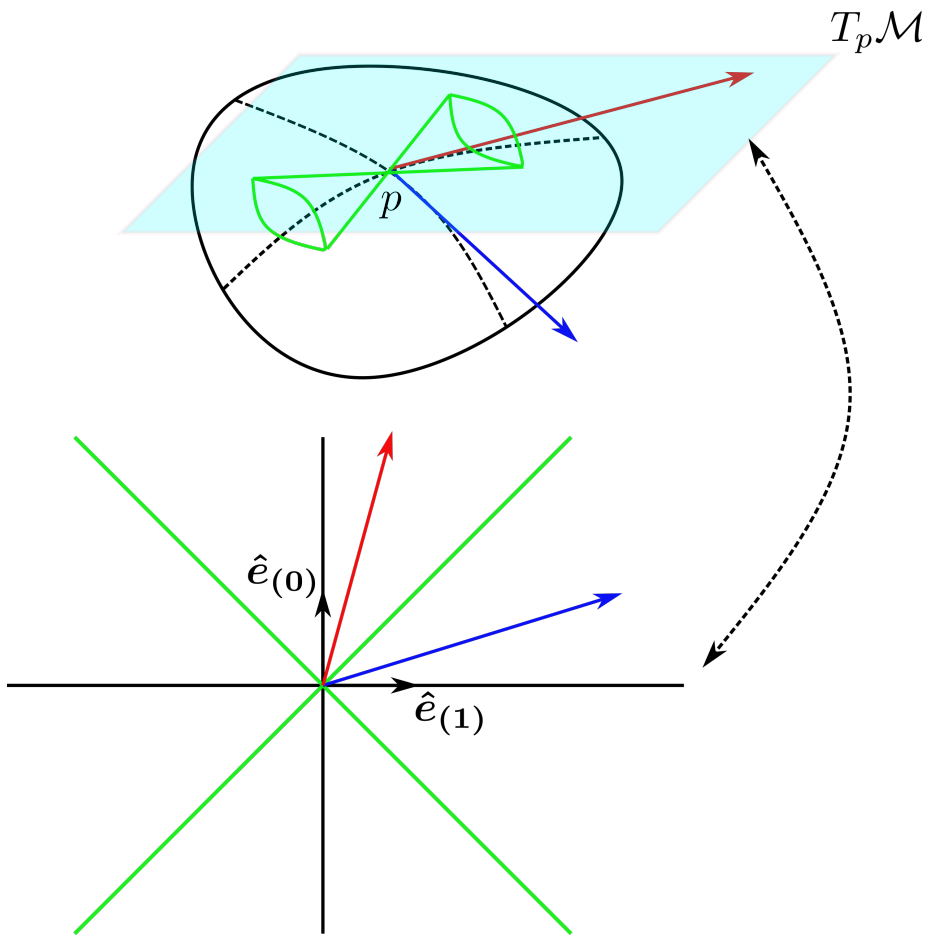


Figure 3.7: Local causal structure at $p \in \mathcal{M}$. Vectors in $T_p\mathcal{M}$ are either spacelike (blue), timelike (red) or lightlike (green lightcone). The curves tangent to them are then also spacelike, timelike or lightlike at the point p. In local inertial coordinates, this structure is exactly identical to the one in Special Relativity (bottom). The lightcone that determines the local causal structure of spacetime changes from point to point *a priori*. It looks like a $\pm\pi/4$ cone only in the local inertial frame but its shape is coordinate dependent.

timelike across the entire region. Let $\boldsymbol{u}$ be such a vector field with:

$$\forall p \in U,\ \boldsymbol{g_p}\left(\boldsymbol{u(p)}, \boldsymbol{u(p)}\right) < 0\ . \tag{3.61}$$

Then, $\boldsymbol{u}$ defines a time-orientation on U and:

- $\boldsymbol{V}$ timelike or lightlike is *future-directed* in U iff $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{V}) < 0$;
- $\boldsymbol{V}$ timelike or lightlike is *past-directed* in U iff $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{V}) > 0$.

This follows from evaluating the scalar product in the local inertial frame.

As an application, consider two points p and $p + \delta p$ on the manifold, such that both belong to the same local chart (U, ϕ). They have coordinates x^μ and $x^\mu + \delta x^\mu$ respectively. Let us consider a curve C through both points and assume that they are infinitesimally close. Let f be a function on $\mathcal{M}$. Let λ be a parameter along the curve with tangent vector $\boldsymbol{V}$:

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} = \boldsymbol{V}(f) \ . \tag{3.62}$$

If, to go from p to $p + \mathrm{d}p$, one needs a change in λ given by $\mathrm{d}\lambda$, then, we can define the tangent vector at p:

$$\mathbf{d}\boldsymbol{p} = \mathrm{d}\lambda \boldsymbol{V} \in T_p\mathcal{M} \ , \tag{3.63}$$

which is tangent to the curve C by construction. We note that:

$$\begin{aligned} \mathbf{d}\boldsymbol{p}(f) =& \mathrm{d}\lambda \boldsymbol{V}(f) && (3.64)\\ =& \mathrm{d}\lambda \frac{\mathrm{d}f}{\mathrm{d}\lambda} = \mathrm{d}f(\lambda) && (3.65)\\ =& f(p + \mathrm{d}p) - f(p) \ , && (3.66) \end{aligned}$$

which shows that $\mathbf{d}\boldsymbol{p}$ is independent of λ and only depends on the points p and $p + \mathrm{d}p$. We call it the *infinitesimal displacement* between the two points. In the local chart, we get:

$$\begin{aligned} \mathbf{d}\boldsymbol{p}(f) =& f\left(x^\mu + \delta x^\mu\right) - f\left(x^\mu\right) && (3.67)\\ =& \frac{\partial f}{\partial x^\mu} \delta x^\mu \ , && (3.68) \end{aligned}$$

so that we can write:

$$\mathbf{d}\boldsymbol{p} = \delta x^\mu \boldsymbol{e}_{(\mu)} \ . \tag{3.69}$$

Now, we can calculate the 'length' of the infinitesimal displacement:

$$\begin{aligned} \mathrm{d}s^2 =& \boldsymbol{g}\left(\delta x^\mu \boldsymbol{e}_{(\mu)}, \delta x^\nu \boldsymbol{e}_{(\nu)}\right) && (3.70)\\ =& \boldsymbol{g}\left(\boldsymbol{e}_{(\mu)}, \boldsymbol{e}_{(\nu)}\right) \delta x^\mu \delta x^\nu && (3.71)\\ =& g_{\mu\nu} \delta x^\mu \delta x^\nu \ , && (3.72) \end{aligned}$$

which represents the square of the length of the infinitesimal displacement. This defines the *space time interval*:

$$\boxed{ds^2 = g_{\mu\nu} dx^\mu dx^\nu} \ . \tag{3.73}$$

Be careful that in this expression, dx^μ is not the canonical basis one-form.

3.4.3 Metric duality

Let us see another useful property of the metric tensor. At a point $p \in \mathcal{M}$, given a vector $\boldsymbol{U} \in T_p\mathcal{M}$, the induced map $\boldsymbol{g}(\boldsymbol{U}, \cdot) : T_p\mathcal{M} \to \mathbb{R}$ is clearly linear, and it is a one-form: $\boldsymbol{\omega_U} = \boldsymbol{g}(\boldsymbol{U}, \cdot) \in T_p^*\mathcal{M}$. Hence, the metric $\boldsymbol{g}$ naturally introduces an isomorphism between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$: to any vector $\boldsymbol{U} \in T_p\mathcal{M}$, we can associate uniquely a one-form $\boldsymbol{g}(\boldsymbol{U}, \cdot) \in T_p^*\mathcal{M}$[7]. In terms of coordinates, this is used to 'transform' contravariant indices into covariant indices. In a local chart (U, ϕ) around p, we can write: $\boldsymbol{X} = X^\mu \boldsymbol{e}_{(\mu)}$ and $\boldsymbol{g} = g_{\mu\nu} dx^\mu \otimes dx^\nu$. Then,

$$\boldsymbol{\omega_X} = \boldsymbol{g}(\boldsymbol{X}, .) = g_{\mu\nu} X^\mu dx^\nu \in T_p^*\mathcal{M} \ . \tag{3.74}$$

Then, by writing

$$\boldsymbol{\omega_X} = \omega_\mu dx^\mu \ , \tag{3.75}$$

we get:

$$\omega_\nu = g_{\mu\nu} X^\mu \ . \tag{3.76}$$

Usually, when there is no confusion possible, this is noted $\omega_\nu = X_\nu$. This can be generalised to tensors of arbitrary orders, e.g.:

$$T_{\mu\nu} = g_{\mu\rho} g_{\nu\sigma} T^{\rho\sigma} \ , \tag{3.77}$$

allows to transform $\boldsymbol{T} = T^{\mu\nu} \boldsymbol{e}_{(\mu)} \otimes \boldsymbol{e}_{(\nu)}$, tensor of order $(2, 0)$, into a tensor of order $(0, 2)$. Because the map between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ is an isomorphism, it has an inverse which associates a unique vector of $T_p\mathcal{M}$ to any one-form. By identifying the components $g_{\mu\nu}$ with the entries of an $n \times n$ matrix, we can then construct an inverse matrix, denoted $g^{\mu\nu}$ such that:

[7] To see that it is an isomorphism, note that it maps the basis vectors $\boldsymbol{e}_{(\mu)} = \frac{\partial}{\partial x^\mu}$ onto a basis of $T_p^*\mathcal{M}$.

Inverse metric

$$g_{\mu\rho}g^{\rho\nu} = \delta^{\nu}{}_{\mu} \ . \tag{3.78}$$

This defines a $(2,0)$ tensor acting on one-forms and called the *inverse metric*:

$$\boldsymbol{g^{-1}} = g^{\mu\nu}\frac{\partial}{\partial x^{\mu}} \otimes \frac{\partial}{\partial x^{\nu}} \ . \tag{3.79}$$

Then, it is easy to see that, given a one-form $\boldsymbol{\omega}$, the vector associated with it via the isomorphism between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ is given, in terms of coordinates, by $\boldsymbol{X} = \omega^{\mu}\boldsymbol{e}_{(\mu)}$ such that:

$$\omega^{\mu} = g^{\mu\nu}\omega_{\nu} \ . \tag{3.80}$$

Hence, the natural isomorphism also allows to transform covariant indices into contravariant ones. And this extends to tensors of arbitrary orders in the same way as before. This is why physicists say that indices are lowered and raised using the metric and its inverse.

3.4.4 The metric in the weak field limit

For pedagogical reasons, we will introduce here the metric tensor associated with a weak gravitational field. This result is rigorously derived in subsection 5.2.3.

Metric of spacetime in the Newtonian limit

In presence of a weak, slowly varying, Newtonian gravitational potential $\Phi_N(x^{\mu})$, there is an orthonormal coordinate system (t,x,y,z) for which the metric of spacetime takes the simple form, at leading order in Φ_N:

$$\boldsymbol{g} = -(1+2\Phi_N)\,\mathrm{d}t\otimes\mathrm{d}t + (1-2\Phi_N)\left[\mathrm{d}x\otimes\mathrm{d}x + \mathrm{d}y\otimes\mathrm{d}y + \mathrm{d}z\otimes\mathrm{d}z\right] \ . \tag{3.81}$$

The line element is then:

$$\mathrm{d}s^2 = -(1+2\Phi_N)\,\mathrm{d}t^2 + (1-2\Phi_N)\,\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ . \tag{3.82}$$

In this framework, every quantity must be understood as being valid at first order in Φ_N, so that they must all be expanded at first order in terms of the potential or its derivatives. In the remainder

of this chapter, we will see that it gives a good description of Newtonian gravitation in a General Relativistic language. This will help us illustrate the concepts and techniques of General Relativity in a familiar context.

3.5 Kinematics

The trajectories of particles, massless or massive, follow quite naturally from their counterparts in Special Relativity. Most defining properties listed in section 2.5 will remain valid provided one substitute $\boldsymbol{g}$ for $\boldsymbol{\eta}$ in the definitions.

3.5.1 Lightlike curves

Massless particles such as photons follow *lightlike curve*:

Lightlike curve

A curve $C \subset \mathcal{M}$ is lightlike iff its tangent vector field is lighlike. This property does not depend on the parametrisation. Given $c : \lambda \in \mathbb{R} \mapsto c(\lambda) \in \mathcal{M}$ such a parametrisation, this means that the tangent vector field $\boldsymbol{k}(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda} = k^\mu \frac{\partial}{\partial x^\mu}$ satisfies:

$$\boldsymbol{g}\left(\boldsymbol{k}, \boldsymbol{k}\right) = g_{\mu\nu} k^\mu k^\nu = 0 \,. \tag{3.83}$$

All lightlike curves through a point $p \in \mathcal{M}$ are tangent to the local lightcone of Fig. 3.7.

3.5.2 Timelike curves

Massive particles follow *timelike curves*:

Timelike curve

A curve $C \subset \mathcal{M}$ is timelike iff its tangent vector field is timelike. This property does not depend on the parametrisation. Given $c : \lambda \in \mathbb{R} \mapsto c(\lambda) \in \mathcal{M}$ such a parametrisation, this means that the tangent vector field $\boldsymbol{U}(\lambda) = \frac{\mathrm{d}}{\mathrm{d}\lambda} = U^\mu \frac{\partial}{\partial x^\mu}$ satisfies:

$$\boldsymbol{g}\left(\boldsymbol{U}, \boldsymbol{U}\right) = g_{\mu\nu} U^\mu U^\nu < 0 \,. \tag{3.84}$$

All timelike curves through a point $p \in \mathcal{M}$ point inside to the local lightcone of Fig.3.7. As in the special relativistic case, there is a preferred parametrisation along timelike curves: the proper time measured by the particle/observer along the timelike curve:

Proper time

Along a timelike curve $C \subset \mathcal{M}$ parametrised by $\lambda \in \mathbb{R}$, the proper time τ is defined by:

$$\begin{aligned} \mathrm{d}\tau =& \sqrt{-\boldsymbol{g}\left(\boldsymbol{U},\boldsymbol{U}\right)}\mathrm{d}\lambda && (3.85)\\ =& \sqrt{-g_{\mu\nu}U^\mu U^\nu}\mathrm{d}\lambda && (3.86)\\ =& \sqrt{-g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}}\mathrm{d}\lambda\ . && (3.87) \end{aligned}$$

For $\lambda = \tau$, writing $\boldsymbol{u} = \frac{\mathrm{d}}{\mathrm{d}\tau}$ for the tangent vector, we see immediately that $\boldsymbol{g}\left(\boldsymbol{u},\boldsymbol{u}\right) = -1$, so that $\boldsymbol{u}$ is a unit vector. It is called the 4-velocity of the particle along its worldline C. The 4-momentum of the particle is then just $\boldsymbol{p} = m\boldsymbol{u}$, so that we still have:

$$\boldsymbol{g}\left(\boldsymbol{p},\boldsymbol{p}\right) = g_{\mu\nu}p^\mu p^\nu = -m^2\ . \tag{3.88}$$

3.5.3 Observers and observables

We will call *observer* any object whose worldline is a timelike curve parametrised by its proper time τ. In this subsection, we will present the tools necessary to understand what quantities an observer measures locally.

Simultaneity and local rest state

Consider an observer O with 4-velocity $\boldsymbol{u} = u^\mu\frac{\partial}{\partial x^\mu}$ and proper time τ, along a worldline $\mathcal{L}$. In order for O to give a time to events along its worldline it is enough for them to pick a reference time at which they can set $\tau = 0$ and then to count the proper time elapsed from this event to any other event. But what of events outside their worldline? Then in relativity, there is no unique way to say. One way though, is practical because it is operational: it can be done in real experiment. Consider that the observer O, in addition to a clock measuring proper time along their worldline, also has the ability to shoot light rays and to receive them. Consider an event $A \in \mathcal{L}$ along the observer's worldline, at proper time τ, and $P \in \mathcal{M}$ that is *not* on $\mathcal{L}$. At a (proper) time τ_1 along their worldline,

O sends a light ray towards P. Upon receiving it, P reflects (say with a mirror) this light ray towards O, so that it is received at τ_2 along $\mathcal{L}$; see Fig. 3.8. We assume that light rays are not affected by anything but gravitation: they are free-falling. We define:

Einstein's simultaneity

P is said to be *simultaneous* with A iff:

$$\tau = \frac{\tau_1 + \tau_2}{2} \,. \tag{3.89}$$

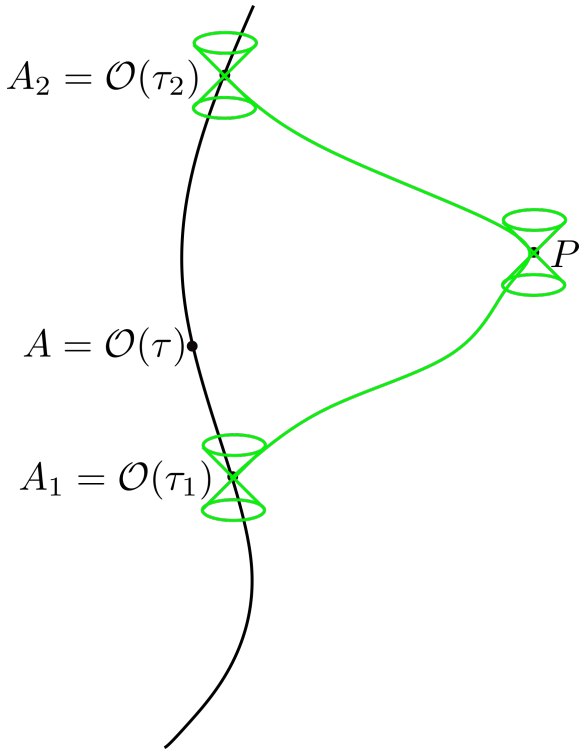


Figure 3.8: The distant event P is connected to O's worldline by two light rays (in green here): one emitted by O at its proper time τ_1, and the other one received by O at τ_2.P is simultaneous to and event A along O's worldline iff the proper time at A is $(\tau_1 + \tau_2)/2$.

This definition is purely local: it depends only on time measurements along the observer's worldline, without involving anything evaluated at the distant event P. The set of all events $P \in \mathcal{M}$ simultaneous with $A \in \mathcal{L}$ forms a hypersurface through A. It is the *simultaneity hypersurface of A*

for O. It is orthogonal to $\mathcal{L}$ and is thus spacelike. Indeed, let us consider that P is infinitesimally close to A so that they can be connected (in $T_A\mathcal{M}$) by infinitesimal displacements. By construction, we have $\tau_1 - \tau = \tau - \tau_2 = \delta\tau$. The infinitesimal displacement between A_1 and A is $\delta\tau\boldsymbol{u}$, and so is the one between A and A_2. Let $\boldsymbol{k}_1$ be the lightlike vector connecting A_1 to P and $\boldsymbol{k}_2$ the one connecting P to A_2. By defining $\boldsymbol{d}$ the vector connecting A to P, we have:

$$\begin{cases} \boldsymbol{k_1} = \delta\tau\boldsymbol{u} + \boldsymbol{d} & (3.90) \\ \boldsymbol{k_2} = \delta\tau\boldsymbol{u} - \boldsymbol{d} \ . & (3.91) \end{cases}$$

Since $\boldsymbol{g}\left(\boldsymbol{k_1}, \boldsymbol{k_1}\right) = \boldsymbol{g}\left(\boldsymbol{k_2}, \boldsymbol{k_2}\right) = 0$, we get:

$$\begin{cases} -(\delta\tau)^2 + \boldsymbol{g}\left(\boldsymbol{d}, \boldsymbol{d}\right) + 2\delta\tau\boldsymbol{g}\left(\boldsymbol{u}, \boldsymbol{d}\right) = 0 & (3.92) \\ -(\delta\tau)^2 + \boldsymbol{g}\left(\boldsymbol{d}, \boldsymbol{d}\right) - 2\delta\tau\boldsymbol{g}\left(\boldsymbol{u}, \boldsymbol{d}\right) = 0 \ , & (3.93) \end{cases}$$

which implies that $\boldsymbol{g}\left(\boldsymbol{u}, \boldsymbol{d}\right) = 0$ and $\boldsymbol{d}$ is thus spacelike. Therefore, the simultaneity hypersurface of A for O is orthogonal to $\boldsymbol{u}$ at A. Vectors tangent to it at A are spacelike. The set of such spacelike vectors orthogonal to $\boldsymbol{u}$ at A is a vector subspace of $T_A\mathcal{M}$ called the *local rest space* of O at A, denoted $\mathcal{R}_O(A)$. Then, the tangent space $T_A\mathcal{M}$ naturally splits into a direct sum:

$$T_A\mathcal{M} = \mathrm{Span}(\boldsymbol{u}) \oplus \mathcal{R}_O(A) \ . \tag{3.94}$$

Given any vector $\boldsymbol{V}$ at A, it can always be decomposed into a part tangent to $\boldsymbol{u}$, $\boldsymbol{V}_{||}$, and a part orthogonal to it, $\boldsymbol{V}_{\perp}$, by using the *orthogonal projector* onto the rest space; see Fig 3.9:

$$\boldsymbol{Pr} = \boldsymbol{Id} + \boldsymbol{u} \otimes \boldsymbol{u}^* \ . \tag{3.95}$$

In terms of components:

$$Pr^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + u^{\mu}u_{\nu} \ . \tag{3.96}$$

You can check that:

$$\begin{cases} \boldsymbol{V}_{\perp} = \boldsymbol{Pr}(\cdot, \boldsymbol{V}) & (3.97) \\ \boldsymbol{V}_{||} = \boldsymbol{V} - \boldsymbol{V}_{\perp} \ . & (3.98) \end{cases}$$

In the same way as $T_A\mathcal{M}$ splits into a direction along $\boldsymbol{u}$ and the rest space according to a direct sum, we can split the dual space and any tensor space, so it makes sense to project one-forms and

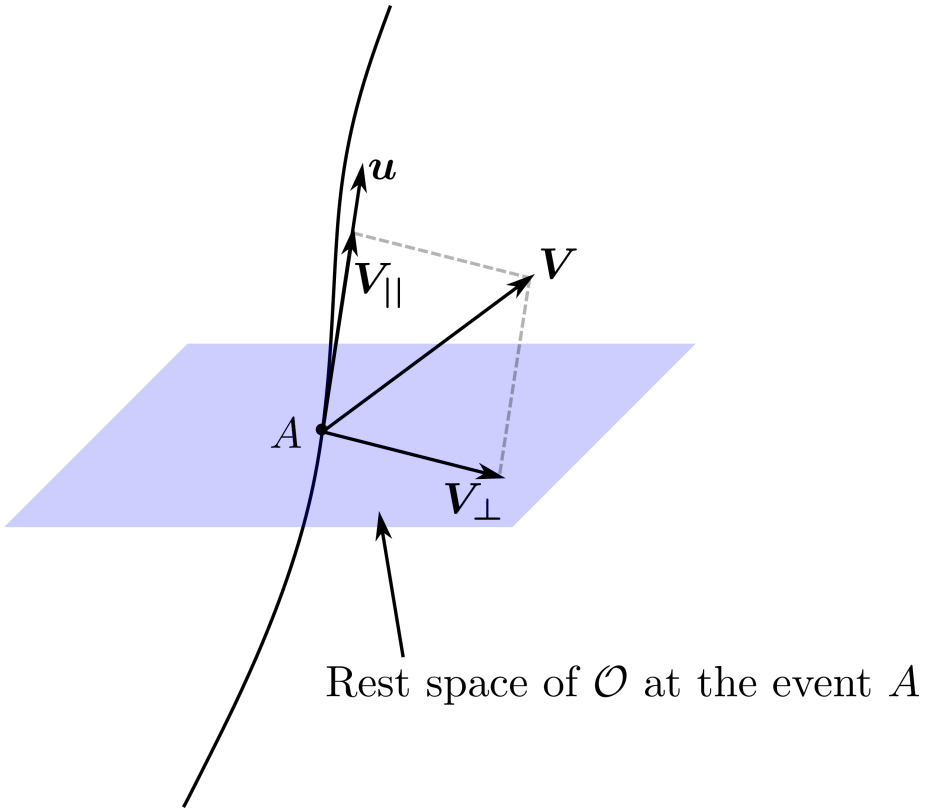


Figure 3.9: Any vector $\boldsymbol{V}$ can be projected onto the rest space of the observer $\mathcal{O}$ at an event A using the projection operator $\boldsymbol{Pr}$ defined in Eq. (3.95).

tensors of any rank along $\boldsymbol{u}$ or orthogonally to it by applying the projection operator and its dual the appropriate number of times. Incidently, we have shown that at any event A along the worldline of an observer with 4-velocity $\boldsymbol{u}$, a lightlike vector $\boldsymbol{k}$, can be decomposed into:

$$\boldsymbol{k} \propto \boldsymbol{u} + \boldsymbol{n} \,, \tag{3.99}$$

where $\boldsymbol{n}$ is spacelike and orthogonal to $\boldsymbol{u}$. This will be important later.

3.5.4 Local Lorentz factor

Consider two observers $\mathcal{O}$ and $\mathcal{O}'$, each following worldlines $\mathcal{L}$ and $\mathcal{L}'$ with 4-velocity $\boldsymbol{u}$ and $\boldsymbol{u'}$, and crossing at an event A. Let τ be the proper time along $\mathcal{L}$ at A and τ' the one along $\mathcal{L}'$ at the

same event. After an infinitesimal interval $\mathrm{d}\tau'$ of its proper time, O' is at the event A' along its worldline. O determine the 'time' of A' in its own frame by the simultaneity procedure described above, and gives it the time $\tau + \mathrm{d}\tau$. The *Lorentz factor* of O' with respect to O, γ, is defined by:

$$\gamma = \frac{\mathrm{d}\tau}{\mathrm{d}\tau'} \ . \tag{3.100}$$

Let B be the event along $\mathcal{L}$ simultaneous to A' according to O; see Fig. 3.10. If we denote by

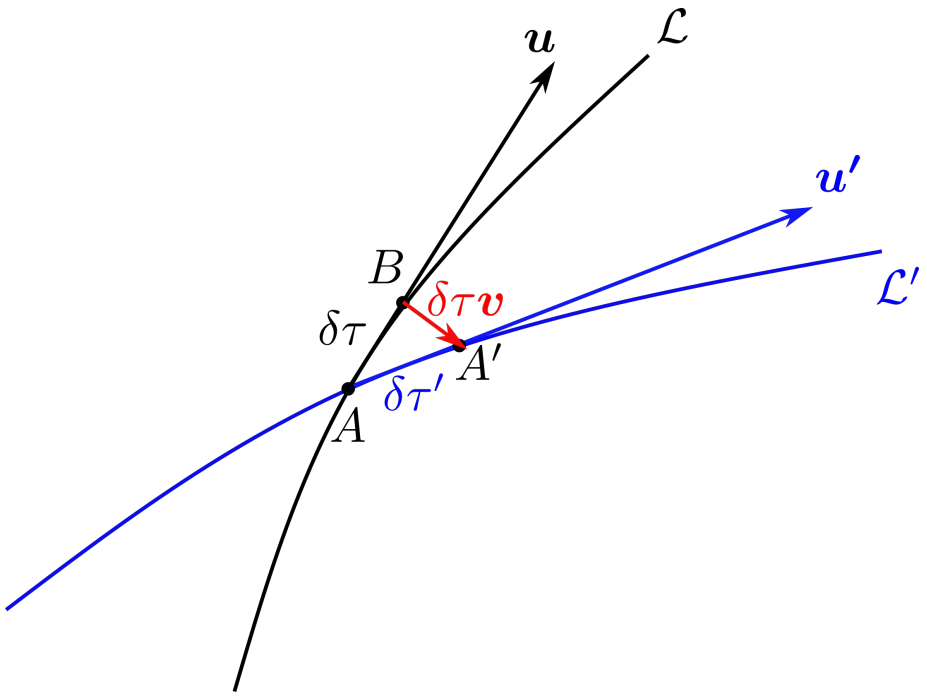


Figure 3.10: Construction to calculate the Lorentz factor between two observers.

$\mathrm{d}\tau\boldsymbol{v}$ the spacelike infinitesimal displacement between B and A', we can write:

$$\mathrm{d}\tau'\boldsymbol{u'} = \mathrm{d}\tau\boldsymbol{u} + \mathrm{d}\tau\boldsymbol{v} \ , \tag{3.101}$$

so that:

$$\boldsymbol{u'} = \gamma\,[\boldsymbol{u} + \boldsymbol{v}] \ . \tag{3.102}$$

Since $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = 0$, we get:

$$\gamma = -\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}') \tag{3.103}$$

$$= \frac{1}{\sqrt{1 - \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})}} \, . \tag{3.104}$$

Note the similarity with the special relativistic formula. Also note that since $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) > 0$, we obtain $\gamma > 1$. This is the general relativistic analogue to the 'time dilation'.

3.5.5 Measurements

Let O be an observer moving along a timelike worldline $\mathcal{L}_O$ with 4-velocity $\boldsymbol{u}$ and proper time τ. A particle, massless or massive, moves along a worldline, timelike of lightlike, $\mathcal{L}$ with 4-momentum $\boldsymbol{k}$ and encounters the observer O at the event C along its worldline. Then, the energy of the particle as measured by O is given by:

$$E = -\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{k}) \ , \tag{3.105}$$

where this quantity must be evaluated at the event C. To see this, one can simply go to the local inertial frame at C and use the special relativistic result. Then, we can define a spacelike vector (it is orthogonal to $\boldsymbol{u}$, which is timelike) living in O's local rest space at C:

$$^{(3)}\boldsymbol{k} = \boldsymbol{k} + \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{k})\, \boldsymbol{u} = \boldsymbol{k} - E\boldsymbol{u} \ , \tag{3.106}$$

called the *particle's* 3*-momentum*.

Massless particles

If the particle is a photon, then $\boldsymbol{g}(\boldsymbol{k}, \boldsymbol{k}) = 0$ and we see that:

$$\boldsymbol{g}\left(^{(3)}\boldsymbol{k}, {}^{(3)}\boldsymbol{k}\right) = E^2 \, . \tag{3.107}$$

We can define a unit spacelike vector $\boldsymbol{n} = {}^{(3)}\boldsymbol{k}/E$ corresponding to the instantaneous direction of propagation of the photon in the observer's rest frame so that:

Decomposition of photon 4-momentum

$$\boldsymbol{k} = E\,[\boldsymbol{u} + \boldsymbol{n}] \ , \tag{3.108}$$

where:

- $\boldsymbol{u}$ is the 4-velocity of the observer;
- $E = -\boldsymbol{g}(\boldsymbol{k}, \boldsymbol{u})$ is the energy of the photon as measured by the observer;
- $\boldsymbol{n}$ is a spacelike unit vector ($\boldsymbol{g}(\boldsymbol{n}, \boldsymbol{n}) = 1$ and $\boldsymbol{g}(\boldsymbol{n}, \boldsymbol{u}) = 0$) corresponding to the direction in which the photon propagates in the observer's rest frame at the event of measurement.

This is pictured in Fig. 3.11.

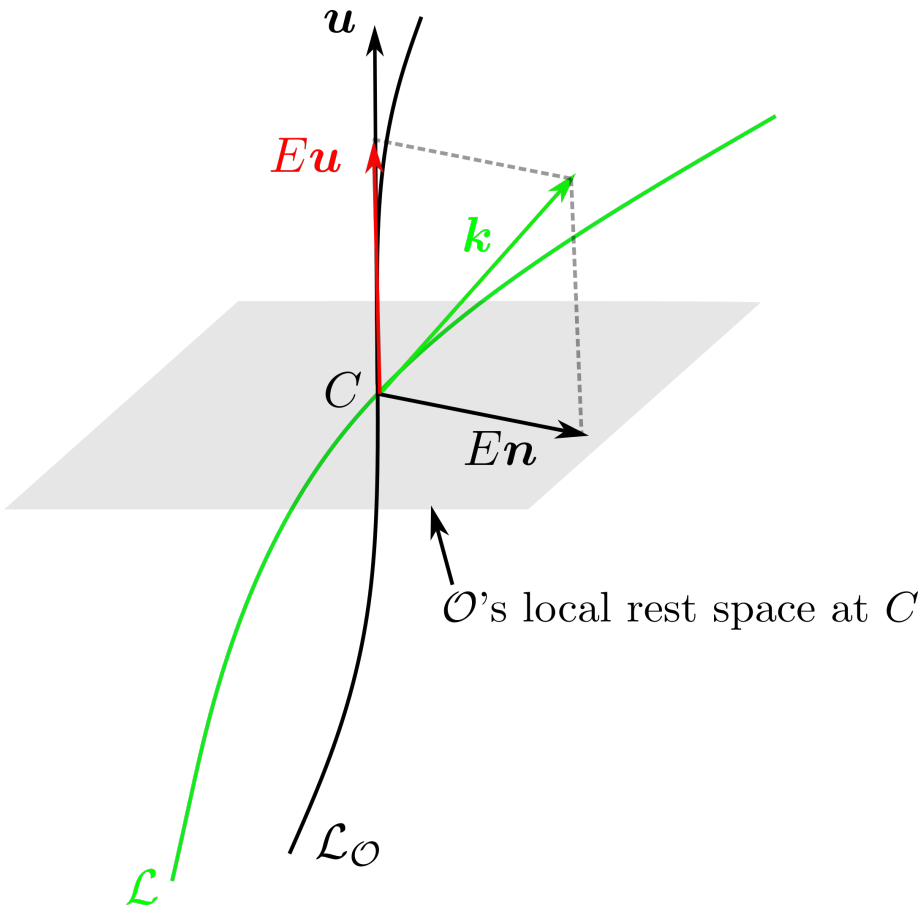


Figure 3.11: The 4-momentum of a photon can be decomposed onto a component along the observer's 4-velocity, giving the observed energy, E, and a component in the observer's local rest space, giving the direction of propagation of the photon at the event of measurement, $\boldsymbol{n}$.

Massive particles

If the particle is massive, of mass m, then $\boldsymbol{k} = m\boldsymbol{U}$ where $\boldsymbol{U}$ is the particle's 4-velocity. Then, the energy is just:

$$E = -m\boldsymbol{g}\left(\boldsymbol{u}, \boldsymbol{U}\right) = m\gamma \ , \tag{3.109}$$

and we recover, formally, Einstein's special relativistic formula. On the other hand, computing $\boldsymbol{g}\left(\boldsymbol{k}, \boldsymbol{k}\right) = -m^2$ in terms of the decomposition (3.106), we get:

$$E^2 = m^2 + \boldsymbol{g}\left({}^{(\boldsymbol{3})}\boldsymbol{k}, {}^{(\boldsymbol{3})}\boldsymbol{k}\right) \ , \tag{3.110}$$

again obtaining a formula formally identical to the special relativistic one. These formal equivalences come from the fact that energy and 3-momentum are purely local quantities. Thus, they could be computed in the local inertial frame and their expressions in an arbitrary coordinate system obtained by a simple change of coordinates. The genuine effects of gravitation will show up when we start taking about non-local quantities.

3.6 Parallel transport, affine connection and the geodesic equation

In this section we develop a way to propagate vectors along curves in spacetime. This leads to the concept of affine connection. We introduce the metric-compatible connection, which provides us with a well-defined notion of derivative along vector fields that generalises partial derivatives to arbitrary coordinate systems. Finally, we obtain the geodesic equation and we show that it allows one to obtain the equations of motions of particles in free-fall.

3.6.1 Parallel transport: a qualitative discussion

We are guided by the fact that two vectors (respectively one-forms, or tensors of any kind) at different points of the manifold cannot be compared directly, since they belong to different tangent spaces (respectively cotangent spaces, etc.). So we are looking for a way to 'glue' tangent spaces together, for a rule that allows one to go from one tangent space to another one associated to a point 'infinitely' closed to the first point. In $\mathbb{R}^n$, with its canonical, Cartesian basis, things are pretty simple. Given a vector field $\boldsymbol{V} = V^i \boldsymbol{e}_{(\boldsymbol{i})}$, we can define its partial derivatives by:

$$\frac{\partial V^i}{\partial x^j} = \lim_{\Delta x^j \to 0} \frac{V^i(x^1, ..., x^{j-1}, x^j + \Delta x^j, x^{j+1}, ..., x^n) - V(x^1, ..., x^n)}{\Delta x^j} \ . \tag{3.111}$$

In the numerator, the first term is defined at the point $x + \Delta x = (x^1, ..., x^{j-1}, x^j + \Delta x^j, x^{j+1}, ..., x^n)$, and the second term at the point $x = (x^1, ..., x^n)$. Therefore we can try to transport $V^i(x+\Delta x)$ to the point x is order to perform the subtraction. Such a method is called *parallel transport*. In this case, which is the one from usual calculus, we simply suppose that $V(x)$ transported to $x + \Delta x$ has the same component as $V(x)$: this follows from the rules defining a vector space. Indeed, in a vector space, vectors are not attached to a point, they can be attached to any point by being drawn parallel to themselves (this is just $(v + a) - a = v$). But for a manifold, such rules are absent and one is free to specify what is meant by parallel transporting vectors. So let us pick up a manifold $\mathcal{M}$, and $\boldsymbol{V}$ a vector field on this manifold, as well as a local chart $\{x^i\}$. Let us write $\tilde{\boldsymbol{V}}(x + \Delta x)$ the result of the parallel transport of the vector $\boldsymbol{V}(x)$ from x to $x + \Delta x$. The first rules we impose are the following:

1. $\forall i \in \{1, ..., n\}$, $\tilde{V}^i(x + \Delta x) - V^i(x) \propto \Delta x$;
2. $\forall i \in \{1, ..., n\}$, $\left(\widetilde{V^i + W^i}\right)(x + \Delta x) = \tilde{V}^i(x + \Delta x) + \tilde{W}^i(x + \Delta x)$, where $\boldsymbol{W}$ is another vector field.

The first condition just expresses the fact that the change has to be infinitesimal for an infinitesimal displacement. The second condition tells us that the parallel transport is a linear operation. Using the first condition, we see that we can write:

$$\tilde{V}^i(x + \Delta x) = V^i(x) + A^i{}_j\,(x, V(x))\,\Delta x^j. \tag{3.112}$$

The $A^i{}_j$'s are assumed to depend on V and x, a priori, because nothing prevents it. If we use the second condition we then get:

$$A^i{}_j(x, (V + W)(x)) = A^i{}_j(x, V(x)) + A^i{}_j(x, W(x))\ , \tag{3.113}$$

so they must be linear functions of the components of the vector:

$$A^i{}_j(x, V(x)) = -\Gamma^i{}_{jk}(x)V^k(x)\ . \tag{3.114}$$

The $\Gamma^i{}_{jk}$ thus defined are the *connection coefficients*. Using them, we have that:

$$\tilde{V}^i(x + \Delta x) = V^i(x) - \Gamma^i{}_{jk}(x)\Delta x^j V^k(x)\ . \tag{3.115}$$

Then, by analogy with the derivatives in $\mathbb{R}^n$, we can define the *covariant derivative* of $\boldsymbol{V}$ with respect to x^j, denoted $\nabla_j \boldsymbol{V}$, by:

$$\nabla_j \boldsymbol{V} = \lim_{\Delta x^j \to 0} \frac{V^i(x+\Delta x) - \tilde{V}^i(x+\Delta x)}{\Delta x^j} \frac{\partial}{\partial x^i} \tag{3.116}$$

$$= \left(\frac{\partial V^i}{\partial x^j} + \Gamma^i{}_{jk} V^k \right) \frac{\partial}{\partial x^i} \ . \tag{3.117}$$

As one can see, it is a vector, since it is expressed as a linear combination of the coordinate basis vectors of $T_{x+\Delta x}\mathcal{M}$. Of course, for $\boldsymbol{V}$ fixed, one can construct a one-form $\boldsymbol{\nabla V} = \nabla_j \boldsymbol{V} \mathrm{d}x^j$ whose components are the covariant derivatives with respect to the x^i's. The case of the vector space $\mathbb{R}^n$ with Cartesian coordinates described above just corresponds to $\Gamma^i{}_{jk} = 0$ for any (i, j, k). Beware that the connection coefficients cannot follow the laws of transformation of tensor, as the example below show.

Let us consider the Euclidean space in 2 dimensions, E^2. We define the parallel transport in the usual sense (case discussed above for $\mathbb{R}^n$): $\tilde{\boldsymbol{V}}(x+\Delta x, y+\Delta y) = \boldsymbol{V}(x,y)$. In Cartesian coordinates, all the connections are identically zero. If we now switch to polar coordinates (r, θ) via

$$\psi^{-1} : \begin{cases} \mathbb{R}_+ \times [0, 2\pi[& \to & \mathbb{R} \times \mathbb{R} \\ (r, \theta) & \mapsto & (x, y) = (r\cos\theta, r\sin\theta) \end{cases} , \tag{3.118}$$

we can write the vector $\boldsymbol{V}$ as:

$$V = V^x \boldsymbol{e_x} + V^y \boldsymbol{e_y} = V^r \boldsymbol{e_r} + V^\theta \boldsymbol{e_\theta} \ , \tag{3.119}$$

where $\{\boldsymbol{e_x}, \boldsymbol{e_y}\} = \left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right\}$ is the Cartesian basis for vectors of E^2, and $\{\boldsymbol{e_r}, \boldsymbol{e_\theta}\} = \left\{\frac{\partial}{\partial r}, \frac{\partial}{\partial \theta}\right\}$ is the polar canonical basis. Figure 3.12 represents the parallel transport of a vector of the plane viewed in polar coordinates.

We know that:

$$\begin{cases} \boldsymbol{e_r}(r,\theta) = \cos\theta \boldsymbol{e_x} + \sin\theta \boldsymbol{e_y} & (3.120) \\ \boldsymbol{e_\theta}(r,\theta) = -r\sin\theta \boldsymbol{e_x} + r\cos\theta \boldsymbol{e_y} \ . & (3.121) \end{cases}$$

Therefore, at first order in the small displacements:

$$\begin{cases} \boldsymbol{e_r}(r+\Delta r, \theta+\Delta\theta) = \boldsymbol{e_r}(r,\theta) + \dfrac{\Delta\theta}{r} \boldsymbol{e_\theta}(r,\theta) & (3.122) \\ \boldsymbol{e_\theta}(r+\Delta r, \theta+\Delta\theta) = \boldsymbol{e_\theta}(r,\theta) - r\Delta\theta \boldsymbol{e_r}(r,\theta) + \dfrac{\Delta r}{r} \boldsymbol{e_\theta}(r,\theta) \ . & (3.123) \end{cases}$$

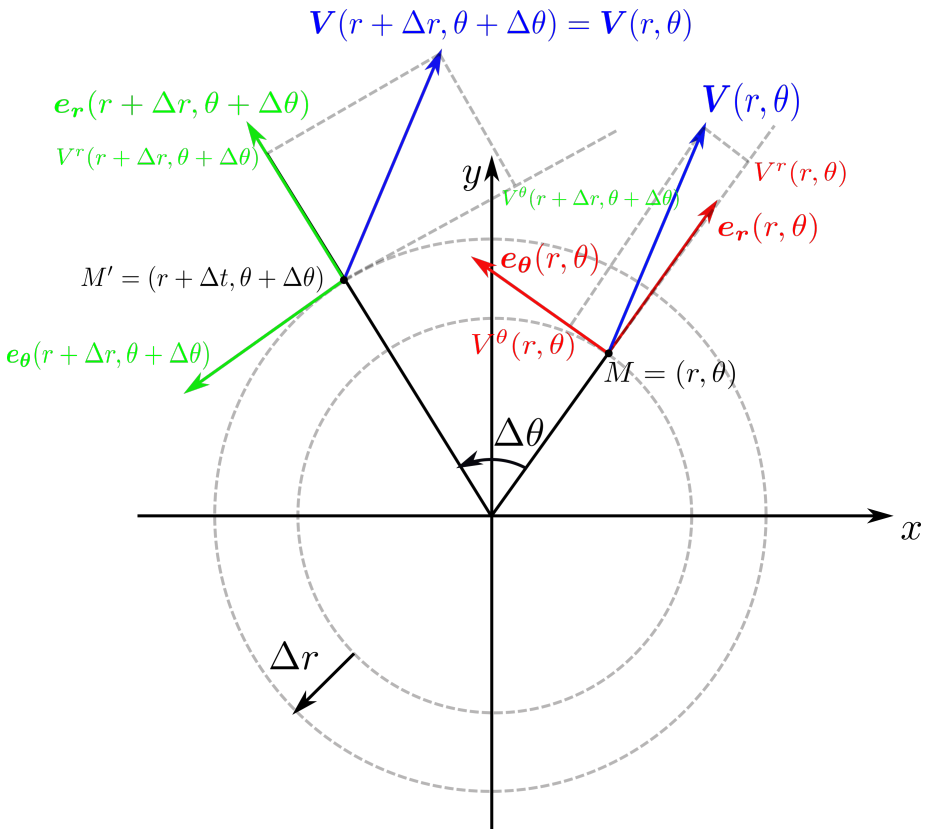


Figure 3.12: The standard parallel transport of the vector $\boldsymbol{V}(M)$ (here in blue) in the plane, to a point M': $\boldsymbol{V}(M') = \boldsymbol{V}(M)$, Clearly the components of the vector in the moving basis are altered by the transport. When Δr and $\Delta\theta$ are infinitesimal, the changes are encapsulated in the connection coefficients.

Then, since we said that: $\tilde{\boldsymbol{V}}(x+\Delta x, y+\Delta y) = \boldsymbol{V}(x,y)$, we must also have: $\tilde{\boldsymbol{V}}(r+\Delta r,\theta+\Delta\theta) = \boldsymbol{V}(r,\theta)$ (properties of vectors don't depend on the coordinate system chosen). So, writing $\boldsymbol{V}(r,\theta) = V^r\boldsymbol{e_r}(r,\theta) + V^\theta\boldsymbol{e_\theta}(r,\theta)$, after a bit of algebra, we find that:

$$\begin{cases} \tilde{V}^r = V^r + rV^\theta\Delta\theta & (3.124) \\ \tilde{V}^\theta = V^\theta - \dfrac{V^\theta}{r}\Delta r - \dfrac{V^r}{r}\Delta\theta\ . & (3.125) \end{cases}$$

Therefore, by coming back to the definition of the connection coefficients, we have that:

$$\begin{cases} \Gamma^r{}_{rr} = \Gamma^r_{r\theta} = \Gamma^r{}_{\theta r} = \Gamma^\theta{}_{\theta\theta} = 0 & (3.126) \\ \Gamma^r{}_{\theta\theta} = -r \text{ and } \Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \dfrac{1}{r} \,. & (3.127) \end{cases}$$

A few remarks:

- There is nothing physical in connection coefficients since they can be made to appear or vanish by a simple change of coordinates.
- It is clear that connection coefficients are not the components of a tensor: they are all zero in Cartesian coordinates, but not in polar coordinates.
- The resulting connection coefficients are symmetric in their lower indices: $\Gamma^i{}_{jk} = \Gamma^i{}_{kj}$;
- In defining this particular case of parallel transport, we have conserved the direction of the vector but also its norm.

Connection coefficients that verify the last two points define a *Levi-Civita connection*.

3.6.2 The affine connection

Definition

Now that we have studied connections 'by-hand', we are ready to give a general definition, As previously, we focus on the four dimensional case of interest in General Relativity and we use the relativistic notations.

Affine connection

If we denote by $\mathcal{X}(\mathcal{M})$ the set of vector fields on spacetime $\mathcal{M}$, an *affine connection* on $\mathcal{M}$ is a map:

$$\boldsymbol{\nabla} : \begin{cases} \mathcal{X}(\mathcal{M}) \times \mathcal{X}(\mathcal{M}) & \to & \mathcal{X}(\mathcal{M}) \\ (\boldsymbol{X}, \boldsymbol{Y}) & \mapsto & \boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{Y} \end{cases} , \tag{3.128}$$

such that, for all $(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) \in \mathcal{X}(\mathcal{M})^3$ and all $f \in \mathcal{F}(\mathcal{M})$:

(i) $\boldsymbol{\nabla}_{\boldsymbol{X}}(\boldsymbol{Y} + \boldsymbol{Z}) = \boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{Y} + \boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{Z}$;

(ii) $\boldsymbol{\nabla}_{(\boldsymbol{X}+\boldsymbol{Y})} \boldsymbol{Z} = \boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{Z} + \boldsymbol{\nabla}_{\boldsymbol{Y}} \boldsymbol{Z}$;

(iii) $\boldsymbol{\nabla}_{(f\boldsymbol{X})}\boldsymbol{Y} = f\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y}$;

(iv) $\boldsymbol{\nabla}_{\boldsymbol{X}}(f\boldsymbol{Y}) = \boldsymbol{X}[f]\boldsymbol{Y} + f\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y}$.

(i) and (ii) mean that the affine connection is a bilinear map, while (iii) and (iv) spell its properties as a differential operator. Of course, we can look at the effect of an affine connection on the vectors of a coordinate basis, once a local chart has been chosen. Let (U, ϕ) be a local chart around $p \in \mathcal{M}$ such that $\{x^\mu\} = \phi(p)$. Given an affine connections $\boldsymbol{\nabla}$ on $\mathcal{M}$, we define the $4^3 = (\dim \mathcal{M})^3 = 64$ functions $\Gamma^\rho{}_{\mu\nu}$, called *connection coefficients* by:

$$\boldsymbol{\nabla}_{\boldsymbol{e}_{(\mu)}}\boldsymbol{e}_{(\nu)} = \Gamma^\rho{}_{\mu\nu}\boldsymbol{e}_{(\rho)} \ , \tag{3.129}$$

where $\{\boldsymbol{e}_{(\mu)}\} = \{\frac{\partial}{\partial x^\mu}\}$ is the canonical basis of $T_p\mathcal{M}$ associated with the local coordinates $\{x^\mu\}$. Usually, one denotes $\boldsymbol{\nabla}_{\boldsymbol{e}_{(\mu)}} = \boldsymbol{\nabla}_\mu$, and this operator is called the *covariant derivative* associated with the given affine connection. This terminology will become clear a bit later. The connection coefficients specify how the basis vectors of the tangent spaces change from one point of the manifold to another when they are parallel transported. This is exactly the same as what we described for the polar basis above. Once we have fixed the action of the connection on the basis, we can calculate its action on any vector field. Let $\boldsymbol{X} = X^\mu\boldsymbol{e}_{(\mu)}$ and $\boldsymbol{Y} = Y^\mu\boldsymbol{e}_{(\mu)}$ be two vector fields. Then, according to the properties of the connection:

$$\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y} = X^\mu\boldsymbol{\nabla}_{\boldsymbol{e}_{(\mu)}}\left(Y^\nu\boldsymbol{e}_{(\nu)}\right) = X^\mu\left(\boldsymbol{e}_{(\mu)}[Y^\nu]\boldsymbol{e}_{(\nu)} + Y^\nu\boldsymbol{\nabla}_{\boldsymbol{e}_{(\mu)}}\boldsymbol{e}_{(\nu)}\right) \tag{3.130}$$

$$= X^\mu\left(\frac{\partial Y^\nu}{\partial x^\mu} + \Gamma^\nu{}_{\mu\rho}Y^\rho\right)e_\nu \ . \tag{3.131}$$

Note that we recover the results of our 'hand-wavy' argument above. By definition, $\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y}$ is a vector field with components equal to the RHS of the previous equation: $\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{Y} = (\nabla_X Y)^\mu\,\boldsymbol{e}_{(\mu)}$. This is the *covariant derivative* of $\boldsymbol{Y}$ along $\boldsymbol{X}$.The components of the covariant derivative of $\boldsymbol{Y}$ along $\boldsymbol{X}$ can be read off Eq. (3.131):

$$(\nabla_X Y)^\nu = X^\mu\left(\nabla_\mu Y\right)^\nu = X^\mu\left(\frac{\partial Y^\nu}{\partial x^\mu} + \Gamma^\nu{}_{\mu\rho}Y^\rho\right) \ . \tag{3.132}$$

Note that in General Relativity, the components of the covariant derivative of a vector field are usually denoted:

$$\nabla_\mu X^\nu = \frac{\partial Y^\nu}{\partial x^\mu} + \Gamma^\nu{}_{\mu\rho}Y^\rho \ , \tag{3.133}$$

instead of the more rigorous $\left(\nabla_\mu X\right)^\nu$. We will make use of both notations.

Action of the connection on arbitrary objects

Once a connection has been chosen, via its action on coordinate basis vector fields of $T\mathcal{M}$, we can readily generalise its action on arbitrary geometric objects such as functions, one-form fields and tensor fields. To simplify expressions, It is customary to keep the same notation for the action on every object. Let us fix a vector field $\boldsymbol{X} \in \mathcal{X}(\mathcal{M})$, and define the action of a connection $\boldsymbol{\nabla_X}$ on the various tensorial objects.

For functions $f \in \mathcal{F}(\mathcal{M})$, the connection is just the directional derivative:

$$\boldsymbol{\nabla_X} : \left\{ \begin{array}{ccc} \mathcal{F}(\mathcal{M}) & \to & \mathcal{F}(\mathcal{M}) \\ f & \mapsto & \boldsymbol{\nabla_X} f = \boldsymbol{X}[f] \end{array} \right. . \tag{3.134}$$

For general tensors, we need to impose a Leibniz rule to ensure that the connection remains a derivative operator:

$$\boldsymbol{\nabla_X} \left(\boldsymbol{T_1} \otimes \boldsymbol{T_2}\right) = \left(\boldsymbol{\nabla_X} \boldsymbol{T_1}\right) \otimes \boldsymbol{T_2} + \boldsymbol{T_1} \otimes \left(\boldsymbol{\nabla_X} \boldsymbol{T_2}\right) \ , \tag{3.135}$$

where $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ are tensor fields of arbitrary orders. For example, for a one-form field $\boldsymbol{\omega}$, we define the connection as:

$$\boldsymbol{\nabla_X} : \left\{ \begin{array}{ccc} \Omega(\mathcal{M}) & \to & \Omega(\mathcal{M}) \\ \boldsymbol{\omega} & \mapsto & \boldsymbol{\nabla_X} \boldsymbol{\omega} \end{array} \right. . \tag{3.136}$$

For any vector field $\boldsymbol{Y}$, we have $\boldsymbol{\omega}(\boldsymbol{Y}) \in \mathcal{F}(\mathcal{M})$, and therefore, we know how to a[ply the connection to it:

$$\boldsymbol{\nabla_X} \left(\boldsymbol{\omega}(\boldsymbol{Y})\right) = \boldsymbol{X}\left[\boldsymbol{\omega}(\boldsymbol{Y})\right] = \boldsymbol{X}[\omega_\mu Y^\mu] \ . \tag{3.137}$$

Besides, if we apply the Leibniz rule above, we must have:

$$\boldsymbol{\nabla_X} \left(\boldsymbol{\omega}(\boldsymbol{Y})\right) = \left(\boldsymbol{\nabla_X} \boldsymbol{\omega}\right)(\boldsymbol{Y}) + \boldsymbol{\omega}\left(\boldsymbol{\nabla_X} \boldsymbol{Y}\right) \ . \tag{3.138}$$

Substituting for $\boldsymbol{\nabla_X Y}$ and for $\boldsymbol{X}(\omega_\mu Y^\mu)$, and equating these two relations, we find that the components of the one-form $\boldsymbol{\nabla_X \omega} = (\nabla_X \omega)_\mu dx^\mu$ verify:

$$\boxed{\left(\nabla_X \omega\right)_\mu = X^\nu \left(\nabla_\nu \omega\right)_\mu = X^\nu \left(\frac{\partial \omega_\mu}{\partial x^\nu} - \Gamma^\rho{}_{\nu\mu} \omega_\rho\right) \ .} \tag{3.139}$$

Note that in General Relativity, as for vectors, the components of the covariant derivative of a one-form field are usually denoted:

$$\nabla_\mu \omega_\nu = \frac{\partial \omega_\mu}{\partial x^\nu} - \Gamma^\rho{}_{\nu\mu}\omega_\rho \ , \tag{3.140}$$

instead of the more rigorous $\left(\nabla_\mu \omega\right)_\nu$. We will make use of both notations.
For $\boldsymbol{X} = \boldsymbol{e}_{(\nu)}$:

$$(\nabla_{e_{(\nu)}}\omega)_\mu = \frac{\partial \omega_\mu}{\partial x^\nu} - \Gamma^\rho{}_{\nu\mu}\omega_\rho \ . \tag{3.141}$$

Further, if $\boldsymbol{\omega} = \mathrm{d}x^i$, then:

$$\nabla_\nu \mathrm{d}x^\mu = -\Gamma^\mu{}_{\nu\rho}\mathrm{d}x^\rho \ . \tag{3.142}$$

This generalises to tensors of arbitrary types in the same way. We define the connection on tensors of type (r, s) by:

$$\boldsymbol{\nabla}_{\boldsymbol{X}} : \begin{cases} \mathcal{T}^r_s(\mathcal{M}) & \to & \mathcal{T}^r_s(\mathcal{M}) \\ \boldsymbol{T} & \mapsto & \boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{T} \end{cases} \ . \tag{3.143}$$

Using the connection on vectors and one-forms, we then obtain a generalised Leibniz rule[8]:

$$\begin{aligned} (\boldsymbol{\nabla}_{\boldsymbol{Y}}\boldsymbol{T})\,(\boldsymbol{\omega_1}, \cdots, \boldsymbol{\omega_r}, \boldsymbol{X_1}, \cdots, \boldsymbol{X_s}) = &\boldsymbol{Y}\,[\boldsymbol{T}\,(\boldsymbol{\omega_1}, \cdots, \boldsymbol{\omega_r}, \boldsymbol{X_1}, \cdots, \boldsymbol{X_s})] \\ &- \boldsymbol{T}\,(\boldsymbol{\nabla_Y}\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_r}, \boldsymbol{X_1}, \cdots, \boldsymbol{X_s}) \\ &- \cdots - \boldsymbol{T}\,(\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\nabla_Y}\boldsymbol{\omega_r}, \boldsymbol{X_1}, \cdots, \boldsymbol{X_s}) \\ &- \boldsymbol{T}\,(\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_r}, \boldsymbol{\nabla_Y}\boldsymbol{X_1}, \cdots, \boldsymbol{X_s}) \\ &- \cdots - \boldsymbol{T}\,(\boldsymbol{\omega_1}, \boldsymbol{\omega_2}, \cdots, \boldsymbol{\omega_r}, \boldsymbol{X_1}, \cdots, \boldsymbol{\nabla_Y}\boldsymbol{X_s}) \ . \end{aligned} \tag{3.144}$$

In terms of components, this gives:

$$\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{T} = X^\mu \boldsymbol{\nabla}_\mu \boldsymbol{T} = X^\mu \left(\boldsymbol{\nabla}_\mu \boldsymbol{T}\right)^{\nu_1\cdots\nu_r}{}_{\rho_1\cdots\rho_s}\boldsymbol{e}_{(\boldsymbol{\nu_1})} \otimes \cdots \otimes \boldsymbol{e}_{(\boldsymbol{\nu_r})} \otimes \mathrm{d}x^{\rho_1} \otimes \cdots \otimes \mathrm{d}x^{\rho_s} \ , \tag{3.145}$$

where:

[8] To convince yourself of this rule, do the calculation for a simple tensor, say of type $(0, 2)$.

$$\begin{aligned}\left(\nabla_\mu T\right)^{\nu_1\cdots\nu_r}{}_{\rho_1\cdots\rho_s} =& \frac{\partial T^{\nu_1\cdots\nu_r}{}_{\rho_1\cdots\rho_s}}{\partial x^\mu} \\ &+ \Gamma^{\nu_1}{}_{\mu\lambda} T^{\lambda\nu_2\cdots\nu_r}{}_{\rho_1\cdots\rho_s} + \cdots + \Gamma^{\nu_r}{}_{\mu\lambda} T^{\nu_1\cdots\nu^{r-1}\lambda}{}_{\rho_1\cdots\rho_s} \\ &- \Gamma^{\lambda}{}_{\mu\rho_1} T^{\nu_1\cdots\nu^r}{}_{\lambda\rho_2\cdots\rho_s} - \cdots - \Gamma^{\lambda}{}_{\mu\rho_s} T^{\nu_1\cdots\nu^r}{}_{\rho_1\cdots\rho_{s-1}\lambda} \,.\end{aligned} \tag{3.146}$$

We will often denote $\left(\nabla_\mu T\right)^{\nu_1\cdots\nu_r}{}_{\rho_1\cdots\rho_s} = \nabla_\mu T^{\nu_1\cdots\nu_r}{}_{\rho_1\cdots\rho_s}$.

Transformation of the connection coefficients

Let $p \in \mathcal{M}$ and (U, ϕ) and (V, φ) be two local charts such that $p \in U \cap V$. Let $x = \{x^\mu\} = \phi(p)$ and $\tilde{x} = \{\tilde{x}^\mu\} = \varphi(p)$. We also denote by $\left\{\boldsymbol{e}_{(\mu)}\right\} = \left\{\frac{\partial}{\partial x^\mu}\right\}$ and $\left\{\tilde{\boldsymbol{e}}_{(\mu)}\right\} = \left\{\frac{\partial}{\partial \tilde{x}^\mu}\right\}$ the two coordinate bases of $T_p\mathcal{M}$ associated with the coordinate functions ϕ and φ, respectively. Finally, for a fixed connection, we note Γ the connection coefficients in the coordinate system given by ϕ and $\tilde{\Gamma}$ these coefficients in the coordinate system given by φ. Then, we have:

$$\boldsymbol{\nabla}_{\tilde{\boldsymbol{e}}_{(\mu)}} \tilde{\boldsymbol{e}}_{(\nu)} = \tilde{\Gamma}^{\rho}{}_{\mu\nu} \tilde{\boldsymbol{e}}_{(\rho)} \,. \tag{3.147}$$

Besides:

$$\tilde{\boldsymbol{e}}_{(\mu)} = \frac{\partial x^\nu}{\partial \tilde{x}^\mu} \boldsymbol{e}_{(\nu)} \,, \tag{3.148}$$

so that, by taking the connection of this expression, we get:

$$\boldsymbol{\nabla}_{\tilde{\boldsymbol{e}}_{(\mu)}} \tilde{\boldsymbol{e}}_{(\nu)} = \boldsymbol{\nabla}_{\tilde{\boldsymbol{e}}_{(\mu)}} \left(\frac{\partial x^\rho}{\partial \tilde{x}^\nu} \boldsymbol{e}_{(\rho)} \right) \tag{3.149}$$

$$= \frac{\partial^2 x^\rho}{\partial \tilde{x}^\mu \partial \tilde{x}^\nu} \boldsymbol{e}_{(\rho)} + \frac{\partial x^\lambda}{\partial x^\mu} \frac{\partial x^\rho}{\partial x^\nu} \boldsymbol{\nabla}_\lambda \boldsymbol{e}_{(\rho)} \tag{3.150}$$

$$= \left(\frac{\partial^2 x^\rho}{\partial \tilde{x}^\mu \partial \tilde{x^\nu}} + \frac{\partial x^\lambda}{\partial \tilde{x}^\mu} \frac{\partial x^\sigma}{\partial \tilde{x}^\nu} \Gamma^{\rho}{}_{\lambda\sigma} \right) \boldsymbol{e}_{(\rho)} \,. \tag{3.151}$$

Hence, by comparing both expressions for $\boldsymbol{\nabla}_{\tilde{\boldsymbol{e}}_{(\mu)}} \tilde{\boldsymbol{e}}_{(\nu)}$, we find that, under a coordinate change, the connection coefficients transform as:

$$\tilde{\Gamma}^{\mu}{}_{\nu\rho} = \frac{\partial x^\lambda}{\partial \tilde{x}^\nu} \frac{\partial x^\sigma}{\partial \tilde{x}^\rho} \frac{\partial \tilde{x}^\mu}{\partial x^\delta} \Gamma^{\delta}{}_{\lambda\sigma} + \frac{\partial \tilde{x}^\mu}{\partial x^\sigma} \frac{\partial^2 x^\sigma}{\partial \tilde{x}^\nu \partial \tilde{x}^\rho} \,. \tag{3.152}$$

It is apparent that this is not the way tensors transforms, because of the second term. Therefore, despite the fact that the connection coefficients carry indices, *they are not the components of any tensor*.

3.6.3 Parallel transport and the geodesic equation

Let us now define the parallel transport of a vector along a curve.

Parallel transport

Let $c :]a, b[\subseteq \mathbb{R} \rightarrow U \subset \mathcal{M}$ be a parametrised curve on an open subset U of $\mathcal{M}$. Let ϕ be a coordinate chart on U. Let us note $x(c(\lambda)) = \{x^\mu(c(\lambda))\} = \phi(c(\lambda))$ for $\lambda \in]a, b[$. Let $\boldsymbol{X} \in \mathcal{X}(\mathcal{M})$ be an arbitrary vector field defined along the curve, such that:

$$\boldsymbol{X}(\boldsymbol{c}(\boldsymbol{\lambda})) = X^\mu(c(\lambda))\boldsymbol{e}_{(\mu)} \ . \tag{3.153}$$

Let:

$$\boldsymbol{V} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\boldsymbol{e}_{(\mu)} \tag{3.154}$$

be the vector tangent to the curve and associated with the parameter λ. Then, if:

$$\forall \lambda \in]a, b[, \boldsymbol{\nabla}_{\boldsymbol{V}} \boldsymbol{X} = 0 \ , \tag{3.155}$$

we say that $\boldsymbol{X}$ is *parallel transported* along the curve c in the open set U. Component-wise, this reads:

$$\frac{\mathrm{d}X^\mu}{\mathrm{d}\lambda} + \Gamma^\mu_{\nu\rho} V^\nu X^\rho = 0 \ , \tag{3.156}$$

or, equivalently:

$$V^\nu \frac{\partial X^\mu}{\partial x^\nu} + \Gamma^\mu_{\nu\rho} V^\nu X^\rho = 0 \ . \tag{3.157}$$

Note that, in the equation for parallel transport (3.157), although each separate piece of the LHS does not transform as a tensor under a coordinate transformation, the overall LHS does: the connection coefficients transform exactly how they should to cancel out the changes in the directional derivative $V^\nu \frac{\partial X^\mu}{\partial x^\nu}$. Once we have the notion of parallel transport, we can define the geodesics of the manifold with connection:

Geodesics

Given a parametrised curve $c :]a, b[\subseteq \mathbb{R} \rightarrow \mathcal{M}$, with parameter λ and associated tangent vector field $\boldsymbol{V}$, if $\boldsymbol{V}$ is parallel transported along c, i.e., if:

$$\boldsymbol{\nabla}_{\boldsymbol{V}} \boldsymbol{V} = 0 \,, \tag{3.158}$$

we say that c is a *geodesic*.

In terms of components, a geodesic is thus characterised by:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu{}_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\rho}{\mathrm{d}\lambda} = 0 \,. \tag{3.159}$$

Geodesics are the generalisation, to manifolds with connection, of the notion of straight line in E^n with the standard connection. Which means that, in a way, they are the 'most direct' path from on point of the manifold to another. But their main definition is that if a curve is a geodesic, along its own flow, the tangent vector does not 'change direction' (with respect to the given connection): it is transported from tangent space to tangent space while remaining 'the same'. Therefore, one could object that the condition $\boldsymbol{\nabla}_{\boldsymbol{V}} \boldsymbol{V} = 0$ is too restrictive. Indeed, if we choose the weaker condition:

$$\exists f \in \mathcal{F}(\mathcal{M}), \boldsymbol{\nabla}_{\boldsymbol{V}} \boldsymbol{V} = f \boldsymbol{V} \,, \tag{3.160}$$

then, the variation of the tangent vector remains parallel to the tangent vector, which is also fine for defining parallelism. Nevertheless, in that case, under a reparametrisation of the curve $\lambda \rightarrow \lambda'$, we get:

$$\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda'} \frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda} \,, \tag{3.161}$$

and

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} = \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda'^2} \left(\frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda} \right)^2 + \frac{\mathrm{d}^2 \lambda'}{\mathrm{d}\lambda^2} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda'} \,. \tag{3.162}$$

Therefore, we see that, by choosing the reparametrisation such that:

$$\frac{\mathrm{d}^2 \lambda'}{\mathrm{d}\lambda^2} = f(c(\lambda)) \frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda} \,, \tag{3.163}$$

with $\mathrm{d}\lambda'/\mathrm{d}\lambda \neq 0$, we can always reparametrise the curve in order for the weaker condition to reduce to $\boldsymbol{\nabla}_{\boldsymbol{V}} \boldsymbol{V} = 0$. This means that our definition of geodesics is very robust.

The connection of General Relativity

The affine connection we have defined so far is too general for our purpose. Because we have a metric on our spacetime, we can restrict it further to suit our needs by imposing two conditions:

- First, we will require that the connection be *metric compatible*, i.e. that when we parallel transport two vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ along a third one, $\boldsymbol{V}$, the scalar product $\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})$ remains unchanged:

$$\boldsymbol{\nabla}_{\boldsymbol{V}}\left[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})\right]=0\,. \tag{3.164}$$

  Geometrically, this ensures that 'lengths' and 'angles' are preserved when vectors are parallel transported.
  Then, we impose that:

$$\begin{aligned}0=&\boldsymbol{\nabla}_{\boldsymbol{V}}\left[\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})\right]=V^{\rho}\left[\left(\boldsymbol{\nabla}_{\rho}\boldsymbol{g}\right)(\boldsymbol{X},\boldsymbol{Y})+\boldsymbol{g}\left(\boldsymbol{\nabla}_{\rho}\boldsymbol{X},\boldsymbol{Y}\right)+\boldsymbol{g}\left(\boldsymbol{X},\boldsymbol{\nabla}_{\rho}\boldsymbol{Y}\right)\right] &(3.165)\\ =&V^{\rho}X^{\mu}Y^{\nu}\left(\boldsymbol{\nabla}_{\rho}\boldsymbol{g}\right)_{\mu\nu}\,, &(3.166)\end{aligned}$$

  where we set $V^{\mu}\boldsymbol{\nabla}_{\mu}\boldsymbol{X}=V^{\mu}\boldsymbol{\nabla}_{\mu}\boldsymbol{Y}=0$ according to the fact that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are parallel transported along $\boldsymbol{V}$. Since this relation must hold for any vectors X, Y and V, we get:

Metric connection

An affine connection $\boldsymbol{\nabla}$ is said to be a *metric connection*, or *compatible with the metric* $\boldsymbol{g}$ iff, in a local chart:

$$\forall(\mu,\nu,\rho)\in\{0,1,2,3\}^3,\ \left(\nabla_{\rho}g\right)_{\mu\nu}=0\,. \tag{3.167}$$

  Equivalently, using the general rule for the covariant derivative of tensor component, this last condition can be written as:

$$\frac{\partial g_{\mu\nu}}{\partial x^{\rho}}-\Gamma^{\lambda}{}_{\rho\mu}g_{\lambda\nu}-\Gamma^{\lambda}{}_{\rho\nu}g_{\lambda\mu}=0\,. \tag{3.168}$$

  Then, starting from this relation, performing permutations of all the indices and combining the results, we get that the connection coefficients of a metric connection are given by:

$$\Gamma^{\mu}{}_{\nu\rho}=\left\{\begin{array}{c}\mu\\ \nu\rho\end{array}\right\}+K^{\mu}{}_{\nu\rho}\,, \tag{3.169}$$

where the $\left\{ \begin{array}{c} \mu \\ \nu\rho \end{array} \right\}$ are symmetric in their lower indices and called the *Christoffel symbols*, given by:

$$\left\{ \begin{array}{c} \mu \\ \nu\rho \end{array} \right\} = \frac{1}{2} g^{\mu\lambda} \left(\partial_\nu g_{\rho\lambda} + \partial_\rho g_{\nu\lambda} - \partial_\lambda g_{\nu\rho} \right) \ , \tag{3.170}$$

and $K^\mu{}_{\nu\rho}$ are the components of the *contorsion tensor*, which is totally antisymmetric in its lower indices.

- Finally, in General Relativity, we impose that $K^\mu{}_{\nu\rho} = 0$, choosing the *unique symmetric, metric compatible connection*. Its connection coefficients in a given coordinate system are then its Christoffel symbols:

Connection coefficients of the general relativistic connection

$$\Gamma^\mu{}_{\nu\rho} = \frac{1}{2} g^{\mu\lambda} \left(\partial_\nu g_{\lambda\rho} + \partial_\rho g_{\nu\lambda} - \partial_\lambda g_{\nu\rho} \right) \ . \tag{3.171}$$

Geodesics and the equivalence principle

We started our exploration of General Relativity by emphasising the central role of the equivalence principle. This stipulates that at any event in spacetime, there exists a local inertial frame in which the laws of physics are those of Special Relativity:

Local inertial frame

At any event $C \in \mathcal{M}$, we can find local coordinates X^μ such that:

$$\boldsymbol{g}_{|C} = \eta_{\mu\nu} \mathrm{d}X^\mu \otimes \mathrm{d}X^\mu \ , \tag{3.172}$$

and:

$$\forall (\mu, \nu, \rho) \in \{0, 1, 2, 3\}^3, \ \frac{\partial g_{\mu\nu}}{\partial X^\rho}(C) = 0 \ . \tag{3.173}$$

The associated frame is called a *local inertial frame* at C. In this frame, the laws of physics are those of Special Relativity.

The second condition should now be clear. Indeed, if at the event $p \in \mathcal{M}, \forall(\mu,\nu,\rho) \in \{0,1,2,3\}^3$, $\frac{\partial g_{\mu\nu}}{\partial X^\rho} = 0$, then, the connection coefficients are all identically zero and the geodesic equation reduces to:

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} = 0 \,, \tag{3.174}$$

which is simply the equation of motion of a free particle in Special Relativity. In other words, *geodesics are the trajectories of free falling particles*. Immediately, we have two interesting kind of geodesics[9]:

- Photons (and other massless particles) follow *lightlike* or *null* geodesics, characterised by their tangent vector $\boldsymbol{k}$ such that:

$$\boldsymbol{\nabla}_{\boldsymbol{k}} \boldsymbol{k} = 0 \;\Leftrightarrow\; \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu{}_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\rho}{\mathrm{d}\lambda} = 0 \tag{3.175}$$

$$\boldsymbol{g}\left(\boldsymbol{k},\boldsymbol{k}\right) = 0 \;\Leftrightarrow\; g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0 \,. \tag{3.176}$$

- Massive particles follow *timelike* geodesics, characterised by their 4-velocity $\boldsymbol{u}$ such that:

$$\boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{u} = 0 \;\Leftrightarrow\; \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\nu\rho} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\rho}{\mathrm{d}\tau} = 0 \tag{3.177}$$

$$\boldsymbol{g}\left(\boldsymbol{u},\boldsymbol{u}\right) = -1 \;\Leftrightarrow\; g_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x^\nu}{\mathrm{d}\tau} = -1 \,, \tag{3.178}$$

  where τ is the proper time along the geodesics.

