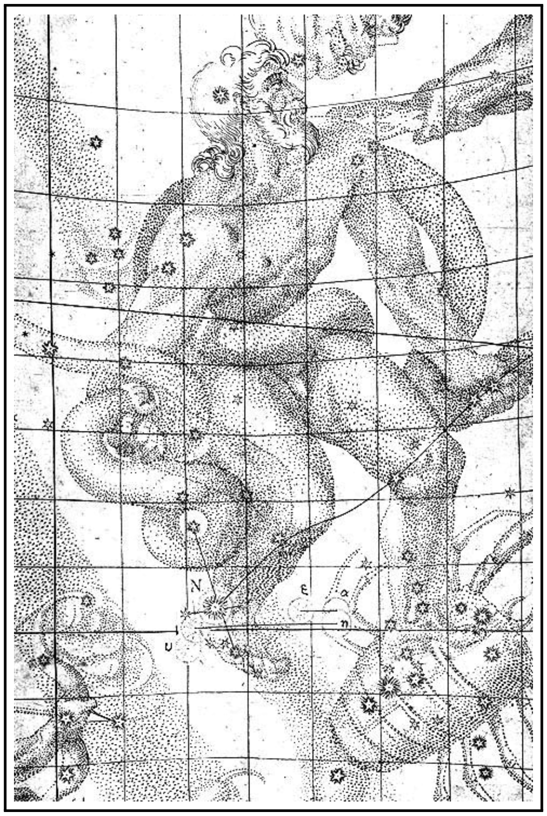


Lecture Notes on General Relativity

S. Hollands, Ko Sanders

These lecture notes were prepared by Stefan Hollands, Stefanie Riedel, and Ko Sanders using LaTeX 2ε.

Preface

These lecture notes on General Relativity intend to give an introduction to all aspects of Einstein's theory: ranging form the conceptual via the mathematical to the physical. In the first part we discuss Special Relativity, focusing on the re-examination of the structure of time and space. In the second part we cover General Relativity, starting with an introduction to the necessary mathematical tools and then explaining the structure of the theory, the conceptual motivations for it and its relation between spacetime and gravity. Finally, in the third part, we focus on physical applications of Einstein's theory at several length scales: gravitational waves, cosmology, black holes and GPS. Whereas the third application is the most practical, the first three are the best illustration for how General Relativity has influenced our understanding of the world around us.

Contents

I Special Relativity **6**

1 Introduction **6**

2 Time and Space in Classical Mechanics **7**

3 Electromagnetism and Poincaré Invariance **11**

4 Spacetime in Special Relativity **14**

5 Mathematics of Minkowski Spacetime **17**

6 Mechanics in Special Relativity **21**

7 Observer Dependence and Paradoxes **25**

II General Relativity **29**

8 Introduction **29**

9 Mathematical Preliminaries **30**
9.1 Calculus of Tensors 31
9.2 Manifolds 34
9.3 Tangent, Cotangent and Tensor Bundles 40
9.4 Covariant Derivatives 45
9.5 Metrics and Geodesics 50

10 General Relativity **58**
10.1 Kinematics of General Relativity 58
10.2 Dynamics of General Relativity 64
10.3 Properties of General Relativity 68
10.4 Stress-Tensor and Energy Conditions 71

III Applications of General Relativity **75**

11 Spacetime Symmetries **75**

12 Cosmological Solutions to Einstein's Equation (Maximally Symmetric Universes) **78**
12.1 Redshift in cosmological spacetimes 87
12.2 Particle Horizons 91

13 Black Holes **92**
13.1 Derivation of the Schwarzschild Solution 93
13.2 The redshift effect 95
13.3 Geodesics 96
13.4 Kruskal extension 101

14 Linearized Gravity and Gravitational Radiation **106**
14.1 Gravitational waves in empty space 107
14.2 Sources of gravitational waves 112

15 The Global Positioning System **120**
15.1 Introduction 120
15.2 Modelling the gravitational field of Earth 121
15.3 Choice of coordinates 123
15.4 Time measurement on a satellite 125
15.5 How to synchronise clocks on different satellites. 128
15.6 Communication with users on Earth 129
15.7 Conclusions 130

Part I
Special Relativity

1 Introduction

Time and space are two of the most basic concepts in describing the world around us. Even young children soon master the difference between "here" and "there", and between "now" and "later" or "earlier". However, strictly speaking, the concepts of time and space are only derived notions, which we introduce as a theoretical construct to help order our observations of events in the world around us.

Because our intuitions about time and space are so strong, it is hard to imagine a description of the world without them, or to explain in detail how those constructs come about. It should not be surprising, therefore, that time and space were long thought to have a fixed structure, which has been formalized by Newton, Galileo and Leibniz in the early days of modern science. These notions are now called "absolute time" and "absolute space" (or "absolute distance"). Immanuel Kant could not imagine how time and space could possibly have any other structure and he even went so far as to raise that structure to a fundamental ordering principle of human thought.

Interestingly, Newton himself noted that absolute time and distance were not entirely without problems. He illustrated this in the form of the "rotating bucket" paradox. However, this paradox did not seem to admit any experimental investigation, so it remained unresolved for a long time and didn't diminish the impression that the notions of absolute time and space were correct.

It should be considered a major achievement of human thought that the notions of absolute time and space were found to be faulty and that they could be replaced by something better: the concept of spacetime. It was only possible after Maxwell completed the unification of the theories of electricity and magnetism into a single theory of electromagnetism, based on Faraday's concept of a field, which avoids action at a distance. Maxwell's theory brought to light a new problem with the notions of absolute time and space: the theory was not invariant under all Galilean transformations. Einstein[1] found that this discrepancy could be resolved in the most elegant way

[1] Albert Einstein won the 1921 Nobel Prize in physics "for his services to theoretical

by noting that the concepts of absolute time and space were not well motivated in an operational way. In his Special Theory of Relativity he replaced them by the new concept of spacetime.

2 Time and Space in Classical Mechanics

For comparison with Special Relativity, it will be useful to give a fairly detailed analysis of the structure of time and space in the classical mechanics of Newton. Let us denote by M the set of all possible events in the universe. Here an event can be identified uniquely by a time and place, indicating when and where the event happens. We usually encode the time and place into numbers, which we call coordinates.

To find out when an event $p \in M$ happens, we may use a clock: when p happens we read off the time, $t(p)$. In this way the clock defines a function $t : M \to \mathbb{R}$. To understand the notion of absolute time, we need to understand how much of the number $t(p)$ depends on the properties of the clock we use, and how much of it is independent of those properties.

Let us first agree to use an ideal clock: it runs for ever and is not influenced by anything happening around it. (Of course such ideal clocks don't really exist, but they help to clarify the notion of absolute time.) Ideal clocks go a long way to exhibit the structure of absolute time, but two ideal clocks may still give different readings for two reasons:

1. Two ideal clocks can disagree which events happen at $t = 0$.

2. Two ideal clocks can use different units of time.

In other words, given an ideal clock $t : M \to \mathbb{R}$, there is no special physical significance to events p with $t(p) = 0$, or on the time difference $t(p_1) - t(p_2)$ for any two events p_1, p_2. However, all ideal clocks do agree on the following:

Tenet 2.1 (Absolute Time) *In classical mechanics, for arbitrary events $p_1, p_2, p'_1, p'_2 \in M$, all ideal clocks $t : M \to \mathbb{R}$ agree on*

1. *whether $t(p_1) > t(p_2)$, $t(p_1) < t(p_2)$ or $t(p_1) = t(p_2)$,*

2. *the value of $\frac{t(p_1)-t(p_2)}{t(p'_1)-t(p'_2)}$ when $t(p'_1) \neq t(p'_2)$.*

---

physics, and especially for his discovery of the law of the photoelectric effect".

The first property defines a time orientation on M: the notions of before, after and simultaneous are well defined for all events, independent of the ideal clock. In the second property, the denominator $t(p_1') - t(p_2')$ essentially fixes the units of time. When the units of time are fixed, all ideal clocks agree on the time interval between any two events. Together these two properties fully characterize absolute time. (Note in particular that we can compare the duration of any two time intervals, regardless of when or where the events are located.)

Exercise 2.2 *Let $t : M \to \mathbb{R}$ and $t' : M \to \mathbb{R}$ be time coordinates defined by ideal clocks. Show that $t'(p) = rt(p) + t_0'$ for some $r > 0$ and $t_0' \in \mathbb{R}$.*

In an analogous way one may identify places in M using ideal measuring rods, which are infinitely extended and which are not influenced by any physical processes happening in the universe. These ideal measuring rods can be used to form an infinite grid, where the rods intersect at right angles. The grid defines coordinate maps $\mathbf{x} : M \to \mathbf{R}^3$, where $\mathbf{x} = (x^1, x^2, x^3)$ in components.

For a given absolute time coordinate t we denote by $M_{t=c}$ the subset of M of all events taking place at time $t = c$. All these simultaneous events make up the entire space at the moment $t = c$. The notion of absolute space in classical mechanics can then be formulated as follows:

Tenet 2.3 (Absolute Space) *In classical mechanics, for any $c \in \mathbb{R}$ and any events $p_1, p_2, p_1', p_2' \in M_{t=c}$, all ideal measuring rods agree on the following:*

1. *the value of $\frac{\|\mathbf{x}(p_1) - \mathbf{x}(p_2)\|}{\|\mathbf{x}(p_1') - \mathbf{x}(p_2')\|}$, when $\|\mathbf{x}(p_1') - \mathbf{x}(p_2')\| \neq 0$,*

2. *$M_{t=c}$ is a three-dimensional Euclidean space, i.e. we may construct a grid of ideal measuring rods such that the distance between p_1 and p_2 satisfies Pythagoras' Theorem*

$$\|\mathbf{x}(p_1) - \mathbf{x}(p_1)\|^2 = \sum_{i=1}^{3} (x^i(p_1) - x^i(p_2))^2.$$

Again the origin of the coordinates $\mathbf{x}$ does not have any physical significance, nor does the orientation. The first property means that all ideal grids agree on the (absolute) distances, up to a change of units. The second property fixes the global shape of space $M_{t=c}$.

Exercise 2.4 *For fixed $c \in \mathbb{R}$, let $\mathbf{x} : M_{t=c} \to \mathbb{R}^3$ and $\mathbf{x}' : M_{t=c} \to \mathbb{R}^3$ be bijections, such that*

$$\frac{\|\mathbf{x}'(p_1) - \mathbf{x}'(p_2)\|}{\|\mathbf{x}'(p_1') - \mathbf{x}'(p_2')\|} = \frac{\|\mathbf{x}(p_1) - \mathbf{x}(p_2)\|}{\|\mathbf{x}(p_1') - \mathbf{x}(p_2')\|}$$

for all $p_1, p_2, p_1', p_2' \in M_{t=c}$ with $\|\mathbf{x}'(p_1')) - \mathbf{x}'(p_2'))\| \neq 0$. Show that

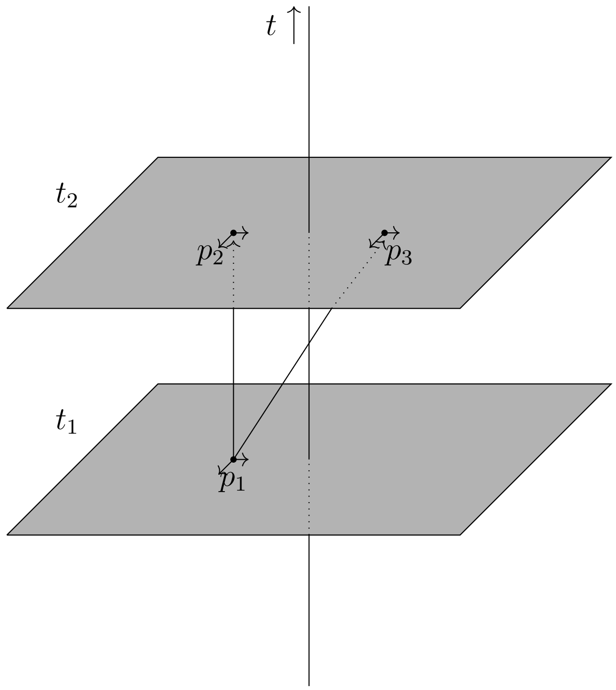
$$\mathbf{x}'(p) = rA \cdot \mathbf{x}(p) + \mathbf{x}_0'$$

for some $\mathbf{x}_0' \in \mathbb{R}^3$ and some orthogonal matrix A and some $r > 0$. (Hint: first identify $\mathbf{x}_0'$ and $r > 0$. Then consider the map $\psi : \mathbb{R}^3 \to \mathbb{R}^3 : X \mapsto r^{-1}(\mathbf{x}' \circ \mathbf{x}^{-1}(X) - \mathbf{x}_0')$. Using the real polarisation identity for the standard inner product, $2\langle X, Y\rangle = \|X\|^2 + \|Y\|^2 - \|X - Y\|^2$ for all $X, Y \in \mathbb{R}^3$, show that ψ must preserve the inner product, because it preserves all distances. Use this to prove that ψ is linear and to find A.)

We can use an ideal clock and an orthogonal grid of ideal measuring rods to define a coordinate system $(t, \mathbf{x}) : M \to \mathbb{R}^4$. In this way, each event in M is assigned a unique set of coordinates. Coordinates which are obtained in this way are called *classical inertial coordinates*. The following is a central tenet of classical mechanics:

Tenet 2.5 (Galilean Relativity Principle) *In all classical inertial coordinate systems, the physics of classical mechanics is described by Newton's laws in terms of absolute time and distance.*

In other words, the classical inertial coordinate systems form a preferred class of coordinate systems and Newton's laws are invariant under the change of coordinates from one classical inertial coordinate system to another. However, within classical mechanics, no classical inertial coordinate system is preferred over any other, simply because one cannot distinguish them by any experimental procedure satisfying Newton's laws. E.g., (approximate) classical inertial coordinates attached to the Earth are no better or worse than the (approximate) classical inertial coordinates attached to a train moving with a constant velocity in a fixed direction. Our natural intuition to favor the first coordinate system, which is more familiar, cannot be justified by classical mechanics.



Because classical inertial coordinate systems are defined by ideal clocks and rods, they are insensitive to any outside forces, so (by Newton's laws) they have zero acceleration with respect to each other. If we express the inertial coordinates $(t', \mathbf{x}')$ in the same physical units as the inertial coordinates $(t, \mathbf{x})$ (which may require a rescaling), then one can show that they must be related by a *Galilean transformation*:

$$t'(p) = t(p) + t'_0, \qquad \mathbf{x}'(p) = A \cdot \mathbf{x}(p) - t(p)\mathbf{v} + \mathbf{x}'_0 \tag{1}$$

for an orthogonal matrix A and a vector $\mathbf{v} \in \mathbb{R}^3$. Conversely any Galilean transformation applied to a classical inertial coordinate system yields a new classical inertial coordinates system.

Exercise 2.6 *Show that two classical inertial coordinate systems which employ the same units are related by a Galilean transformation with uniquely determined A, $\mathbf{v}$, $\mathbf{x}'_0$ and t'_0. (Hint: use the previous exercises and the fact that a trajectory in M which is linear in one classical inertial coordinate system is linear in every other classical inertial coordinate system, due to Newton's laws.)*

Exercise 2.7 *Show that the Galilean transformations form a group (i.e. the composition and the inverses of such transformations are Galilean transformations).*

Exercise 2.8 *Let* $(t, \mathbf{x})$ *be classical inertial coordinates and let* p_1 *and* p_2 *be the events such that* $\mathbf{x}(p_1) = \mathbf{x}(p_2)$*, but* $t(p_2) > t(p_1)$*. Does it make sense to say that the event* p_2 *happens at the same place as* p_1 *but at a later time? (I.e. would all inertial coordinate systems agree that* p_2 *is at the same place as* p_1*?) Do points of space preserve their identity in the course of time?*

Example 2.9 (Rotating bucket paradox) *Let* $(t, \mathbf{x})$ *be a classical inertial coordinate system and consider new coordinates* $(t', \mathbf{x}')$ *which rotate around a fixed axis with a constant angular velocity, e.g.:*

$$t'(p) = t(p), \qquad \mathbf{x}'(p) = R_{t(p)} \cdot \mathbf{x}(p),$$

where $R_{t(p)}$ *is a rotation around the* x^3*-axis over an angle* $t(p)$*. Note that* $(t', \mathbf{x}')$ *is not a classical inertial coordinate system. Expressing physical laws in the coordinates* $(t', \mathbf{x}')$ *leads to different formulae than in the coordinates* $(t, \mathbf{x})$ *(including a centrifugal force).*

Newton already noted that a bucket of water in an otherwise empty universe, which stands still w.r.t. the coordinates $(t', \mathbf{x}')$*, rotates w.r.t.* $(t, \mathbf{x})$*. The water in the bucket experiences a centrifugal force, which causes it to rise against the walls of the bucket. Without rotation, the water would remain flat. Note that the difference cannot be caused by any other matter in the universe, so it must be due to rotations w.r.t. empty space. This effect is surprising, because a rotation w.r.t. empty space is difficult to see. This problem is called "Newton's rotating bucket paradox".*

3 Electromagnetism and Poincaré Invariance

A first modification of our understanding of time and space was initiated when Maxwell completed the unification of the theories of electricity and magnetism. The theory of electromagnetism concerns an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$ on the spacetime M, which satisfy Maxwell's equations of motion. In vacuum these equations take the form

$$\begin{aligned} \nabla \times \mathbf{E} = -\partial_t \mathbf{B}, &\qquad \nabla \cdot \mathbf{B} = 0 \\ \nabla \cdot \mathbf{E} = 0, &\qquad \nabla \times \mathbf{B} = \frac{1}{c^2} \partial_t \mathbf{E}. \end{aligned} \tag{2}$$

expressed in inertial coordinates $(t, \mathbf{x})$ fixed to Earth.

One of the nice aspects of Maxwell's theory of electromagnetism is that it is a field theory[2], which implements the intuition that physical influences should propagate from point to neighbouring point. Indeed, Maxwell's equations imply that all components of $\mathbf{E}$ and $\mathbf{B}$ satisfy the wave equation:

$$\Box \mathbf{E} := -\frac{1}{c^2}\partial_t^2 \mathbf{E} + \Delta \mathbf{E} = 0, \tag{3}$$

where $\Delta := \sum_{i=1}^{3} \partial_{x^i}^2$ is the Laplace operator and $\Box$ is called the d'Alembert operator. This equation ensures that disturbances in the (components of the) fields can propagate no faster than a certain maximum speed c, the speed of light in vacuum.

Maxwell's equations have another property, however, which seems very puzzling in comparison to classical mechanics: the equations are not invariant under all Galilean transformations. Although there is no problem with translations or rotations in space, the equations are not invariant under uniform motion. In particular, the speed of light in vacuum follows from Maxwell's equations, so if these equations were invariant, all inertial coordinate systems would have to agree on this speed. However, for a coordinate system moving fast in the opposite direction of a light ray, one might expect the (relative) speed of that light ray to be bigger!

Exercise 3.1 *Consider a Galilean transformation to a coordinate system $(t', \mathbf{x}')$ which moves at a constant speed v in the x^1 direction. Show that the wave equation (3) transforms to the equation*

$$\Box' \mathbf{E} = v^2 \partial_{x'}^2 \mathbf{E} + 2v\partial_{t'}\partial_{x'}\mathbf{E},$$

where $\Box'$ is defined in a similar way as $\Box$, but with respect to the new coordinates. Show in a similar way that Maxwell's equations transform to

$$\begin{aligned} \nabla' \times \mathbf{E} = -\partial_{t'}\mathbf{B} - v\partial_{x'}\mathbf{B}, \qquad & \nabla' \cdot \mathbf{B} = 0 \\ \nabla' \cdot \mathbf{E} = 0, \qquad & \nabla' \times \mathbf{B} = \frac{1}{c^2}(\partial_{t'}\mathbf{E} + v\partial_{x'}\mathbf{E}). \end{aligned}$$

so Maxwell's equations are not invariant under this Galilean transformation. (This conclusion remains true, even if we argue that the components of $\mathbf{E}$

[2] The concept of a field is due to Faraday.

and $\mathbf{B}$ *are components of vectors, so they should also transform under the change of coordinates. We will see in Section 9.1 that several components of the fields should be unaffected by the change of coordinates, but the wave equation for those coordinates is not invariant.)*

The mathematical modifications needed to reconcile electromagnetism with absolute time and space were soon understood: Maxwell's equations are invariant under *Poincaré transformations*:

$$\begin{aligned} t'(p) &= \gamma\left(t(p) - c^{-2}v(A\cdot\mathbf{x}(p))_{\|}\right) + t'_0, \\ \mathbf{x}'(p) &= (A\cdot\mathbf{x}(p))_{\perp} + \gamma\left((A\cdot\mathbf{x}(p))_{\|} - t(p)v\right)\mathbf{n} + \mathbf{x}'_0 \\ \gamma &:= (1 - c^{-2}\|\mathbf{v}\|^2)^{-\frac{1}{2}}, \end{aligned} \tag{4}$$

where A is an orthogonal matrix, $\mathbf{v} = v\mathbf{n}$ a velocity vector with speed $v = \|\mathbf{v}\| < c$, and $_{\|}$ and $_{\perp}$ denote the components of a vector parallel or perpendicular to the unit vector $\mathbf{n}$. When $\mathbf{v} = 0$ the two transformations coincide, but for uniform motion this is not so. When the inhomogeneous terms t'_0 and $\mathbf{x}'_0$ vanish, we speak of a *Lorentz transformation*. Lorentz transformations for which $A = I$ are called *boosts* with relative velocity $\mathbf{v} \neq 0$.

Exercise 3.2 *Show that any Poincaré transformation can be written as a composition of an orthogonal transformation* A *of the spatial coordinates, a boost with relative velocity* $\mathbf{v}$ *and a translation in spacetime along* $(t'_0, \mathbf{x}'_0)$*. Show that* A*,* $\mathbf{v}$ *and* $(t'_0, \mathbf{x}'_0)$ *are uniquely determined.*

Exercise 3.3 *Show that the wave equation (3) is invariant under Poincaré transformations. (Hint: by the previous exercise it suffices to consider orthogonal transformations, boosts and translations.)*

Exercise 3.4 *Consider a boost with relative velocity* $\mathbf{v}$ *and find the* 4×4 *matrix* L *such that*

$$\begin{pmatrix} t' \\ \mathbf{x}' \end{pmatrix} = L \cdot \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix}.$$

Consider the limit $c \to \infty$ *with* $c^{-1}x^0$ *remaining finite and show that one recovers a Galilean transformation.*

One may argue that electromagnetism enables us to do experiments which go beyond classical mechanics and which allow us to distinguish certain inertial reference frames from others. However, there are several objections to this argument. Firstly, by Galilean Invariance the speed of light should differ for classical inertial coordinates moving in different directions, but the famous Michelson-Morley experiment[3] failed to detect such differences. Secondly, classical mechanics describes the motion of rigid bodies whose internal forces and collision forces are predominantly electromagnetic. In a uniformly moving frame at velocity v, Maxwell's equations take a different form, allowing us to detect v. It would be very remarkable if in some miraculous way the effect of electromagnetism on rigid bodies should cancel out completely to ensure that the Galilean Relativity Principle is satisfied.

Einstein realised that the conflict of classical mechanics with electromagnetism can be resolved most elegantly by noting that the notions of absolute time and distance are not operationally justified and need modification. E.g., let $p_1, p_2 \in M$ be two events which are supposedly simultaneous, but separated by a large distance. To verify that they really are simultaneous, one would need to be able to communicate between the places where the events are taking place at arbitrarily high speeds. Certainly no electromagnetic communication channel could satisfy this, because electromagnetic signals travel at a finite speed of propagation.

In his Special theory of Relativity, Einstein took electromagnetism as more fundamental than the notions of absolute time and distance, and he declared the speed of light in vacuum to be a universal physical constant. This forced him to weaken the assumptions on the properties of time and space, leading to the new concept of spacetime, which we investigate next.

4 Spacetime in Special Relativity

Let x^μ be coordinates on M, where $\mu = 0, \dots 3$ and $x^0 := ct$, and suppose that Newton's first law and Maxwell's equations are both valid in these coordinates. We call these coordinates (relativistic) *inertial coordinates*, or a (relativistic) *inertial frame*. (That such coordinates exist, to a high degree of accuracy, is known from experience.) Now let x'^μ be any new set of coordinates, obtained from x^μ by a Poincaré transformation. Then Maxwell's equations will also hold in the new coordinates, and so will many essential aspects of Newton's laws, because the change of coordinates is linear. In

[3]Albert Michelson won the 1907 Nobel Prize in physics "for his optical precision instruments and the spectroscopic and metrological investigations carried out with their aid".

particular, the straight paths of free particles will map to straight paths, which means that the new coordinates are also inertial coordinates. However, two sets of inertial coordinates may no longer agree on time differences between all events or on the question whether two events are simultaneous, let alone on the distance between simultaneous events, masses of particles or the strengths of forces. This means that Newtonian mechanics will need some modification.

In order to preserve the equations of electromagnetism and (much of) classical mechanics, Einstein proposed to remove absolute time and distance from Galileo's Relativity Principle:

Tenet 4.1 (Special Relativity Principle) *In all inertial coordinate systems, the physics of classical mechanics and of electromagnetism are described respectively by (an adaptation of) Newton's laws and Maxwell's equations. In addition, nothing moves faster than c, the speed of light in vacuum.*

Despite the absence of absolute time and distances, there are still some statements that all inertial coordinate systems agree on. To formulate these, we introduce a standard basis on $\mathbb{R}^4$, e_μ with $\mu = 0, \ldots, 3$, so we may write $x = \sum_{\mu=0}^{3} x^\mu e_\mu$. In addition we introduce a bilinear symmetric form on $\mathbb{R}^4$,

$$\eta(x, y) := -x^0 y^0 + (x^1 y^1 + x^2 y^2 + x^3 y^3), \tag{5}$$

which is called the *Lorentz inner product* (or rather, pseudo-inner product). Note that $\eta(x, x)$ is not always non-negative, but η is non-degenerate: when $\eta(x, y) = 0$ for all y, then $x = 0$. We say that $x, y \in \mathbb{R}^4$ are orthogonal when $\eta(x, y) = 0$.

We define the *spacetime interval* (in appropriate units) by

$$\sigma(p_1, p_2) := \eta(x(p_1) - x(p_2), x(p_1) - x(p_2)). \tag{6}$$

When a physical signal travels from event p_1 to event p_2, then

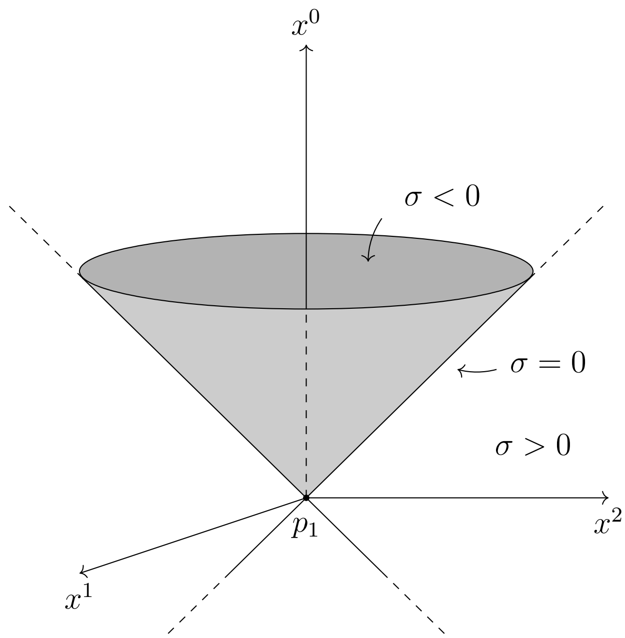
$$\begin{aligned} \sigma(p_1, p_2) = -(x^0(p_1) - x^0(p_2))^2 + \|\mathbf{x}(p_1) - \mathbf{x}(p_2)\|^2 &\leq 0, \\ x^0(p_2) - x^0(p_1) &\geq 0, \end{aligned}$$

because the signal cannot travel faster than the speed of light c. We have equality in the first line exactly when the signal travels at the speed of light and along a straight line.

Definition 4.2 *Two events p_1, p_2 are called* time-like*, resp.* light-like*, resp.* space-like related *when $\sigma(p_1, p_2) < 0$, resp. $\sigma(p_1, p_2) = 0$, resp. $\sigma(p_1, p_2) > 0$. Two events are called* causally related *when they are not space-like related ($\sigma(p_1, p_2) \leq 0$).*

When p_1, p_2 are causally related, p_1 lies to the future*, resp.* past *of p_2 when $x^0(p_1) \geq x^0(p_2)$, resp. $x^0(p_1) \leq x^0(p_2)$.*

Note that when p_1 and p_2 are causally related and $x^0(p_1) = x^0(p_2)$, then $p_1 = p_2$.



Tenet 4.3 (Causal Structure of Spacetime) *In Special Relativity, for any events $p_1, p_2, p_1', p_2' \in M$, all inertial coordinate systems agree on*

1. *whether p_1 and p_2 are time-like, light-like or space-like related,*
2. *whether p_1 lies to the future or past of p_2, when p_1, p_2 are causally related, and*
3. *the value of $\frac{\sigma(p_1,p_2)}{\sigma(p_1',p_2')}$, when the denominator is non-zero.*

The second statement means that all inertial coordinate systems agree on the time-ordering between any two causally related events. In the last statement, the denominator is once again used to fix units. Note that in Special Relativity it suffices to fix time units only, because corresponding spatial units are then obtained by multiplication with the universal constant c.

Exercise 4.4 *Show that Poincaré transformations preserve the spacetime interval σ and the causal structure of Tenet 4.3.*

Note that we have not defined simultaneity between events. In Special Relativity such a notion can only be defined in a coordinate dependent way.

One might worry that the choice of an inertial coordinate system x^μ is already more than can be justified in an operational way. E.g., if an observer remains at position $\mathbf{x} = 0$, how can he possibly ascribe coordinates to events with $\mathbf{x} \neq 0$, or synchronize watches at different places? The answers to such questions were elaborated by Einstein in an entirely operational way, using only light signals and their reflections. This means that inertial coordinates can be constructed by a procedure that is independent of the observer, and for any two observers the resulting sets of inertial coordinates are indeed related by a Poincaré transformation. For the details of these procedures we refer to interested reader to most text books on Special Relativity.

5 Mathematics of Minkowski Spacetime

Using inertial coordinates x^μ we identify M with $\mathbb{R}^4$ and the structure of spacetime is then entirely encoded in η. The pair $(\mathbb{R}^4, \eta)$ is called *Minkowski space* (or rather: *Minkowski spacetime*). We will now formulate the structure of M and of the Poincaré transformations systematically in this four-dimensional formulation, using the coordinates x^μ. Most of the results are formulated in terms of exercises.

The terminology for causal relations on spacetime M can also be used on $\mathbb{R}^4$, using the identification via inertial coordinates. E.g. we say that a vector $x \in \mathbb{R}^4$ is time-like when $\eta(x, x) < 0$. (This is equivalent to saying that the event in M corresponding to the coordinates x is time-like related to the event corresponding to the coordinates $0 \in \mathbb{R}^4$.) Similarly, x is future pointing when it is causal (i.e. $\eta(x, x) \leq 0$) and $x^0 \geq 0$. In addition we call the set of all light-light vectors the *light cone* and the set of all causal vectors the *causal cone*. These cones can be split up into the *forward* and *backward* cones (each containing the vector 0).