We can now clarify the link between geodesics and straight lines in Special Relativity:

Timelike geodesics as extremal curves

Let $\mathcal{L}$ be a timelike curve connecting two events p and q. Let λ be a parameter along $\mathcal{L}$ which is *not* the proper time. Then, the proper time elapsed along $\mathcal{L}$ between p and q is:

$$\tau\left(p,q\right) = \int_{\lambda(p)}^{\lambda(q)} \mathrm{d}\tau = \int_{\lambda(p)}^{\lambda(q)} \sqrt{-g_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda}} \mathrm{d}\lambda \,. \tag{3.179}$$

Then $\mathcal{L}$ is a timelike geodesics iff it extremises this proper time.

[9] Spacelike geodesics are also interesting of course, but they do not correspond to the trajectory of any particles.

To obtain this result, we perturb the curve in the class of timelike curves: $x^\mu(\lambda) = \bar{x}^\mu(\lambda) + \delta x^\mu(\lambda)$, where we choose the same parameter along every timelike curve and $\bar{x}^\mu$ denotes the curve that extremises the proper time. Then, using a dot to denote derivatives with respect to λ:

$$\tau(p,q)\,[\bar{x}+\delta x] = \int_{\lambda(p)}^{\lambda(q)} \sqrt{-\left(g_{\alpha\beta}(\bar{x}) + \frac{\partial g_{\alpha\beta}}{\partial x^\mu}\delta x^\mu\right)\left(\dot{\bar{x}}^\alpha + \dot{\delta x}^\alpha\right)\left(\dot{\bar{x}}^\beta + \dot{\delta x}^\beta\right)}\mathrm{d}\lambda \tag{3.180}$$

$$= \int_{\lambda(p)}^{\lambda(q)} \sqrt{\underbrace{-g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta}_{=F^2(\bar{x})} \underbrace{- g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\alpha\dot{\delta x}^\beta + g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\beta\dot{\delta x}^\alpha}_{=2g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\alpha\dot{\delta x}^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta\delta x^\gamma}\mathrm{d}\lambda \tag{3.181}$$

$$= \int_{\lambda(p)}^{\lambda(q)} F(\bar{x})\mathrm{d}\lambda\sqrt{1 - F^{-2}(\bar{x})\left[2g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\alpha\dot{\delta x}^\beta + \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}(\bar{x})\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta\delta x^\gamma\right]} \tag{3.182}$$

$$= \tau(p,q)[\bar{x}] - \frac{1}{2}\int_{\lambda(p)}^{\lambda(q)} \frac{\mathrm{d}\lambda}{F(\bar{x})}\left[2g_{\alpha\beta}(\bar{x})\dot{\bar{x}}^\alpha\dot{\delta x}^\beta + \frac{\partial g_{\alpha\beta}}{\partial x^\gamma}(\bar{x})\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta\delta x^\gamma\right] . \tag{3.183}$$

Thus, after an integration by part:

$$\delta\tau(p,q) = \tau(p,q)[\bar{x}+\delta x] - \tau(p,q)[\bar{x}] = -\frac{1}{2}\int_{\lambda(p)}^{\lambda(q)}\left[-2\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(F^{-1}g_{\alpha\gamma}\dot{\bar{x}}^\alpha\right) + F^{-1}\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta\right]\delta x^\gamma . \tag{3.184}$$

Since this must be true for every perturbation in the class of timelike curve, we get:

$$\delta\tau(p,q) = 0 \;\Rightarrow\; -2\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(F^{-1}g_{\alpha\gamma}\dot{\bar{x}}^\alpha\right) + F^{-1}\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\dot{\bar{x}}^\alpha\dot{\bar{x}}^\beta = 0 . \tag{3.185}$$

Besides, $F^2(\bar{x}) = \left(\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right)^2$ by construction, so that:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} = F\frac{\mathrm{d}}{\mathrm{d}\tau} . \tag{3.186}$$

Hence, Eq. (3.185) becomes:

$$g_{\alpha\beta}\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} + \frac{\mathrm{d}g_{\alpha\gamma}}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau} - \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0 . \tag{3.187}$$

Thus:

$$g_{\alpha\beta}\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\tau^2} + \frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau} - \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0 . \tag{3.188}$$

Contracting with $g^{\gamma\delta}$:

$$\frac{\mathrm{d}^2x^\delta}{\mathrm{d}\tau^2} + g^{\gamma\delta}\frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau} - \frac{1}{2}g^{\gamma\delta}\frac{\partial g_{\alpha\beta}}{\partial x^\gamma}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0 . \tag{3.189}$$

Finally, we can split the second term by noticing that α and β are dummy indices:

$$g^{\gamma\delta}\frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau} = \frac{1}{2}g^{\gamma\delta}\frac{\partial g_{\alpha\gamma}}{\partial x^\beta}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau} + \frac{1}{2}g^{\gamma\delta}\frac{\partial g_{\beta\gamma}}{\partial x^\alpha}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau}\,, \tag{3.190}$$

so that we arrive at the equation for the curve extremising the proper time:

$$\frac{\mathrm{d}^2\bar{x}^\delta}{\mathrm{d}\tau^2} + \Gamma^\delta{}_{\alpha\beta}\frac{\mathrm{d}\bar{x}^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{x}^\beta}{\mathrm{d}\tau} = 0\,, \tag{3.191}$$

which is exactly the geodesic equation.

This argument does not work for lightlike geodesics of course, but it can be amended by extremising the proper time of arrival of the light ray at the observer for a fixed source. This is the general relativistic version of Fermat's principle of optics.

3.6.4 Application: static, weak field limit

As an illustration, let us consider the metric for a static, weak gravitational field such as the one at the surface of the Earth. We assumed that it was given by the form:

$$\mathrm{d}s^2 = -\left(1+2\Phi(x,y,z)\right)\mathrm{d}t^2 + \left(1-2\Phi(x,y,z)\right)\left[\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2\right]\,, \tag{3.192}$$

with $|\Phi| \ll 1$, so we must expand everything at first order in Φ. It is a good exercise to calculate the connection coefficients. The only non-zero ones are:

Connection coefficients in the static, weak field limit

$$\Gamma^0{}_{0i} = \Gamma^0{}_{i0} = \frac{\partial\Phi}{\partial x^i}\,,\quad \Gamma^i{}_{00} = \delta^{ij}\frac{\partial\Phi}{\partial x^j} \tag{3.193}$$

$$\Gamma^i{}_{jk} = -\frac{\partial\Phi}{\partial x^j}\delta^i{}_k - \frac{\partial\Phi}{\partial x^k}\delta^i{}_j + \frac{\partial\Phi}{\partial x^l}\delta^{il}\delta_{jk}\,. \tag{3.194}$$

The geodesic equations then take the form:

$$\begin{cases} \dfrac{\mathrm{d}^2t}{\mathrm{d}\lambda^2} + 2\dfrac{\partial\Phi}{\partial x^i}\dfrac{\mathrm{d}x^i}{\mathrm{d}\tau}\dfrac{\mathrm{d}t}{\mathrm{d}\lambda} = 0 & (3.195)\\[2ex] \dfrac{\mathrm{d}^2x^i}{\mathrm{d}\tau^2} + \delta^{ij}\dfrac{\partial\Phi}{\partial x^j}\left(\dfrac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 + \left[\delta^{il}\dfrac{\partial\Phi}{\partial x^l}\delta_{jk} - \dfrac{\partial\Phi}{\partial x^j}\delta^i{}_k - \dfrac{\partial\Phi}{\partial x^k}\delta^i{}_j\right]\dfrac{\mathrm{d}x^j}{\mathrm{d}\tau}\dfrac{\mathrm{d}x^k}{\mathrm{d}\tau} = 0\,. & (3.196) \end{cases}$$

Timelike geodesics

Timelike geodesics are then described by their 4-velocity:

$$\boldsymbol{u} = \bar{\boldsymbol{u}} + \boldsymbol{\delta u} \ , \tag{3.197}$$

where $\bar{\boldsymbol{u}}$ is the 4 velocity in absence of gravitational field and $\boldsymbol{\delta u} = O(\Phi)$. Therefore, we can write:

$$\bar{\boldsymbol{u}} = \gamma \left[\frac{\partial}{\partial t} + \bar{v}^i \frac{\partial}{\partial x^i} \right] \ , \tag{3.198}$$

with $\bar{v}^i = \frac{\mathrm{d}x^i}{\mathrm{d}t}$ and $\gamma^{-1} = \sqrt{1 - \|\vec{v}\|^2}$. Let us work with a particle which in absence of any gravitational field, would be at rest in the local frame[10], so that $\bar{v}^i = 0$ and $\gamma = 1$. In that case $\bar{\boldsymbol{u}} = \frac{\partial}{\partial t}$, so that:

$$\frac{\mathrm{d}}{\mathrm{d}\bar{\tau}} = \bar{u}^\mu \frac{\partial}{\partial x^\mu} = \frac{\mathrm{d}}{\mathrm{d}t} \ , \tag{3.199}$$

and:

$$\boldsymbol{u} = (1 + \delta u^0) \frac{\partial}{\partial t} + v^i \frac{\partial}{\partial x^i} \ , \tag{3.200}$$

with $v^i = \frac{\mathrm{d}x^i}{\mathrm{d}\tau} = \frac{\mathrm{d}x^i}{\mathrm{d}t} = O(\Phi)$. The condition $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = -1$ at first order then gives $\delta u^0 = -\Phi$. Using τ as parameter, Eq. (3.195) then gives:

$$\frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2} = 0 \ . \tag{3.201}$$

Besides, the third term in Eq. (3.196) is of order 3 and can be neglected, so that we get:

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = -\delta^{ij} \frac{\partial \Phi}{\partial x^j} \ . \tag{3.202}$$

This is Newton's law for a particle falling in the gravitational field, as it should be.

Lightlike geodesics

For lightlight geodesics with 4-momentum $\boldsymbol{k}$, we get:

$$\boldsymbol{k} = \bar{\boldsymbol{k}} + \boldsymbol{\delta k} \ . \tag{3.203}$$

In absence of gravitational field, we know that the energy is constant and:

$$\bar{k}^0 = E \ \text{ and } \ \bar{k}^i = E n^i \ , \tag{3.204}$$

[10] The case $\bar{v}^i \neq 0$ can be treated similarly, with a few more technical steps but nothing conceptually more subtle.

where n^i fixes the direction of propagation in absence of gravitational field. The constraint $\boldsymbol{g}(\boldsymbol{k},\boldsymbol{k}) = 0$ at first order then gives:

$$\frac{\delta k^0 - n^i \delta k_i}{E} = -2\Phi \ . \tag{3.205}$$

The geodesic equations become:

$$\begin{cases} \dfrac{\mathrm{d}\delta k^0}{\mathrm{d}\lambda} + 2E^2 n^i \dfrac{\partial \Phi}{\partial x^i} = 0 & (3.206) \\ \dfrac{\mathrm{d}\delta k^i}{\mathrm{d}\lambda} + 2E^2 \left[\delta^{ij} \dfrac{\partial \Phi}{\partial x^j} - \left(n^j \dfrac{\partial \Phi}{\partial x^j} \right) n^i \right] = 0 \ . & (3.207) \end{cases}$$

We can make some progress by noting that at leading order in these equations:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} = \bar{k}^\mu \frac{\partial}{\partial x^\mu} = E \left[\frac{\partial}{\partial t} + n^i \frac{\partial}{\partial x^i} \right] \ . \tag{3.208}$$

And since Φ does not depend explicitly on t:

$$E n^i \frac{\partial \Phi}{\partial x^i} = \frac{\mathrm{d}\Phi}{\mathrm{d}\lambda} \ . \tag{3.209}$$

Therefore:

$$\frac{\mathrm{d}\delta k^0}{\mathrm{d}\lambda} = -2E \frac{\mathrm{d}\Phi}{\mathrm{d}\lambda} \ . \tag{3.210}$$

Considering a photon emitted by a source at λ_S and received by an observer comoving with the coordinates at λ_O, we get:

$$\delta k^0(\lambda_O) - \delta k^0(\lambda_S) = 2E \left[\Phi(\lambda_S) - \Phi(\lambda_O) \right] \ . \tag{3.211}$$

Because the observer is comoving, $\boldsymbol{u} = (1-\Phi)\frac{\partial}{\partial t}$, and we can calculate the gravitational redshift:

$$1 + z = \frac{\left(k^\mu u_\mu\right)_S}{\left(k^\mu u_\mu\right)_O} \ . \tag{3.212}$$

Since

$$\boldsymbol{g}(\boldsymbol{k},\boldsymbol{u}) = k^\mu u_\mu = -E \left[1 + \Phi + \frac{\delta k^0}{E} \right] \ , \tag{3.213}$$

we get:

$$1 + z = 1 + \Phi(O) - \Phi(S) \ . \tag{3.214}$$

To compare with Eq. (2.339) expressing the same effects purely from the equivalence principle, we note that $P = S$, $Q = O$:

$$1 + z = \frac{E(S)}{E(O)} = \frac{\Delta\tau_Q}{\Delta\tau_P} = 1 + \Phi(O) - \Phi(S) \ , \tag{3.215}$$

so the two methods agree.
The second geodesic equation can also be used to calculate the deviation of light but this will be done in details in chapter 4 so we leave it for now.

3.7 Gravitation is curvature

As we just showed, free-falling particles follow geodesics. However, locally, we can always make a geodesics look like a straight line in spacetime by going to the local inertial frame around that geodesic. Are there effects of gravitation that cannot be made to vanish by such a change of frame? Yes, and as we are going to see, these are linked to tidal effects, i.e. differential in the value of the gravitational field in an extended (even small) region of spacetime. Let us start by considering a timelike geodesic $\mathcal{L}_0$ parametrised by an affine parameter λ. The observer following it can always cancel the effect of the gravitational field along its worldline. But if another free falling particle passes nearby along its own geodesic, it will not in general be at rest in the observer's rest space, unless the gravitational field does not vary between their respective positions. If the field varies across the region, effects of the gravitational field will be visible in the local rest frame. The two particles will be moving farther or closer: these are exactly tidal effects. This discussion induces us into studying the variations in the relative positions between geodesics.

3.7.1 Geodesic deviation equation

Let $\mathcal{L}_0$ be a timelike or lightlike geodesics parametrised by an affine parameter λ, and with a tangent vector field $\boldsymbol{X} = \frac{\mathrm{d}}{\mathrm{d}\lambda}$. We consider a continuous family of geodesics of the same kind, $\mathcal{L}(s)$, indexed by $s \in \mathbb{R}$, in the neighbourhood of $\mathcal{L}_0$, such that $\mathcal{L}_0 = \mathcal{L}(0)$ and all parametrised by the same affine parameter λ as the reference geodesic $\mathcal{L}_0$. Locally, events can be parametrised by the two numbers (s, λ) telling us on which geodesic they are (s) and where they are on that geodesic (λ); see Fig. 3.13.

To go from the point $x^\alpha(\lambda, 0)$ on the reference geodesic $\mathcal{L}_0$ to the point on the geodesics $\mathcal{L}(\mathrm{d}s)$ with the same affine parameter: $x^\alpha(\lambda, \mathrm{d}s)$, we move along the vector field $\xi = \frac{\mathrm{d}}{\mathrm{d}s}$ locally tangent to the curves parametrised by s:"

$$x^\alpha(\lambda, \mathrm{d}s) = x^\alpha(\lambda, 0) + \xi^\alpha(\lambda)\mathrm{d}s \ . \tag{3.216}$$

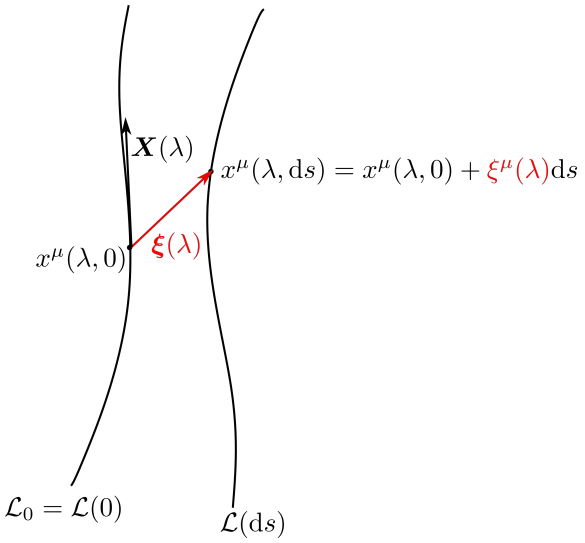


Figure 3.13: The deviation vector connecting geodesics into the neighbourhood of a reference geodesic $\mathcal{L}_0$ to that reference.

The vector field ξ defined along the reference geodesic is called the *separation vector*. The vector field, defined along the geodesic of reference:

$$\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{\xi} = \frac{\mathrm{D}\xi^\mu}{\mathrm{D}\lambda}\boldsymbol{e}_{(\mu)} \tag{3.217}$$

$$= \left(\frac{\mathrm{d}\xi^\mu}{\mathrm{d}\lambda} + \Gamma^\mu{}_{\nu\rho}X^\nu\xi^\rho\right)\boldsymbol{e}_{(\mu)} \tag{3.218}$$

can be thought of as the 'velocity' of this vector. Note that, for the convenience of this section, we defined the new derivative operator "covariant derivative along the geodesic" acting on components of vectors:

$$\frac{\mathrm{D}}{\mathrm{D}\lambda} = X^\mu\nabla_\nu \ . \tag{3.219}$$

Similarly, we have an acceleration:

$$\boldsymbol{\nabla}_{\boldsymbol{X}}\left[\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{\xi}\right] = \frac{\mathrm{D}^2\xi^\mu}{\mathrm{D}\lambda^2}\boldsymbol{e}_{(\mu)} \ . \tag{3.220}$$

It is the second quantity that is relevant. Let us calculate it:

$$\frac{\mathrm{D}^2\xi^\mu}{\mathrm{D}\lambda^2} = \boldsymbol{\nabla}_{\boldsymbol{X}}\left[\boldsymbol{\nabla}_{\boldsymbol{X}}\xi^\mu\right] \tag{3.221}$$

$$= X^\alpha\nabla_\alpha\left[X^\beta\nabla_\beta\xi^\mu\right] \ . \tag{3.222}$$

First, we need to note a remarkable properties of the vectors $\boldsymbol{X}$ and $\boldsymbol{\xi}$. Because they are associated to independent parameters, we have:

$$X^\alpha \nabla_\alpha \xi^\beta - \xi^\alpha \nabla_\alpha X^\beta = X^\alpha \frac{\partial \xi^\beta}{\partial x^\alpha} - \xi^\alpha \frac{\partial X^\beta}{\partial x^\alpha} \tag{3.223}$$

$$= \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\frac{\partial \xi^\beta}{\partial x^\alpha} - \frac{\mathrm{d}x^\alpha}{\mathrm{d}s}\frac{\partial X^\beta}{\partial x^\alpha} \tag{3.224}$$

$$= \frac{\mathrm{d}\xi^\beta}{\mathrm{d}\lambda} - \frac{\mathrm{d}X^\beta}{\mathrm{d}s} \tag{3.225}$$

$$= \frac{\mathrm{d}^2 x^\beta}{\mathrm{d}s\mathrm{d}\lambda} - \frac{\mathrm{d}^2 x^\beta}{\mathrm{d}\lambda \mathrm{d}s} = 0\ . \tag{3.226}$$

Thus:

$$\boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{\xi} = \boldsymbol{\nabla}_{\boldsymbol{\xi}} \boldsymbol{X}\ . \tag{3.227}$$

Therefore:

$$\frac{\mathrm{D}^2 \xi^\mu}{\mathrm{D}\lambda^2} = \boldsymbol{\nabla}_{\boldsymbol{X}} \left[\boldsymbol{\nabla}_{\boldsymbol{\xi}} X^\mu\right] \tag{3.228}$$

$$= X^\alpha \nabla_\alpha \left(\xi^\beta\right) \nabla_\beta X^\mu + X^\alpha \xi^\beta \nabla_\alpha \nabla_\beta X^\mu \tag{3.229}$$

$$= \left(\xi^\alpha \nabla_\alpha X^\beta\right) \nabla_\beta X^\mu + X^\alpha \xi^\beta \nabla_\alpha \nabla_\beta X^\mu \tag{3.230}$$

$$= \xi^\alpha \nabla_\alpha \left(\underbrace{X^\beta \nabla_\beta X^\mu}_{=0}\right) - \xi^\alpha X^\beta \nabla_\alpha \nabla_\beta X^\mu + X^\alpha \xi^\beta \nabla_\alpha \nabla_\beta X^\mu \tag{3.231}$$

$$= \xi^\beta X^\alpha \left[\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha\right] X^\mu\ . \tag{3.232}$$

This quantity is intimately related to the *Riemann curvature tensor* of the metric connection, which is defined for any one-form field $\boldsymbol{\omega}$ and three vector fields $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ as:

Riemann curvature tensor

$$\boldsymbol{R}\left(\boldsymbol{\omega}, \boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{Y}\right) = \boldsymbol{\omega}\left[\boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{\nabla}_{\boldsymbol{Y}} \boldsymbol{Z} - \boldsymbol{\nabla}_{\boldsymbol{Y}} \boldsymbol{\nabla}_{\boldsymbol{X}} \boldsymbol{Z} - \boldsymbol{\nabla}_{[\boldsymbol{X},\boldsymbol{Y}]} \boldsymbol{Z}\right]\ . \tag{3.233}$$

In a coordinate basis, its components are:

$$R^\alpha{}_{\rho\beta\gamma} = \frac{\partial \Gamma^\alpha{}_{\gamma\rho}}{\partial x^\beta} - \frac{\partial \Gamma^\alpha{}_{\beta\rho}}{\partial x^\gamma} + \Gamma^\sigma{}_{\gamma\rho}\Gamma^\alpha{}_{\beta\sigma} - \Gamma^\sigma{}_{\beta\rho}\Gamma^\alpha{}_{\gamma\sigma}\ . \tag{3.234}$$

Indeed, we see that we have:

Geodesic deviation equation

$$\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{\nabla}_{\boldsymbol{X}}\boldsymbol{\xi} = \boldsymbol{R}\left(\cdot, \boldsymbol{X}, \boldsymbol{X}, \boldsymbol{\xi}\right) \ , \tag{3.235}$$

or in terms of components:

$$\frac{\mathrm{D}^2\xi^\alpha}{\mathrm{D}\lambda^2} = R^\alpha{}_{\rho\beta\gamma}X^\rho X^\beta \xi^\gamma \ . \tag{3.236}$$

Equivalently, in a less covariantly beautiful but often useful form:

$$\frac{\mathrm{d}^2\xi^\alpha}{\mathrm{d}\lambda^2} + 2\Gamma^\alpha{}_{\nu\rho}X^\nu\frac{\mathrm{d}\xi^\rho}{\mathrm{d}\lambda} + \xi^\gamma\partial_\gamma\Gamma^\alpha{}_{\beta\rho}X^\beta X^\rho = 0 \ . \tag{3.237}$$

We see that tidal effects, i.e. relative motions of free-falling particles with respect to each other are encoded in this Riemann curvature tensor, so we will now spend some efforts understanding it a bit more and discovering a few of its properties.

3.7.2 The Riemann curvature tensor

Geometric interpretation

Before talking about the spacetime manifold, let us focus for a moment on the example of the 2-sphere S^2. We define the parallel transport along the great circles (i.e., equivalently, the connection) by requiring that the angle in the ambient space (so in the Euclidean sense) between a vector and the tangent to the great circle remains constant when we move the vector along the great circle. Consider two points p and q on the equator of S^2 (for simplicity) that are diametrically opposite. Then, there are two great circles through p and q, the 'equal latitude' one, and the 'equal longitude' one. We can observe that, the result of transporting a vector $\boldsymbol{V}$ from p to q according to our connection rule along the 'equal latitude' circle is very different from the result of transporting the same initial vector along the 'equal longitude' circle; see Fig. 3.14.

The fact that the result of transporting a vector depends on the path chosen to transport it is what characterises the curvature, and it is clearly independent on the coordinates chosen to represent the manifold.

In an arbitrary spacetime manifold with a metric connection, consider an infinitesimal 'parallelogram' $PQRS$, were, in a local chart (U, ϕ) such that $PQRS \subset U$, we can set $\phi(P) = x$, $\phi(Q) = x+\epsilon$, $\phi(S) = x+\delta$ and $\phi(S) = x+\epsilon+\delta$. The situation described in what follows is depicted on Fig. 3.15.

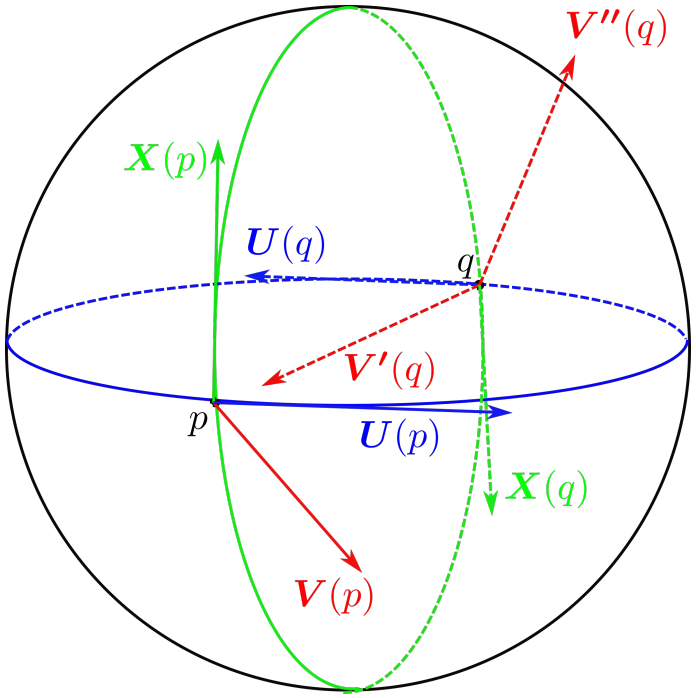


Figure 3.14: On the one hand, the vector $\boldsymbol{V(p)}$ is transported along the great circle tangent to $\boldsymbol{U}$ until q while preserving the Euclidean angle between the two vectors, producing $\boldsymbol{V'(q)}$. On the other hand, it is transported along the great circle tangent to $\boldsymbol{X}$ applying the same rule, producing $\boldsymbol{V''(q)}$. The difference between the two vectors comes from the curvature of the connection corresponding to our rule for parallel transport.

Let us parallel transport a vector $\boldsymbol{V(P)} \in T_P\mathcal{M}$ along PQ and then along QR. First, we obtain a vector $\boldsymbol{V_C(Q)} \in T_Q\mathcal{M}$ whose components in the local coordinate basis are given by:

$$V_C^\mu(Q) = V^\mu(P) - \Gamma^\mu{}_{\nu\rho}(P)V^\rho(P)\epsilon^\nu \ . \tag{3.238}$$

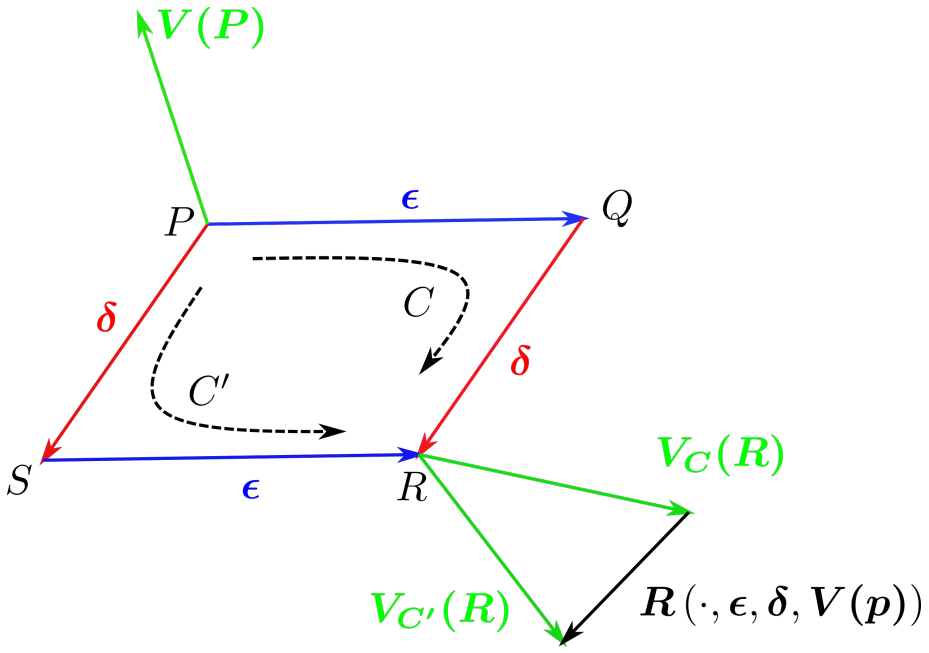


Figure 3.15: A same vector $\boldsymbol{V(P)}$is parallel transported to R along two different paths closing in a parallelogram. The mismatch between the two vectors obtained at R is quantified by the Riemann tensor.

Then, we transport this vector all the way to R, and we get:

$$\begin{aligned} V^\mu_C(R) &= V^\mu_C(Q) - \Gamma^\mu{}_{\nu\rho}(Q)V^\rho_C(Q)\delta^\nu && (3.239)\\ &= V^\mu(P) - \Gamma^\mu{}_{\nu\rho}(P)V^\rho(P)\epsilon^\nu - \left(\Gamma^\mu{}_{\nu\rho}(P) + \partial_\lambda\Gamma^\mu{}_{\nu\rho}(P)\epsilon^\lambda\right) \\ &\quad \times \left(V^\rho(P) - \Gamma^\rho{}_{\lambda\sigma}V^\sigma\epsilon^\lambda\right)\delta^\nu && (3.240)\\ &= V^\mu(P) - \Gamma^\mu{}_{\nu\rho}(P)V^\rho(P)\left(\epsilon^\nu + \delta^\nu\right) \\ &\quad - \left(\partial_\lambda\Gamma^\mu{}_{\nu\rho}(P) - \Gamma^\mu{}_{\nu\rho}(P)\Gamma^\sigma{}_{\lambda\rho}(P)\right)V^\rho(P)\epsilon^\lambda\delta^\nu\ , && (3.241)\end{aligned}$$

where we have used a Taylor expansion of the connection coefficients and we have only kept terms up to second order. We also used the notation: $\partial_\mu = \frac{\partial}{\partial x^\mu}$. Similarly, we can obtain for the vector transported from P to R via PS followed by SR:

$$\begin{aligned} V^\mu_{C'}(R) &= V^\mu(P) - \Gamma^\mu{}_{\nu\rho}(P)V^\rho(P)\left(\epsilon^\nu + \delta^\nu\right) \\ &\quad - \left(\partial_\nu\Gamma^\mu{}_{\lambda\rho}(P) - \Gamma^\mu{}_{\lambda\sigma}(P)\Gamma^\sigma{}_{\nu\rho}(P)\right)V^\rho(P)\epsilon^\lambda\delta^\nu\ . && (3.242)\end{aligned}$$

Hence, the two vectors at R differ by:

$$V^\mu_{C'}(R) - V^\mu_C(R) = R^\mu{}_{\rho\nu\lambda}(P)V^\rho(P)\epsilon^\nu\delta^\lambda\ . \tag{3.243}$$

In other words, the Riemann tensor measures the difference between two vectors resulting from the parallel transport of one vector in two directions along an infinitesimal parallelogram[11].

We will say that spacetime is *flat* iff $\boldsymbol{R} = 0$, otherwise spacetime is said to be *curved*. This characterisation does not depend on the local charts chosen as it is tensorial.

Properties of the Riemann tensor

Since the Riemann tensor is the crucial object, holding the properties of the gravitational field in General Relativity, we are going to present some of its most important properties. First, we have

[11]The Lie bracket, that we introduced incidently early without comment, measures the inability to close a parallelogram by following the flows of two vector fields in one way or the opposite. Once we have a connection, thanks to the Riemann tensor, we can actually say what it 'costs' for vectors to be transported along closed parallelograms.

some symmetry properties of its components:

$$\begin{cases} R^{\mu}{}_{\nu\sigma\rho} = -R^{\mu}{}_{\nu\rho\sigma} & (3.244) \\ g_{\mu\lambda}R^{\lambda}{}_{\nu\rho\sigma} = R_{\mu\nu\rho\sigma} = -R_{\nu\mu\rho\sigma} = -g_{\nu\lambda}R^{\lambda}{}_{\mu\rho\sigma} & (3.245) \\ R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu} & (3.246) \\ R^{\mu}{}_{\nu\rho\sigma} + R^{\mu}{}_{\sigma\nu\rho} + R^{\mu}{}_{\rho\sigma\nu} = 0 \text{ (first Bianchi identities).} & (3.247) \end{cases}$$

These can be verified by careful substitutions and inspections. The Riemann tensor has $4^4 = 256$ components, but, thanks to these symmetries, only 20 are independent.
Next, we have a important relation called *the second Bianchi identities*:

$$(\nabla_X R)(\cdot, U, Y, Z) + (\nabla_Z R)(\cdot, U, X, Y) + (\nabla_Y R)(\cdot, U, Z, X) = 0;, \tag{3.248}$$

valid for any vector fields $\boldsymbol{U}$, $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$. It is better known in its expression in terms of components:

$$\nabla_\mu R^{\nu}{}_{\lambda\rho\sigma} + \nabla_\rho R^{\nu}{}_{\lambda\sigma\mu} + \nabla_\sigma R^{\nu}{}_{\lambda\mu\rho} = 0\ . \tag{3.249}$$

3.7.3 Application: weak field limit

We return to the static, weak field gravitational field, for which the metric is given by Eq. (3.192). As an exercise, one can compute the components of the Riemann tensor. The only non-zero ones are:

Riemann tensor in the static, weak field limit

$$R^{0}{}_{ij0} = -R^{0}{}_{i0j} = \frac{\partial^2\Phi}{\partial x^i \partial x^j} \tag{3.250}$$

$$R^{i}{}_{0j0} = -R^{i}{}_{00j} = \delta^{ik}\frac{\partial^2\Phi}{\partial x^k \partial x^j} \tag{3.251}$$

$$R^{i}{}_{klj} = R^{i}{}_{jlk} = \frac{\partial^2\Phi}{\partial x^k \partial x^j}\delta^{i}{}_{l} - \frac{\partial^2\Phi}{\partial x^k \partial x^l}\delta^{i}{}_{j} + \frac{\partial^2\Phi}{\partial x^l \partial x^m}\delta^{im}\delta_{jk} - \frac{\partial^2\Phi}{\partial x^j \partial x^m}\delta^{im}\delta_{lk}\ . \tag{3.252}$$

Consider a massive particle in free-fall in this gravitational field. It follows a timelike geodesic like the ones we studied in subsection 3.6.4 . In particular, its 4-velocity is given by:

$$\boldsymbol{u} = (1-\Phi)\frac{\partial}{\partial t} + v^i\frac{\partial}{\partial x^i}\ , \tag{3.253}$$

with $v^i = \frac{\mathrm{d}x^i}{\mathrm{d}t} = O(\Phi)$. Thus, at first order, the geodesic deviation equation. Eq. (3.236), reads:

$$\frac{\mathrm{D}^2\xi^\mu}{\mathrm{D}\tau^2} = R^\mu{}_{00\nu}\xi^\nu \ . \tag{3.254}$$

Note that, at first order:

$$\frac{\mathrm{D}\xi^i}{\mathrm{D}\tau} = \frac{\mathrm{d}\xi^i}{\mathrm{d}\tau} + \Gamma^i{}_{0\nu}\xi^\nu \ , \tag{3.255}$$

and:

$$\frac{\mathrm{D}^2\xi^i}{\mathrm{D}\tau^2} = \frac{\mathrm{d}^2\xi^i}{\mathrm{d}\tau^2} \ , \tag{3.256}$$

so that the spatial separation between two neighbouring free-falling massive particles obeys:

$$\boxed{\frac{\mathrm{d}^2\xi^i}{\mathrm{d}\tau^2} = -\delta^{ik}\frac{\partial^2\Phi}{\partial x^k\partial x^j}\xi^j \ .} \tag{3.257}$$

This is exactly what a Newtonian calculation would give for the tidal force acting on two nearby free-falling objects.

3.7.4 Constructing local inertial frames: Riemann and Fermi normal coordinates

As we have repeatedly seen, local inertial frames, i.e. local reference systems attached to free-falling observers and in which the laws of special relativity hold, are key to formulating the laws of general relativity. So far, we have assumed that we could always construct such frames[12]. Here, we show how to explicitly build such frames. We start with Riemann normal coordinates defined at an event. The strategy is the following:

1. We pick an event O in a spacetime with metric $\boldsymbol{g}$, along the worldline of a free-falling observer. We call x^μ a set of local coordinates.

2. We pick a timelike unit vector $\hat{\boldsymbol{e}}_{(\mathbf{0})}$. This could be the 4-velocity of the free-falling observer passing by O, but it does not have to be.

[12] Actually, we have once constructed one, in subsection 2.7.2, for Rindler observers uniformly accelerating in Minkowski spacetime. By the equivalence principle, this is clearly analogous to a free falling observer in a uniform gravitational field.

3. We pick 3 linearly independent, orthogonal spacelike unit vectors $\hat{\boldsymbol{e}}_{(i)}$. These could be in the local rest space of the observer passing by O, but they don't have to be. In the local coordinate system around O, we have (orthonormality condition):

$$g_{\mu\nu}\hat{e}^{\mu}_{(\alpha)}\hat{e}^{\nu}_{(\beta)} = \eta_{\alpha\beta} \ . \tag{3.258}$$

4. We choose another event P and draw a geodesic segment $\mathcal{S}$ connecting O to P. Under reasonable assumptions, we can assume that this geodesic segment is unique, provided P is in a sufficiently small neighbourhood of O. If the geodesic is spacelike, we choose the proper distance s as a parameter λ, and if it is timelike, we take the proper time τ. Up to a simple redefinition, we can choose $\lambda(O) = 0$ and $\lambda(P) = \lambda_P$. Let us call $\boldsymbol{n}$ the tangent vector to this geodesic. Letting $\boldsymbol{e}_{(\mu)} = \frac{\partial}{\partial x^{\mu}}$, we have:

$$\boldsymbol{n} = n^{\mu}\boldsymbol{e}_{(\mu)} \tag{3.259}$$

$$= \hat{n}^{\alpha}\hat{e}_{(\alpha)} = \hat{n}^{\alpha}\hat{e}^{\mu}_{(\alpha)}\boldsymbol{e}_{(\mu)} \ , \tag{3.260}$$

so that, in the local coordinate system at O:

$$n^{\mu}(O) = \hat{n}^{\alpha}_{O}\hat{e}^{\mu}_{(\alpha)} \ . \tag{3.261}$$

The four numbers $\hat{n}^{\alpha}_{O}$ give the direction of the geodesic segment $\mathcal{S}$ relative to the tetrad $\{\hat{\boldsymbol{e}}_{(\alpha)}\}$.

5. We define the *Riemann normal coordinates of* P in the local tetrad $\{\hat{\boldsymbol{e}}_{(\alpha)}\}$ to be the four numbers $\hat{x}^{\alpha}$ such that:

$$\boxed{\hat{x}^{\alpha} = \lambda_P \hat{n}^{\alpha}_{O} \ .} \tag{3.262}$$

For the sake of clarity, let us assume that $\mathcal{S}$ is spacelike. Let us denote by $\hat{g}_{\alpha\beta}$ the components of the metric tensor in Riemann normal coordinates in the neighbourhood of O. Since $\boldsymbol{n}$ is a unit vector, we have along $\mathcal{S}$:

$$\boldsymbol{g}(\boldsymbol{n}, \boldsymbol{n}) = 1 \ , \tag{3.263}$$

which, in the local coordinates x^{μ}, and evaluated at O reads:

$$g_{\mu\nu}(O)n^{\mu}(O)n^{\nu}(O) = 1 \tag{3.264}$$

$$= g_{\mu\nu}(O)\hat{n}^{\alpha}_{O}\hat{e}^{\mu}_{(\alpha)}\hat{n}^{\beta}_{O}\hat{e}^{\nu}_{(\beta)} \tag{3.265}$$

$$= \eta_{\alpha\beta}\hat{n}^{\alpha}_{O}\hat{n}^{\beta}_{O} \ . \tag{3.266}$$

Besides, in Riemann normal coordinates, Eq. (3.263) evaluated at O reads:

$$\hat{g}_{\alpha\beta}(O)\hat{n}_O^\alpha \hat{n}_O^\beta = 1 \ . \tag{3.267}$$

Thus:

$$\hat{g}_{\alpha\beta}(O)\hat{n}_O^\alpha \hat{n}_O^\beta = \eta_{\alpha\beta}\hat{n}_O^\alpha \hat{n}_O^\beta \ , \tag{3.268}$$

and since the metric at a point cannot depend on the direction:

$$\boxed{\hat{g}_{\alpha\beta}(O) = \eta_{\alpha\beta} \ .} \tag{3.269}$$

At the event O, Riemann normal coordinates are orthonormal. However, there is more: this is also true in the neighbourhood of O in a precise sense, as we will see now. The geodesic segment $\mathcal{S}$ is tangent to the vector field $\boldsymbol{n}$, which in Riemann normal coordinates decomposes as:

$$\boldsymbol{n} = \hat{n}^\alpha \frac{\partial}{\partial \hat{x}^\alpha} \ , \tag{3.270}$$

with constant components along the geodesic segment:

$$\hat{n}^\alpha = \frac{\mathrm{d}\hat{x}^\alpha}{\mathrm{d}s} = \frac{\mathrm{d}}{\mathrm{d}s}\left(s\hat{n}_O^\alpha\right) = \hat{n}_O^\alpha \ . \tag{3.271}$$

Let us write the geodesic equation along $\mathcal{S}$ in Riemann normal coordinates:

$$\frac{\mathrm{d}\hat{n}^\alpha}{\mathrm{d}s} + \hat{\Gamma}^\alpha{}_{\beta\gamma}\hat{n}^\beta \hat{n}^\gamma = 0 \ , \tag{3.272}$$

and since $\hat{n}^\alpha = \hat{n}_O^\alpha$ are constant, we find that:

$$\hat{\Gamma}^\alpha{}_{\beta\gamma}\hat{n}^\beta \hat{n}^\gamma = 0 \ . \tag{3.273}$$

This relation is valid everywhere along $\mathcal{S}$ and since at O the connection coefficients cannot depend on direction, we get:

$$\hat{\Gamma}^\alpha_{\beta\gamma}(O) = 0 \ . \tag{3.274}$$

Clearly, this implies that, in Riemann normal coordinates[13]:

$$\boxed{\frac{\partial \hat{g}_{\alpha\beta}}{\partial \hat{x}^\gamma}(O) = 0 \ .} \tag{3.275}$$

[13] It is sufficient to realise that in any coordinate system: $\partial_\gamma g_{\alpha\beta} = g_{\alpha\mu}\Gamma^\mu{}_{\beta\gamma} + g_{\beta\mu}\Gamma^\mu{}_{\alpha\gamma}$.

Thus, we see that, by Taylor expanding the metric coefficients in Riemann normal coordinates, in the neighbourhood of O we have:

$$\begin{cases} \hat{g}_{\alpha\beta}\left(\hat{x}^{\gamma}\right) = \hat{g}_{\alpha\beta}(O) + \hat{x}^{\gamma}\dfrac{\partial \hat{g}_{\alpha\beta}}{\partial \hat{x}^{\gamma}}(O) + O\left(\hat{x}^{2}\right) & (3.276) \\ \qquad\quad = \eta_{\alpha\beta} + O\left(\hat{x}^{2}\right) , & (3.277) \end{cases}$$

confirming that the frame of Riemann normal coordinates is indeed locally inertial. Corrections appear at the next order and to evaluate them, we need to use the geodesic deviation equation . First, let us note that if we vary slightly the numbers $\hat{n}^{\alpha}_{O}$, we define new geodesics from O so that we have a family of deviation vectors with components:

$$\eta^{\mu}_{(\alpha)} = \frac{\partial \hat{x}^{\alpha}}{\partial \hat{n}^{\alpha}_{O}} \, . \tag{3.278}$$

Be careful that (α) here is not a spacetime index: it is merely a label. Expanding in the neighbourhood of O:

$$\hat{\Gamma}^{\mu}{}_{\nu\rho} = 0 + \partial_{\sigma}\hat{\Gamma}^{\mu}{}_{\nu\rho}(O)\hat{x}^{\sigma} + O\left(\hat{x}^{2}\right) \, , \tag{3.279}$$

we get:

$$\frac{\mathrm{D}\eta^{\mu}_{(\alpha)}}{\mathrm{D}s} = \delta^{\mu}{}_{\alpha} + s^{2}\partial_{\rho}\hat{\Gamma}^{\mu}{}_{\alpha\lambda}(O)\hat{n}^{\lambda}_{O}\hat{n}^{\rho}_{O} + O\left(s^{3}\right) \, . \tag{3.280}$$

Taking a second derivative[14]:

$$\frac{\mathrm{D}^{2}\eta^{\mu}_{(\alpha)}}{\mathrm{D}s^{2}} = 3s\partial_{\rho}\hat{\Gamma}^{\mu}{}_{\alpha\lambda}(O)\hat{n}^{\lambda}_{O}\hat{n}^{\rho}_{O} + O(s) \, . \tag{3.281}$$

Thus, using the geodesic deviation equation, Eq. (3.236), and once again using the fact that the result ought to be independent from the direction, and remembering the symmetries of the Riemann tensor:

$$3\frac{\partial \hat{\Gamma}^{\mu}{}_{\alpha\lambda}}{\partial \hat{x}^{\rho}}(O) = -\hat{R}^{\mu}{}_{\rho\alpha\lambda}(O) \, . \tag{3.282}$$

Hence:

$$\partial_{\rho}\hat{\Gamma}^{\mu}{}_{\alpha\lambda}(O) + \partial_{\lambda}\hat{\Gamma}^{\mu}{}_{\alpha\rho}(O) = -\frac{1}{3}\left[\hat{R}^{\mu}{}_{\rho\alpha\lambda}(O) + \hat{R}^{\mu}{}_{\lambda\alpha\rho}(O)\right] \, . \tag{3.283}$$

[14]This is where you have to be careful that (α) is not a coordinate index here.

Permuting indices 2 by two and combining the results we get:

$$\partial_\lambda \hat{\Gamma}^\mu{}_{\alpha\rho}(O) = -\frac{1}{3}\left[\hat{R}^\mu{}_{\rho\alpha\lambda}(O) + \hat{R}^\mu{}_{\alpha\rho\lambda}(O)\right] \ . \tag{3.284}$$

From there, we can easily extract that:

$$\boxed{\frac{\partial^2 \hat{g}_{\alpha\beta}}{\partial\hat{x}^\lambda \partial\hat{x}^\rho}(0) = -\frac{1}{3}\left[\hat{R}_{\alpha\rho\beta\lambda} + \hat{R}_{\beta\rho\alpha\lambda}\right] \ .} \tag{3.285}$$

Thus, we find that, at dominant order:

$$\boxed{\hat{g}_{\alpha\beta}\left(\hat{x}^\gamma\right) = \eta_{\alpha\beta} - \frac{1}{3}\hat{R}_{\alpha\rho\beta\lambda}\hat{x}^\rho\hat{x}^\lambda + O\left(\hat{x}^3\right) \ .} \tag{3.286}$$

We recover that the Riemann tensor encodes tidal effects, which are the true gravitational degrees of freedom and the only gravitational effects present in a local inertial coordinate system.

Riemann normal coordinates are adapted to the definition of an inertial frame in one given event. If one wants to define an inertial frame that is carried around by an observer, one needs to be able to consistently define orthonormal coordinates along the worldline of this observer. This is what *Fermi normal coordinates* are. In order to define them, let us pick an observer $\mathcal{A}$ following a timelike geodesic γ. We parametrise the geodesic by the proper time τ of the observer. For any event P outside γ, we draw a spacelike geodesic β that passes through P and intersect γ orthogonally at $A(\tau)$, such that P is in the local rest space of $A(\tau)$. Picking $A(\tau)$ as origin, we choose the tangent vector to γ at $A(\tau)$, $\bar{\boldsymbol{e}}_{(\mathbf{0})}(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}$ as the time direction and we pick three orthonormal spacelike vectors $\bar{\boldsymbol{e}}_{(\boldsymbol{i})}(\tau)$ in the tangent rest space of $A(\tau)$. The geodesic β is parametrised by the proper distance s so that $s = 0$ corresponds to $A(\tau)$ et $s = s_P$ to P. The tangent vector to β is the unit vector $\boldsymbol{n}$. At $A(\tau)$, this vector, which is spacelike, can be decomposed on the local tetrad and its components is the local coordinates x^μ are:

$$n^\alpha(\tau) = \bar{n}^i \bar{e}^\alpha_{(i)} \ . \tag{3.287}$$

$\bar{n}^0 = 0$ because $\boldsymbol{g}\left(\boldsymbol{n}, \bar{\boldsymbol{e}}_{(\mathbf{0})}\right) = 0$ and the tetrad is orthonormal. The *Fermi normal coordinates* of P along γ are then $\{\bar{x}^\alpha\}$ such that:

$$\boxed{\begin{cases} \bar{x}^0 = \tau & (3.288) \\ \bar{x}^i = s_P \bar{n}^i \ . & (3.289) \end{cases}}$$

Let us, for the sake of the demonstration assume that we have a set of Riemman normal coordinates $\{\hat{x}^\mu\}$ at $O \in \gamma$ constructed as above and that $O = A(0)$. We suppose that the vector $\hat{\boldsymbol{e}}_{\boldsymbol{0}}$ of the Riemman normal coordinates is aligned with the tangent vector to γ at $\tau = 0$. We set:

$$\bar{\boldsymbol{e}}_{(\mu)}(0) = \hat{\boldsymbol{e}}_{(\mu)} \ . \tag{3.290}$$

The basis vectors at any other event along γ are the obtained by parallel transport along γ. Writing:

$$\bar{e}^\mu_{\ (\alpha)}(\tau) = \bar{e}^\mu_{\ (\alpha)}(0) + \tau\dot{\bar{e}}^\mu_{\ (\alpha)}(0) + \frac{1}{2}\tau^2\ddot{\bar{e}}^\mu_{\ (\alpha)}(0) + O\left(\tau^2\right) \ , \tag{3.291}$$

where a dot denotes differentiation along γ (i.e. w.r.t. τ). We can use that the basis is a Riemann normal one at $\tau = 0$ to set $\bar{e}^\mu_{\ (\alpha)}(0) = \delta^\alpha_{\ \mu}$. Moreover, the parallel transport of the basis vectors $\bar{\boldsymbol{e}}_{(\mu)}$ along γ in Riemann normal coordinates reads:

$$\frac{\mathrm{d}\bar{e}^\mu_{\ (\nu)}}{\mathrm{d}\tau} + \hat{\Gamma}^\mu_{\ \lambda\rho} e^\lambda_{\ (\nu)} \frac{\mathrm{d}\hat{x}^\rho}{\mathrm{d}\tau} = 0 \ . \tag{3.292}$$

Note that these are components on the Fermi normal frame vectors in the Riemann normal frame, not in the local coordinates $\{x^\mu\}$ but notations have their limits here. This can be solved order by order in powers of τ. We find:

$$\begin{cases} \dot{\bar{e}}^\mu_{\ (\nu)}(0) = 0 & (3.293) \\ \ddot{\bar{e}}^\mu_{\ (\nu)}(0) = \frac{1}{3}\hat{R}^\mu_{\ 0\nu 0} \ . & (3.294) \end{cases}$$

Thus:

$$\bar{e}^\mu_{\ (\nu)}(\tau) = \delta^\mu_{\ \nu} + \frac{1}{6}\tau^2\hat{R}^\mu_{\ 0\nu 0} + O\left(\tau^3\right) \ . \tag{3.295}$$

This completes the construction of the Fermi basis along γ in the vicinity of O. Along the spacelike geodesic β between P and $O(\tau)$, we have the geodesic equation in Riemann normal coordinates:

$$\frac{\mathrm{d}\hat{x}^\mu}{\mathrm{d}s} + \hat{\Gamma}^\mu_{\ \nu\rho}\frac{\mathrm{d}\hat{x}^\nu}{\mathrm{d}s}\frac{\mathrm{d}\hat{x}^\rho}{\mathrm{d}s} = 0 \ . \tag{3.296}$$

Inserting into this equation the Taylor expansion:

$$\hat{x}^\mu(s) = \hat{x}^\mu(0) + s\dot{\hat{x}}^\mu + \frac{1}{2}s^2\ddot{\hat{x}}^\mu(0) + \frac{1}{6}s^3\frac{\mathrm{d}^3\hat{x}^\mu}{\mathrm{d}s^3}(0) + O\left(s^3\right) \ , \tag{3.297}$$

where a dot is now a derivative w.r.t s, we can again solve order by order, remembering that $\hat{x}^0(0) = \tau$, $\hat{x}^i(0) = 0$ and $\dot{\hat{x}}^\mu(0) = \hat{n}^\mu(O)$, to get:

$$\begin{cases} \hat{x}^0(s) = \tau + \frac{1}{3}\tau s^2\hat{R}_{0p0q}(O)\hat{n}^p\hat{n}^q + \dots & (3.298) \\ \hat{x}^i(s) = s\hat{n}^i + \frac{1}{6}\tau^2 s\hat{R}^i_{\ 0p0}()\hat{n}^p + \frac{1}{3}\tau s^2\hat{R}^i_{\ pq0}\hat{n}^p\hat{n}^q + \dots \ . & (3.299) \end{cases}$$

The Riemann tensor here is evaluated at O. Finally, this gives the coordinate transformation from Fermi to Riemann normal coordinates:

$$\begin{cases} \hat{x}^0(s) = \bar{x}^0 + \frac{1}{3}\bar{x}^0 \hat{R}_{0p0q}\bar{x}^p\bar{x}^q + \dots & (3.300) \\ \hat{x}^i(s) = \bar{x}^i + \frac{1}{6}\left(\bar{x}^0\right)^2 \hat{R}^i{}_{0p0}\bar{x}^p + \frac{1}{3}\bar{x}^0\hat{R}^i{}_{pq0}\bar{x}^p\bar{x}^q + \dots \ . & (3.301) \end{cases}$$

The metric components in Fermi normal coordinates can then be obtained from those in Riemann normal coordinates by a change of coordinates, and one finds:

$$\begin{cases} \bar{g}_{00} = -1 - \hat{R}_{0p0q}(\tau)\bar{x}^p\bar{x}^q + O\left(\bar{x}^3\right) & (3.302) \\ \bar{g}_{0i} = \frac{2}{3}\hat{R}_{ipq0}(\tau)\bar{x}^p\bar{x}^q + O\left(\bar{x}^3\right) & (3.303) \\ \bar{g}_{ij} = \delta_{ij} - \frac{1}{3}\hat{R}_{ipjq}(\tau)\bar{x}^p\bar{x}^q + O\left(\bar{x}^3\right) \ . & (3.304) \end{cases}$$

Note that, in principle, the components of the Riemann tensor here are evaluated at O, not that $A(\tau)$. But, notice that moving from O to $A(\tau)$ does not change the metric components at order 2 in $\bar{x}$, nor does the coordinate transformation at third orderSo the error on the components of the Riemann tensor are negligible at third order. Finally, note that these components, when expressed in Riemann or Fermi normal coordinates, differ at most by terms of order $\bar{x}^2$, so that, in the metric components, one can use the components in either frame to express the metric components, at the order considered here.

We see that, along γ, with $\bar{x}^i = 0$, we have:

$$\bar{g}_{\mu\nu}(\gamma) = \eta_{\mu\nu} \quad \text{and} \quad \bar{\Gamma}^{\mu}{}_{\nu\rho}(\gamma) = 0 \ . \tag{3.305}$$

We have thus constructed a free-falling frame carried by $\mathcal{A}$ in its motion.

3.8 Energy, momentum and the energy-momentum tensor

So far, we have discussed how matter, in the form of massless or massive point particles, reacts to a *given* gravitational field. This means that we have treated matter as a collection of test-particles we could use to probe the geometry of spacetime. As we have discussed, in General Relativity, this is encapsulated in the metric of spacetime, parallel transport and the associated curvature of

the affine connection. To have a complete theory of gravitation, we still have to determine how the gravitational field is *generated* by the matter present in the Universe. This will properly be the topic of the next section; but before we can establish this link, we have to introduce the correct way to describe matter in General Relativity. This is the goal of this section.

3.8.1 The energy-momentum tensor

In General Relativity, on sufficiently large scales, the metric of spacetime is not sensitive to all the details in the distribution of mater. It only feels a few coarse-grained properties of that distribution, namely its *energy* and its averaged *momentum*. So although, by the equivalence principle and the relativistic equivalence between mass and energy, it is sensitive to the distribution of *all matter*, particles but also fields such as the electromagnetic field, it is only sensitive to some of their collective properties. These are fully encapsulated in the *energy-momentum tensor*.

Energy-momentum tensor

The energy-momentum tensor of matter is a symmetric $(0,2)$-tensor field $\boldsymbol{T}$.

As we have seen, the notions of energy and momentum are observer-dependent for particles, so we expect them to also depend on the observer for the collective distribution of matter. Let $\mathcal{O}$ be an observer, with 4-velocity $\boldsymbol{U_O}$, and a local basis for its rest frame $\left\{\boldsymbol{e_{(i)}}\right\}$. Then:

1. $\rho_{\mathcal{O}} = \boldsymbol{T}\left(\boldsymbol{U_O}, \boldsymbol{U_O}\right)$ is the *energy density* of the matter measured by $\mathcal{O}$.

2. $p^i_{\mathcal{O}} = -\boldsymbol{T}(\boldsymbol{e_{(i)}}, \boldsymbol{U_O})$ are the components of the *momentum density* of the matter measured by $\mathcal{O}$, $\boldsymbol{p_O} = p^i_{\mathcal{O}} \boldsymbol{e_{(i)}}$.

3. $q^i_{\mathcal{O}} = -\boldsymbol{T}(\boldsymbol{U_O}, \boldsymbol{e_{(i)}})$ are the components of the *energy flux* of the matter measured by $\mathcal{O}$, $\boldsymbol{q_O} = q^i_{\mathcal{O}} \boldsymbol{e_{(i)}}$. The energy crossing a surface element in the observer's rest frame with area $\mathrm{d}S$ and normal $\vec{n}$, in a time $\mathrm{d}t$, is thus:

$$\mathrm{d}E = n_i q^i_{\mathcal{O}} \mathrm{d}S\mathrm{d}t \ . \tag{3.306}$$

   By symmetry of the energy-momentum tensor, in units of $c = 1$, we get $\boldsymbol{q_O} = \boldsymbol{p_O}$, which is a consequence of the relativistic equivalence between mass and energy.

4. $\Pi^{O}_{ij} = \boldsymbol{T}\left(\boldsymbol{e}_{(i)}, \boldsymbol{e}_{(j)}\right)$ are the components of the *stress* of the matter measured by O, $\boldsymbol{\Pi}_{\boldsymbol{O}}$ i.e. the force exerted by matter along the direction $\boldsymbol{e}_{(i)}$ on the surface normal to $\boldsymbol{e}_{(j)}$. It is the part of the energy-momentum tensor acting on the rest space of the observer.

By duality, we can construct a $(2,0)$-tensor, $\boldsymbol{T}^{*}$ equivalent to the energy-momentum tensor as defined above. In terms of the quantities observed by O:

$$\boldsymbol{T}^{*} = \rho_{O}\boldsymbol{U}_{O} \otimes \boldsymbol{U}_{O} + \boldsymbol{p}_{O} \otimes \boldsymbol{U}_{O} + \boldsymbol{U}_{O} \otimes \boldsymbol{q}_{O} + \boldsymbol{\Pi}_{O} \ . \tag{3.307}$$

The energy-momentum tensor is obtained from it by using the dual of each vector and tensor:

$$\boldsymbol{T} = \rho_{O}\boldsymbol{U}^{*}_{O} \otimes \boldsymbol{U}^{*}_{O} + \boldsymbol{p}^{*}_{O} \otimes \boldsymbol{U}^{*}_{O} + \boldsymbol{U}^{*}_{O} \otimes \boldsymbol{q}^{*}_{O} + \boldsymbol{\Pi}^{*}_{O} \ . \tag{3.308}$$

In terms of components in a local basis, we get:

$$T_{\mu\nu} = \rho_{O}U_{O\mu}U_{O\nu} + p_{\mu}U_{O\nu} + p_{\nu}U_{O\mu} + \Pi_{O\mu\nu} \ . \tag{3.309}$$

The stress can be further decomposed into an isotropic part and an anisotropic part by isolating its trace:

$$\Pi_{O\mu\nu} = P_{O}\left(g_{\mu\nu} + u_{\mu}u_{\nu}\right) + \hat{\Pi}_{O\mu\nu} \ , \tag{3.310}$$

where the trace is $3P_{O}$ with P_{O} the *pressure* measured by O, and $\hat{\boldsymbol{\Pi}}_{\boldsymbol{O}}$ is traceless.

3.8.2 Energy-momentum tensor for a fluid

If matter can be described by a continuous fluid, we can define the 4-velocity of the fluid elements $\boldsymbol{u}$, and the energy momentum tensor gets a natural decomposition by using the quantities measured by an observer *comoving* with the fluid: $\boldsymbol{U}_{\boldsymbol{O}} = \boldsymbol{u}$. Then the energy-momentum tensor can be written:

$$\boldsymbol{T} = (\rho + P)\boldsymbol{u}^{*} \otimes \boldsymbol{u}^{*} + P\boldsymbol{g} + \boldsymbol{q}^{*} \otimes \boldsymbol{u}^{*} + \boldsymbol{u}^{*} \otimes \boldsymbol{q}^{*} + \hat{\boldsymbol{\Pi}}^{*} \ . \tag{3.311}$$

The quantities defined here are:

1. the energy-density of the fluid: ρ;
2. the pressure of the fluid: P
3. the energy flux of the fluid: $\boldsymbol{q} = q^{\mu}\boldsymbol{e}_{(\mu)}$ with $q^{\mu}u_{\mu} = 0$;

4. the anisotropic stress of the fluid: $\hat{\boldsymbol{\Pi}} = \hat{\Pi}^{\mu\nu}\boldsymbol{e}_{(\mu)}\boldsymbol{e}_{(\nu)}$ with $\hat{\Pi}^{\mu}{}_{\mu} = 0$ and $\hat{\Pi}^{\mu\nu}u_{\nu} = 0$.

For a perfect fluid: $\boldsymbol{q} = 0$ and $\hat{\boldsymbol{\Pi}} = 0$, so that:

Energy-momentum tensor for a perfect fluid

$$\boldsymbol{T} = (\rho + P)\boldsymbol{u}^* \otimes \boldsymbol{u}^* + P\boldsymbol{g} \ . \tag{3.312}$$

In terms of components is a coordinate basis:

$$T_{\mu\nu} = (\rho + P)\, u_{\mu}u_{\nu} + P g_{\mu\nu} \ . \tag{3.313}$$

Note that a perfect fluid may not appear perfect to an observer O not comoving with the fluid. For a perfect fluid of non-relativistic particles sourcing a weak gravitational field:

$$u_{\mu} = (1 + \Phi)\delta_{\mu 0} + v^{i}\delta_{\mu i} \tag{3.314}$$

$$P \ll \rho = O(\Phi) \ , \tag{3.315}$$

so that:

$$T_{00} = \rho \tag{3.316}$$

$$T_{0i} = 0 \tag{3.317}$$

$$T_{ij} = 0 \ . \tag{3.318}$$

3.8.3 Conservation of energy and momentum

In Special Relativity, we know that energy and momentum are conserved, which reads:

$$\frac{\partial T^{\mu}{}_{\nu}}{\partial x^{\mu}} = 0 \ . \tag{3.319}$$

In General Relativity, this must remain true in local inertial frames, so it must be replaced by its covariant form involving the covariant derivative:

$$\nabla_{\mu} T^{\mu}{}_{\nu} = 0 \ . \tag{3.320}$$

3.9 From source to geometry: Einstein field equations

In Newtonian mechanics, the gravitational field, Φ, is determined by its source, namely the mass density, ρ, via Poisson's equation:

$$\Delta\Phi = 4\pi G\rho \ . \tag{3.321}$$

Therefore, if the metric of spacetime is to accurately capture the properties of the gravitational field, we expect it to be linked to the energy-momentum content via a *tensorial* equation:

$$\boldsymbol{E}[\boldsymbol{g}] = \kappa \boldsymbol{T} \ , \tag{3.322}$$

where κ is a constant and $\boldsymbol{E}$ is a $(0,2)$-tensor. The fact that the equation is a relation between tensors ensures that it remains valid in any coordinate system, thus satisfying one of General Relativity's fundamental assumption; this is often called *general covariance*. From our point of view however, this is just the fact that the natural framework to incorporate gravitation to a relativistic theory is the one of spacetime manifolds. We would like Eq. (3.322) to reduce to Eq. (3.321) for a non-relativistic, weak source. Therefore, we can assume that $\boldsymbol{E}$ only depends on $\boldsymbol{g}$ and its first and second derivatives at most. Specifically, it must be a function of $\boldsymbol{g}$ and the Riemann tensor. Moreover, since Eq. (3.320) must be satisfied for any reasonable source, we must have:

$$\nabla_\mu E^\mu{}_\nu = 0 \ . \tag{3.323}$$

Starting from Bianchi identities, Eq. (3.249), it is actually possible to construct such a tensor. It turns out to be unique up to a simple term but we will not attempt to prove this uniqueness here. Let us thus start from Bianchi identities:

$$\nabla_\mu R^\nu{}_{\lambda\rho\sigma} + \nabla_\rho R^\nu{}_{\lambda\sigma\mu} + \nabla_\sigma R^\nu{}_{\lambda\mu\rho} = 0 \ , \tag{3.324}$$

and define the *Ricci tensor* as a symmetric (0,2)-tensor obtained by contracting the Riemann tensor on its first and second entries, or in terms of its components:

$$\boxed{R_{\mu\nu} = R^\rho{}_{\mu\rho\nu}} \ . \tag{3.325}$$

Then, taking a trace of Eq.(3.324) on the second and fourth indices[15]:

$$\nabla_\mu R_{\lambda\sigma} + \nabla_\nu R^\nu{}_{\lambda\sigma\mu} + \nabla_\sigma R^\nu{}_{\lambda\mu\nu} = 0 \ . \tag{3.326}$$

[15]Remember that $\nabla_\alpha g_{\beta\gamma} = 0$ according to metric compatibility, so we are always free to contract inside covariant derivatives and take traces as we wish.

Using that $R^{\nu}{}_{\lambda\mu\nu} = -R^{\nu}{}_{\lambda\nu\mu} = -R_{\lambda\mu}$, we get:

$$\nabla_\mu R_{\lambda\sigma} + \nabla_\nu R^{\nu}{}_{\lambda\sigma\mu} - \nabla_\sigma R_{\lambda\mu} = 0 \, . \tag{3.327}$$

Next, let us contract this equation with $g^{\lambda\sigma}$:

$$\nabla_\mu R + \nabla_\nu \left[g^{\lambda\sigma} R^{\nu}{}_{\lambda\sigma\mu} \right] - g^{\lambda\sigma} \nabla_\sigma R_{\lambda\mu} = 0 \, , \tag{3.328}$$

where we used that $\nabla_\alpha g_{\beta\gamma} = 0$ and we defined the *Ricci scalar* as the trace of the Ricci tensor:

$$\boxed{R = g^{\mu\nu} R_{\mu\nu}} \, . \tag{3.329}$$

The second term in Eq. (3.328) can be simplified by noting that:

$$\begin{aligned} g^{\lambda\sigma} R^{\nu}{}_{\lambda\sigma\mu} =& g^{\lambda\sigma} g^{\nu\rho} R_{\rho\lambda\sigma\mu} && (3.330)\\ =& - g^{\lambda\sigma} g^{\nu\rho} R_{\lambda\rho\sigma\mu} && (3.331)\\ =& - g^{\nu\rho} R^{\sigma}{}_{\rho\sigma\mu} && (3.332)\\ =& - g^{\nu\rho} R_{\rho\mu} \, . && (3.333) \end{aligned}$$

Then, we get:

$$\nabla_\mu R - g^{\nu\rho} \nabla_\nu R_{\rho\mu} - g^{\lambda\sigma} \nabla_\sigma R_{\lambda\mu} = 0 \, , \tag{3.334}$$

or, relabelling dummy indices:

$$\nabla_\mu R - 2 g^{\nu\rho} \nabla_\nu R_{\rho\mu} = 0 \, . \tag{3.335}$$

Thus:

$$g^{\nu\rho} \nabla_\nu R_{\rho\nu} - \frac{1}{2} g_{\mu\rho} g^{\rho\nu} \nabla_\nu R = 0 \, , \tag{3.336}$$

where we used $\delta_\mu{}^\nu = g_{\mu\rho} g^{\rho\nu}$. Finally:

$$g^{\nu\rho} \nabla_\nu \left[R_{\rho\mu} - \frac{1}{2} R g_{\rho\mu} \right] = 0 \, . \tag{3.337}$$

Therefore, the *Einstein tensor*, $\boldsymbol{G}$ defined, in components, by:

$$\boxed{G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}} \tag{3.338}$$

satisfies:

$$\nabla_\mu G^\mu{}_\nu = 0 \,, \tag{3.339}$$

and could be used as the tensor $\boldsymbol{E}$ we were looking for. However, it needs to be supplemented by another terms for completeness. As we have noted several times, $\nabla_\rho g_{\mu\nu} = 0$, so that we can always add a term proportional to $g_{\mu\nu}$ to the Einstein's tensor and obtain a tensor satisfying our requirements, Thus, we will choose:

$$\boldsymbol{E} = \boldsymbol{G} + \Lambda \boldsymbol{g} \,, \tag{3.340}$$

where $\Lambda \in \mathbb{R}$ is a constant known as the *cosmological constant*. It appears as a constant of the theory, to be determined by experiment or observations. Currently, it manifests itself at cosmological scales only and its observed value is:

$$\Lambda \simeq 10^{-52}\mathrm{m}^{-2} \,. \tag{3.341}$$

Because it is so small, it has no discernible effects on length scales much smaller than cosmological ones. This is why we will neglect it everywhere, except in chapter 6. To summarize, we arrived at the following form for our gravitational field equations:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \kappa T_{\mu\nu} \,. \tag{3.342}$$

To complete our task, we need to fix the coupling constant κ. We expect it to be proportional to G, Newton's constant, but we want to determine how exactly. To do that, we impose that these equations lead to Poisson's equation for a static, weak gravitational field generated by a non-relativistic source. In that case:

$$T_{\mu\nu} = \rho \delta_{\mu 0} \delta_{\nu 0} \,, \tag{3.343}$$

and the non-zero components of the Ricci tensor reads:

Ricci tensor in the static, weak field limit

$$R_{00} = \Delta\Phi \tag{3.344}$$

$$R_{ij} = \Delta\Phi \delta_{ij} \,, \tag{3.345}$$

so that the Ricci scalar becomes:

$$R = 2\Delta\Phi \,. \tag{3.346}$$

The Einstein tensor is thus:

$$G_{00} = 2\Delta\Phi \quad \text{and} \quad G_{ij} = G_{0i} = 0 \ . \tag{3.347}$$

Thus, neglecting the cosmological constant, we get[16]:

$$2\Delta\Phi = \kappa \left[c^4\right] \rho \ , \tag{3.348}$$

and thus:

$$\frac{\kappa \left[c^4\right]}{2} = 4\pi G \ , \tag{3.349}$$

i.e.:

$$\kappa = \frac{8\pi G}{c^4} = 8\pi G \ . \tag{3.350}$$

Finally, we get the:

Einstein Field Equations

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \ , \tag{3.351}$$

with:

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \ . \tag{3.352}$$

Note that by taking the trace of this equation we get:

$$-R + 4\Lambda = 8\pi G T \ , \tag{3.353}$$

where $T = T^{\mu}{}_{\mu}$ is the trace of the energy-momentum tensor. Thus, the Einstein field equations are sometimes written in the equivalent form:

$$R_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi G \left[T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right] \ , \tag{3.354}$$

[16] We have re-established factors of c to get units right. In the Poisson equation, ρ is the mass density. Thus, the energy density in the relativistic energy-momentum tensor is $c^2\rho$. Two more factors of c come from the fact that $\Phi \to \Phi/c^2$ in the metric. This can be directly checked form the Einstein field equations. the energy-momentum tensor components have units of energy per unit volume and the LHS has units of inverse area, so the constant κ must have units of length over energy. But G has units of length to the power 5 over energy times time to the four, so it must be divided by c^4 to get units of length over energy.

which is particularly useful in vacuum, when $T_{\mu\nu} = 0$. These Einstein field equations quantify how matter and energy act on the spacetime geometry to generate the metric encoding the gravitational field. Note that by symmetries, they consist in $4\times(4+1)/2 = 10$ coupled partial differential equations for the 10 metric independent components. Of course, a choice of arbitrary coordinate system can fix 4 of these metric components, so we are left with 10 coupled equations for 6 actual degrees of freedom. This simply means that some of these equations can be interpreted as constraints. However, this system is so complex that, in general, and without any assumptions on the symmetries of the system and the nature of the sources, there is little hope to arrive at any practical solution. The following chapters will all be devoted to such assumptions in various physically relevant contexts.

4

Stars and black holes: the Schwarzschild solution

Contents

4.1 Introduction

The simplest way to illustrate the link between geometry and matter that underpins general relativity and allowed us to build the theory in chapter 3 is to study the gravitational field generated by a highly symmetrical distribution of matter: that of a perfectly spherical, isolated clump of non-relativistic matter, i.e. a ball of mass. This is certainly the simplest model we can build of astrophysical objects such as stars and planets.

A few things need to be understood from the start. We are going to review the approach to be followed in the Newtonian context first, so that we will be able to contrast with the relativistic case later. The spherical distribution of mass plays the role of a *source* of the gravitational field. If we choose the origin of the coordinate system at the centre of the distribution and we use spherical coordinates, we can describe that source by a density:

$$\rho(t,r) = \rho(r)\Theta(r - R(t)) \ , \tag{4.1}$$

where $\rho(r)$ describes the profile of density and depends only on r by symmetry. $\Theta(u)$ is the Heaviside functions which is 1 for $u \leq 0$ and 0 otherwise. The function $R(t)$ is the radius of the object at time t; by giving it a time dependence, we allow for the object to "pulse", as long as it does so while keeping its spherical symmetry. What we will focus on is the field generated by this source outside the source itself, i.e. *in vacuum*: $r > R(t)$. In Newtonian physics, this amounts to solving the Laplace equation:

$$\Delta\Phi(t,r,\theta,\phi) = 0 \ , \tag{4.2}$$

for the gravitational potential $\Phi(t,r,\theta,\phi)$. Actually, by symmetry, we must have $\phi = \phi(t,r)$ and the Laplacian thus simplifies, so that Eq. (4.2) becomes:

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2\frac{\partial\Phi}{\partial r}\right] = 0 \ . \tag{4.3}$$

This can be straightforwardly solved to obtain:

$$\Phi(t,r) = -\frac{C_1(t)}{r} + C_2(t) \ , \tag{4.4}$$

where $C_1(t)$ and $C_2(t)$ are arbitrary functions of time. They need to be set by imposing some boundary conditions. Using the fact that the action of the mass distribution should fade at infinity:

$$\lim_{r\to+\infty}\Phi(t,r) = 0 \ , \tag{4.5}$$

we impose $C_2(t) = 0$. To fix $C_1(t)$, we need to examine the other boundary, that is, the surface of the star: $r = R(t)$. Inside the star, the gravitational field obeys the Poisson equation:

$$\Delta\Phi = 4\pi G\rho \ . \tag{4.6}$$

Using spherical symmetry, we can get that in the form:

$$\frac{\partial\Phi}{\partial r} = \frac{GM(<r)}{r^2} \ , \tag{4.7}$$

where $M(<r)$ is the mass enclosed in a sphere of radius $r < R(t)$:

$$M(<r) = 4\pi \int_0^r u^2\rho(u)\mathrm{d}u \ . \tag{4.8}$$

Differentiating Eq. (4.4) with respect to r and evaluating the result at $r = R(t)$, we get:

$$\frac{\partial\Phi}{\partial r}(t, R(t)) = \frac{C_1(t)}{R^2(t)} \ , \tag{4.9}$$

while the same quantity obtained from Eq. (4.7) gives:

$$\frac{\partial\Phi}{\partial r}(t, R(t)) = \frac{GM}{R^2(t)} \ , \tag{4.10}$$

where $M = M(<R)$ is the total mass of the star. Thus, we find that $C_1(t) = GM$. This is a simplified version of *Gauss's theorem*: the potential outside a spherical gravitational object only depends on the total mass of the object:. The form of the potential is given by:

$$\Phi(t, r) = -\frac{GM}{r} \ . \tag{4.11}$$

It is interesting to note that we did not need to suppose that the source was static. As long as it pulses while retaining its symmetry, the potential outside the source remains time-independent.
We are now going to apply the same strategy that we deployed above to tackle the same problem in General Relativity. Let us consider the spacetime around a spherically symmetric distribution of mass. What is its geometry? To describe the system, we are going to make a certain number of assumptions:

(a) Asymptotic flatness: We assume that the system is isolated so that far enough from the source (in a way that will be made clear later), the geometry of spacetime is that of Minkowski spacetime, i.e. the one we get in absence of any gravitational field.

(b) Vacuum solution: We concentrate on the geometry outside the sources: $T_{\mu\nu} = 0$.

(c) Spherical symmetry: By the principle according to which the symmetries of the sources dictate the symmetries of the solution, we will assume that the spacetime is *spherically symmetric*.

The last assumption leads us to choose a coordinate chart outside the star that respects the symmetries.

One can show that, by an appropriate choice of coordinates, the most general spherically symmetric spacetime can always be decomposed (foliated) in a sequence of spacelike hypersurfaces labelled by a time coordinate $t \in \mathbb{R}$, that are themselves foliated by concentric spheres of radii $r \in \mathbb{R}$; see Fig. 4.1.

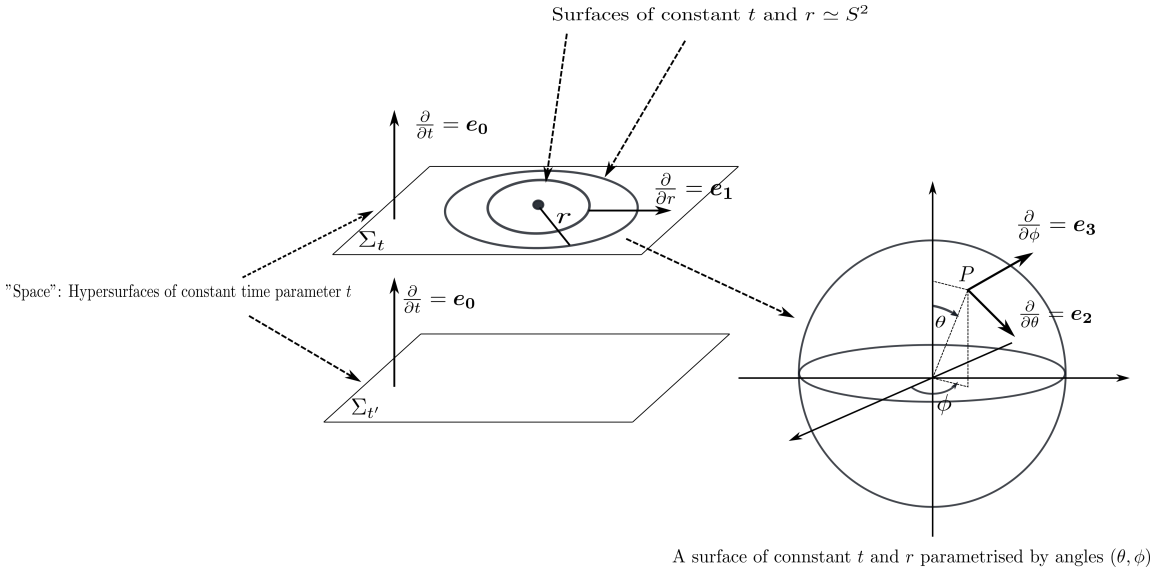


Figure 4.1: The chart (t, r, θ, ϕ) adapted to spherically symmetric spacetime and the associated foliation. The black thick dot represents the centre of symmetry at the time t. Any event in spacetime P sits on a 2-sphere of radius r at fixed time t that is embedded in an hypersurfaces at constant t orthogonal to $\frac{\partial}{\partial t}$. It is then labelled by its the angular position (θ, ϕ) on that sphere.

The metric then takes the form:

$$\boldsymbol{g} = -\mathrm{e}^{\nu(t,r)}\mathrm{d}t \otimes \mathrm{d}t + \mathrm{e}^{\lambda(t,r)}\mathrm{d}r \otimes \mathrm{d}r + r^2\left[\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta\,\mathrm{d}\phi \otimes \mathrm{d}\phi\right] , \tag{4.12}$$

and the line element reads:

$$\mathrm{d}s^2 = -\mathrm{e}^{\nu(t,r)}\mathrm{d}t^2 + \mathrm{e}^{\lambda(t,r)}\mathrm{d}r^2 + r^2\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right] \ , \tag{4.13}$$

where $\nu(t,r)$ and $\lambda(t,r)$ are arbitrary functions. The exponential form used here is arbitrary but it serves to remember the proper signs of the metric components and it will simplify calculations further down. Proofs of this fact can be found in [21] and in box 23.3 of [16]. In the language of appendix B, the space at constant t is isotropic around the centre but *not* homogeneous. Here $(\theta,\phi) \in [0,\pi)\times[0,2\pi)$ are the usual angles on the sphere S^2, r is the area distance, i.e. the distance such that the area of the sphere at constant t and r is given by $4\pi r^2$.

4.2 Spacetime outside a spherical star: The Schwarzschild solution

4.2.1 Solution to the Einstein field equations

We can now use assumption (b) above to solve the Einstein field equations for the metric (4.12). We aim to solve the vacuum equations which can always be written as:

$$R_{\mu\nu} = 0 \ . \tag{4.14}$$

We first need to compute the connection coefficients for the metric 4.12. The only ones that are non-zero are:

Connection coefficients for the metric (4.12)

$$\Gamma^0{}_{00} = \frac{1}{2}\frac{\partial \nu}{\partial t} \; ; \; \Gamma^0{}_{01} = \Gamma^0{}_{10} = \frac{1}{2}\frac{\partial \nu}{\partial r} \; ; \; \Gamma^0{}_{11} = \frac{e^{\lambda-\nu}}{2}\frac{\partial \lambda}{\partial t} \; ; \tag{4.15}$$

$$\Gamma^1{}_{00} = \frac{e^{\nu-\lambda}}{2}\frac{\partial \nu}{\partial r} \; ; \; \Gamma^1{}_{11} = \frac{1}{2}\frac{\partial \lambda}{\partial r} \; ; \; \Gamma^1{}_{01} = \Gamma^1{}_{10} = \frac{1}{2}\frac{\partial \lambda}{\partial t} \; ; \tag{4.16}$$

$$\Gamma^1{}_{22} = -re^{-\lambda} \; ; \; \Gamma^1{}_{33} = -re^{-\lambda}\sin^2\theta \; ; \tag{4.17}$$

$$\Gamma^2{}_{12} = \Gamma^2{}_{21} = \frac{1}{r} \; ; \; \Gamma^2{}_{33} = -\sin\theta\cos\theta \tag{4.18}$$

$$\Gamma^3{}_{31} = \Gamma^3{}_{13} = \frac{1}{r} \; ; \; \Gamma^3{}_{32} = \Gamma^3{}_{23} = \frac{\cos\theta}{\sin\theta} \; . \tag{4.19}$$

Using these, we obtain the non-zero components of the Ricci tensor:

Components of the Ricci tensor for the metric (4.12)

$$R_{00} = -\frac{1}{2}\left[\frac{\partial^2\lambda}{\partial t^2} + \frac{1}{2}\left(\frac{\partial \lambda}{\partial t}\right)^2 - \frac{1}{2}\frac{\partial \nu}{\partial t}\frac{\partial \lambda}{\partial t}\right] + \frac{e^{\nu-\lambda}}{2}\left[\frac{\partial^2\nu}{\partial r^2} + \frac{1}{2}\left(\frac{\partial \nu}{\partial r}\right)^2 - \frac{1}{2}\frac{\partial \nu}{\partial r}\frac{\partial \lambda}{\partial r} + \frac{2}{r}\frac{\partial \nu}{\partial r}\right] \; ; \tag{4.20}$$

$$R_{11} = -\frac{1}{2}\left[\frac{\partial^2\nu}{\partial r^2} + \frac{1}{2}\left(\frac{\partial \nu}{\partial r}\right)^2 - \frac{1}{2}\frac{\partial \lambda}{\partial r}\frac{\partial \nu}{\partial r} - \frac{2}{r}\frac{\partial \lambda}{\partial r}\right] + \frac{e^{\lambda-\nu}}{2}\left[\frac{\partial^2\lambda}{\partial t^2} + \frac{1}{2}\left(\frac{\partial \lambda}{\partial t}\right)^2 - \frac{1}{2}\frac{\partial \nu}{\partial t}\frac{\partial \lambda}{\partial t}\right] \; ; \tag{4.21}$$

$$R_{01} = R_{10} = \frac{1}{r}\frac{\partial \lambda}{\partial t} \; ; \; R_{22} = 1 - e^{-\lambda}\left[1 + \frac{r}{2}\left(\frac{\partial \nu}{\partial r} - \frac{\partial \lambda}{\partial r}\right)\right] \; ; \; R_{33} = \sin^2\theta R_{22} \; . \tag{4.22}$$

From $R_{01} = 0$, we can then obtain simply:

$$\lambda = \lambda(r) \; . \tag{4.23}$$

Then, we form:

$$R_{00} + \mathrm{e}^{\nu-\lambda} R_{11} = 0 \,, \tag{4.24}$$

to get:

$$\frac{\partial \nu}{\partial r} + \frac{\mathrm{d}\lambda}{\mathrm{d}r} = 0 \,, \tag{4.25}$$

so that:

$$\nu(t, r) = -\lambda(r) + f(t) \,, \tag{4.26}$$

for an arbitrary function $f(t)$. Next, we use that $R_{22} = 0$ to get:

$$\mathrm{e}^{-\lambda} \left[1 - r \frac{\mathrm{d}\lambda}{\mathrm{d}r} \right] = 1 \,, \tag{4.27}$$

which can be rewritten:

$$\frac{\mathrm{d}}{\mathrm{d}r} \left(r \mathrm{e}^{-\lambda(r)} \right) = 1 \,, \tag{4.28}$$

so that:

$$\mathrm{e}^{-\lambda} = 1 + \frac{C}{r} \quad \text{for } C \in \mathbb{R} \,. \tag{4.29}$$

To conclude the calculation, we need to fix the function $f(t)$. This can be done by using assumption (a) above. From Eq. (4.29) we get:

$$\lim_{r \to +\infty} \lambda(r) = 0 \,, \tag{4.30}$$

which implies that:

$$\lim_{r \to +\infty} \nu(t, r) = f(t) \,. \tag{4.31}$$

Then, asymptotic flatness imposes that:

$$\lim_{r \to +\infty} \nu(t, r) = f(t) = 0 \,. \tag{4.32}$$

Then we arrive at a family of solutions indexed by an arbitrary constant $C \in \mathbb{R}$:

$$\boldsymbol{g} = -\left(1 + \frac{C}{r} \right) \mathrm{d}t \otimes \mathrm{d}t + \left(1 + \frac{C}{r} \right)^{-1} \mathrm{d}r \otimes \mathrm{d}r + r^2 \left[\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2 \theta \, \mathrm{d}\phi \otimes \mathrm{d}\phi \right] \,. \tag{4.33}$$

Note that we obtained *static* solutions although the source is not necessarily static: as long as it stays spherically symmetric, the source could evolve and the metric outside would remain static[1].

[1] What do we mean by static here? We mean that it has a timelike Killing vector field, $\frac{\partial}{\partial t}$ here; see appendix B

4.2.2 Birkhoff-Jebsen theorem

To fully characterise the field around a spherical star, we still have to find a way to fix C. At the very least, far from the star, when the field is weak, we expect to recover Newton's law of motion for a massive test-particle. At first order in $1/r$, the metric becomes:

$$\boldsymbol{g} \simeq -\left(1+\frac{C}{r}\right)\mathrm{d}t\otimes\mathrm{d}t+\left(1-\frac{C}{r}\right)\mathrm{d}r\otimes\mathrm{d}r+r^2\left[\mathrm{d}\theta\otimes\mathrm{d}\theta+\sin^2\theta\,\mathrm{d}\phi\otimes\mathrm{d}\phi\right]\ . \tag{4.34}$$

Let us consider a test-particle of mass m and 4-velocity $\boldsymbol{u} = u^\mu \boldsymbol{e}_{(\mu)}$. At first order, we develop $u^\mu = \bar{u}^\mu + \delta u^\mu$ where $\bar{u}^\mu$ are the components of the 4-velocity in absence of any gravitational field from the star , and $\delta u^\mu = O(1/r)$ is the perturbation due to the presence of the star. Then, we expand at first order the geodesic equation:

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0 \ \Leftrightarrow\ u^\alpha\partial_\alpha u^\mu + \Gamma^\mu{}_{\alpha\beta}u^\alpha u^\beta = 0\ . \tag{4.35}$$

At dominant order, we recover the equation of a straight line in Minkowski spacetime, expressed in spherical coordinates. If we start at rest, we stay at rest since we are in Minkowski spacetime. Since we focus on motion caused by the presence of the massive star, we can thus assume that $\bar{u}^\mu = \delta^\mu_0$. The first order, perturbative equation of motion then becomes:

$$\frac{\mathrm{d}\delta u^\mu}{\mathrm{d}t} + \delta\Gamma^\mu{}_{00} = 0\ . \tag{4.36}$$

Since:

$$\delta\Gamma^\mu{}_{00} \simeq -\frac{C}{2r^2}\delta^{\mu 1}\ , \tag{4.37}$$

we get a radial motion with:

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{C}{2r^2}\ . \tag{4.38}$$

To recover Newton's second law, we must thus set $C = -2GM$ where M is the *total mass* of the star. The quantity:

$$\boxed{R_S = 2GM\left[=\frac{2GM}{c^2}\right]}\ , \tag{4.39}$$

is called the *Schwarzschild radius* of the star. It is the only free parameter entering the metric generated in vacuum by a spherical object. There is no time-dependence left and given the mass of the object, the metric is fully determined as long as it stays spherical.

We have proven the:

Birkhoff-Jebsen Theorem

There exists a unique spherically symmetric, vacuum and asymptotically flat solution to the Einstein field equations. This solution is static and is given by the *Schwarzschild geometry*:

$$\boldsymbol{g} = -\left(1 - \frac{R_S}{r}\right) \mathrm{d}t \otimes \mathrm{d}t + \left(1 - \frac{R_S}{r}\right)^{-1} \mathrm{d}r \otimes \mathrm{d}r + r^2 \left[\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta \, \mathrm{d}\phi \otimes \mathrm{d}\phi\right] \ . \tag{4.40}$$

The coordinates (t, r, θ, ϕ) are known as the *Schwarzschild coordinates*. Note that the departure from Minkowski spacetime in the metric coefficients is simply:

$$-\frac{R_S}{r} = 2\Phi(r) \ , \tag{4.41}$$

i.e. twice the Newtonian potential.

Equivalently, we get the *Schwarzschild line element*:

$$\mathrm{d}s^2 = -\left(1 - \frac{R_S}{r}\right) \mathrm{d}t^2 + \left(1 - \frac{R_S}{r}\right)^{-1} \mathrm{d}r^2 + r^2 \left[\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right] \tag{4.42}$$

$$= -\left(1 + 2\Phi(r)\right) \mathrm{d}t^2 + \left(1 + 2\Phi(r)\right)^{-1} \mathrm{d}r^2 + r^2 \left[\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi^2\right] \ . \tag{4.43}$$

This line element is singular for $r = 0$ and for $r = R_S$. We will see in section 4.6, that the first singularity is serious while the other one is not. But in what follows we first want to use Eq. (4.42) to describe the geometry outside a star so we have to check that these singularities are not a problem for us. Clearly, $r = 0$ is inside the star and as such, is not a covered by the geometry given by Eq. (4.40), which describes the vacuum region outside the star. On the other hand, for a star of mass M, the Schwarzschild radius can be expressed as:

$$R_S \simeq 3 \frac{M}{M_\odot} \ \mathrm{km} \ , \tag{4.44}$$

to be compared with the Sun's radius: $R_\odot \simeq 7 \times 10^5$ km. For a stellar object, the Schwarzschild radius is always much smaller than the size of the object and the coordinate singularity of the Schwarzchild metric at $r = R_S$ is irrelevant as far as describing the geometry outside the star is concerned. In what follows we will always restrict ourselves to $r > R_S$. We will allow for values of r close to R_S to illustrate some important relativistic effects when discussion geodesics, although

such effects would not be present for realistic Solar-type stars, because they become important for compact starts such as neutron stars. In section 4.6 we will explore what happens if we remove the central object and continue the Schwarzschild geometry past the hypersurface $r = R_S$ to discover a new class of objects: black holes.

4.3 Geodesics of the Schwarzschild geometry

4.3.1 Geodesic equations in Schwarzschild coordinates

To explore the physics associated to a geometry, the best way is to study how test particles in free fall behave in that geometry. This means one has to study the timelike and lightlike geodesics. Let us remember that these are characterised by the *geodesic equation*:

$$\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0 \;\Leftrightarrow\; \frac{\mathrm{d}u^\mu}{\mathrm{d}\lambda} + \Gamma^\mu{}_{\nu\rho}u^\mu u^\rho = 0\,, \tag{4.45}$$

where λ is an affine parameter such that $\mathrm{d}\lambda = -u_\mu \mathrm{d}x^\mu$. Moreover:

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\,, \tag{4.46}$$

are the components of the components of the vector field tangent to the geodesics given parametrically by $x^\mu(\lambda)$. The nature of the geodesics is determined by:

$$\boldsymbol{g}(\boldsymbol{u},\boldsymbol{u}) = g_{\mu\nu}u^\mu u^\nu = \varepsilon\,, \tag{4.47}$$

with $\varepsilon = -1$ for a timelike geodesics with the parameter chosen to be the proper time along the geodesic ($\lambda = \tau$) and $\varepsilon = 0$ for a lightlike one.
Let us first rewrite the connection coefficients (4.15)-(4.19) for the Schwarzschild metric (4.40):

Connection coefficients for the Schwarzschild metric

$$\Gamma^0{}_{01} = \Gamma^0{}_{10} = \frac{R_S}{2r(r-R_S)} \; ; \tag{4.48}$$

$$\Gamma^1{}_{00} = \frac{R_S(r-R_S)}{2r^3} \; ; \; \Gamma^1{}_{11} = -\frac{R_S}{2r(r-R_S)} \; ; \tag{4.49}$$

$$\Gamma^1{}_{22} = R_S - r \; ; \; \Gamma^1{}_{33} = (R_S - r)\sin^2\theta \; ; \tag{4.50}$$

$$\Gamma^2{}_{12} = \Gamma^2{}_{21} = \frac{1}{r} \; ; \; \Gamma^2{}_{33} = -\sin\theta\cos\theta \tag{4.51}$$

$$\Gamma^3{}_{12} = \Gamma^3{}_{31} = \frac{1}{r} \; ; \; \Gamma^3{}_{23} = \Gamma^3{}_{32} = \frac{\cos\theta}{\sin\theta} \; . \tag{4.52}$$

Then, the geodesic equations read:

$$\begin{cases} \dfrac{\mathrm{d}^2t}{\mathrm{d}\lambda^2} + \dfrac{R_S}{r(r-R_S)}\dfrac{\mathrm{d}t}{\mathrm{d}\lambda}\dfrac{\mathrm{d}r}{\mathrm{d}\lambda} = 0 & (4.53) \\ \dfrac{\mathrm{d}^2r}{\mathrm{d}\lambda^2} + \dfrac{R_S}{2r^3}\left(\dfrac{\mathrm{d}t}{\mathrm{d}\lambda}\right)^2 - \dfrac{R_S}{2r(r-R_S)}\left(\dfrac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 - (r-R_S)\left[\left(\dfrac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right)^2 + \sin^2\theta\left(\dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2\right] = 0 & (4.54) \\ \dfrac{\mathrm{d}^2\theta}{\mathrm{d}\lambda^2} + \dfrac{2}{r}\dfrac{\mathrm{d}r}{\mathrm{d}\lambda}\dfrac{\mathrm{d}\theta}{\mathrm{d}\lambda} - \sin\theta\cos\theta\left(\dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right) = 0 & (4.55) \\ \dfrac{\mathrm{d}^2\phi}{\mathrm{d}\lambda^2} + \dfrac{2}{r}\dfrac{\mathrm{d}r}{\mathrm{d}\lambda}\dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda} + \dfrac{2\cos\theta}{\sin\theta}\dfrac{\mathrm{d}\theta}{\mathrm{d}\lambda}\dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda} = 0 \; . & (4.56) \end{cases}$$

4.3.2 Conserved quantities

These equations are highly coupled and there is no hope that we will be able to solve them exactly. However, we can make some progress by looking for *conserved quantities*. This can be done by looking for combinations of Eqs. (4.53)-(4.56) that can be written as total derivatives. In a more clever way, we can also use *Killing vector fields* of the geometry. From appendix B, we know that a vector field $\boldsymbol{\xi}$ is a Killing vector field, i.e. that it generates a local isometry of spacetime iff it

satisfies the Killing equation:

$$\nabla_{(\mu}\xi_{\nu)} = 0 \,. \tag{4.57}$$

Consider such a Killing vector $\boldsymbol{\xi}$ and a geodesics with tangent vector $\boldsymbol{u}$. Then:

$$\begin{aligned}
\boldsymbol{\nabla_u}\left(\boldsymbol{g}(\boldsymbol{\xi},\boldsymbol{u})\right) &= u^\alpha \nabla_\alpha \left[g_{\beta\gamma}\xi^\beta u^\gamma\right] && (4.58)\\
&= u^\alpha \nabla_\alpha \left[\xi_\beta u^\beta\right] && (4.59)\\
&= \left(u^\alpha \nabla_\alpha \xi_\beta\right) u^\beta + \xi_\beta \underbrace{u^\alpha \nabla_\alpha u^\beta}_{=0} && (4.60)\\
&= \frac{1}{2}\left(\nabla_\alpha \xi_\beta + \nabla_\beta \xi_\alpha\right) u^\alpha u^\beta \quad \text{(symmetry } \alpha \leftrightarrow \beta\text{)} && (4.61)\\
&= \nabla_{(\alpha}\xi_{\beta)} u^\alpha u^\beta = 0 \,. && (4.62)
\end{aligned}$$

Thus we have the result:

Conserved quantities

If $\boldsymbol{\xi}$ is a Killing vector field for the geometry, then $\boldsymbol{g}(\boldsymbol{\xi},\boldsymbol{u})$ is conserved along the geodesic tangent to $\boldsymbol{u}$.

In the case of the Schwarzschild metric, we have 4 Killing vector fields:

- a timelike Killing vector field $\boldsymbol{\xi_{(0)}} = \boldsymbol{e_{(0)}} = \frac{\partial}{\partial t}$ which can be found by noticing that the metric components in the coordinate system (t, r, θ, ϕ) does not depend on t; see appendix B;
- three spacelike Killing vector fields coming from spherical symmetry of space; see appendix B. These three vectors are:

$$\begin{cases}
\boldsymbol{\xi_{(1)}} = \sin\phi \dfrac{\partial}{\partial\theta} + \dfrac{\cos\theta\cos\phi}{\sin\theta}\dfrac{\partial}{\partial\phi} \,; & (4.63)\\
\boldsymbol{\xi_{(2)}} = \cos\phi \dfrac{\partial}{\partial\theta} - \dfrac{\cos\theta\sin\phi}{\sin\theta}\dfrac{\partial}{\partial\phi} \,; & (4.64)\\
\boldsymbol{\xi_{(3)}} = \dfrac{\partial}{\partial\phi} \,. & (4.65)
\end{cases}$$

The conservation of $\boldsymbol{g}(\boldsymbol{\xi_{(0)}},\boldsymbol{u})$ along the geodesics gives us an analogue to the conservation of energy, while the conservation of $\boldsymbol{g}(\boldsymbol{\xi_{(3)}},\boldsymbol{u})$ provides a conservation akin to the one of angular momentum. Thus, we find the following conserved quantities in the Schwarzschild geometry:

$$\begin{cases} e = \left(1 - \dfrac{R_S}{r}\right) \dfrac{\mathrm{d}t}{\mathrm{d}\lambda} & (4.66) \\ l = r^2 \sin^2\theta \dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda} & (4.67) \end{cases}$$

The other two Killing vector fields can be combined to get:

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{1}{r^2}\left[\cos\phi \boldsymbol{g}(\boldsymbol{\xi_{(2)}}, \boldsymbol{u}) + \sin\phi \boldsymbol{g}(\boldsymbol{\xi_{(1)}}, \boldsymbol{u})\right] \ . \tag{4.68}$$

If we start initially with $\theta(\lambda_i) = \theta_i$ and $\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}(\lambda_i) = 0$, then Eq. (4.68) must be satisfied for any value of ϕ, which implies that $\boldsymbol{g}(\boldsymbol{\xi_{(2)}}, \boldsymbol{u}) = \boldsymbol{g}(\boldsymbol{\xi_{(1)}}, \boldsymbol{u}) = 0$. These two quantities are conserved along the geodesics and must remain zero. Therefore:

$$\forall\lambda \in \mathbb{R}\ , \ \frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = 0\ , \ \text{i.e.}\ \forall\lambda \in \mathbb{R}\ , \ \theta = \theta_i\ . \tag{4.69}$$

Note that, using Eq. (4.55) and Eq. (4.67) , this is only possible for $\theta_i = \pi/2$, but one can always rotate the coordinate system to satisfy this condition. We see that the motion happens *in a plane*, which we can take to be the plane $\theta = \pi/2$. In that case, the conserved quantities become:

$$\begin{cases} e = \left(1 - \dfrac{R_S}{r}\right) \dfrac{\mathrm{d}t}{\mathrm{d}\lambda} & (4.70) \\ l = r^2 \dfrac{\mathrm{d}\phi}{\mathrm{d}\lambda}\ , & (4.71) \end{cases}$$

and the geodesic equations read:

$$\begin{cases} \dfrac{\mathrm{d}^2 t}{\mathrm{d}\lambda^2} + \dfrac{eR_S}{(r - R_S)^2} \dfrac{\mathrm{d}r}{\mathrm{d}\lambda} = 0 & (4.72) \\ \dfrac{\mathrm{d}^2 r}{\mathrm{d}\lambda^2} - \dfrac{R_S}{2r(r - R_S)} \left(\dfrac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + \dfrac{e^2 R_S}{2r(r - R_S)^2} - \dfrac{(r - R_S)l^2}{r^4} = 0 & (4.73) \\ \dfrac{\mathrm{d}^2\phi}{\mathrm{d}\lambda^2} + \dfrac{2l}{r^3} \dfrac{\mathrm{d}r}{\mathrm{d}\lambda} = 0\ . & (4.74) \end{cases}$$

Besides, using $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = \varepsilon$ gives:

Master equation for geodesic motion

$$\mathcal{E} \equiv \frac{e^2+\varepsilon}{2} = \frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + V_{\rm eff}(r)\ , \tag{4.75}$$

with the *effective one-dimensional potential*:

$$V_{\rm eff}(r) = \varepsilon\frac{R_S}{2r} + \frac{l^2}{2r^2} - \frac{R_S l^2}{2r^3}\ . \tag{4.76}$$

The master equation (4.75) allows one to analyse the motion purely in terms of the radial coordinate r. Formally, it has the form of an equation of conservation of the total energy per unit mass $\mathcal{E}$ for a test-particle of velocity $\frac{\mathrm{d}r}{\mathrm{d}\lambda}$ in the one-dimensional potential $V_{\rm eff}(r)$. Once properties of $r(\lambda)$ are determined by using this "energy method", then $\phi(\lambda)$ can, in principle, be obtained by solving:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\lambda} = \frac{l}{r^2(\lambda)}\ . \tag{4.77}$$

Trajectories are characterised by two dimensionless constants, $\mathcal{E}$ and $\mu = l/R_S$ which will be key to our subsequent analysis.

4.4 Motion of massive bodies around a spherical star

4.4.1 General properties of trajectories

Let us first illustrate the power of the "energy method" above for timelike geodesics. In that case, $\varepsilon = -1$, $\lambda = \tau$, the proper time along the geodesics, and the effective potential reads:

$$V_{\rm eff}(r) = \underbrace{\underbrace{\underbrace{-\frac{GM}{r}}_{\text{Newtonian potential}} + \underbrace{\frac{l^2}{2r^2}}_{\text{centrifugal barrier}}}_{\text{Newtonian part}} - \underbrace{\frac{R_S l^2}{2r^3}}_{\text{GR effect}}}\ . \tag{4.78}$$

It has three distinct contribution:

- the attractive Newtonian gravitational potential $-GM/r$ which dominates for large values of r;

- a centrifugal term $l^2/2r^2$ which is always repulsive and is also present in Newtonian mechanics;

- a new, general relativistic term, $-R_S l^2/2r^3$ which becomes important for small values of r.

This third term is entirely responsible for deviations from the Keplerian trajectories of Newtonian mechanics. We represent this potential for a few values of the ratio $\mu = l/R_S$ on Fig. 4.2 where we already see a very important new feature: for small values of $r \gtrsim R_S$, the relativistic term dominates over the centrifugal one and the *relativistic potential barrier has a finite height*; we will talk about consequences of this fact in a moment. In terms of the variable $\hat{r} = r/R_S$, the potential becomes:

$$V_{\text{eff}}(\hat{r}) = -\frac{1}{2\hat{r}} + \frac{\mu^2}{2\hat{r}^2} - \frac{\mu^2}{2\hat{r}^3} \ . \tag{4.79}$$

Thus, the potential has extrema iff:

$$\frac{\partial V_{\text{eff}}}{\partial \hat{r}}(\hat{r}_{\text{ext}}) = 0 \Rightarrow \hat{r}_{\text{ext}}^2 - 2\mu^2 \hat{r}_{\text{ext}} + 3\mu^2 = 0 \ . \tag{4.80}$$

The discriminant of this polynomial is simply:

$$\Delta = 4\mu^2 \left(\mu^2 - 3\right) \ . \tag{4.81}$$

This results in three possible configurations:

1. If $l < R_S\sqrt{3}$, then the potential does not have any extremum. It is a monotonously increasing function of r. The particle's angular momentum is not sufficient to keep the mass away from the star's attraction and whatever its "total energy" $\mathcal{E}$, it ultimately falls onto the central star. The bound orbits of Newtonian mechanics disappear due to the presence of the general relativistic term which dominates for small values of r. This is illustrated on Fig. 4.3 and Fig. 4.4. The motion happens along horizontal lines at $\mathcal{E} = \text{cst}$ which must necessarily obey:

$$\mathcal{E} \geq V_{\text{eff}}(r(\lambda)) \ . \tag{4.82}$$

   Since $V_{\text{eff}} < 0$, all trajectories with $\mathcal{E} > 0$ are free: the particle can either fall onto the central star or evade it. On the other hand, all trajectories with $\mathcal{E} < 0$ are bounded and actually fall onto the central star.

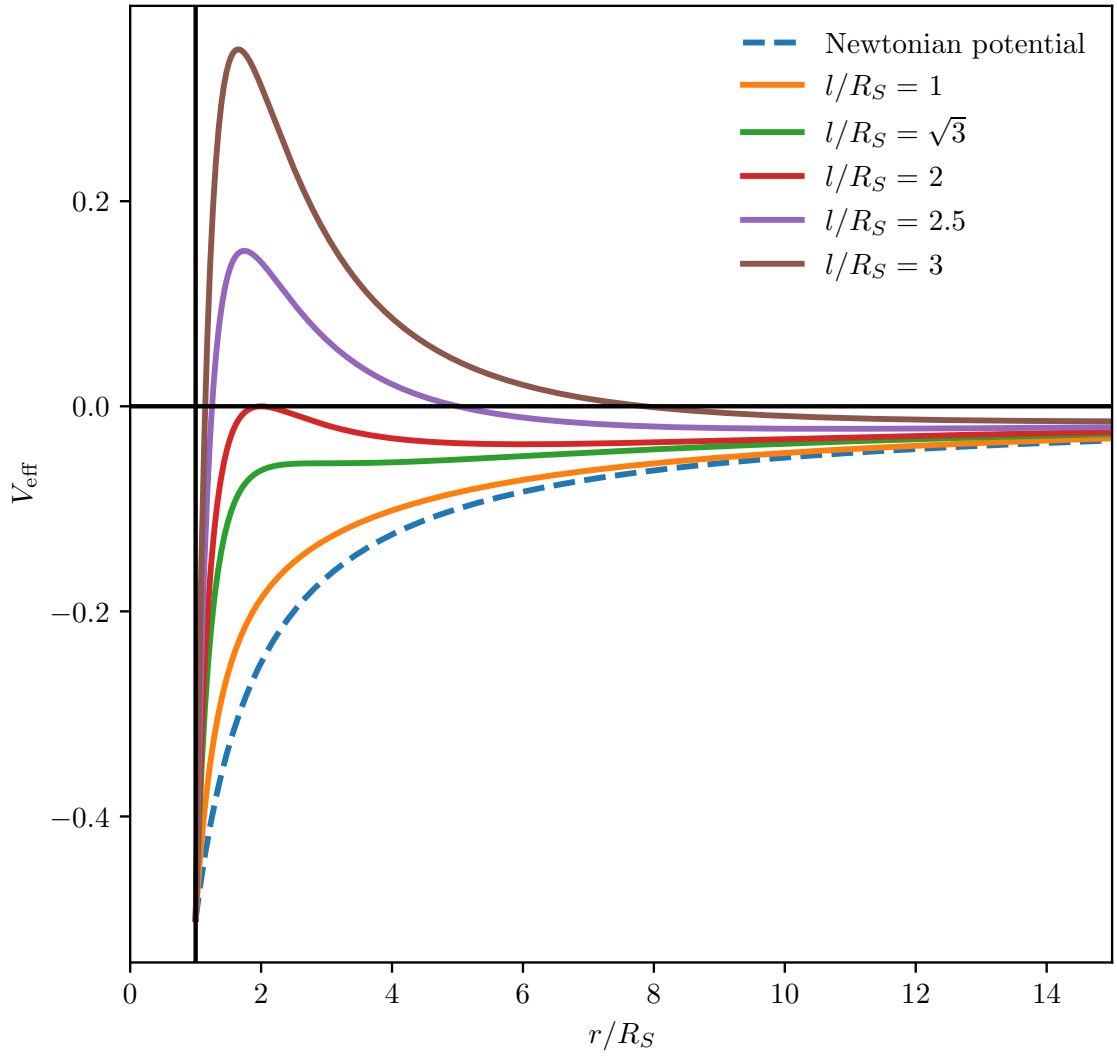


Figure 4.2: The potential V_{eff} from Eq. (4.78) for a few values of $\mu = l/R_S$ as a function of $\hat{r} = r/R_S$.

2. If $l = R_S\sqrt{3}$, there is a unique extremum at $r_{\text{ISCO}} = 3R_S$. It is attained for $\mathcal{E}_{\text{ISCO}} = V_{\text{ISCO}}$, i.e. $r(\lambda) = r_{\text{ISCO}}$ constant. This corresponds to a "marginally" stable circular orbit ($V''_{\text{eff}}(r_{\text{ISCO}}) = 0$) and is called the Innermost Stable Circular Orbit or ISCO because no object can be maintained on a circular orbit at a distance $r < r_{\text{ISCO}}$ (see below). For negative $\mathcal{E} \neq \mathcal{E}_{\text{ISCO}}$, particles fall onto the central star; see Fig. 4.5.

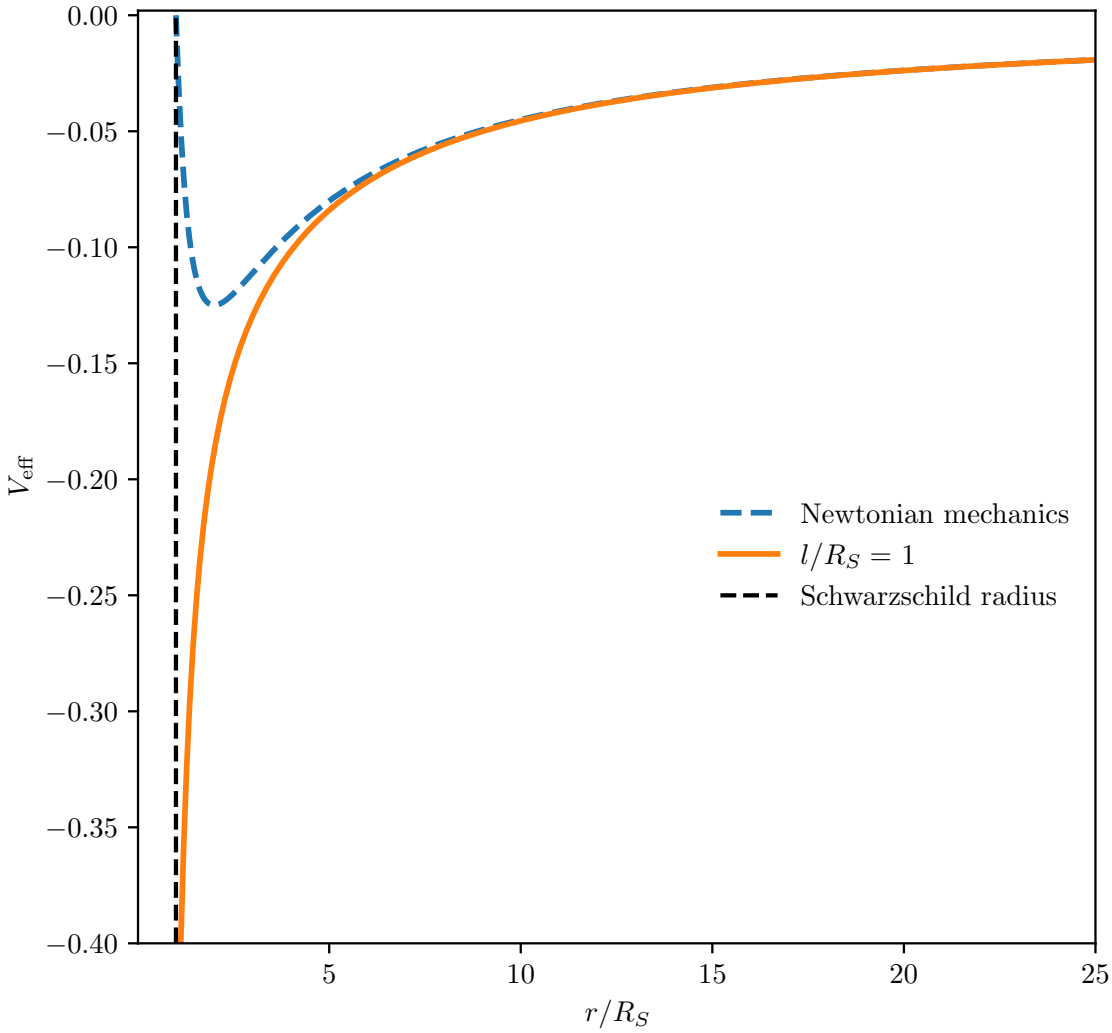


Figure 4.3: The potential V_{eff} from Eq. 4.78 for $\mu = l/R_S = 1 < \sqrt{3}$ as a function of $\hat{r} = r/R_S$. The relativistic term dominates at small r and prevents the existence of the usual bounded Newtonian potential.

3. If $l > R_S\sqrt{3}$, then the potential has two distinct extrema:

   - a minimum at:

$$r_{\text{min}} = R_S\left[\mu^2 + \mu\sqrt{\mu^2 - 3}\right] . \tag{4.83}$$

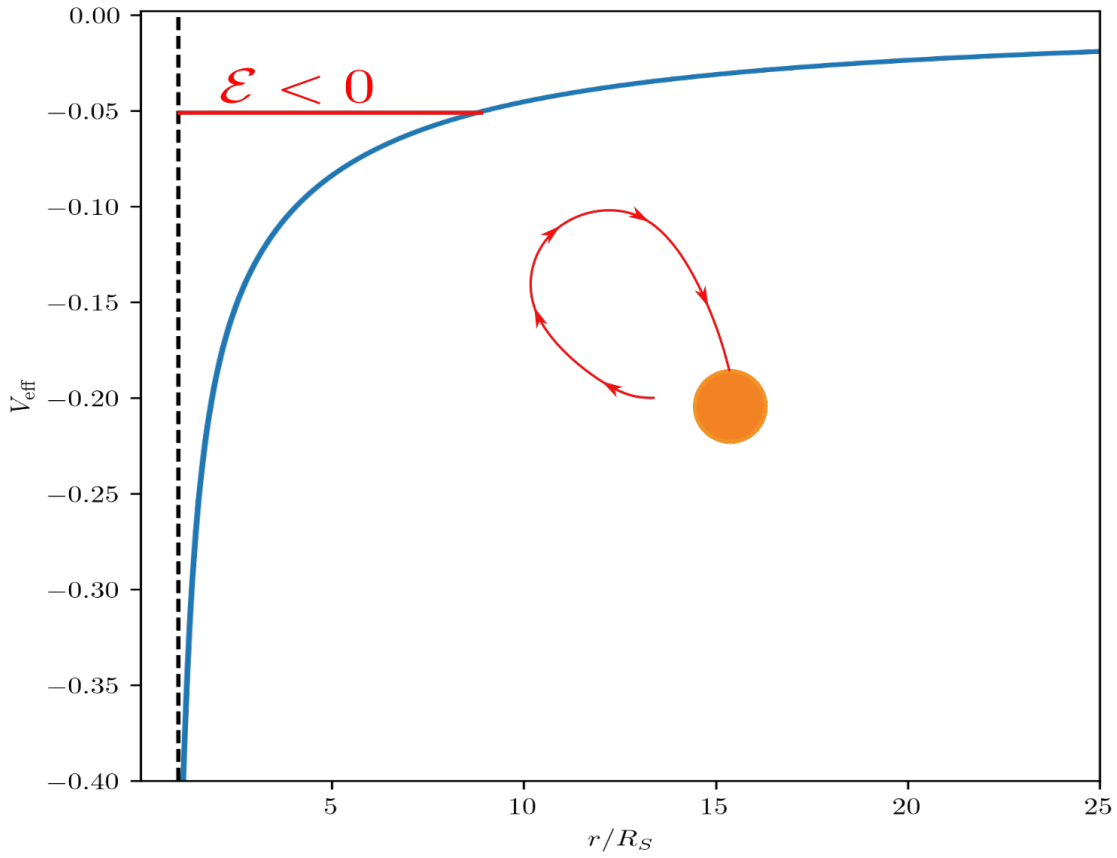


Figure 4.4: Trajectories with $\mathcal{E} < 0$ in V_{eff} from Eq. (4.78) for $\mu = l/R_S < \sqrt{3}$. The relativistic term dominates at small r and prevents the existence of the usual bound orbits of Newtonian mechanics.

This corresponds to a *stable circular orbit* around the central star, attained for $\mathcal{E} = V_{\text{min}} = V_{\text{eff}}(r_{\text{min}})$. Note that we always have $r_{\text{min}} > r_{\text{ISCO}}$, which justifies the name.

- a maximum at:

$$r_{\text{max}} = R_S \left[\mu^2 - \mu\sqrt{\mu^2 - 3}\right] . \tag{4.84}$$

This corresponds to an unstable circular orbit marking the height of the potential barrier. Unlike in the Newtonian case, relativistic orbits with large l can still plunge into the central object if $\mathcal{E} > V_{r_{\text{min}}} = V_{\text{max}}$. Note that $V_{\text{max}} > 0$ iff $l > 2R_S$. Outside of the two circular orbits, we find three types of trajectories if $l > 2R_S$ (and only 2 if $l < 2R_S$). This is summarised in Fig. 4.6:

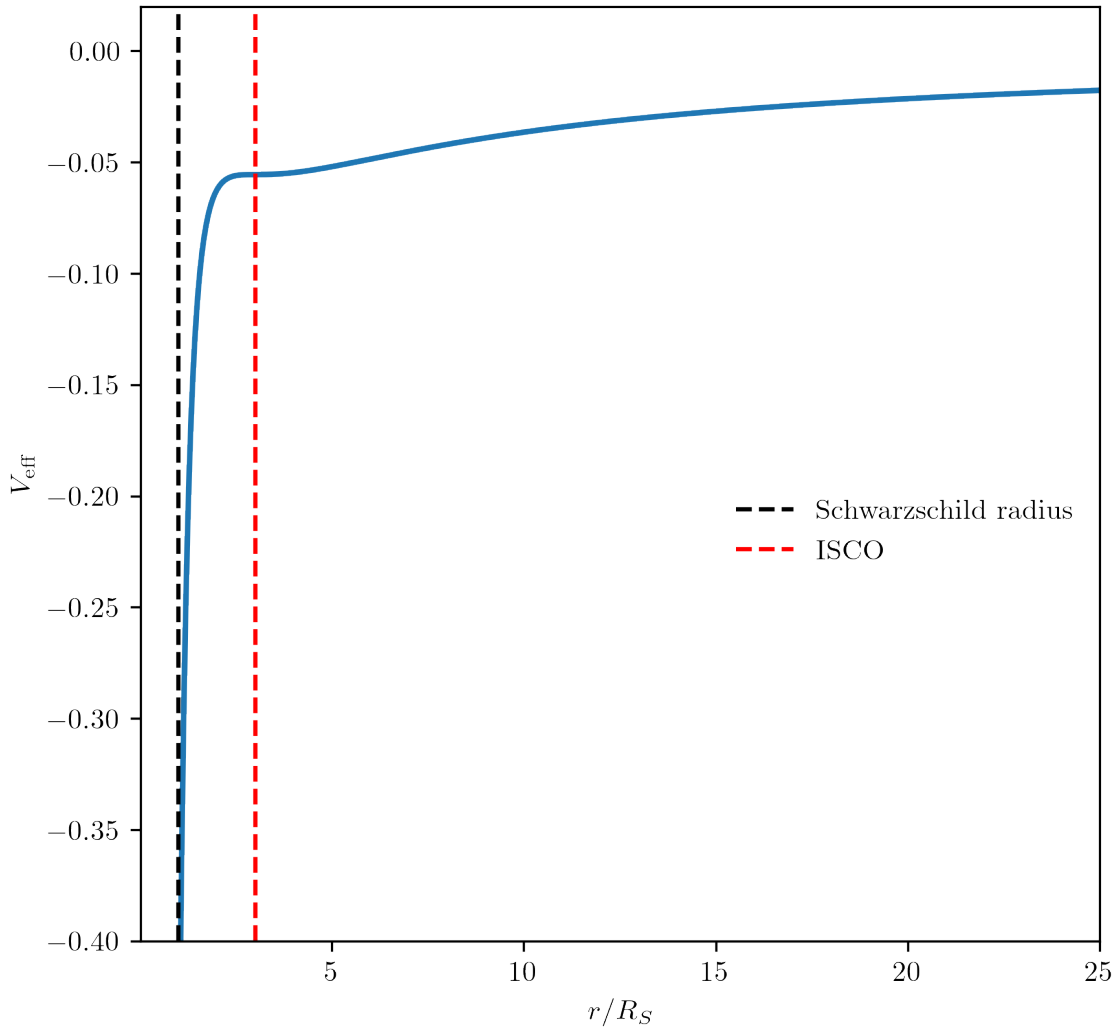


Figure 4.5: Potential for $l = R_S\sqrt{3}.$, showing the ISCO.

- For $\mathcal{E} > V_{\text{max}}$, particles are free to move away from the central object or to plunge directly into it, depending on their initial radial velocity.
- For $0 < \mathcal{E} < V_{\text{max}}$, particles either directly escape to infinity (if there initial radial velocity is positive), or approach the central object up to a minimum distance b determined by $V_{\text{eff}}(b) = \mathcal{E}$ before escaping to infinity. This is a collision, or scattering.

· For $V_{\text{min}} < \mathcal{E} < 0$, trajectories are orbits. Radial distances to the central object are bounded from below and above respectively by r_{peri} (for periastron) and r_{apo} (for apastron). This bounds are determined by solving:

$$V_{\text{eff}}(r) = \mathcal{E} < 0 \, . \tag{4.85}$$

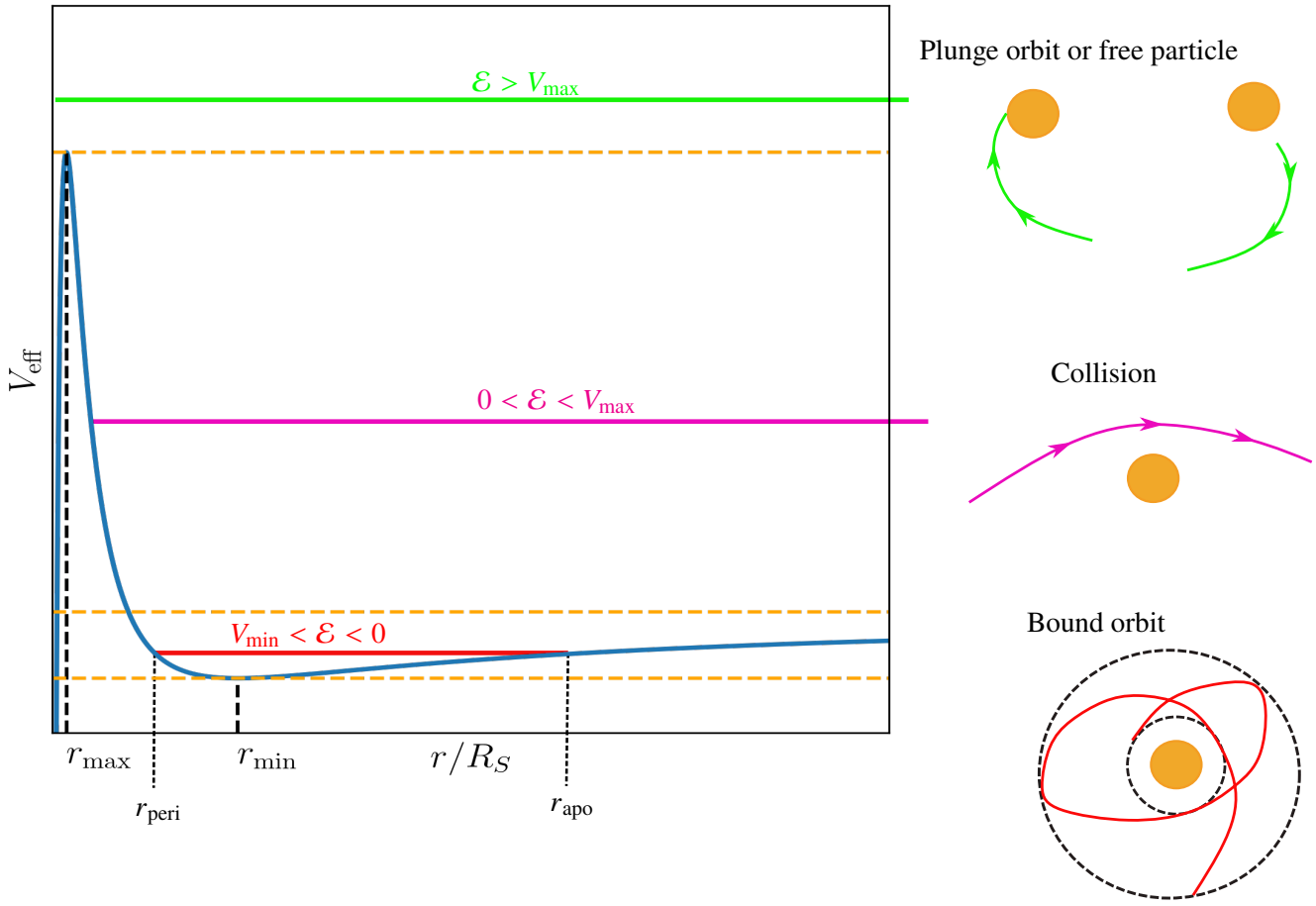


Figure 4.6: Possible trajectories for $l > R_S\sqrt{3}$. Here we present the case $l > 2R_S$ but the discussion is not altered by the less restrictive bound; the collisions simply disappear.

Note that for $V_{\text{min}} < \mathcal{E} < V_{\text{max}}$, if the particle starts with $r < r_{\text{max}}$, then the particle is trapped and necessarily falls towards the central object.

4.4.2 Kepler's law for the stable circular orbit

Let us look in more details at the stable circular orbit at $r_{\min}$ given by Eq. (4.83) when $l > R_S\sqrt{3}$. Let Ω be the angular velocity of a test particle placed on this circular orbit, as measured by an observer $\mathcal{O}$ at infinity i.e. infinitely far from the star, so that we can consider that locally around the observer, the metric is that of Minkowski spacetime. How would $\mathcal{O}$ measure Ω?

Let us assume that the test particle emits a photon radially towards $\mathcal{O}$ at a point with coordinates $\left(t_{1,\text{emit}}, r_{\min}, \pi/2, \phi_0\right)$. Let us further assume that this photon is received by $\mathcal{O}$ at $\left(t_{1,\text{rec}}, r_{\text{rec}}, \pi/2, \phi_0\right)$ with $r_{\text{rec}} \gg r_{\min}$. Next, the test particle emits a second photon radially towards $\mathcal{O}$ after a complete orbit, i.e. at the event of coordinates $\left(t_{2,\text{emit}}, r_{\min}, \pi/2, \phi_0 + 2\pi\right)$. Since we have that:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{1}{e}\left(1 - \frac{R_S}{r_{\min}}\right)\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{l}{e r_{\min}^2}\left(1 - \frac{R_S}{r_{\min}}\right) = \text{cst}\ , \tag{4.86}$$

we get:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{2\pi}{t_{2,\text{emit}} - t_{1,\text{emit}}}\ . \tag{4.87}$$

On the other hand, along radial null geodesics, we get:

$$\mathrm{d}t = \left(1 - \frac{R_S}{r}\right)\mathrm{d}r\ . \tag{4.88}$$

Integrating this equation for the two trajectories, we notice that the RHS is identical in both cases. Therefore:

$$t_{2,\text{rec}} - t_{2,\text{emit}} = t_{1,\text{rec}} - t_{1,\text{emit}}\ . \tag{4.89}$$

Finally, the proper time measured by $\mathcal{O}$ with their own clock is t since they are located at infinity. Thus, they measure the angular velocity:

$$\Omega = \frac{2\pi}{t_{2,\text{rec}} - t_{1,\text{rec}}} \tag{4.90}$$

$$= \frac{2\pi}{t_{2,\text{emit}} - t_{1,\text{emit}}} \tag{4.91}$$

$$= \frac{\mathrm{d}\phi}{\mathrm{d}t}\ . \tag{4.92}$$

Hence, we get:

$$\Omega = \frac{l}{e r_{\min}^2}\left(1 - \frac{R_S}{r_{\min}}\right)\ . \tag{4.93}$$

Using that on the circular orbit:

$$\frac{e^2-1}{2} = V_{\text{eff}}\left(r_{\text{min}}\right) \ , \tag{4.94}$$

we get that:

$$\frac{e^2}{l^2} = \left(1-\frac{R_S}{r_{\text{min}}}\right)\left(\frac{1}{l^2}+\frac{1}{r_{\text{min}}^2}\right) \ . \tag{4.95}$$

Then, Eq. (4.83) gives:

$$\frac{r_{\text{min}}}{R_S} = \left(\frac{l}{R_S}\right)^2 + \frac{l}{R_S}\sqrt{\left(\frac{l}{R_S}\right)^2-3} \ , \tag{4.96}$$

from which we get:

$$\frac{R_S}{r_{\text{min}}} = \frac{R_S^2}{l^2\left[1+\sqrt{1-3\frac{R_S^2}{l^2}}\right]} \tag{4.97}$$

$$= \frac{R_S^2}{l^2\left[1-\left(1-3R_S^2/l^2\right)\right]}\left[1-\sqrt{1-\left(\frac{R_S^2}{l^2}\right)}\right] \tag{4.98}$$

$$= \frac{1}{3}\left[1-\sqrt{1-3\frac{R_S^2}{l^2}}\right] \ . \tag{4.99}$$

Thus, taking the square:

$$\left(\frac{R_S}{r_{\text{min}}}\right)^2 = \frac{1}{3}\left[2\frac{R_S}{r_{\text{min}}}-\frac{R_S^2}{l^2}\right] \ , \tag{4.100}$$

from which we get:

$$\frac{R_S^2}{l^2} = 2\frac{R_S}{r_{\text{min}}}-3\frac{R_S^2}{r_{\text{min}}^2} \ . \tag{4.101}$$

Plugging this back into Eq. (4.95), we arrive at:

$$\frac{l}{e} = \frac{1}{1-R_S/r_{\text{min}}}\sqrt{\frac{R_S r_{\text{min}}}{2}} \ , \tag{4.102}$$

so that we arrive at

$$\Omega = \sqrt{\frac{R_S}{2r_{\text{min}}^3}} \ . \tag{4.103}$$

Introducing the observed period $T = 2\pi/\Omega$, we get *Kepler's law for the stable circular orbit*:

$$\boxed{T^2 \propto r_{\min}^3 \,.} \tag{4.104}$$

All the relativistic terms have vanished and we are left with the Newtonian result. This is a pure coincidence without any obvious physical meaning. This is apparent in the sense that r is a mere arbitrary coordinate. With a different radial coordinate, the result would be altered.

4.4.3 Non-circular bound orbits

In the rest fo this section, we are going to concentrate on the more refined properties of bound orbits. We consider a trajectory with $l > R_S\sqrt{3}$ and $V_{\min} < \mathcal{E} < 0$. We are going to write an equation for $r(\phi)$. We can rewrite Eq. (4.75) as:

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 = 2\mathcal{E} + \frac{R_s}{r} - \frac{l^2}{r^2} + \frac{R_S l^2}{r^3} \,. \tag{4.105}$$

Then, using that:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{l}{r^2} \,, \tag{4.106}$$

we get that:

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = \frac{\mathrm{d}r}{\mathrm{d}\phi}\frac{\mathrm{d}\phi}{\mathrm{d}\tau} = \frac{l}{r^2}\frac{\mathrm{d}r}{\mathrm{d}\phi} \,. \tag{4.107}$$

We are now going to use Binet's approach and define[2]:

$$x = \frac{2l^2}{R_S r} \,, \tag{4.108}$$

such that:

$$\frac{\mathrm{d}r}{\mathrm{d}\phi} = -\frac{2l^2}{R_S x^2}\frac{\mathrm{d}x}{\mathrm{d}\phi} \,. \tag{4.109}$$

Substituting in Eq. (4.107) and then in Eq. (4.105), we obtain after some simplifications:

$$\left(\frac{\mathrm{d}x}{\mathrm{d}\phi}\right)^2 = \frac{8\mathcal{E}l^2}{R_S^2} + 2x - x^2 + \frac{R_S^2}{2l^2}x^3 \,. \tag{4.110}$$

[2]Note that with this normalisation, $x = 1$ at the Newtonian circular orbit.

Differentiating with respect to ϕ and simplifying, we arrive at the *parametric equation of bound orbits for massive particles*:

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\phi^2} = 1 - x + \frac{3R_S^2}{4l^2}x^2 \ . \tag{4.111}$$

This is the equation of a harmonic oscillator with angular frequency $\omega_0 = 1$ and forcing term $1 + \frac{3R_S^2}{4l^2}x^2$. The last term is absent in Newtonian mechanics and encodes the general relativistic effects. In absence of this term, the Newtonian equation is exactly solvable so it may be possible to make progress if we assume that we are in the limit of small corrections. Note that the coefficient:

$$\alpha = \frac{3R_S^2}{4l^2} = \frac{3}{4\mu^2} < \frac{1}{4} \ . \tag{4.112}$$

Let us assume that we are in the regime $\alpha \ll 1$, so that we can look for a first order expansion of the solution:

$$x(\phi) \simeq x_0(\phi) + x_1(\phi) \ , \tag{4.113}$$

where $x_0(\phi)$ is solution to:

$$\frac{\mathrm{d}^2 x_0}{\mathrm{d}\phi^2} = 1 - x_0 \ , \tag{4.114}$$

and $x_1 = O(\alpha)$. The solution to Eq. (4.114) is well-known and, up to a phase that we are free to fix, the solution is given by:

$$x_0(\phi) = 1 + \tilde{e}\cos\phi \ , \tag{4.115}$$

where $\tilde{e} = 1 - b^2/a^2 < 1$ is the eccentricity of the ellipse with a focus at the star, a and b are the semi-major and semi-minor axes, respectively. Plugging this into Eq. (4.111) and developing to first order in α only, we get an equation for $x_1(\phi)$:

$$\frac{\mathrm{d}^2 x_1}{\mathrm{d}\phi^2} + x_1 = \alpha\left[1 + \tilde{e}\cos\phi\right]^2 = \alpha\left[1 + \frac{\tilde{e}^2}{2} + 2\tilde{e}\cos\phi + \frac{\tilde{e}^2}{2}\cos(2\phi)\right] \ . \tag{4.116}$$

Noting that:

$$\begin{cases} \dfrac{\mathrm{d}^2}{\mathrm{d}\phi^2}(\phi\sin\phi) + \phi\sin\phi = 2\cos\phi & (4.117) \\ \dfrac{\mathrm{d}^2}{\mathrm{d}\phi^2}(\cos 2\phi) + \cos(2\phi) = -3\cos(2\phi) \ , & (4.118) \end{cases}$$

we see that a solution to Eq. (4.116) is provided by:

$$x_1(\phi) = \alpha \left[1 + \frac{\tilde{e}^2}{2} - \frac{\tilde{e}^2}{6} \cos(2\phi) + \tilde{e}\phi \sin\phi \right] . \tag{4.119}$$

At first order in α, the orbit is thus defined by the parametric equation:

$$x(\phi) \simeq 1 + \tilde{e}\cos\phi + \alpha \left[1 + \frac{\tilde{e}^2}{2} - \frac{\tilde{e}^2}{6} \cos(2\phi) + \tilde{e}\phi \sin\phi \right] . \tag{4.120}$$

Note that at first order in α, we can write:

$$\cos\left[(1-\alpha)\phi\right] = \cos\phi\cos(\alpha\phi) + \sin\phi\sin(\alpha\phi) \tag{4.121}$$

$$\simeq \cos\phi + \alpha\phi\sin\phi , \tag{4.122}$$

so that we can write:

$$x(\phi) \simeq 1 + \tilde{e}\cos\left[(1-\alpha)\phi\right] + \alpha \left[1 + \frac{\tilde{e}^2}{2} - \frac{\tilde{e}^2}{6} \cos(2\phi) \right] . \tag{4.123}$$

For $\alpha = 0$, the solution is periodic, with period $T_0 = 2\pi$. What is the period of the first order solution? Let us write:

$$T = 2\pi + \Delta\phi = 2\pi + \alpha\beta , \tag{4.124}$$

and look to determine β. By definition of the period, we have:

$$x(\phi + T) = x(\phi) . \tag{4.125}$$

Since:

$$\alpha\cos\left(2(\phi+T)\right) = \alpha\cos(2\phi) + O(\alpha^2) \tag{4.126}$$

$$\cos\left[(1-\alpha)(\phi+T)\right] \simeq \cos\phi + \alpha\phi\sin\phi + \alpha(2\pi-\beta)\sin\phi \tag{4.127}$$

$$\simeq \cos\left((1-\alpha)\phi\right) + \alpha(2\pi-\beta)\sin\phi , \tag{4.128}$$

this is only satisfied if $\beta = 2\pi$ Thus after a complete period that brings the particle back to its original position, the polar angle at which it happens has shifted by:

$$\Delta\phi = 2\pi\alpha = \frac{3\pi R_S^2}{2l^2} . \tag{4.129}$$

In particular, the point of closest approach, i.e. the periastron advances by an angle $\Delta\phi$ after each revolution of the particle. At lowest order, we can replace the angular momentum per unit mass l by its value on the Newtonian ellipse. We have:

$$2a = r_{\text{peri}} + r_{\text{apo}} = r(0) + r(\pi) = \frac{2l^2}{R_S(1+\tilde{e})} + \frac{2l^2}{R_S(1-\tilde{e})} \ . \tag{4.130}$$

This gives:

$$l^2 = \frac{R_S}{2}(1-\tilde{e}^2)a \ . \tag{4.131}$$

Therefore, restoring units, we find that the *periastron* of an object such as a planet orbiting a star of mass M *advances after each revolution by an angle*:

$$\boxed{\Delta\phi = \frac{6\pi GM}{(1-\tilde{e}^2)ac^2} \text{ per period} \ .} \tag{4.132}$$

This is illustrated on Fig. 4.7.

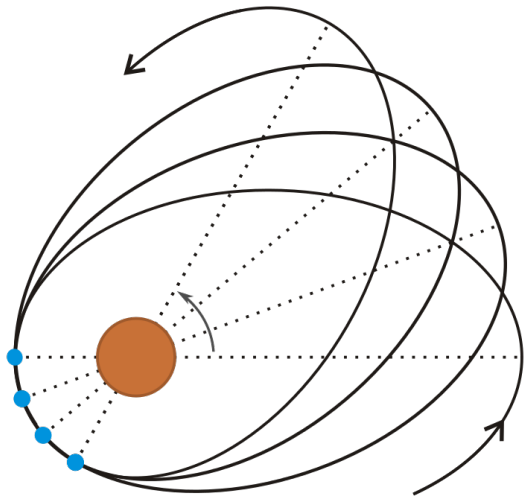


Figure 4.7: Precession of the orbit of a planet around a star. The orbit looks like a slowly advancing ellipse.

If we apply this formula to the case of Mercury, which has the largest eccentricity in the Solar

System, with:

$$\begin{cases} \dfrac{GM}{c^2} = \dfrac{GM_\odot}{c^2} \simeq 1.48 \times 10^3 \text{ m} & (4.133) \\ a \simeq 5.79 \times 10^{10} \text{ m} & (4.134) \\ \tilde{e} \simeq 0.2056 \,, & (4.135) \end{cases}$$

we find:

$$\Delta\phi \simeq 5.01 \times 10^{-7} \text{ per period} \simeq 0.103'' \text{ per period} \,. \tag{4.136}$$

Given that Mercury orbits around the Sun in 88 (Earth) days, the advance of its perihelion in a century (on Earth) is given by:

$$\Delta\phi \simeq 42.72'' \text{ per century} \,. \tag{4.137}$$

The perihelion of Mercury actually precesses by $574.10''$ per century, a fact that was well know by astronomers in the 19^{th} century. All but about $43''$ could be explained within Newtonian mechanics by taking into account the perturbations to the orbit due to the other planets in the Solar System, as well as the non-sphericity of the Sun. Einstein's calculation of the relativistic correction that made up the missing $43''$ was a powerful motivation to adopt General Relativity.

4.5 Light rays around a spherical star

Let us turn our attention to the trajectories of photons. We start with some general properties and then move to some physical applications.

4.5.1 General properties of light rays in Schwarzschild spacetime

Effective potential

For lightlike geodesics, $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = \varepsilon = 0$, so that the effective potential (4.76) becomes:

$$V_{\text{eff}}(r) = \frac{l^2}{2r^2} - \frac{R_S l^2}{2r^3} \,, \tag{4.138}$$

which, in terms of $\hat{r} = r/R_S$ reads:

$$V_{\text{eff}}(\hat{r}) = \frac{\mu^2}{2\hat{r}^2} - \frac{\mu^2}{2\hat{r}^3} \,. \tag{4.139}$$

This potential is represented for a few values of μ on Fig. 4.8

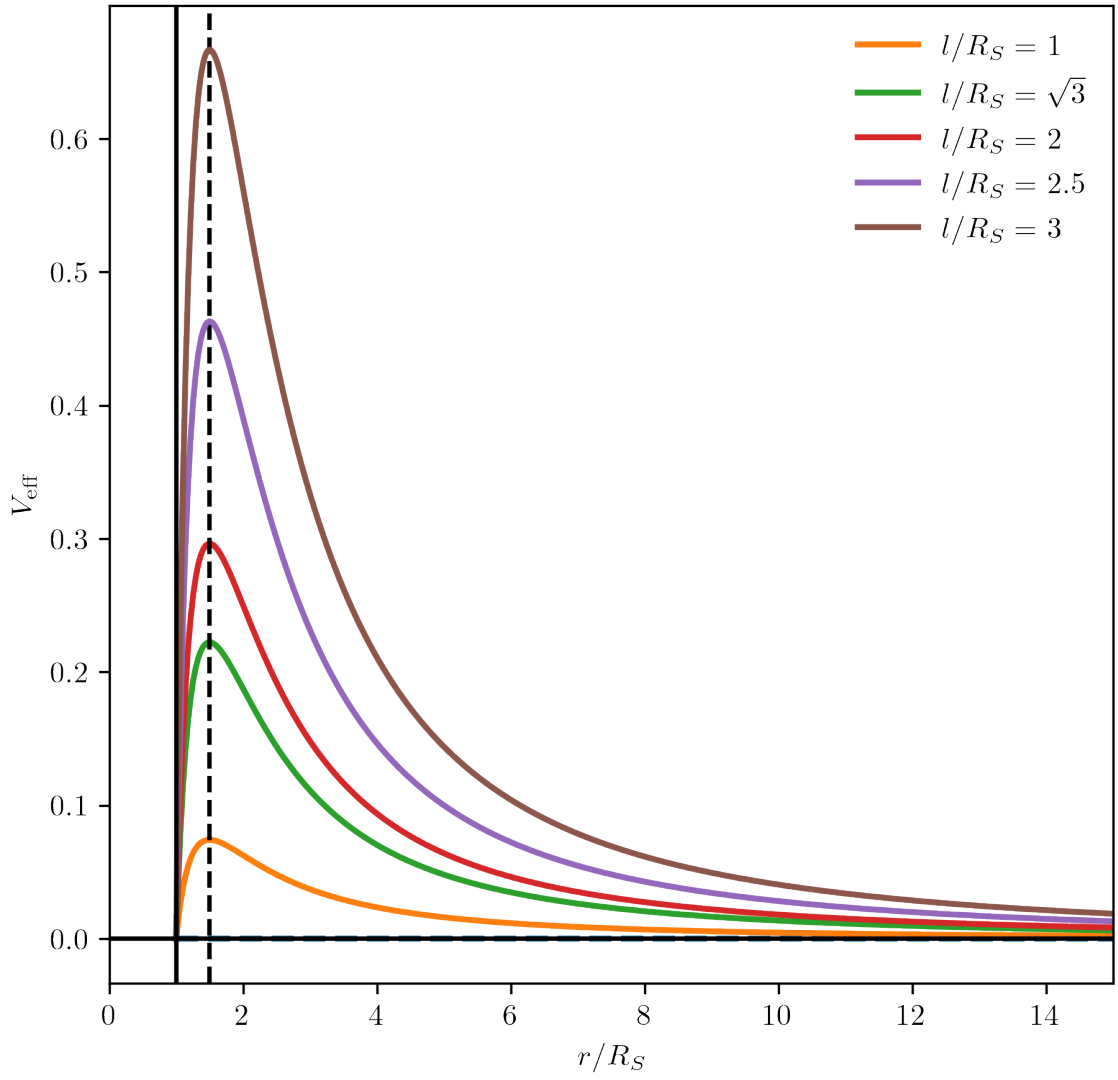


Figure 4.8: The effective potential for light rays. The vertical line at $r = 3R_S/2$ indicates the unstable circular orbit.