Exercise 5.1 *Give a sketch of $\mathbb{R}^4$, indicating the sets of space-like vectors and time-like vectors. Also indicate the forward and backward light cones and causal cones.*

Exercise 5.2 *Recall the standard basis e_μ of $\mathbb{R}^4$, such that $x = \sum_{\mu=0}^{3} x^\mu e_\mu$. Show that*

$$\eta(e_\mu, e_\nu) = \begin{cases} -1 & \mu = \nu = 0 \\ 1 & \mu = \nu \neq 0 \\ 0 & \mu \neq \nu \end{cases} .$$

Exercise 5.3 *Let $x \in \mathbb{R}^4$ be a time-like vector. Show that all vectors orthogonal to x (w.r.t. the Lorentz inner product η) are space-like or zero.*

Exercise 5.4 *Let $x, y \in \mathbb{R}^4$ be two future pointing causal vectors. Show that $\eta(x, y) \leq 0$, that $z := x + y$ is a future pointing causal vector and that $\eta(x, y) = 0$ if and only if z is light-like if and only if x and y are light-like and parallel to each other.*

Exercise 5.5 *Find all vectors in $\mathbb{R}^4$ which are orthogonal to themselves (w.r.t. the Lorentz inner product η).*

We will now focus on Poincaré transformations.

Exercise 5.6 *Show that any Poincaré transformation is of the form*

$$x'^\mu(p) = L \cdot x^\mu(p) + x_0'^\mu,$$

for a unique $x_0' \in \mathbb{R}^4$ and a unique 4×4-matrix L such that

$$\eta(L \cdot x, L \cdot y) = \eta(x, y)$$

for all $x, y \in \mathbb{R}^4$. Express L in terms of A and $\mathbf{v}$ of Ex. 3.2.

Exercise 5.7 *Show that the Poincaré transformations form a group.*

Exercise 5.8 *Show that two inertial coordinate systems which employ the same units are related by a Poincaré transformation.*

It is sometimes helpful to also introduce the standard Euclidean inner product $\langle x, y\rangle = \sum_{\mu=0}^{3} x^\mu y^\mu$. Any elements $x, y \in \mathbb{R}^4$ then have

$$\eta(x, y) = \langle x, \eta \cdot y\rangle$$

for the diagonal matrix

$$\eta := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{7}$$

Note, however, that the standard basis or inner product have no special physical significance. (They are not invariant under Poincaré transformations.)

The 4×4-matrices satisfying $\eta(L \cdot x, L \cdot y) = \eta(x, y)$ form a (Lie-)group $\mathcal{L}$, called the *Lorentz group*. Equivalently, it consists of all matrices satisfying $L^T \cdot \eta \cdot L = \eta$. Any element $L \in \mathcal{L}$ has $\det L = \pm 1$ and, expressed in the basis e_μ, the entry $L_{00} \neq 0$, because

$$-L_{00}^2 + \sum_{i=1}^{3} L_{i0}^2 = \langle L \cdot e_0, \eta \cdot L \cdot e_0\rangle = \langle e_0, \eta \cdot e_0\rangle = -1.$$

We may therefore decompose $\mathcal{L}$ into the four disjoint subsets (which are not necessarily subgroups)

$$\begin{aligned} \mathcal{L}_+^\uparrow &= \{L \in \mathcal{L} |\ \det L = +1,\ L_{00} > 0\}, \\ \mathcal{L}_-^\uparrow &= \{L \in \mathcal{L} |\ \det L = -1,\ L_{00} > 0\}, \\ \mathcal{L}_+^\downarrow &= \{L \in \mathcal{L} |\ \det L = +1,\ L_{00} < 0\}, \\ \mathcal{L}_-^\downarrow &= \{L \in \mathcal{L} |\ \det L = -1,\ L_{00} < 0\}. \end{aligned}$$

One may show that all these subsets are connected. We can go from any of these subsets to any other by multiplication with one of the following special elements:

- parity operation ("spatial reflection"): $L = -\eta$,
- time reversal operation: $L = \eta$,
- spacetime reflection: $L = -I$.

Exercise 5.9 *Show that $L \in \mathcal{L}$ defines a Lorentz transformation (which preserves the causal structure of Tenet 4.3) if and only if L is in the* orthochronous Lorentz group

$$\mathcal{L}^{\uparrow} := \mathcal{L}_{+}^{\uparrow} \cup \mathcal{L}_{-}^{\uparrow}.$$

Other subgroups of $\mathcal{L}$ are the *proper Lorentz group* $\mathcal{L}_{+} = \{L \in \mathcal{L}| \ \det L = +1\}$ and the *proper, orthochronous Lorentz group* $\mathcal{L}_{+}^{\uparrow}$. (The latter would have been the group of interest if we had also introduced an orientation on Minkowski spacetime, in addition to the time orientation.)

Exercise 5.10 *Let $x \in \mathbb{R}^4$ be a future pointing time-like vector. Show that there is a matrix $L \in \mathcal{L}$ such that $L \cdot x = e_0$, the standard basis vector which is future pointing and time-like.*

In the remaining part of this section we give some exercises about boosts, which illustrate the sometimes counter-intuitive consequences of the absence of absolute time.

Exercise 5.11 *Show that a boost with relative velocity $\mathbf{v} = ve_1$ is given by the matrix $L \in \mathcal{L}$ such that*

$$L = \begin{pmatrix} \cosh(\theta) & -\sinh(\theta) & 0 & 0 \\ -\sinh(\theta) & \cosh(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and $\mathbf{v} = c \tanh(\theta) e_1$.

The parameter θ is called the *rapidity* of the boost. Another often used notation is $\beta = c^{-1}\mathbf{v}$, or $\beta = c^{-1}\|\mathbf{v}\|$.

Exercise 5.12 *Consider two boosts L_i, $i = 1, 2$, in the same direction with rapidities θ_i. Show that $L_3 := L_1 \cdot L_2$ is another boost with rapidity $\theta_3 = \theta_1 + \theta_2$. Show that the corresponding velocities $v_i = c \tanh(\theta_i)$ satisfy the relativistic velocity addition theorem*

$$v_3 = \frac{v_1 + v_2}{1 + \frac{v_1 v_2}{c^2}}.$$

6 Mechanics in Special Relativity

Let us now explain how classical mechanics can be reformulated in the context of Special Relativity. The trajectory that any point-like object follows in the course of time forms a curve in the spacetime M, which is called a *world line*. We begin this section by considering curves in M.

Let $I = (a, b) \subset \mathbb{R}$ be an open interval and consider a (parameterized) curve $\xi : I \to M$. We choose some inertial coordinate system x^μ to make an identification $X : M \to \mathbb{R}^4$, so that $X(\xi(s)) = \sum_{\mu=0}^{3} \xi^\mu(s) e_\mu$ with coordinates $\xi^\mu := x^\mu \circ \xi$. We also assume that the ξ^μ are C^1 (continuously differentiable) and we denote the derivatives w.r.t. the parameter $s \in I$ by a dot, e.g. $\dot{\xi}^\mu(s)$. We will also write $\dot{\xi}(s) := \sum_{\mu=0}^{3} \dot{\xi}^\mu(s) e_\mu$. Curves are a natural way to model the trajectories of particles or other physical objects.

We may distinguish several kinds of curves, which have a definite causal character and time orientation:

Definition 6.1 *A C^1 curve $\xi : I \to M$ is called* space-like*, resp.* light-like*, resp.* time-like*, resp.* causal*, when for all $s \in I$, $\eta(\dot{\xi}(s), \dot{\xi}(s)) > 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) = 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) < 0$, resp. $\eta(\dot{\xi}(s), \dot{\xi}(s)) \leq 0$.*

A causal C^1 curve ξ is future*, resp.* past directed *when for all $s \in I$, $\dot{\xi}^0(s) \geq 0$, resp. $\dot{\xi}^0(s) \leq 0$.*

Now consider a C^1 bijection $s : I' \to I$ with a C^1 inverse, where $I' = (a', b')$. We call the curve $\xi' : I' \to M$ defined by $\xi'(s') := \xi(s(s'))$ a reparametrization of ξ. The image of ξ' coincides with the image of ξ. The speed with which this image is traced out may differ, but the velocity vectors are always parallel:

$$\dot{\xi}'^\mu(s') = \dot{\xi}^\mu(s(s'))\partial_{s'} s(s').$$

We say that ξ' has the same direction as ξ when $s : I' \to I$ preserves the orientation, i.e. when $s(a') < s(b')$. Otherwise we say that ξ' has the opposite direction as ξ.

Exercise 6.2 *Show that the causal character (space-like, light-like, time-like or causal) of a C^1 curve is independent of the choice of the parametrization. Show that the time-orientation of a causal curve is preserved when the reparametrization preserves the direction of the curves.*

Exercise 6.3 *All curves in this exercise are assumed to have C^1 coordinates:*

1. *Find a curve which is neither space-like nor causal.*
2. *Find a spacelike curve* $\xi : (-2, 2)$ *such that* $\xi(1)$ *lies to the future of* $\xi(-1)$.
3. *Find a causal curve which is neither light-like not time-like.*
4. *Find a causal curve which is neither future nor past directed.*
5. *Show that every time-like curve is either future or past directed.*

In analogy to the length of a curve in Euclidean geometry, we can define the *arc length* of a space-like C^1 curve by:

$$l(\xi) := \int_I \sqrt{\eta(\dot{\xi}(s), \dot{\xi}(s))} \mathrm{d}\, s$$

For a causal C^1 curve we similarly define the *proper time* to be

$$\tau(\xi) := \frac{1}{c} \int_I \sqrt{-\eta(\dot{\xi}(s), \dot{\xi}(s))} \mathrm{d}\, s.$$

Using the arc length and proper time we may find preferred parametrizations for all causal and space-like C^1 curves:

Definition 6.4 *A* C^1 *space-like curve is* parameterized by arc-length *when* $\eta(\dot{\xi}, \dot{\xi}) = 1$.

A C^1 *time-like curve is* parameterized by proper time *when* $\eta(\dot{\xi}, \dot{\xi}) = -c^2$.

Theorem 6.5 *A space-like curve* ξ *can always be parameterized by arc-length, without changing its direction. A time-like curve* ξ *can always be parameterized by proper time, without changing its direction (or time-orientation). These parametrizations are independent of the choice of inertial coordinates (up to a choice of units).*

Proof: In the space-like case we may choose $s'(s)$ such that $\partial_s s'(s) = \sqrt{\eta(\dot{\xi}(s), \dot{\xi}(s))}$ and in the time-like case such that $\partial_s s'(s) = \frac{1}{c}\sqrt{-\eta(\dot{\xi}(s), \dot{\xi}(s))}$. In this way we find orientation preserving changes of parameter and it is straight-forward to check that they implement the desired conditions. Independence of the choice of inertial coordinates follows from the independence of the spacetime interval (up to a choice of units). $\square$

Exercise 6.6 *Show that the arc length and proper time are independent of the choice of coordinates x^μ and of the parametrization of ξ. This shows that the parametrizations by arc length or proper time are independent of the choice of inertial coordinates.*

When a future directed time-like curve $\xi(\tau)$ is parameterised by proper time, we may define its velocity and acceleration as:

$$\begin{aligned}\dot{\xi}(\tau) &:= \sum_{\mu=0}^{3} \dot{\xi}^\mu(\tau) e_\mu, \\ \ddot{\xi}(\tau) &:= \sum_{\mu=0}^{3} \ddot{\xi}^\mu(\tau) e_\mu.\end{aligned}$$

The expressions on the left are independent of the choice of inertial coordinates, if we let a change of coordinates also affect the basis e_μ. By the *(rest) mass* m_0 of a particle we mean the mass that it has in an inertial coordinate system where it is at rest, i.e. a frame where it maintains a fixed position in space, so that it follows the trajectory $\tau \mapsto (c\tau, \mathbf{x}_0)$ for some fixed $\mathbf{x}_0$. When such a particle traverses a general, future directed, time-like curve ξ, we define its *energy-momentum* vector to be

$$P^\mu(\tau) := m_0 \dot{\xi}^\mu(\tau), \tag{8}$$

where τ is the proper time of the curve. In Special Relativity, Newton's first law then takes the form

$$F^\mu(\tau) = \partial_\tau P^\mu(\tau) \tag{9}$$

Let us now fix any choice of inertial coordinates and express the formulae above in terms of the coordinate time x^0, rather than the proper time τ. If a curve is time-like, then $\dot{\xi}^0(s) \neq 0$ for all $s \in I$ and (after reversing the orientation of the parameter, if necessary) we may assume that $\dot{\xi}^0(s) > 0$. We may then introduce a new parameter t, defined by $t(s) := c^{-1} \int_{s_0}^{s} \dot{\xi}^0$ for some $s_0 \in I$. (The map $s \mapsto t(s)$ will be an orientation preserving diffeomorphism on I, because $\dot{\xi}^0 > 0$.) The new parametrization leads to

$$\dot{\xi}^0(s) = \partial_s \xi'^0(t(s)) = \dot{\xi}'^0(t)\dot{t}(s)$$

and hence $\dot{\xi}'^0(t) = c$. This means that, after changing parametrization, we may assume that $\xi^0(t) = ct$ and $\xi^\mu(t) = (ct, \mathbf{x}(t))$.

The parameter derivative is now the same as the time derivative w.r.t. the inertial time coordinate $t = c^{-1}x^0$:

$$\dot{\xi}^\mu(t) = (c, \dot{\mathbf{x}}(t)) =: (c, \mathbf{v}(t)),$$

The formula above then shows that (in an arbitrary choice of inertial coordinates) the speed of the time-like curve satisfies $\|\mathbf{v}(t)\| < c$. This shows that time-like, future pointing curves model the trajectories of massive particles (also in the presence of forces). When ξ is a light-like curve (but $\dot{\xi} \neq 0$) we find in a similar way that $\|\mathbf{v}(t)\| = c$, which models the trajectory of a massless particle or light ray. Similarly, space-like curves may have speeds $\|\mathbf{v}(t)\| > c$, or even infinite speeds. Such curves do not model any known matter in the universe. Hypothetical particles that move faster than light are called *tachyons*.

A comparison of the parametrisation $\xi^\mu(t)$ with the parametrisation by proper time yields:

$$\begin{aligned}
\dot{\xi}^\mu(\tau) &= \partial_\tau t(\tau)\dot{\xi}^\mu(t(\tau)) = \gamma(\tau)\ \left(c, \mathbf{v}(t(\tau))\right), \\
\partial_\tau t(\tau) &= \gamma(\tau) := (1 - c^{-2}\|\mathbf{v}(t(\tau))\|^2)^{-\frac{1}{2}},
\end{aligned}$$

where the second line follows from the normalisation $c^2 = -\eta(\dot{\xi}^\mu(\tau), \dot{\xi}^\mu(\tau))$. It follows that

$$P^\mu(\tau) = m_0\gamma(\tau)\dot{\xi}^\mu(t(\tau)) = m_0\gamma(\tau)\ \left(c, \mathbf{v}(t(\tau))\right) =: (c^{-1}E(t(\tau)), \mathbf{P}(t(\tau))),$$

so that spatial components of the energy-momentum vector $P^\mu(\tau)$ in the given inertial frame take the form $\mathbf{P}(t) = m(t)\mathbf{v}(t)$, where the apparent (or inertial) mass in this coordinate frame is $m(t) = \gamma(\tau(t))m_0$. (Note that a high velocity implies a high apparent mass.) The energy in the given inertial frame satisfies

$$E(t) = \gamma(\tau(t))m_0c^2 = m_0c^2 + \frac{m_0\|\mathbf{v}(t)\|^2}{2} + \ldots,$$

where we made a Taylor expansion in $c^{-1}\|\mathbf{v}\|$. From this we see that the total energy in the given coordinate frame consists not only of the kinetic energy, but also a contribution from the rest mass m_0c^2 and higher order corrections. Note that the total energy is frame dependent: in the rest frame of the particle, only the first term remains.

Taking a further derivative we find

$$\begin{aligned} F^\mu(\tau) &= \partial_\tau P^\mu(\tau) = \gamma(\tau)\ \left(c^{-1}\partial_t E(t(\tau)), \partial_t \mathbf{P}(t(\tau))\right) \\ &= \gamma(\tau)\ \left(c^{-1}\partial_t E(t(\tau)), \mathbf{F}(t(\tau))\right), \end{aligned}$$

where $\mathbf{F}(t) = \partial_t \mathbf{P}(t)$ corresponds to the force appearing in Newton's second law (with a time-dependent mass $m(t)$) in the given inertial frame.

7 Observer Dependence and Paradoxes

To conclude we present some examples of effects in Special Relativity which are counter-intuitive and whose resolution relies on the fact that some of the quantities used are defined in a coordinate dependent way.

Throughout this section we will consider two sets of inertial coordinates, x^μ and x'^μ, related by a boost in the x^1-direction with relative speed v.

Length Contraction In the coordinate system x^μ we may use x^0 to measure time and $\mathbf{x}$ to measure distances and we consider a plank of length $l > 0$, whose endpoints at $x^0 = s$ are located at $(s, 0, 0, 0)$ and $(s, l, 0, 0)$, respectively, so the plank is at rest. In the coordinates x'^μ, the left and right endpoints are located at $(s\cosh(\theta), -s\sinh(\theta), 0, 0)$ and $(s\cosh(\theta) - l\sinh(\theta), l\cosh(\theta) - s\sinh(\theta), 0, 0)$, respectively, where θ is the rapidity of the boost.

If we now use x'^0 to measure time and $\mathbf{x}'$ to measure distances, then we can find the endpoints of the plank at a fixed time $x'^0 = s'$ as follows. For the left endpoint we set $s' = s\cosh(\theta)$, which leads to

$$(s\cosh(\theta), -s\sinh(\theta), 0, 0) = (s', -s'\tanh(\theta), 0, 0).$$

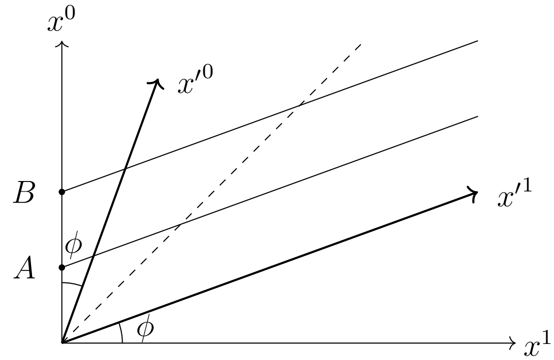
For the right endpoint we set $s' = s\cosh(\theta) - l\sinh(\theta)$, which leads to $s = s'\cosh(\theta)^{-1} + l\tanh(\theta)$ and hence

$$\begin{aligned} & (s\cosh(\theta) - l\sinh(\theta), l\cosh(\theta) - s\sinh(\theta), 0, 0) \\ = \ & (s', l\cosh(\theta)^{-1} - s'\tanh(\theta), 0, 0). \end{aligned}$$

In the primed coordinate system, the plank moves with a constant speed $v = c\tanh(\theta)$ in the negative x'^1-direction, as expected, but the length of the plank is

$$l' = l\cosh(\theta)^{-1} = l\gamma^{-1} < l. \tag{10}$$

This illustrates that length is a coordinate dependent notion. The plank is longest in the inertial coordinates in which it is at rest. The effect that in a boosted inertial frame all lengths in the direction of the boost are reduced, is called *length contraction*.[4] (Note that this effect is mutual: a plank which is at rest in the frame x'^μ will also appear contracted in the frame x^μ.)



Remark 7.1 *In many ways length contractions by boosts are very similar to the following situation: when a plank of length l, which lies along the x^1-axis, is rotated along the x^3-axis, say, then its projection onto the x^1-axis will have a contracted length.*

Time Dilation Consider an observer O who is at rest in the x^μ coordinate system and located at $\mathbf{x} = 0$, so his world line is $t \mapsto (ct, 0, 0, 0)$ and who carries a clock that measure the (proper) time $t = c^{-1}x^0$. We will use the fact that O can determine the coordinates x^μ of any event using operational

[4]Before Einstein proposed Special Relativity, Lorentz and Fitzgerald had already proposed that objects undergo a length contraction in their direction of motion. However, their interpretation was rather different: They assumed the existence of absolute time and space and, in addition, that a particular classical inertial frame can be singled out by the existence of a substance called ether, which pervades all space and is static in this frame. Length contractions were argued to be a physical process, caused by the motion w.r.t. the ether, and the underlying mechanisms were sought in electromagnetism. A consistent treatment of this idea leads to a theory which makes exactly the same predictions as Special Relativity, but which has a number of superfluous concepts that have no operational meaning: ether, absolute time and absolute space.

procedures (involving sending light rays), so he may use x^0 as a global time coordinate.

Now we consider a similar observer O' in the x'^μ coordinate system, so that each observer sees the other one moving away with a speed v. Let $A = (ct_1, 0, 0, 0)$ and $B = (ct_2, 0, 0, 0)$ be two events in the coordinates x^μ, which differ by a time interval $(t_2 - t_1)$ according to O. We may express these events in the coordinate frame x'^μ as

$$A = (ct_1 \cosh(\theta), -ct_1 \sinh(\theta), 0, 0), \qquad B = (ct_2 \cosh(\theta), -ct_2 \sinh(\theta), 0, 0)$$

in the primed coordinate system. According to O', the two events are therefore separated by a time interval

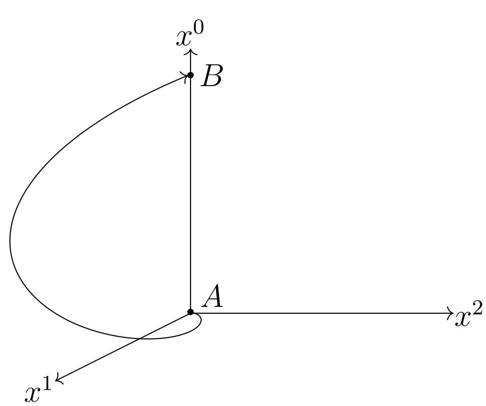
$$T' = \cosh(\theta)(t_2 - t_1) = \gamma T, \tag{11}$$

where $T = t_2 - t_1$. According to O', more time has elapsed between the two events, so O's clock is slow. In a similar way, O will find that the clock that O' uses is slow!

This effect is called *time dilation*. The fact that both observers find the other observer's clock to be slow violates the intuition that one clock must be faster than the other. However, this intuition is based on the false assumption that there exists an absolute time (and that both clocks run at a fixed rate compared with absolute time).

Exercise 7.2 *Consider an observer O', who starts running at event $A = (0, 0, 0, 0)$ along a curve of the form*

$$\chi(t) := (ct, 1 - \cos(\omega t), \sin(\omega t), 0).$$



For what values of ω is this curve time-like? For what values of ω is O' also present at the event $B = (ct_1, 0, 0, 0)$? What is the proper length of this curve between A and B? For what values of ω is this proper time less than half of the proper time t_1 along the straight line $t \mapsto (ct, 0, 0, 0)$? How fast must O' run to age only half as fast?

The Twin Paradox Consider the same Observer O in the coordinate system x^μ and an observer O' traveling through $A = (ct_1, 0, 0, 0)$ to $B = (ct_2, 0, 0, 0)$ via any C^1 time-like and future pointing curve $\xi : I \to M$. Then the following holds:

Theorem 7.3 *Let $\tau_\xi(A, B)$ be the proper time interval between A and B along ξ. Then $\tau_\xi(A, B) \leq (t_2 - t_1)$, with equality if and only if the (proper) velocity satisfies $\dot{\xi}^\mu \equiv (c, 0, 0, 0)$ between A and B.*

Proof: Because ξ is time-like and future pointing we have $\dot{\xi}^0(s) > 0$ for all $s \in I$, so we may change the parametrisation such that $t = c^{-1}x^0$ becomes the new parameter and $\xi^\mu(t) = (ct, \mathbf{X}(t))$. The proper time interval between A and B is then

$$
\begin{aligned}
\tau_\xi(A, B) &= c^{-1} \int_{t_1}^{t_2} \sqrt{c^2 - \|\dot{\mathbf{X}}(t)\|^2} \mathrm{d}\, t \\
&\leq c^{-1} \int_{t_1}^{t_2} c = (t_2 - t_1),
\end{aligned}
$$

which proves the estimate. Note that we have equality if and only if $\dot{\mathbf{X}} \equiv 0$, which means that $\dot{\xi}^\mu = (c, 0, 0, 0)$. In that case t is the proper time coordinate along ξ and we have $\xi^\mu(t) = (ct, 0, 0, 0)$, so ξ does indeed go from A to B. $\square$

Theorem 7.3 is completely analogous to the familiar result in Euclidean space that two points A and B can be connected by curves of different lengths, and there is a unique shortest curve. However, in the present context, the interpretation of the theorem is that the observer O, who followed the linear curve, has aged more than the observer O', who followed the curve ξ. This exhibits a violation of the idea of absolute time. The effect is most striking when O' first travels away from O at a constant speed and then turns around to return. ξ is then also a linear curve most of the way, so in view of time dilation, which is symmetric, it is then paradoxical that both observers agree on the fact that O has aged more than O'.

The resolution of the paradox lies in the fact that we cannot (in general) choose an inertial frame in which O' is at rest all the way, so there is no symmetry between O and O'. We will see in Section 9.5 that the essential aspect of Theorem 7.3 is that the linear curve that O follows is a geodesic, whereas ξ is not, in general. (Using this idea one may reformulate Theorem 7.3 in a way that is independent of the choice of inertial coordinates: all that matters is that O follows the linear curve between A and B in any inertial coordinate frame.)

In the case where ξ is linear most of the time, at least two inertial coordinate frames are relevant. In this setting it is tempting to use the time coordinates of those inertial frames globally, to find out "when" the aging of O with respect to O' takes place. Some interpretations ascribe this aging to the acceleration of O' at the turning point. However, the main issue is that this question is ill-posed, because the term "when" makes no global sense in Special Relativity, especially not when several inertial frames are involved. Indeed, it is a non-trivial issue for O' to compute inertial (time) coordinates for events taking place elsewhere (i.e. not on his world line) and the coordinates that he computes (using light signals) depend on his state of motion. The change of inertial frame by O' thus forces a change in the notion of simultaneity, which causes a jump in the age of O, as computed by O'.

Part II
General Relativity

8 Introduction

Special Relativity was already a wonderful scientific revelation, which unified Newtonian mechanics and electromagnetism by weakening the assumed intrinsic structure of space and time. However, Newton's theory of gravitation does not fit into this new framework, because it does not behave well under Poincaré transformations. Indeed, Newtonian gravity involves an instantaneous action at a distance, but when the notion of simultaneity at non-zero distances is no longer available, such an action makes no sense anymore. Newton's theory had been criticised before for its action at a distance, notably by Descartes. Descartes had proposed an alternative theory of gravity, based on vortices, which, however, contradicted empirical evidence.

It was not until Einstein formulated his General Theory of Relativity that Newton's theory was replaced by a field theory of gravity. Moreover, with General Relativity Einstein achieved a variety of other deep and subtle goals. It drops e.g. the assumption that the set M of all events should admit a bijection onto $\mathbb{R}^4$. It also explains why the (heavy) mass appearing in Newton's law of gravitation is the same as the (inertial) mass that appears in his first law of mechanics (the *weak equivalence* principle). This equivalence means that the trajectory of a freely falling body is completely determined by its initial position and velocity and it is independent of the object's mass or shape. In General Relativity this is explained by the fact that these preferred free-fall trajectories are a part of the structure of the spacetime M.

Note that all bodies are influenced by gravity, so it is now inconceivable that one may produce the idealized clocks and measuring rods which make up the inertial coordinate systems of Special Relativity. Instead, General Relativity will treat all coordinate systems on an equal footing. Only on very small scales, where variations in the gravitational field can be neglected, can we consider inertial coordinates, which are associated to freely falling clocks and rods. As before we require that all such local inertial frames are equivalent, i.e. the outcome of any local experiment in a freely falling laboratory is independent of the initial position and velocity of the laboratory. (Together with the weak equivalence principle, this forms the *strong equivalence principle.*)

The important conceptual step that makes all this possible is to turn the background structure of spacetime, which determines the spacetime intervals, into a dynamical structure, like a physical field, which must satisfy Einstein's equation of motion. This conceptual change also resolves Newton's "rotating bucket" paradox: the bucket is not rotating with respect to empty space, but with respect to the gravitational field, which is a physical quantity in its own right.

9 Mathematical Preliminaries

In this section we will present the mathematical tools needed to formulate General Relativity. These mathematical tools had been developed before General Relativity, in particular by Riemann (building on earlier work by

Gauss). Most of these tools deal with the idea that the set of all events M may no longer be identifiable with $\mathbb{R}^4$, which requires the mathematical theory of manifolds (i.e. differential geometry). The idea that there should be a field which tells us how to measure time intervals and distances locally, requires the theory of (pseudo)-Riemannian manifolds. First, however, we will describe some notations for tensors, that will make the subsequent developments run more smoothly.

9.1 Calculus of Tensors

Let V be a finite dimensional real vector space of dimension $n \in \mathbb{N}$ and let $\{e_1, \ldots, e_n\}$ be a basis for V. Then each vector $v \in V$ can be written in a unique way as $v = \sum_{\mu=1}^{n} v^\mu e_\mu$ and we call the real numbers v^μ the components of v in the basis $\{e_\mu\}$. To simplify our notations we introduce the following convention:

Convention 9.1 (Einstein's Summation Convention) *Whenever an expression contains an index that appears once as a superscript and once as a subscript, then a summation over the range of this index (e.g.* $1, \ldots, n$*) is implied, unless explicitly stated otherwise.*

This means that we may write $v = v^\mu e_\mu$, dropping the summation symbol.

Now let V^* be the *dual vector space* of V, i.e. the vector space of all linear maps $\omega : V \to \mathbb{R}$, where the vector space structure is given by pointwise addition and multiplication:

$$(\lambda_1\omega_1 + \lambda_2\omega_2)(v) := \lambda_1\omega_1(v) + \lambda_2\omega_2(v).$$

There is a natural basis of V^* which is dual to e_μ, namely $e^{*1}, \ldots, e^{*n}$ such that

$$e^{*\mu}(e_\nu) = \delta^\mu{}_\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases} \tag{12}$$

Because any $\omega \in V^*$ is uniquely determined by its values on the basis vectors e_μ, we see that $\{e^{*\mu}\}$ is indeed a basis and that V^* also has the dimension n. We may write $\omega = \omega_\mu e^{*\mu}$ with unique components ω_μ in the basis $\{e^{*\mu}\}$. Note that

$$\omega(v) = \omega_\mu v^\nu e^{*\mu}(e_\nu) = \omega_\mu v^\mu,$$

because the bases of V and V^* are dual to each other.

Exercise 9.1 *Let V^{**} be the double dual vector space, i.e. the dual vector space of Vv. Show that there is a linear map $\iota : V \to V^{**}$ defined by $\iota(v)(\omega) = \omega(v)$ for all $\omega \in V^*$. Moreover, show that ι is an isomorphism of vector spaces, which is independent of any choice of basis for V. (For this reason it suffices to consider V and V^* and no further duals are needed.)*

By a *tensor T of type* (k, l) we will mean a multi-linear map[5]

$$T : \underbrace{V^* \times \ldots \times V^*}_{k \text{ times}} \times \underbrace{V \times \ldots \times V}_{l \text{ times}} \to \mathbb{R},$$

i.e. a map $T(\omega_1, \ldots, \omega_k, v_1 \ldots, v_l)$ which is linear in each ω_i and v_j when all other arguments are fixed. The space of all such tensors is denoted by

$$\underbrace{V \otimes \ldots \otimes V}_{k \text{ times}} \otimes \underbrace{V^* \otimes \ldots \otimes V^*}_{l \text{ times}}$$

and it carries a natural vector space structure (by pointwise addition and scalar multiplication).