Note that the Newtonian gravitational potential has disappeared, as expected. For all values of $l \neq 0$, the potential has a maximum at:

$$r_c = \frac{3}{2} R_S \,, \tag{4.140}$$

corresponding to an unstable circular orbit, attained for $e_c = 2\mu/3\sqrt{3}$. For $e > e_c$, the trajectories

correspond to plunge orbits or photons escaping to infinity while for $0 < e < e_c$ we obtain a standard scattering if $r > 3R_S/2$; if $r < 3R_S/2$, photons cannot escape and plunge onto the central object. Before exploring some consequences of the structure of these trajectories, we will treat the particular case of radial trajectories, with $l = 0$.

Radial trajectories

As one can readily see, radial orbits, with $l = 0$, i.e. $\phi = \phi_0 =$ cst, are a bit particular since they correspond to an effective potential:

$$V_{\text{eff}} = 0 \,, \tag{4.141}$$

which results in a radial equation:

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = e^2 \,. \tag{4.142}$$

This fixes a relation between r and the affine parameter λ:

$$\mathrm{d}\lambda = \pm\frac{\mathrm{d}r}{e} \,, \tag{4.143}$$

with a + for infalling photons and a − for outgoing photons. These geodesics are more easily studied by going back to the line element:

$$0 = -\left(1 - \frac{R_S}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{R_S}{r}\right)^{-1}\mathrm{d}r^2 \,. \tag{4.144}$$

Hence:

$$\mathrm{d}t = \pm\left(1 - \frac{R_S}{r}\right)^{-1}\mathrm{d}r \,, \tag{4.145}$$

which is of course consistent with Eq. (4.143) given the definition of e. Eq. (4.145) is easy enough to integrate:

$$t = \pm\int \frac{\mathrm{d}r}{1 - R_S/r} = \pm R_S \int \frac{x\mathrm{d}x}{x-1} \quad \text{with } x = \frac{r}{R_S} \tag{4.146}$$

$$= \pm\, R_S \int \frac{x - 1 + 1}{x-1}\mathrm{d}x = \pm R_S\left[\int \mathrm{d}x + \int \frac{\mathrm{d}x}{x-1}\right] \tag{4.147}$$

$$= \pm\, R_S\left[x + \ln|x-1|\right] + C \quad \text{for } C \in \mathbb{R} \tag{4.148}$$

$$= \pm\left[r + R_S \ln\left|\frac{r}{R_S} - 1\right|\right] + C \quad \text{for } C \in \mathbb{R} \,. \tag{4.149}$$

Note that this expression is also valid for $r < R_S$ and we will come back to that important fact in section 4.6. For now, we have arrived at two families of radial lightlike geodesics in the region $r > R_S$:

Radial lightlike geodesics

- *Outgoing radial* light like geodesics for which $\frac{\mathrm{d}r}{\mathrm{d}t} > 0$, given by:

$$t = r + R_S \ln\left(\frac{r}{R_S} - 1\right) + C \quad \text{for } C \in \mathbb{R} \,. \tag{4.150}$$

- *Infalling radial* light like geodesics for which $\frac{\mathrm{d}r}{\mathrm{d}t} < 0$, given by:

$$t = -r - R_S \ln\left(\frac{r}{R_S} - 1\right) + C \quad \text{for } C \in \mathbb{R} \,. \tag{4.151}$$

Some of these radial geodesics are represented on Fig. 4.9. One can see that, as expected, far from the central object, we recover the geodesics of Minkowski spacetime, i.e. straight lines:

$$t = \pm r + C \quad \text{for } C \in \mathbb{R} \,. \tag{4.152}$$

Since every point on the spacetime diagram is effectively a 2-sphere, the two radial lightlike geodesics actually generate a local lightcone. We represented in green the future-directed part of a few of these lightcones. Note that we implicitly chose $\boldsymbol{e}_{(\mathbf{0})} = \frac{\partial}{\partial t}$ as our future timelike direction in the region $r > R_S$ as this corresponds to the proper time of an observer located very far from the central object. As one can see, the local causal structure tends to the one of Minkowski for $r \gg R_S$, with local lightcones with an opening angle of $\pi/2$. However, as we approach the central region and $r = R_S$, the lightcones in (t, r) coordinates close up. We still have some outgoing and infalling rays, but it gets "harder and harder" for outgoing rays to escape the central region. Infalling rays also seem to be "grazing" the hypersurface $r = R_S$. Timelike curves are always locally inside these lightcones. This geodesic structure will be very useful when we extend the Schwarzschild geometry to describe black holes in section 4.6. But before we focus on this problem, we can come back to generic lightlike geodesics for $r > R_S$ and deduce some physical, observable effects of the propagation of light in the field of a central object.

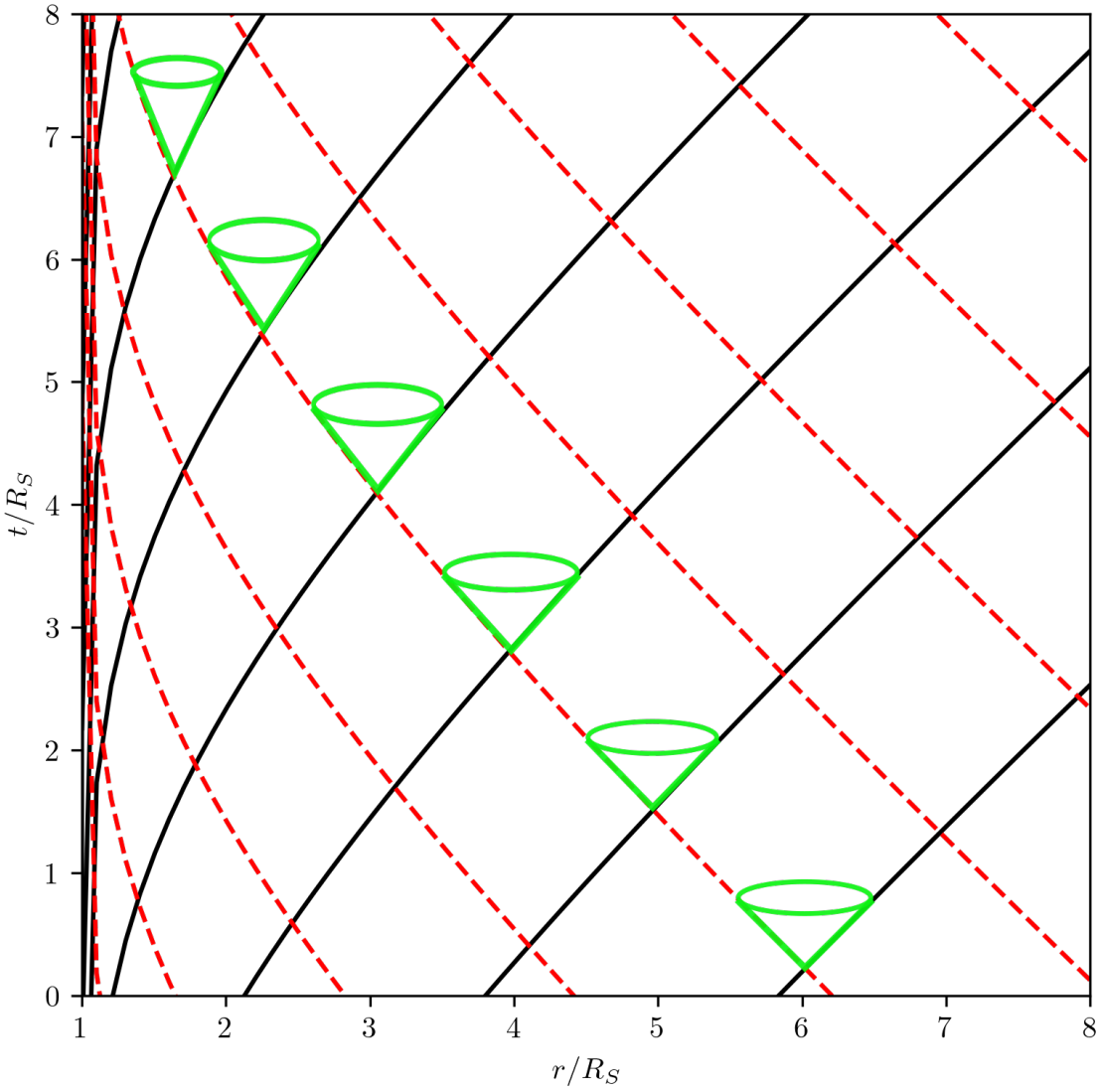


Figure 4.9: Radial lightlike geodesics of the Schwarzschild metric in the region $r > R_S$ in a space-time diagram. Infalling geodesics are represented in red dashed lines and outgoing ones in solid black.

4.5.2 Gravitational redshift

First, consider a set of observers at rest in the Schwarzschild coordinates. If we call $\boldsymbol{U} = U^\mu \boldsymbol{e}_{(\mu)}$ the field of their 4-velocity, we have:

$$U^1 = U^2 = U^3 = 0 \ . \tag{4.153}$$

Since $\boldsymbol{g}(\boldsymbol{U}, \boldsymbol{U}) = -1$, this results in:

$$U^0 = \left(1 - \frac{R_S}{r}\right)^{-1/2} \ . \tag{4.154}$$

Assume that such an observer emits a photon at $(t_1, r_1, \pi/2, \phi_1)$ that is received by another such observer at $(t_2, r_2, \pi/2, \phi_2)$. Let $\boldsymbol{k}$ be the tangent vector to the lighlike geodesics followed by the photon. At any point (t, r, θ, ϕ), the frequency, $\omega(r)$, of the photon as measured by an observer at rest in Schwarzschild coordinates is given by:

$$\hbar\omega(r) = -\boldsymbol{g}(\boldsymbol{U}, \boldsymbol{k}) \tag{4.155}$$

$$= -g_{00} U^0 k^0 \tag{4.156}$$

$$= \sqrt{1 - \frac{R_S}{r}} \frac{\mathrm{d}t}{\mathrm{d}\lambda} \ , \tag{4.157}$$

where λ is an affine parameter along the lightlike geodesics. Using the conservation of $\boldsymbol{g}(\boldsymbol{e}_{(\boldsymbol{0})}, \boldsymbol{k}) = -e$, we get:

$$\hbar\omega(r) = -\frac{e}{\sqrt{1 - \frac{R_S}{r}}} \ . \tag{4.158}$$

Therefore, the observed photon at r_2 has a frequency that is shifted with respect to its observed frequency at r_1, by a factor:

$$1 + z = \frac{\omega(r_1)}{\omega(r_2)} = \sqrt{\frac{1 - R_S/r_2}{1 - R_S/r_1}} \ . \tag{4.159}$$

This expression is general and does not depend on whether or not the geodesics is radial. z is the gravitational spectral shift. It is given by the fractional change in the measured wavelength of the photon between emission and reception:

$$z = \frac{\lambda_2 - \lambda_1}{\lambda_1} \ , \tag{4.160}$$

so that $z > 0$ corresponds to a *redshift* and $z < 0$ to a *blueshift*. Using Eq. (4.159), we see that a photon moving away from the central object is redshifted, while it is blueshifted if it falls onto the central object.

In the weak field limit, with $r \gg R_S$, we get:

$$1 + z \simeq \left(1 - \frac{R_S}{2r_2}\right)\left(1 + \frac{R_S}{2r_1}\right) \tag{4.161}$$

$$\simeq 1 - \frac{GM}{r_2} + \frac{GM}{r_1} \tag{4.162}$$

$$\simeq 1 + \Phi(r_2) - \Phi(r_1) \ , \tag{4.163}$$

which agrees with the result we obtained using the equivalence principle in subsection 2.8.2. It is also consistent with the work in the static, weak field limit in subsection 3.6.3, as it should be because the Schwarzschild metric reduces to the static, weak field limit in the approximations we have made here.

4.5.3 Deviation of light

Assume that a light ray coming from infinity falls onto the central object with $0 < e < e_c$. It approaches the central object until it reaches a minimum distance $\bar{r}$ given by:

$$\frac{e^2}{2} = V_{\text{eff}}(\bar{r}) \Leftrightarrow \frac{\bar{r}^3}{b^2} - \bar{r} + R_S = 0 \ , \tag{4.164}$$

where we defined the *impact parameter* $b = l/e$ (more on this name later). Note that the condition $e < e_c$ translates into $b > 3\sqrt{3}R_S/2$. Then, the light ray bounces back on the potential barrier and moves away from the central object, going to infinity in a direction different from its infalling one. This is a classical scattering problem. We propose to estimate the angle by which the outgoing direction deviates from the ingoing one. We introduce Cartesian axes $[Ox)$ and $[Oy)$ in the plane of motion such that O coincides with the centre of the star, and the x direction is aligned with the initial direction of propagation of the photons; see Fig. 4.10.

Initially, the photon is emitted at a point $(t_{\text{in}}, r_{\text{in}}, \pi/2, \phi_{\text{in}}) = (t_{\text{in}}, x_{\text{in}}, y_{\text{in}}, 0)$ and is very far from the central object, so we can take the limit $r_{\text{in}} \to +\infty$. It has 4-momentum with $k^0_{\text{in}} \simeq e$ and $k^1_{\text{in}} \simeq -e$ (using the radial equation in the limit of large r). Besides, we have that: $y_{\text{in}} = r_{\text{in}} \sin\phi_{\text{in}} \simeq r_{\text{in}}\phi_{\text{in}}$ because $y_{\text{in}} \ll r_{\text{in}}$ so that ϕ_{in} is small. Taking a derivative, we get:

$$\left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{\text{in}} \simeq \frac{y_{\text{in}}}{r^2_{\text{in}}} \ , \tag{4.165}$$

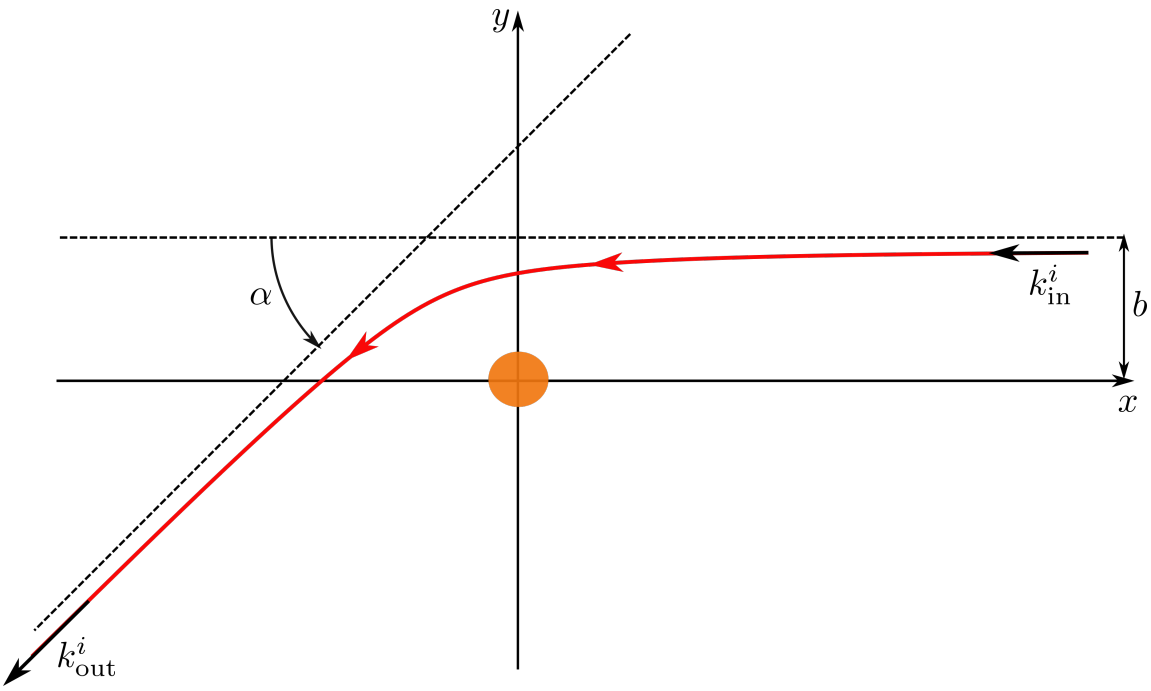


Figure 4.10: Geometry of the problem for the deviation of light by a central object.

where we used that:

$$\left.\frac{\mathrm{d}r}{\mathrm{d}t}\right|_{\rm in} = \frac{k^1_{\rm in}}{k^0_{\rm in}} \simeq 1 \ . \tag{4.166}$$

On the other hand:

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = \frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} \tag{4.167}$$

$$= \frac{l}{er^2}\left(1 - \frac{R_S}{r}\right) \ , \tag{4.168}$$

so that:

$$\left.\frac{\mathrm{d}\phi}{\mathrm{d}t}\right|_{\rm in} \simeq \frac{b}{r^2_{\rm in}} \ . \tag{4.169}$$

We can conclude that $y_{\rm in} = b$, thus the name impact parameter for b.

The radial equation of motion can then be written in terms of the angle ϕ and Binet's variable $u = b/r$, to give (exercise):

$$\frac{\mathrm{d}^2u}{\mathrm{d}\phi^2} + u = \frac{3R_S}{2b}u^2 \ , \tag{4.170}$$

similarly to what was done in subsection 4.4.3 to obtain Eq. (4.111). In that equation, the term on the RHS is the general relativistic term. Then, we assume that the photons remain far from the Schwarzschild radius of the object, so that we are in weak field. This amounts to assuming that $R_S \ll b$ so that the general relativistic term is small compared to the others. Hence, we look for a solution in the form:

$$u(\phi) \simeq u_0(\phi) + u_1(\phi) \ , \tag{4.171}$$

where $u_1(\phi) = O\left(R_S/b\right)$ and u_0 satisfies:

$$\frac{\mathrm{d}^2 u_0}{\mathrm{d}\phi^2} + u_0 = 0 \ . \tag{4.172}$$

Since this corresponds to the trajectory of light without deviation, i.e. to the straight line $y = b$, we must have: $r_0 \sin\phi = b$, i.e. $u_0 = \sin\phi$, which clearly satisfies Eq. (4.172). Injecting that in Eq. (4.170) and expanding at first order in R_S/b we obtain an equation for $u_1(\phi)$:

$$\frac{\mathrm{d}^2 u_1}{\mathrm{d}\phi^2} + u_1 = \frac{3R_S}{2b}\sin^2\phi \ . \tag{4.173}$$

This can be solved to get the solution with the right boundary condition[3], namely $u(0) = 0$:

$$u_1(\phi) = \frac{3R_S}{4b}\left(1 + \frac{1}{3}\cos(2\phi)\right) - \frac{R_S}{b}\cos\phi \ . \tag{4.174}$$

The deviation α is such that on the outward branch of the trajectory we have, asymptotically $\phi_{\text{out}} = \pi + \alpha$. Thus imposing $u(\pi + \alpha) = 0$ leads, at first order, to:

$$0 = \sin(\pi + \alpha) + \frac{3R_S}{4b}\left(1 + \frac{1}{3}\cos\left(2(\pi + \alpha)\right)\right) - \frac{R_S}{b}\cos(\pi + \alpha) \tag{4.175}$$

$$= -\alpha + \frac{R_S}{b} + \frac{R_S}{b} \ , \tag{4.176}$$

so that we have light is deflected by an angle:

$$\boxed{\alpha = \frac{2R_S}{b} \ .} \tag{4.177}$$

[3]The general solution to the non-homogeneous problem is the general solution to the homogeneous problem, $A\sin\phi + B\cos\phi$, added to a particular solution to the non-homogeneous problem, here $\frac{3R_S}{4b}\left(1 + \frac{1}{3}\cos(2\phi)\right)$. Since we know that $u(0) = 0$ and $u_0(0) = 0$, this fixes B. A is irrelevant here has it gets re-absorbed in the zeroth-order solution and is zero asymptotically anyway.

In 1919, Eddington and his team set to measure this angular deviation by observing the apparent shift in position on the sky of stars located behind the Sun during a Solar eclipse: by observing the stars at night before the eclipse (while the Sun was not interfering with light coming from them) and then again during the eclipse, they were able to measure α. For rays grazing the Sun, we can take $b \simeq R_\odot$ and we find:

$$\alpha \simeq 1.75'' \ . \tag{4.178}$$

Eddington's observations were in agreement with that prediction and this was the first real test of General Relativity (on a genuine, unexpected prediction). It made Einstein famous worldwide overnight. The study of deviations of light rays by clumps of matter is now a fully developed branch of astrophysics known as gravitational lensing which is used on multiple scales to probe gravity, the growth of structure in the Universe and the nature of Dark Matter; see Fig. 4.11 for a stunning recent image exhibiting many gravitationally lensed images.

Fig. 4.12 summarises modern constraints on General Relativity obtained via gravitational lensing across a wide range of scales and using various instruments. The agreement is stunning.

4.5.4 Shapiro time delay

Not only are light rays bent by the presence of mass along their path, but electromagnetic signals are also retarded by gravitational field. This is known as gravitational (or Shapiro) time-delay. Let us consider a light ray sent from Earth to a distant satellite, bouncing back on the satellite and being sent back to Earth. We choose our Schwarzschild coordinates (t, r, θ, ϕ) centred on the Sun with $\theta = \pi/2$ the ecliptic. We denote by $r_\oplus$ and r_* the radial coordinates of Earth and the satellite respectively. Let r_0 be the closest distance to the Sun along the light ray. The situation is summarised on Fig. 4.13. Along the lightlike geodesic connecting Earth to the satellite, as well as along the one connecting the satellite to Earth, using Eqs. (4.75)-(4.138), we have:

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 = e^2 - \frac{l^2}{r^2} + \frac{R_S}{l^2 r^3} \tag{4.179}$$

$$= \frac{l^2}{b^2} - \frac{l^2}{r^2} + \frac{R_S}{l^2 r^3} \ . \tag{4.180}$$

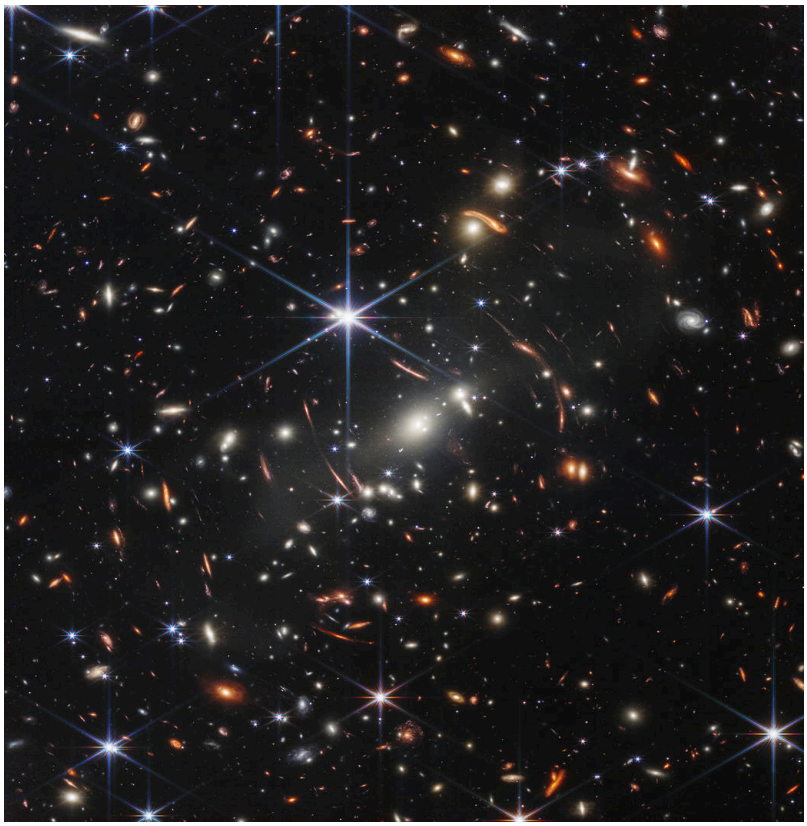


Figure 4.11: Image of the SMACS 0723 galaxy cluster by the James Webb Space Telescope, Credit:NASA, ESA, CSA, and STScI. There are many images of gravitationally lensed galaxies in this image. They appear distorted by the massive cluster situated on the line of sight. Try and find them.

Then:

$$\frac{\mathrm{d}t}{\mathrm{d}r} = \frac{\mathrm{d}t}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}r} \tag{4.181}$$

$$= \left(1 - \frac{R_S}{r}\right)^{-1} \left(1 - \frac{b^2}{r^2}\left(1 - \frac{R_S}{r}\right)\right)^{-1/2} . \tag{4.182}$$

Given that at $r = r_0$, $\frac{\mathrm{d}r}{\mathrm{d}\lambda} = 0$ by definition, we can related b to r_0:

$$b^2 = r_0^2 \left(1 - \frac{R_S}{r}\right)^{-1} . \tag{4.183}$$

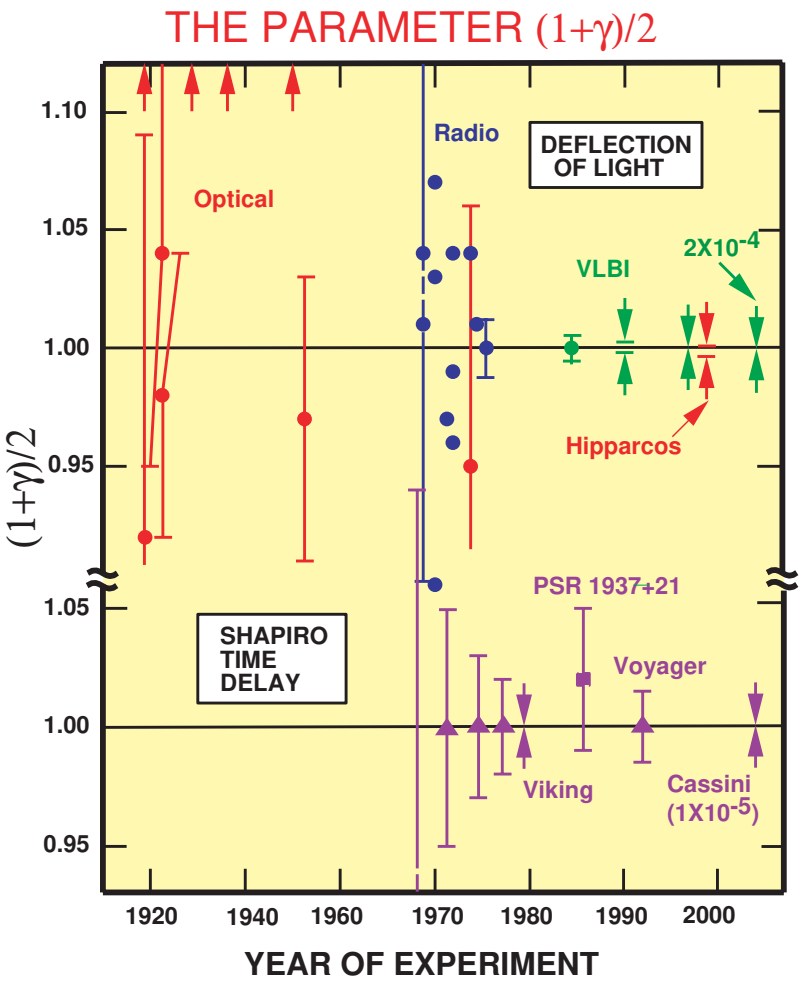


Figure 4.12: Constraints on General Relativity from lensing observables: deviation of light in the upper part and time delays in the lower part. General Relativity corresponds to $\gamma = 1$. Figs taken from [22].

Thus, we arrive that:

$$\frac{\mathrm{d}t}{\mathrm{d}r} = \left(1 - \frac{R_S}{r}\right)^{-1} \left[1 - \frac{r_0^2}{r^2}\frac{1 - R_S/r}{1 - R_S/r_0}\right]^{-1} . \tag{4.184}$$

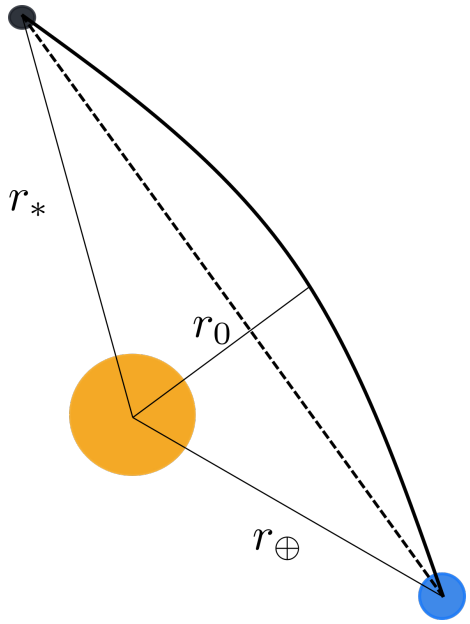


Figure 4.13: Geometry of the problem for the Shapiro effect. The dashed line corresponds to the trajectory light would have in special relativity, in absence of the Sun, while the solid curve is the actual trajectory.

In all applicable situations, we have $r_0 \ll r_\oplus$, $r_0 \ll r_*$ and $R_S \ll r$, so we can simplify this expression by a series of Taylor expansions at first order in small quantities:

$$\frac{\mathrm{d}t}{\mathrm{d}r} \simeq \left(1+\frac{R_S}{r}\right)\left(1-\frac{r_0^2}{r^2}\right)^{-1/2}\left[1+\frac{r_0^2}{r^2-r_0^2}\left(\frac{R_S}{r}-\frac{R_S}{r_0}\right)\right]^{-1/2} \tag{4.185}$$

$$\simeq \left(1+\frac{R_S}{r}\right)\left(1-\frac{r_0^2}{r^2}\right)^{-1/2}\left[1+\frac{r_0}{2\,(r_0+r)}\frac{R_S}{r}\right] \tag{4.186}$$

$$\simeq \left(1-\frac{r_0^2}{r^2}\right)^{-1/2}\left[1+\frac{R_S}{r}+\frac{r_0}{2\,(r_0+r)}\frac{R_S}{r}\right] \tag{4.187}$$

$$\simeq \frac{r}{\sqrt{r^2-r_0^2}}\left[1+\frac{R_S}{r}+\frac{r_0}{2\,(r_0+r)}\frac{R_S}{r}\right]\,. \tag{4.188}$$

We can use this expression to get the time taken by a photon to travel between a source at r_1 towards an observer at r_2:

$$t\left(r_1, r_2\right) = \int_{r_1}^{r_2} \frac{r}{\sqrt{r^2 - r_0^2}} \left[1 + \frac{R_S}{r} + \frac{r_0}{2\left(r_0 + r\right)} \frac{R_S}{r}\right] \mathrm{d}r \ . \tag{4.189}$$

Equivalently:

$$t\left(r_1, r_2\right) = \underbrace{\int_{r_1}^{r_2} \frac{r}{\sqrt{r^2 - r_0^2}} \mathrm{d}r}_{= \left[\sqrt{r^2 - r_0^2}\right]_{r_1}^{r_2}} + R_S \underbrace{\int_{r_1}^{r_2} \frac{\mathrm{d}r}{\sqrt{r^2 - r_0^2}}}_{= \left[\ln\left(\frac{r}{r_0} + \sqrt{\frac{r^2}{r_0^2} - 1}\right)\right]_{r_1}^{r_2}} + \frac{R_S}{2} \underbrace{\int_{r_1}^{r_2} \frac{r_0 \mathrm{d}r}{\sqrt{r - r_0}\left(r + r_0\right)^{3/2}}}_{= \left[\sqrt{\frac{r - r_0}{r + r_0}}\right]_{r_1}^{r_2}} \ . \tag{4.190}$$

The travel time between Earth and the point of closest approach is therefore:

$$t\left(r_\oplus, r_0\right) = \sqrt{r_\oplus^2 - r_0^2} + R_S \ln\left(\frac{r_\oplus}{r_0} + \sqrt{\frac{r_\oplus^2}{r_0^2} - 1}\right) + \frac{R_S}{2} \sqrt{\frac{r_\oplus - r_0}{r_\oplus + r_0}} \ , \tag{4.191}$$

and an identical expression for the travel time between the point of closest approach and the satellite by replacing $r_\oplus$ by r_*[4]. The travel time is thus given by:

$$T = T_{\text{Mink}} + \Delta T \ , \tag{4.192}$$

where:

$$T_{\text{Mink}} = 2\left[\sqrt{r_\oplus^2 - r_0^2} + \sqrt{r_*^2 - r_0^2}\right] \tag{4.193}$$

is the travel time in Minkowski spacetime. It is twice the sum of the travel time between Earth and the point of closest approach in a straight line and the travel time between the point of closest approach and the satellite in a straight line. However, in Minkowski spacetime, light travels along straight lines and follows the dashed curve on Fig. 4.13, which means that its travel time between Earth and the satellite and back is:

$$T_{\text{Mink}}^{\text{True}} = 2\left[\sqrt{r_\oplus^2 - b^2} + \sqrt{r_*^2 - b^2}\right] \ . \tag{4.194}$$

[4]Although it is not apparent in the final expression (4.190), because the spacetime is static, and because we can assume that the positions of the satellite and Earth have not changed significantly during the return trip of photons, we have $t\left(r_0, r_*\right) = t\left(r_*, r_0\right)$. This is called the "inverse return of light" in geometric optics. We will use these relations in what follows.

But at leading order, using Eq. (4.183), we get:

$$T_{\text{Mink}}^{\text{True}} \simeq T_{\text{Mink}} - \left[t\left(r_\oplus, r_0\right) \frac{R_S r_0}{r_\oplus^2} + t\left(r_*, r_0\right) \frac{R_S r_0}{r_*^2} \right] . \tag{4.195}$$

The difference between the two expressions, which is known as the geometric time delay for clear reasons, is of order two in our small parameters and can thus be neglected.
Therefore, the *Shapiro time delay* is given by:

$$\Delta T = R_S \left[2 \ln \left(\frac{\left(r_\oplus + \sqrt{r_\oplus^2 - r_0^2}\right)\left(r_* + \sqrt{r_*^2 - r_0^2}\right)}{r_0^2} \right) + \sqrt{\frac{r_\oplus - r_0}{r_\oplus + r_0}} + \sqrt{\frac{r_* - r_0}{r_* + r_0}} \right] . \tag{4.196}$$

This expression is always positive so this is indeed a delay. The expression given here is actually too general since it does not take into account the fact that we used Taylor expansions to obtain it. If we do Taylor expand Eq. (4.196), we get, at dominant order[5]:

$$\boxed{\Delta T = 2 R_S \left[\ln \left(\frac{4 r_\oplus r_*}{r_0^2} \right) + 1 \right] .} \tag{4.197}$$

The measurement of this time delay with the Cassini probe in 2003 confirmed that gravitational agreed with General Relativity to one part in 2×10^{-5}[4].
In 2010, the Shapiro effect was also used to measure the mass of a neutron star. The neutron star is actually observed as a pulsar named PSR J1614-2230 and belongs to a binary system with a white dwarf. Once every revolution, the white dwarf passes between the neutron star and us. This means that by observing carefully the intervals between the pulses emitted by the neutron star during the passage of the white dwarf and immediately before or after, one can measure a delay in their arrival time on Earth. This was measured to be 25 μs for this system, which leads to an estimate of the mass of the white dwarf and, using orbital parameters, to a value for the mass of the neutron star: $M = 1.97 \pm 0.04\, M_\odot$. This is a very high value, the highest ever recorded for a neutron star, which provides a lot of information on the properties of the dense matter forming the neutron star.

[5]Note that, the argument of the logarithm is very large so neglecting terms proportional to $r_0/r_\oplus$ or r_0/r_* is easily justified. However, this argument receives another contribution, equal to $\left(r_\oplus^2 + r_*^2\right)/(r_\oplus r_*)$ which is not small. Nevertheless, using $r_\oplus \sim r_*$, we see that this term is order unity, so much smaller than the dominant term that we kept here. Technically, we performed a Laurent expansion rather than a Taylor expansion.

4.6 The Schwarzschild black hole

In the previous sections of this chapter, we studied the gravitational field around an isolated, spherical source like a star. If we describe a star as a ball of hot gas with density $\rho(t, r)$ and pressure $p(t, r)$, the stability of the star depends on the balance between two forces: the self-gravitation of the matter which tends to make the star contract and heats up the gas, and the internal pressure of the gas, which increases with temperature and tends to support the matter, opposing the gravitational collapse. However, as nucleosynthesis proceeds, the chemical composition of the star changes and this equilibrium evolves. Eventually, after a series of complicated stages, nuclear fusion stops and the internal process are no longer able to oppose the gravitational collapse: the star dies. Its end state depends on how much mass remains in the core of the star: dwarfs for low mass remnants or neutron stars for $M_\odot \lesssim M \lesssim 3M_\odot$. These compact objects are supported by the quantum pressure generated by Pauli exclusion principle of fermions: electrons for white dwarfs and neutrons for neutron stars. However, if the mass of the remnant exceeds about $3M_\odot$, even these quantum effects are not strong enough to maintain the stability of the object: it collapses completely and forms a black hole.

In this section, we will discuss in fair details the simplest black hole solution, the Schwarzschild black hole. We will not attempt any discussion of the formation of black holes, nevertheless, we can make a few remarks in this introduction. Imagine a ball of self-gravitating, non-relativistic matter with radius (in Schwarzschild coordinates) $R(t)$ collapsing isotropically under the influence of its own gravity. As we have seen, outside the star, everything is described by the exterior Schwarzschild solution we studied in the previous sections. As long as $R(t) > R_S$ everything is all right. As we are going to see next, it turns out that nothing special happens to timelike geodesics at $r = R_S$ so that the star continues to shrink past that limit. At that point however, it forms a black hole. What happens to matter as it continues falling beyond the Schwarzschild radius towards an infinite concentration at the centre of the spatial coordinates in the asymptotic future is subject to contemporary debates and can only be assessed in quantum gravity, which is way beyond the scope of these notes.

4.6.1 Beyond the Schwarzschild radius: a conundrum

Let us go back to the Schwarzschild metric in Schwarzschild coordinates:

$$\boldsymbol{g} = -\left(1 - \frac{R_S}{r}\right) \mathrm{d}t \otimes \mathrm{d}t + \left(1 - \frac{R_S}{r}\right)^{-1} \mathrm{d}r \otimes \mathrm{d}r + r^2 \left[\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta \, \mathrm{d}\phi \otimes \mathrm{d}\phi\right] . \tag{4.198}$$

So far, we looked at the geometry for the exterior region $r > R_S$ but we can note that this metric is perfectly well defined for $0 < r < R_S$ which we will call the interior region. As an illustration, the radial lightlike geodesics that we characterised in subsection 4.5.1 via Eqs. (4.151)-(4.150) can be constructed for $0 < r < R_S$; they simply exchange their role and we get:

Radial lightlike geodesics: interior and exterior

- *Outward radial* lightlike geodesics ($\frac{\mathrm{d}r}{\mathrm{d}t} > 0$):

$$t = r + R_S \ln\left(\frac{r}{R_S} - 1\right) + C \quad \text{for } C \in \mathbb{R} \quad \text{for} \quad r > R_S \tag{4.199}$$

$$t = -r - R_S \ln\left(1 - \frac{r}{R_S}\right) + C \quad \text{for } C \in \mathbb{R} \quad \text{for} \quad 0 < r < R_S \, . \tag{4.200}$$

- *Inward radial* lightlike geodesics ($\frac{\mathrm{d}r}{\mathrm{d}t} < 0$):

$$t = -r - R_S \ln\left(\frac{r}{R_S} - 1\right) + C \quad \text{for } C \in \mathbb{R} \quad \text{for} \quad r > R_S \tag{4.201}$$

$$t = r + R_S \ln\left(1 - \frac{r}{R_S}\right) + C \quad \text{for } C \in \mathbb{R} \quad \text{for} \quad 0 < r < R_S \, . \tag{4.202}$$

A few of these geodesics are represented on Fig. 4.14. Note that each point on this diagram is effectively a 2-sphere. Note that we have changed their names from "outgoing" and "infalling" to "outward" and "inward". In the exterior region, the notions overlap because we have chosen $\boldsymbol{e}_{(\mathbf{0})} = \frac{\partial}{\partial t}$ for our future direction. But what of the interior region? There, we have:

$$\begin{cases} \boldsymbol{g}\left(\boldsymbol{e}_{(\mathbf{0})}, \boldsymbol{e}_{(\mathbf{0})}\right) = -\left(1 - \dfrac{R_S}{r}\right) > 0 & (4.203) \\ \boldsymbol{g}\left(\boldsymbol{e}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{1})}\right) = \left(1 - \dfrac{R_S}{r}\right)^{-1} < 0 \, , & (4.204) \end{cases}$$

so that $\boldsymbol{e}_{(\mathbf{1})} = \frac{\partial}{\partial r}$ is timelike while $\boldsymbol{e}_{(\mathbf{0})} = \frac{\partial}{\partial t}$ is spacelike. Therefore, how do we choose a time orientation to be able to draw the local future lightcones and study causality in this region? Can we do it in a way that is consistent with the causal structure in the region $r > R_S$, so that we can "glue" these regions smoothly through the hypersurface $r = R_S$ and obtain one spacetime covering both regions?

We won't be able to address these questions in (t, r) coordinates because they are pathological as

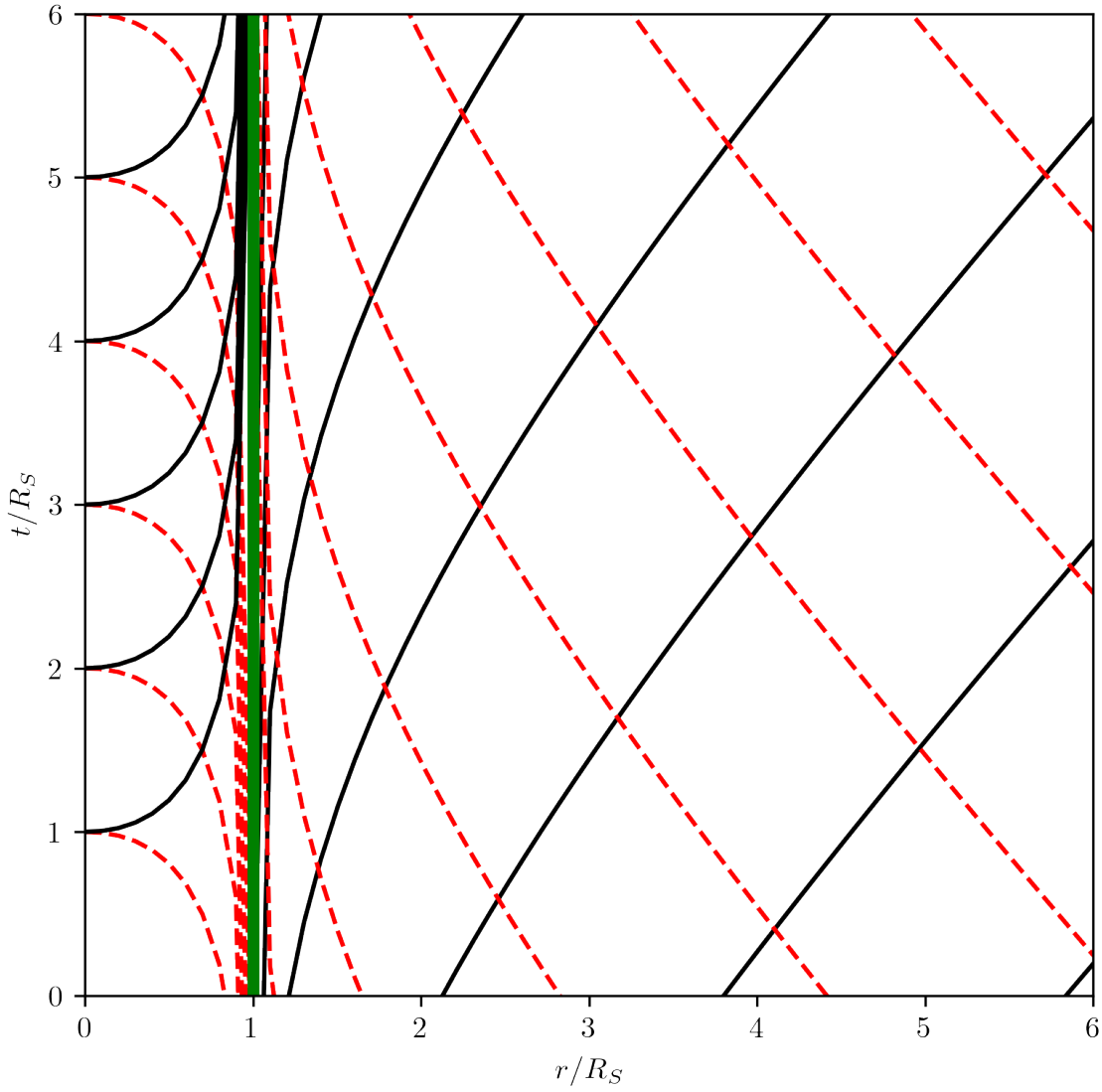


Figure 4.14: Radial null geodesics of the Schwarzschild metric both in the exterior region, $r > R_S$ and the interior region, $0 < r < R_S$. The hypersurface at $r = R_S$ is represented in green. Black solid curves are outward ($\mathrm{d}r/\mathrm{d}t > 0$) while dashed red ones are inward ($\mathrm{d}r/\mathrm{d}t < 0$).

we approach $r = R_S$. Before we explain the strategy we are going to follow to tackle these questions, we need to make sure that they may even be tackled. Let us consider a massive object free falling radially onto the central region from infinity, where it started at rest. Then, we have $\mathrm{d}t/\mathrm{d}\tau = 1$ at

infinity and $e = 1$, where τ is the proper time along the object's geodesic. Since $l = 0$, we get that (the minus sign means it is infalling):

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = -\sqrt{\frac{R_S}{r}} \ . \tag{4.205}$$

The 4-velocity along the geodesics is then:

$$\boldsymbol{u} = \left(1 - \frac{R_S}{r}\right)^{-1} \frac{\partial}{\partial t} - \sqrt{\frac{R_S}{r}} \frac{\partial}{\partial r} \ . \tag{4.206}$$

The local inertial frame of the free-falling observer is $\left\{\boldsymbol{u} = \frac{\partial}{\partial \tau}, \hat{\boldsymbol{e}}_{(\mathbf{1})} = \frac{\partial}{\partial R}, \frac{1}{r}\frac{\partial}{\partial \theta}, \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}\right\}$ where we define the new coordinate R such that:

$$\frac{\partial}{\partial R} = A\frac{\partial}{\partial t} + B\frac{\partial}{\partial r} \tag{4.207}$$

with:

$$\boldsymbol{g}\left(\frac{\partial}{\partial R}, \frac{\partial}{\partial R}\right) = 1 \tag{4.208}$$

$$\boldsymbol{g}\left(\frac{\partial}{\partial \tau}, \frac{\partial}{\partial R}\right) = 0 \ . \tag{4.209}$$

This gives:

$$\frac{\partial}{\partial R} = -\left(1 - \frac{R_S}{r}\right)^{-1} \sqrt{\frac{R_S}{r}} \frac{\partial}{\partial t} + \frac{\partial}{\partial r} \ . \tag{4.210}$$

In this coordinate system, the non-zero components of the Riemann curvature tensor read[6]:

$$R^{\hat{0}}{}_{\hat{1}\hat{0}\hat{1}} = \frac{R_S}{r^3} \ , \ R^{\hat{0}}{}_{\hat{2}\hat{0}\hat{2}} = R^{\hat{0}}{}_{\hat{3}\hat{0}\hat{3}} = -\frac{R_S}{2r^3} \tag{4.211}$$

$$R^{\hat{2}}{}_{\hat{3}\hat{2}\hat{3}} = \frac{R_S}{r^5} \ , \ R^{\hat{1}}{}_{\hat{2}\hat{1}\hat{2}} = R^{\hat{1}}{}_{\hat{3}\hat{1}\hat{3}} = -\frac{R_S}{2r^3} \ . \tag{4.212}$$

Thus, the geodesic deviation equation for the deviation vector $\boldsymbol{\xi}$ that describes the deformation of the object by tidal forces becomes:

$$\frac{\mathrm{d}^2\xi^{\hat{\alpha}}}{\mathrm{d}\tau^2} = R^{\hat{\alpha}}{}_{\hat{0}\hat{0}\hat{\gamma}}\xi^{\hat{\gamma}} \tag{4.213}$$

$$= -\eta^{\hat{\alpha}\hat{\beta}} R_{\hat{0}\hat{\beta}\hat{0}\hat{\gamma}}\xi^{\hat{\gamma}} \tag{4.214}$$

$$= \eta^{\hat{\alpha}\hat{\beta}} R^{\hat{0}}{}_{\hat{\beta}\hat{0}\hat{\gamma}}\xi^{\hat{\gamma}} \ . \tag{4.215}$$

[6]This result is quite cumbersome to obtain. One first needs to calculate the components of the Riemann tensor in Schwarzschild coordinates and then apply the law of transformations for the components of a tensor.

These forces remains thus completely finite in the neighbourhood of the hypersurface $r = R_S$. A free falling observer that approaches that hypersurface does not experience any special gravitational effect so it looks like from a physical point of view, we may be able to bring the hypersurface $r = R_S$ back into the physical spacetime by a clever change of coordinates and some continuation procedure. As you can see, this is clearly not the case for the $r = 0$ singularity: a similar change of coordinate in the $r < R_S$ region shows that tidal forces diverge as one approaches $r = 0$. We say that $r = 0$ is a *physical singularity*. Strictly speaking it does not even belong to the spacetime and General Relativity fails there.

4.6.2 Exploring the Schwarzschild black hole

Eddington-Filkenstein coordinates

We can now turn to the main issue of this section: the construction of a consistent geometry spanning both the exterior and interior region of the Schwarzschild spacetime. We are going to construct a new coordinate system that patches together interior and exterior regions in a continuous way, covering the $r = R_S$ region in a smooth, regular way.

The strategy consists in exploring the causal structure of spacetime, i.e. in following lightlike geodesics and connecting them across the apparent singular region. We start in the exterior region, $r > R_S$ and we follow the (future-directed) radial lightlike geodesics. Because we want to extend our spacetime towards the $r < R_S$ region, we focus on infalling radial null geodesics. We will talk about the outgoing ones later. We define the *infalling Eddington-Filkenstein coordinate* v to be constant along the inward radial geodesics in the exterior region:

$$v = t + r + R_S \ln\left|\frac{r}{R_S} - 1\right| . \tag{4.216}$$

This definition is valid in both exterior and interior region ans this will be useful later. If we construct the chart (v, r, θ, ϕ), then at constant θ and ϕ, v = constant lines will correspond to light rays. In the exterior regions, these will be the infalling rays (red dashed curves on Fig. 4.14) What does the metric look like in this chart? We have:

$$\mathrm{d}v = \mathrm{d}t + \left(1 - \frac{1}{1 - r/R_S}\right)\mathrm{d}r \tag{4.217}$$

$$= \mathrm{d}t + \frac{1}{1 - R_S/r}\mathrm{d}r . \tag{4.218}$$

Thus:

$$\mathrm{d}t^2 = \mathrm{d}v^2 - 2\left(1 - \frac{R_S}{r}\right)^{-1} \mathrm{d}v\mathrm{d}r + \left(1 - \frac{R_S}{r}\right)^{-2} \mathrm{d}r \ . \tag{4.219}$$

Substituting in the line element (4.42), we get it expressed in the new coordinate system, known as the *Eddington-Finkelstein infalling coordinates* :

$$\boxed{\mathrm{d}s^2 = -\left(1 - \frac{R_S}{r}\right) \mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r + r^2 \left[\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\phi\right] \ .,} \tag{4.220}$$

with the bounds $v \in \mathbb{R}$, $r \in \mathbb{R}_+^*$, $\theta \in [0, \pi)$ and $\phi \in [0, 2\pi)$.

First let us note that the metric is perfectly regular at $r = R_S$. The determinant of $\boldsymbol{g}$ is given by:

$$\det \boldsymbol{g} = -r^4 \sin^2\theta \ , \tag{4.221}$$

so that the metric is invertible at all points with $r > 0$ (the apparent singularity at $\theta = 0$ and $\pi/2$ is an artefact of the spherical coordinates). The chart (v, r, θ, ϕ) covers the interior, exterior and $r = R_S$ regions in a perfectly smooth way. The coordinate v tends to $+\infty$ when t, or r, or both, tend to $+\infty$, and to $-\infty$ when $t \longrightarrow -\infty$ at fixed r.

Causal structure

What do radial lightlike geodesics look like? Let us set:

$$\mathrm{d}s^2 = -\left(1 - \frac{R_S}{r}\right) \mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}r = 0 \ . \tag{4.222}$$

We have two sets of radial lightlike geodesics:

$$\begin{cases} \text{Type Iv:} \quad \mathrm{d}v = 0 \implies \forall r \in \mathbb{R}_+^*, \ v = \text{cst} & (4.223) \\ \text{Type IIv:} \quad \mathrm{d}r = \dfrac{1}{2}\left(1 - \dfrac{R_S}{r}\right) \mathrm{d}v \implies v = 2r + 2R_S \ln\left|\dfrac{r}{R_S} - 1\right| + \text{cst} \ . & (4.224) \end{cases}$$

Type Iv corresponds to the inward radial geodesics in the region $r > R_S$ and to the outward ones in the region $r < R_S$ (to see that, you can express $v = \text{cst}$ in terms of (t, r)). Therefore, in terms of the parameter v, these two geodesics connect through the hypersurface $r = R_S$. Photons travelling radially from the $r > R_S$ region towards the central region cross the hypersurface $r = R_S$ and continue towards $r = 0$. The outward geodesics in the region $r < R_S$ are travelled from $r = R_S$

down to $r = 0$, with $\mathrm{d}r/\mathrm{d}\lambda < 0$. How can we understand this counter-intuitive fact?
If we want to speak of direction of travel along (timelike or lightlike) geodesics, we need to introduce a time orientation. Since we already chose one outside by requiring that future-directed vectors be "aligned" with $\boldsymbol{e}_{(\mathbf{0})} = \frac{\partial}{\partial t}$, we need to choose an orientation inside that is compatible with this. First, since we have two sets of coordinates, we need to be a bit careful here. Remember that truly:

$$\boldsymbol{e}_{(\mathbf{1})} = \left.\frac{\partial}{\partial r}\right|_{(t,\theta,\phi)} , \tag{4.225}$$

while, if we call $\hat{\boldsymbol{e}}_{(\boldsymbol{\mu})}$ the coordinate basis associated to the chart (v, r, θ, ϕ), we have:

$$\hat{\boldsymbol{e}}_{(\mathbf{1})} = \left.\frac{\partial}{\partial r}\right|_{(v,\theta,\phi)} \neq \boldsymbol{e}_{(\mathbf{1})} . \tag{4.226}$$

In fact, we have that $\hat{\boldsymbol{e}}_{(\mathbf{1})}$ is globally lightlike:

$$\boldsymbol{g}\left(\hat{\boldsymbol{e}}_{(\mathbf{1})}, \hat{\boldsymbol{e}}_{(\mathbf{1})}\right) = 0 . \tag{4.227}$$

Therefore, $\pm\hat{\boldsymbol{e}}_{(\mathbf{1})}$ can be used to define a time-orientation in the $r < R_S$ region. To decide which sign to use, we need to ensure that whichever we choose to be future directed inside is also future-directed outside. Since for $r > R_S$ and for any function f independent of θ and ϕ:

$$\left.\frac{\partial f}{\partial t}\right|_{(r,\theta,\phi)} = \left.\frac{\partial v}{\partial t}\right|_{(r,\theta,\phi)} \left.\frac{\partial f}{\partial v}\right|_{(r,\theta,\phi)} + \left.\frac{\partial r}{\partial t}\right|_{(r,\theta,\phi)} \left.\frac{\partial f}{\partial r}\right|_{(v,\theta,\phi)} \tag{4.228}$$

$$= \left.\frac{\partial f}{\partial v}\right|_{(r,\theta,\phi)} , \tag{4.229}$$

we have $\hat{\boldsymbol{e}}_{(\mathbf{0})} = \boldsymbol{e}_{(\mathbf{0})}$. Thus:

$$\boldsymbol{g}\left(-\hat{\boldsymbol{e}}_{(\mathbf{1})}, \boldsymbol{e}_{(\mathbf{0})}\right) = \boldsymbol{g}\left(-\hat{\boldsymbol{e}}_{(\mathbf{1})}, \hat{\boldsymbol{e}}_{(\mathbf{0})}\right) = -1 < 0 , \tag{4.230}$$

we see that $-\hat{\boldsymbol{e}}_{(\mathbf{1})}$ is future-directed in the exterior region. We can then pick it as our future direction inside. Clearly, that means that photons on the $v =$ cst curves in the interior region fall towards decreasing r. Physically, $-\hat{\boldsymbol{e}}_{(\mathbf{1})}$ is oriented along the future-travelling photons' path both outside and inside.

On the other hand, type IIv corresponds to outward geodesics in the region $r > R_S$ and to the inward ones in the region $r < R_S$ (again just re-express the condition in terms of t and r). This

translates the fact that lightcones bend as they approach the hypersurface $r = R_S$ and flips once they pass that limit. Actually, $r = R_S$ is itself generated by type IIv lightlike geodesics. The future light rays of type IIv in the region $r < R_S$ are oriented upward because denoting $\boldsymbol{K}$ their tangent vector we have that:

$$\underbrace{\boldsymbol{g}\left(\boldsymbol{K}, -\hat{\boldsymbol{e}}_{(\mathbf{1})}\right)}_{<0} = -g_{vr}K^v = -K^v = -\frac{\mathrm{d}v}{\mathrm{d}\lambda} = -\underbrace{\frac{\mathrm{d}v}{\mathrm{d}r}}_{<0}\frac{\mathrm{d}r}{\mathrm{d}\lambda}\,, \tag{4.231}$$

so that $\frac{\mathrm{d}r}{\mathrm{d}\lambda} = U^r < 0$. All this is summarized on Fig. 4.15.

The event horizon and the nature of a black hole

The hypersurface $r = R_S$, that we will denote $\mathcal{H}$ is called the *event horizon*. The vector $\hat{\boldsymbol{e}}_{(\mathbf{0})} = \frac{\partial}{\partial v}$ is lightlike at $r = R_S$ and it is both orthogonal and tangent to $\mathcal{H}$ which makes this hypersurface a lightlike one. The event horizon, not the central singularity is what makes of the Schwarzschild spacetime a black hole. Although future-directed curves in the region $r > R_S$ can either enter the interior region or escape to infinity, future-directed curves in the interior region and on the horizon are trapped. This can be summarised by the following result:

Characterisation of the event horizon $\mathcal{H}$

Let $x^\mu(\lambda)$ be any future-directed causal curve (not necessarily a geodesic). If $r(\lambda_0) \leq R_S$ for some λ_0, then $r(\lambda) \leq R_S$ for any $\lambda \geq \lambda_0$.

Indeed, let us pick up one such future-directed curve parametrised by λ, with $r(\lambda_0) \leq R_S$ and a non-zero tangent vector $\boldsymbol{U}$. Since it is future directed, we have:

$$\boldsymbol{g}\left(-\hat{\boldsymbol{e}}_{(\mathbf{1})}, \boldsymbol{U}\right) = -g_{r\mu}U^\mu = -U^v = -\frac{\mathrm{d}v}{\mathrm{d}\lambda} \leq 0\,. \tag{4.232}$$

Along the curves, v is thus constant or increasing. Besides, we have:

$$\boldsymbol{g}\left(\boldsymbol{U}, \boldsymbol{U}\right) = -\left(1 - \frac{R_S}{r}\right)\left(\frac{\mathrm{d}v}{\mathrm{d}\lambda}\right)^2 + 2\frac{\mathrm{d}v}{\mathrm{d}\lambda}\frac{\mathrm{d}r}{\mathrm{d}\lambda} + r^2\left(\frac{\mathrm{d}\Omega}{\mathrm{d}\lambda}\right)^2\,, \tag{4.233}$$

where:

$$\left(\frac{\mathrm{d}\Omega}{\mathrm{d}\lambda}\right)^2 = \left(\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right)^2 + \sin^2\theta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\lambda}\right)^2\,. \tag{4.234}$$

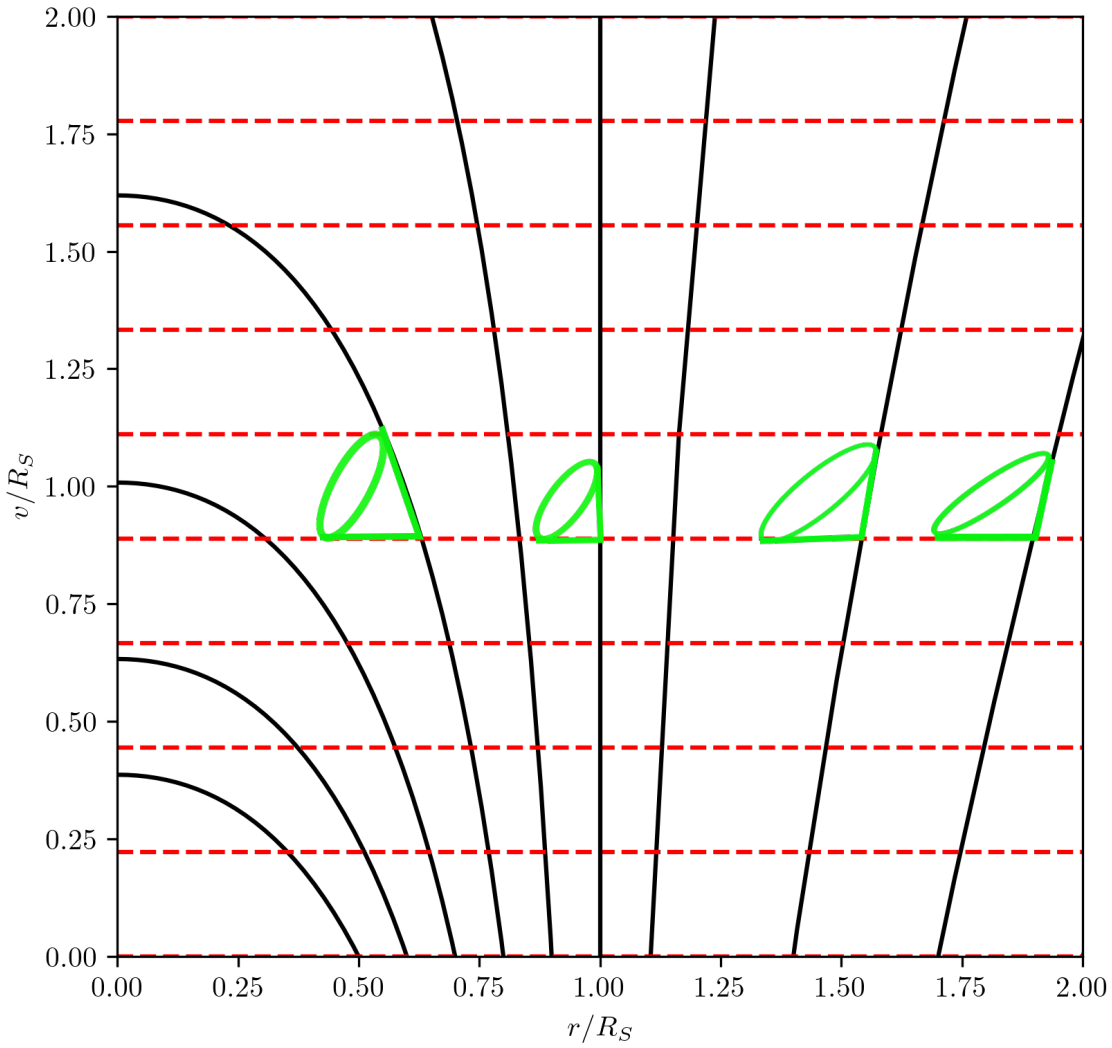


Figure 4.15: Radial null geodesics of the Schwarzschild metric in Eddington-Filkenstein infalling coordinate. Type Iv geodesics are represented in dashed red and type IIv in solid black. We see that local lightcones gradually bend towards the central singularity as one moves along infalling radial geodesics. When they pass $r = R_S$, no radial light rays are outgoing and the entire lightcone points towards the central singularity at $r = 0$. At $r = R_S$, the type IIv align with the hypersurface at $r = R_S$: these light rays get trapped at constant r.

Thus:

$$-2\frac{\mathrm{d}v}{\mathrm{d}\lambda}\frac{\mathrm{d}r}{\mathrm{d}\lambda} = -\boldsymbol{g}\left(\boldsymbol{U},\boldsymbol{U}\right) - \left(1-\frac{R_S}{r}\right)\left(\frac{\mathrm{d}v}{\mathrm{d}\lambda}\right)^2 + r^2\left(\frac{\mathrm{d}\Omega}{\mathrm{d}\lambda}\right)^2 . \tag{4.235}$$

In the region $r \leq R_S$ all the terms on the RHS are positive or zero so that we have:

$$\frac{\mathrm{d}v}{\mathrm{d}\lambda}\frac{\mathrm{d}r}{\mathrm{d}\lambda} \leq 0 . \tag{4.236}$$

If $\frac{\mathrm{d}r}{\mathrm{d}\lambda} > 0$, then we must have $\frac{\mathrm{d}v}{\mathrm{d}\lambda} = 0$ to satisfy Eq. (4.232). But then, Eq. (4.235) imposes that $\boldsymbol{g}\left(\boldsymbol{U},\boldsymbol{U}\right) = \left(\frac{\mathrm{d}\Omega}{\mathrm{d}\lambda}\right)^2 = 0$. Thus:

$$\boldsymbol{U} = \frac{\mathrm{d}r}{\mathrm{d}\lambda}\hat{\boldsymbol{e}}_{(\mathbf{1})} , \tag{4.237}$$

the coefficient being positive, so that $\boldsymbol{U}$ is past -directed, which is a contradiction. Thus, we must have:

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} \leq 0 \tag{4.238}$$

for any future-directed causal curve in the interior region. This inequality must actually be strict for $r < R_S$, otherwise, the vector would be identically zero, a contradiction. Therefore, we have shown that in the region $r < R_S$ $r(\lambda)$ is a *monotonously decreasing function along any future-directed causal curve*.

For $r = R_S$, things need to be studied separately. Let us assume that $r\left(\lambda_0\right) = R_S$. If $\frac{\mathrm{d}r}{\mathrm{d}\lambda} < 0$ at $\lambda = \lambda_0$, then we get $r < R_S$ immediately after and we are back to the previous reasoning. Thus, we can restrict our analysis to $\frac{\mathrm{d}r}{\mathrm{d}\lambda} = 0$ at $\lambda = \lambda_0$. If $\frac{\mathrm{d}r}{\mathrm{d}\lambda} = 0$ for any $\lambda > \lambda_0$, then the curve stays trapped on $r = R_S$ and we are done. So let us assume that $\frac{\mathrm{d}r}{\mathrm{d}\lambda} > 0$ for any λ slightly "later" that λ_0 (if it became negative we would be back to the case $r < R_S$). At $\lambda = \lambda_0$, we have $\frac{\mathrm{d}v}{\mathrm{d}\lambda} \neq 0$ (otherwise $\boldsymbol{U}$ would be identically zero), so it must be $\frac{\mathrm{d}v}{\mathrm{d}\lambda} > 0$ at $\lambda = \lambda_0$. Locally, we can thus use v as a parameter along the curve. Denoting $v_0 = v\left(\lambda_0\right)$ and dividing Eq. (4.235) by $\left(\frac{\mathrm{d}v}{\mathrm{d}\lambda}\right)^2$, we get:

$$-2\frac{\mathrm{d}r}{\mathrm{d}v} \geq \frac{R_S}{s} - 1 \;\Rightarrow\; 2\frac{\mathrm{d}r}{\mathrm{d}v} \leq 1 - \frac{R_S}{s} . \tag{4.239}$$

So for $v_2 > v_1 > v_0$:

$$2\int_{r(v_1)}^{r(v_2)} \frac{\mathrm{d}r}{1-R_S/r} \leq v_2 - v_1 . \tag{4.240}$$

In the limit $v_1 \to v_0$, $r(v_1) \to R_S$ and the LHS diverges while the RHS remains finite. This is a contradiction and thus, the condition $\frac{\mathrm{d}r}{\mathrm{d}\lambda} > 0$ in the neighbourhood of $\lambda = \lambda_0$ is impossible. This concludes the proof.

What we have shown is that one cannot find a future-directed causal curve, geodesic or not, connecting an event with $r \leq R_S$ to another event with $r > R_S$. Causal curves can cross $\mathcal{H}$ from the exterior to the interior but not the other way around: the event horizon protects causally the exterior region from the interior one. This is what we call a *black hole*. Note that the central singularity $r = 0$ is not part of the spacetime. It should also not be thought of as a point lying "somewhere" in space at the centre of the black hole. Indeed, as we saw, it is rather in the future of the causal curves that cross the horizon and of those who start inside the horizon so it has to be thought of as lying "sometime" in the future of the other points in the Schwarzschild spacetime.

Motion of a massive particle in the black hole region

Consider a massive test particle inside the event horizon. It is not necessarily in free fall and we call τ its proper time.

(a) Show that its radial coordinate must decrease at a minimum rate given by:

$$\left|\frac{\mathrm{d}r}{\mathrm{d}\tau}\right| \geq \sqrt{\frac{R_S}{r} - 1}\,. \tag{4.241}$$

1. (b) Determine the maximum lifetime for such a particle starting at $r = R_S$ before it reaches the singularity at $r = 0$.
   Hint: We have:
   $$\int \sqrt{\frac{x}{1-x}}\mathrm{d}x = \arcsin(\sqrt{x}) - \sqrt{(1-x)x}\,. \tag{4.242}$$

(c) Show that this maximum lifetime is attained for a certain class of free-falling particles in radial orbits. Comment.

4.6.3 Extending the trip: the white hole region

Starting in the exterior region $r > R_S$ and following ingoing radial lightlike geodesics into the future, we arrived in the black hole region by crossing the $r = R_S$ event horizon. Following outgoing radial lightlike geodesics, we would have ended "at infinity" into the Minkowski region. But what if we followed these geodesics into the past? Specifically, where do photons escaping to infinity come from? To do that, it is better to introduce the *outgoing Eddington-Filkenstein coordinate*:

$$u = t - r - R_S \ln\left|\frac{r}{R_S} - 1\right|\,, \tag{4.243}$$

that is constant along outgoing lightlike rays in the region $r > R_S$. The metric becomes:

$$\mathrm{d}s^2 = -\left(1 - \frac{R_S}{r}\right)\mathrm{d}u^2 - 2\mathrm{d}u\mathrm{d}r + r^2\mathrm{d}\Omega^2 \ . \tag{4.244}$$

Radial light rays are thus of two types:

- Type Iu. These are the rays with:

$$\frac{\mathrm{d}u}{\mathrm{d}r} = -2\left(1 - \frac{R_S}{r}\right)^{-1} \ . \tag{4.245}$$

  Integrating this equation leads to the equation of the rays and one sees that they correspond to the ingoing rays at $\mathrm{d}v = 0$, i.e. to type Iv.

- Type IIu. These obey $\mathrm{d}u = 0$, i.e. $u = \text{cst}$ and they correspond to the outgoing rays, by construction, i.e. to type IIv.

If we follow type Iu rays into the future, we arrive in the black hole region, as we saw in the previous subsection. Following them into the past lead us to the asymptotically flat Minkowski region far from the black hole so we do not learn anything (see more on that later). Following type IIu rays into the future also leads us to this Minkowski region, although in a different corner (see below). But what if we follow them into the past?[7]

We denote by $\tilde{\boldsymbol{e}}_{(0)} = \frac{\partial}{\partial u}$ and $\tilde{\boldsymbol{e}}_{(1)} = \frac{\partial}{\partial r}\Big|_{(u,\theta,\phi)}$, and we can see that in the region $r > R_S$, $\tilde{\boldsymbol{e}}_{(0)} = \boldsymbol{e}_{(0)}$. Following the rays at $u = \text{cst}$ into the past amounts to decreasing r along the geodesics while also decreasing t: $\frac{\mathrm{d}r}{\mathrm{d}\lambda} < 0$ and $\frac{\mathrm{d}t}{\mathrm{d}\lambda} < 0$. These can be followed backwards up to $r = 0$, crossing a $r = R_S$ hypersurface smoothly. But you can see that the region $r < R_S$ in which we arrive is *not* the same as the black hole region. Indeed, if we now run the film towards the future, we see radial lightlike geodesics escape from the $r < R_S$ region into the $r > R_S$ one. This can not happen in the black hole region, as we have seen above. We have discovered a new region of spacetime! It can also be covered by local (t, r), Schwarzschild coordinates, with the usual metric components, but its nature is completely unexpected. To get a time orientation in this region that is consistent with the one outside we must now choose $+\tilde{\boldsymbol{e}}_{(1)} = \frac{\partial}{\partial r}\Big|_{(u,\theta,\phi)}$ so that future-directed photons on the $\mathrm{d}u = 0$

[7] Be careful, we are talking about radial rays here. Clearly some infalling non-radial light rays do not end up in the black hole; think about those that scatter back to infinity and for which we calculated a deviation angle; some are also trapped in orbit.

curves must move towards larger values of r. Moreover, let us call $\boldsymbol{K}$ the future-directed tangent vector to a type Iu ray. In the $r > R_S$ region, we have:

$$\underbrace{\boldsymbol{g}\left(\boldsymbol{K}, \tilde{\boldsymbol{e}}_{(\mathbf{0})}\right)}_{<0} = g_{uu} K^u \tag{4.246}$$

$$= -\left(1 - \frac{R_S}{r}\right) \frac{\mathrm{d}u}{\mathrm{d}\lambda} \tag{4.247}$$

$$= 2\frac{\mathrm{d}r}{\mathrm{d}\lambda} < 0\,, \tag{4.248}$$

so these rays are ingoing. Similarly, in the $r < R_S$ region:

$$\underbrace{\boldsymbol{g}\left(\boldsymbol{K}, \tilde{\boldsymbol{e}}_{(\mathbf{1})}\right)}_{<0} = g_{ur} K^r \tag{4.249}$$

$$= 2\underbrace{\left(1 - \frac{R_S}{r}\right)^{-1}}_{<0} \underbrace{\frac{\mathrm{d}r}{\mathrm{d}\lambda}}_{>0}\,, \tag{4.250}$$

so the ray are outgoing.