A tensor of type $(0, 1)$ is an element of V^*, whereas a tensor of type $(1, 0)$ is an element of $V^{**} \simeq V$. In this way tensors generalize vectors and dual vectors. Note that the tensor product $V \otimes V$ has a natural basis determined by the e_μ, namely $e_\mu \otimes e_\nu$ with $\mu, \nu = 1, \ldots, n$. (The number of such vectors is n^2, which is indeed the dimension of $V \otimes V$.) Any element of $w \in V \otimes V$ can therefore be written in terms of components as $w = w^{\mu\nu} e_\mu \otimes e_\nu$. Similarly, for $L \in V \otimes V^*$ we may write $L = L^\mu{}_\nu e_\mu \otimes e^{*\nu}$ and an analogous expansion into components works for arbitrary tensors.

Of course the components of tensors depend heavily on the choice of basis e_μ. Nevertheless, it is possible to write down expressions which are largely independent of this choice of basis. To see how this goes we will now consider how a change of basis acts on the various components.

Let $\{\tilde{e}_1, \ldots, \tilde{e}_n\}$ be another basis of V. Then there is a unique, invertible linear map $L : V \to V$ such that $\tilde{e}_\mu = L \cdot e_\mu$ for all $\mu = 1, \ldots, n$. We can view L as an element of $V \otimes V^*$, using

$$L : V^* \times V \to \mathbb{R} : (\omega, v) \mapsto \omega(L(v)),$$

[5]The symbol $\times$ denotes the Cartesian product of sets, i.e. the set whose elements are of the form $(\omega_1, \ldots, \omega_k, v_1 \ldots, v_l)$.

which is bilinear. The components of L in the basis $e_\mu \otimes e^{*\nu}$ are

$$L^\mu{}_\nu = e^{*\mu}(\tilde{e}_\nu).$$

Note in particular that

$$\tilde{e}_\nu = L^\mu{}_\nu e_\mu, \qquad e^{*\mu} = L^\mu{}_\nu \tilde{e}^{*\nu},$$

where the second equality follows from the fact that both sides of the equation take the same values on the basis $\tilde{e}_\rho$.

For any element $v \in V$ we may now compare the components v^μ and $\tilde{v}^\mu$ in the two bases:

$$v^\mu e_\mu = v = \tilde{v}^\nu \tilde{e}_\nu = L^\mu{}_\nu \tilde{v}^\nu e_\mu \qquad \Leftrightarrow \qquad v^\mu = L^\mu{}_\nu \tilde{v}^\nu.$$

Similarly, for any $\omega \in V^*$ we have

$$\tilde{\omega}_\mu \tilde{e}^{*\mu} = \omega = \omega_\nu e^{*\nu} = \omega_\mu L^\mu{}_\nu \tilde{e}^{*\nu} \qquad \Leftrightarrow \qquad \tilde{\omega}_\nu = \omega_\mu L^\mu{}_\nu.$$

After interchanging the roles of e_μ and $\tilde{e}_\mu$ and using L^{-1} instead of L this leads to:

$$\omega_\nu = (L^{-1})^\mu{}_\nu \tilde{\omega}_\mu.$$

A similar relation can be obtained for arbitrary tensors:

$$T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} = L^{\mu_1}{}_{\rho_1} \cdots L^{\mu_k}{}_{\rho_k} (L^{-1})^{\sigma_1}{}_{\nu_1} \cdots (L^{-1})^{\sigma_l}{}_{\nu_l} \tilde{T}^{\rho_1 \cdots \rho_k}{}_{\sigma_1 \cdots \sigma_l}. \tag{13}$$

This is called the *tensor transformation law.*

Given two tensors T and S of the same type (k, l), the equality $T = S$ of multi-linear maps is equivalent to the equality of the components

$$T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} = S^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} \tag{14}$$

using a basis e_μ of V. This is because the tensors are uniquely determined by their values on the basis elements, and these values are exactly the components. Note that this equality holds in any basis e_μ. Sometimes, however, it is convenient to write equality (14) for the components of a tensor T in some given basis e_μ and a set of numbers $S^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}$ which may not be related to a tensor, but which emerge in some other way from a particular physical problem. Such an equality does depend on the choice of basis, because the numbers in S do not transform as a tensor. To distinguish true tensor equations from other equalities using indices we introduce an additional convention:

Convention 9.2 (Abstract Index Notation) *The components of a tensor in a certain basis will be denoted by Greek indices. When the choice of basis is arbitrary, we will use a symbol with latin indices, e.g.* $T^{a_1\cdots a_k}{}_{b_1\cdots b_l}$*, to denote the type of the tensor. (These are not numbers or components of the tensor in some basis.) In particular, an equality between tensors, which holds in any basis, will be written using latin indices: it is an equality between multi-linear maps, rather than between real numbers.*

For general developments it is often nicer to use abstract indices, but in concrete examples it is often easier to choose a particular set of coordinates, which is adapted to the symmetries of the problem. This is why we will use abstract indices for now, but we will mostly revert to particular coordinates when considering applications.

There are two widely used operations on tensors, which we will now describe. Firstly, given a tensor S of type (k, l) and a tensor T of type (k', l') we may define the *outer product* as the tensor of type $(k + k', l + l')$ defined by

$$(S \otimes T)^{a_1\cdots a_{k+k'}}{}_{b_1\cdots b_{l+l'}} := S^{a_1\cdots a_k}{}_{b_1\cdots b_l} T^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}}.$$

(Note that this definition is independent of the choice of basis.) Furthermore, given any tensor T of type (k, l) with $k \geq 1$ and $l \geq 1$ we may define the contraction CT of the ith upper index and the jth lower index as the tensor

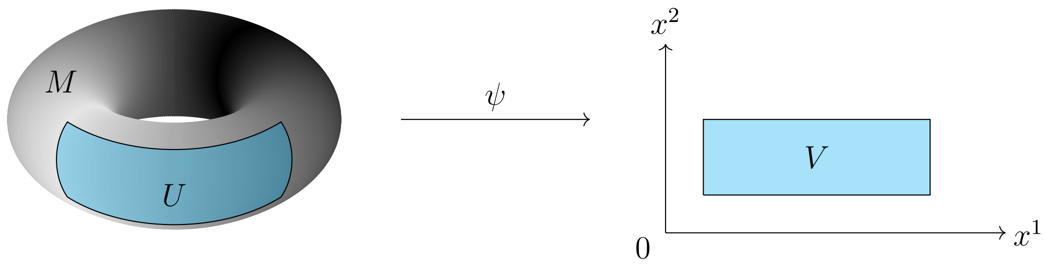
$$(CT)^{a_1\cdots\widehat{a_i}\cdots a_k}{}_{b_1\cdots\widehat{b_j}\cdots b_l} := T^{a_1\cdots c\cdots a_k}{}_{b_1\cdots c\cdots b_l},$$

where we recall that a sum is implied. As an example we consider a vector v^a and a dual vector ω_b, for which the outer product is $v^a\omega_b$ and the contraction of the outer product is $v^a\omega_a = \omega(v)$.

Exercise 9.2 *Verify that the equality* $v^a\omega_a = \omega(v)$ *holds in any basis* e_μ*.*

9.2 Manifolds

In order to describe the set M of all events with as few unphysical assumptions as possible, we introduce the mathematical concept of a manifold. This is ultimately based on ideas from cartography.



Definition 9.3 *A C^∞* manifold of dimension $n \in \mathbb{N}$ *is a non-empty set M together with a collection of bijective maps $\psi_i : U_i \to V_i$, where $U_i \subset M$, $V_i \subset \mathbb{R}^n$ is an open set and $i \in \mathcal{I}$ some index set, such that the collection of maps $\{\psi_i\}_{i\in\mathcal{I}}$ satisfies the following conditions:*

1. *the sets U_i cover M, $\bigcup_{i\in\mathcal{I}} U_i = M$, i.e. each $x \in M$ lies in some U_i,*
2. *if $U_i \cap U_j \neq \emptyset$ for some $i, j \in \mathcal{I}$, then $\psi_i(U_i \cap U_j)$ and $\psi_j(U_i \cap U_j)$ are open subsets of $\mathbb{R}^n$ and $\psi_j \circ \psi_i^{-1}$ is a C^∞ map with a C^∞ inverse between these sets.*

The maps $\psi_i : U_i \to V_i$ are called charts *(or* coordinate systems*) and the collection $\{\psi_i\}_{i\in\mathcal{I}}$ is called an* atlas *for M. The dimension n of M is written as* $\dim(M)$.

Let us consider some examples, to illustrate this notion:

Example 9.4
- *$M = \mathbb{R}^n$ is an n-dimensional manifold if we choose the atlas to consist of the single chart $\psi : M \to \mathbb{R}^n$, where ψ is the identity map.*
- *The unit circle $\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 |\ x^2 + y^2 = 1\}$ is a manifold of dimension 1. We may cover the set by four charts*

$$\begin{aligned}
\psi_{y+} : \{(x,y) \in \mathbb{S}^1 |\ y > 0\} \to (-1,1) \quad &: \quad (x,y) \mapsto x \\
\psi_{y-} : \{(x,y) \in \mathbb{S}^1 |\ y < 0\} \to (-1,1) \quad &: \quad (x,y) \mapsto x \\
\psi_{x+} : \{(x,y) \in \mathbb{S}^1 |\ x > 0\} \to (-1,1) \quad &: \quad (x,y) \mapsto y \\
\psi_{x-} : \{(x,y) \in \mathbb{S}^1 |\ x < 0\} \to (-1,1) \quad &: \quad (x,y) \mapsto y.
\end{aligned}$$

The compatibility of these charts is straightforward to check. E.g., for ψ_{x+} and ψ_{y+} the intersection of the domains of definition is $\{(x, y) \in$

$\mathbb{S}^1|\ x > 0, y > 0\}$ and we have $\psi_{x+} \circ \psi_{y+}^{-1}(x) = \sqrt{1-x^2}$, which is C^∞ and has the C^∞ inverse $\psi_{y+} \circ \psi_{x+}^{-1}(y) = \sqrt{1-y^2}$.

- *The n-dimensional unit sphere $\mathbb{S}^n = \{x \in \mathbb{R}^{n+1}|\ \sum_{k=1}^{n+1}(x^k)^2 = 1\}$ is a manifold of dimension n. It can be covered by charts in a similar way as $\mathbb{S}^1$, but now $2(n+1)$ charts are needed.*

Notice that a general manifold cannot be covered by a single chart.

Exercise 9.5 *Show that $\mathbb{S}^2$ is a manifold, give the coordinate charts in analogy to $\mathbb{S}^1$ and show that the changes of coordinate charts are smooth. (You may use symmetries to simplify the problem.)*

In addition to specific examples, there are general constructions which allow us to construct new manifolds from old ones. We mention the most important ones.

First, we call a set $O \subset M$ an *open subset* if and only if $\psi_i(O \cap U_i)$ is an open subset of $\mathbb{R}^n$ for all charts in the atlas. Any open subset $O \subset M$ is a manifold in its own right, where the charts are given by $\psi_i|_O : (O \cap U_i) \to \psi_i(O \cap U_i)$ with $i \in \mathcal{I}$.

Given two manifolds M and M', the product set $M \times M'$ is a manifold, where the charts are given by all maps $\psi_i \times \psi'_j : U_i \times U'_j \to V_i \times V_j$, where $\psi_i : U_i \to V_i$ is any chart on M and $\psi'_i : U'_i \to V'_i$ on M'. Note that $\dim(M \times M') = \dim(M) + \dim(M')$.

Example 9.6

The n-dimensional torus is $\mathbb{T}^n := \mathbb{S}^1 \times \ldots \times \mathbb{S}^1$ (n factors). Note that $\mathbb{T}^1$ is just the circle and $\mathbb{T}^2$ has the shape of a doughnut.

Exercise 9.7 *Sketch $\mathbb{T}^2$ as a subset of $\mathbb{R}^3$ and sketch a typical product chart obtained from the charts of $\mathbb{S}^1$ given in Example 9.4.*

A map $f : M \to M'$ between two manifolds is called k-times continuously differentiable, or C^k when $\psi'_i \circ f \circ \psi_j^{-1}$ is k-times continuously differentiable for all charts ψ_j in the atlas for M and all charts ψ'_i in the atlas for M'. As an example we note that, in any chart $\psi_i : U_i \to V_i$, the coordinates $x^k : \mathbb{R}^n \to \mathbb{R}$ can be used to define smooth maps $x^k \circ \psi_i$ on U_i. Other than such local coordinates, we will mostly be concerned with curves, $\xi : I \to M$, where $I \subset \mathbb{R}$ is an open interval (which is also a manifold of dimension 1).

All of our manifolds will have some additional properties, which we mention here without a very detailed discussion, because they will not be needed explicitly:

1. The atlas of a manifold M is called *maximal* when it has the following property: Let $\psi : U \to V$ be a bijective map between some $U \subset M$ and an open set $V \subset \mathbb{R}^n$ and suppose that for all $i \in \mathcal{I}$, ψ is compatible with ψ_i in the sense of the second condition in the definition of a manifold. Then ψ is already contained in the atlas $\{\psi_i\}_{i\in\mathcal{I}}$, i.e. $\psi = \psi_j$ for some $j \in \mathcal{I}$. Any atlas of a manifold M can always be extended in a unique way to a *maximal* one. We will always assume that our manifolds are equipped with a maximal atlas.

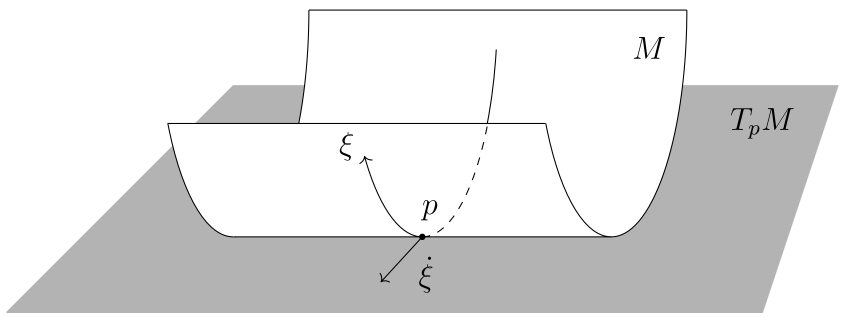
2. All our manifolds are path-connected. This means that for any two points $p_1, p_2 \in M$ there is a continuous curve $\xi : [0, 1] \to M$ such that $\xi(0) = p_1$ and $\xi(1) = p_2$.

3. All our manifolds are *Hausdorff* topological spaces. This means that for any two points $p_1, p_2 \in M$ there are open sets $U_1, U_2 \subset M$ with $p_i \in U_i$, $i = 1, 2$, but $U_1 \cap U_2 = \emptyset$. (This is an additional assumption on M, which we will always make.)

4. All our manifolds are *second countable*. This means that there is a countable collection $\{O_n\}_{n\in\mathbb{N}}$ of open sets $O_n \subset M$ such that every open set $O \subset M$ contains some O_n.

All of the examples of manifolds M we have seen so far can be embedded into $\mathbb{R}^m$ (for some $m \geq \dim(M)$). However, the importance of manifolds is that we can investigate them in a framework which is independent of this embedding. For example, $\mathbb{S}^1$ is defined as a subset of $\mathbb{R}^2$ and by taking products $\mathbb{T}^2$ can be viewed as a subset of $\mathbb{R}^4$. However, $\mathbb{T}^2$ can also be viewed as a subset of $\mathbb{R}^3$. Moreover, for the set M that models all events in the universe, it is not clear a priori if it can be embedded into any $\mathbb{R}^m$ at all! Also the shape of M is not a priori clear – why should we assume that it be covered by a single chart like Minkowski space? For these reasons, any information about a manifold should really be formulated in terms of the intrinsic structure of the manifold itself, independent of any embedding. We have already done this for the set M and its topological and differential structure (i.e. we know what C^∞ maps on M are). We now turn to the notion of tangent vectors.

A C^1 curve $\xi : I \to \mathbb{R}^n$ with $0 \in I$ has a tangent vector at $p = \xi(0)$, which is given by $\dot{\xi}^a(0)$ (where the derivative is taken component-wise in some basis). The tangent vectors at $p \in \mathbb{R}^n$ form a vector space of dimension n, just by adding their components.