It is a white hole. It is a region of spacetime from which future-directed causal curves can escape but which cannot be entered by them. It lies into the past of the rest of the Schwarzschild spacetime, including the black hole region and its asymptotically Minkowski surroundings. It contains a singularity at $r = 0$ in local Schwarzschild coordinates; this is not the same set as the one with $r = 0$ in the black hole region. The situation is depicted in Fig. 4.16·

4.6.4 A bird's eye view of the Schwarzschild geometry

Let us recap what we have learned about Schwarzschild spacetime, i.e. the unique spherically symmetric, vacuum and asymptotically flat spacetime.

1. It has an exterior region, $r > R_S$ that is asymptotically Minkowski. This is the region we studied when probing the behaviour of matter and light around a compact object such as a star. We can call it *region I*.

2. If one follows the radial light rays emanating from that region and falling towards the $0 < r < R_S$ region, one ends up in a trapped region of spacetime, from which no future-directed

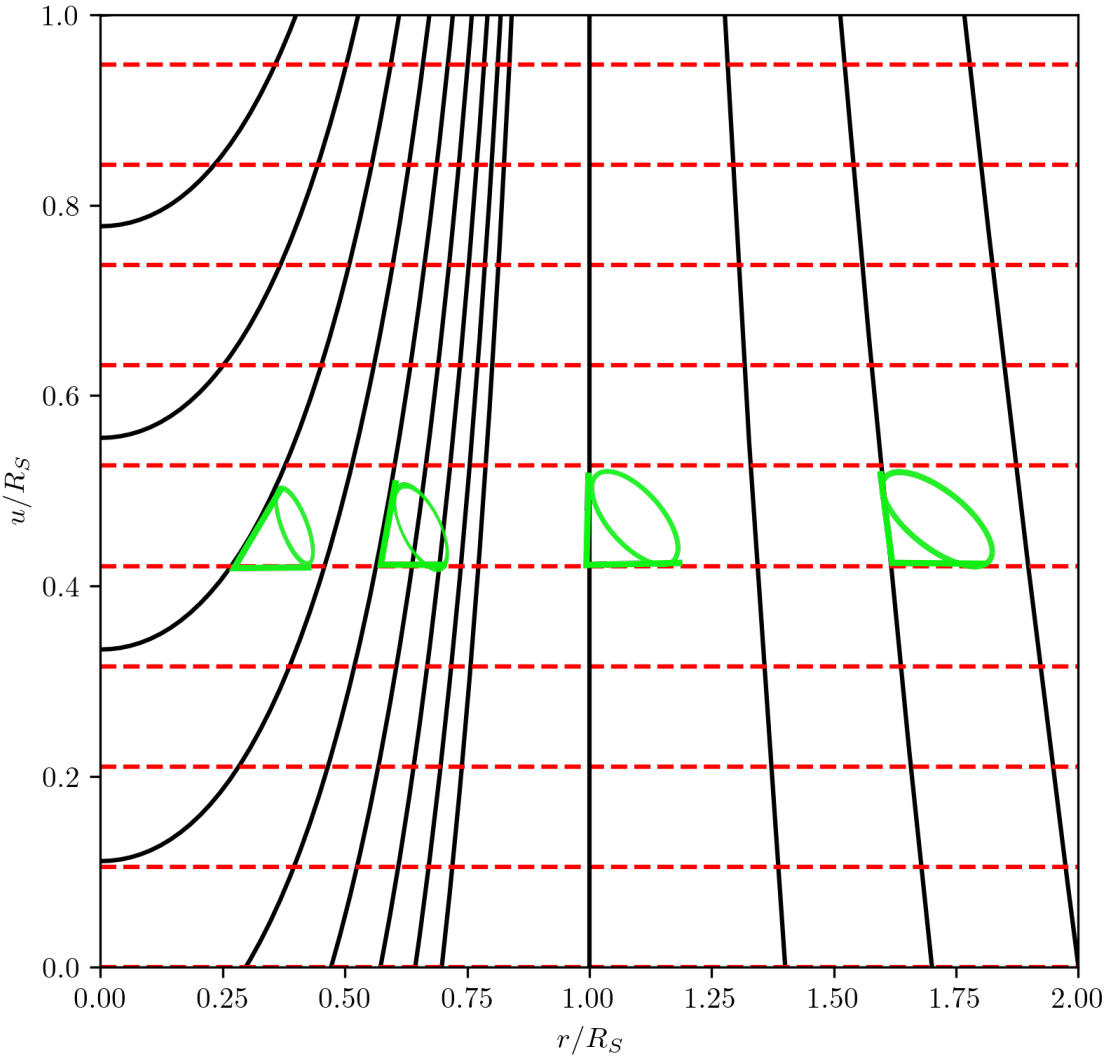


Figure 4.16: The white hole region of the Schwarzschild spacetime. Radial lightrays with $\mathrm{d}u = 0$ are in dashed red and type I ones in solid black. Future-directed lightcones open up while exiting the white hole region, here spanned by $0 < r < R_S$ so that the exterior region lies in the future of the interior one. The white hole singularity at $r = 0$ is in the past of the Schwarzschild spacetime.

causal curve, timelike of lightlike, can escape. This region is bounded by a lightlike surface, the event horizon $\mathcal{H}$, at $r = R_S$ that acts as a fictitious one-way membrane. It contains a true

singularity that sits in the future of all events inside the horizon and of all infalling geodesics emanating from region I. This is the black hole region; let us call it *region II*.

3. If one follows radial light rays emanating from region I and going away from the black hole region by tracing them back into the past, one arrives at a new region that can also be spanned by $0 < r < R_S$ but that is not the black hole region. Rather, it is in the past of regions I and II. Let us call it *region III*. It contains a singularity at $r = 0$ that is in the past of some causal curves ending up in regions I and II. It is surrounded by a hypersurface at $r = R_S$ that acts as a membrane in the opposite way to the horizon $\mathcal{H}$: no future-directed causal curve can cross it from region II. It is called the white hole region.

To discover these regions, we used two sets of coordinates, one to look into the future of region I (using the infalling Eddington-Filkenstein coordinate v) and one to look into its past (using the outgoing Eddington-Filkenstein coordinate u). Here, we would like to find a coordinate system that allows us to cover all those regions at once. A first attempt might be to use u and v to get a chart, i.e. to use radial light rays to label points in spacetime. In that case, we obtain the following expression for the line element:

$$\mathrm{d}s^2 = -\left(1 - \frac{R_S}{r}\right)\mathrm{d}u\mathrm{d}v + r^2(u,v)\mathrm{d}\Omega^2 \ . \tag{4.251}$$

The problem is that the hypersurfaces at $r = R_S$, which are so important to understanding this spacetime are sent to infinity in these coordinates: $u(r \to R_S) = +\infty$ going towards the black hole regions and $v(r \to R_S) = -\infty$ going towards the white hole region. This is most unsavoury and we will want to "bring them back closer". Therefore, let us introduce the coordinates (U, V) defined in patches. For region I:

$$\begin{cases} U = -\mathrm{e}^{-u/2R_S} & (4.252) \\ V = \mathrm{e}^{v/2R_S} \ , & (4.253) \end{cases}$$

for region II:

$$\begin{cases} U = \mathrm{e}^{-u/2R_S} & (4.254) \\ V = \mathrm{e}^{v/2R_S} \ , & (4.255) \end{cases}$$

and for region III:

$$\begin{cases} U = -\mathrm{e}^{-u/2R_S} & (4.256) \\ V = -\mathrm{e}^{v/2R_S} \ . & (4.257) \end{cases}$$

In terms of (t, r) in region I, we get:

$$\begin{cases} U = -\sqrt{\dfrac{r}{R_S} - 1}\, \mathrm{e}^{(r-t)/2R_S} & (4.258) \\ V = \sqrt{\dfrac{r}{R_S} - 1}\, \mathrm{e}^{(t+r)/2R_S} \ , & (4.259) \end{cases}$$

and similar expressions for the other regions. The limits above are clearly satisfied, the crossing from one region to another happens smoothly, and we retain that radial light rays fall into two categories: $U = \text{cst}$ or $V = \text{cst}$. Note that in all regions:

$$UV = \left(1 - \frac{r}{R_S}\right) \mathrm{e}^{r/R_S} \ , \tag{4.260}$$

so that $r = R_S$ corresponds to $UV = 0$. We have $U = 0$ if $\mathrm{e}^{(r-t)/2R_S}$ does not diverge when we approach $r = R_S$, i.e. we must have $t \to +\infty$: this is the event horizon $\mathcal{H}$. Conversely, $V = 0$ happens on the white hole "anti-horizon". The line element then becomes (exercise):

$$\mathrm{d}s^2 = -\frac{4R_S^3}{r(U,V)} \mathrm{e}^{-r(T,R)/R_S} \mathrm{d}U\mathrm{d}V + r^2(U,V)\mathrm{d}\Omega^2 \ , \tag{4.261}$$

and it is perfectly regular for $(U, V) \in \mathbb{R}^2$. But regions I, II and III only cover three quadrants in the plane and there is thus a fourth one that is accessible in these coordinates and that we have not explored yet, with $U > 0$ and $V < 0$. What is it?

U and V are null coordinates and it would be nice, for intuition purposes to get timelike and spacelike coordinates instead. Let us simply write:

$$\begin{cases} T = \dfrac{1}{2}(V + U) & (4.262) \\ R = \dfrac{1}{2}(V - U) \ . & (4.263) \end{cases}$$

Then, we have:

$$-\mathrm{d}T^2 + \mathrm{d}R^2 = -\mathrm{d}U\mathrm{d}V \ , \tag{4.264}$$

so that the line element in *Kruskal coordinates* becomes:

$$\mathrm{d}s^2 = \frac{4R_S^3}{r(T,R)} \mathrm{e}^{-r(T,R)/R_S} \left(-\mathrm{d}T^2 + \mathrm{d}R^2\right) + r^2(T,R)\mathrm{d}\Omega^2 \ . \tag{4.265}$$

Finally note the useful relation:

$$T^2 - R^2 = UV = \left(1 - \frac{r}{R_S}\right) \mathrm{e}^{r/R_S} \ . \tag{4.266}$$

These coordinates are extremely powerful because as long as we ignore the angular part of the metric, which is fine for a spherically symmetric spacetime when we want to study the causal structure, the spacetime is *conformally related* to Minkowski spacetime, i.e. conformally flat. Then, for radial light rays: $\mathrm{d}s^2 = 0$ and we simply have: $\mathrm{d}T^2 = \mathrm{d}R^2$. This means that the conformal spacetime diagram, obtained by setting $\mathrm{d}\Omega = 0$ in Eq. (4.265) and by ignoring the non-zero prefactor, is going to look very simple since local lightcones will always be straight lines at $\pm\pi/4$. Let us construct this diagram and for that, list a few properties:

- Radial null geodesics are given by $T = \pm R + \mathrm{cst}$.
- Hypersurfaces $r = R_S$ are given by $T^2 - R^2 = 0$, i.e. $T = \mp R$.
- Hypersurfaces $r = \mathrm{cst}$, i.e. worldline of static observers are hyperbolæ with $T^2 - R^2 = \left(1 - \frac{r}{R_S}\right) \mathrm{e}^{r/R_S}$.
- Hypersurfaces $t = \mathrm{cst}$, are at:

$$\frac{T}{R} = \tanh\left(\frac{t}{2R_S}\right) \quad \text{if} \quad r > R_S \tag{4.267}$$

$$= 1/\tanh\left(\frac{t}{2R_S}\right) \quad \text{if} \quad r < R_S \ . \tag{4.268}$$

- The coordinates T and R are not allowed to run in the all of $\mathbb{R}^2$ because of the physical singularity at $r = 0$. Imposing $r > 0$ results in:

$$T^2 - R^2 = \left(1 - \frac{r}{R_S}\right) \mathrm{e}^{r/R_S} < 1 \ , \tag{4.269}$$

  so that, for any value of $R \in \mathbb{R}$, we must have:

$$T^2 < R^2 + 1 \ . \tag{4.270}$$

This is summarised in the *Kruskal diagram*, Fig. 4.17. We discover that there is a fourth region in the Schwarzschild spacetime, that denoted region IV on the diagram, corresponding to $\{U >$

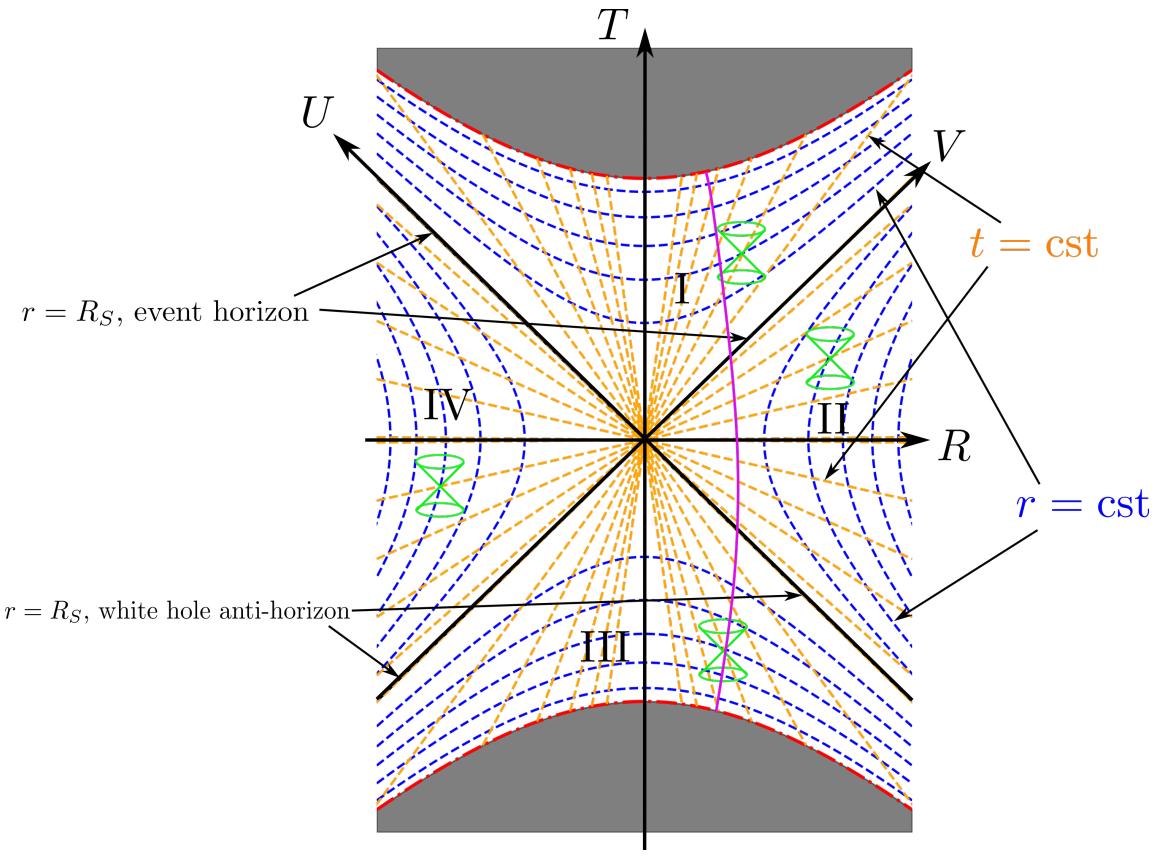


Figure 4.17: The Kruskal diagram of the Schwarzschild spacetime. Blue dashed lines represent hypersurfaces with r = cst and dashed orange ones hypersurfaces with t = cst. Local lightcones are always at an angle of $\pm\pi/4$ and are lines with U = cst or V = cst; a few are represented in green here. The U and V axes represent the surfaces $r = R_S$, clearly showing their lightlike nature. The singularities are in dot-dashed red. The four regions are shown as quadrants in the planes, marked I, II, III and IV. The grey, shaded regions are not part of the spacetime. The purple curve is a timelike curve emerging from the white hole singularity into the exterior region I before plunging into the black hole region and hitting the black hole singularity.

$0\,,\, V < 0\}$. It is completely identical, locally, to region I, i.e. the exterior region we started from: it is asymptotically Minkowski and it continues to the black hole region in the future and to the white hole region in the past. We can say that it is another exterior "Universe". Note that no causal signal or observer can cross from region II to region IV or the other way around so these two "Universes"

are completely isolated[8].

The fact that the black hole and white hole singularities are extended curves located respectively in the future and past of region I and IV is now completely apparent. Look at the purple curve, which represents the trajectory of a massive particle (not in free fall) emerging from the white hole and plunging into the black hole. The Kruskal coordinate system gives the maximal extension of the Schwarzschild spacetime in the sense that it covers all the points accessible by following causal curves: we have extended timelike and lightlike geodesics as far as we could. It corresponds to the spacetime of an *eternal black hole*.

As it turns out, there are still points whose status is a bit unclear on this diagram: the end points of geodesics that escape to infinity in regions II and IV. There is a technique to bring them back at "finite distance" on paper, called the construction of the Carter-Penrose diagram of the spacetime. These very useful object will not be studied here, by lack of time.

4.6.5 Astrophysical black holes

As we have seen, the Kruskal extension corresponds to an eternal black hole. Astrophysical black holes, on the other hand, are formed by the collapse of a star. This means that the causal past of points in the exterior region II does not contain a white hole. Instead. it contains... a star. Besides, the event horizon will only form when the radius of this star passes below $r = R_S$. The Kruskal diagram will thus look something like what is show in Fig 4.18. Regions III and IV have disappeared.

[8]They are, in fact, connected by an Einstein-Rosen bridge, or Schwarzschild wormhole that can be constructed by slicing the Kruskal diagram with t = cst straight lines. Since these are spatial hypersurfaces though, the brigde, which connects both regions at $U = V = 0$ is spacelike and cannot be crossed.

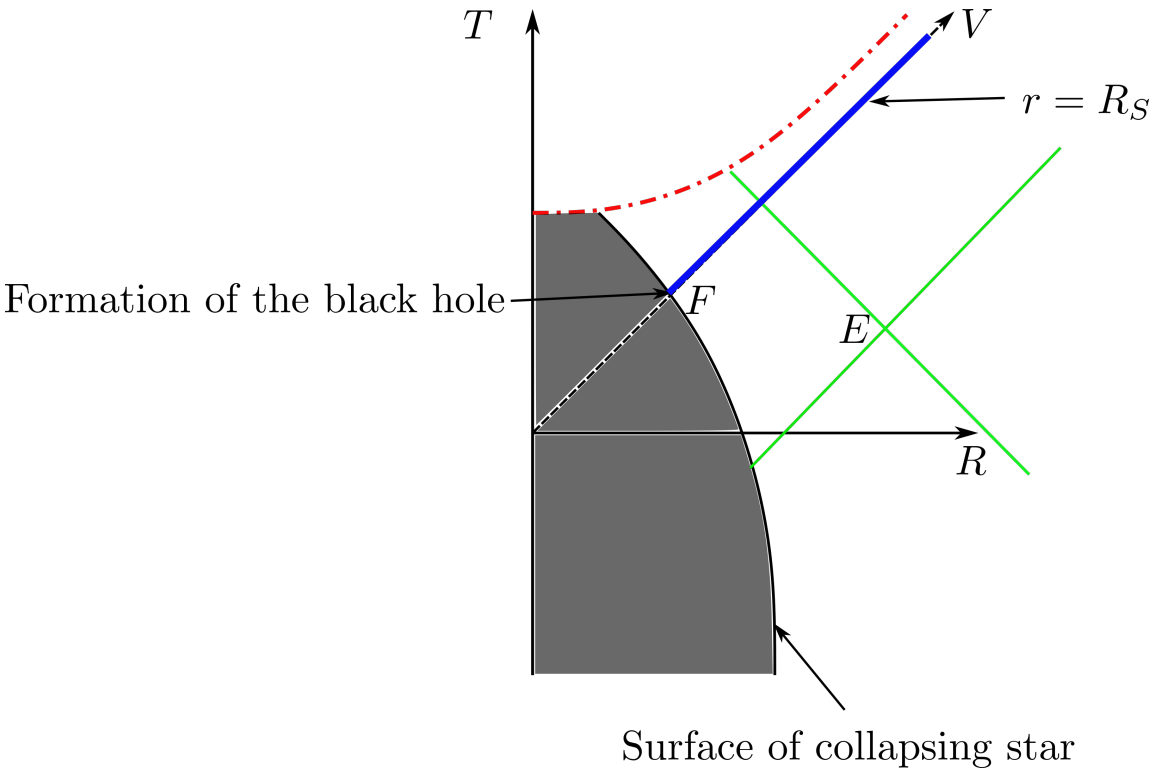


Figure 4.18: Collapse of a star in Kruskal coordinates. The shaded region corresponds to the inside of the star and is not described by the Schwarzschild solution. The black hole only forms when the star collapses below its Schwarzschild radius, at event F, the event horizon then appears (blue thick line). Before that, the spacetime is just exterior Schwarzschild with a star at its centre. The green lines at an angle $\pm\pi/4$ represent the radial light rays reaching or leaving the event E outside the star. We see that outgoing rays do not fall into a white hole in the past, but rather intersect the surface of the star. Only regions I and II of the Kruskal extension survive.

5

Free field solutions: gravitational waves

Contents

5.1 Introduction

In this chapter, we first need to understand what it means for a spacetime with a given geometry to be "close" to another one. That will lead us to the notion of a gauge[1]. As an application, we will see how to recover the weak field geometry we have been using in these notes so far. Then, we will develop the theory of gravitational waves. As every classical field theory, for example electromagnetism, General Relativity admits freely propagating solutions in vacuum, aka free waves. But unlike electromagnetic waves, which can be produced by charge dipoles, gravitational waves can only be produced by at least quadrupolar distributions of mass-energy. We will show that they have two degrees of polarisation, study how they impact matter, and try to understand how they are produced. Gravitational waves are very, very weak, and only those produced by extreme astrophysical systems like merging binary black holes or binaries made of neutron stars and black holes can be detected on Earth. Actually, these have only been detected for the first time in 2015 by the LIGO experiment [1], despite having been predicted by Einstein in 1916 [9] (article amended in 1918 [10]). However, for the past 6 years, we have been detecting them almost routinely, and we start being able to do some astrophysics with their observations. Perturbation theory is also central to the development of modern cosmology, as we will see in M2.

5.2 Perturbation theory

5.2.1 Perturbing a spacetime

It is usual in physics to try and approach complex systems lacking any apparent symmetry by trying to describe them as only slightly non-symmetrical, and related to a highly symmetrical, well-known, physical system by a small perturbation. For example, as a first approximation, the surface of the Earth is well-approximated by a sphere, and departures from sphericity such as ellipticity due to rotation, mountains and valleys etc. can be described as small hierarchical perturbations around a perfectly spherical, idealised Earth. The spherical Earth model is what we will call a background geometry, while the corrections to sphericity will be called perturbations of the geometry. The advantage of such a description is that a sphere is a highly symmetrical object, thus quantities and dynamics can be easily calculated exactly on it (equations are easier to solve on a sphere that on a gen-

[1] Be careful: this is a related, but distinct concept from the usual "gauge" of field theory.

eral "bumpy" surface). Then corrections to these quantities and dynamics due to the non-sphericity can be calculated order by order in importance of the perturbers on various relevant scales.
Now, there is an ambiguity here due to the very symmetric nature of a sphere. One can label points on the sphere by their latitude and longitude but of course, these are completely arbitrary in the sense that latitudes depend on identifying poles, while longitudes depend on selecting a reference meridian. Thus, given a mountain on Earth considered as a small perturbation on the shape of the surface, locating it at a given latitude and longitude is completely arbitrary. This means that whatever impact the mountain has on physical quantities cannot depend on the point of the idealised spherical model at which we have anchored it: the symmetries of the "background" model introduce some indetermination in the perturbed model, and this indetermination has to be removed (physicists say "gauged" out) once physical quantities are constructed. In essence this is the gauge problem in General Relativity. Let us see how it works in details.
We start with a highly-symmetrical background spacetime $(\bar{M}, \bar{\boldsymbol{g}})$, where $\bar{M}$ is a differentiable manifold and $\bar{\boldsymbol{g}}$ a Lorentzian metric on M which is a known, exact solution of the Einstein Field equations. $\bar{\boldsymbol{g}}$ is usually highly symmetrical, e.g. Minkowski, Schwarzschild, Friedmann-Lemaître-Robertson-Walker etc. We consider a second differentiable manifold M, which is diffeomorphic to $\bar{M}$ so they could be treated as the same manifold, up to identifying points in M and points in $\bar{M}$. Let us pick such an identification by selecting a specific diffeomorphism $\phi : \bar{M} \to M$. This choice is arbitrary, and this will be important in what follows. Next, we pick a Lorentzian metric $\boldsymbol{g}$ on M. We would like to make precise the following statement:

$$\text{The manifold } (M, \boldsymbol{g}) \text{ is close to the manifold } \left(\bar{M}, \bar{\boldsymbol{g}}\right).$$

Let $p \in \bar{M}$ and a local chart $(U, \bar{\varphi})$ around p such that $\bar{\varphi}(p) = \bar{x}$. Let (V, φ) be a local chart of M containing $\phi\,(U)$, such that $\varphi\,(\phi(p)) = x$. In order to compare the metric $\boldsymbol{g}$ to the metric $\bar{\boldsymbol{g}}$, we are going to pullback $\boldsymbol{g}$ onto $\bar{M}$, using our selected diffeomorphism ϕ. A pictorial representation of the set-up can be found in figure 5.1. Given:

$$\boldsymbol{g} = g_{\alpha\beta} \mathrm{d}x^{\alpha} \otimes \mathrm{d}x^{\beta}\,, \tag{5.1}$$

we get the pullback metric on $\bar{M}$, $\boldsymbol{\phi}^{*}\boldsymbol{g}$ defined, for any two vector fields $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ on $\bar{M}$, by:

$$\left(\boldsymbol{\phi}^{*}\boldsymbol{g}\right)\left(\bar{\boldsymbol{X}}, \bar{\boldsymbol{Y}}\right) \equiv \boldsymbol{g}\left(\boldsymbol{\phi}_{*}\bar{\boldsymbol{X}}, \boldsymbol{\phi}_{*}\bar{\boldsymbol{Y}}\right)\,, \tag{5.2}$$

where $\boldsymbol{\phi}_{*}\bar{\boldsymbol{X}}$ and $\boldsymbol{\phi}_{*}\bar{\boldsymbol{Y}}$ are the pushforward of $\bar{\boldsymbol{X}}$ and $\bar{\boldsymbol{Y}}$ onto M; see appendix B. Then, the symmetric, rank 2 tensor $\boldsymbol{\phi}^{*}\boldsymbol{g}$ is well-defined on $\bar{M}$ and defines a new metric tensor on the background, which

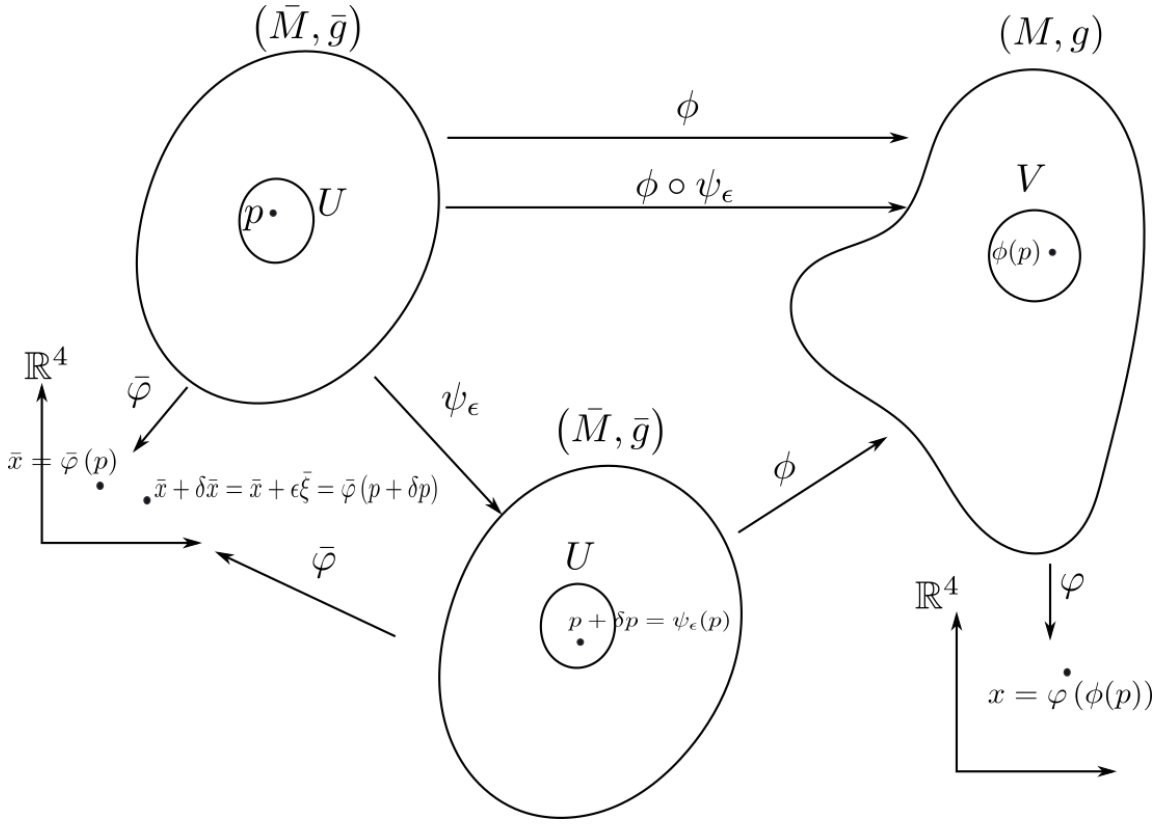


Figure 5.1: Sets and maps necessary to set-up the gauge transformations.

can thus be compared to the background metric $\bar{\boldsymbol{g}}$ pointwise. We define the difference between the two metrics as:

$$\boldsymbol{h} = \boldsymbol{\phi}^* \boldsymbol{g} - \bar{\boldsymbol{g}} \tag{5.3}$$

as a symmetric rank-two tensor on the background $\bar{M}$, such that:

$$h_{\mu\nu}(\bar{x})\, \mathrm{d}\bar{x}^\mu \otimes \mathrm{d}\bar{x}^\nu = \left[(\phi^* g)_{\mu\nu}(\bar{x}) - \bar{g}_{\mu\nu}(\bar{x})\right] \mathrm{d}\bar{x}^\mu \otimes \mathrm{d}\bar{x}^\nu \ . \tag{5.4}$$

We will say that $(M, \boldsymbol{g})$ is close to $(\bar{M}, \bar{\boldsymbol{g}})$, or is a perturbed spacetime with respect to the background $(\bar{M}, \bar{\boldsymbol{g}})$ if and only if we can find one diffeomorphism ϕ between $\bar{M}$ and M such that the components of $\boldsymbol{h}$ are small (compared to 1). In that case, $\boldsymbol{h}$ is called a perturbation to the background metric $\bar{\boldsymbol{g}}$. Note that there is no reason whatsoever for the components $\left|h_{\mu\nu}\right|$ to be small for an arbitrary diffeomorphism ϕ. Nevertheless, if we assume that there is such a ϕ that leads to a small difference tensor $\boldsymbol{h}$, then there is an infinite number of diffeomorphisms between $\bar{M}$ and M which keep the metrics $\boldsymbol{g}$ and $\bar{\boldsymbol{g}}$ close. Indeed, consider an arbitrary vector field on $\bar{M}$:

$$\bar{\boldsymbol{\xi}} = \bar{\xi}^\mu \frac{\partial}{\partial \bar{x}^\mu} \ . \tag{5.5}$$

Let $\epsilon \in \mathbb{R}$, small. Then we can define a one-parameter family of diffeomorphisms $\psi_\epsilon : \bar{M} \to \bar{M}$ by displacing points of $\bar{M}$ along the flow of $\bar{\boldsymbol{\xi}}$ by an amount ϵ:

$$\forall \bar{x} = \bar{\varphi}(p) \in \bar{\varphi}(U),\ \bar{y}^\mu = [\bar{\varphi}(p + \delta p)]^\mu = \bar{x}^\mu + \delta \bar{x}^\mu = \bar{x}^\mu + \epsilon \bar{\xi}^\mu \ . \tag{5.6}$$

Then, by construction, $\phi \circ \psi_\epsilon$ will also be a diffeomorphism between $\bar{M}$ and M, for $|\epsilon| \ll 1$. Thus we can pick up any one of those to define our metric perturbation $\boldsymbol{h}$, so that we are left with a family of perturbations, indexed by a choice of the vector field $\bar{\boldsymbol{\xi}}$:

$$\begin{aligned} \boldsymbol{h}^{(\epsilon)} &\equiv (\boldsymbol{\phi} \circ \boldsymbol{\psi}_\epsilon)^* \boldsymbol{g} - \bar{\boldsymbol{g}} && (5.7) \\ &= \left(\boldsymbol{\psi}_\epsilon^* \left(\boldsymbol{\phi}^* \boldsymbol{g}\right)\right) - \bar{\boldsymbol{g}} \ . && (5.8) \end{aligned}$$

How are members of this family of perturbations related to each other?
We can notice that:

$$\begin{aligned} \boldsymbol{h}^{(\epsilon)} &= \psi_\epsilon^* (\boldsymbol{h} + \bar{\boldsymbol{g}}) - \bar{\boldsymbol{g}} && (5.9) \\ &= \boldsymbol{\psi}_\epsilon^* \boldsymbol{h} + \boldsymbol{\psi}_\epsilon^* \bar{\boldsymbol{g}} - \bar{\boldsymbol{g}} \ \text{(linearity of pullback)} && (5.10) \\ &= \boldsymbol{h} + \boldsymbol{\psi}_\epsilon^* \bar{\boldsymbol{g}} - \bar{\boldsymbol{g}} \ \left(\boldsymbol{\psi}_\epsilon^* \boldsymbol{h} = \boldsymbol{h} \text{ at leading order since} \epsilon,\ \|\boldsymbol{h}\| \ll 1\right) && (5.11) \\ &= \boldsymbol{h} + \epsilon \frac{\boldsymbol{\psi}_\epsilon^* \bar{\boldsymbol{g}} - \bar{\boldsymbol{g}}}{\epsilon} \ . && (5.12) \end{aligned}$$

Note that we see the Lie derivative of $\bar{\boldsymbol{g}}$ along $\bar{\boldsymbol{\xi}}$ appear. Indeed, by definition:

$$\mathcal{L}_{\bar{\xi}}\bar{\boldsymbol{g}} = \lim_{\epsilon\to 0}\frac{\boldsymbol{\psi}_{\epsilon}^{*}\bar{\boldsymbol{g}} - \bar{\boldsymbol{g}}}{\epsilon}\ . \tag{5.13}$$

Let us calculate this term. For the ease of notation, let us define:

$$\boldsymbol{G}_{\epsilon} = \boldsymbol{\psi}_{\epsilon}^{*}\bar{\boldsymbol{g}}\ . \tag{5.14}$$

Then, by definition for two arbitrary vector fields $\bar{X}$ and $\bar{Y}$ on $\bar{M}$:

$$\boldsymbol{G}_{\epsilon|p}\left(\bar{\boldsymbol{X}},\bar{\boldsymbol{Y}}\right) \equiv \bar{\boldsymbol{g}}_{|p+\delta p}\left(\boldsymbol{\psi}_{\epsilon,*}\bar{\boldsymbol{X}},\boldsymbol{\psi}_{\epsilon,*}\bar{\boldsymbol{Y}}\right)\ . \tag{5.15}$$

Then, we also have:

$$\begin{aligned}
\left[\boldsymbol{\psi}_{\epsilon,*}\bar{\boldsymbol{X}}\right]^{\mu} &= \frac{\partial\left(\bar{x}^{\mu}+\epsilon\bar{\xi}^{\mu}\right)}{\partial\bar{x}^{\nu}}\bar{X}^{\nu} && (5.16)\\
&= \left(\delta^{\mu}_{\nu}+\epsilon\frac{\partial\bar{\xi}^{\mu}}{\partial\bar{x}^{\nu}}\right)\bar{X}^{\nu} && (5.17)\\
&= \bar{X}^{\mu}+\epsilon\bar{X}^{\nu}\frac{\partial\bar{\xi}^{\mu}}{\partial\bar{x}^{\nu}}\ . && (5.18)
\end{aligned}$$

In particular:

$$\left[\psi_{\epsilon,*}\frac{\partial}{\partial\bar{x}^{\mu}}\right]^{\alpha} = \delta^{\alpha}_{\mu}+\epsilon\frac{\partial\bar{\xi}^{\alpha}}{\partial\bar{x}^{\mu}}\ . \tag{5.19}$$

Finally:

$$\begin{aligned}
\left(G_{\epsilon}\right)_{\mu\nu} &\equiv \boldsymbol{G}_{\epsilon}\left(\frac{\partial}{\partial\bar{x}^{\mu}},\frac{\partial}{\partial\bar{x}^{\nu}}\right) && (5.20)\\
&= \bar{\boldsymbol{g}}(p+\delta p)\left[\psi_{\epsilon,*}\frac{\partial}{\partial\bar{x}^{\mu}},\psi_{\epsilon,*}\frac{\partial}{\partial\bar{x}^{\nu}}\right] && (5.21)\\
&= \bar{g}_{\alpha\beta}\left(\bar{x}^{\sigma}+\epsilon\bar{\xi}^{\sigma}\right)\left[\delta^{\alpha}_{\mu}+\epsilon\frac{\partial\bar{\xi}^{\alpha}}{\partial\bar{x}^{\mu}}\right]\left[\delta^{\beta}_{\nu}+\epsilon\frac{\partial\bar{\xi}^{\beta}}{\partial\bar{x}^{\nu}}\right] && (5.22)\\
&= \left(\bar{g}_{\alpha\beta}\left(\bar{x}^{\sigma}+\epsilon\bar{\xi}^{\sigma}\frac{\partial\bar{g}_{\alpha\beta}}{\partial\bar{x}^{\sigma}}\right)\right)\left(\delta^{\alpha}_{\mu}\delta^{\beta}_{\nu}+\epsilon\delta^{\beta}_{\nu}\frac{\partial\bar{\xi}^{\alpha}}{\partial\bar{x}^{\mu}}+\epsilon\delta^{\alpha}_{\mu}\frac{\partial\bar{\xi}^{\beta}}{\partial\bar{x}^{\nu}}\right) && (5.23)\\
&= \bar{g}_{\mu\nu}\left(\bar{x}\right)+\epsilon\left[\xi^{\sigma}\frac{\partial\bar{g}_{\mu\nu}}{\partial\bar{x}^{\sigma}}+\bar{g}_{\alpha\nu}\frac{\partial\bar{\xi}^{\alpha}}{\partial\bar{x}^{\mu}}+\bar{g}_{\mu\alpha}\frac{\partial\bar{\xi}^{\alpha}}{\partial\bar{x}^{\nu}}\right] && (5.24)\\
&= \bar{g}_{\mu\nu}\left(\bar{x}\right)+2\nabla_{(\mu}\bar{\xi}_{\nu)}\ . && (5.25)
\end{aligned}$$

Thus, we see that:

Gauge transformation

$$h^{(\epsilon)}_{\mu\nu} = h_{\mu\nu} + 2\epsilon\nabla_{(\mu}\bar{\xi}_{\nu)} \ . \tag{5.26}$$

Note that this change at order ϵ vanishes if $\bar{\boldsymbol{\xi}}$ is a Killing vector field of the background metric $\bar{\boldsymbol{g}}$; see appendix B. Such changes in the components of the metric perturbation under an infinitesimal diffeomorphism of $\bar{M}$ along a vector field that is not a Killing vector field of $\bar{\boldsymbol{g}}$ are called gauge transformations for the perturbation. Every two metric perturbations related to each other by a gauge transformation (5.26) for an appropriate choice of field $\bar{\boldsymbol{\xi}}$ represent the same physical configuration, since physical properties cannot depend on the arbitrary choice of $\bar{\boldsymbol{\xi}}$. Therefore, one usually performs perturbative calculations in a specific gauge, i.e. choosing a specific form of the perturbation $\boldsymbol{h}$, but in the end, one must make sure to relate everything that has been calculated to observables from which gauge degrees of freedom have been removed.

Finally, note that we have chosen to present gauge transformations from an active viewpoint, i.e. by shifting points of $\bar{M}$ around while keeping the local charts fixed. One could arrive at the same gauge transformations (5.26) by adopting a passive viewpoint and changing the local charts along the flow of $\bar{\boldsymbol{\xi}}$ while keeping the points fixed, via: $\bar{x}^\mu \mapsto \bar{x}^\mu - \epsilon\bar{\xi}^\mu$.

In the rest of this chapter, we will be interested in perturbations around a Minkowski background, so we will set $\bar{\boldsymbol{g}} = \boldsymbol{\eta}$.

Gauge transformation: the passive viewpoint

Instead of adopting the active viewpoint exposed above, one could use a passive gauge transformation:

$$\bar{x}^\mu \mapsto \bar{x}^\mu - \epsilon\bar{\xi}^\mu \ . \tag{5.27}$$

This amounts to keeping points fixed but changing the local chart along the flow of $-\vec{\bar{\xi}}$. Then:

$$\begin{aligned} \boldsymbol{g} &= \left[\bar{g}_{\alpha\beta}(\bar{x}) + h_{\alpha\beta}(\bar{x})\right] \mathrm{d}\bar{x}^\alpha \otimes \mathrm{d}\bar{x}^\beta & (5.28)\\ &= \left[\bar{g}_{\alpha\beta}(x) - \varepsilon\xi^\gamma \frac{\partial \bar{g}_{\alpha\beta}}{\partial x^\gamma}(\bar{x}) + h_{\alpha\beta}(x)\right] \\ &\quad \times \left[\mathrm{d}x^\alpha + \varepsilon\frac{\partial \xi^\alpha}{\partial x^\gamma}\mathrm{d}x^\gamma\right] \otimes \left[\mathrm{d}x^\beta + \varepsilon\frac{\partial \xi^\beta}{\partial x^\delta}\mathrm{d}x^\delta\right] \ . & (5.29) \end{aligned}$$

Expanding at first order in ε and relabelling dummy indices, we get:

$$\boldsymbol{g} = \left[\bar{g}_{\alpha\beta}(x) + h^{(\varepsilon)}_{\alpha\beta}(x)\right] \mathrm{d}x^{\alpha} \otimes \mathrm{d}x^{\beta} \ , \tag{5.30}$$

with:

$$h^{(\epsilon)}_{\mu\nu} = h_{\mu\nu} + 2\epsilon\nabla_{(\mu}\bar{\xi}_{\nu)} \ . \tag{5.31}$$

5.2.2 Perturbative degrees of freedom

Scalar, vector, tensor decomposition

By construction, the perturbation tensor $\boldsymbol{h}$ is a symmetric rank $(0,2)$ tensor so that it has 10 independent components. In an arbitrary coordinate system, the line element reads:

$$\mathrm{d}s^2 = \left(\eta_{\mu\nu} + h_{\mu\nu}\right) \mathrm{d}x^{\mu}\mathrm{d}x^{\nu} \ . \tag{5.32}$$

If we choose the coordinate system (t, x^i) so that $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$, i.e. to be orthonormal, we can then write:

$$\mathrm{d}s^2 = (-1 + h_{00})\,\mathrm{d}t^2 + 2h_{0i}\mathrm{d}x^i\mathrm{d}t + \left(\delta_{ij} + h_{ij}\right)\mathrm{d}x^i\mathrm{d}x^j \ . \tag{5.33}$$

For clarity and convenience, since the problem is invariant under under spacelike rotations in the hypersurface spanned by $\left\{\boldsymbol{e}_{(i)}\right\}_{i\in\{1,2,3\}}$, the perturbations can be decomposed into scalars, vectors and tensors[2].

- First, we have a scalar under rotations Φ:

$$h_{00} = -2\Phi \ . \tag{5.34}$$

- Then, we have:

$$h_{0i} = \partial_i w + \hat{w}_i \ , \tag{5.35}$$

  where w is a scalar and $\hat{w}^i$ is a divergence-free (also known as transverse) vector:

$$\partial_i \hat{w}^i = 0 \ . \tag{5.36}$$

  This is the usual decomposition of a 3-vector into a gradient and a divergence-free vector (Helmholtz theorem).

[2] Technically, these are the irreducible representations of the action of the rotation group $SO(3)$ on the span of $\left\{\boldsymbol{e}_{(i)}\right\}_{i\in\{1,2,3\}}$.

- Finally:

$$h_{ij} = -2\Psi\delta_{ij} + 2s_{ij} \ , \tag{5.37}$$

where $\mathrm{Tr}[h_{ij}] = -6\Psi$ with Ψ a scalar and s_{ij} is a traceless tensor which is further decomposed into:

$$s_{ij} = D_{ij}E + \partial_{(i}\hat{E}_{j)} + \hat{E}_{ij} \ , \tag{5.38}$$

with E a scalar, $\hat{E}^j$ a divergence-less vector with $\partial_i\hat{E}^i = 0$ and E_{ij} a divergence-less and tracefree tensor: $\partial_i\hat{E}^i{}_j = 0$ and $\hat{E}^i{}_i = 0$. We have also defined the traceless differential operator:

$$D_{ij}E = \partial_i\partial_j E - \frac{1}{3}\Delta E\delta_{ij} \ . \tag{5.39}$$

If we count the degrees of freedom, we thus have:

- 4 scalars which are 4 functional degrees of freedom: Φ, Ψ, w, E;
- 2 transverse vectors which are 4 functional degrees of freedom (3 components with a constraint each): $\hat{w}^i$ and $\hat{E}^i$;
- 1 symmetric transverse trace-free tensor which is 2 degrees of freedom (9 components with 3 (symmetry)+3 (transverse)+1 (trace-free)=7 constraints): $\hat{E}_{ij}$.

This leaves us with 10 functional degrees of freedom. The advantage of this decomposition is that, at linear order in the perturbations, the Einstein field equations separate into scalar, (transverse) vector and (transverse and trace-free) tensor parts that are satisfied separately so that these are genuinely the true, independent degrees of freedom. Since we have 10 Einstein field equations, it may look like we have a perfectly well-defined system to determine the metric degrees of freedom. Of course, the Einstein field equations are actually separated into evolution and constraints because the invariance of physics under the choice of coordinates, which, in perturbation theory reduces to the gauge freedom we explored in the previous section, allows us to fix 4 functional degrees of freedom, leaving us with only 6 truly independent functions among the 10 we identified previously. Let us see how it works by working out how the scalar, vector and tensor degrees of freedom transform under a gauge transformation[3]. Let us choose a vector field $\boldsymbol{\xi}$ so that. We decompose it, like our

[3]Note that we restrict ourselves to gauge transformations because we want to stay "close to Minkowski". A general, non infinitesimal transformation may lead to metric components that are large, thus breaking the perturbative approach

perturbations, into two scalars ξ^0 and ξ and a transverse vector $\hat{\xi}$:

$$\boldsymbol{\xi} = \xi^0 \boldsymbol{e}_{(\mathbf{0})} + \left(\partial^i \hat{\xi} + \hat{\xi}^i\right) \boldsymbol{e}_{(\boldsymbol{i})} \text{ with } \partial_i \hat{\xi}^i = 0 \ . \tag{5.40}$$

Then, using Eq. (5.26), we get:

$$\begin{cases} \Phi' = \Phi + \varepsilon \partial_0 \xi^0 & (5.41) \\ w' = w + \varepsilon \partial_0 \hat{\xi} - \varepsilon \xi^0 & (5.42) \\ \hat{w}_i' = \hat{w}_i + \varepsilon \partial_0 \hat{\xi}_i & (5.43) \\ \Psi' = \Psi - \dfrac{\varepsilon}{3} \Delta \hat{\xi} & (5.44) \\ E' = E + \varepsilon \hat{\xi} & (5.45) \\ \hat{E}_i' = \hat{E}_i + \varepsilon \hat{\xi}_i & (5.46) \\ \hat{E}_{ij}' = \hat{E}_{ij} \ . & (5.47) \end{cases}$$

We see that the tensor degree of freedom is gauge invariant while everything else is affected in a general gauge transformation.

Einstein field equations for perturbations

The Einstein field equations:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{5.48}$$

can be written for a generic energy momentum tensor:

$$T_{\mu\nu} = (\rho + P)\, u_\mu u_\nu + P g_{\mu\nu} + 2 q_{(\mu} u_{\nu)} + \Pi_{\mu\nu} \ , \tag{5.49}$$

where u^μ is the 4-velocity of the matter fluid, ρ and P its energy density and pressure respectively; $q_\mu = q_i \delta^i_\mu$, a 3-vector is its heat flux and $\Pi_{\mu\nu} = \Pi_{ij} \delta^i_\mu \delta^j_\nu$ its anisotropic stress, which is a traceless 3-tensor. However, since we are interested in a spacetime that can be written as a perturbation of the vacuum Minkowski spacetime, the sources of the field ought to be weak as well. This means that:

$$\begin{cases} \rho,\ P = O(\Phi) & (5.50) \\ u^\mu = \delta^\mu_0 + v^\mu \text{ with } v^\mu = O(\Phi) & (5.51) \\ q_i = \partial_i q + \hat{q}_i \text{ with } \partial_i \hat{q}^i = 0 \text{ and } q,\ \hat{q}^i = O(\Phi) & (5.52) \\ \Pi_{ij} = D_{ij}\Pi + \partial_{(i}\hat{\Pi}_{j)} + \hat{\Pi}_{ij} \text{ with } \Pi,\ \hat{\Pi}^i,\ \hat{\Pi}_{ij} = O(\Phi) \ . & (5.53) \end{cases}$$

One could write the Einstein field equations in an arbitrary gauge, keeping all 10 metric degrees of freedom but it is quite cumbersome. It is better to first define a specific gauge and then write the field equations in that particular gauge.

Some gauges

To illustrate how to pick up a gauge, let us give two examples here. We start with a description of a spacetime "close to Minkowski" in an arbitrary gauge defined through Eqs. (5.34)-(5.37). Then, we construct the vector field $\boldsymbol{\xi}$ necessary to get to the get we wish to define.
Our first example is the *synchronous gauge*. It is defined as the gauge in which observers comoving with the coordinate system ($u^i = 0 \;\Leftrightarrow\; v^\mu = 0$) have proper timer t. This implies that, in that gauge $\Phi_{\rm sync} = 0$. This can clearly be achieved by choosing ξ^0 such that:

$$\varepsilon \partial_0 \xi^0 = -\Phi \; . \tag{5.54}$$

Clearly, this is not enough to fix a gauge completely since we still have 3 degrees of freedom to fix. We do this by requiring:

$$w^i_{\rm sync} = 0 \; , \tag{5.55}$$

i.e.:

$$w_{\rm sync} = 0 \;\; \text{and} \;\; \hat{w}^i_{\rm sync} = 0 \; . \tag{5.56}$$

this can be done by imposing:

$$\varepsilon \left[\partial_0 \hat{\xi} - \xi^0 \right] = -w \;\; \text{and} \;\; \varepsilon \partial_0 \hat{\xi}_i = -\hat{w}_i \; . \tag{5.57}$$

In the end, the line element reads:

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \left((1 - 2\Psi_{\rm sync})\delta_{ij} + 2D_{ij}E_{\rm sync} + 2\partial_{(i}\hat{E}^{\rm sync}_{j)} + 2\hat{E}_{ij} \right) \mathrm{d}x^i \mathrm{d}x^j \; . \tag{5.58}$$

Another useful gauge is known as the *transverse gauge* or *longitudinal gauge*. It is the gauge in which the field equations look the closest to the Newtonian equations (see below). It is obtained by imposing:

$$\partial_i s^{ij}_{\rm trans} = 0 \;\; \text{and} \; \partial_i w^i_{\rm sync} = 0 \; , \tag{5.59}$$

which completely fixes the gauge. In details, we have:

$$E_{\rm trans} = 0 \; , \;\; \hat{E}^j_{\rm trans} = 0 \;\; \text{and} \;\; w_{\rm trans} = 0 \; . \tag{5.60}$$

This is achieved by specifying the gauge transformation:

$$\varepsilon\hat{\xi} = -E \tag{5.61}$$

$$\varepsilon\hat{\xi}^i = -\hat{E}^i \tag{5.62}$$

$$\varepsilon\left[\partial_0\hat{\xi} - \xi^0\right] = -w \ . \tag{5.63}$$

The line element then becomes:

$$\mathrm{d}s^2 = -\left(1 + 2\Phi_{\mathrm{trans}}\right)\mathrm{d}t^2 + 2\hat{w}_i^{\mathrm{trans}}\mathrm{d}x^i\mathrm{d}t + \left[\left(1 - 2\Psi_{\mathrm{trans}}\right)\delta_{ij} + 2\hat{E}_{ij}\right]\mathrm{d}x^i\mathrm{d}x^j \ . \tag{5.64}$$

5.2.3 Quasi-Newtonian limit

We can now prove that the line element in a weak, slowly varying gravitational field is given, as we claimed by:

$$\mathrm{d}s^2 = -\left(1 + 2\Phi_N\right)\mathrm{d}t^2 + \left(1 - 2\Phi_N\right)\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j \ , \tag{5.65}$$

where Φ_N is the Newtonian gravitational potential.
We work in the transverse (longitudinal) gauge in which the metric is given by Eq. (5.64). We drop the label trans and use:

$$\mathrm{d}s^2 = -\left(1 + 2\Phi\right)\mathrm{d}t^2 + 2\hat{w}_i\mathrm{d}x^i\mathrm{d}t + \left[\left(1 - 2\Psi\right)\delta_{ij} + 2\hat{E}_{ij}\right]\mathrm{d}x^i\mathrm{d}x^j \ . \tag{5.66}$$

Writing the Einstein field equations in absence of the cosmological constant we get, at first order and separating the degrees of freedom:

$$\begin{cases} \Delta\Psi = 4\pi G\rho & (5.67) \\ \Delta\hat{w}_i = -16\pi G\hat{q}_i & (5.68) \\ \partial_i\left[\partial_0\Psi - 4\pi G q\right] = 0 & (5.69) \\ \left(\delta_{ij}\Delta - \partial_i\partial_j\right)\left(\Phi - \Psi\right) + 2\partial_0^2\Psi\delta_{ij} = 8\pi G\left[p\delta_{ij} + D_iD_j\Pi\right] & (5.70) \\ \partial_0\partial_{(i}\hat{w}_{j)} = 8\pi G\partial_{(i}\hat{\Pi}_{j)} & (5.71) \\ \Box\hat{E}_{ij} = -8\pi G\hat{\Pi}_{ij} \ . & (5.72) \end{cases}$$

In the Newtonian limit, we assume that the fields are varying much more slowly in time than in space: $c|\partial_0 f| \ll |\partial_i f|$, and the source is non relativistic, so that: $P \simeq 0$, $q^i \simeq 0$, $\Pi_{ij} \simeq 0$. The

equations then become simply:

$$\begin{cases} \Delta\Psi = 4\pi G\rho & (5.73) \\ \left(\delta_{ij}\Delta - \partial_i\partial_j\right)(\Phi - \Psi) = 0 & (5.74) \\ \Delta\hat{w}_i = 0 & (5.75) \\ \Delta\hat{E}_{ij} = 0 & (5.76) \end{cases}$$

If we impose that the solutions are regular at infinity and do not diverge, the only solutions to Eqs. (5.75)-(5.76) are:

$$\hat{w}_i = 0 \tag{5.77}$$

$$\hat{E}_{ij} = 0\ . \tag{5.78}$$

Besides, Eq.(5.75) implies that $\Phi - \Psi$ is a pure function of time which can always be reabsorbed by a redefinition of the time coordinate, so that:

$$\Phi = \Psi\ . \tag{5.79}$$

Since Eq. (5.73) is simply the Poisson equation with solution the Newtonian potential, the result follows.

5.3 Gravitational waves: the plane wave solution

We have seen, at the end of the previous section, that the weak field limit of General Relativity for a slowly moving, non-relativistic matter source was fully characterised by the Newtonian gravitational potential via the perturbed metric (5.65). In the rest of this chapter, we want to study the somewhat opposite end of the spectrum among weak field solutions of Einstein field equations: the freely propagating waves. We could start from the analysis in terms of degrees of freedom presented in the previous section, but we will follow a more classical approach when dealing with gravitational waves and start anew.

5.3.1 The field equations for freely propagating gravitational radiation

We seek a solution of the Einstein field equations in absence of the cosmological constant far from the source, i.e. in vacuum:

$$R_{\mu\nu} = 0\ , \tag{5.80}$$

when the geometry is assumed "close" to Minkowski:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \left(\eta_{\mu\nu} + h_{\mu\nu}\right)\mathrm{d}x^\mu \mathrm{d}x^\nu \ , \tag{5.81}$$

for $\left|h_{\mu\nu}\right| \ll 1$ in some appropriate gauge. To simplify calculations, we restrict our analysis to orthonormal coordinates, such that $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$.
First, let us notice that, by expanding $g_{\mu\rho}g^{\rho\nu} = \delta_\rho{}^\nu$, we have:

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} \ , \tag{5.82}$$

where:

$$h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma} \ . \tag{5.83}$$

Since $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$, we have that $\forall(\nu,\alpha,\beta)$, $\partial_\nu g_{\alpha\beta} = \partial_\nu h_{\alpha\beta}$, so that, at first order in the perturbation h:

$$\Gamma^\mu_{\nu\rho} = \frac{1}{2}\eta^{\mu\sigma}\left[\partial_\nu h_{\sigma\rho} + \partial_\rho h_{\nu\sigma} - \partial_\sigma h_{\nu\rho}\right] + O(h^2) \ . \tag{5.84}$$

The Riemann tensor at first order is then given by:

$$R^\mu{}_{\nu\rho\sigma} = \partial_\rho\Gamma^\mu{}_{\nu\sigma} - \partial_\sigma\Gamma^\mu{}_{\nu\rho} + \underbrace{\Gamma^\mu{}_{\varepsilon\rho}\Gamma^\varepsilon{}_{\nu\sigma} - \Gamma^\mu{}_{\varepsilon\sigma}\Gamma^\varepsilon{}_{\nu\rho}}_{=O\left(h^2\right)} \tag{5.85}$$

$$= \partial_\rho\Gamma^\mu{}_{\nu\sigma} - \partial_\sigma\Gamma^\mu{}_{\nu\rho} + O\left(h^2\right) \tag{5.86}$$

$$= \frac{1}{2}\eta^{\mu\lambda}\partial_\rho\left[\partial_\nu h_{\lambda\sigma} + \partial_\sigma h_{\nu\lambda} - \partial_\lambda h_{\nu\sigma}\right] - \frac{1}{2}\eta^{\mu\lambda}\partial_\sigma\left[\partial_\nu h_{\lambda\rho} + \partial_\rho h_{\nu\lambda} - \partial_\lambda h_{\nu\rho}\right] + O\left(h^2\right) \tag{5.87}$$

$$= \frac{1}{2}\eta^{\mu\lambda}\left[\partial_\rho\partial_\nu h_{\lambda\sigma} - \partial_\sigma\partial_\nu h_{\lambda\rho} - \partial_\rho\partial_\lambda h_{\nu\sigma} + \partial_\sigma\partial_\lambda h_{\nu\rho}\right] + O\left(h^2\right) \ . \tag{5.88}$$

Therefore, the Ricci tensor is:

$$R_{\nu\sigma} = R^\mu{}_{\nu\mu\sigma} \tag{5.89}$$

$$= \frac{1}{2}\left[\partial_\mu\partial_\nu h^\mu{}_\sigma - \partial_\sigma\partial_\nu h^\mu{}_\mu - \Box h_{\nu\sigma} + \partial_\sigma\partial_\mu h_\nu{}^\mu\right] + O\left(h^2\right) \ , \tag{5.90}$$

where, as usual, $\Box\cdot = \eta^{\mu\nu}\partial_\mu\partial_\nu\cdot = \partial^\lambda\partial_\lambda\cdot$. Dropping the $O\left(h^2\right)$ from now on and working consistently at linear order, the Einstein Field equations (5.80) become:

$$\frac{1}{2}\left[\partial_\lambda\partial_\mu h^\lambda{}_\nu + \partial_\nu\partial_\lambda h^\lambda{}_\mu - \Box h_{\mu\nu} - \partial_\nu\partial_\mu h_\lambda{}^\lambda\right] = 0 \ . \tag{5.91}$$

Rearranging the terms, we get:

$$\Box h_{\mu\nu} - \partial_\mu \left[\partial_\lambda h^\lambda{}_\nu - \frac{1}{2}\partial_\nu h^\lambda{}_\lambda\right] - \partial_\nu \left[\partial_\lambda h^\lambda{}_\mu - \frac{1}{2}\partial_\mu h^\lambda{}_\lambda\right] = 0 \,. \tag{5.92}$$

Therefore, we have:

Gravitational wave equations

$$\Box h_{\mu\nu} - 2\partial_{(\mu} V_{\nu)} = 0 \,, \tag{5.93}$$

with:

$$V_\alpha = \partial_\lambda h^\lambda{}_\alpha - \frac{1}{2}\partial_\alpha h^\lambda{}_\lambda \,. \tag{5.94}$$

So far, we have worked in an arbitrary gauge, so that Eq. (5.93) contains non-physical degrees of freedom. The one-form $\boldsymbol{V}$ defined by Eq. (5.94) has exactly 4 degrees of freedom, so we can fix a gauge by requiring:

$$V_\alpha = \partial_\lambda h^\lambda{}_\alpha - \frac{1}{2}\partial_\alpha h^\lambda{}_\lambda = 0 \,. \tag{5.95}$$

If we denote by $\tilde{h}$ the original perturbation, this can be achieved by choosing a vector $\boldsymbol{\xi}$ such that:

$$\partial_\lambda h^\lambda{}_\alpha - \frac{1}{2}\partial_\alpha h^\lambda{}_\lambda = 0 \tag{5.96}$$

$$\partial_\lambda \left[\tilde{h}^\lambda{}_\alpha + \partial^\lambda \xi_\alpha + \partial_\alpha \xi^\lambda\right] - \frac{1}{2}\partial_\alpha \left[\tilde{h}^\lambda{}_\lambda + 2\partial_\lambda \xi^\lambda\right] = 0 \tag{5.97}$$

$$\partial_\lambda \tilde{h}^\lambda{}_\alpha + \Box \xi_\alpha + \partial_\lambda \partial_\alpha \xi^\lambda - \frac{1}{2}\partial_\alpha \tilde{h}^\lambda{}_\lambda - \partial_\alpha \partial_\lambda \xi^\lambda = 0 \,. \tag{5.98}$$

Rearranging this we get:

$$\Box \xi_\alpha = \frac{1}{2}\partial_\alpha \tilde{h}^\lambda{}_\lambda - \partial_\lambda \tilde{h}^\lambda{}_\alpha = -\tilde{V}_\alpha \,. \tag{5.99}$$

In that new gauge, let us call it the *Lorenz gauge*, we have the equations:

Gravitational wave equations in Lorenz gauge

$$\Box h_{\mu\nu} = 0 \,, \tag{5.100}$$

with the constraints:

$$\partial_\mu h^\mu{}_\nu - \frac{1}{2}\partial_\nu h = 0 \,, \tag{5.101}$$

where we defined the trace $h = h^{\lambda}{}_{\lambda}$. To simplify the following discussion, let us define:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} h \eta_{\mu\nu} \,, \tag{5.102}$$

called the opposite trace perturbations because $\bar{h} = \bar{h}^{\mu}{}_{\mu} = h - 2h = -h$. In terms of this new variable, Eqs.(5.100)-(5.101) become:

$$\Box \bar{h}_{\mu\nu} = 0 \tag{5.103}$$

$$\partial_{\mu} \bar{h}^{\mu}{}_{\nu} = 0 \,. \tag{5.104}$$

In principle, we have fixed our gauge by selecting a gauge transformation vector $\boldsymbol{\xi}$ via Eq. (5.99). However, let us apply another gauge transformation, generated by a second vector $\hat{\xi}$. Under this transformation, the variable $\bar{h}_{\mu\nu}$ transforms as:

$$\bar{h}_{\mu\nu} \;\rightarrow\; \bar{h}_{\mu\nu} + 2\partial_{(\mu}\hat{\xi}_{\nu)} - \partial_{\alpha}\hat{\xi}^{\alpha}\eta_{\mu\nu} \,, \tag{5.105}$$

so that:

$$\partial^{\mu}\bar{h}_{\mu\nu} \;\rightarrow\; \partial^{\mu}\bar{h}_{\mu\nu} + \Box\hat{\xi}_{\nu} + \underbrace{\partial^{\mu}\partial_{\nu}\hat{\xi}_{\mu} - \partial_{\nu}\partial_{\mu}\hat{\xi}^{\mu}}_{=0} \,. \tag{5.106}$$

But this means that any vector field such that $\Box\hat{\xi}_{\nu} = 0$ preserves both the field equations $\Box\bar{h}_{\mu\nu} = 0$ and the constraints $\partial_{\mu}\bar{h}^{\mu}{}_{\nu} = 0$. Since it also preserves the original gauge change (5.99), that means that instead of $\boldsymbol{\xi}$ prescribed by solving this equation, we could have chosen $\boldsymbol{\xi} + \hat{\boldsymbol{\xi}}$ with $\Box\hat{\xi}_{\alpha} = 0$. Thus, we have 4 extra free degrees of freedom that we can get rid of by a gauge choice. We will get back to that later.

5.3.2 Plane wave solution

Let us look for a solution to Eqs (5.103)-(5.104) in the form of a plane wave:

$$\bar{h}_{\mu\nu} = A_{\mu\nu} \mathrm{e}^{\mathrm{i}k_{\mu}x^{\mu}} \,, \tag{5.107}$$

where $A_{\mu\nu}$ is a constant tensor and $\boldsymbol{k} = k^{\mu}\boldsymbol{e}_{(\mu)}$ is the wave vector. Plugging this in Eqs (5.103)-(5.104) we get:

$$\begin{cases} k_{\mu}k^{\mu} = 0 \;:\; \text{the wave propagates at the speed of light;} & (5.108) \\ k^{\mu}A_{\mu\nu} = 0 \;:\; \text{the wave is transverse.} & (5.109) \end{cases}$$

The fact that the wave is transverse does reduce the number of free parameters in the solution from 10 to 6 as expected. But as we know, there is an extra freedom in this solution, given by any harmonic vector field: $\Box\hat{\xi}^\mu = 0$. Certainly,

$$\hat{\xi}^\mu = B^\mu \mathrm{e}^{\mathrm{i}k_\mu x^\mu} \ , \tag{5.110}$$

with B^μ constants is an harmonic vector field. Let us remember that under a gauge transformation generated by $\hat{\boldsymbol{\xi}}$, $\bar{h}_\mu$ transforms as (see Eq. (5.105)):

$$\bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} + 2\partial_{(\mu}\hat{\xi}_{\nu)} - \partial_\alpha \hat{\xi}^\alpha \ . \tag{5.111}$$

Substituting the solution and the harmonic gauge vector (5.110) in this, we get:

$$A'_{\mu\nu} = A_{\mu\nu} + i\left[k_\mu B_\nu + k_\nu B_\mu - k_\alpha B^\alpha \eta_{\mu\nu}\right] \ . \tag{5.112}$$

Let us try to impose:

$$\eta^{\mu\nu} A'_{\mu\nu} = 0 \ \text{ i.e. that } \bar{h}'_{\mu\nu} \ \text{ is traceless} \tag{5.113}$$

$$A'_{0i} = 0 \ . \tag{5.114}$$

This leads to a system of 4 equations for the 4 unknowns B^μ:

$$\begin{cases} k_\alpha B^\alpha = -\dfrac{i}{2} A^\mu_\mu & (5.115) \\ -k_j B^0 + k_0 B_j = i A_{0i} \ . & (5.116) \end{cases}$$

The determinant of the system is $k_0^2\left(k_0^2 + k_1^2 + k_2^2 + k_3^2\right) = 2k_0^4 \neq 0$ so there is always a unique solution for the vector components B^μ and thus for the vector $\hat{\boldsymbol{\xi}}$. This shows that we are allowed to suppose (suppressing the prime for ease of notations):

$$\begin{cases} \bar{h}_{0i} = 0 \ \Rightarrow \ A_{0i} = 0 & (5.117) \\ \bar{h}^\mu{}_\mu = 0 \ \Rightarrow A^\mu{}_\mu = 0 \ . & (5.118) \end{cases}$$

This is known as the *radiation gauge* or the *TT gauge* for transverse traceless. Indeed $\bar{h}^\mu_\mu = 0$ implies $h^\mu_\mu = 0$ and thus $\bar{h}_{\mu\nu} = h_{\mu\nu}$, so that the Lorenz gauge condition can be read:

$$\partial_\mu h^\mu{}_\nu = 0 \ . \tag{5.119}$$

In particular, for $\nu = 0$, we get that:

$$\partial_\mu h^\mu{}_0 = \partial_i \underbrace{h^i{}_0}_{=0} + \partial_0 h^0{}_0 = 0 \,, \tag{5.120}$$

But this reads:

$$k_0 A^0{}_0 = 0 \,, \tag{5.121}$$

and since $k_0 \neq 0$, we have $A_{00} = 0$, i.e. $h_{00} = 0$. To sum up, the plane wave solution in the TT gauge reads:

Gravitational wave solution in the TT gauge

$$h_{\mu\nu} = A_{\mu\nu} \mathrm{e}^{i k_\alpha x^\alpha} \,, \tag{5.122}$$

with:

$$k_\mu k^\mu = 0 \,, \quad k^\mu A_{\mu\nu} = 0 \,, \tag{5.123}$$

$$A_{0\mu} = 0 \,, \quad A^\mu{}_\mu = 0 \,. \tag{5.124}$$

This is a total of 9 constraints on $A_{\mu\nu}$, but they are not all independent since making $\nu = 0$ in $k^\mu A_{\mu\nu} = 0$ results in an equation that is true by virtue of $A_{0\mu} = 0$ and is thus not independent. Therefore, we have 8 constraints on 10 degrees of freedom, so we are left with 2 independent degrees of freedom. This is a very important result:

Gravitational waves degrees of freedom

A freely propagating plane gravitational wave is fully determined by two independent degrees of freedom, called its polarisation states.

To try and be more specific, let us pick up a direction of propagation, say along $\boldsymbol{e}_{(3)}$, then: $k^1 = k^2 = 0$ and $k^0 = \pm k^3 = -\omega$. Let us restrict our attention to $k^3 = -k^0 = \omega$, so that $k_\mu x^\mu = \omega(t - z)$ where we have also relabelled $x^1 = x$, $x^2 = y$ and $x^3 = z$. The Lorenz gauge condition reads:

$$0 = \omega \left(-A_{0\nu} + A_{3\nu}\right) = \omega A_{3\nu} \,, \tag{5.125}$$

so we have that $A_{30} = A_{31} = A_{32} = A_{33} = 0$ and by symmetry $A_{23} = A_{13} = 0$. The only non-zero components are thus A_{11}, A_{22} and $A_{12} = A_{21}$. Besides, since $A^{\mu}{}_{\mu} = 0$, we must have $A_{22} = -A_{11}$. Denoting $A_{11} = h_+$ and $A_{12} = h_\times$, we arrive at the plane wave solution propagating along the z-axis:

$$h^{\mu}{}_{\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \mathrm{e}^{\mathrm{i}\omega(t-z/c)} + \text{c.c.} \,. \tag{5.126}$$

h_+ and $h_\times$ are the *two canonical polarisation states* of the freely propagating plane gravitational wave.

5.4 Physical effects of gravitational waves

Let us turn to the problem of the effects gravitational waves such as the one described by Eq. (5.126) have on physical observables.

5.4.1 Effects of a gravitational wave on matter

Consider some massive test particle with worldline $x^\mu(\tau)$ in the TT gauge. We denote its 4-velocity by:

$$u^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau} = \bar{u}^\mu + \delta u^\mu \,, \tag{5.127}$$

where $\bar{u}^\mu$ is its 4-velocity in the Minkowski background, in absence of gravitational wave, and δu^μ the perturbation induced by the wave. Let us assume that the particle is at rest in Minkowski: $\bar{u}^0 = 1$ and $\bar{u}^i = 0$. Then expanding $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = -1$ at first order, we get $\delta u^0 = 0$. Moreover, the geodesic equation gives:

$$\frac{\mathrm{d}u^i}{\mathrm{d}t} + \Gamma^i{}_{00} = 0 \,, \tag{5.128}$$

and since $\Gamma^i_{00} = 0$, we get: $u^i = 0$. In other words, a particle initially at rest (before the wave passes through) stays at rest. Is there a problem? Does it mean that gravitational waves do not have any observational effects? Certainly not. It simply means that the coordinates of the TT gauge are comoving by construction: free-falling particles remain at constant values of the TT coordinates.

But coordinate systems do not encode any physics. What matters, when we talk about gravitational effects is curvature and tidal forces, i.e. the geodesic deviation equation. However, the geodesic deviation equation written in the TT gauge is not of much help.

Indeed, let us consider an infinitesimal ring of free-falling, massive test particles centred on a co-moving observer in the TT gauge. What we have is a bundle of matter around a timelike geodesics with 4-velocity $\boldsymbol{u} = \boldsymbol{e}_{(\mathbf{0})}$ and proper time $\tau = t$. The deviation vector $\boldsymbol{\xi}$, connecting the reference geodesics to the one of the nearby test particles and that describes how these geodesics move with respect to each other, tells us how the small ring of matter is deformed in the field of the wave; it obeys the geodesic deviation equation:

$$\frac{\mathrm{D}^2\xi^\mu}{\mathrm{D}\tau^2} = R^\mu{}_{\nu\rho\sigma}u^\mu u^\rho \xi^\sigma \ . \tag{5.129}$$

Note that the LHS of this equation must be understood as a double covariant derivative of the deviation vector components *along the central geodesics*, so that, using that $u^\nu = \delta^\nu{}_0$ and $\Gamma^\mu{}_{0\rho} = \frac{1}{2}\partial_t h^\mu{}_\rho$, and developing at first order in $\boldsymbol{h}$:

$$\begin{aligned}
\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}\tau^2} =& u^\nu\nabla_\nu\left[u^\rho\nabla_\rho\xi^\mu\right] && (5.130)\\
=& u^\nu\nabla_\nu\left[\frac{\mathrm{d}\xi^\mu}{\mathrm{d}t} + \frac{1}{2}\xi^\rho\partial_t h^\mu{}_\rho\right] && (5.131)\\
=& \frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}t^2} + \frac{1}{2}\xi^\rho\partial_t^2 h^\mu{}_\rho + \partial_t h^\mu{}_\rho\frac{\mathrm{d}\xi^\rho}{\mathrm{d}t} \ . && (5.132)
\end{aligned}$$

Thus, we have:

$$\frac{\mathrm{d}^2\xi^\mu}{\mathrm{d}t^2} + \frac{1}{2}\xi^\rho\partial_t^2 h^\mu{}_\rho + \partial_t h^\mu{}_\rho\frac{\mathrm{d}\xi^\rho}{\mathrm{d}t} = R^\mu{}_{00\sigma}\xi^\sigma \ , \tag{5.133}$$

with, from Eq. (5.88):

$$R^\mu{}_{00\sigma} = \frac{1}{2}\eta^{\mu\lambda}\partial_0^2 h_{\lambda\sigma} \ . \tag{5.134}$$

For the spatial displacements, we thus get:

$$\frac{\mathrm{d}^2\xi^i}{\mathrm{d}t^2} = -\partial_t h^i{}_j\frac{\mathrm{d}\xi^j}{\mathrm{d}t} \ . \tag{5.135}$$

Therefore, if the particles are initially at rest in the TT gauge, which they must be for them to be on geodesics, as we have seen, then, the only solution is for the spatial deviation ξ^i to remain constant: in the TT gauge, the physical effects of gravitational waves are locally fully re-absorbed in a change of coordinates. On the other hand, in General Relativity, when we want to talk about distances

between objects, we need to do it *with respect to a specific frame*, for example, one in which some rulers are kept fixed. Such rulers cannot be rigidly free-falling in the laboratory if the gravitational field varies on the scales of the experiment. Thus, let us assume that the local free-falling observer O at the centre of the ring of particles uses their local inertial frame constructed as Fermi normal coordinates as in subsection 3.7.4, $\left\{\hat{\boldsymbol{e}}_{(\mu)}\right\}$ such that $\hat{\boldsymbol{e}}_{(\mathbf{0})} = \boldsymbol{u}$ and $\hat{\boldsymbol{e}}_{(\boldsymbol{i})}$ are three orthonormal vectors spanning the local rest frame of O. We denote by $\left\{\hat{t}, \hat{x}^i\right\}$ local coordinates in that frame, such that $\hat{x}^i = 0$ at $O(\tau)$. Then, at O, the metric takes its Minkowski form:

$$\mathrm{d}s^2_{|O} = -\mathrm{d}\hat{t}^2 + \eta_{ij}\mathrm{d}\hat{x}^i\mathrm{d}\hat{x}^j \ , \tag{5.136}$$

with $\partial_\mu \hat{g}_{\rho\sigma}\left(t, \vec{0}\right) = 0$: this is the equivalence principle. The linear coordinate transformation is actually easy to find and we get:

$$\hat{x}^0 = x^0 = t \tag{5.137}$$

$$\hat{x}^i = x^i + \frac{1}{2} h^i{}_j\left(t, \vec{0}\right) x^j \ , \tag{5.138}$$

at leading order in x^i. The metric in the neighbourhood of O then differs from the Minkowski metric by terms of order $\hat{x}^\mu\hat{x}^\nu$ at most. These are examples of *local Fermi coordinates*; see subsection 3.7.4. Free-falling masses following timelike geodesics will not remain at rest in these coordinates.
Let us imagine that the wave propagates in the $x^3 = z$ direction, like in Eq. (5.126) and that the ring of matter is in the plane $z = 0$. Then, the deformation of this ring in the local Fermi coordinates is determined by the timelike geodesic deviation equation written in these coordinates:

$$\frac{\mathrm{D}^2\hat{\xi}^\mu}{\mathrm{D}\tau^2} = \hat{R}^\mu{}_{00\nu}\hat{\xi}^\nu \ , \tag{5.139}$$

where we used that on the central worldline of O, $\hat{u}^\mu = \delta^\mu_0$. However, note that, at first order in $\boldsymbol{h}$, the Riemann tensor is invariant by a change of coordinates (it is gauge-invariant), so that:

$$\hat{R}^\mu{}_{\nu\rho\sigma} = R^\mu{}_{\nu\rho\sigma} \ , \tag{5.140}$$

where the RHS are the components of the Riemann tensor in the TT gauge. Moreover, in the Fermi coordinates, the connection coefficients are zero along the central worldline, by construction. Therefore, we can write:

$$\frac{\mathrm{D}^2\hat{\xi}^\mu}{\mathrm{D}\tau^2} = \frac{\mathrm{d}^2\hat{\xi}^\mu}{\mathrm{d}\tau^2} \ . \tag{5.141}$$

Finally, we obtain the evolution of the spatial displacements between the masses and the centre in local Fermi coordinates:

$$\frac{\mathrm{d}^2\hat{\xi}^i}{\mathrm{d}\tau^2} = R^i{}_{00j}\hat{\xi}^j \ , \tag{5.142}$$

or, equivalently:

$$\frac{\mathrm{d}^2\hat{\xi}^i}{\mathrm{d}t^2} = \frac{1}{2}\partial_t^2 h^i{}_j(t,0)\hat{\xi}^j \ . \tag{5.143}$$

Let us solve Eq. (5.143) for each polarisation state separately[4]. We will take or wave to be:

$$h^\mu{}_\nu = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \sin\left(\omega(t-z)\right) \ , \tag{5.144}$$

so that:

$$\partial_t^2 h^i{}_j(t,0) = -\omega^2 H \sin(\omega t) \ , \tag{5.145}$$

for $H = \pm h_+$ or $H = h_\times$ appropriately. First, let us assume that $h_\times = 0$. Then, we get:

$$\begin{cases} \dfrac{\mathrm{d}^2\hat{\xi}^1}{\mathrm{d}t^2} = -\dfrac{\omega^2}{2} h_+ \sin(\omega t)\,\hat{\xi}^1 & (5.146) \\ \dfrac{\mathrm{d}^2\hat{\xi}^2}{\mathrm{d}t^2} = \dfrac{\omega^2}{2} h_+ \sin(\omega t)\,\hat{\xi}^2 \ . & (5.147) \end{cases}$$

This can be solved at leading order in the perturbation h_+. Let us write $\hat{\xi}^i(t) = \hat{\xi}^i(0) + \delta\hat{\xi}^i$ since clearly the solution for $h_+ = 0$ is constant. Then, plugging this into the equations and expanding at first order, we get:

$$\begin{cases} \dfrac{\mathrm{d}^2\delta\hat{\xi}^1}{\mathrm{d}t^2} = -\dfrac{\omega^2}{2} h_+ \sin(\omega t)\,\hat{\xi}^1(0) & (5.148) \\ \dfrac{\mathrm{d}^2\delta\hat{\xi}^2}{\mathrm{d}t^2} = \dfrac{\omega^2}{2} h_+ \sin(\omega t)\,\hat{\xi}^2(0) \ . & (5.149) \end{cases}$$

These are readily solved to give:

[4]Note that the solution to the geodesic deviation equation in the local Fermi frame could also have been obtained by a change of coordinates from the TT gauge, in which the components ξ^i are constants. However, the integration directly in the Fermi frame has some pedagogical values.

Deformation of a small ring of matter in the + polarisation state

$$\begin{cases} \hat{\xi}^1(t) = \left(1 + \dfrac{1}{2} h_+ \sin(\omega t)\right) \hat{\xi}^1(0) & (5.150) \\ \hat{\xi}^2(t) = \left(1 - \dfrac{1}{2} h_+ \sin(\omega t)\right) \hat{\xi}^2(0) \ . & (5.151) \end{cases}$$

A similar approach gives the solution for the other polarisation state:

Deformation of a small ring of matter in the × polarisation state

$$\begin{cases} \hat{\xi}^1(t) = \hat{\xi}^1(0) + \dfrac{1}{2} h_\times \sin(\omega t)\, \hat{\xi}^2(0) & (5.152) \\ \hat{\xi}^2(t) = \hat{\xi}^2(0) + \dfrac{1}{2} h_\times \sin(\omega t)\, \hat{\xi}^1(0) \ . & (5.153) \end{cases}$$

These deformations are illustrated on Fig. 5.2 if the particles are initially distributed on a circle. It

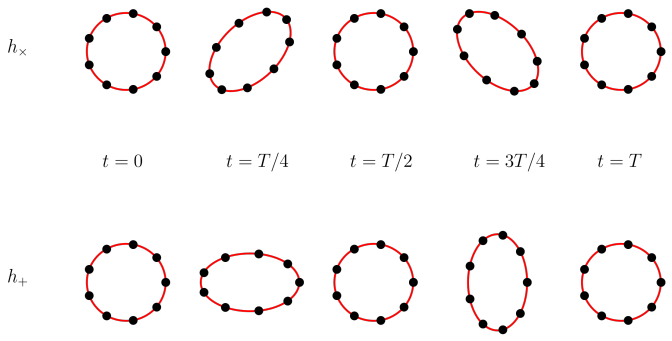


Figure 5.2: A small ring (red line) of test particles (black dots) in the $z = 0$ plane is deformed by the passing of a gravitational wave propagating along the z-axis with period $T = 2\pi/\omega$. Each polarisation mode is depicted separately.

is easy to prove that the ring deforms into ellipses. Indeed, let us assume that $\hat{\xi}^1(0) = R\cos\theta$ and $\hat{\xi}^2(0) = R\sin\theta$. Then, for the + polarisation, we clearly have that:

$$\frac{\left(\hat{\xi}^1(t)\right)^2}{R^2\left[1 + \frac{1}{2}h_+ \sin\omega t\right]^2} + \frac{\left(\hat{\xi}^2(t)\right)^2}{R^2\left[1 - \frac{1}{2}h_+ \sin\omega t\right]^2} = 0 \ , \qquad (5.154)$$

which is the equation of an ellipse with semi-minor and semi-major axes along the coordinate axes and alternating over time. The semi-major axis is along the $\hat{x}^1$-axis when $\sin\omega t = 1$ is maximum, i.e. at $t = (2k+1)\pi/2\omega = (2k+1)T/4$ for $k \in \mathbb{Z}$, at which times, the semi-minor axis is along the $\hat{x}^2$-axis. When $\sin\omega t = -1$, the situation is reversed and this happens at $t = 3(2k+1)\pi/2\omega = 3(2k+1)T/4$ for $k \in \mathbb{Z}$. In between, at $t = kT/2$ for $k \in \mathbb{Z}$, $\sin\omega t = 0$ and the ring is a circle. The same thing happens for the $\times$ polarisation but the axes are rotated by $\pi/4$.

5.4.2 Effects on the path of light

We see that gravitational waves do have an impact on the local distribution of matter but as usual in General Relativity, if we want to measure this effect, we need to design an *operational* way to measure distances (which is an ambiguous concept). Let us assume that we have two masses A and B that are free-falling in the local gravitational field determined by the plane gravitational wave of Eq. (5.126). An observer $\mathcal{O}$ attached to A sends a light signal towards B at a proper time t_1 along its worldline (event $A\,(t_1)$). This light signal is received at $B\,(t')$ at a proper time t' along the worldline of B and reflected towards A, where it is again received at a proper time t_2 as measured by $\mathcal{O}$ (event $A\,(t_2)$); see Fig. 5.3. We can then *define* the distance travelled by the light signal between the masses A and B using Einstein's simultaneity and converting the time of flight into a distance using the speed of light :

$$L = \frac{c}{2}\,(t_2 - t_1)\ . \tag{5.155}$$

This says that $B\,(t')$ is simultaneous (In Einstein's sense) to the event $A(t)$: $t = t'$ in TT gauge. Note that if A and B are infinitesimally close, then $L^2 = \boldsymbol{g}\,(\boldsymbol{A(t)B\,(t')}, \boldsymbol{A(t)B\,(t')})$, i.e. that L is indeed the distance between $A(t)$ and $B\,(t')$ as measured locally using the metric $\boldsymbol{g}$. Let us centre our TT coordinates on $\mathcal{O}$ so that $x^i_A = 0 = \text{cst}$ and $x^i_{B(t')} = \text{cst} \neq 0$. Thus, using $x^0_{A(t)} = x^0_{B(t')} = 0$, and denoting $B' = B\,(t')$ and $A = A(t)$ for simplicity, we can write:

$$\begin{aligned} L^2 =& g_{\mu\nu}\left(x^\mu_{B'} - x^\mu_A\right)\left(x^\nu_{B'} - x^\nu_A\right) && (5.156)\\ =& \left(\delta_{ij} + h_{ij}\right) x^i_{B'} x^j_{B'}\ . && (5.157)\end{aligned}$$

If we let $x^i_{B'} = L_0 n^i$, where $\boldsymbol{n}$ is the spatial vector connecting A and B' and $L_0 = \delta_{ij} x^i_{B'} x^j_{B'}$ is the distance travelled by the light signal in absence of gravitational wave, then, we get:

$$\frac{\delta L}{L_0} = \frac{1}{2} h_{ij} n^i n^j\ . \tag{5.158}$$

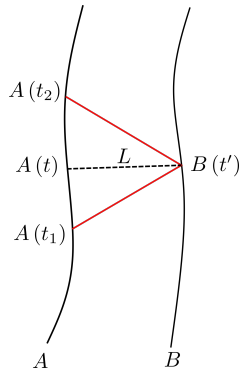


Figure 5.3: Using the Einsteinian definition of simultaneity to define a measurable notion of distance and to study the effects of gravitational waves, calculated in the TT gauge. A sends light towards B at t_1. This light is instantaneously reflected back towards A who receives it at t_2. It defines the time $t = (t_1 + t_2)/2$, which in the TT gauge is the time at which B reflected the light back. Then, $L = ct$.

This change in travel length is what the interferometers such as LIGO measure. It is called the *strain*. You can see one of the two LIGO detectors on Fig. 5.4: it is a giant interferometer with arms 4 km long. Note that what is really measured by the interferometer is a difference in (proper) time of arrival at the centre of the interferometer, not a length. This is in line with the fact that observers can only measure things along their worldline.

Figure 5.4: Photography of the Hanford detector of LIGO. Credit: LIGO Laboratory

5.5 Sources of gravitational waves: the quadrupole formula

5.5.1 General expression

Now, we are able to describe the propagation of a plane gravitational wave in vacuum, and we understand its physical effect on matter and light travel. How are these waves generated? What kind of sources can we study with them and how do we link properties of the sources with properties of the waves generated? We will consider only weakly-gravitating sources, so that our linear perturbation theory remains valid, even at the sources. In the previous section, we determined Eqs. (5.93)-(5.94) that govern the general propagation equations for gravitational waves in vacuum, before fixing any gauge, and we introduced the opposite trace perturbations:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} h \eta_{\mu\nu} \,, \tag{5.159}$$

with $h = h^{\alpha}{}_{\alpha}$. It turns out that, like in the vacuum case, this opposite trace perturbation is more suited to the study of gravitational waves in presence of some sources, which is what we will focus on in this section. However, we have to be careful with gauge fixing. Indeed, we have that the Ricci tensor, Eq. (5.90) reads:

$$R_{\mu\nu} = \frac{1}{2} \left[-\Box h_{\mu\nu} + 2\partial_{(\mu} V_{\nu)} \right] \,, \tag{5.160}$$

and thus:

$$R = R_{\mu\nu}\eta^{\mu\nu} = \frac{1}{2} \left[-\Box h + 2\partial_{\mu} V^{\mu} \right] \,. \tag{5.161}$$

Therefore:

$$\begin{aligned} G_{\mu\nu} =& R_{\mu\nu} - \frac{1}{2} R \eta_{\mu\nu} && (5.162)\\ =& \frac{1}{2} \left[-\Box h_{\mu\nu} + 2\partial_{(\mu} V_{\nu)} \right] - \frac{1}{4} \left[-\Box h + 2\partial_{\alpha} V^{\alpha} \right] && (5.163)\\ =& -\frac{1}{2} \Box \left[h_{\mu\nu} - \frac{1}{2} h \eta_{\mu\nu} \right] + \partial_{(\mu} V_{\nu)} - \frac{1}{2} \partial_{\alpha} V^{\alpha} \eta_{\mu\nu} && (5.164)\\ =& -\frac{1}{2} \Box \bar{h}_{\mu\nu} + \partial_{(\mu} V_{\nu)} - \frac{1}{2} \partial_{\alpha} V^{\alpha} \eta_{\mu\nu} \,. && (5.165) \end{aligned}$$

Let us also recall that:

$$V_{\mu} = \partial_{\alpha} \bar{h}^{\alpha}{}_{\mu} \,. \tag{5.166}$$

Therefore, the Einstein field equations read:

$$\Box\bar{h}_{\mu\nu} - 2\partial_{(\mu}V_{\nu)} + \partial_\alpha V^\alpha \eta_{\mu\nu} = -16\pi G T_{\mu\nu} \ , \tag{5.167}$$

with:

$$V_\mu = \partial_\alpha \bar{h}^\alpha{}_\mu \ . \tag{5.168}$$

Under a gauge transformation generated by a vector field $\boldsymbol{\xi}$, we have:

$$\bar{h}_{\mu\nu} \mapsto \ \bar{h}_{\mu\nu} + 2\partial_{(\mu}\xi_{\nu)} - \partial_\alpha \xi^\alpha \eta_{\mu\nu} \tag{5.169}$$

$$V_\mu \mapsto \ V_\mu + \Box\xi_\mu \ . \tag{5.170}$$

Hence, given a perturbative spacetime, by choosing $\boldsymbol{\xi}$ such that:

$$\Box\xi^\mu = -V^\mu = -\partial_\alpha \bar{h}^{\alpha\mu} \ , \tag{5.171}$$

we can set the *Lorenz gauge*:

$$\partial_\mu \bar{h}^\mu{}_\nu = 0 \ . \tag{5.172}$$

Then the *Einstein field equations in the Lorenz gauge* are:

$$\begin{cases} \Box\bar{h}_{\mu\nu} = -\,16\pi G T_{\mu\nu} & (5.173) \\ \partial_\mu \bar{h}^\mu{}_\nu = 0 \ . & (5.174) \end{cases}$$

So far, this is identical to what we did in vacuum. However, we cannot make the extra jump to the TT gauge. Indeed, taking the trace of Eq. (5.173), we get:

$$\Box h = 16\pi G T^\mu{}_\mu \ . \tag{5.175}$$

Generically, the source is not of zero trace and scalar modes of perturbations are sourced, so we cannot set $h = 0$ without neglecting some physics[5]: we cannot focus solely on the freely propagating degrees of freedom encoded in $\hat{E}_{ij}$; see subsection 5.2.2. To solve Eq. (5.173), we use the Green's function of the d'Alembert's operator, $\Box$:

$$G\left(x^\mu - y^\mu\right) = -\frac{1}{4\pi\left|\vec{x}-\vec{y}\right|}\delta^{(D)}\left[\left|\vec{x}-\vec{y}\right| - \left(x^0 - y^0\right)\right] H(x^0 - y^0) \ , \tag{5.176}$$

where $\delta^{(D)}(x)$ is the Dirac delta "function", $H(x)$ the Heaviside function, and where we used the notation $\vec{x} = \left(x^1, x^2, x^3\right)$.

[5]This is also true for vectors: since, generally, $T_{0i} \neq 0$, vector modes of perturbations are also sourced.

Green function of d'Alembert operator

The Green function of the d'Alembert operator is defined as the solution of:

$$\Box_x G\left(x^\mu - y^\mu\right) = \delta^{(D)}\left(x^\mu - y^\mu\right) \ , \tag{5.177}$$

where the subscript in $\Box_x$ indicates the variable on which the operator acts. Fo an equation of the form:

$$\Box_x f\left(x^\mu\right) = S\left(x^\mu\right) \ , \tag{5.178}$$

the function:

$$f\left(x^\mu\right) = \int G\left(x^\mu - y^\mu\right) S\left(y^\mu\right) \mathrm{d}^4 y \ , \tag{5.179}$$

is a solution. Indeed:

$$\Box_x G\left(x^\mu - y^\mu\right) = \delta^{(D)}\left(x^\mu - y^\mu\right) \tag{5.180}$$

$$\Rightarrow \int \mathrm{d}^4 y S\left(y^\mu\right) \Box_x G\left(x^\mu - y^\mu\right) = \int \mathrm{d}^4 y S\left(y^\mu\right) \delta^{(D)}\left(x^\mu - y^\mu\right) \tag{5.181}$$

$$\Rightarrow \Box_x \int \mathrm{d}^4 y S\left(y^\mu\right) G\left(x^\mu - y^\mu\right) = S\left(x^\mu\right) \tag{5.182}$$

$$\Rightarrow \Box_x f\left(x^\mu\right) = S\left(x^\mu\right) \ . \tag{5.183}$$

The form of the Green function (5.176) is derived in details in appendix C. Using this Green function, we find the wave generated by a source $T_{\mu\nu}$:

$$\bar{h}_{\mu\nu}\left(t, \vec{x}\right) = 4G \int \frac{1}{|\vec{x} - \vec{y}|} T_{\mu\nu}\left[t - |\vec{x} - \vec{y}|, \vec{y}\right] \mathrm{d}^3 y \ . \tag{5.184}$$

Note the presence in the source of the *retarded time*: the field at the event $(t, \vec{x})$ is fully determined by the source, integrated in space, at the retarded time $t_r = t - |\vec{x} - \vec{y}|$. This is reminiscent of the fact that gravitational waves propagate at the speed of light. One can easily check that the Lorenz gauge condition (5.174) is satisfied by this solution because of the conservation of energy-momentum at first order: $\partial_\mu T^\mu{}_\nu = 0$.

Let us now assume that we are interested in the expression for the gravitational wave far from the

source, and that the source is itself isolated and small. Precisely, we assume that:

$$\vec{y} = \vec{y}_0 + \delta\vec{y} \quad \text{with} \quad |\delta\vec{y}| \ll |\vec{y}_0| \tag{5.185}$$

$$r = |\vec{x} - \vec{y}_0| \gg \vec{\delta y} \ . \tag{5.186}$$

In that case, denoting $\vec{r} = \vec{x} - \vec{y}_0$:

$$|\vec{x} - \vec{y}|^2 = (\vec{r} - \delta\vec{y})\,(\vec{r} - \delta\vec{y}) \tag{5.187}$$

$$= r^2 - 2\vec{r}\cdot\delta\vec{y} + O\left(|\delta\vec{y}|^2\right) \ . \tag{5.188}$$

Thus:

$$\frac{1}{|\vec{x} - \vec{y}|} = \frac{1}{r} + \frac{\vec{r}\cdot\delta\vec{y}}{r^3} + O\left(|\delta\vec{y}|^2\right) \ . \tag{5.189}$$

We can also neglect the term in $1/r^3$ and write:

$$\frac{1}{|\vec{x} - \vec{y}|} \sim \frac{1}{r} \ . \tag{5.190}$$

Then:

$$\bar{h}_{\mu\nu} = \frac{4G}{r}\int T_{\mu\nu}\,[t - r, \vec{y}]\,\mathrm{d}^3 y \ . \tag{5.191}$$

Now, let us consider the energy-momentum conservation at first order:

$$\partial_\mu T^{\mu\nu} = 0 \ . \tag{5.192}$$

For $\nu = 0$, we get:

$$\frac{\partial T^{00}}{\partial t} + \frac{\partial T^{i0}}{\partial x^i} = 0 \tag{5.193}$$

$$\Rightarrow \quad \frac{\partial^2 T^{00}}{\partial t^2} = -\frac{\partial}{\partial t}\frac{\partial T^{i0}}{\partial x^i} = -\frac{\partial}{\partial x^i}\frac{\partial T^{0i}}{\partial t} \ . \tag{5.194}$$

On the other hand, for $\nu = i$:

$$\frac{\partial T^{0i}}{\partial t} + \frac{\partial T^{ki}}{\partial x^k} = 0 \tag{5.195}$$

$$\Rightarrow \quad \frac{\partial T^{0i}}{\partial t} = -\frac{\partial T^{ki}}{\partial x^k} \ . \tag{5.196}$$

Hence:

$$\frac{\partial^2 T^{00}}{\partial t^2} = \frac{\partial^2 T^{lk}}{\partial x^l \partial x^k} \ . \tag{5.197}$$

We can then multiply this relation by $x^i x^j$ and integrate over space (going back to $\vec{y}$ to label points of the source):

$$\frac{\partial^2}{\partial t^2}\int y^i y^j T^{00}\left(t_r,\vec{y}\right)\mathrm{d}^3 y = \int y^i y^j \frac{\partial^2}{\partial y^l \partial y^k} T^{lk}\left(t_r,\vec{y}\right)\mathrm{d}^3 y \tag{5.198}$$

$$= -\int \frac{\partial}{\partial y^l}\left(y^i y^j\right)\frac{\partial T^{lk}}{\partial y^k}\mathrm{d}^3 y \tag{5.199}$$

$$= -\int \left(y^i \delta^j{}_l + y^j \delta^i{}_l\right)\frac{\partial T^{lk}}{\partial y^k}\mathrm{d}^3 y \tag{5.200}$$

$$= \int \left(\delta^i{}_k \delta^j{}_l + \delta^j{}_k \delta^i{}_l\right) T^{lk}\mathrm{d}^3 y \tag{5.201}$$

$$= 2\int T^{ij}\left(t_r,\vec{y}\right)\mathrm{d}^3 y \ . \tag{5.202}$$

Note that when performing integrations by parts, we assumed the source isolated, so that the boundary terms vanish. We conclude that:

$$2\int T_{ij}\left[t-r,\vec{y}\right]\mathrm{d}^3 y = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\int T_{00}\left[t-r,\vec{y}\right] y_i y_j \mathrm{d}^3 y \ . \tag{5.203}$$

If the source we consider is non-relativistic, we can write:

$$T_{00}\left[t-r,\vec{y}\right] = \rho\left[t-r,\vec{y}\right] \ , \tag{5.204}$$

where ρ is the mass density of the source. Then, we introduce the *second mass moment of the source* (or tensor of inertia):

$$I_{ij}\left[t'\right] = \int \rho\left[t',\vec{y}\right] y_i y_j \mathrm{d}^3 y \ , \tag{5.205}$$

as well as the total mass of the source:

$$M(t') = \int \rho\left[t',\vec{y}\right]\mathrm{d}^3 y \ . \tag{5.206}$$

Thus, far from the source, and placing the source at the origin of the coordinates to simplify the expressions ($\vec{y}_0 = \vec{0}$), we have :

Gravitational wave far from the source

$$\begin{cases} \bar{h}_{00}\,(t,\vec{r}) = \dfrac{4GM(t-r)}{r} & (5.207) \\ \bar{h}_{ij}\,(t,\vec{r}) = \dfrac{2G}{r}\ddot{I}_{ij}(t-r)\ , & (5.208) \end{cases}$$

where a dot denotes a derivative of the function I_{ij} with respect to its argument.

The h_{00} component is exactly $-2\Phi_N$, i.e. twice the opposite of the retarded Newtonian potential generated by the source considered as a point mass, as expected. For a source that is very far, this is almost constant in space and by the equivalence principle, it should not play any role in the physics of the wave, locally. It can be re-absorbed in the local definition of the time coordinate. We can therefore set it to zero without loss of generality. Finally, let us conclude this calculation by remembering that we are interested in the wave in the TT gauge, since this is the one we know how to relate to observables. We can obtain it from Eq. (5.208) by using the projection operator on the plane orthogonal to $\vec{n} = \vec{r}/r$:

$$P_{ij} = \delta_{ij} - n_i n_j\ . \tag{5.209}$$

P_{ij} gives the components of vectors and tensors in the plane orthogonal to the vector $\vec{n}$. Thus, by removing the trace, we obtain the perturbation that is traceless and transverse to the direction of propagation $\vec{n}$:

$$h^{TT}_{ij}\,(t,\vec{r}) = \left(P^k{}_i P^k{}_j - \frac{1}{2}P^{kl}P_{ij}\right)\bar{h}_{ij}\,(t,\vec{r})\ . \tag{5.210}$$

Finally, instead of using I_{ij}, it is often better to use its traceless part only, which will not affect Eq. (5.210):

$$Q_{ij} = I_{ij} - \frac{1}{3}I^i{}_i\delta_{ij}\ , \tag{5.211}$$

which can be written:

$$Q_{ij}(t') = \int \rho\,[t',\vec{y}]\left(y_i y_j - \frac{1}{3}\,|\vec{y}|^2\,\delta_{ij}\right)\mathrm{d}^3y\ , \tag{5.212}$$

and is known as the *mass quadrupolar moment* of the source. This is the quantity that naturally appears in the multipolar development of the source Newtonian potential:

$$\Phi_N\,(t,\vec{r}) = -\frac{GM}{r} + \frac{3Q_{ij}n^i n^j}{2r^3} + \dots\ . \tag{5.213}$$

In the end, we obtain the *Quadrupole formula*

$$h_{ij}^{TT}(t,\vec{r}) = \frac{2G}{r}\left(P^k{}_i P^k{}_j - \frac{1}{2}P^{kl}P_{ij}\right)\ddot{Q}_{ij}(t-r)\ . \tag{5.214}$$

5.5.2 Example: binary stars

Consider two identical stars A and B, of mass M in circular orbit in the (x^1, x^2)-plane, at a distance R from their centre-of-mass, taken as origin of the coordinate system. Then, assuming the dynamics of the source is Newtonian, we have their velocity:

$$v = \sqrt{\frac{GM}{2R}}\ . \tag{5.215}$$

Orbits have a period of $T = 2\pi R/v$, which gives the angular frequency:

$$\Omega = \sqrt{\frac{GM}{2R^3}}\ . \tag{5.216}$$

Thus, the stars have positions:

$$x_A^1 = R\cos\Omega t\ ;\ x_A^2 = R\sin\Omega t \tag{5.217}$$

$$x_B^2 = -R\cos\Omega t\ ;\ x_B^2 = -R\sin\Omega t\ , \tag{5.218}$$

and we get the mass density:

$$\rho(t,\vec{x}) = M\delta^D\left(x^3\right)\left[\delta^D\left(x^1 - R\cos\Omega t\right)\delta^D\left(x^2 - \sin\Omega t\right) + \delta^D\left(x^1 + R\cos\Omega t\right)\delta^D\left(x^2 + \sin\Omega t\right)\right]\ . \tag{5.219}$$

This allows one to calculate the quadrupolar moment:

$$Q_{ij}(t_r) = MR^2\begin{pmatrix} 2/3 + \cos 2\Omega t & \sin 2\Omega t & 0 \\ \sin 2\Omega t & -2/3 - \cos 2\Omega t & 0 \\ 0 & 0 & 0 \end{pmatrix}\ . \tag{5.220}$$

And finally:

$$h_{ij}^{TT}(t,\vec{r}) = -\frac{8GMR^2\Omega^2}{r}\begin{pmatrix} \cos\left[2\Omega(t-r)\right] & \sin\left[2\Omega(t-r)\right] & 0 \\ \sin\left[2\Omega(t-r)\right] & -\cos\left[2\Omega(t-r)\right] & 0 \\ 0 & 0 & 0 \end{pmatrix}\ , \tag{5.221}$$

or, equivalently:

$$h_{ij}^{TT}\,(t,\vec{r}) = -\frac{4(GM)^2}{rR}\begin{pmatrix}\cos\left[2\Omega(t-r)\right] & \sin\left[2\Omega(t-r)\right] & 0\\ \sin\left[2\Omega(t-r)\right] & -\cos\left[2\Omega(t-r)\right] & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{5.222}$$

A few comments are in order.

- The strain of the gravitational wave, as measured in a detector, is given by Eq. (5.158):

$$h = \frac{\delta L}{L_0} = h_{ij}^{TT}\,(t,\vec{r})\,N^i N^i\ , \tag{5.223}$$

  where N^i is the direction of the arm of the interferometer. Let us write $\vec{N} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ by introducing spherical coordinates centred on the source. Clearly:

$$h = -\frac{4(GM)^2}{rR}\cos\left[2\Omega(t-r) - 2\phi\right]\sin^2\theta\ . \tag{5.224}$$

  It is maximum for $\theta = \pi/2$, i.e. when the arms are in a plane parallel to the source plane, i.e. orthogonal to the x^3-axis and cancels for planes orthogonal to the source plane: most of the gravitational wave is emitted in the direction orthogonal to the plane of motion.

- Let us pick up a binary made of neutron stars at a cosmological distance, say $r \simeq 100$ Mpc $\simeq 3\times 10^{25}$ m, with $M \simeq 2M_\odot$, so that $GM \simeq 2\times 10^3$ m . Let us assume that they orbit at approximately $R \simeq 10R_S \simeq 4\times 10^4$ m. Then, the strain of the gravitational wave is:

$$h = \left|h_{ij}\right| \simeq 10^{-23}\ . \tag{5.225}$$

  This is very small. For an interferometer with an arm length $L_0 \simeq 10$ km, we get $\delta L \simeq 10^{-20}$m $\ll a_0 \simeq 5\times 10^{-11}$ m, the Bohr radius, which is the typical size of an atom. Not that this apparent paradox (measuring a displacement smaller than the typical size of an atom) is only really apparent: what is measured by interferometers like LIGO is not a length but a time, namely the proper time, for a free-falling observer at the centre of the interferometer, that it takes photons to traverse an arm and come back.

- The frequency of the gravitational wave is:

$$f = \frac{\Omega}{\pi} = c\sqrt{\frac{GM}{2\pi^2 R^3}}\ . \tag{5.226}$$

  With the previous numbers, we get $f \simeq 378$ Hz.

Of course, the binary system loses energy via emission of gravitational waves, which means that over time, the binary becomes harder and harder until the objects merge. This is what is actually observed by gravitational wave detectors but our perturbative analysis can only cover the early stage of the inspiral. The late stages are highly non-linear and need to be modelled via numerical simulations. Figure 5.5 shows mass characteristics of the gravitational wave sources observed in the third campaign of LIGO, Virgo and KAGRA. Each triplet of points show the masses of the binary objects infalling and the mass of the final product of the merger, usually a black hole.

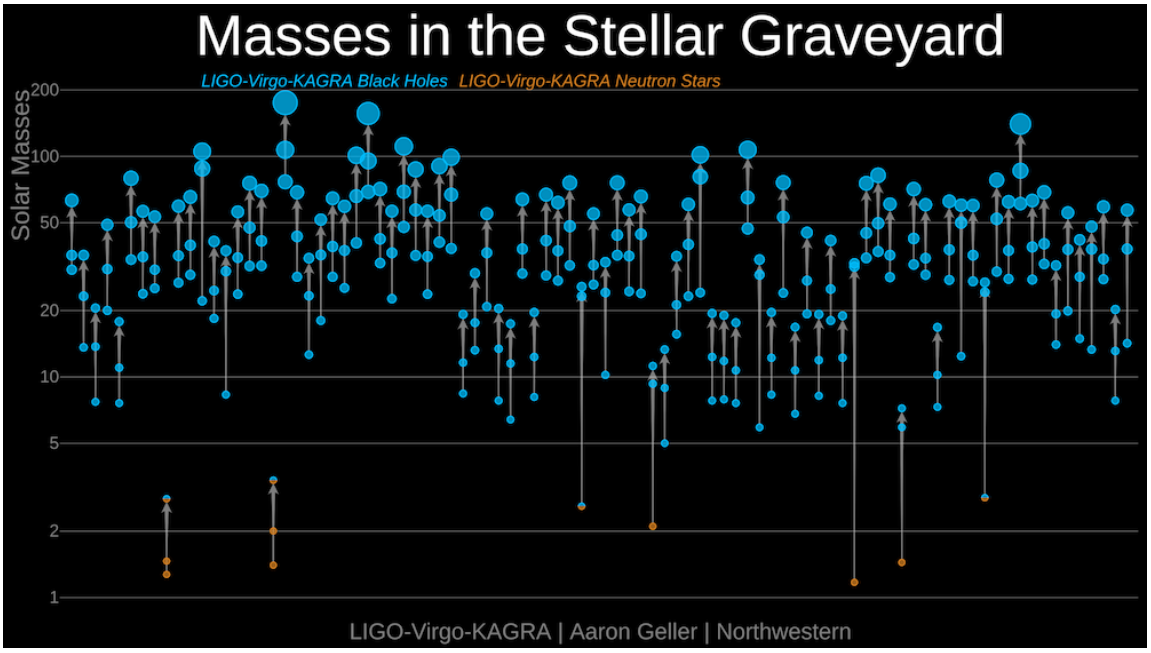


Figure 5.5: Catalogue of sources of gravitational waves observed by LIGO-Virgo-KAGRA during their third round of observations. Mergers are shows as triplets of points with two points showing the masses of the black holes or neutron stars before merger, and the third point showing the mass of the remnant black hole. Credit: LIGO-Virgo/Aaron Geller/Northwestern

6

The homogeneous and isotropic universe: the Friedman-Lemaître-Robertson-Walker spacetime

Contents

The purpose of this chapter is to introduce the basic properties of the modern description of the Universe on large scales. The second year Cosmology course will go into much more details. In particular, in these notes we will stick to the homogeneous and isotropic Universe.

6.1 What is cosmology, and what is it not?

What is the world made of? What is its shape? Did it have a beginning or always existed? Does it have boundaries or not? What is its size? What is its fate? Where are we in it?
All these questions have helped shape human cultures. They are questions about the Universe and our place inside it. They are at the heart of Cosmology: they are central to any attempt, mythical, mystical, religious, metaphysical etc., at finding our place in existence. However with the advent of modern science in the 17^{th} century, some of these questions have started to receive scientific, rather than metaphysical or mythical answers, they have be incorporated into the scientific discourse. Of course, to this day, some of those questions have remained outside the purview of science, such as, e.g. the notion of origin of the Universe. The story these notes aim to tell is about the ones who can, partially or in full, receive scientific answers, within the context of current physical theories. This means answers that are revocable, subject to the tribunal of observations, experiments and theoretical arguments. This means that the model presented here is only temporary and constantly being tested and revised, at least in its minute details.
For all those scientific questions, we can use the word cosmology, dropping the capital letter.[1]
Before we start diving into physical cosmology, it is worth reflecting on the origin and meaning of the word cosmology. 'Cosmos' is a Greek word ('$\kappa\acute{o}\sigma\mu o\varsigma$') that originally means order, good order, but also jewellery or (physical) ornament. This makes for an a priori surprising relation that we still encounter today in the proximity of words such as cosmology and cosmetics. This probably comes from an analogy drawn between the bright stars 'embellishing' the night sky and jewellery such as pearl necklaces etc., also used to embellish the earthly body.
The Universe is full of these wondrous embellishments and I hope this course will help illuminate some of those: behind the sometimes dry and tedious calculations, one must try never to forget the magnificent and awesome realities that we are trying to describe.

[1] Thanks are due to J.-P. Uzan for having introduced this use in his book 'Big-bang, Comprendre l'univers depuis ici et maintenant', Flammarion 2018.

In this short introduction, we will list a few basic observational facts about our Universe. Primarily this will be useful to set the characteristic scales that will be studied in the notes. In addition, the two main facts we will encounter, i.e. the recession of distant galaxies and the statistical isotropy of the distribution of matter around us, will be the basic starting points for the construction of a model of the Universe that will be explored in the rest of these notes.

6.2 The Observed Universe: basic facts

Admittedly, modern cosmology started with the discovery by Slipher, Hubble and others at the beginning of the XX^{th} century, that the distant nebulæ of old times were in fact distant galaxies, in all points similar to our own[2]. Very quickly, these first physical cosmologists measured the velocities of these galaxies using the redshift experienced by spectroscopic lines in the light they emit. This led to the discovery of the universal recession of distant galaxies: seen from our point of view, distant galaxies appear to be moving away from us, with a velocity proportional to their distance to us. This is *Hubble-Lemaître's law*; see Fig. 6.1:

$$v = H_0 d \ , \tag{6.1}$$

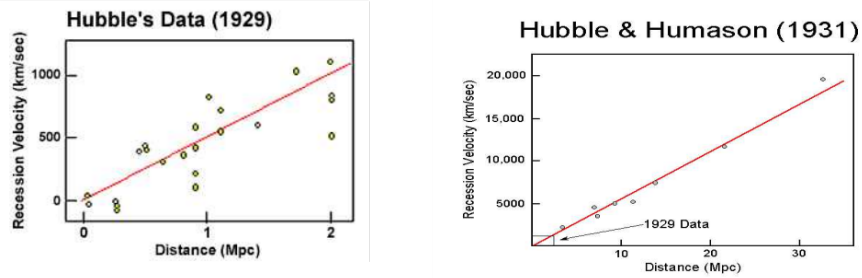
where H_0 is known as the *Hubble constant*. Its modern value is currently the topic of a controversy but it is in the range:

$$H_0 = 65 - 75 \text{km/s/Mpc}. \tag{6.2}$$

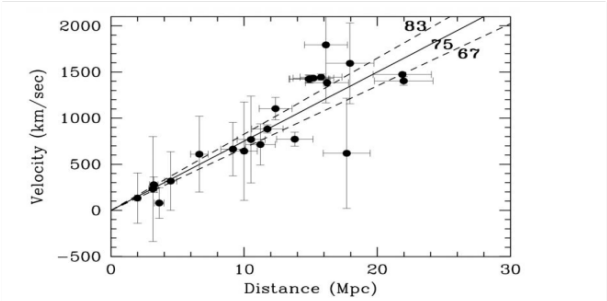
It means that an object located at a distance from us of 1 Mpc, moves away from us with a velocity of 65 to 75 km/s. This law tells us that the Universe is expanding around us: it is a dynamical, evolving object for which we can try and uncover a history. Writing this history is the task of cosmology. As you can see, a new unit has appeared here: the parsec, symbol pc. It is a very useful and common unit in cosmology. It is defined as the distance at which an "object" that measures 1 AU subtends an angle on the sky of 1 arcsecond:

$$1 \text{ pc} = \frac{648000}{\pi} \text{ AU} \simeq 3.1 \times 10^{13} \text{ m} \simeq 3.26 \text{ light-years}. \tag{6.3}$$

[2]Kant contemplated such an idea with his island Universes, presented in one of his first book: "Universal Natural History and Theory of the Heavens", published in 1755; this was the first attempt to apply Newton's theory of gravitation to the building of a cosmology.



(https://starchild.gsfc.nasa.gov/docs/StarChild/questions/redshift.html)



(Freedman et al (2001), ApJ 553, 47)

Figure 6.1: The Hubble law, then and now. Recent measurements from [11].

The star nearest to the Sun, Proxima Centauri, is at 1.3 pc from here. The disk of the Milky Way is some 30 kpc wide, and the Sun is located approximately 8 kpc from the centre of the Milky Way. The nearest galaxy is Andromeda, and it is about 780 kpc from us. Going further away, the nearest large cluster of galaxy, the Virgo cluster is about 17 Mpc from here. It has a typical size of 1 Mpc. Large scale structures such as filaments and walls along which galaxies align in the Universe can be several Gpc across, and the visible Universe has a radius of approximately 50 Gpc. These notes are concerned with the dynamics of the Universe and structures found in it on scales typically larger than 1 Mpc, all the way up to the size of the visible Universe. This means that the physics we will describe has to span approximately 4 orders of magnitude in physical size today. What happens when we look on the largest of these scales, that is if we are only concerned with describing the Universe smoothed on scales of a few hundreds of Mpc? The Universe appears extremely regular, when looked at on such large enough scales. This is visible in surveys of galaxies, which simply count the number of distant objects; see Fig. 6.2. But this is even more stricking when looking as far back as possible, and measuring the background relic radiation known as the Cosmic Microwave Background (CMB), the

remnant of an epoch known as decoupling, when photons decoupled from matter and became free to propagate in the Universe, creating a thermal bath which today consists of approximately 400 to 500 photons per cubic cm^3 all over the Universe; see Fig. 6.2. This background radiation corresponds to a black body radiation of temperature $T_0 \simeq 2.725$ K, which is extraordinarily isotropic around us: fluctuations in this temperature do not exceed 1 part in 100 000. Therefore the Universe is **statistically isotropic around us**. Observed on large enough scales, it is isotropic, and on top of this isotropic background, one can detect small fluctuations on small scales which average out when smoothed appropriately.

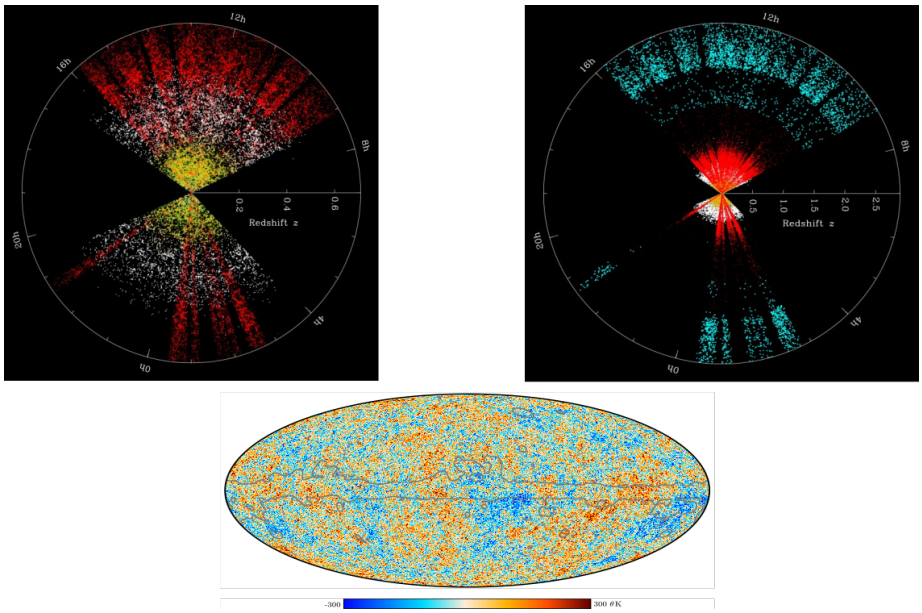


Figure 6.2: Top left: Combination of SDSS normal galaxies (yellow dots), SDSS Luminous Red Galaxies (white dots) and BOSS LUminous Red Galaxies (red dots). Each point is a galaxy. On such diagrams, we are located at the centre and the distance to this centre denotes the redshift, which is linked to the distance to us; see later: the farther an object, the larger its redshift. Top Right: Same as top left, but with BOSS quasars added (blue dots), probing a much deeper Universe. Credit: M. Blanton/SDSS. Bottom: Temperature anisotropies in the Cosmic Microwave Background measured by PLANCK.

So here is our task: building a cosmological model that can describe a Universe smooth on large

scales, but full of structures on small scales. Let us start our journey.

6.3 The Friedmann-Lemaître-Robertson-Walker Universe

6.3.1 Metric

As emphasised in the previous section, on large enough scales, the Universe appears remarkably isotropic around us: the temperature of the cosmic microwave background radiation does not vary by more that one part in 100000 over the whole sky, and the distribution of galaxies in the late Universe is also very isotropic when smoothed on sufficiently large scales. This fact will be of great use to simplify the description of our Universe on large scales. Indeed, the prospect of solving the equations of General Relativity without any hypothesis on the symmetries of the solution is absolutely daunting (6 independent, coupled, non-linear partial differential equations!), so any guidance towards simplifying assumptions is very welcome indeed. Let us thus assume the following:

Observed isotropy

On average, our Universe is statistically isotropic around us.

Unfortunately, we do not have any direct access to what the Universe could look like to distant observers, located in other galaxies and to move forward, we have to assume the Copernican principle that is not directly based on as simple observational facts as isotropy[3]:

The Copernican principle

We are typical cosmological observers; i.e. we do not occupy a special spatial location in our Universe.

This means that whatever properties of our Universe we observe, on average, any other observer should observe the same properties. In particular, since we have assumed average isotropy around us, the Universe must appear isotropic, on average, to any other typical observer. This is known as the cosmological principle and as we are going to see, taken as a strong statement, it determines the geometry of our Universe unambiguously.

Let us, for now, simplify our description a bit further and drop the "average" qualification from these

[3]However, one can now try and test this principle; see, e.g. [8] for a review.

statements. That is, let us assume that the Universe is perfectly isotropic for any typical observer. Let us populate our spacetime with a family of such typical observers, each with its own worldline thus defining a field of timelike vectors u that are the 4-velocities of these observers. This preferred set of observers is really important and is also known as the set of fundamental observers. Let us call t the proper time measured by these observers along their worldlines. Their flow in spacetime defines a preferred foliation of spacetime into hypersurfaces Σ_t orthogonal to the field u at every point, such that the metric tensor takes the form:

$$\boldsymbol{g} = -\boldsymbol{u} \otimes \boldsymbol{u} + \boldsymbol{\gamma}(t) \ , \tag{6.4}$$

where $\boldsymbol{\gamma}(t)$ is the induced metric on the spatial slice Σ_t at fixed proper time t. According to the cosmological principle, the hypersurfaces Σ_t ought to be isotropic around each of their points. This means that any quantity defined on Σ_t is spherically symmetric around each point. But this implies that the hypersurfaces Σ_t are also homogeneous, i.e. that each quantity defined on them is invariant by translation as well. Thus the Copernican principle, combined with isotropy around fundamental observers implies that the hypersurfaces Σ_t are invariant under rotations and translations, i.e. maximally symmetric; see appendix B for details about such 3 dimensional hypersurfaces. Combining all this, on large enough scales, the geometry of the Universe is well-approximated by the *Friedmann-Lemaître-Robertson Walker* metric (hereafter FLRW metric):

The Friedmann-Lemaître-Robertson Walker (FLRW) metric

$$\boldsymbol{g} = -\mathrm{d}t \otimes \mathrm{d}t + a^2(t)\left[\frac{1}{1-Kr^2}\mathrm{d}r \otimes \mathrm{d}r + r^2\left(\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta\,\mathrm{d}\phi \otimes \mathrm{d}\phi\right)\right] \tag{6.5}$$

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-Kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \ , \tag{6.6}$$

where:

1. t is the *proper time measured by fundamental observers* (those seeing an isotropic and homogeneous Universe);
2. r is the *coordinate radial distance*;
3. $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2$ is the round metric on the 2-sphere S^2, also called the "celestial sphere" (sky) of fundamental observers;

4. $a(t)$is the *scale factor*;

5. $K \in \mathbb{R}$ is the *scalar curvature of space*.

These coordinates are not the ones we introduce in appendix B, but we will see how they are mapped into each other below. The non-zero connection coefficients of the FLRW metric in (t, r, θ, ϕ) coordinates are:

Connection coefficients of the FLRW metric in (t, r, θ, ϕ) coordinates

$$\Gamma^0{}_{ij} = \frac{\dot{a}}{a} g_{ij} \; ; \; \Gamma^1{}_{01} = \Gamma^1{}_{10} = \frac{\dot{a}}{a} \tag{6.7}$$

$$\Gamma^1{}_{11} = \frac{Kr}{1 - Kr^2} \; ; \; \Gamma^1{}_{22} = -r(1 - Kr^2) \; ; \; \Gamma^2{}_{33} = -r(1 - Kr^2) \sin^2\theta \tag{6.8}$$

$$\Gamma^2{}_{02} = \Gamma^2{}_{20} = \frac{\dot{a}}{a} \; ; \; \Gamma^2{}_{12} = \Gamma^2{}_{21} = \frac{1}{r} \; ; \; \Gamma^2{}_{33} = -\sin\theta\cos\theta \tag{6.9}$$

$$\Gamma^3{}_{03} = \Gamma^3{}_{30} = \frac{\dot{a}}{a} \; ; \; \Gamma^3{}_{13} = \Gamma^3{}_{31} = \frac{1}{r} \; ; \; \Gamma^3{}_{23} = \Gamma^3{}_{32} = \frac{\cos\theta}{\sin\theta} \; . \tag{6.10}$$

Here and afterwards, a dot will denote a derivative with respect to the time coordinate t.

6.3.2 Kinematics

We can now explore the basic geometric properties of the FLRW metric. t is the proper time measured by fundamental observers along their worldlines defined by $\mathrm{d}r = \mathrm{d}\theta = \mathrm{d}\phi = 0$. It is often called the *cosmic time*. The 4-velocity of fundamental observers is then simply, in those coordinates:

$$u^\mu = \delta^\mu_0 . \tag{6.11}$$

In the following, we will denote by γ_{ij} the components of the metric of conformal space:

$$\gamma_{ij} = \frac{1}{\sqrt{1 - Kr^2}} \delta^r_i \delta^r_j + r^2 \delta^\theta_i \delta^\theta_j + r^2 \sin^2\theta \delta^\phi_i \delta^\phi_j \; . \tag{6.12}$$

Now, consider two fundamental observers, spatially separated ($\mathrm{d}t = 0$), and located at r and $r + \Delta r$, $\phi = \phi_0$ and $\theta = \theta_0$, with $\Delta r \ll 1$. Then, the physical distance between these observers is given by:

$$\Delta d_{\mathrm{phys}}(r, t) = a(t) \frac{\Delta r}{\sqrt{1 - Kr^2}} \; . \tag{6.13}$$

The number K represents the curvature of the spacelike hypersurfaces at constant t (denoted κ in subsection B.5.2 of appendix B) and we can already see that for physical reasons, r has a finite range in the case $K > 0$. Although this coordinate system, (t, r, θ, ϕ) is natural from a physical point of view, one can introduce a new set of coordinates that proves much more useful from a mathematical and physical point of view. First, let us introduce the *conformal time* η, such that:

$$\mathrm{d}\eta = \frac{\mathrm{d}t}{a(t)} \,, \tag{6.14}$$

or in integral form:

$$\eta - \eta_0 = \int_{t_0}^{t} \frac{\mathrm{d}t'}{a(t')} \,. \tag{6.15}$$

This time coordinate allows one to "factor out" the scale factor and rewrite the line element:

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + \frac{\mathrm{d}r^2}{1-Kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \,, \tag{6.16}$$

where, as usual, we have used the physicist's abuse of notation and set $a(\eta) \equiv a\left(t(\eta)\right)$, with $t(\eta)$ obtained by inverting the relation $\eta(t)$ coming from Eq. (6.15). Next, we introduce a radial coordinate χ adapted to the type of spatial curvature K, such that:

$$\mathrm{d}\chi = \frac{\mathrm{d}r}{\sqrt{1-Kr^2}} \,, \tag{6.17}$$

or equivalently, setting $\chi = 0$ for $r = 0$:

$$\chi = \int_0^r \frac{\mathrm{d}r'}{\sqrt{1-Kr'^2}} \,. \tag{6.18}$$

As a matter of fact, the integration in this case is quite easy to perform, and one gets:

$$r(\chi) \equiv S_K(\chi) = \begin{cases} \frac{1}{\sqrt{K}}\sin\left(\sqrt{K}\chi\right) & \text{for } K > 0 \\ \chi & \text{for } K = 0 \\ \frac{1}{\sqrt{-K}}\sinh\left(\sqrt{-K}\chi\right) & \text{for } K < 0 \end{cases} \,. \tag{6.19}$$

Note that this makes apparent what the admissible range of the radial coordinate is:

- For $K > 0$, as $r \in \left[0, 1/\sqrt{K}\right]$, $\chi \in [0, \pi]$;
- For $K \leq 0$, as $r \in [0, +\infty)$, $\chi \in [0, +\infty)$.

This is the same χ than the one introduced in subsection B.5.2 of appendix B. Then, the line element finally reads:

$$\mathrm{d}s^2 = a^2(\eta)\left[-\mathrm{d}\eta^2 + \mathrm{d}\chi^2 + S_K^2(\chi)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right)\right] \ . \tag{6.20}$$

This form is both remarkable and convenient for various reasons.

- Spatial hypersurfaces of constant η are, up to a conformal factor $a^2(\eta)$ the simplest constant curvature 3-dimensional manifolds. For $K = 0$, we recover the standard Euclidean "flat" space with its flat metric in spherical coordinates, $\mathbb{E}^3$. For $K > 0$, this is simply the round metric on a 3-sphere $\mathbb{S}^3$ of radius $1/\sqrt{K}$. And, finally, for $K < 0$, this is the standard metric on hyperbolic space $\mathbb{H}^3$. Note that this form also makes it clear that the radius of a 2-sphere at coordinate distance χ from the origin is given by $S_K(\chi)$ in the sense that the physical area of such a 2-sphere (at $\mathrm{d}\eta = \mathrm{d}\chi = 0$) is exactly $4\pi S_K^2(\chi)$.
- Radial light rays ($\mathrm{d}s^2 = \mathrm{d}\theta = \mathrm{d}\phi = 0$) are straight lines at $\pm\pi/4$ angles: $\chi - \chi_0 = \pm\,(\eta - \eta_0)$.

In these coordinates, the non-zero connection coefficients are: (symmetry in lower indices is implicit)

Connection coefficients of the FLRW metric in $(\eta, \chi, \theta, \phi)$ coordinates

$$\Gamma^0{}_{11} = \frac{a'}{a} \ ; \ \Gamma^0{}_{22} = \frac{a'}{a}S_K^2 \ ; \ \Gamma^0{}_{33} = \sin^2\theta\,\Gamma^0{}_{22} \tag{6.21}$$

$$\Gamma^1{}_{01} = \frac{a'}{a} \ ; \ \Gamma^1{}_{22} = -S_K\frac{\mathrm{d}S_K}{\mathrm{d}\chi} \ ; \ \Gamma^1{}_{33} = \sin^2\theta\,\Gamma^1{}_{22} \tag{6.22}$$

$$\Gamma^2{}_{02} = \frac{a'}{a} \ ; \ \Gamma^2{}_{12} = \frac{1}{S_K}\frac{\mathrm{d}S_K}{\mathrm{d}\chi} \ ; \ \Gamma^2{}_{33} = -\sin\theta\cos\theta \tag{6.23}$$

$$\Gamma^3{}_{03} = \frac{a'}{a} \ ; \ \Gamma^3{}_{13} = \frac{1}{S_K}\frac{\mathrm{d}S_K}{\mathrm{d}\chi} \ ; \ \Gamma^3{}_{23} = \frac{\cos\theta}{\sin\theta} \ . \tag{6.24}$$

Here and from now on, a prime will denote a derivative with respect to conformal time, η. Fundamental observers have 4-velocity $\boldsymbol{u} = \frac{\partial}{\partial t}$ with components $(1,0,0,0)_{(t,r,\theta,\phi)}$, thus, in this new coordinate system, $u^\mu = \left(\frac{1}{a},0,0,0\right) = \frac{1}{a}\delta^\mu_0$. If we consider two such fundamental observers located in space at $\vec{x}_1$ and $\vec{x}_2$, their physical separation at time t is given by:

$$\vec{r}_{12} = a(t)\,(\vec{x}_1 - \vec{x}_2) \ . \tag{6.25}$$

The position vectors $\vec{x}_1$ and $\vec{x}_2$ are constant in time by definition of fundamental observers.Thus:

$$\frac{\mathrm{d}}{\mathrm{d}t}\vec{r}_{12} = \dot{a}\,(\vec{x}_1 - \vec{x}_2) = \frac{\dot{a}}{a}\vec{r}_{12}. \tag{6.26}$$

The function:

$$H(t) \equiv \frac{\dot{a}}{a} \tag{6.27}$$

is called the *Hubble rate* and Eq. (6.26) is the *Hubble-Lemaître's law*. Written at present time, $t = t_0 \sim 13.7$ Gyr, it gives the historical Hubble-Lemaître's law, and reads:

$$\vec{v} = H_0\vec{r}\,, \tag{6.28}$$

with $H_0 = H\,(t_0)$ the *Hubble constant*. It expresses the fact that cosmological objects like galaxies move with respect to each other with a velocity that is greater the farther they are from each other. In an expanding Universe, $H_0 > 0$ and the velocity is a recession velocity: distant galaxies move away from each other.

Finally, let us focus on the trajectories and properties of light rays in the FLRW Universe. In the geometric optics limit (i.e. when the wavelength of the light considered is small with respect to the typical curvature radius of spacetime), valid in the cosmological context, the propagation of electromagnetic waves is well-approximated by the properties of light rays, i.e. null curves with tangent vector field $\boldsymbol{k}$ with components $k^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ satisfying:

$$\begin{cases} \boldsymbol{g}\,(\boldsymbol{k},\boldsymbol{k}) = k_\mu k^\mu = 0 & (6.29) \\ \nabla_{\boldsymbol{k}}\boldsymbol{k} = k^\nu\nabla_\nu k^\mu = 0\,. & (6.30) \end{cases}$$

Here, λ is an affine parameter along the light ray considered. Let $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$ be the components of the *projection tensor* $\boldsymbol{h} = \boldsymbol{g} + \boldsymbol{u}\otimes\boldsymbol{u}$ which projects orthogonally on hypersurfaces of constant t (or equivalently constant η). Then, in cosmic time coordinates:

$$h_{\mu\nu} = g_{\mu\nu} + \delta_{\mu 0}\delta_{\nu 0}\,, \tag{6.31}$$

so that $h_{0\mu} = 0$. Then, let $E = -k^\mu u_\mu$ and $p^\mu = h^\mu{}_\nu k^\nu$. For a future directed light ray, E is the energy of the light ray (for a past directed light ray it is minus the energy) as measured in the rest-frame of the fundamental observer, and p^μ/E are the components of the instantaneous direction of propagation of the light ray in the same rest frame; it is everywhere orthogonal to the 4-velocity of

the observers: $p^\mu u_\mu = 0$. p^μ are simply the components of the 3-momentum of the photons. Then, as we have seen in chapter 3, we can write uniquely:

$$k^\mu = E u^\mu + p^\mu \ . \tag{6.32}$$

Using $\boldsymbol{g}\left(\boldsymbol{k}, \boldsymbol{k}\right) = 0$, we get:

$$-E^2 + a^2 \gamma_{ij} p^i p^j = 0. \tag{6.33}$$

Then, projecting the null geodesic equation along $\boldsymbol{u}$:

$$u_\nu \left(k^\mu \nabla_\mu k^\nu\right) = 0 \ , \tag{6.34}$$

we get:

$$E\dot{E} + a^2 H \gamma_{ij} p^i p^j = 0 \ . \tag{6.35}$$

Hence, using Eq. (6.33), we obtain:

$$\frac{\dot{E}}{E} = -H = -\frac{\dot{a}}{a} \ , \tag{6.36}$$

which is trivial to integrate, to get:

$$E = \frac{C_0}{a} \ , \ C_0 \in \mathbb{R} \ . \tag{6.37}$$

For a light ray with frequency ν, the energy of a photon is given by $E = h\nu$, so that the frequency of light is affected by cosmic expansion along the trajectory of photons, according to:

$$\nu(t) = \frac{a\left(t_e\right)}{a(t)} \nu\left(t_e\right) \ , \tag{6.38}$$

where t_e is the time at which the photons have been emitted by their source located at $(t_e, \chi_e, \theta_e, \phi_e)$. If a fundamental observer located at the centre of the coordinate system ($\chi = 0$) receives this light today, at $t = t_0$, the redshift z is defined by:

$$z \equiv \frac{\lambda\left(t_0\right) - \lambda\left(t_e\right)}{\lambda\left(t_e\right)} \ , \tag{6.39}$$

and in the FLRW context:

$$\boxed{1 + z = \frac{a\left(t_0\right)}{a\left(t_e\right)}} \ . \tag{6.40}$$

The name redshift is justified by the fact that, in an expanding universe, $a\left(t_0\right) > a\left(t_e\right)$, so that $\lambda\left(t_0\right) > \lambda\left(t_e\right)$: the wavelength of the light has been moved to higher values, towards the redder

part of the spectrum. In a purely expanding FLRW Universe, there is a one-to-one and onto relationship between times of emission and redshifts at observation, so that one can interchangeably use either t or $z(t) = a(t_0)/a(t) - 1$ to characterise past events. We will use this freedom extensively in what follows. Moreover, note that scale factor and coordinate radial distance are only defined simultaneously up to an overall scaling. This means that by setting the units for radial distances appropriately at time $t = t_0$ ("today"), one can always set $a_0 = a(t_0) = 1$. From now on, we will fix units this way. We will also, by convention, agree that a subscript 0 attached to any function corresponds to the value of that function at the value of the proper time today, t_0, or equivalently at the present value of the conformal time η_0 (or equivalently at $z = 0$).

6.3.3 Distances

The coordinate distances given by the radial coordinates r and χ, such as the one used to derive the Hubble-Lemaître's law are not measurable quantities in General Relativity, as they are calculated purely by the spacelike separation of two events in spacetime.Physically meaningful distances ought to be related to observable quantities involving causal processes; in cosmology such physically relevant distances are obtained by determining distances measured down the past lightcone of an observer, because almost every piece of information we get about the distant Universe is obtained via electromagnetic observations. We will define two relevant distances, related respectively to the luminosity of sources and to their angular size. But before we define these physical distances, it is convenient to introduce one coordinate distance that is important in deriving them: the comoving radial distance.

Comoving radial distance

Consider a fundamental observer located at $\chi = 0$, receiving at $\eta = \eta_0$ light that was emitted by a distant source at a time t corresponding to a redshift z. By an appropriate choice of our coordinate system, we can ensure that the light ray propagates radially, with $\mathrm{d}\theta = \mathrm{d}\phi = 0$. Then, along the light ray propagating from the source to the observer, we have $\mathrm{d}s^2 = 0$, i.e.:

$$\mathrm{d}\chi = -\frac{\mathrm{d}t}{a(t)} = -\frac{\mathrm{d}a}{a^2 H} \,. \tag{6.41}$$

The minus sign ensures that the ray propagates forward in time from the source at $\chi > 0$ to the observer at $\chi = 0$. Using a as a "time" variable instead of t is safe as long as they are related

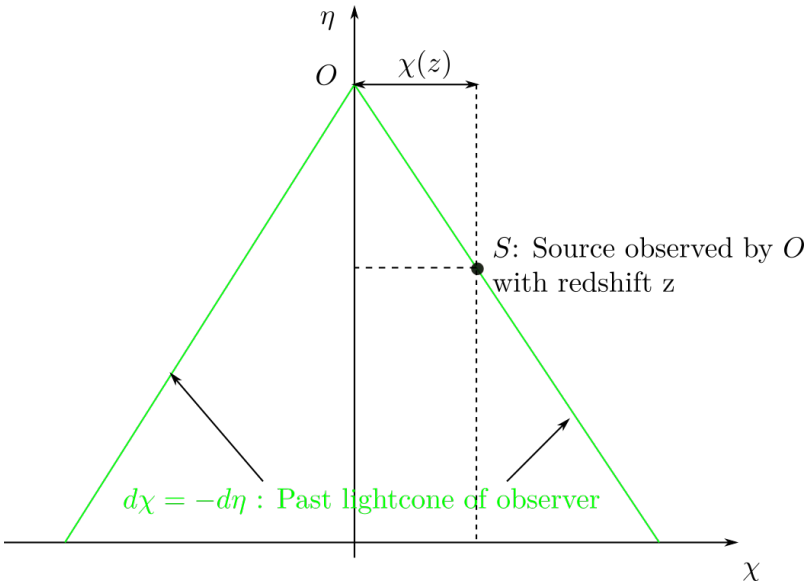


Figure 6.3: Definition of the comoving radial distance.

in a monotonous way, which is true in an expanding Universe (see dynamics below). Then, using $1 + z = 1/a$, we get:

$$\mathrm{d}\chi = \frac{\mathrm{d}z}{H(z)} \,. \tag{6.42}$$

The *comoving radial distance* between the source and the observer is then simply the change in χ along the light ray between source and observer (see Fig. 6.3), and it is obtained by integrating the previous differential relation:

$$\chi(z) \equiv \int_0^z \frac{\mathrm{d}z'}{H(z')} \,. \tag{6.43}$$

It is not an observable. It can be used to defined another unobservable, but important distance: the *comoving angular distance*. Consider the comoving 2-sphere at $\mathrm{d}\eta = \mathrm{d}\chi = 0$ at $\chi = \chi(z)$, then, its round metric gives the line element (it is comoving so we ignore the scale factor):

$$\mathrm{d}s^2_{com} = S^2_K(\chi(z)) \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2\right) \,. \tag{6.44}$$

A small source located on that sphere and observed at the centre under a small solid angle $\mathrm{d}\Omega^2_{obs}$ subtends a small transverse area portion of the sphere $\mathrm{d}S^{source}_{com}$ such that:

$$\mathrm{d}S^{source}_{com} = S^2_K(\chi(z))\mathrm{d}\Omega^2_{obs} \,; \tag{6.45}$$

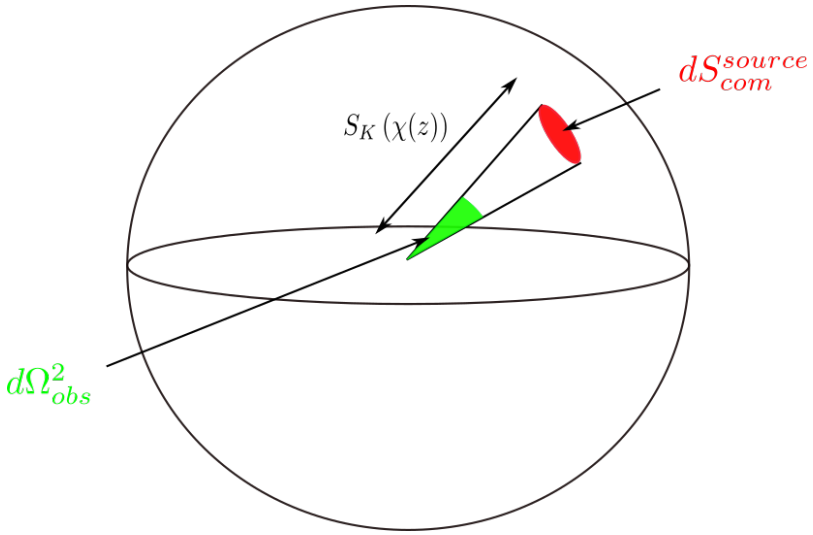


Figure 6.4: Definition of the comoving angular distance.

see Fig. 6.4 for a detail of the geometry.

The *comoving angular distance* between the source at redshift z and the observer is then defined as the ratio:

$$R^2_{ang}(z) \equiv \frac{\mathrm{d}S^{source}_{com}}{\mathrm{d}\Omega^2_{obs}}. \tag{6.46}$$

Thus:

$$\boxed{R_{ang}(z) = S_K(\chi(z)) \ .} \tag{6.47}$$

The effect of curvature on this comoving angular distance is summarised on Fig. 6.5. The green curves represent light rays coming from the boundary of the small distant object and reaching the observer at the point of convergence. An object of the same size, located at the same coordinate distance $\chi(z)$ will have a different observed angular size in spaces of different curvature. The black dotted lines represent the opening angle observed in each case. We see that because $\sin(u)/u < 1$ and $\sinh u/u > 1$, the observed angle will be larger in the $K > 0$ case and smaller in the $K < 0$ case, compared to the $K = 0$ case.

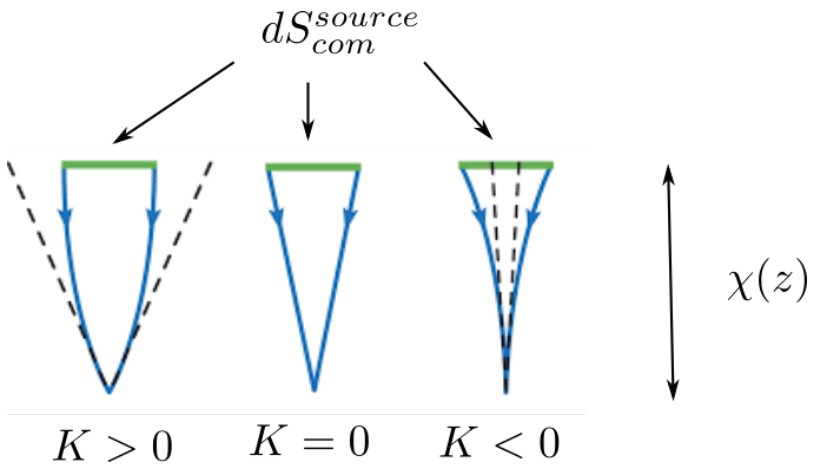


Figure 6.5: Effect of spatial curvature on the angular size of distant objects.