However, to formulate the idea of tangent vectors on a manifold it is helpful to take a different perspective. Let x^μ be Cartesian coordinates on $\mathbb{R}^n$ (with a corresponding basis e_μ). Any tangent vector v at p, with components v^μ, defines a directional derivative operator

$$v^\mu \partial_{x^\mu} : C^\infty(\mathbb{R}^n, \mathbb{R}) \to \mathbb{R} : f \mapsto v^\mu \partial_{x^\mu} f(p).$$

This operator is linear and satisfies the Leibniz rule. Conversely, one may show that any such directional derivative operator corresponds to a unique tangent vector.

Exercise 9.8 *Show that the formula $v^\mu \partial_{x^\mu}$ is independent of the choice of basis e_μ.*

For a manifold M we now make the following

Definition 9.9 *A* tangent vector *v at $p \in M$ is an operator $v : C^\infty(M, \mathbb{R}) \to \mathbb{R}$ such that, for all $f_1, f_2 \in C^\infty(M, \mathbb{R})$ and $c_1, c_2 \in \mathbb{R}$,*

1. *$v(c_1 f_1 + c_2 f_2) = c_1 v(f_1) + c_2 v(f_2)$, and*
2. *$v(f_1 f_2) = f_1(p) v(f_2) + v(f_1) f_2(p)$ (Leibniz rule).*

The set of all tangent vectors at $p \in M$ is denoted by T_pM.

We note that T_pM forms a vector space, where

$$(c_1 v_1 + c_2 v_2)(f) := c_1 v_1(f) + c_2 v_2(f).$$

One way to obtain some examples $X_\mu \in T_pM$ is by fixing a chart $\psi : U \to V$ with $p \in U$ and setting

$$X_\mu(f) := \partial_{x^\mu}(f \circ \psi^{-1})(\psi(p)),$$

where the x^μ are Cartesian coordinates on $\mathbb{R}^n$. It is not hard to verify that the X_μ are indeed in T_pM. Moreover, one may show that they form a basis of T_pM, so that T_pM is a vector space of dimension $n = \dim(M)$. (For this one first shows that v only depends on the values of f in any (small) neighbourhood of p. One may then use a chart near p to turn the question into a problem in ordinary calculus.)

In fact, we may use the chart $\psi : U \to V$ to define tangent vectors like X_μ at any point $p \in U$. In this way we obtain a basis for T_pM for any $p \in U$, which is called a *coordinate basis*. If we use a different chart $\psi' : U' \to V'$ near p, then the vectors X'_μ are related to the X_μ by the chain rule:

$$X'_\mu = \frac{\partial x^\nu(x')}{\partial x'^\mu} X_\nu,$$

where $x^\nu(x')$ is short-hand for $x^\nu(\psi \circ \psi'^{-1}(x'))$, which describes how the map $\psi \circ \psi'$ changes the coordinates x'^μ into coordinates x^ν. The quotient can also be written in the matrix notation $D^\nu{}_\mu(\psi \circ \psi'^{-1})$ (with bases corresponding to the Cartesian coordinates x^μ and x'^ν).

Similarly, any vector $v \in T_pM$ can be written as $v = v^\mu X_\mu = v'^\mu X'_\mu$ with

$$v'^\mu = \frac{\partial x'^\mu(x)}{\partial x^\nu} v^\nu. \tag{15}$$

This is just the tensor transformation rule applied to a vector, except that the matrix involved may now depend on the point $p \in M$.

Another way to obtain tangent vectors in T_pM is to consider a C^1 curve $\xi : I \to M$ with $0 \in I$ and such that $\xi(0) = p$ and setting

$$D_0\xi(f) := \dot{\xi}_0(f) := \partial_s(f \circ \xi)(0).$$

In a chart $\psi : U \to V$ with $p \in U$, we can express ξ in terms of its components $\xi^\mu := x^\mu \circ \psi \circ \xi$ in Cartesian coordinates x^μ:

$$\begin{aligned} D_0\xi(f) &= \partial_s f \circ \psi^{-1} \circ \psi \circ \xi(0) \\ &= \partial_\mu (f \circ \psi^{-1})(\psi(p)) \cdot \partial_s \xi^\mu(0) \\ &= \dot{\xi}^\mu(0) X_\mu(f), \end{aligned}$$

i.e.

$$D_0\xi = \dot{\xi}^\mu(0) X_\mu.$$

Many curves can define the same tangent vector in T_pM, but every tangent vector in T_pM is of this form for some curve ξ.

More generally, let $\chi : M \to M'$ be a smooth map such that $\chi(p) = p'$. Any tangent vector $v \in T_pM$ gives rise to a tangent vector $D_p\chi(v) \in T_{p'}M'$ defined by

$$(D_p\chi(v))(f) := v(f \circ \chi). \tag{16}$$

Note that the map $D_p\chi : T_pM \to T_{p'}M'$ is linear in v. We recover $D_0\xi$ as a special case, where our notation suppresses the vector $v = e_1 \in T_0\mathbb{R}$, which is the unit vector which points in the positive direction.

As a general warning we emphasize that for a general manifold there is no natural way to identify tangent vectors at some point $p \in M$ with tangent vectors at some other point $q \in M$. We know that this is possible in $\mathbb{R}^n$, simply by applying a translation. (More precisely, there is a unique translation $\tau : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto x - (p - q)$ which maps p to q and $D_p\tau$ can be used to identify $T_p\mathbb{R}^n$ with $T_q\mathbb{R}^n$.) However, such translations are not defined for general manifolds and there is no natural analog.

9.3 Tangent, Cotangent and Tensor Bundles

In order to systematically keep track of all tangent vectors at all points we introduce the tangent bundle:

Definition 9.10 *The* tangent bundle TM *of a manifold* M *of dimension* n *is the set*

$$TM := \bigcup_{p \in M} T_pM = \{(p, v) |\ p \in M,\ v \in T_pM\},$$

where the union is disjoint (so $(p, v) = (p', v')$ *if and only if* $p = p'$ *and* $v = v'$*). We view* TM *as a manifold of dimension* $2n$ *with a maximal atlas* *containing all charts of the form*

$$D\psi : TU \to V \times \mathbb{R}^n : (p, v) := (\psi(p), D_p\psi(v))$$

such that $v = (D_p\psi(v))^\mu X_\mu$ *in the coordinate basis of* T_pM *determined by* ψ.

To see that the manifold structure of TM is well defined we notice that for any charts $\psi : U \to V$ and $\psi' : U' \to V'$ with $U \cap U' \neq \emptyset$, the change of charts from $D\psi$ to $D\psi'$ is

$$(D\psi \circ (D\psi')^{-1})(\psi'(p), D_p\psi'(v)) = (\psi(p), D_p\psi(v)),$$

which is a diffeomorphism. To see this one may write it in component form and use Equation (15) for the vector components.

If $\chi : M \to M'$ is a C^∞ map, we may define the *tangent map*

$$D\chi : TM \to TM' : D\chi(p, v) := (\chi(p), D_p\chi(v)),$$

where $D_p\chi$ was defined in Equation (16).

Definition 9.11 *A* C^∞ vector field *is a* C^∞ *map* $v : M \to TM$ *such that* $v(p) \in T_p(M)$. *The space of all* C^∞ *vector fields is denoted by* $\Gamma^\infty(M, TM)$.

Because the atlas of TM is closely related to that of M, the smoothness condition can be formulated as follows: for any chart $\psi : U \to V$ in the atlas of M and the corresponding coordinate bases X_μ of T_pM, $p \in U$, the coefficient functions v^μ appearing in

$$v(p) = v^\mu(p) X_\mu,$$

must be C^∞. This local condition is independent of the choice of chart, so it gives rise to a global condition on v. The simplest example of vector fields, at least on the domain U of a chart ψ, are the coordinate basis vector fields X_μ.

For any $p \in M$, T_pM is a vector space, so we may apply the constructions of Section 9.1. We denote the dual space by T_p^*M and for the space of tensors of type (k, l) we introduce the notation

$$T_p^{(k,l)}M := \underbrace{T_pM \otimes \ldots \otimes T_pM}_{k \text{ times}} \otimes \underbrace{T_p^*M \otimes \ldots \otimes T_p^*M}_{l \text{ times}}.$$

We will now show that these linear spaces may be pasted together to form new manifolds, just like the tangent bundle.

Let us first consider the space T_p^*M, which is called the *cotangent space* of M at $p \in M$. The dimension of T_p^*M equals that of T_pM, which is $n = \dim(M)$. We may obtain interesting examples of elements in T_p^*M by choosing a function $f \in C^\infty(M, \mathbb{R})$ and defining $\mathrm{d}_p f \in T_p^*M$ by a funny reversal of perspective in the definition of tangent vectors:

$$\mathrm{d}_p f : T_pM \to \mathbb{R} : v \mapsto v(f). \tag{17}$$

$\mathrm{d}_p f$ is called the *differential* of f at $p \in M$.

Using a chart $\psi : U \to V$ with $p \in U$ we may apply this procedure to the coordinate functions $x^\mu \circ \psi$, which leads to a basis

$$X^{*\mu} := \mathrm{d}\,(x^\mu \circ \psi).$$

At each $p \in U$, this basis is dual to X_μ:

$$X^{*\mu}(X_\nu) = X_\nu(x^\mu \circ \psi) = \partial_{x^\nu}(x^\mu) = \delta^\mu{}_\nu.$$

We therefore call it the *dual coordinate basis* determined by ψ.

Recall that a change of chart at p induces a change of the basis X_μ, which can be written as a matrix multiplication. The change in the dual basis is then given by a multiplication with the inverse matrix, as in the tensor transformation law (13).

Definition 9.12 *The* cotangent bundle T^*M *of a manifold* M *of dimension* n *is the set*

$$T^*M := \bigcup_{p \in M} T_p^*M = \{(p, \omega) |\; p \in M,\; \omega \in T_P^*M\},$$

where the union is disjoint (so $(p, \omega) = (p', \omega')$ *if and only if* $p = p'$ *and* $\omega = \omega'$*). We view* T^*M *as a manifold of dimension* $2n$ *with a maximal atlas containing all charts of the form*

$$D^*\psi : T^*U \to V \times \mathbb{R}^n : (p, \omega) := (\psi(p), D_p^*\psi(\omega))$$

such that $(D_p^*\psi(\omega))_\mu = \omega(X_\mu)$*.*

A C^∞ cotangent vector field *(or* dual vector field *or* 1-form field*) is a* C^∞ *map* $\omega : M \to T^*M$ *such that* $v(p) \in T_p^*(M)$*. The space of all* C^∞ *dual vector fields is denoted by* $\Gamma^\infty(M, T^*M)$*.*

Note that $\omega(X_\mu)$ are the components of ω in the dual coordinate basis, $\omega = \omega(X_\mu)X^{*\mu}$, as may be checked by letting both sides act on the coordinate basis X_ν. The dual coordinate basis $X^{*\mu}$ defines C^∞ cotangent vector fields on the domain U of the given chart and a general C^∞ cotangent vector field can be expressed as

$$\omega(p) = \omega_\mu(p)X^{*\mu}$$

with C^∞ coefficients ω_μ.

General tensors may be treated in an analogous way:

Definition 9.13 *The* tensor bundle $T^{(k,l)}M$ of type (k,l) *on a manifold M of dimension n is the set*

$$T^{(k,l)}M := \bigcup_{p\in M} T_p^{(k,l)}M,$$

where the union is disjoint. We view T^*M as a manifold of dimension n^{k+l+1} with a maximal atlas containing all charts of the form

$$D^{(k,l)}\psi : T^{(k,l)}U \to V \times \mathbb{R}^{n^{k+l}} : (p,S) := (\psi(p), (D_p^{(k,l)}\psi)S)$$

such that the components of $((D_p^{(k,l)}\psi)S)^{\mu_1\cdots\mu_k}{}_{\nu_1\cdots\nu_l} = S(X^{*\mu_1},\ldots,X^{*\mu_k},X_{\nu_1},\ldots,X_{\nu_l})$.

A C^∞ tensor field of type (k,l) *is a C^∞ map $S : M \to T^{(k,l)}M$ such that $S(p) \in T_p^{(k,l)}(M)$. The space of all C^∞ tensor fields of type (k,l) is denoted by $\Gamma^\infty(M, T^{(k,l)}M)$.*

Again the smoothness of a tensor field S means that the components of S in the basis obtained from X_μ and $X^{*,\nu}$ are smooth functions on M.

A chart $\psi : U \to V$ determines a coordinate basis at each $p \in U$ for the tensor bundle. This basis consists of tensor products of X_μ and their duals $X^{*,\mu}$. We may express the components of a tensor field T in terms of this basis, but we may also use the abstract index notation. Recall that a formula in the abstract index notation is valid when the abstract symbols are replaced by the components of the tensor in any coordinate basis (cf. Convention 9.2).

The outer product and contraction of tensors can also be defined for tensor fields in a point-wise fashion. Given a tensor field S of type (k,l) and a tensor field T of type (k',l'), the outer product is a tensor field of type $(k+k', l+l')$, given by

$$(S\otimes T)^{a_1\cdots a_{k+k'}}{}_{b_1\cdots b_{l+l'}} := S^{a_1\cdots a_k}{}_{b_1\cdots b_l} T^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}},$$

whereas the contraction over the ith upper index and the jth lower index is given by

$$(CT)^{a_1\cdots\widehat{a_i}\cdots a_k}{}_{b_1\cdots\widehat{b_j}\cdots b_l} := T^{a_1\cdots c\cdots a_k}{}_{b_1\cdots c\cdots b_l}.$$

These equations hold point-wise as identities between linear maps. In addition, given any $p \in M$ we may choose a chart $\psi : U \to V$ with $p \in U$ and the equations above (in abstract index notation) imply corresponding equalities for the components in the coordinate basis of ψ. These equations are independent of the choice of the chart ψ, but they only make sense for $p \in U$ and not for the entire manifold M. This is an important reason to use abstract indices: because the equations are independent of the choice of chart and basis X_μ, they make sense on the entire manifold.

Example 9.14 *For a vector field v^a and a dual vector field ω_b we can construct the $(1,1)$-tensor field $(v \otimes \omega)^a{}_b = v^a\omega_b$ and its contraction $(Cv \otimes \omega) = v^a\omega_a$. (In this case there is only one choice of indices that we can contract.) The result is a C^∞ function on M, which at each point $p \in M$ equals the value of $\omega(p)$ when acting on $v(p)$.*

Example 9.15 *For any tensor T of type $(0,l)$ we can define a fully anti-symmetric tensor aT by*

$$(aT)_{b_1\cdots b_l} := \frac{1}{l!}\sum_{\pi\in S_l} \varepsilon(\pi) T_{b_{\pi(1)}\cdots b_{\pi(l)}},$$

where S_l is the group of all permutations π of the numbers $(1,\ldots,l)$ and $\varepsilon(\pi) = \pm 1$ according to whether the permutation is odd or even. Such fully anti-symmetric tensors are called differential forms *in the mathematical literature. Note that aT is indeed anti-symmetric in all its indices and that $a(aT) = aT$. As a matter of notation we will write*

$$T_{[b_1\cdots b_l]} := (aT)_{b_1\cdots b_l},$$

where the square brackets [ ] indicate that the indices should be anti-symmetrized over.

In a similar way we can also define a fully symmetric tensor sT by

$$(sT)_{b_1\cdots b_l} := \frac{1}{l!}\sum_{\pi\in S_l} T_{b_{\pi(1)}\cdots b_{\pi(l)}},$$

which we will write as

$$T_{(b_1\cdots b_l)} := (sT)_{b_1\cdots b_l}.$$

where the round brackets ( ) *indicate that the indices should be symmetrized over.*

With a slight abuse of notation the basis vectors X_μ and their duals $X^{*\mu}$ are often written as

$$\frac{\partial}{\partial x^\mu} := X_\mu, \qquad \mathrm{d}\, x^\mu := X^{*\mu}.$$

We have avoided this notation so far, because it suppresses the choice of the ψ and treats the tangent vectors $\frac{\partial}{\partial x^\mu}$ on V as if they were tangent vectors on U. (The purpose of the charts ψ is of course to translate the usual differential calculus on V into a differential calculus on U, thereby justifying that the abusive notation is indeed valid.)