Angular diameter distance

The comoving angular distance is not directly observable because it depends on the comoving (coordinate) size of the source which is not observable. However, by relating this comoving size the the actual, physical transverse size of the source:

$$\mathrm{d}S_{phys}^{source} = a^2 \mathrm{d}S_{com}^{source} \ , \tag{6.48}$$

one can obtain an observable distance: the *angular diameter distance* $D_A(z)$ of an object located at redshift z with respect to the observer, defined as:

$$D_A^2 \equiv \frac{\mathrm{d}S_{phys}^{source}}{\mathrm{d}\Omega_{obs}^2} \ . \tag{6.49}$$

This is measurable in principle. Indeed, if the observer can measure the apparent angular size of the source on their sky and if they have an independent knowledge of the absolute physical size of the source (from theoretical modelling), they can deduce the angular diameter distance. This is why measurements of the angular diameter distance require the knowledge of *standard rulers*, i.e. object whose physical size is stable over time and known to great accuracy. We see that the angular diameter distance to a source located at redshift z is thus:

$$D_A(z) = aR_{ang}(z) = \frac{1}{1+z} S_K\left(\chi(z)\right) \ . \tag{6.50}$$

Luminosity distance

The other distance that turns out to be useful in cosmology makes use of another class of objects called *standard candles*. These are objects whose absolute luminosity is assumed well known and stable from independent theoretical models. So assume that an observer at $\chi = 0$ and $t = t_0$ observes such a source located at a comoving radial distance $\chi(z)$ with absolute luminosity L_{source}. Assuming that the source radiates isotropically, the observed flux, Φ_{obs} will correspond to the isotropic flux through a sphere of radius $D_L(z)$:

$$\Phi_{obs} = \frac{L_{source}}{4\pi D_L^2} \ . \tag{6.51}$$

This D_L is the *luminosity distance* between the source and the observer. By definition, the luminosity is the rate of change of energy by units of time:

$$L_{source} = \frac{\Delta E_{emit}}{\Delta t_{emit}} = \frac{\Delta E(z)}{\Delta t(z)} \ . \tag{6.52}$$

Because of the redshift experienced by light between emission and observation, the change of energy observed is given by:

$$\Delta E_{obs} = \frac{\Delta E_{emit}}{1+z} \ . \tag{6.53}$$

Moreover, For two light rays emitted from the source in an interval of conformal time $\Delta\eta_{emit}$ and arriving at the observer in an interval $\Delta\eta_0$, we have: $\Delta\eta_0 = \Delta\eta_{emit}$ (light rays are straight lines in $\eta - \chi$ coordinates). Thus, going to proper time:

$$\Delta t_0 = \frac{1}{a}\Delta t_{emit} = (1+z)\Delta t_{emit} \ . \tag{6.54}$$

Therefore, the observed luminosity is given by:

$$L_{obs} = \frac{\Delta E_{obs}}{\Delta t_0} = \frac{1}{(1+z)^2}\frac{\Delta E_{emit}}{\Delta t_{emit}} = \frac{1}{(1+z)^2} L_{source} \ . \tag{6.55}$$

On the other hand, the observed flux is the ratio of the total luminosity at the time of observation by the surface area over which this luminosity is distributed, S^{phys}. This surface area is the physical area today of the sphere centred on the source of comoving radius $R_{ang}(z) = S_K(\chi(z))$:

$$S^{phys} = a_0^2 S^{com} = 4\pi S_K^2(\chi(z)) \ . \tag{6.56}$$

Thus:

$$\Phi_{obs} = \frac{L_{source}}{(1+z)^2 \times 4\pi S_K^2(\chi(z))} \ . \tag{6.57}$$

Equating the two expression for the observed flux, we get:

$$D_L(z) = (1+z) S_K\left(\chi(z)\right) \ . \tag{6.58}$$

Note that the angular and luminosity distances are related by the *distance-duality relation*:

$$D_L(z) = (1+z)^2 D_A(z) \ . \tag{6.59}$$

This relation is actually true in any spacetime, in any metric theory of gravity, as long as the number of photons is conserved during the propagation of light between source and observer.
In a flat FLRW universe, these distances take the simple integral expressions:

Angular and luminosoty distances in flat FLRW

$$D_A(z) = \frac{1}{1+z}\int_0^z \frac{dz'}{H(z')} \tag{6.60}$$

$$D_L(z) = (1+z)\int_0^z \frac{dz'}{H(z')}. \tag{6.61}$$

These various notions of distance are all equally valid and their use depend on the physical system we want to evaluate the distance to. Fig 6.6 shows the radial comoving distance $\chi(z)$, the angular distance $D_A(z)$ and the luminosity distance $D_L(z)$ as functions of redshift for the nominal cosmology we introduce below; see Eqs. (6.145)-(6.150). Clearly, although they match for small redshifts (left panel), they differ significantly as soon as we probe further back into the past (right panel). In particular, the angular diameter distance exhibits a non-monotonous behaviour which means that after some redshift, objects that are further into the past appear smaller and smaller! Finally, note that we need knowledge of the dynamics on the FLRW Universe between the source and the observer, through the Hubble rate $H(z)$ to determine the behaviour of these distances. This dynamics is what we will focus on next.

6.3.4 Dynamics

To determine the dynamics of the FLRW Universe, one needs to write the Einstein Field Equations:

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \ , \tag{6.62}$$

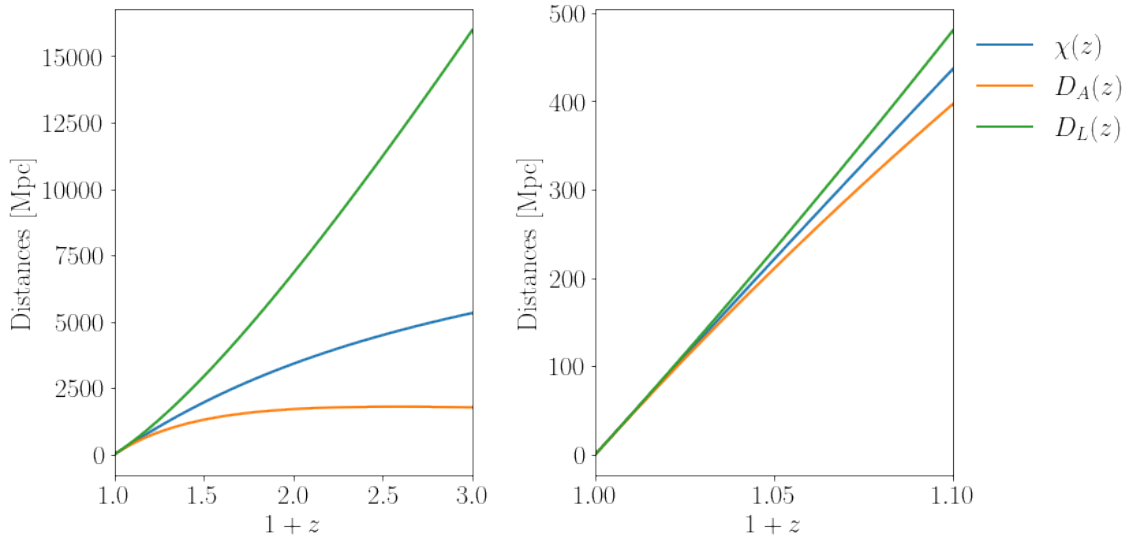


Figure 6.6: Radial comoving distance $\chi(z)$, angular distance $D_A(z)$ and luminosity distance $D_L(z)$ as functions of redshift for the nominal cosmology of Eqs. (6.145)-(6.150).

for the FLRW metric and the appropriate energy-momentum content. For the left-hand side of those equations, we have, in proper time:

Ricci tensor for the FRLRW metric in (t, r, θ, ϕ) coordinates

$$\begin{cases} R_{00} = -3\dfrac{\ddot{a}}{a} & (6.63) \\ R_{ij} = a^2\left(2H^2 + \dfrac{\ddot{a}}{a} + 2\dfrac{K}{a^2}\right)\gamma_{ij} & (6.64) \\ R = 6\left(H^2 + \dfrac{\ddot{a}}{a} + \dfrac{K}{a^2}\right) . & (6.65) \end{cases}$$

Energy-momentum content

But what of $T_{\mu\nu}$? In principle, we should include all possible particles and fields present in the Universe, photons, electrons, protons, all atoms once they have formed, neutrinos, exotic sources like Dark Matter and Dark Energy (see below) etc. Usually, these are treated as independent, non-

interacting fluids, with energy densities $\rho_i(t)$ and pressure $p_i(t)$. By symmetry, they ought to be comoving and their common 4-velocity sets the 4-velocity field of fundamental observers. Then one can show easily that these fluids ought to be perfect (no heat flux or anisotropic pressure), thus having energy-momentum tensors:

$$T^{(i)}_{\mu\nu} = (\rho_i + p_i)\, u_\mu u_\nu + p_i g_{\mu\nu} \ , \tag{6.66}$$

which separately obey a conservation equation (non-interacting):

$$\nabla_\mu T^{(i)\mu}{}_\nu = 0 \ . \tag{6.67}$$

Then, for each fluid, we can define an equation of state:

$$w_i = \frac{p_i}{\rho_i} \ , \tag{6.68}$$

and Eq. (6.67) leads to:

$$\dot{\rho}_i + 3\,(1 + w_i)\, H\rho_i = 0 \ , \tag{6.69}$$

for each individual fluid. The total energy-momentum content is then an effective fluid with effective, total density, pressure and equation of state:

$$\begin{cases} \rho = \displaystyle\sum_i \rho_i & (6.70) \\ p = \displaystyle\sum_i p_i & (6.71) \\ w = \dfrac{p}{\rho} \ , & (6.72) \end{cases}$$

modelled by the *total energy-momentum tensor*, with components in (t, r, θ, ϕ) coordinates:

$$T_{\mu\nu} = (\rho + p)\, \delta_{\mu 0}\delta_{\nu 0} + p g_{\mu\nu} \ . \tag{6.73}$$

The conservation of this total energy-momentum tensor then leads to:

$$\dot{\rho} + 3\,(1 + w)\, H\rho = 0 \ . \tag{6.74}$$

Usually, in cosmology, the various standard fluids are separated into two main classes:

Non-relativistic fluids: These are fluids whose internal velocity dispersion is small. Individual particles of the fluid move slowly compared with the speed of light. For these fluids, the pressure $p_i \sim 0$, so that $w_i \sim 0$. Standard baryonic and leptonic matter fall into this category for most of the history of the Universe. So do neutrinos in the very late-time Universe. Cold Dark Matter is also non-relativistic throughout the history of the Universe. These non-relativistic fluids are often called dust or simply matter when the context is clear.

Relativistic fluids: These are fluids with internal particle velocities close to the speed of light. In that case, $p_i \simeq \frac{1}{3}\rho_i$ so that $w_i = 1/3$. Photons are such particles. So are neutrinos for most of the history of the Universe.

But it is common to consider more exotic fluids. For example, taking the cosmological constant from the LHS to the RHS of the Einstein field equations, one can formally rewrite its effect as that of a perfect fluid with $p_\Lambda = -\rho_\Lambda$, thus $w_\Lambda = -1$. Perfect fluids with a constant equation of state are called barotropic. So dust and relativistic fluids are barotropic fluids; so is the cosmological constant if it is interpreted as a fluid. They are widely used in cosmology as they provide gvery good approximations to the actual content of the Universe.
Solving Eq. (6.69) for non-relativistic and relativistic fluids we see that:

$$\rho_{NR}(a) = \rho_{NR,0}a^{-3} \tag{6.75}$$

$$\rho_R(a) = \rho_{R,0}a^{-4} . \tag{6.76}$$

Therefore, in an expanding universe, dust is diluted by a factor proportional to the volume increase; this simply means that the number of particles (thus the total energy) in a given physical volume remains constant while the volume increases. Relativistic matter on the other hand receives an extra dilution in $1/a$; this comes from the redshift of the energy of individual photons in the fluid. It is common to write a subscript m for non-relativistic matter, and r for relativistic matter, which is what we will do from now on. For the cosmological constant, we get:

$$\rho_\Lambda = \text{cst} = \frac{\Lambda}{8\pi G} . \tag{6.77}$$

Dynamical equations

We are now ready to write the equations governing the dynamics of the FLRW Universe with a total matter content given by ρ and p. Combining the Ricci tensor and its trace from Eqs. (6.63)-

(6.65) and the total energy-momemtum tensor, Eq. (6.73) within the Einstein field equations, we get:

FLRW dynamics in cosmic time

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2} + \frac{\Lambda}{3} \quad \text{(Friedmann Eq.)} \tag{6.78}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \quad \text{(Raychaudhury Eq.)} \tag{6.79}$$

$$\dot{\rho} = -3H(\rho + p) \quad \text{(Continuity Eq.)}. \tag{6.80}$$

Note that these three equations are not independent (show it), so we only truly have two independent equations for three unknown functions. Thus, we need to assume an equation of state $p(\rho)$ to be able to solve this system. In conformal time, using the connection coefficients from Eqs. (6.21)-(6.24) and introducing the *conformal Hubble rate*:

$$\mathcal{H} = \frac{a'}{a} = aH\,, \tag{6.81}$$

we obtain the following dynamical equations:

FLRW dynamics in conformal time

$$\mathcal{H}^2 = \left(\frac{a'}{a}\right)^2 = \frac{8\pi G}{3}\rho a^2 - K + \frac{\Lambda}{3}a^2 \quad \text{(Friedmann Eq.)} \tag{6.82}$$

$$\mathcal{H}' = -\frac{4\pi G}{3}(\rho + 3p)\,a^2 + \frac{\Lambda}{3}a^2 \quad \text{(Raychaudhury Eq.)} \tag{6.83}$$

$$\rho' = -3\mathcal{H}(\rho + p) \quad \text{(Continuity Eq.)}. \tag{6.84}$$

Let us assume first, for simplicity, that the total fluid is barotropic, i.e. with a constant equation of state w: $p = w\rho$. Then the continuity equation can be easily solved:

$$\rho(a) = \rho_0 a^{-3(1+w)}\,. \tag{6.85}$$

In that case, assuming $K = \Lambda = 0$ and $w \neq 1$, we can solve the Friedmann equation and retain only the expanding solution:

$$a(t) = \left(\frac{t}{t_0}\right)^{\frac{2}{3(1+w)}} \text{ and } a(\eta) = \left(\frac{\eta}{\eta_0}\right)^{\frac{2}{1+3w}} , \tag{6.86}$$

and also:

$$H(t) = \frac{2}{3(1+w)t} \text{ or } \mathcal{H}(\eta) = \frac{2}{(1+3w)\eta} . \tag{6.87}$$

One notes that:

$$\begin{cases} a \propto t^{2/3} \propto \eta^2 \text{ for a non-relativistic fluid} & (6.88) \\ a \propto t^{1/2} \propto \eta \text{ for a relativistic fluid.} & (6.89) \end{cases}$$

Also, for a cosmological constant $\Lambda \neq 0$ only:

$$a(t) = \exp\left[\sqrt{\frac{\Lambda}{3}}\,(t - t_0)\right] . \tag{6.90}$$

These scalings wil be very important throughout.

Let us now introduce dimensionless density parameters:

$$\begin{aligned} \Omega_i(z) &\equiv \frac{8\pi G\rho_i(z)}{3H^2(z)} & (6.91) \\ \Omega(z) &\equiv \frac{8\pi G\rho(z)}{3H^2(z)} = \sum_i \Omega_i(z) \text{ (Total energy content)} & (6.92) \\ \Omega_\Lambda(z) &= \frac{8\pi G\rho_\Lambda}{3H^2(z)} & (6.93) \\ \Omega_K(z) &= -\frac{K}{a^2(z)H^2(z)}. & (6.94) \end{aligned}$$

Then the Friedmann equation becomes simply a balancing equation valid at all time/reshift:

$$\Omega + \Omega_\Lambda + \Omega_K = 1 . \tag{6.95}$$

In particular, today:

$$\sum_i \Omega_{i,0} + \Omega_{\Lambda,0} + \Omega_{K,0} = 1 . \tag{6.96}$$

For each barotropic fluid of constant equation of state w_i:

$$\Omega_i(z) = \Omega_{i,0}\left(\frac{H_0}{H(z)}\right)^2 (1+z)^{3(1+w_i)} . \tag{6.97}$$

Thus, we can introduce the dimensionless expansion rate:

$$E(z) \equiv \frac{H}{H_0} \,, \tag{6.98}$$

so that:

$$E^2(z) = \sum_i \Omega_{i,0}(1+z)^{3(1+w_i)} + \Omega_{K,0}(1+z)^2 + \Omega_{\Lambda,0} \,. \tag{6.99}$$

Cosmological eras

Finally, let us assume that the Universe is filled with a non-relativistic fluid and a relativistic one, as well as a cosmological constant. For simplicity, let us set $K = 0$. We can introduce the critical density of the Universe:

$$\rho_{c,0} = \frac{3H_0^2}{8\pi G} \,, \tag{6.100}$$

so that we have:

$$\rho_m = \Omega_{m,0}\rho_{c,0}a^{-3} = \Omega_{m,0}\rho_{c,0}(1+z)^3 \tag{6.101}$$

$$\rho_r = \Omega_{r,0}\rho_{c,0}a^{-4} = \Omega_{r,0}\rho_{c,0}(1+z)^4 \tag{6.102}$$

$$\rho_\Lambda = \Omega_{\Lambda,0}\rho_{c,0} \,. \tag{6.103}$$

Thus, as illustrated on Fig. 6.7, we see that in an expanding Universe, for generic choices of the parameters today, the Universe goes through three distinct phases:

1. $\rho(a) \sim a^{-4}$ as $a \to 0$. This is a *Radiation Dominated Era* (RDE): when the energy content and the dynamics of the Universe are dominated by the relativistic fluid;

2. At some point, the non-relativistic fluid starts to dominate the energy content and $\rho \sim a^{-3}$. This is a *Matter Dominated Era* (MDE);

3. Finally, provided one waits for long enough, since all energy densities decay except the one coming from the cosmological constant, a final epoch starts when the expansion of the Universe is governed by the cosmological constant. This is the *Dark Energy Dominated Era* (ΛDE). In the asymptotic future, when all the fluids have been infinitely diluted, the Universe is in a steady state called the de Sitter Universe.

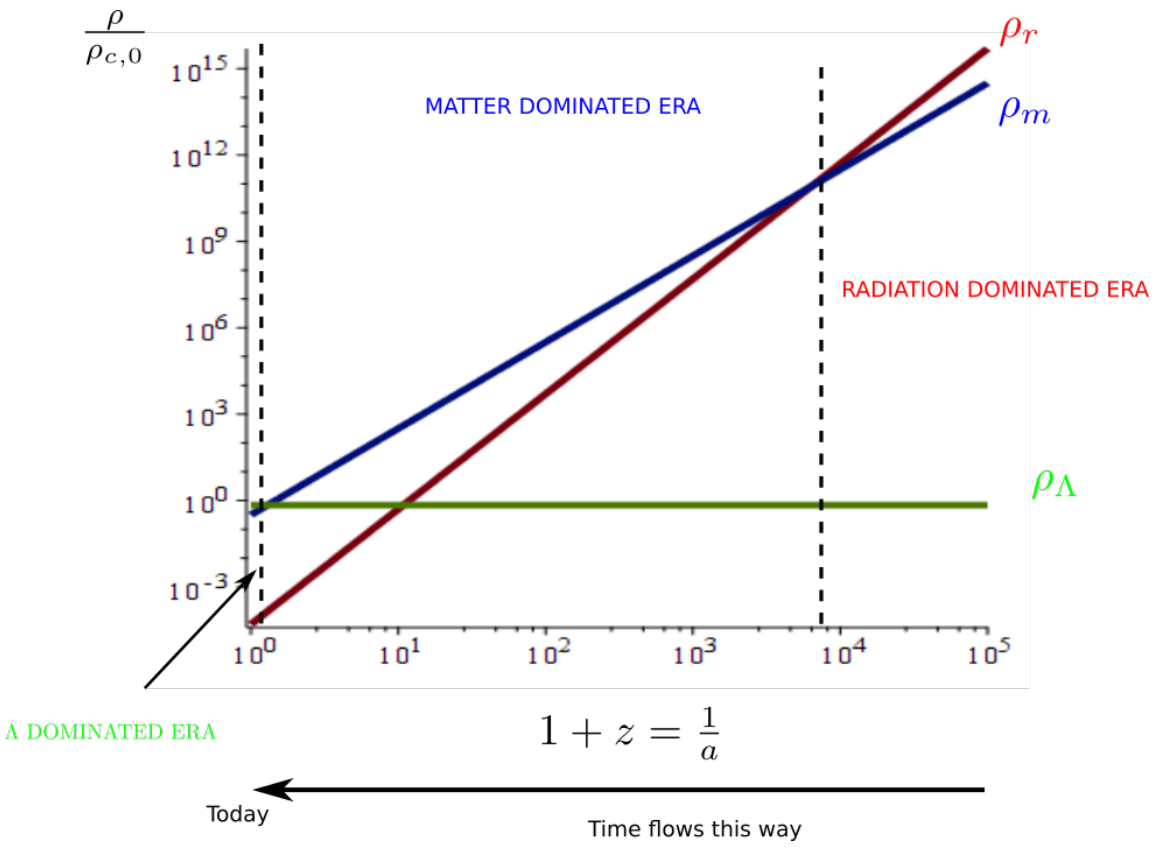


Figure 6.7: Log-log plot of the densities in a cosmology with $\Omega_{m,0} = 0.317$, $\Omega_\Lambda = 0.683$ and $\Omega_{r,0} = 2.10^{-5}$. The densities are expressed in units of the critical density. Typical values for the critical density are of the order $\rho_c \simeq 1.10^{-27}$ kg.m^{-3}.

These three phases in the history of an expanding Universe will be key for our analysis of the growth of large-scale structure next year. A last piece of information we will need about the background is the only characteristic scale that enters this extremely symmetric model: the Hubble radius. The *physical Hubble radius* is the (time-dependent) length:

$$R_H = cH^{-1} \, . \tag{6.104}$$

So we see that this scale grows during a MDE and a RDE as: $R_H(t) \propto t$ and is constant during a ΛDE. Actually, it will be more natural to consider the *comoving Hubble scale*:

$$R_{\mathcal{H}} = c\mathcal{H}^{-1}. \tag{6.105}$$

During a RDE, it goes like $R_{\mathcal{H}} = c\eta$, and during a MDE: $R_{\mathcal{H}} = c\eta/2$. On the other hand, it decreases in a ΛDE: $R_{\mathcal{H}} = R_{\mathcal{H},i}\eta_i/\eta$. We will see next year that this behaviour is key to the formation of structure.

6.3.5 The hot Big-Bang model

From the behaviour of the FLRW scale factor in presence of relativistic and non-relativistic matter fluids, we deduced that an expanding Universe, i.e. a Universe that was smaller with denser fluids in the past, ought to have undergone a transition between two phases: its early history is characterised by a Radiation Dominated era, followed by a Matter Dominated era. The relativistic fluid that dominates the dynamics during the Radiation Dominated era has a density:

$$\rho_r \propto a^{-4} \propto (1+z)^4 \, . \tag{6.106}$$

Assuming that this fluid is in thermodynamical equilibrium at temperature T and zero chemical potential (for simplicity), the energy density in terms of the distribution function, $f\,(p,T)$, of the particles in the fluid is given by:

$$\rho = \int f\,(p,T)\, E(p) d^3p \, . \tag{6.107}$$

Thus, for relativistic particles with $T \gg m$, whether particles are fermions or bosons:

$$\rho \propto T^4 \, . \tag{6.108}$$

The expanding Universe was hotter in the past (when it was also denser). This is why one talks of a *Hot Big-Bang* model. Thus, the temperature of the relativistic fluid (mostly photons) in the past is given, in terms of redshift by:

$$T(z) = T_0(1+z) \ , \tag{6.109}$$

where $T_0 \simeq 2.725$ K is the temperature of the CMB today. Strictly speaking, this is the common temperature of all matter species in the Universe only as long as all forms of matter remain in thermal equilibrium. For example, baryons only remain coupled with photons until recombination and decoupling, after which their temperature starts to deviate from the one of photons. However, it is common to call the temperature of the CMB the 'temperature of the Universe' and to use it as a clock to describe the thermal history of the Universe. Note that during the Radiation Dominated era:

$$H(T) \propto \sqrt{\rho(T)} \propto T^2 \ . \tag{6.110}$$

Thus the typical timescale of expansion of the Universe evolves as:

$$\tau_H = H^{-1} \propto T^{-2} \ . \tag{6.111}$$

Let us consider an interaction between particles with rate Γ (units of inverse time). As long as $\Gamma \gg H$, the interaction remains efficient, the particles involved in the interaction have enough time to interact before being separated by the cosmic expansion, and they remain in thermal equilibrium. However, as soon as $\Gamma < H$, the interactions freeze and the various particles involved start evolving independently: they decouple. Considering that the content of our Universe is well-described by the standard model of particle physics, this leads to an elegant thermal history of the Universe[4]:

1. $T > 100$ GeV; $z > 10^{15}$; $t < 20$ ps: Quantum Gravity; Inflation; Baryogenesis. This very early period is not described adequately by the standard model of particle physics and its details remain the topic of conjectures and speculations. For reasons to be explored later, it seems to include a phase of accelerated expansion of the Universe called inflation, or something that would produce similar signatures on the later Universe. It also needs to include a

[4]We used that $T_0 = 2.275$ K $\simeq 2 \cdot 10^{-4}$ eV and that $T(z) = T_0(1+z)$ to determine the redshifts from the temperatures. The time t is the cosmic time, conventionally set to 0 at the Big-Bang, i.e. the time at which the model becomes singular. As will become apparent when we introduce inflation, this reference time is actually quite arbitrary in standard cosmology, as the Big-Bang singularity disappears from the physical Universe and potentially even completely.

mechanism responsible for the asymmetry between matter and anti-matter that we observe today.

2. $T = 100$ GeV; $z = 10^{15}$; $t = 20$ ps: Electroweak phase transition. The electromagnetic and weak interactions separate via the Higgs mechanism, and particles acquire their masses.

3. $T = 150$ MeV; $z = 10^{12}$; $t = 20\,\mu$s: QCD phase transition. Above that temperature, quarks are asymptotically free, i.e. they are only subjected to the weak interaction. But below that temperature, the strong interaction kicks in and quarks and gluons form bound states: baryons (three quarks) and mesons (pairs quark-antiquark).

4. $T = 1$ MeV; $z = 6 \cdot 10^9$; $t = 1$ s: neutrinos decoupling. Weak interactions are no longer fast enough to maintain neutrinos in thermal equilibrium with the rest of matter. They decouple and form an hypothetical cosmic neutrino background that should permeate the whole Universe today (but has not yet been observed) with its own temperature.

5. $T = 500$ keV; $z = 2 \cdot 10^9$; $t = 6$ s: electron-positron annihilation. Electrons and positrons cannot be maintained in thermal equilibrium with photons and annihilate, releasing energies in the photon fluid (reason why the CMB has a different temperature than the cosmic neutrino background). A small asymmetry between matter and anti-matter is necessary to keep some electrons around after this phase.

6. $T = 100$ keV; $z = 4 \cdot 10^8$; $t = 3$ min: Big Bang Nucleosynthesis (BBN). Some protons and neutrons escape the thermal equilibrium and bound to form atomic nuclei via a complex network of nuclear reactions. Only the light elements are formed in any significant quantity: deuterium, helium, lithium and beryllium. The amount of each element formed during this primordial phase can be calculated very accurately in the standard model and the agreement of these predictions with observations constitutes one of the most robust pillar of the Hot Big-Bang model.

7. $T = 0.75$ eV; $z = 3400$; $t = 60$ kyr: Matter-Radiation Equality. The energy densities of relativistic and non-relativistic matter coincide.

8. $T = 0.26 - 0.33$ eV; $z = 1100 - 1400$; $t = 260 - 380$ kyr: Recombination. Electrons and baryons (mostly protons and helium nuclei) combine to form atoms (neutral hydrogen, helium

atoms) via e.g. $e^- + p \rightarrow H + \gamma$ once the converse reaction is energetically disfavoured. Matter becomes neutral and the mean-free path of photons increases rapidly. This leads to:

9. $T = 0.23 - 0.28$ eV; $z = 1000 - 1200$; $t = 380$ kyr: Photon decoupling also called simply decoupling. Before recombination, photons and electrons are tightly coupled via Thomson scattering: $e^- + \gamma \rightarrow e^- + \gamma$. However, when atoms start to form and matter becomes neutral, free electrons become scarce and Thomson scattering becomes inefficient. Therefore, the photons mean free path increases rapidly and they decouple from the rest of matter, forming a thermal bath of radiation that free streams and permeates the Universe: this is the Cosmic Microwave Background. In parallel, ordinary matter is now free from the influence of the radiation fluid and can start falling in the gravitational wells of Dark Matter that have already started to form under their own gravitational pull: structures start to form in the Universe.

10. $T = 2.6 - 7$ meV; $z = 11 - 30$; $t = 100 - 400$ Myr: Reionisation. The formation of the first stars lead to bursts of energetic radiation which gradually re-ionise the neutral hydrogen formed during recombination.

11. $T = 0.33$ meV; $z = 0.4$; $t = 9$ Gyr: Dark Energy-Matter equality. The cosmological constant starts to dominate the dynamics of the Universe. See below.

12. $T = 0.24$ meV; $z = 0$; $t = 13.8$ Gyr: Today.

6.4 The dark sector

In addition to the matter-energy content provided by the standard model of particle physics, the standard model of cosmology needs to introduce at least two new sources of the gravitational field to account for the behaviour of the Universe and objects inside it. Because these new sources are, to date, only felt through their gravitational interaction, and do not seem to interact significantly via electromagnetic interactions, they are called dark.

6.4.1 Dark Matter

The first dark component that one needs to introduce is an additional fluid of non-relativistic particles known as Dark Matter. The nature of Dark Matter has not yet been determined and this is a true

puzzle for fundamental physics. However, as we will see, it is clear that at cosmological/extragalactic scales, something peculiar happens that needs to be explained. The standard lore is to assume the presence of Dark Matter and to hope that its constituents will be identified at some point, be they fundamental particles, condensates of fundamental particles, or even small black holes formed in the primordial phases of the history of the Universe and remaining to this day. Alternatives consider that gravity and/or inertia itself is modified to account for the unexpected phenomena. Although these are puzzling and interesting possibilities, we will not explore them in this introductory course.

The first evidence for Dark Matter comes from the observations of distant spiral galaxies. The visible part of a spiral galaxy forms a thin disc of radius $R_d \sim$ a few kpc, with stars orbiting in quasi-circular orbits. Newtonian mechanics applied to the motion of these stars leads to a profile of velocity as a function of the distance to the centre of the galaxy r given by:

$$\frac{v^2(r)}{r} = \frac{GM\,(<r)}{r^2}\,, \tag{6.112}$$

where $M\,(<r)$ is the total mass contained within a shell of radius r. Thus, at distances $r \geq R_d$ beyond the size of the disc, if all the mass of the galaxy is contained into stars (and interstellar gas), $M\,(<r) \to M$ reaches a constant value, and the velocity profile should scale like:

$$v(r) \propto \frac{1}{\sqrt{r}}\,. \tag{6.113}$$

But observations do not support such a decrease. Instead, the velocity profile reaches a constant value $v_\infty \neq 0$ as r becomes large. This is illustrated for a specific galaxy on Fig. 6.8.

Such a profile requires the presence of additional matter beyond the observable disc of the galaxy, with a distribution of mass going as:

$$M(r) \propto r \text{ for } r \geq R_d\,, \tag{6.114}$$

which, for a spherically distributed halo, corresponds to an additional density of matter $\rho(r) \sim 1/r^2$. This is Dark Matter on galactic scales. The presence of such halos has also been confirmed by gravitational lensing of distant light by galaxies and clusters of galaxies. Finally, let us mention that Dark Matter is also needed on cosmological scales:

- BBN gives us a precise measurement of the ratio of baryonic matter to radiation in the Universe and the amount of radiation can be inferred from observation of the CMB. These facts

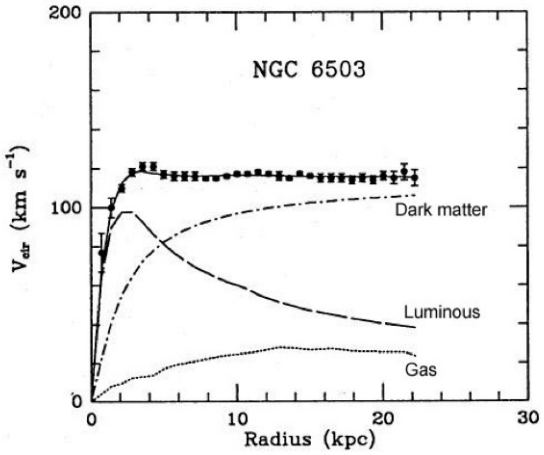


Figure 6.8: Rotation Velocity in the galaxy NGC6503, together with the respective contributions from diffuse gas, stars (labelled luminous), and the Dark Matter halo necessary to account for the observed profile. From [3].

in combination lead to a small energy density of baryons, too small to constitute the entire energy budget in non-relativistic particles.

- We will see that during the Matter Dominated era, small-scale matter overdensities grow like the scale factor: $\delta \propto a \propto 1/(1+z)$. However, baryons can only start to grow structure after they decouple from photons This means that, if the non-relativistic fluid only consisted of baryons, an overdensity of size 1 today should have been of size $\sim 10^{-3}$ at decoupling. This is 2 orders of magnitude larger than the overdensities in the photon-baryon plasma at decoupling inferred from the observations of the CMB. Thus structures have had to start forming earlier, in a fluid that did not feel the pressure waves of the plasma: a weakly interacting Dark Matter component does just that.

6.4.2 Late-time Universe: Λ

Dark Matter is thus required to explain the formation and behaviour of structure in the Universe. On the largest scales and latest times, on the other hand, another problem arises. Let us introduce

the *deceleration parameter*:

$$q_0 = -\frac{\ddot{a}}{aH^2}_{|t=t_0} . \tag{6.115}$$

Note that, neglecting radiation in the late Universe:

$$q_0 = \frac{1}{2}\Omega_{m,0} - \Omega_{\Lambda,0} . \tag{6.116}$$

We can then Taylor expand all quantities around the present time, e.g., at the relevant, dominant orders:

$$a(t) \simeq 1 + H_0 (t - t_0) - \frac{1}{2} q_0 H_0^2 (t - t_0)^2 \tag{6.117}$$

$$z(t) \simeq -H_0 (t - t_0) \tag{6.118}$$

$$E(z) \simeq 1 + (1 + q_0) z . \tag{6.119}$$

Thus, the luminosity distance of a distant object at small redshift behaves like:

$$D_L(z) \simeq H_0^{-1} \left(z + \frac{1 - q_0}{2} z^2 \right) . \tag{6.120}$$

It is possible to calibrate the luminosity curves of Type 1a Supernovæ and use them as standard candles, i.e. as distant objects whose intrinsic luminosity can be determined. Then, one can measure their apparent luminosity on Earth and determine their luminosity distance. By measuring their redshift, one can thus determine a distance-redshift relation $D_L(z)$ and constrain cosmology. Actually, the quantity that is usually being reported in the distance modulus:

$$\mu(z) - M = -2.5 \log \left[\frac{\phi(z)}{\phi\,(10\ \mathrm{pc})} \right] , \tag{6.121}$$

where $\phi(z)$ is the flux of a source located at redshift z and $\phi\,(10\ \mathrm{pc})$ the one of a source at 10 pc. The factor -2.5 is arbitrary and was chosen to match the definition of magnitude given by Hipparcos for stars. $\mu(z)$ is the apparent, measured, magnitude of the object, and M its absolute magnitude defined with respect to the magnitude of the Sun:

$$M = -2.5 \log \left(\frac{L}{3.8 \times 10^{26}\ \mathrm{W}} \right) + 4.75 . \tag{6.122}$$

Such observations have been performed with greater and greater accuracy since 1998, and consistently report $q_0 < 0$, i.e. a relation whose second derivative at the origin is larger than H_0^{-1}. But

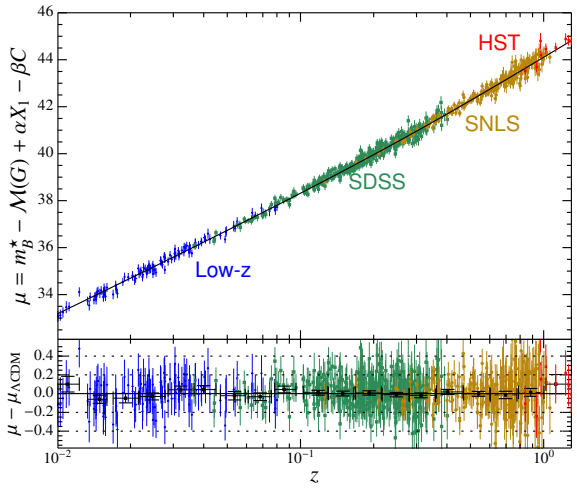


Figure 6.9: Distance modulus of distant Supernovæ 1a and residuals with respect to a flat FLRW Universe with Cold Dark Matter and $K = 0$. From [5].

this is only possible if $\Omega_{\Lambda,0} \neq 0$, in other words, if $\Lambda \neq 0$. Moreover, it means that the expansion of the Universe is currently accelerating: $\ddot{a} > 0$, a phenomenon that cannot emerge from any standard source of the gravitational field. Thus, if one were to assume $\Lambda = 0$, one would have to introduce some non-standard, exotic matter source (or modify gravity) to ensure $\ddot{a} > 0$; this is what is dubbed Dark Energy. So far, there is no evidence favouring an exotic Dark Energy over a simple cosmological constant so in what follows we will limit our discussion to this simple scenario. Fig. 6.9 summarises measurements of the distance-redshift relation from various recent projects. Note that cosmological evidence for the presence of a cosmological constant are now numerous and we do not only rely on these $m(z)$ diagrams.

6.5 Limits of the model: Inflation

The hot Big-Bang model we just described has been extraordinarily successful at explaining a wide range of observations, as well as at predicting some quantities that were measured later. By any measure, it is a very successful scientific model. However, it suffers from a few shortcomings that have to do with its initial state. The initial singularity is clearly a problem, but we are going to see that it is not just a mathematical one. Rather, it comes with some physical implications that

are quite puzzling and need to be overcome. This will be the role played by a phase in the history of the Universe taking place before the radiation dominated epoch and known as cosmic inflation. Let us stress immediately that although the principles of inflation and its overall phenomenology are very useful in solving the problems of the hot Big-Bang model, inflation as a model does not enjoy the same status as the rest of the cosmological model. In particular, it is not as well tested and constrained as the hot Big-Bang phase. There are essentially four problems with the standard Big-Bang model:

- The causality problem. In the hot Big-Bang, regions of spacetime that appear extremely similar to us did not have enough time to interact with each other. But then, why are they so similar?
- The flatness problem. In the standard, ΛCDM model, the Universe appears to be close to spatially flat today. In the Hot Big-Bang model, that means it must have started extremely flat at the Big-Bang. How can it be?
- The relic problem. At high energies, close to the initial singularity, phase transitions should have produced topological defects with very high densities. Why don't we see them around us?
- The origin of structures problem. How are the seeds for structure formation generated?

Inflation will somehow solve all these problems at once. In this section, we will highlight the problems of the standard model listed above and sketch how inflation solves the first three of them, that is, the ones which have to do with the background expansion history, rather than with structures. A somewhat more detailed treatment of inflation can be found in Chapter 5, in particular as far as the origin of structures is concerned (which will not be treated it).

6.5.1 The causality problem

Let us consider an observer O ('Us') today (at $\eta = \eta_0$), observing the Cosmic Microwave background emitted at η_{dec}. The situation is summarised on Fig. 6.10 in an (η, χ) diagram. The surface of last scattering for O[5] appears as a sphere of radius given by the comoving radial distance

[5]This is the surface obtained as the section of the space at time t_{dec} at which photons decouple from baryonic matter by the past lightcone of the observer O. Strictly speaking, decoupling is not instantaneous, and last-scattering for an observer is not quite a surface, but this does not modify the argument and we will ignore this subtlety.

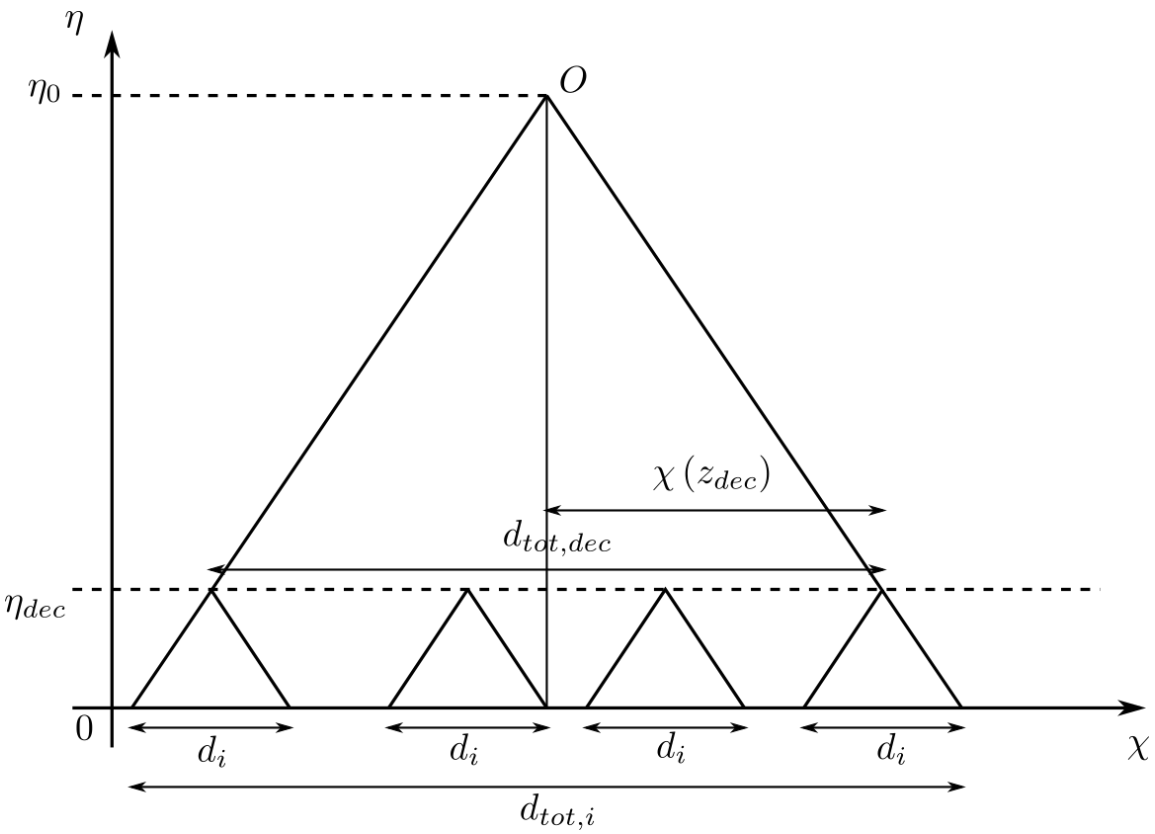


Figure 6.10: Spacetime diagram to illustrate the causality problem

$\chi(z_{dec}) = \int_0^{z_{dec}} dz'/H\left(z'\right)$. Thus, the diameter represented on the diagram is given by:

$$d_{tot,dec} = 2\int_0^{z_{dec}} \mathrm{d}z'/H\left(z'\right) \simeq 2 \times 1.93 H_0^{-1}\ , \tag{6.123}$$

where we used standard values for the cosmological parameters and we neglected the effect of the cosmological constant on the expansion history (the argument is not affected by this approximation). Let us now consider events at $\eta = \eta_{dec}$ located on or inside the past lightcone of O. The regions of space at the initial time (at the Big-Bang), $\eta = 0$, which have had time to influence these events at η_{dec} are balls at $\eta = 0$ with (comoving) diameters:

$$d_i = 2\int_{z_{dec}}^{+\infty} \frac{\mathrm{d}z'}{H\left(z'\right)} \simeq 2 \times 4 \cdot 10^{-2} H_0^{-1}\ . \tag{6.124}$$

On the other hand, the intersection of the past lightcone of O with the initial space slice at $\eta = 0$, which gives the set of all the points that actually influenced the events on or inside the last scattering

surface seen by O, delimits a ball of (comoving) diameter:

$$d_{tot,i} = 2\int_0^{+\infty} \frac{\mathrm{d}z'}{H(z')} \simeq 2 \times 1.98 H_0^{-1} \ . \tag{6.125}$$

Therefore, the number of disconnected regions at the Big Bang, each able to influence a different point on or inside the last scattering surface is roughly given by:

$$N \simeq \left(\frac{d_{tot,i}}{d_i}\right)^3 \simeq 10^5 \ . \tag{6.126}$$

The corresponding points at η_{dec} have not had time to interact in any causal way but if we live in an almost FLRW Universe, they ought to have almost the same temperature, as seen in the CMB temperature anisotropies which are of the order of 10^{-5}. Unless the initial conditions at $\eta = 0$ were set extremely precisely (fine-tuned) to ensure this coincidence at η_{dec}, this is not possible.

One might be worried that this argument depends on the Copernican principle, since we talk about events located inside our past lightcone at last scattering, so events that we do not observe. We can turn things around and examine what happens on the last scattering surface only. Consider now an event located at $\eta = 0$. The intersection of the inside of its future lightcone with the hypersurface at last scattering will be a ball of proper diameter:

$$D_i = \frac{1}{1+z_{dec}} d_i \simeq 2 \times 4 \times 10^{-5} H_0^{-1} \ . \tag{6.127}$$

If it intersects the last scattering surface, it does so on a patch with typical size D_i. On the other hand, the distance form 0 to the last scattering surface is given by:

$$D = \frac{1}{1+z_{dec}} d_{tot,dec} \simeq 2 \times 10^{-3} H_0^{-1} \ . \tag{6.128}$$

This means that the angular size, as seen from 0, of a patch of the last scattering surface that has been influenced by an event at the Big-Bang is given by:

$$\Delta\theta \simeq \frac{D_i}{D} \simeq 2 \times 10^{-2} \sim 1^o \ . \tag{6.129}$$

The number of such disconnected patches on the CMB sky is roughly given by the ratio of the solid angles:

$$N' \simeq \frac{\Delta\theta^2}{4\pi} \simeq 10^4 \ . \tag{6.130}$$

All these patches have not had time to thermalise by causal contact and yet, they exhibits remarkably similar properties on the sky observed by 0. How is this possible?

6.5.2 The flatness problem

In a hot Big-Bang scenario, still neglecting the effects of Λ for simplicity, we can write the evolution of the curvature parameter as:

$$\Omega_K(z) = \frac{\Omega_{K,0}}{\Omega_{m,0}(1+z) + \Omega_{r,0}(1+z)^2} \ . \tag{6.131}$$

The problem is that this function is decreasing: since we observe a small curvature parameter today, typically $\left|\Omega_{K,0}\right| < 10^{-2}$, the effect of curvature needs to have been even smaller in the past. In the early Universe, close to the Big-Bang:

$$\Omega_K(z) \sim \frac{\Omega_{K,0}}{\Omega_{r,0}}(1+z)^{-2} \text{ when } z \to +\infty \ . \tag{6.132}$$

Thus, using $\Omega_{r,0} \sim 10^{-5}$:

$$\left|\Omega_{K,i}\right| < 10^3 \ (1+z_i)^{-2} \ . \tag{6.133}$$

At BBN, this bound is of order 10^{-7} and it reaches 10^{-61} at the Planck time. Therefore, the Universe needs to start in an extremely flat configuration in order to get a very flat Universe today. Of course, this is only a problem if one considers that this is an unnatural initial state; in absence of a measure giving us the likelihood of a given curvature, this is impossible to assess. Therefore, this problem with the hot Big-Bang is of a different nature than the causality problem. Whereas the latter is really linked to a physical difficulty, the former is only a problem as far as "taste" for "natural" initial conditions is concerned.

6.5.3 The relic problem

As we have seen, as the Universe cools down, some phase transitions occur when fundamental symmetries are broken. If Grand Unified scenarii are correct, when the Grand Unification theory breaks down, at the very early stages of the Radiation Dominated epoch, some topological defects such as monopoles are created. These carry a very large amount of energy density that, if present, would completely dominate the expansion of the Universe and change the expansion history that we know. So, why are these topological defects not around and dominating the expansion of the Universe?

6.5.4 Origin of structure

Finally, as we mentioned before, we need to find a way to generate density fluctuations in the early Universe that are large enough to give rise to the structures we observe via gravitational infall. Moreover, because of the behaviour of the Hubble radius, we know that, in a Universe with only a matter dominated and a radiation dominated eras, all physical scales on which we observe fluctuations in the matter distribution today will eventually exit the Hubble radius if we trace them backward in time far enough. This means that these fluctuations cannot have been generated causally in the Hot Big-Bang model (because the Hubble radius fixes approximately the scale below which causal processes are efficient in the Universe; see below). How is this possible?

6.5.5 The idea of inflation

Let us get back to the comoving distance between a point at an initial time t_i for the expansion of the Universe and a point at time = t further in the future:

$$\chi(t) = \int_{t_i}^{t} \frac{\mathrm{d}t'}{a\,(t')} = \int_{\ln a_i}^{\ln a} \mathcal{H}^{-1}\,(a')\,\mathrm{d}\ln a' \,, \tag{6.134}$$

where we have written $a_i = a(t_i)$. Note that here, since we want to replace the Big-Bang by something else, we do not yet assume that $a_i = 0$. For a perfect fluid with $w = cst$, the comoving Hubble scale $\mathcal{H}^{-1} = (aH)^{-1}$ behaves as:

$$\mathcal{H}^{-1} = (aH)^{-1} \propto a^{(1+3w)/2} \,. \tag{6.135}$$

Thus, for standard matter, with $1 + 3w > 0$, this scale increases with the expansion of the Universe. But this means that the integral in Eq. (6.134) is dominated by its upper limit and receives a vanishing contribution from the early times. Indeed, performing the integral (and using the fact that we are tracing lightrays, so that $d\chi = -d\eta$), we get:

$$\chi(a) = \eta - \eta_i \propto a^{(1+3w)/2} - a_i^{(1+3w)/2} \,, \tag{6.136}$$

with $\eta_i \propto a_i^{(1+3w)/2}$. Note that $\chi(a)$ is always finite and that $\eta_i \to 0$ when $a_i \to 0$, i.e. in case of a Big-Bang singularity. But what happens if, at early times, i.e. before the radiation dominated era, there is an era with $1 + 3w < 0$? In that case, we have that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}^{-1} \propto \frac{1+3w}{2} a^{(3w-1)/2}\mathcal{H} < 0 \,. \tag{6.137}$$

Therefore, the comoving Hubble scale $\mathcal{H}^{-1}$ now decreases as a increases. But this means that, in that case, the integral in Eq. (6.134) is dominated by its lower bound, and that the Big-bang sigularity gets pushed to negative values of the conformal time:

$$\eta_i \propto \frac{2}{1+3w} a_i^{(1+3w)/2} \rightarrow -\infty \text{ when } a_i \rightarrow 0 \, . \tag{6.138}$$

In principle, by choosing this early phase to be arbitrarily long, one can push the Big-Bang singularity arbitrarily far into the past, thus asymptotically ridding the cosmological model of the Big-Bang singularity. This means one has "much more conformal time available" between the singularity and decoupling, allowing for regions to interact causally. The comoving distance between the Big-Bang and decoupling can now be made arbitrarily large. This early phase during which $\mathcal{H}^{-1}$ is a decreasing function of time is known as inflation, since:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}^{-1} = -\frac{\ddot{a}}{\dot{a}^2} < 0 \Rightarrow \ddot{a} > 0 \, , \tag{6.139}$$

meaning that the expansion is actually accelerating. The behaviour of causally connected regions in a Universe with an early inflationary phase is presented in Fig. 6.11, to be contrasted with what we saw in a standard Big-Bang model. Fig. 6.12 also presents the behaviour of the comoving Hubble scale and of physical scales in such a Universe. Note that during inflation, scales that were initially sub-Hubble are expelled for the comoving Hubble scale and only re-enter later, during the standard hot-Big-Bang phase, either when radiation or matter dominate the expansion. That will explain why structures that are sub-Hubble today but were super-Hubble in the past actually formed causally: they were actually sub-Hubble in an even more distant past, during inflation. How much inflation do we need to solve the causality problem? At the very least, we need the observable Universe today to fit into the comoving Hubble radius at the beginning of inflation. This will ensure that all the points in our Hubble volume today will have been in causal contact at some point during inflation, before separating out later. Note that this is much more conservative than using the comoving distance to decoupling (which is what we really need to ensure causality), because we have $\chi\left(z_{dec}\right) > H_0^{-1}$ so if our condition is satisfied, so is the condition on the comoving size of the last scattering surface. Calculations are simpler this way. Our condition corresponds to (keeping $a_0 = 1$ for symmetry in the expressions):

$$\left(a_0 H_0\right)^{-1} < \left(a_I H_I\right)^{-1} \, . \tag{6.140}$$

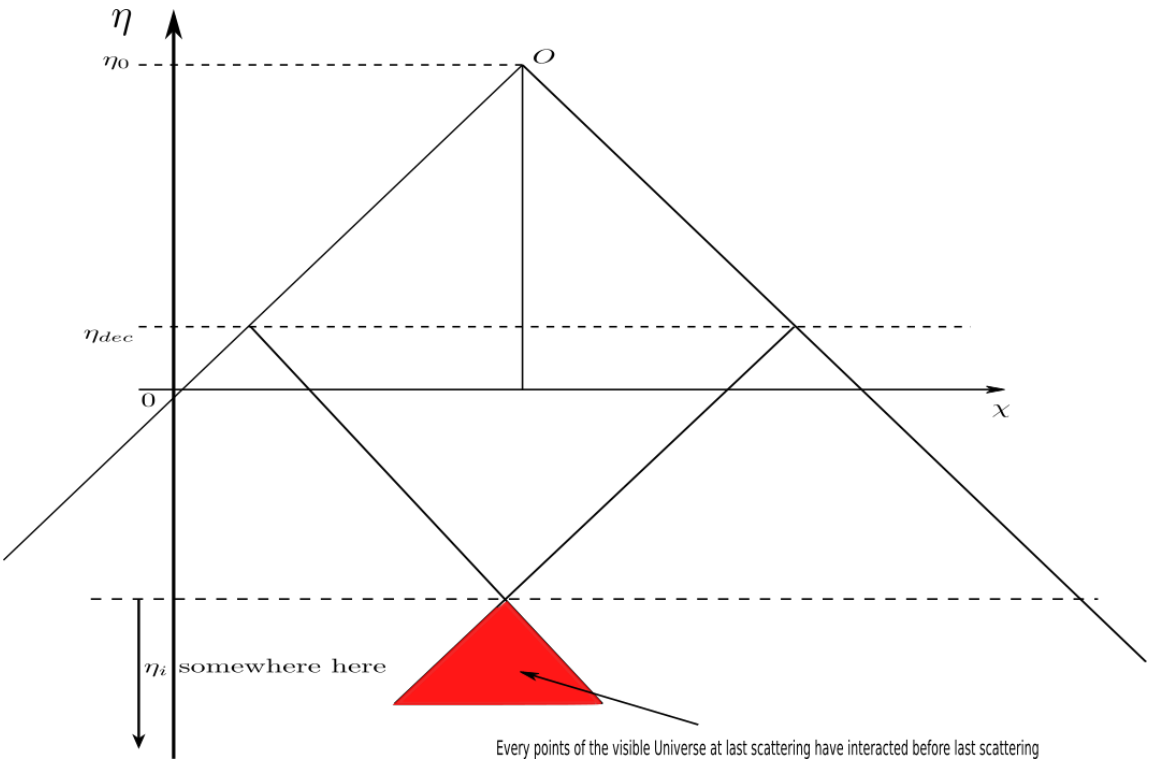


Figure 6.11: How the causality problem is resolved by an early phase of inflation. In the red region, two antipodal points on the last scattering surface which would have been totally causally disconnect in the standard Big-Bang scenario, can now have interacted in their past, thus thermalising by physical process and sharing a nearly equal temperature, as observed. The choice of η_i must be made such that at least antipodal points have interacted; this ensures that other points on the last scattering surface will also have had time to interact.

Now, neglecting the matter dominated and Λ dominated phases (which lower the comoving Hubble radius compared to keeping only radiation, so our bound is stronger here), we get:

$$\frac{a_0 H_0}{a_E H_E} \simeq \frac{a_0}{a_E}\left(\frac{a_E}{a_0}\right)^2 = \frac{a_E}{a_0} = \frac{T_0}{T_E} \ . \tag{6.141}$$

Assuming that the end of inflation is around the Grand Unified Theory scale (which ensures that the monopoles get diluted by inflation and thus also solves the relic problem), so that $T_E \sim 10^{15} - 10^{16}$ GeV, we find that:

$$(a_I H_I)^{-1} > 10^{28}\,(a_E H_E)^{-1} \ , \tag{6.142}$$

thus, the comoving Hubble radius must shrink by 28 orders of magnitude during inflation. For an almost constant Hubble rate, this implies that the number of e-folds must be:

$$N \equiv \ln\left(\frac{a_E}{a_I}\right) > 64 \ . \tag{6.143}$$

Note that in terms of physical distance, this corresponds to a physical Hubble radius H^{-1} increasing dramatically. Such a huge amount of inflation, in addition to solving the causality problem and the monopole problem (because the volume increases so much that the density of monopoles, if they exist, decreases dramatically), also addresses the flatness problem. This is because during inflation, the parameter $\Omega_K(a)$ actually decreases dramatically. Hence any curvature present at the beginning of inflation would have been wiped out by a factor 10^{-56}:

$$\frac{\Omega_K\,(a_E)}{\Omega_K\,(a_I)} = \left(\frac{a_I H_I}{a_E H_E}\right)^2 < \left(10^{-28}\right)^2 = 10^{-56} \ . \tag{6.144}$$

The physical volume of the Universe increases so much during inflation that the curvature becomes very small.

6.6 A concordance model

The FLRW Universe with $\Lambda \neq 0$, some Cold Dark Matter, and flat spatial sections ($K = 0$) is called the *concordance model of cosmology*. In addition to the parameters of the standard model of particle physics (that are considered determined and fixed in the concordance model), it contains a certain number of free parameters that need to be determined by observations or principles. The 6 cosmological parameters that are left free and to be determined in the concordance model are usually:

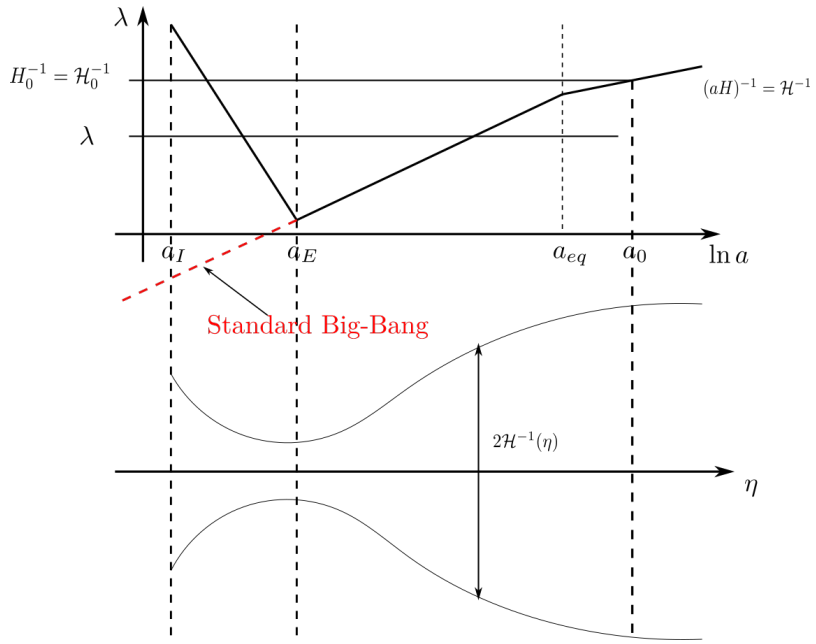


Figure 6.12: Upper part: Behaviour of comoving scales in an inflationary Universe. Inflation starts at a_I and ends at a_E, after which the standard hot Big-Bang expansion starts: rafiation dominated era followed by a matter dominated era (the effects of the cosmological constants are ignored for illustrative purposes here). Comoving scales $\lambda < a_I H_I$ start sub-Hubble and are expelled from the Hubble sphere during inflation. They only re-enter the Hubble radius during the Hot Big-Bang phase. Lower part: Qualitative behaviour of a section of the comoving Hubble sphere. The transition between inflation and the radiation dominated phase is called reheating and is yet poorly understood.

1. the physical baryon density: $\Omega_{b,0}h^2$, where $h = H_0/(100\ \text{km/s/Mpc})$;
2. the physical CDM density: $\Omega_{c,0}h^2 = \left(\Omega_{m,0} - \Omega_{b,0}\right) h^2$;
3. the age of the Universe: t_0;
4. the optical depth of reionisation τ;
5. the scalar spectral index n_s (a parameter of inflation; see below);
6. the amplitude of initial curvature perturbations Δ^2_ζ (a parameter of inflation; see below).

The cosmological parameters that are fixed by default in the concordance model are:

1. the curvature parameter: $K = 0$;
2. the tensor to scalar ratio: $r = 0$ (a parameter of inflation; see below);
3. the running of the spectral index: $\frac{dn_s}{d\ln k} = 0$;
4. the sum of the masses of neutrinos: $\sum m_\nu = 0.06\ \text{eV/c}^2$;
5. the effective number of relativistic degrees of freedom: $N_{eff} = 3.046$.

All other parameters can be determined by calculations. Today, due to the not so small number of these parameters, and to intrinsic degeneracies between them in observables, the most precise determination of these parameters, or any different combination of those and extra parameters that one may want to leave free, comes from combining constraints that can be inferred from different observations, e.g., CMB anisotropies, supernovæ 1a, Baryon Acoustic Oscillations, Weak lensing shear surveys, BBN, galaxy number counts etc. This is why the model is called concordant: it provides the minimal, "simplest" model that can account for most (if not all) of the current observations available on our Universe. Over the last decade, as observations became more and more precise, some tensions started to appear in this concordance model. Careful scrutiny and more and more precise observations have not led to any resolution of these tensions but it remains unclear whether or not such issues can be attributed to new physics, beyond the minimal ΛCDM model, to systematic biases due to our inability to accurately model non-linear physics on multiple scales to fit the model to observations, or to observational errors. As a matter of fact, there is not a single model that can currently account for all these tensions at once at still pass with success all the other tests

that ΛCDM passed. Therefore, for pedagogical purposes, we can concentrate on this model. Deviations from it are small and, although they might prove very important from a conceptual level, they will most likely not alter the big picture significantly. The interested students will find an extensive review of these recent issues in [2].

We will use the following nominal values for background cosmological parameters, unless otherwise stated:

Nominal background parameters

$$\Omega_{K,0} = 0 \tag{6.145}$$

$$\Omega_{m,0} = 0.32 \tag{6.146}$$

$$\Omega_{b,0} = 0.05 \tag{6.147}$$

$$\Omega_{\Lambda,0} = 0.68 \tag{6.148}$$

$$\Omega_{r,0} = 10^{-4} \tag{6.149}$$

$$H_0 = 67 \text{ km/s/Mpc.} \tag{6.150}$$

Appendices

A

Mathematical preliminaries

Contents

This appendix sums up some preliminary notions that must already be known by students in one context or another. We are merely presenting them to refresh their memories and lay down some notations. It can be omitted by readers with a physicist's mind.
Proofs are omitted and students wishing to access some of them are encouraged to use mathematics textbooks.

A.1 Maps

Map

Consider two (abstract) sets X and Y. A *map* (or *mapping*) f between X and Y is a rule that assigns an element $y \in Y$ to each element $x \in X$. This is denoted:

$$f : X \to Y . \tag{A.1}$$

The action of the map f on elements of X is summarised as:

$$f : x \mapsto y . \tag{A.2}$$

Note that the same $y \in Y$ may correspond to more that one element of X via the map f. A subset of X whose elements are mapped to $y \in Y$ is called the *inverse image* of y by f, and is denoted $f^{-1}(y) = \{x \in X,\ f(x) = y\}$. The set X is called the *domain* of the map f, while Y is called the *range* of f. The *image* of the map is the set:

$$f(X) = \{y \in Y,\ \exists x \in X,\ y = f(x)\} \subseteq Y . \tag{A.3}$$

It is important to realise that the domain and the range are an integral part of the definition of a map. Consider, for example $f : x \mapsto \exp(x)$. If $X = Y = \mathbb{R}$, then -1 has no inverse image, whereas, if $X = Y = \mathbb{C}$, we have, $f^{-1}(-1) = \{(2k+1)\pi\mathrm{i},\ k \in \mathbb{Z}\}$.

Type of maps

Consider a map $f : X \to Y$.

- f is called *injective* iff $\forall (x, x') \in X^2,\ x \neq x' \implies f(x) \neq f(x')$.
- f is called *surjective* iff $\forall y \in Y,\ \exists x \in X,\ f(x) = y$.

- f is called *bijective* iff it is both injective and surjective. This translates into: $\forall y \in Y,\ \exists! x \in X,\ f(x) = y$.

Given a map $f : X \to Y$, and $A \subset X$, we can define the *restriction* of f to A, denoted $f_{|A} : A \to Y$, by $f_{|A}(a) = f(a)$ for any $a \in A \subset X$.
Given three sets X, Y and Z, and two maps $f : X \to Y$ and $g : Y \to Z$, we can define the *composition* of f and g, denoted $g \circ f : X \to Z$ or $gf : X \to Z$, by:

$$\forall x \in X,\ (g \circ f)(x) = g(f(x)) \ . \tag{A.4}$$

Now, let us consider two sets X and Y, and suppose that some algebraic structures (i.e. operations, e.g. additions, products etc.) are given on X and Y. A mapping $f : X \to Y$ that preserves these algebraic structures is called an *homomorphism*. For example, suppose that both X and Y are endowed with a product. Then, if, $\forall (x, x') \in X^2,\ f(xx') = f(x)f(x')$, f is an homomorphism.
An important class of homomorphisms is the one of group homomorphisms: suppose that X and Y are two groups with operations $*$ and $+$ respectively (be careful, these are not necessarily a multiplication and an addition), then, if $f : X \to Y$ is an homomorphism ($f(a * b) = f(a) + f(b)$), it is called a *group homomorphism*; in essence, it preserves the group structure. An easy example of group homomorphism is given by the exponential map $\exp : \mathbb{R} \to \mathbb{R}$, for which: $\forall (a, b) \in (\mathbb{R}, +)$, $\exp(a + b) = \exp(a).\exp(b)$; therefore, it is a group homomorphism between the group of real number with addition and the group of positive real numbers with multiplication.

If an homomorphism $f : X \to Y$ is bijective, it is called an *isomorphism*, and X and Y are said to be *isomorphic*. This is denoted $X \cong Y$.

A.2 Vector spaces and linear algebra

A.2.1 Vector spaces

Vector space

A vector space V over the field of numbers K ($\mathbb{R}$ or $\mathbb{C}$ for example) is a set of elements v, called vectors of V, with two operations:

- an *addition*, denoted $+$, which, to any pair $(v, w) \in V^2$ associates a third element $u \in V$

written $u = v + w$,

- and a *scalar multiplication*, which, to any elements $v \in V$ and $a \in K$, assigns an element $w \in V$ such that: $w = av$.

Moreover, the addition is supposed to have a *neutral element* $0 \in V$ such that the following axioms are satisfied:

- $\forall (v, w) \in V^2,\ v + w = w + v$ (Commutativity of +);
- $\forall (u, v, w) \in V^3,\ (u + v) + w = u + (v + w)$ (Associativity of +);
- $\forall v \in V,\ v + 0 = v$ (0 is the neutral element for the addition);
- $\forall v \in V,\ \exists ! - v,\ v + (-v) = 0$ (Existence of an inverse for the addition);
- $\forall a \in K,\ \forall (v, w) \in V^2,\ a(v + w) = av + aw$ (Distributivity of the multiplication with respect to the addition);
- $\forall (a, b) \in K^2,\ \forall v \in V,\ (a + b)v = av + bv$ (Distributivity of the scalar multiplication with respect to the field addition);
- $\forall v \in V,\ 1v = v$ (1 is the neutral element for the scalar multiplication).

Such a vector space will be denoted $(V, K, +, .)$. Usually, the numbers $a \in K$ are called *scalars*.

As far as this course is concerned, *we will limit our study to vector spaces of finite dimensions*. The typical example of a vector space is $\mathbb{R}^n = \left\{v = (x^1, x^2, ..., x^n),\ \forall p \in \{1, 2, ..., n\},\ x^p \in \mathbb{R}\right\}$, considered as the set of ordered lists of n elements of $\mathbb{R}$, together with the field $\mathbb{R}$ and the addition and scalar multiplication given by:

- $\forall v = (a^1, a^2, ..., a^n) \in \mathbb{R}^n,\ \forall w = (b^1, b^2, ..., b^n) \in \mathbb{R}^n,\ v + w = (a^1 + b^1, a^2 + b^2, ..., a^n + b^n)$;
- $\forall v = (a^1, a^2, ..., a^n) \in \mathbb{R}^n,\ \forall c \in \mathbb{R},\ cv = (ca^1, ca^2, ..., ca^n)$.

In particular, in the case $n = 2$, $(\mathbb{R}^2, \mathbb{R}, +, .)$ defined as above can be simply identified with the Cartesian plane, and the vectors of $\mathbb{R}^2$ are the usual vectors of plane geometry. In the same way, $\mathbb{R}^3$ can be viewed as the vector space of vectors in three dimensions.

It is obvious that the same properties apply to $\mathbb{C}^n$. In particular, $\mathbb{R}$ and $\mathbb{C}$ are vector spaces, hence, the field K on which our vector spaces are defined can always be seen as a vector space. An important concept in the study of vector spaces is the existence of linear relations between vectors.

Linear combinations

Let V be a vector space on the field K. Let $r \in \mathbb{N}^*$. Let $\forall i \in \{1, 2, ..., r\}$, $v_i \in V$ and $\forall i \in \{1, 2, ..., r\}$, $a_i \in K$. The vector of V defined by $w = \sum_{i=1}^{r} a_i v_i$ is said to be a *linear combination* of the vectors $v_1, v_2, ..., v_r$.
The set $(v_1, v_2, ..., v_r)$ is said to be *linearly independent* if and only if:

$$\sum_{i=1}^{r} a_i v_i = 0 \implies \forall i \in \{1, 2, ..., r\},\ a_i = 0. \tag{A.5}$$

If this is not the case, the set $(v_1, v_2, ..., v_r)$ is said to be linearly dependent.

For example, consider the vector space $\mathbb{R}^2$ over the field $\mathbb{R}$. Consider the vectors $(0, 1)$, $(1, 0)$ and $(0, 2)$. The neutral element for the addition in $\mathbb{R}^2$ is $0 = (0, 0)$. We clearly have $\forall (a, b) \in \mathbb{R}^2$, $a(0, 1) + b(1, 0) = (b, a)$. Hence, $\forall (a, b) \in \mathbb{R}^2$, $a(0, 1) + b(1, 0) = 0 \implies (a = 0 \text{ and } b = 0)$. So, the set $S = \{(0, 1), (1, 0)\}$ is linearly independent. In the same way, the set $\{(1, 0), (0, 2)\}$ is also linearly independent. On the contrary $\{(0, 1), (0, 2)\}$ is not linearly independent, since $-2(0, 1) + (0, 2) = 0$.

As a simple consequence, we see that two vectors are linearly dependent iff one is a multiple of the other. The maximum number of linearly independent vectors in a vector space V is called the *dimension* of V over the field K, and is often denoted $\dim_K(V)$, or $\dim V$ if there is no ambiguity on the field K. We can recall the definition of a basis of V.

Basis

Let V be a vector space on the field K. A subset $S \subseteq V$ is basis of V iff:

- S is linearly independent;
- Every element of V is a linear combination of elements of S.

Moreover, for any $v \in V$, the linear combination expressing v in terms of the elements of S

is unique, up to the ordering of the terms. The unique scalars occurring as coefficients in the linear combination are called the components of v with respect to the basis S.

As a by-product, we can then state that a subset $S \subseteq V$ is a basis of V iff every element of V can be written uniquely as a linear combination of elements of S. Moreover, any two bases of a vector space V have the same number of elements, equal to the dimension of V.

As an example, in $\mathbb{R}^3$, the set $\{(1,0,0),(0,1,0),(0,0,1)\}$ is obviously a basis. It is usually called a Cartesian basis, since one can construct from it a set of Cartesian coordinates (x,y,z), such that any vector v in $\mathbb{R}^3$ can be written uniquely as $v = x(1,0,0) + y(0,1,0) + z(0,0,1)$. Of course, there exist other standard basis of $\mathbb{R}^3$. For example, $\{(1,0,0),(0,1,1),(0,0,1)\}$ also forms a basis. The dimension of $\mathbb{R}^3$ is thus 3, as expected.
Another example is the field K on which the vector space is defined. In our case K is $\mathbb{R}$ or $\mathbb{C}$. Both are vector spaces of dimension 1 on themselves ($\mathbb{C}$ is a vector field of dimension 2 on $\mathbb{R}$).

We call $W \subseteq V$ a *subspace* of the vector space V iff W is itself a vector space with its structure inherited from the one of V. For example $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$. We have the obvious result that $\dim W \leq \dim V$ if W is a subspace of V.

A.2.2 Linear maps; Matrices

Linear map

Let V and W be two vector spaces over the field K, and let $f : V \to W$ be a map. f is a *linear map* (or linear mapping) of V into W iff

$$\forall (v_1, v_2) \in V^2,\ \forall (a,b) \in K^2,\ f(av_1 + bv_2) = af(v_1) + bf(v_2)\ . \tag{A.6}$$

A linear map is said to be an isomorphism if it is 1-1 onto (injective and surjective). Two vector spaces V and W that are linked by an isomorphism are said to be isomorphic, and we write $V \simeq W$. Linear functions are supposed to be well-known. For completeness, the reader may refer to standard textbooks of linear algebra such as *Fundamentals of Linear Algebra*, K. Nomizu, Mc Graw-Hill eds.