9.4 Covariant Derivatives

We now turn our attention to derivatives of tensor fields on a manifold M. We have already seen that any C^∞ function $f : M \to \mathbb{R}$ gives rise to a one-form $\mathrm{d}\, f$ defined by $\mathrm{d}\, f(v) := v(f)$. In a chart ψ the one-form locally has the expression $(\mathrm{d}\, f)_\mu = \partial_{x^\mu} f$. Under a change of chart, this expression is multiplied by a Jacobi matrix, as usual for a dual vector. For higher order derivatives, however, an expression like $\partial_{x^\mu}\partial_{x^\nu} f$ can be used to define a tensor, but the resulting tensor depends on the choice of coordinates used. In other words, the formula $\partial_{x^\mu}\partial_{x^\nu} f$ is not preserved under changes of coordinates, because we also obtain terms involving derivatives of the Jacobi matrix. We now discuss how to eliminate, or at least keep track of, this dependence on the chart, using the notion of (covariant) derivative operators.

Definition 9.16 *A* (covariant) derivative operator ∇ *on a manifold M is a map (or rather a set of maps, for each (k, l))*

$$\nabla : \Gamma^\infty(M, T^{(k,l)}M) \to \Gamma^\infty(M, T^{(k,l+1)}M) : T^{a_1\cdots a_k}{}_{b_1\cdots b_l} \to \nabla_{b_0} T^{a_1\cdots a_k}{}_{b_1\cdots b_l}$$

with the following properties:

1. linearity*: for all $c_i \in \mathbb{R}$ and tensor fields T_i of type (k, l), $\nabla(c_1T_1 + c_2T_2) = c_1\nabla T_1 + c_2\nabla T_2$, i.e.*

   $$\nabla_{b_0}(c_1T_1 + c_2T_2)^{a_1\cdots a_k}{}_{b_1\cdots b_l} = c_1\nabla_{b_0}(T_1)^{a_1\cdots a_k}{}_{b_1\cdots b_l} + c_2\nabla_{b_0}(T_2)^{a_1\cdots a_k}{}_{b_1\cdots b_l},$$

2. Leibniz rule*: for all tensor fields T and S, $\nabla(T \otimes S) = (\nabla T) \otimes S + T \otimes (\nabla S)$, i.e.*

$$\begin{aligned}\nabla_{b_0}(T^{a_1\cdots a_k}{}_{b_1\cdots b_l} S^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}}) &= \nabla_{b_0} T^{a_1\cdots a_k}{}_{b_1\cdots b_l} S^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}} + \\ &\quad T^{a_1\cdots a_k}{}_{b_1\cdots b_l} \nabla_{b_0} S^{a_{k+1}\cdots a_{k+k'}}{}_{b_{l+1}\cdots b_{l+l'}}\end{aligned}$$

if T is of type (k,l) and T' of type (k',l').

3. commutativity with contractions*: for any tensor field T of type (k,l) with $k,l \geq 1$, $(\nabla CT) = C'(\nabla T)$, where C contracts the ith upper and jth lower index and C' the ith upper and $j+1$st lower index:*

$$\nabla_{b_0}(T^{a_1\cdots c\cdots a_k}{}_{b_1\cdots c\cdots b_l} = \nabla_{b_0} T^{a_1\cdots c\cdots a_k}{}_{b_1\cdots c\cdots b_l},$$

4. consistency with differentials*: for any C^∞ function f, $\nabla f = d f$, i.e.*

$$\nabla_{b_0} f = (d f)_{b_0}.$$

5. torsion free*: for any C^∞ function f, $\nabla\nabla f$ is a symmetric tensor, i.e.*

$$\nabla_{b_0}\nabla_{b_1} f = \nabla_{b_1}\nabla_{b_0} f.$$

All these properties are familiar from calculus in $\mathbb{R}^n$ (except perhaps the commutativity with contractions). In fact, using a chart $\psi : U \to V$ and expressing tensors in their coordinate basis, one may choose $\nabla_\mu := \partial_{x^\mu}$ to define a derivative operator on U. This derivative, however, depends on the choice of coordinates. That covariant derivatives exist globally will be seen later, when we prove the existence of a preferred derivative operator.

For a general derivative operator, the action at $p \in M$ on a general tensor field T depends only on the values of T in an infinitesimal neighborhood of p. This follows form the Leibniz rule and $(\nabla T)(p) = (\nabla(fT))(p)$ when f is a C^∞ function with $f(p) = 1$ and $\mathrm{d} f(p) = 0$. Expanding T in a coordinate basis, we may use linearity and the Leibniz rule to express ∇T in terms of derivatives of the basis vector and dual vector fields. Because the derivative operator commutes with contractions and its value on functions is known, we also have

$$\omega_b \nabla_a v^b = \nabla_a(v^b \omega_b) - v^b \nabla_a \omega_b.$$

When ω_b ranges over a basis, this allows us to express the action on any vector field in terms of the action on dual vector fields and functions. Thus, the only freedom we have, is to modify the action on dual vector fields.

Theorem 9.17 *Given two derivative operators on M, ∇ and ∇', there is a unique tensor field C of type $(1,2)$ such that $C^c{}_{ab} = C^c{}_{ba}$ and*

$$\nabla'_a \omega_b - \nabla_a \omega_b = C^c{}_{ab} \omega_c \tag{18}$$

for all dual vector fields ω_a. Conversely, given any derivative operator ∇' and a tensor field $C^c{}_{ab}$ which is symmetric in its lower indices, Equation (18) defines a derivative operator ∇.

Proof: At any point $p \in M$, $\nabla_a \omega_b - \nabla'_a \omega_b$ is a tensor of type $(0,2)$ which depends linearly on ω_a. We will show that it actually only depends on the values $\omega_a(p)$ and not on any derivatives. For this purpose we fix any chart $\psi : U \to V$ with $p \in U$ and we define the coordinate functions $x^\mu \circ \psi$ on U, $\mu = 1, \ldots, n$, where the x^μ are Cartesian coordinates on $\mathbb{R}^n$. The local basis $X^{*\mu}$ can be expressed in an abstract index notation as

$$(X^{*\mu})_a = (\mathrm{d}\,(x^\mu \circ \psi))_a = \nabla_a (x^\mu \circ \psi) = \nabla'_a (x^\mu \circ \psi)$$

and we may write $(X_\mu)^c$ in a similar way. We then view the components ω_μ as functions on U and write $\omega_a = \omega_\mu (X^{*\mu})_a = \omega_\mu \nabla_a (x^\mu \circ \psi)$ and $\omega_\mu = \omega_c (X_\mu)^c$. Then we compute with Leibniz rule and linearity:

$$\begin{aligned}
\nabla'_a \omega_b - \nabla_a \omega_b &= (\nabla'_a \omega_\mu - \nabla_a \omega_\mu)(X^{*\mu})_b + \omega_\mu (\nabla'_a \nabla'_b (x^\mu \circ \psi) - \nabla_a \nabla_b (x^\mu \circ \psi)) \\
&= \omega_c (X_\mu)^c (\nabla'_a \nabla'_b (x^\mu \circ \psi) - \nabla_a \nabla_b (x^\mu \circ \psi)).
\end{aligned}$$

Note that the first term on the right-hand side of the first line vanishes, because all derivative operators have the same action on the functions ω_μ. We have now shown that $(\nabla'_a \omega_b - \nabla_a \omega_b)(p)$ is a linear map of $\omega_a(p)$ without any derivatives. This entails the existence of the tensor field $C^c{}_{ab}$. We even obtain a formula for this tensor field,

$$C^c{}_{ab} := (X_\mu)^c (\nabla'_a \nabla'_b (x^\mu \circ \psi) - \nabla_a \nabla_b (x^\mu \circ \psi)),$$

(which is actually independent of the choice of chart).

Conversely, given any derivative operator ∇' and a tensor field $C^c{}_{ab}$ which is symmetric in its lower indices, one may define

$$\begin{aligned}
\nabla_{b_0} T^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} &:= \nabla'_{b_0} T^{a_1 \cdots a_k}{}_{b_1 \cdots b_l} + \sum_{i=1}^{k} C^{a_i}{}_{b_0 c} T^{a_1 \cdots c \cdots a_k}{}_{b_1 \cdots b_l} \\
&\quad - \sum_{j=1}^{l} C^c{}_{b_0 b_j} T^{a_1 \cdots a_k}{}_{b_1 \cdots c \cdots b_l}.
\end{aligned}$$

We will omit the straightforward verification that ∇ is indeed a derivative operator. $\square$

In a local chart we may choose ∇' to be a coordinate derivative. The tensor field ${C^c}_{b_0 b_j}$ is then typically written as ${\Gamma^c}_{b_0 b_j}$ and it is called the *Christoffel symbol*.

Given a covariant derivative operator, it is natural to consider tensor fields whose covariant derivative vanishes. A very useful way of doing this is as follows. Let $\xi : I \to M$ be C^∞ curve with $0 \in I$ and $p := \xi(0)$ and let $v^a(p) \in T_p M$. Given a covariant derivative operator, we may define unique vectors $v^a(q)$ at all points q on ξ such that

$$\dot{\xi}^a \nabla_a v^b = 0$$

on the curve. To see why this is true, we choose a chart and write the equation as

$$0 = \dot{\xi}^\mu (\partial_\mu v^\nu + {\Gamma^\nu}_{\mu\rho} v^\rho) = \partial_s (v^\nu \circ \xi) + \dot{\xi}^\mu {\Gamma^\nu}_{\mu\rho} v^\rho.$$

This is a first order differential equation for the components $v^\nu \circ \xi$ as a function of the parameter s, so given the initial values, it admits a unique solution. The vectors $v^a(q)$ are called the *parallel transports* of $v^a(p)$. Of course a similar result holds for any tensor. (The general result may be obtained by linearity and taking tensor products and duals.)

Given a curve $\xi : I \to M$ that goes through $p, p' \in M$, we can use parallel transports to define a map $C_{\xi;p,p'} : T_p M \to T_{p'} M$. Notice that this map is linear, because the parallel transport equation is linear. Moreover, it is invertible, because we can traverse the curve ξ in the opposite direction to find that $C_{\xi;p',p}$ is the inverse. This way of identifying tangent spaces at different points is called a *connection* in the mathematical literature.

As a warning we emphasise that the linear isomorphism $C_{\xi;p,p'} : T_p M \to T_{p'} M$ depends on the choice of the curve ξ.

The fact that connections depend on the choice of a curve ξ is unfamiliar from calculus in $\mathbb{R}^n$. Another marked difference between covariant derivatives and ordinary derivatives is the fact that mixed derivatives do not always commute. We do have $\nabla_a \nabla_b f = \nabla_b \nabla_a f$ for functions f, but for tensors this may no longer be true. This leads to the concept of curvature.

Theorem 9.18 *Given a derivative operator ∇ on a manifold M, there is a tensor field ${R_{abc}}^d$ of type $(1,3)$ such that*

$$\nabla_a \nabla_b \omega_c - \nabla_b \nabla_a \omega_c = {R_{abc}}^d \omega_d$$

for all one-form fields ω_c.

Proof: The left-hand side is a tensor which depends linearly on ω_c. The main point is to show that the value at any $p \in M$ depends only on the value $\omega_c(p)$ and not on any derivatives. For this purpose one may simply work in local coordinates and express ∇ in terms of the coordinate derivatives and a Christoffel symbol. A straightforward computation, which we omit, shows that no derivatives of ω at p appear and

$$R_{\mu\nu\rho}{}^{\sigma} = -\partial_\mu \Gamma^\sigma{}_{\nu\rho} + \Gamma^\tau{}_{\mu\rho}\Gamma^\sigma{}_{\nu\tau} - (\mu \leftrightarrow \nu).$$

$\square$

Using linearity and the Leibniz rule one may show that for general tensors,

$$\begin{aligned}(\nabla_a\nabla_b - \nabla_b\nabla_a)T^{c_1\cdots c_k}{}_{d_1\cdots d_l} &= -\sum_{i=1}^{k} R_{abe}{}^{c_i} T^{c_1\cdots e\cdots c_k}{}_{d_1\cdots d_l} \\ &\quad + \sum_{j=1}^{l} R_{abd_j}{}^{e} T^{c_1\cdots c_k}{}_{d_1\cdots e\cdots d_l}.\end{aligned}$$

Definition 9.19 *The tensor $R_{abc}{}^{d}$ is called the* Riemann curvature tensor *of ∇. The tensor $R_{adc}{}^{d}$ of type $(0,2)$ is called the* Ricci curvature tensor*.*

Theorem 9.20 *The Riemann curvature tensor has the following properties:*

1. $R_{abc}{}^{d} = -R_{bac}{}^{d}$,
2. $R_{[abc]}{}^{d} = 0$,
3. $\nabla_{[a}R_{bc]d}{}^{e} = 0$ *(Bianchi identity).*

Proof: The first property of $R_{abc}{}^{d}$ is obvious. For the second we note that $\nabla_{[\mu}\omega_{\nu]} = \partial_{[\mu}\omega_{\nu]}$, because the Christoffel symbol is symmetric in its lower indices. Similarly,

$$\nabla_{[\mu}\nabla_\nu\omega_{\rho]} = \partial_{[\mu}\partial_\nu\omega_{\rho]} = 0$$

by Equation (19) and the fact that ordinary derivatives commute. Thus, $R_{[\mu\nu\rho]}{}^{\sigma}\omega_\sigma = 2\nabla_{[\mu}\nabla_\nu\omega_{\rho]} = 0$. For the third property we consider

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\nabla_c\omega_d = R_{abc}{}^{e}\nabla_e\omega_d + R_{abd}{}^{e}\nabla_c\omega_e$$

and

$$\nabla_a(\nabla_b\nabla_c - \nabla_c\nabla_b)\omega_d = (\nabla_a R_{bcd}{}^e)\omega_e + R_{bcd}{}^e\nabla_a\omega_e.$$

Antisymmetrizing over the indices a, b, c, the left-hand sides become equal. Using the properties of the Riemann curvature that we have already shown on the right-hand side leads to

$$(\nabla_{[a}R_{bc]d}{}^e)\omega_e = R_{[ab|d|}{}^e\nabla_{c]}\omega_e - R_{[bc|d|}{}^e\nabla_{a]}\omega_e = 0.$$

As ω_e is arbitrary, the Bianchi identity follows. □

9.5 Metrics and Geodesics

So far we do not have sufficient structure on our manifolds to formulate physically relevant questions: What is the distance between two points? What is the shortest path between two points? And is there a preferred choice of covariant derivative?

Let us start with the definition of a (pseudo-) metric:

Definition 9.21 *A (pseudo-)*metric *on a manifold M is a smooth tensor field g_{ab} of type $(0,2)$ such that*

1. *$g_{ab} = g_{ba}$ (symmetry),*
2. *if $v^a \in T_pM$ has $g_{ab}(p)v^a w^b = 0$ for all $w^b \in T_pM$, then $v^a = 0$ (non-degeneracy).*

A pair (M, g_{ab}) consisting of a manifold with a (pseudo-)metric is called a pseudo-Riemannian manifold.

For any $v^a \in TM$, $g_{ab}v^a v^b$ can be viewed as the squared norm of the vector v^a. However, g_{ab} need not be positive definite, i.e. it need not be true that $g_{ab}v^a v^b \geq 0$ for all $v^a \in TM$.