Ex: An important property of linear maps

Prove that if f is a linear mapping of V into W, we have:

$$f(0_V) = 0_W \,. \tag{A.7}$$

Here, we will just cite some results that will be important in the text:

Vector space of linear maps

Let V and W be two vector spaces. The set of all linear functions of V into W, denoted $L(V,W)$ forms a vector space, if we define the sum and the scalar products as follow:

$$\forall v \in V,\ \forall (f,g) \in L(V,W),\ (f+g)(v) = f(v) + g(v) \tag{A.8}$$

$$\forall v \in V,\ \forall a \in K,\ (af)(v) = af(v)\,. \tag{A.9}$$

Now, we would like to see how to express linear functions when basis have been chosen in V and W. Let us suppose that $\dim V = n$ and $\dim W = m$, and let us call $(e_i)_{i\in\{1,2,\ldots,n\}}$ and $(w_i)_{i\in\{1,2,\ldots,m\}}$ the chosen basis in V and W, respectively. At this point, it will be convenient to introduce a summation convention called Einstein summation convention, in order to avoid the constant occurrence of summation symbols. A vector $v \in V$ can be decomposed on the basis $(e_i)_{i\in\{1,2,\ldots,n\}}$ as:

$$v = \sum_{i=1}^{n} v^i e_i \,, \tag{A.10}$$

where the v^i's are the component of the vector in the given basis. We will define the summation convention to be:

$$\forall v \in V,\ v = v^i e_i \,, \tag{A.11}$$

where a repeated index up and down means that a summation has to be performed over all the possible values of the index:

$$v^i e_i = \sum_{i=1}^{n} v^i e_i . \tag{A.12}$$

The same applies of course in W. The effect of a linear function on the basis of V can then be summarized as follows:

$$\forall i \in \{1,2,\ldots,n\},\ f(e_i) = F^j{}_i w_j \,, \tag{A.13}$$

where the ${F^j}_i$ are $m \times n$ coefficients. Indeed, for any $i \in \{1, 2, .., n\}$, $f(e_i)$ is an element of W, and, as such, can be decomposed on the basis $(w_i)_{i\in\{1,2,...,m\}}$ of W. Now, consider a vector $v \in V$. Then, we can write:

$$f(v) = f\left(v^i e_i\right) = v^i f(e_i) = v^i {F^j}_i w_j = \sum_{i=1}^{n}\sum_{j=1}^{m} {F^j}_i w_j v^i \ . \tag{A.14}$$

The scalars ${F^j}_i$ are then the components of the matrix associated with f in the basis $(e_i)_{i\in\{1,2,...,n\}}$ and $\left(w_j\right)_{j\in\{1,2,...,m\}}$. Hence, in the same way as we think of the components of a vector in a given basis as the coordinates of this vector, we can say that the entries of a matrix are the coordinates of the associated linear function in the two basis chosen for the two vector spaces linked by the linear function. The upper index counts the rows, and the lower index counts the columns.

Consider for example the vector space $\mathbb{R}^3$ with the basis $\boldsymbol{e} = \{(1,0,0),(0,1,0),(0,0,1)\}$, For any $v \in \mathbb{R}^3$, we write $v = v^i e_i$. Consider the map:

$$f : \left\{ \begin{array}{ccc} \mathbb{R}^3 & \to & \mathbb{R}^3 \\ v & \mapsto & \left(2v^1 + v^2\right) e_1 + \left(v^3 - v^2\right) e_2 + \left(3v^1 - v^3\right) e_3 \end{array} \right. \tag{A.15}$$

This function is trivially linear. Now, one has:

$$f(e_1) = 2e_1 + 3e_3 \tag{A.16}$$

$$f(e_2) = e_1 - e_2 \tag{A.17}$$

$$f(e_3) = e_2 - e_3 \ . \tag{A.18}$$

Hence, because $\forall i \in \{1,2,3\}$, $f(e_i) = {F^j}_i e_j$, in the basis **e** (on both sides), the matrix associated to f in this basis is:

$$F = \begin{pmatrix} 2 & 1 & 0 \\ 0 & -1 & 1 \\ 3 & 0 & -1 \end{pmatrix} . \tag{A.19}$$

Hence, one can see that, to determine a linear function, one needs $n \times m$ scalars. This suggests that the vector space $L(V,W)$ has dimension $n \times m$. This is indeed true, as summarized in the following proposition.

Space of linear maps: basis

- $\dim L(V,W) = \dim V \times \dim W$;
- Let $(e_i)_{i\in\{1,2,\ldots,n\}}$ be a basis of V and $(w_j)_{j\in\{1,2,\ldots,m\}}$ be a basis of W. Then, a basis of $L(V,W)$ is given by the linear functions $E^j{}_i$ such that: $E^i{}_j(e_k) = \delta^i_k w_j$, where δ^i_k is the Kronecker symbol, equal to 1 if $i = k$ and to 0 otherwise.
- If $\left(F^j{}_i\right)_{i\in\{1,2,\ldots,n\};j\in\{1,2,\ldots,m\}}$ is the matrix associated to $f \in L(V,W)$ in the basis $\mathbf{e}$ and $\mathbf{w}$, then, the expression for f in terms of the basis $\{E^i{}_j\}$ of $L(V,W)$ is:

$$f = F^j{}_i E^i{}_j \ . \tag{A.20}$$

Here is an important result about linear functions. Let V and W be two vector spaces on a field K.

Rank-nullity theorem

Let $f \in L(V,W)$. Then:

$$\dim V = \dim \mathrm{Im} f + \dim \mathrm{Ker} f, \tag{A.21}$$

where $\mathrm{Im} f = \{w \in W,\ \exists v \in V,\ w = f(v)\}$ is the image of f, and $\mathrm{Ker} f = \{v \in V,\ f(v) = 0\}$.

A.2.3 Inner product

On a vector space V over the field K ($\mathbb{R}$ or $\mathbb{C}$), one can define a new operation known as an inner product.

Inner product

Let V be a vector space over K ($\mathbb{R}$ or $\mathbb{C}$). An *inner product* on V is a function $\langle .,. \rangle : V \times V \to K$ such that:

(i) Conjugate symmetry: $\forall (x,y) \in V^2,\ \langle x,y\rangle = \overline{\langle y,x\rangle}$;

(ii) Linearity: $\forall (x,y,z) \in V^3,\ \forall a \in K,\ \langle ax+y,z\rangle = a\langle x,z\rangle + \langle y,z\rangle$;

(iii) Non degenerate: $\forall x \in V,\ \langle x,y\rangle = 0 \implies y = 0$.

Note that in the case $K = \mathbb{R}$, the conjugate symmetry becomes a simple symmetry, and the inner product is then a bilinear form (*cf* section on tensor algebra). Also, in that case, if we have $\forall x \in V,\ x \neq 0,\ \langle x, x\rangle > 0$ (*resp.* $\forall x \in V,\ x \neq 0,\ \langle x, x\rangle < 0$), the inner product is said to be positive (*resp.* negative) definite.

A very important type of inner product for us is the standard inner product on $\mathbb{R}^n$, usually dubbed the dot product, or Euclidean inner product. If $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n)$ are two arbitrary vectors of $\mathbb{R}^n$, the dot product is defined by:

$$x \cdot y = \sum_{j=1}^{n} x_i y_j \ . \tag{A.22}$$

Check that it verifies all the conditions to be an inner product. Check that this is a positive definite inner product. This dot product in turn is used to define a *norm* for the vectors of $\mathbb{R}^n$, called the Euclidean norm:

$$\forall x \in \mathbb{R}^n,\ \|x\|_E = \sqrt{x \cdot x} \ . \tag{A.23}$$

This norm tells us how 'long' is a given vector[1].

Inner products are a particular case of bilinear functions that we will study in more details in what follows. It allows one to define the notion of *orthogonality*.

Orthogonality

Let V be a vector space of dimension n equipped with an inner product $\boldsymbol{g} = \langle ., . \rangle$. Two vectors $u \in V$ and $v \in V$ are *g-orthogonal*, or simply *orthogonal* if there is no possible confusion on the inner product, iff:

$$\langle u, v\rangle = 0 \ . \tag{A.24}$$

A basis of V, say $\{v_1, ..., v_n\}$ is *g-orthogonal*, or *orthogonal* if there is no possible confusion on the inner product considered, iff:

$$\forall (i, j) \in \{1, ..., n\}^2,\ i \neq j \ \Rightarrow \ \langle v_i, v_j \rangle = 0 \ . \tag{A.25}$$

[1]Note that in these notes, we do not make the difference between the Euclidean space E^n which is the space in which one can do geometry with lines, planes, circles, triangles etc., and $\mathbb{R}^n$ equipped with the dot product. These two structures are actually different but in one-to-one correspondence. We chose to identify them for simplicity, because it avoids the introduction of cumbersome (but important) subtleties and serves our purposes perfectly.

For example, consider the vector space $\mathbb{R}^3$. Then, $\{(1,0,0),(0,1,0),(0,0,1)\}$ is orthogonal. More generally, in the vector space $\mathbb{R}^n$, let e_i be a vector whose components are all zero, except the ith one, which is 1. Then, $\{e_1,...,e_n\}$ is an orthonormal basis of $\mathbb{R}^n$, called the *canonical basis* of $\mathbb{R}^n$.

Orthogonal complement

If W is a *subspace* of V, i.e. W vector space and $W \subset V$, then the *orthogonal complement* of W in $V$is defined as:

$$W^{\perp} = \{v \in V,\ \forall w \in W,\ \langle v, w\rangle = 0\}\ . \tag{A.26}$$

From now on, unless otherwise stated, inner products will be defined on real vector spaces and will therefore be at values in $\mathbb{R}$.

Quadratic form

The *quadratic form* associated with an inner product $\boldsymbol{g}$ on V is the function: $\boldsymbol{q} : V \longrightarrow \mathbb{R}$ such that:

$$\forall v \in V,\ \boldsymbol{q}(v) = \boldsymbol{g}(v,v)\ . \tag{A.27}$$

For an inner product $\boldsymbol{g}$ on a vector space V, and its associated quadratic form, we have:

$$\forall (u,v) \in V^2,\ \boldsymbol{g}(u,v) = \frac{1}{2}\left[\boldsymbol{q}(u+v) - \boldsymbol{q}(u) - \boldsymbol{q}(v)\right]\ . \tag{A.28}$$

Moreover, two distinct inner products on V cannot give the same quadratic form. This shows that one can equivalently define the structure on the vector space via the inner product or the quadratic form.

Unit vectors

A vector $v \in V$ for which $|q(v)| = 1$ is called a *unit vector*. A basis of V that is orthogonal and made of unit vectors is called *orthonormal*.

Using unit vectors, we can then create orthonormal bases.

Existence and structure of orthonormal bases

Let V be a vector space of finite dimension $n \in \mathbb{N}$ on which an inner product $\boldsymbol{g} : V \times V \longrightarrow \mathbb{R}$

is defined. Then, there exists at least an orthonormal basis of V, $\{e_i\}_{i\in\{1,...,n\}}$. Moreover, the number of basis vectors e_i for which $\boldsymbol{q}\,(e_i) = -1$ is the same for any such basis.

The number r of basis vectors e_i's for which $q\,(e_i) = -1$ is called the *index of the inner product* g.

A.2.4 An important set of linear functions: Orthogonal transformations

Here, we restrict our attention to the vector space $\mathbb{R}^n$ with the usual addition of vectors and the usual scalar multiplication, and with the dot product and its associated norm $\|.\|_E$.

Orthogonal transformations

We call *orthogonal transformations* the maps $\phi : \mathbb{R}^n \to \mathbb{R}^n$ such that:

$$\forall (x, y) \in \mathbb{R}^n,\ \phi(x) \cdot \phi(y) = x \cdot y\ . \tag{A.29}$$

For example, the *antipodal transformation* $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ such that: $\forall x \in \mathbb{R}^n,\ \mathcal{A}(x) = -x$ is orthogonal since:

$$\mathcal{A}(x) \cdot \mathcal{A}(y) = (-x) \cdot (-y) = x \cdot y\ . \tag{A.30}$$

Can you justify the name 'antipodal' by considering the effect of this transformation in $\mathbb{R}^2$ and $\mathbb{R}^3$?

Orthogonal transformation and orthogonal bases

Let $\{e_1, ..., e_n\}$ be the canonical basis of $\mathbb{R}^n$.
A function $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is an *orthogonal* transformation iff ϕ is *linear* and $\{\phi(e_1), ..., \phi(e_n)\}$ is an orthonormal basis of $\mathbb{R}^n$.

Now that we know that orthogonal transformations are linear, we associate matrices to them.

Orthogonal matrices

A real $n\times n$ matrix O is *orthogonal* iff it is associated to an orthogonal transformation ϕ defined by: $\forall x \in \mathbb{R}^n,\ \phi(x) = Ox$. Therefore, an orthogonal matrix O is such that: $\forall x \in \mathbb{R}^n,\ Ox = x$. The set of orthogonal matrices is denoted $O(n)$.

We have the following result:

Properties of orthogonal matrices

Let O be a real $n \times n$ matrix. Then, the following propositions are equivalent:

(i) O is orthogonal;

(i) The columns of O form an orthonormal basis of $\mathbb{R}^n$;

(iii) $O^t O = OO^t = I$, where t stands for transposition, and I is the identity matrix;

(iv) The rows of O form an orthonormal basis of $\mathbb{R}^n$.

Orthogonal transformations are the rotations of $\mathbb{R}^n$. Try and justify that in the cases $n = 2$ and $n = 3$.

A.3 Multilinear algebra and tensors

Here, we introduce multilinear functions and tensors.

A.3.1 Dual space

Among the linear functions, some are of particular interest in this course: those with values in the vector space K.

Dual space

Let V be a vector space on the field K. Then, the vector space $L(V, K)$ is called the *dual space* of V. It is usually denoted V^*. Elements of V^* are called *one-forms* or *covectors* on V.

We then have the following property, as long as we restrict ourselves to finite dimensional vector spaces:

$$\dim V^* = \dim V \ . \tag{A.31}$$

Consider now that $\dim V = n$, and pick up a basis $\{e_i\}_{i\in\{1,2,\ldots,n\}}$ of V. A natural basis for K is the neutral element for the field multiplication, 1. Then, we have seen that $L(V, K) = V^*$ has a

basis $\{\omega^i\}_{i\in\{1,2,...,n\}}$, such that:

$$\omega^i e_j = \delta^i{}_j \ . \tag{A.32}$$

The linear maps $\omega^i : V \to K$ defined above are called the *dual basis* to the basis $\{e_i\}$.
Let $w \in V^*$. Then, by definition, w is a linear scalar-valued function on V. That means that we have:

$$\forall v \in V,\ w(v) = w^i \omega_i(v) = w^i \omega_i \left(v_j e^j\right) = w^i v_j \delta^i{}_j = w^i v_i \ . \tag{A.33}$$

Elements of the dual basis are the simplest linear functions on a vector space once a basis of this space has been chosen. Indeed, the function ω^j associates to the vector $v \in V$ its j^{th} coordinate, v^j in the basis $\{e_i\}$.

A.3.2 Multilinear functions

So far, apart from the inner product, we have studied linear functions. The notions we have seen can actually be extended to a larger class of functions, called multilinear functions. Consider a set of N vector spaces $\{V_i\}_{i\in\{1,2,...,N\}}$ on the field K, and a vector space W on the field K.

Multilinear map

A map $f : V_1 \times V_2 \times ... \times V_N \to W$ is a *N-linear function* iff:

$$\begin{gathered}\forall i \in \{1,2,..,N\},\ \forall (a,b) \in K^2,\\ \forall (v_1,...,v_{i-1},v_{i+1},...,v_N) \in V_1 \times ... \times V_{i-1} \times V_{i+1} \times ... \times V_N,\\ \forall (v_i^1, v_i^2) \in V_i^2, f(v_1,...,v_{i-1}, a v_i^1 + b v_i^2, v_{i+1},...,v_N) =\\ a f(v_1,...,v_{i-1},v_i^1,v_{i+1},...,v_N) + b f(v_1,...,v_{i-1},v_i^2,v_{i+1},...,v_N).\end{gathered} \tag{A.34}$$

This means that f is linear in any one of its N variables.
If $N = 2$, the function is said to be *bilinear*.

The set of all the multilinear functions mapping the set $V_1 \times V_2 \times ... \times V_N$ into W forms a vector space denoted $L(V_1, ..., V_N; W)$.

A.3.3 Tensor algebra

Now that we know what multilinear functions are, we are going to concentrate on a particular subclass of multilinear functions that will be of relevance in that course: tensors.
Consider, as before, a vector space V of dimension $n \in \mathbb{N}^*$ on a field K.

Tensors

We call *tensor* of type $(r, s) \in \mathbb{N}^2$ over V, a multilinear function $T : V^* \times ... \times V^* \times V \times ... \times V \rightarrow K$ where V^* is repeated r times and V s times. r is called the *contravariant degree* of T and s its *covariant degree*.

In other words, a tensor of type (r, s) is a scalar-valued multilinear function that has r variables in the dual of V and s variables in V. For example, a tensor $T : V^* \times V \rightarrow K$ of type $(1, 1)$ takes a linear scalar-valued function on V and a vector of V as variables, and returns a scalar. The set of all tensors over V of type (r, s) is a vector space, called the *tensor space* over V of type (r, s). It is noted $T^r_s(V)$ or $V \otimes ... \otimes V \otimes V^* \otimes ... \otimes V^*$, where V^* is repeated r times and V s times. A tensor of type $(0, 0)$ is, by definition a scalar, so that we can write: $T^0_0(V) = K$. A tensor of type $(1, 0)$ is called a *contravariant vector*, and a tensor of type $(0, 1)$ a *covariant vector*. A contravariant vector is just an element of V, i.e. a usual vector of V. A covariant vector is a scalar-valued linear function on V, that means, it takes a vector of V, and returns a scalar: it is an element of V^*, i.e. a one-form. It is called a vector because it is seen as an element of V^* considered as a vector space.

By extension, a tensor of type $(r, 0)$ is called a *contravariant tensor*, and one of type $(0, s)$ a *covariant tensor*.
The structure of vector space on T^s_r gives us an addition between tensors of the same type, as well as a multiplication of tensors by scalars. On top of this structure, we will now define the *tensor product*, that allows one to combine tensors of different types.

Tensor product

Let $A \in T^r_s$ and $B \in T^t_u$. The *tensor product* of A and B is a tensor, denoted $A \otimes B$, of type

$(r+t, s+u)$ defined by:

$$\forall (w^1, .., w^{r+t}, v_1, ..., v_{s+u}) \in (V^*)^{r+t} \times V^{s+u},$$
$$A \otimes B(w^1, ..., w^{r+t}, v_1, ..., v_{s+u}) = A(w^1, ..., w^r, v_1, ..., v_s) B(w^{r+1}, ..., w^{r+t}, v_{s+1}, ..., v_{s+u}) \ . \tag{A.35}$$

One can easily check that the tensor product is associative, and that it is also distributive with respect to the addition of tensors:

- $\forall (A, B, C) \in T^r_s \times T^t_u \times T^v_w,\ (A \otimes B) \otimes C = a \otimes (B \otimes C)$;
- $\forall (A, B, C) \in T^r_s \times T^t_u \times T^v_w,\ A \otimes (B + C) = A \otimes B + A \otimes C$;
- $\forall (A, B, C) \in T^r_s \times T^t_u \times T^v_w,\ (A + B) \otimes C = A \otimes C + B \otimes C$.

CAUTION: Usually, the tensor product is not commutative: $A \otimes B \neq B \otimes A$.

We see that, thanks to the tensor product one can define a bilinear functions from two linear functions. Indeed, consider $f \in V^*$ and $g \in V^*$. Then, $f \otimes g$ is an elements of $L(V^*, V^*, K) = T^0_2(V)$, that means, it is a bilinear function on the space of linear functions on V, that takes two vectors of V^* and returns a scalar.

Another important notion is the one of *scalar product*:

Scalar product

We call scalar product of a vector space V, and we note $(.,.)$ the bilinear function:

$$(.,.) : V \times V^* \rightarrow K \ , \tag{A.36}$$

defined by:

$$\forall (v, w) \in V \times V^*,\ (v, w) = w(v) \ . \tag{A.37}$$

By definition, $(.,.) \in T^1_1(V)$.

Now, for $v \in V$, consider the linear function $(v, .) : V^* \rightarrow K$. It is an element of $T^1_0(V) \simeq V^{**}$, i.e. a covariant vector. We see that the name vector is then justified: vectors of V can be seen as acting on V^*, through the scalar product; in essence, there is a one-to-one relation between vectors of V

and a subset of elements of V^{**}. One can also see that a scalar product can be generated via an inner product by using a linear map L from V into V^*, then the bilinear function $(v, w) \in V \times V^* \mapsto \langle ., L. \rangle$ is a scalar product.

Up to now, we have introduced tensors and characterized a few of them, namely tensors of type $(1,0)$, $(0,1)$, $(2,0)$ and $(1,1)$. Tensors of type $(0,1)$ are linear functions on V. We have seen that they admit a decomposition into scalar components once a basis in chosen in V, and the dual basis constructed in V^*. The same thing happens to general tensors. Let e_i be a basis of V and ω^i the dual basis in V^*. then we have:

Bases of tensors

Let $T \in T^r_s(V)$. Then T is uniquely determined by its values on e_i and ω^i. Precisely, we have:

- $\left\{e_{i_1} \otimes e_{i_2} \otimes ... \times e_{i_s} \otimes w^{j_1} \otimes w^{j_2} \otimes ... \otimes w^{j_r}\right\}_{i_k \in \{1,...,N\}, j_l \in \{1,...,N\}}$ is a basis of $T^r_s(V)$;
- $T = T^{j_1 \cdots j_r}{}_{i_1 \ldots is} e_{i_1} \otimes e_{i_2} \otimes ... \times e_{i_s} \otimes w^{j_1} \otimes w^{j_2} \otimes ... \otimes w^{j_r}$, where:

$$T^{j_1 \cdots j_r}{}_{i_1 \ldots is} = T(w^{j_1}, ..., w^{j_r}, e_{i_1}, ..., e_{i_s}). \tag{A.38}$$

As an obvious corollary, we have:

$$\dim T^r_s(V) = (\dim V)^{r+s} \quad . \tag{A.39}$$

Finally, we can give a new, simple interpretation of a tensor in $T^1_1(V)$:

Matrices as tensors

Let $T \in T^1_1(V)$, with $T = T^i_j e_i \otimes \omega^j$. Then, the T^i_j are:

- the components of $T \in T^1_1(V)$ with respect to the basis $e_i \otimes \omega^j$;
- for a fixed $v \in V$, the matrix entries of the matrix of the linear function:

$$T_1 : \begin{array}{ccc} V^* & \to & V^* \\ f & \mapsto & T(f, v) \end{array} \quad , \tag{A.40}$$

with respect to the dual basis ω^i; in that case i is the row index, and j the column index;

- for a fixed $f \in V^*$, the matrix entries of the matrix of the linear function:

$$T_2 : \begin{array}{ccc} V & \to & V \\ v & \mapsto & T(f,v) \end{array} , \tag{A.41}$$

with respect to the basis e_i; in that case i is the row index, and j the column index.

As an illustration, we can look at the representation, on a basis, of an inner product $g : V \times V \to K$. g is a tensor of type $(0,2)$. We have, using the previous bases for V and its dual:

$$\forall (v,w) \in V^2,\ v = v^i e_i,\ w = w^i e_i,\ g(v,w) = \sum_{i,j=1}^{n} v^i w^j g\left(e_i, e_j\right) \ . \tag{A.42}$$

A.4 Topological spaces

Topological spaces are the 'simplest' spaces, in that they are the ones with the most minimal structure: metric spaces are a subset of manifolds and manifolds a subset of topological spaces.

A.4.1 Topological spaces: definitions

Topological space

Let X be a set and $C = \{U_i, i \in I\}$ denote a collection of subsets of X: $\forall i \in I,\ U_i \subseteq X$, where I is a set of indices. Then, (X, C) is a *topological space* iff C satisfies the following conditions:

(i) $\emptyset \in C$;

(ii) $X \in C$;

(iii) If J is any subcollection of indices in I (it may be infinite), then: $\cup_{j \in J} U_j \in C$;

(iv) If K is any finite subcollection of indices in I, then: $\cap_{k \in K} U_k \in C$.

Often, X alone is called a topological space, when there is no ambiguity; the subsets $U_i \in C$ are called the *open sets*, and C is said to give a *topology* to X.

A good way to turn a space into a topological one is to give it a metric.

Metric

Let X be a set. A *metric* $d : X \times X \to \mathbb{R}$ is a function that satisfies:

(i) $\forall (x, y) \in X \times X,\ d(x, y) = d(y, x)$;

(ii) $\forall (x, y) \in X \times X,\ d(x, y) \geq 0$, and equality holds iff $x = y$;

(iii) $\forall (x, y, z) \in X \times X \times X,\ d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

If X is given a metric d, it becomes a topological space whose open sets are the 'open discs of radius ϵ':

$$U_\epsilon(x) = \{y \in X,\ d(x, y) < \epsilon\}, \tag{A.43}$$

together with all their possible unions. The topology C thus defined is called the *metric topology* determined by d, and (X, C) is called a *metric space*.
Given a topological space (X, C) and $A \subset X$. Then C induces the *relative topology* in A: $C_A = \{U_i \cap A,\ U_i \in C\}$.

A.4.2 Continuous maps

Continuity

Let X and Y be topological spaces. A map $f : X \to Y$ is *continuous* if the inverse image of an open set in Y is an open set in X.

This definition is in agreement with our intuitive idea of continuity. Indeed, consider the map $f : \mathbb{R} \to \mathbb{R}$ defined by: $f(x) = -x + 1$ if $x \leq 0$, and $f(x) = -x + 1/2$ if $x > 0$. Using the usual topology of $\mathbb{R}$ (the one from open intervals), then $]a, b[$ is open for any a and $b > a$ by definition. We know that in calculus, we would say that f is discontinuous at 0. Now, consider $]3/2, 2[$, which is open. We have $f^{-1}(]3/2, 2[) =]-1, -1/2[$ which is open. But if we take the open set $]3/4, 5/4[$, we have $f^{-1}(]3/4, 1/4[) =]-1/4, 0]$, which is not open, so f is not continuous.
By taking $f(x) = x^2$, which is a continuous map for the standard topology of $\mathbb{R}$, show that the converse of the previous definition is not true, that is that 'f is continuous if it maps an open set into an open set' is not true. (*Hint*: Consider the image of $]-\epsilon, \epsilon[$).

A.5 Neighbourhoods and Hausdorff spaces

Neighbourhood

Let X be a topological space with a topology C. $N \subset X$ is called a *neighbourhood* of $x \in X$ iff $x \in N$, and N contains at least an open set U_i such that $x \in U_i$.

Note that N does not have to be open. If it is, it is called an open neighbourhood.
A simple example of neighbourhood is provided by $X = \mathbb{R}$ with the usual topology. Then $[-1, 1]$ is a neighbourhood of any point $x \in]-1, 1[$.

A topological space (X, C) is called a *Hausdorff space* if, for an arbitrary pair of distinct points x and x' in X, we can find neighbourhoods U_x and $U_{x'}$ of x and x' respectively, such that $U_x \cap U_{x'} = \emptyset$. In physics, (almost) all spaces that appear are Hausdorff spaces, so in the remainder of these notes, we will always assume that topological spaces are Hausdorff spaces.

Homeomorphism

Let X_1 and X_2 be topological spaces and $f : X_1 \to X_2$ be a continuous map such that $f^{-1} : X_2 \to X_1$ exists and is also continuous. Then, f is called an *homeomorphism* and X_1 and X_2 are said to be *homeomorphic*.

Homeomorphic spaces can thus be deformed into each other continuously, so that, topologically, they are virtually equivalent and can be treated as the same space.

B

Isometries and Killing vector fields

Contents

B.1 Maps and induced maps

Let (M, g) be spacetime manifold manifold. Let $f : M \to M$ be a smooth map of M onto itself. We recall that given a map f such that for $p \in M$, its image is $f(p) \in M$, we can define its action on tangent vectors by introducing the induced map $f_* : T_pM \to T_{f(p)}M$ called the *pushforward* of f. To any vector $\boldsymbol{X}$ tangent to M at p, it associates a tangent vector at $f(p)$, denoted $\boldsymbol{f_*X}$ so that, given a function $g : M \to \mathbb{R}$:

Pushforward of a map f

$$\boldsymbol{f_*X}(g) = \boldsymbol{X}\,(g \circ f)\ . \tag{B.1}$$

In other words, it defines the directional derivative of any function g at $f(p)$ in terms of the one induced by $\boldsymbol{X}$ on the function $g \circ f$ at p. If we call $x = \left(x^0, x^1, x^2, x^3\right) = \varphi(p)$ and $y = \left(y^0, y^1, y^2, y^3\right) = \psi(f(p))$ local charts around p and $f(p)$ respectively, so that $y^\mu = \psi\left(f(\varphi^{-1}(x))\right)$, abusing notations to identify the action of the function on M and its action "down on the charts", i.e. writing $g(y) = g(f(x))$ for the proper $g\left(\psi^{-1}(y)\right) = g\left(f\left(\varphi^{-1}(x)\right)\right)$, Eq. (B.1) becomes:

$$(f_*X)^\mu \frac{\partial g\,(y)}{\partial y^\mu} = X^\nu \frac{\partial g\,(f(x))}{\partial x^\nu} \tag{B.2}$$

$$= X^\nu \frac{\partial y^\mu}{\partial x^\nu}\frac{\partial g}{\partial y^\mu}\ , \tag{B.3}$$

so that we can write:

$$(f_*X)^\mu = \frac{\partial y^\mu}{\partial x^\nu} X^\nu\ . \tag{B.4}$$

A similar induced map, known as the *pullback* of f and denoted $f^*T^*_{f(p)}M \to T^*_pM$ can be defined on one-forms (covectors), in order to obtain one-forms at p given one-forms at $f(p)$ (hence the name pullback). Given any tangent vector X at p and any one-form at $f(p)$, we define the pullback of $\boldsymbol{\omega}$ as:

Pullback of a map f

$$\boldsymbol{f^*\omega}(\boldsymbol{X}) = \boldsymbol{\omega}\,(\boldsymbol{f_*X})\ . \tag{B.5}$$

In terms of components, we get:

$$(f^*\omega)_\mu = \frac{\partial y^\nu}{\partial x^\mu}\omega_\nu\ . \tag{B.6}$$

Note that it is not the same as a local change of coordinate (well, changes of coordinates can be formulated in terms of pullbacks but let us not get unnecessarily complicated here)!
This can be straightforwardly generalised to tensors, in particular, for the metric tensor $\boldsymbol{g}$, we can define another metric tensor at p given the metric at $f(p)$ via:

$$\forall(\boldsymbol{X},\boldsymbol{Y}) \in T_pM \times T_pM,\ \boldsymbol{f}^*\boldsymbol{g}_{|p}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{g}_{|f(p)}(\boldsymbol{f}_*\boldsymbol{X},\boldsymbol{f}_*\boldsymbol{Y})\ . \tag{B.7}$$

B.2 Isometries

Isometry

f is called an *isometry* iff it is a bijective map of $\mathcal{M}$ onto itself (aka a *diffeomorphims*) that preserves the metric:

$$\forall p \in M,\ \boldsymbol{f}^*\boldsymbol{g}_{|p} = \boldsymbol{g}_{|p}\ . \tag{B.8}$$

In other words, the metric properties (angle between vectors, length of vectors) are invariant under the transformation f: if we calculate the angle between two vectors at p, then transport these vectors to $f(p)$ using a pushforward, the angle between the transported vectors remain the same as it was.
If we introduce local coordinates around p and $f(p)$, such as $x = \left(x^0, x^1, x^2, x^3\right) = \varphi(p)$ and $y = \left(y^0, y^1, y^2, y^3\right) = \psi(f(p))$, so that $y = (\psi \circ f \circ \varphi^{-1})(x)$, the definition of an isometry reduces to:

$$\forall(\mu,\nu) \in \{0,...,3\}^2,\ \frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu} g_{\alpha\beta}(f(p)) = g_{\mu\nu}(p)\ . \tag{B.9}$$

It is easy to see from this definition that the set of isometries of a manifold M forms a group under the composition of maps. Because isometries preserve the local 'length' of a vector in T_pM, they can be interpreted as the rigid motions of the manifold. Of course, they apply to Riemannian as well as Lorentzian manifolds. For example, on the manifold $\mathbb{R}^n$ with the standard scalar product as metric, the isometry group is the Euclidean group:

$$\mathbb{E}^n = \{f : x \mapsto Ax + T, A \in SO(n), T \in \mathbb{R}^n\}\ , \tag{B.10}$$

consisting of all the rotations and translations (and their linear combinations).

B.3 Killing vector fields

When a (pseudo-)Riemannian manifold possesses some isometries, one can introduce a new kind of vector fields that generate these isometries in the sense that they represent the infinitesimal version of them: moving a small amount in the direction of the vector field, the metric structure remains unchanged .

Let (M, g) be a spacetime manifold and $X \in \mathcal{X}(M)$. Let $\epsilon \in \mathbb{R}$ be an infinitesimal parameter. Let $p \in M$ and (U, φ) a local chart around p such that, in this chart $x = \varphi(p)$ and $X = X^\mu \frac{\partial}{\partial \mu}$. Let us define the diffeomorphism $f : M \to M$ such that[1] $\varphi(f(p))^\mu = x^\mu + \epsilon X^\mu(p)$. This is just an infinitesimal displacement from p in the direction of X. If f is an isometry, then, by definition of the pushforward:

$$g_{\mu\nu}(x) = \frac{\partial(x^\alpha + \epsilon X^\alpha(p))}{\partial x^\mu} \frac{\partial(x^\beta + \epsilon X^\beta(p))}{\partial x^\nu} g_{\alpha\beta}\left(x^m + \epsilon X^m\right) \ . \tag{B.11}$$

Expanding at first order in ϵ and taking the limit $\epsilon \to 0$, this leads to the *Killing equations*:

$$X^k \partial_k g_{ij} + g_{kj}\partial_i X^k + g_{ik}\partial_j X^k = 0, \tag{B.12}$$

which is equivalent to:

$$\boxed{\nabla_{(\alpha} X_{\beta)} = 0 \ .} \tag{B.13}$$

A vector field X that satisfies this Killing equation is called a *Killing vector field*. Moving along the curves tangent to this vector field, the geometry remains unchanged: the Killing vector field represents the direction of symmetry of a manifold. Killing vector fields are said to be dependent if one of them can be expressed as a linear combination of the others with constant coefficients.

Finally, let us note that from the Killing equation in the form (B.12), it is easy to show that if the metric components do not depend explicitly on a coordinate x^μ when expressed in these coordinates, then $\frac{\partial}{\partial x^\mu}$ is a Killing vector field.

B.4 Example of killing vectors: the sphere S^2

In order to illustrate what Killing vectors are, let us look at the isometries of the standard, two-dimensional sphere.

[1] We can always choose ϵ small enough to ensure that $f(p) \in U$ so that we can use the same chart.

Consider the manifold $S^2 = \left\{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\right\}$. Let us Introduce the standard spherical coordinates $\varphi(p) = (\theta, \phi) \in]0, \pi[\times]0, 2\pi[$ on $U = \varphi^{-1}(]0, \pi[\times]0, 2\pi[) \subset S^2$, which is the sphere to which we removed its poles and the half-meridian connecting them. The standard Euclidean metric of $\mathbb{R}^3$:

$$\boldsymbol{G} = \mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y + \mathrm{d}z \otimes \mathrm{d}z \tag{B.14}$$

induces the Riemannian metric on U given by:

$$\boldsymbol{g} = \mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta\, \mathrm{d}\phi \otimes \mathrm{d}\phi \ . \tag{B.15}$$

Writing the Killing equation (B.13) for this metric, we get:

$$\partial_\theta X^\theta = 0 \tag{B.16}$$

$$\partial_\phi X^\phi = -\frac{\cos\theta}{\sin\theta} X^\theta \tag{B.17}$$

$$\partial_\phi X^\theta = -\sin^2\theta \partial_\theta X^\phi \ . \tag{B.18}$$

Solving these equations, we see that any Killing vector field $\boldsymbol{X}$ of S^2 on U, can be written:

$$\boldsymbol{X} = A\boldsymbol{L_1} + B\boldsymbol{L_2} + C\boldsymbol{L_3} \ , \tag{B.19}$$

where A, B and C are arbitrary real numbers and:

$$\boldsymbol{L_1} = \sin\phi \frac{\partial}{\partial\theta} + \frac{\cos\theta\cos\phi}{\sin\theta}\frac{\partial}{\partial\phi} \tag{B.20}$$

$$\boldsymbol{L_2} = \cos\phi \frac{\partial}{\partial\theta} - \frac{\cos\theta\sin\phi}{\sin\theta}\frac{\partial}{\partial\phi} \tag{B.21}$$

$$\boldsymbol{L_3} = \frac{\partial}{\partial\phi} \tag{B.22}$$

are linearly independent vector fields on U that are Killing vector fields. For any running point $p(t) = (x(t), y(t), z(t)) \in U$, along the integral curves of L_1, L_2 and L_3, we have:

$$\frac{dx}{dt} = 0 \qquad \text{along } L_1 \tag{B.23}$$

$$\frac{dy}{dt} = 0 \qquad \text{along } L_2 \tag{B.24}$$

$$\frac{dz}{dt} = 0 \qquad \text{along } L_3 \ , \tag{B.25}$$

where, in each case, t is the affine parameter along the integral curves of L_1, L_2 and L_3 respectively. Thus, the integral curves of L_1, L_2 and L_3 are circles at x, y and z constant respectively, so that these

vectors correspond to rotations around the x, y and z axes respectively. We see that the sphere has $3 = 2 \times (2+1)/2$ independent Killing vector fields so it is maximally symmetric, as we are going to explain below.

B.5 Maximally symmetric spaces

The maximum number of symmetries on a metric manifold is related to the dimension of the manifold n by $n(n+1)/2$. Manifolds which admit exactly $n(n+1)/2$ independent Killing vector fields are called *maximally symmetric spaces*. Let us see how it works.

B.5.1 Properties of maximally symmetric spaces

We can classify isometries f into two classes: translations and rotations, depending on whether or not they admit a fixed point. A point $p \in \mathcal{M}$ is a fixed point of the isometries f iff $f(p) = p$. Isometries that do not have any fixed point are called *translations*, while an isometry with a fixed point p is called a *rotation around* p. The pushforward of a rotation is exactly what one would call a rotation in the tangent space $T_p\mathcal{M}$: it transforms vectors at p into one another. A maximally symmetric space is such that:

- $(\mathcal{M}, \boldsymbol{g})$ is *homogeneous*. This means that for every pair of points $(p, q) \in \mathcal{M}$, there exists a translation f that maps p into q.
- $(\mathcal{M}, \boldsymbol{g})$ is *isotropic*. This means that for any point $p \in \mathcal{M}$, there exists a rotation that fixes p and such that for any $(\boldsymbol{v}, \boldsymbol{w}) \in T_p\mathcal{M} \times T_p\mathcal{M}$, $\exists \alpha \in \mathbb{R}$, $f_*\boldsymbol{v} = \alpha \boldsymbol{w}$.

Because we will need to talk about maximally symmetric spacetimes, but also three-dimensional spaces, let us revert to generic notations and work on a manifold of dimension $n \in \mathbb{N}^*$ with an arbitrary metric. Let $p \in \mathcal{M}$ and $\{\hat{\boldsymbol{e}}_{(\boldsymbol{i})}\}$ a local g-orthonormal basis such that, for the metric compatible connection:

$$\nabla_j \hat{\boldsymbol{e}}_{(\boldsymbol{i})} = \Gamma^k{}_{ij}(p) \hat{\boldsymbol{e}}_{(\boldsymbol{k})} = 0 \implies \Gamma^k{}_{ij}(p) = 0 \,. \tag{B.26}$$

A local translation at p is associated with a Killing vector $\boldsymbol{\xi}$ such that for the small displacement $\boldsymbol{\delta x} = \varepsilon \xi^i \hat{\boldsymbol{e}}_{(\boldsymbol{i})}$, we satisfy the Killing equation (B.12). In the local orthonormal frame chosen here,

for which connection coefficients vanish at p, this means that:

$$\frac{\partial \xi_j}{\partial \hat{x}^i} = -\frac{\partial \xi_i}{\partial \hat{x}^j} \ . \tag{B.27}$$

In a maximally symmetric space, any point $p + \delta p$ is related to p via a translation. In particular, this must be true for the n small displacements along the "axes", with $\boldsymbol{\delta x} = \varepsilon \hat{\boldsymbol{e}}_{(i)}$ which are thus Killing vectors at p. Any translation can then be written as a linear combination of these linearly independent displacements. Thus, we see that, in a maximally symmetric space, the set of translations at p is generated by n linearly independent Killing vectors.
Infinitesimal rotations at p, on the other hand, are generated by Killing vectors such that:

$$\boldsymbol{\xi}(p) = 0 \ . \tag{B.28}$$

Their action on the tangent space at p can be summarised by their action on the basis vectors. Since any vector must be mappable into another one, we can map any basis vector into another one and we therefore have $n(n-1)/2$ rotations:

$$f_*^{i \to j} \hat{\boldsymbol{e}}_{(i)} = \hat{\boldsymbol{e}}_{(j)} \ , \tag{B.29}$$

each generated by a Killing vector.
Thus, as announced, the total number of Killing vector fields of a maximally symnmetric space is:

$$N_{\max} = \underbrace{n}_{\text{translations}} + \underbrace{\frac{n(n-1)}{2}}_{\text{rotations}} = \frac{n(n+1)}{2} \ . \tag{B.30}$$

Let us now turn to the Riemann curvature of the metric connection:

$$\boldsymbol{R} = R^i{}_{jkl} \hat{\boldsymbol{e}}_{(i)} \otimes \hat{\boldsymbol{\omega}}^{(j)} \otimes \hat{\boldsymbol{\omega}}^{(k)} \otimes \hat{\boldsymbol{\omega}}^{(l)} \ , \tag{B.31}$$

where $\{\hat{\boldsymbol{\omega}}^{(i)}\}$ is the dual basis associated with $\{\hat{\boldsymbol{e}}_{(i)}\}$. Using the metric, let us then define the $(2,2)$ tensor:

$$\boldsymbol{\mathcal{R}} = R^{ij}{}_{kl} \hat{\boldsymbol{e}}_{(i)} \otimes \hat{\boldsymbol{e}}_{(j)} \otimes \hat{\boldsymbol{\omega}}^{(k)} \otimes \hat{\boldsymbol{\omega}}^{(l)} \ . \tag{B.32}$$

It is clearly a symmetric linear map from the set of antisymmetric bilinear forms:

$$W = \left\{ \boldsymbol{w} \in T^0_{p,2}(\mathcal{M}),\ \forall (\boldsymbol{X}, \boldsymbol{Y}) \in T_p\mathcal{M} \times T_p\mathcal{M}, \boldsymbol{w}(\boldsymbol{X}, \boldsymbol{Y}) = -\boldsymbol{w}(\boldsymbol{Y}, \boldsymbol{X}) \right\} \ , \tag{B.33}$$

onto itself. Therefore, it is diagonalisable using an orthonormal basis of W. Let us call $\hat{\boldsymbol{\Omega}}_{(a)} = \hat{\Omega}^{(a)}_{ij}\left[\hat{\boldsymbol{\omega}}_{(i)} \otimes \hat{\boldsymbol{\omega}}_{(j)} - \hat{\boldsymbol{\omega}}_{(j)} \otimes \hat{\boldsymbol{\omega}}_{(i)}\right]$ that basis, with eigenvalues $\lambda_{(a)}$. Note that there are $n(n-1)/2$ such eigenvectors and eigenvalues, the dimension of W. Because the manifold is maximally symmetric, we can use an arbitrary rotation at p to transform the local orthonormal basis of $T_p\mathcal{M}$ and thus the dual basis in $T^*_p\mathcal{M}$. But such an operation will affect the antisymmetric bilinear forms $\hat{\boldsymbol{\Omega}}_{(a)}$ and by appropriate choices of rotations, we must be able to map them into one another. This implies that all the eigenvalues must be equal: $\lambda_{(a)} = \kappa$ for all a. Thus as a mapping on antisymmetric bilinear forms, in the eigenbasis:

$$\boldsymbol{\mathcal{R}} = \kappa \boldsymbol{Id} \ . \tag{B.34}$$

Writing this in a local coordinate basis:

$$R^{ij}{}_{kl} = \kappa \delta^i{}_{[k}\delta^j{}_{l]} \ , \tag{B.35}$$

so that:

$$R^i{}_{jkl} = g_{jm}R^{im}{}_{kl} = \kappa\left[\delta^i{}_k g_{jl} - \delta^i{}_l g_{jk}\right] \ . \tag{B.36}$$

The Ricci tensor is then:

$$R_{ij} = \kappa(n-1)g_{ij} \ , \tag{B.37}$$

and the Ricci scalar:

$$R = n(n-1)\kappa \ , \tag{B.38}$$

so that the arbitrary constant κ is, up to a numerical factor, the curvature of the manifold. As we can see maximally symmetric spaces are thus *constant curvature spaces*.

B.5.2 Riemannian maximally symmetric spaces in 3 dimensions

In the Riemannian case and in 3 dimensions, the maximally symmetric spaces fall into 3 categories that will be important in chapter 6, for cosmology.

1. If $\kappa = 0$, we have *flat space*, $\mathbb{E}^3$, which is homeomorphic to $\mathbb{R}^3$ with the standard euclidean metric, so that, a the Cartesian chart (x, y, z):

$$\mathrm{d}s^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 \ . \tag{B.39}$$

2. If $\kappa > 0$, the manifold is homeomorphic to the 3-*sphere*, $\mathbb{S}^3$, defined as the subset of $\mathbb{R}^4$ such that:

$$\mathbb{S}^3 = \left\{(x, y, z, w) \in \mathbb{R}^4,\ x^2 + y^2 + z^2 + w^2 = 1\right\} \ . \tag{B.40}$$

The standard mapping of $\mathbb{S}^3$ onto $\mathbb{R}^4$ is given by the spherical coordinates:

$$\begin{cases} x = \sin\chi \sin\theta \cos\phi & \text{(B.41)} \\ y = \sin\chi \sin\theta \sin\phi & \text{(B.42)} \\ z = \sin\chi \cos\theta & \text{(B.43)} \\ w = \cos\chi \ , & \text{(B.44)} \end{cases}$$

with $\phi \in [0, 2\pi)$, $\theta \in [0, \pi]$ and $\chi \in [0, \pi]$. The induced metric on $\mathbb{S}^3$ by pullback of the Euclidean metric of $\mathbb{R}^4$ is then:

$$\mathrm{d}s^2 = \mathrm{d}\chi^2 + \sin^2\chi \left[\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\phi^2\right] \ . \tag{B.45}$$

One can readily check that the Riemann tensor of the metric connection is then of the form (B.36) with $\kappa = 1$. To go to the maximally symmetric space of curvature $\kappa > 0$ arbitrary, one simply works with:

$$\mathbb{S}^3_\kappa = \left\{(x, y, z, w) \in \mathbb{R}^4,\ x^2 + y^2 + z^2 + w^2 = \kappa^{-1}\right\} \ , \tag{B.46}$$

and change $\chi \to \sqrt{\kappa}\chi$. The isometry group generated by the Killing vector fields is $O(4)$. The 3-sphere is represented on Fig. B.1.

3. If $\kappa < 0$, the manifold is homeomorphic to the 3-*hyperboloid* with two sheets, $\mathbb{H}^3$. It is the subset of $\mathbb{R}^4$ such that:

$$\mathbb{H}^3 = \left\{(x, y, z, w) \in \mathbb{R}^4,\ x^2 + y^2 + z^2 - w^2 = -1\right\} \ . \tag{B.47}$$

The chart (χ, θ, ϕ) with $\chi \in \mathbb{R}_+$, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$ defined by:

$$\begin{cases} x = \sinh\chi \sin\theta \cos\phi & \text{(B.48)} \\ y = \sinh\chi \sin\theta \sin\phi & \text{(B.49)} \\ z = \sinh\chi \cos\theta & \text{(B.50)} \\ w = \cosh\chi \ , & \text{(B.51)} \end{cases}$$

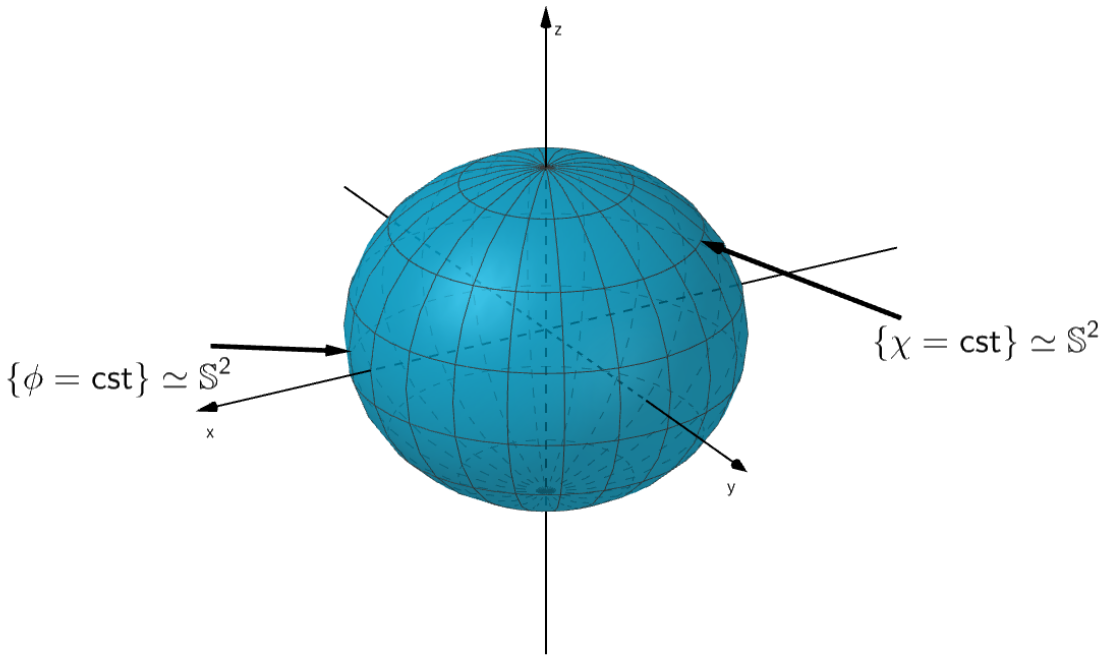


Figure B.1: The 3-sphere embedded in $\mathbb{R}^4$, with one dimension suppressed.

defines a mapping between $\mathbb{H}^3$ and $\mathbb{R}^4$ such that the Minkowski metric on $\mathbb{R}^4$ pulls back to:

$$\mathrm{d}s^2 = \mathrm{d}\chi^2 + \sinh^2\chi\left[\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2\right] \ . \tag{B.52}$$

Note that there is no embedding of $\mathbb{H}^3$ into $\mathbb{R}^4$ equipped with the Euclidean metric. Again, this corresponds to a Riemann tensor of the form (B.36), but with $\kappa = -1$. To get to $\kappa < 0$ arbitrary we once again perform the transformation $\chi \to \sqrt{-\kappa}\chi$. The isometry group of $\mathbb{H}^3$ is the orthochronous Lorentz group, that we introduced in chapter 2. One sheet of the 3-hyperboloid is represented on Fig. B.2.

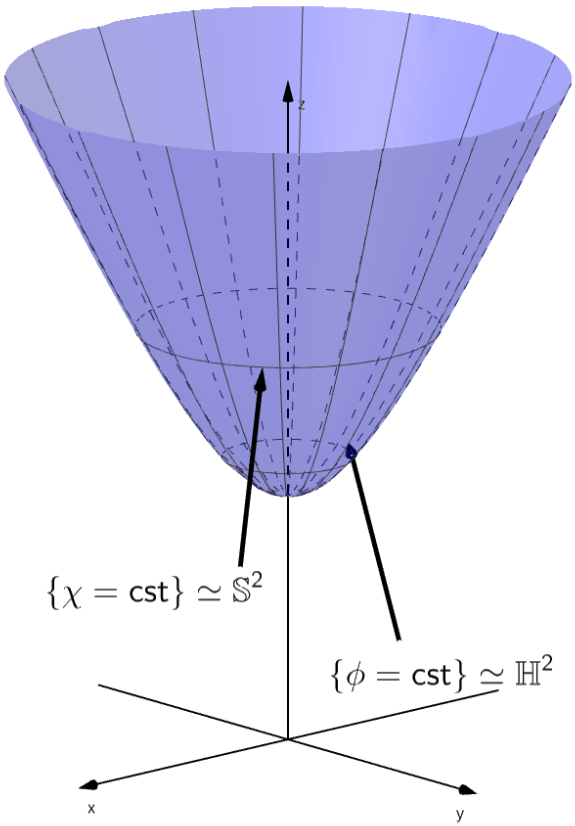


Figure B.2: On sheet of the 3-hyperboloid embedded in $\mathbb{R}^4$, with one dimension suppressed. Note that the embedding as shown here is not isometric since, by force, the Euclidean metric is used here in representing the ambient $\mathbb{R}^4$, instead of the Minkowski metric.

B.5.3 Einsteinian maximally symmetric spaces in 4 dimensions

To conclude this appendix, let us determine the 4 dimensional maximally symmetric geometries that satisfy the Einstein field equations. We call them *Einsteinian maximally symmetric spaces in 4 dimensions*. Writing the Einstein field equations for the Ricci tensor (B.37) and $n = 4$, we get:

$$(\Lambda - 3\kappa)\, g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{B.53}$$

In vacuum, we have immediately:

$$\kappa = \frac{\Lambda}{3} \ . \tag{B.54}$$

If we insist on having matter present, then it must be of a very special kind. For example, if we have a perfect fluid:

$$\rho = -p = \frac{3\kappa - \Lambda}{8\pi G} \ . \tag{B.55}$$

In any case, as in the Riemannian case in 3 dimension, we find 3 different cases.

1. If $\kappa = 0$, the Riemann tensor is zero and the spacetime is flat. It is simply *Minkowski spacetime*.

2. If $\kappa > 0$, the scalar curvature $R = 12\kappa > 0$ and this is *de Sitter spacetime*. It has the topology of $\mathbb{R} \times \mathbb{S}^3$. It can be isometrically embedded in $\mathbb{R}^5$ equipped with the Minkowski metric. Indeed, let us chart $\mathbb{R}^5$ with Cartesian coordinates (t, w, x, y, z) such that:

$$\boldsymbol{G} = -\mathrm{d}t \otimes \mathrm{d}t + \mathrm{d}w \otimes \mathrm{d}w + \mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y + \mathrm{d}z \otimes \mathrm{d}z \ . \tag{B.56}$$

   Defining a chart on de Sitter spacetime $(\tau, \chi, \theta, \phi)$ with $\tau \in \mathbb{R}$, $\chi \in [0, \pi]$, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$, such that we have the embedding map:

$$\begin{cases} \sqrt{\kappa}t = \sinh\left(\sqrt{\kappa}\tau\right) & \text{(B.57)} \\ \sqrt{\kappa}x = \cosh\left(\sqrt{\kappa}\tau\right) \sin\chi \sin\theta \cos\phi & \text{(B.58)} \\ \sqrt{\kappa}y = \cosh\left(\sqrt{\kappa}\tau\right) \sin\chi \sin\theta \sin\phi & \text{(B.59)} \\ \sqrt{\kappa}z = \cosh\left(\sqrt{\kappa}\tau\right) \sin\chi \cos\theta & \text{(B.60)} \\ \sqrt{\kappa}w = \cosh\left(\sqrt{\kappa}\tau\right) \cos\chi \ , & \text{(B.61)} \end{cases}$$

   the induced metric of de Sitter spacetime is:

$$\boldsymbol{g} = -\mathrm{d}\tau \otimes \mathrm{d}\tau + \frac{\cosh^2\left(\sqrt{\kappa}\tau\right)}{\kappa} \left[\mathrm{d}\chi \otimes \mathrm{d}\chi + \sin^2\chi \left(\mathrm{d}\theta \otimes \mathrm{d}\theta + \sin^2\theta \, \mathrm{d}\phi \otimes \mathrm{d}\phi\right)\right] \ . \tag{B.62}$$

   One can easily check that it is of constant curvature κ. In (T, x, y, z, w) coordinates, de Sitter is an hyperboloid of one sheet in $\mathbb{R}^5$ that lies 'along' the t axis.

3. If $\kappa < 0$, the scalar curvature $R = 12\kappa < 0$ and this is *anti-de Sitter spacetime*.Topologically, it is homeomorphic to $\mathbb{R}^4$. It can be embedded isometrically into $\mathbb{R}^2$ equipped with a pseudo-Riemannian metric that is not Lorentzian, but is given by:

$$\boldsymbol{G} = -\mathrm{d}t \otimes \mathrm{d}u - \mathrm{d}v \otimes \mathrm{d}v + \mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y + \mathrm{d}z \otimes \mathrm{d}z \ . \tag{B.63}$$

In these coordinates, anti-de Sitter spacetime is an hyperboloid of one sheet that lies in the subspace orthogonal to both the u and v axes. The map is given by:

$$\begin{cases} \sqrt{-\kappa}u = \sin\left(\sqrt{\kappa}\tau\right)\cosh\chi & \text{(B.64)} \\ \sqrt{-\kappa}v = \cos\left(\sqrt{\kappa}\tau\right)\cosh\chi & \text{(B.65)} \\ \sqrt{-\kappa}x = \sinh\chi\sin\theta\cos\phi & \text{(B.66)} \\ \sqrt{-\kappa}y = \sinh\chi\sin\theta\sin\phi & \text{(B.67)} \\ \sqrt{-\kappa}z = \sinh\chi\cos\theta\,, & \text{(B.68)} \end{cases}$$

with $\tau \in \mathbb{R}$, $\rho \in \mathbb{R}_+$, $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$. The induced metric is of constant curvature κ and reads:

$$\boldsymbol{g} = -\cosh^2\chi\,\mathrm{d}\tau\otimes\mathrm{d}\tau + \frac{1}{\kappa}\left[\mathrm{d}\chi\otimes\mathrm{d}\chi + \sinh^2\chi\left(\mathrm{d}\theta\otimes\mathrm{d}\theta + \sin^2\theta\,\mathrm{d}\phi\otimes\mathrm{d}\phi\right)\right]\,. \qquad \text{(B.69)}$$

We see that the topology is that of $\mathbb{R}\times\mathbb{H}^3$. Note that anti-de Sitter is a static spacetime, since it has a timelike Killing vector field, $\frac{\partial}{\partial\tau}$.

C

Green function of the d'Alembert operator

Contents

We want to find expression (5.176) for the Green function of the d'Alembert operator in MInkowski spacetime. We recall that it is a function G such that:

$$\eta^{\mu\nu}\frac{\partial}{\partial x^\mu \partial x^\nu}G\left(\boldsymbol{x}-\boldsymbol{y}\right)=\delta^D\left(\boldsymbol{x}-\boldsymbol{y}\right)\,. \tag{C.1}$$

C.1 Covariant form

The standard approach consists in using a Fourier expansion:

$$G(\boldsymbol{x}-\boldsymbol{y})=\int \hat{G}(\boldsymbol{k})\mathrm{e}^{\mathrm{i}k_\mu(x^\mu-y^\mu)}\mathrm{d}^4k\,. \tag{C.2}$$

Then:

$$\frac{\partial}{\partial x^\mu \partial x^\nu}G\left(\boldsymbol{x}-\boldsymbol{y}\right)=\int \hat{G}(\boldsymbol{k})\left(-k_\mu k_\nu\right)\mathrm{e}^{\mathrm{i}k_\mu(x^\mu-y^\mu)}\mathrm{d}^4k\,, \tag{C.3}$$

so that Eq. (C.1) becomes:

$$\int \hat{G}(\boldsymbol{k})\left(-k_\mu k^\mu\right)\mathrm{e}^{\mathrm{i}k_\mu(x^\mu-y^\mu)}\mathrm{d}^4k=\delta^{(D)}(\boldsymbol{x}-\boldsymbol{y}) \tag{C.4}$$

$$=\frac{1}{(2\pi)^4}\int \mathrm{e}^{\mathrm{i}k_\mu(x^\mu-y^\mu)}\mathrm{d}^4k\,, \tag{C.5}$$

where we used the Fourier expansion of the delta function. Since this equation must hold for any $\boldsymbol{x}-\boldsymbol{y}$, we get:

$$-k_\mu k^\mu \hat{G}\left(\boldsymbol{k}\right)-\frac{1}{(2\pi)^4}=0\,, \tag{C.6}$$

in other words:

$$G\left(\boldsymbol{k}\right)=\frac{1}{(2\pi)^4\left[\left(k^0\right)^2-\left|\vec{k}\right|^2\right]}\,. \tag{C.7}$$

Let us denote $\omega=k^0$, $k=\left|\vec{k}\right|$, $x^\mu=(t,\vec{x})$ and $y^\mu=(t',\vec{y})$. Then:

$$G\left(\boldsymbol{x}-\boldsymbol{y}\right)=\frac{1}{(2\pi)^4}\int\frac{\exp\left[-\mathrm{i}\omega\left(t-t'\right)+\vec{k}\cdot\left(\vec{x}-\vec{y}\right)\right]}{\omega^2-k^2}\mathrm{d}\omega\mathrm{d}^3k\,. \tag{C.8}$$

We can always introduce some spherical coordinates (θ,ϕ) in Fourier space so that $\theta=0$ when $\vec{k}$ is along $\vec{x}-\vec{y}$. Then, we have:

$$\vec{k}\cdot\left(\vec{x}-\vec{y}\right)=k\left|\vec{x}-\vec{y}\right|\cos\theta \tag{C.9}$$

$$\mathrm{d}^3k=k^2\sin\theta\mathrm{d}k\mathrm{d}\theta\mathrm{d}\phi\,. \tag{C.10}$$

Finally, let $\mu = -\cos\theta$. Then, the Green function becomes:

$$G(\boldsymbol{x}-\boldsymbol{y}) = \frac{1}{(2\pi)^4}\int_{-\infty}^{+\infty}\mathrm{d}\omega\int_0^{+\infty}k^2\mathrm{d}k\int_{-1}^{1}\mathrm{d}\mu\int_0^{2\pi}\mathrm{d}\phi\frac{\mathrm{e}^{-\mathrm{i}\omega(t-t')}\mathrm{e}^{\mathrm{i}k|\vec{x}-\vec{y}|\mu}}{\omega^2-k^2} \tag{C.11}$$

$$= \frac{1}{(2\pi)^3}\int_0^{+\infty}k^2\mathrm{d}k\underbrace{\int_{-\infty}^{+\infty}\frac{\mathrm{e}^{-\mathrm{i}\omega(t-t')}}{\omega^2-k^2}\mathrm{d}\omega}_{=I(k,t-t')}\underbrace{\int_{-1}^{1}\mathrm{d}\mu\mathrm{e}^{\mathrm{i}k|\vec{x}-\vec{y}|\mu}}_{=\frac{\mathrm{e}^{\mathrm{i}k|\vec{x}-\vec{y}|}-\mathrm{e}^{-\mathrm{i}k|\vec{x}-\vec{y}|}}{\mathrm{i}k|\vec{x}-\vec{y}|}=2\frac{\sin(k|\vec{x}-\vec{y}|)}{k|\vec{x}-\vec{y}|}} \tag{C.12}$$

C.2 Some complex integration

Estimating $I(k, t-t')$ requires a little bit of contour integration in the complex plane. First, notice that $\Delta t = t - t' > 0$ because we are only interested in the causal Green function. Thus, the integral we want to evaluate is:

$$I(k,\Delta t) = \int_{-\infty}^{+\infty}\frac{\mathrm{e}^{-\mathrm{i}\omega\Delta t}}{(\omega+k)(\omega-k)}\mathrm{d}\omega\ . \tag{C.13}$$

The complex integrand thus reads:

$$f(z) = \frac{\mathrm{e}^{-\mathrm{i}\Delta t z}}{(z+k)(z-k)}\ . \tag{C.14}$$

It has two simple poles located on the real axis but is holomorphic otherwise. Thus, we will need to use an indented contour together with Cauchy's theorem. Besides, we need to close this contour while ensuring that for $z = R\mathrm{e}^{i\theta}$, $\mathrm{e}^{-\mathrm{i}\Delta t z} \to 0$ when $R \to +\infty$. Clearly $\mathrm{e}^{-\mathrm{i}\Delta t z} = \mathrm{e}^{-\mathrm{i}R\cos\theta}\mathrm{e}^{R\sin\theta\Delta t}$, so that we need to have $\sin\theta < 0$, i.e. we need to close in the lower half-plane. We choose the contour depicted in Fig. C.1: $\gamma = [-R, -k-\varepsilon]\cup\Gamma(-k,\varepsilon)\cup[-k+\varepsilon, k-\varepsilon]\cup\Gamma(k,\varepsilon)\cup[k+\varepsilon, R]\cup(-\Gamma(0,R))$. There is no singularity inside the contour, so we can apply Cauchy's theorem:

$$\int_\gamma f(z)\mathrm{d}z = 0\ . \tag{C.15}$$

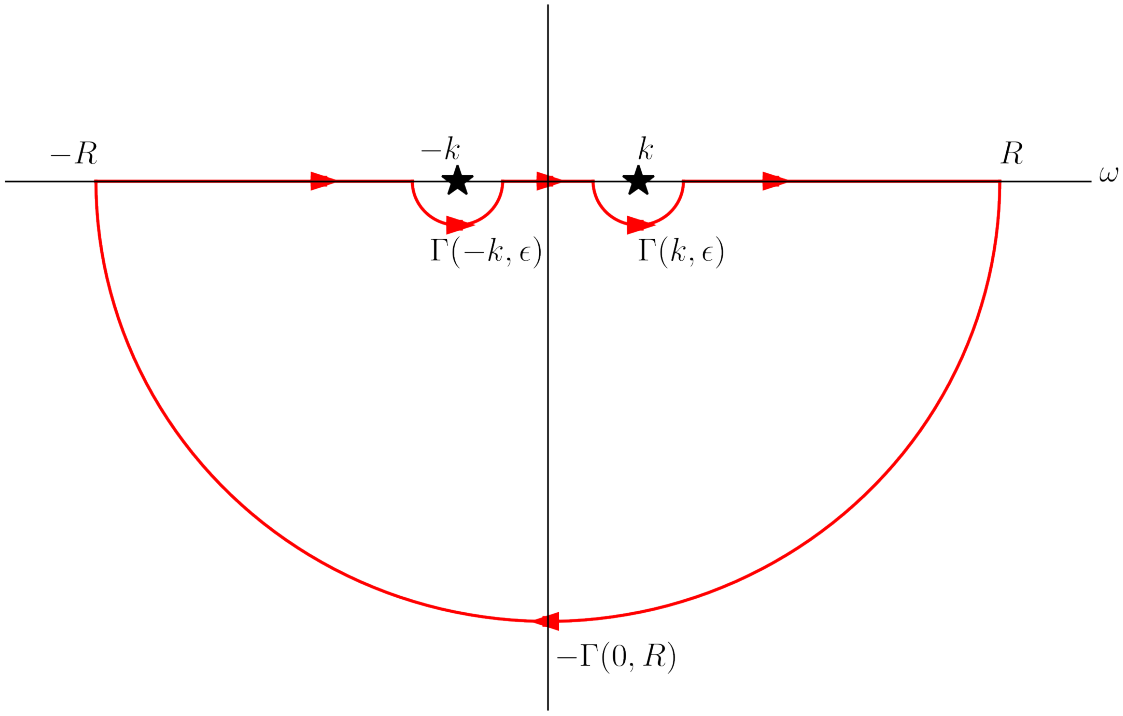


Figure C.1: Contour used to calculate the integral $I(k,\Delta t)$ in Eq. (C.13).

Splitting the integral by bits and taking the limits $\varepsilon \to 0$ and $R \to +\infty$:

$$\underbrace{\int_{[-R,-k-\varepsilon]} f(z)\mathrm{d}z}_{\to\int_{-\infty}^{-k} f(x\omega)\mathrm{d}\omega} + \underbrace{\int_{\Gamma(-k,\varepsilon)} f(z)\mathrm{d}z}_{\to \mathrm{i}\pi\mathrm{res}\{f(z),-k\}(2\pi-\pi)} + \underbrace{\int_{[-k-\varepsilon,k-\varepsilon]} f(z)\mathrm{d}z}_{\to\int_{-k}^{k} f(\omega)\mathrm{d}\omega} +$$
$$\underbrace{\int_{\Gamma(k,\varepsilon)} f(z)\mathrm{d}z}_{\to \mathrm{i}\pi\mathrm{res}\{f(z),k\}(2\pi-\pi)} + \underbrace{\int_{[k+\varepsilon,R]} f(z)\mathrm{d}z}_{\to\int_{k}^{+\infty} f(\omega)\mathrm{d}\omega} + \underbrace{\int_{-\Gamma(0,R)} f(z)\mathrm{d}z}_{\to 0} = 0 \tag{C.16}$$

$$I(k,\Delta t) - \mathrm{i}\pi\frac{\mathrm{e}^{\mathrm{i}k\Delta t}}{2k} + \mathrm{i}\pi\frac{\mathrm{e}^{-\mathrm{i}k\Delta t}}{2k} = 0\,. \tag{C.17}$$

Thus:

$$I(k,\Delta t) = -\pi\frac{\sin(k\Delta t)}{k}H(\Delta t)\,, \tag{C.18}$$

Where $H(x)$ is the Heaviside function: $H(x) = 1$ if $x > 0$ and 0 otherwise.

C.3 Final form

Finally, we can put everything together:

$$G\left(\boldsymbol{x}-\boldsymbol{y}\right) = -\frac{H\left(\Delta t\right)}{4\pi^2\left|\vec{x}-\vec{y}\right|}\int_0^{+\infty}\sin\left(k\Delta t\right)\sin\left(k\left|\vec{x}-\vec{y}\right|\right)\mathrm{d}k \tag{C.19}$$

$$= -\frac{H\left(\Delta t\right)}{4\pi^2\left|\vec{x}-\vec{y}\right|}\int_0^{+\infty}\left\{\cos\left[k\left(\left|\vec{x}-\vec{y}\right|-\Delta t\right)\right]-\cos\left[k\left(\left|\vec{x}-\vec{y}\right|+\Delta t\right)\right]\right\}\mathrm{d}k\ . \tag{C.20}$$

Besides:

$$\int_0^{+\infty}\cos(\alpha u)\mathrm{d}u = \frac{1}{2}\int_{-\infty}^{+\infty}\cos(\alpha u)\mathrm{d}u \tag{C.21}$$

$$= \frac{1}{4}\int_{-\infty}^{+\infty}\left(\mathrm{e}^{\mathrm{i}\alpha u}+\mathrm{e}^{-\mathrm{i}\alpha u}\right)\mathrm{d}u \tag{C.22}$$

$$= \frac{1}{4}\left(2\pi\delta^D(\alpha)+2\pi\delta^D(\alpha)\right) \tag{C.23}$$

$$= \pi\delta^D(\alpha)\ . \tag{C.24}$$

Thus:

$$G\left(\boldsymbol{x}-\boldsymbol{y}\right) = -\frac{1}{4\pi\left|\vec{x}-\vec{y}\right|}\left[\delta^D\left(\left|\vec{x}-\vec{y}\right|-\Delta t\right)-\delta^D\left(\underbrace{\left|\vec{x}-\vec{y}\right|+\Delta t}_{\neq 0}\right)\right]H(\Delta t) \tag{C.25}$$

$$= -\frac{1}{4\pi\left|\vec{x}-\vec{y}\right|}\delta^D\left(\left|\vec{x}-\vec{y}\right|-\Delta t\right)H(\Delta t)\ . \tag{C.26}$$

To conclude, the Green function of the d'Alembert operator in Minkowski spacetime reads:

Green function of d'Alembert operator in Minkowski spacetime

$$G\left(\boldsymbol{x}-\boldsymbol{y}\right) = -\frac{1}{4\pi\left|\vec{x}-\vec{y}\right|}\delta^D\left(\left|\vec{x}-\vec{y}\right|-\left(x^0-y^0\right)\right)H(x^0-y^0)\ . \tag{C.27}$$

Bibliography

[1] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016. 224

[2] Elcio Abdalla et al. Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies. *JHEAp*, 34:49–211, 2022. 300

[3] K. G. Begeman, A. H. Broeils, and R. H. Sanders. Extended rotation curves of spiral galaxies: Dark haloes and modified dynamics. *Mon. Not. Roy. Astron. Soc.*, 249:523, 1991. 287

[4] B. Bertotti, L. Iess, and P. Tortora. A test of general relativity using radio links with the Cassini spacecraft. *Nature*, 425:374–376, 2003. 201

[5] M. Betoule et al. Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples. *Astron. Astrophys.*, 568:A22, 2014. 289

[6] Max Born and Emil Wolf. *Principles of optics*. Cambridge University Press, 1999. 51

[7] Sean M. Carroll. *Spacetime and Geometry*. Cambridge University Press, 7 2019. iii

[8] C. Clarkson. Establishing homogeneity of the universe in the shadow of dark energy. *Comptes Rendus Physique*, 13:682–718, 2012. 262

[9] Albert Einstein. Näherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1916:688–696, 1916. 224

[10] Albert Einstein. Über Gravitationswellen. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1918:154–167, 1918. 224

[11] W. L. Freedman et al. Final results from the Hubble Space Telescope key project to measure the Hubble constant. *Astrophys. J.*, 553:47–72, 2001. 260

[12] Éric Gourgoulhon. *Special Relativity in General Frames*. Graduate Texts in Physics. Springer, Berlin, Heidelberg, 2013. iii

[13] J. B. Hartle. *Gravity: An introduction to Einstein's general relativity*. Pearson, 2003. iii

[14] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2 2011. iii

[15] M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General relativity: An introduction for physicists*. Cambridge University Press, 2006. iii

[16] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973. iii, 165

[17] Isaac Newton. *Philosophiæ Naturalis Principia Mathematica*. Translated by Andrew Motte (1845), England, 1687. 6

[18] Robert V. Pound and Glen A. Rebka, Jr. Apparent Weight of Photons. *Phys. Rev. Lett.*, 4:337–341, 1960. 77

[19] Norbert Straumann. *General Relativity*. Graduate Texts in Physics. Springer, Dordrecht, 2013. iii

[20] Pierre Touboul et al. Space test of the Equivalence Principle: first results of the MICROSCOPE mission. *Class. Quant. Grav.*, 36(22):225006, 2019. 71

[21] Robert M. Wald. *General Relativity*. Chicago University Press, Chicago, USA, 1984. iii, 165

[22] Clifford M. Will. The Confrontation between General Relativity and Experiment. *Living Rev. Rel.*, 17:4, 2014. 198