At any point $p \in M$ we can consider the matrix $g_{\mu\nu}(p)$ in a coordinate basis of some chart. This matrix depends on the choice of coordinates, but it has a number of properties which are independent of this choice. It is always symmetric, so it can be diagonalized and the numbers n_+ of positive eigenvalues and $n_- = n - n_+$ of negative eigenvalues are also coordinate independent. In fact, for a given point $p \in M$ we can always choose coordinates such that $g_{\mu\nu}(p)$ is diagonal with eigenvalues $+1$ and -1 only. However, this may hold only hold at the point p! Finally we note that n_+ and n_- do not depend on the choice of p, because the components $g_{\mu\nu}(p)$ vary smoothly with p and $g_{\mu\nu}$ must remain non-degenerate.

Definition 9.22 *The* signature *of a metric is the pair (n_+, n_-), where n_+ is the number of positive eigenvalues of the matrix $g_{\mu\nu}(p)$ (in a coordinate basis at any point $p \in M$), and n_- is the number of negative eigenvalues.*

We call a metric Riemannian *when its signature is $(n, 0)$. In that case we call the pair (M, g_{ab}) a* Riemannian manifold*.*

We call a metric Lorentzian *when its signature is $(n-1, 1)$ and $n \geq 2$. In that case we call the pair (M, g_{ab}) a* Lorentzian manifold*, or a* spacetime*.*

For a Riemannian metric, $g_{\mu\nu}$ is positive definite. For a Lorentzian metric, $g_{\mu\nu}$ is similar to the matrix η of Equation (7).

For a Riemannian manifold (M, g_{ab}) and a curve $\xi : (a, b) \to M$ we can define the length of ξ by

$$\int_a^b \sqrt{g_{\mu\nu}(\xi(s))\dot{\xi}^\mu(s)\dot{\xi}^\nu(s)}\, \mathrm{d}\, s,$$

i.e. we integrate the length of the tangent vectors along the curve. Note that this expression is independent of the choice of parameter.

For Lorentzian manifolds (M, g_{ab}) we say that a curve $\gamma : (s_1, s_2) \to M$ is *spacelike* when $g_{ab}(\gamma(s))\dot{\gamma}^a(s)\dot{\gamma}^b(s) > 0$ for all s. The *length* of γ is then given by

$$l_\gamma := \int_{s_1}^{s_2} \sqrt{g_{ab}(\gamma(s))\dot{\gamma}^a(s)\dot{\gamma}^b(s)}\mathrm{d}\, s.$$

Similarly, γ is called *timelike* when $g_{ab}(\gamma(s))\dot{\gamma}^a(s)\dot{\gamma}^b(s) < 0$ for all s. The *proper time* of γ is then given by

$$\tau_\gamma := \int_{s_1}^{s_2} \sqrt{-g_{ab}(\gamma(s))\dot{\gamma}^a(s)\dot{\gamma}^b(s)}\mathrm{d}\, s.$$

For curves which are neither timelike nor spacelike we do not define a length or proper time.

Because g_{ab} is non-degenerate, we can use it to define a bijection

$$TM \to T^*M : (p, v^b) \mapsto (p, g_{ab}(p)v^b),$$

which is a linear isomorphism from T_pM to T_p^*M for each $p \in M$. The inverse of this map is a tensor of type $(2,0)$ and will be denoted by g^{ab}, with upper indices, so that

$$g^{ab}g_{bc} = \delta^a{}_c$$

is the identity map on TM. As a matter of notation we will define

$$v_a := g_{ab}(p)v^b, \qquad \omega^a := g^{ab}(p)\omega_b$$

for any $(p, v^b) \in TM$ and $(p, \omega_b) \in T^*M$. In this way we can use the metric to raise and lower indices, i.e. to identify vectors and dual vectors. The notation can be extended in an obvious way to all tensors. As a warning, however, we note that in General Relativity, the Lorentzian metric g_{ab} is a dynamical object, so the identification between vectors and dual vectors is not given a priori. Instead, we must solve Einstein's equations to find it.

Exercise 9.23 *Use the inverse metric g^{ab} to show that $g_{ab}g^{ac}g^{bd} = g^{cd}$, so the notation for raising and lowering indices can also be applied in a consistent way to the metric itself.*

Suppose that $\gamma : I \to M$ is a curve and v^a and w^b are vector fields along the curve which are parallel transports for some covariant derivative operator ∇. The inner product $g_{ab}v^aw^b$ along γ changes as follows:

$$\dot{\gamma}^a\nabla_a(g_{bc}v^bw^c) = \dot{\gamma}^a v^b w^c \nabla_a g_{bc}.$$

This means that the inner product remains constant if $\nabla_a g_{bc} = 0$. We now show that for a given (pseudo-)metric g_{ab} there exists a unique covariant derivative operator which satisfies this additional condition:

Theorem 9.24 *Let g_{ab} be any (pseudo-) metric on a manifold M. Then there exists a unique covariant derivative operator ∇ on M which is compatible with g_{ab} in the sense that $\nabla_a g_{bc} = 0$.*

Proof: In any chart $\psi : U \to V$ we can use the coordinate derivative ∂_μ on U. By Theorem 9.17 we can characterize ∇ on U by its Christoffel symbols $\Gamma^c{}_{ab} = C^c{}_{ab}$. We have

$$0 = \nabla_\mu g_{\nu\rho} = \partial_\mu g_{\nu\rho} - \Gamma^\sigma{}_{\mu\nu} g_{\sigma\rho} - \Gamma^\sigma{}_{\mu\rho} g_{\nu\sigma}$$

and therefore,

$$\begin{aligned}\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu} &= (\Gamma_{\rho\mu\nu} + \Gamma_{\nu\mu\rho}) + (\Gamma_{\rho\nu\mu} + \Gamma_{\mu\nu\rho}) - (\Gamma_{\nu\rho\mu} + \Gamma_{\mu\rho\nu}) \\ &= 2\Gamma_{\rho\mu\nu},\end{aligned}$$

where we used the symmetry of the Christoffel symbol. This shows that any derivative operator which is compatible with g_{ab} must have a Christoffel symbol which is given in any chart by

$$\Gamma^\sigma{}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho}(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}),$$

which proves uniqueness of ∇ on U. Conversely, choosing this Christoffel symbol to define ∇, we also see that ∇ exists on U. We can do this for any chart and on the intersection of any two chart domains, the resulting covariant derivatives must coincide, because both satisfy the compatibility condition. □

The unique covariant derivative operator which is compatible with g_{ab} is called the *Levi-Civita* covariant derivative (or *Levi-Civita connection*). The Levi-Civita derivative operator gives rise to a connection $C_{\xi;p,p'}$ which is not just a linear isomorphism $T_pM \to T_{p'}M$, but it also preserves inner products, i.e. it maps $g_{ab}(p)$ to $g_{ab}(p')$.

We now consider the curvature of the Levi-Civita derivative operator, starting with some additional notations:

Definition 9.25 *The function $R := g^{ac}R_{ac} = g^{ac}R_{abc}{}^b$ is called the (Ricci)* scalar curvature *of the Levi-Civita derivative. The tensor $G_{ab} := R_{ab} - \frac{1}{2}Rg_{ab}$ is called its* Einstein tensor*.*

Theorem 9.26 *The curvature of the Levi-Civita derivative operator satisfies*

$$\begin{aligned} R_{abcd} &= -R_{abdc}, \\ G_{ab} &= G_{ba}, \\ \nabla^a G_{ab} &= 0. \end{aligned}$$

Proof: The first property follows directly from $0 = \nabla_{[a}\nabla_{b]}g_{cd}$ and Equation (19). The second property follows from $R_{[abc]}{}^d = 0$, because

$$R_{ba} = R_{bca}{}^c = -R_{abc}{}^c - R_{cab}{}^c = -g^{cd}R_{abcd} + R_{acb}{}^c = R_{ab}.$$

(The first term vanishes by the first property and symmetry reasons.) For the third we consider the Bianchi identity $\nabla_{[a}R_{bc]d}{}^{e} = 0$ (cf. Theorem 9.20) and we use the symmetries of the Riemann curvature to compute:

$$\begin{aligned} 0 &= 3g^{bd}\nabla_{[a}R_{bc]d}{}^{c} \\ &= g^{bd}\nabla_a R_{bcd}{}^{c} + g^{bd}\nabla_b R_{cad}{}^{c} + g^{bd}\nabla_c R_{abd}{}^{c} \\ &= \nabla_a g^{bd}R_{bd} - \nabla^d R_{acd}{}^{c} - \nabla^c g^{bd}R_{abcd} \\ &= \nabla_a R - 2\nabla^d R_{ad} \\ &= -2\nabla^d(R_{ad} - \frac{1}{2}g_{ad}R) = -2\nabla^d G_{ad}. \end{aligned}$$

Because the Einstein tensor is symmetric, the result follows. □

Definition 9.27 *A* geodesic *in a (pseudo-)Riemannian manifold* (M, g_{ab}) *is a curve* $\xi : I \to M$ *whose tangent vectors* $\dot{\xi}^a$ *are parallelly transported along* ξ *with respect to the Levi-Civita covariant derivative operator.*

(Actually, the definition can be made for arbitrary covariant derivative operators on M, but the Levi-Civita derivative operator will be our main interest.)

We think of geodesics as curves which are as straight as possible. The condition on the tangent vectors can be written as the geodesic equation:

$$\dot{\xi}^a\nabla_a\dot{\xi}^b = 0, \tag{19}$$

or, in a chart and using the parameter s along ξ:

$$\ddot{\xi}^\mu + \Gamma^\mu{}_{\nu\rho}(\xi)\dot{\xi}^\nu\dot{\xi}^\rho = 0.$$

This is a second order (non-linear) ordinary differential equation for the components $\xi^\mu(s)$, so its solution is uniquely determined (for s in some maximal interval) by initial values for ξ^μ and $\dot{\xi}^\mu$. In other words, there exists a unique maximal geodesic through any point $(p, v^a) \in TM$.

Exercise 9.28 *Consider* $\mathbb{R}^n$ *as a Riemannian manifold with the constant metric field* $g_{\mu\nu} = \delta_{\mu\nu}$ *in a global Cartesian chart. Given any tangent vector,* $(p, v^\mu) \in T\mathbb{R}^n$*, show that the unique geodesic* ξ *with initial data* (p, v^μ) *is*

$$\xi^\mu(s) := p^\mu + sv^\mu.$$

Exercise 9.29 *Show that for any geodesic $\xi : I \to M$, $\dot{\xi}^a \dot{\xi}_a$ is constant along the curve ξ.*

For spacelike or timelike curves the geodesic equation (19) can be obtained from a variational principle. The length l_γ, resp. proper time τ_γ, of a spacelike, respectively timelike, curve γ between two fixed points attains a (local) extremum when γ is a geodesic. To verify this one computes the Euler-Lagrange equations of

$$\int \sqrt{|g_{ab}(\gamma(s))\dot{\gamma}^a(s)\dot{\gamma}^b(s)|}\mathrm{d}\, s$$

to be (in local coordinates (x^μ))

$$0 = \frac{-2}{\sqrt{|g_{\mu\nu}(\xi)\dot{\xi}^\mu \dot{\xi}^\nu|}} \left(-\frac{1}{2}(\partial_\rho g_{\mu\nu})(\xi)\dot{\xi}^\mu \dot{\xi}^\nu + \partial_s(g_{\mu\rho}(\xi)\dot{\xi}^\mu) \right)$$

and the term in brackets can be rewritten, using $\partial_s(g_{\mu\rho}(\xi) = (\partial_\nu g_{\mu\rho})\dot{\xi}^\nu$ and relabeling indices, as

$$g_{\mu\rho}\left(\ddot{\xi}^\mu + \Gamma^\mu{}_{\sigma\nu}\dot{\xi}^\sigma \dot{\xi}^\nu\right).$$

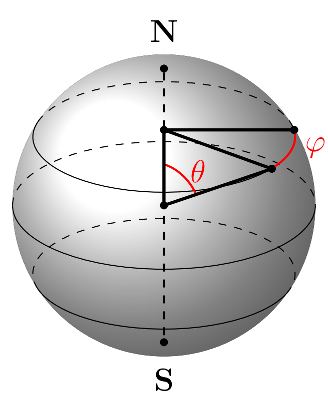
Example 9.30 (Riemann-tensor and geodesics on $\mathbb{S}^2$) *The metric of a 'round' 2-sphere $\mathbb{S}^2$ with radius r is by definition*

$$ds^2 = r^2 \left(d\theta^2 + \sin^2\theta \; d\varphi^2\right)$$

with $\theta = x_1$ and $\varphi = x_2$, or in matrix form

$$(g_{\mu\nu}) = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2\theta \end{pmatrix},$$

$$(g^{\mu\nu}) = \begin{pmatrix} \frac{1}{r^2} & 0 \\ 0 & \frac{1}{r^2\sin^2\theta} \end{pmatrix}.$$



One can check using the definition of $\Gamma^{\sigma}{}_{\mu\nu}$ *that*

$$\begin{aligned}\Gamma^{1}{}_{22} &= -\sin\theta\cos\theta \\ \Gamma^{2}{}_{12} = \Gamma^{2}{}_{21} &= \cot\theta\end{aligned}$$

and that all other components vanish.

$$\begin{aligned}R_{212}{}^{1} &= \partial_1\Gamma^{1}{}_{22} - \partial_2\Gamma^{1}{}_{12} + \Gamma^{1}{}_{1\alpha}\Gamma^{\alpha}{}_{22} - \Gamma^{1}{}_{2\alpha}\Gamma^{\alpha}{}_{12} \quad (\textit{sum over } \alpha = 1, 2!) \\ &= \left(\sin^2\theta - \cos^2\theta\right) - (0) + (0) - (-\sin\theta\cos\theta)(\cot\theta) \\ &= \sin^2\theta.\end{aligned}$$

To get $R_{\mu\nu\sigma\rho}$ *with lower indices, we use* $R_{\mu\nu\sigma\rho} = g_{\rho\alpha}R_{\mu\nu\sigma}{}^{\alpha}$. *Then we have*

$$\begin{aligned}\Rightarrow \quad R_{2121} &= g_{1\alpha}R_{212}{}^{\alpha} \\ &= g_{11}R_{212}{}^{1} + \cancel{g_{12}}^{=0}R_{212}{}^{2} \\ &= r^2\sin^2\theta.\end{aligned}$$

Using the symmetries

$$R_{11\alpha\beta} = R_{22\alpha\beta} = R_{\alpha\beta 11} = R_{\alpha\beta 22} = 0$$

we find $R_{2112} = -R_{2121}, \ldots$ *and so on for all 16 components of* $R_{\alpha\beta\gamma\delta}$.

$$\begin{aligned}R_{11} &= g^{\alpha\beta}R_{1\alpha 1\beta} \\ &= g^{11}\cancel{R_{1111}}^{=0} + g^{22}R_{1212} + 2g^{12}\cancel{R_{1121}}^{=0} \\ &= 1 \\ R_{22} &= g^{\alpha\beta}R_{2\alpha 2\beta} \\ &= g^{11}R_{2121} + g^{22}\cancel{R_{2222}}^{=0} + 2g^{12}\cancel{R_{2122}}^{=0} \\ &= \sin^2\theta\end{aligned}$$

$$\begin{aligned}\Rightarrow \quad R &= g^{11}R_{11} + \cancel{2g^{12}}^{=0}R_{12} + g^{22}R_{22} \\ &= \frac{2}{r^2}.\end{aligned}$$

It shows, that $\mathbb{S}^2$ *has a constant scalar curvature.*

To find the geodesics, let $(\gamma^{\mu}(t)) = (x_1(t), x_2(t)) = (\theta(t), \varphi(t))$. *Then we have the geodesic equation*

$$\frac{d^2}{dt^2}\gamma^{\mu}(t) + \Gamma^{\mu}{}_{\alpha\beta}(\gamma(t))\,\frac{d}{dt}\gamma^{\alpha}(t)\frac{d}{dt}\gamma^{\beta}(t) = 0$$

Written out using $\frac{d}{dt} = \dot{}$ and $\mu = 1$ or $\mu = 2$:

$$\ddot{\theta} - \sin\theta\cos\theta\,(\dot{\varphi})^2 = 0 \quad \text{and} \quad \ddot{\varphi} + 2\cot\theta\;\dot{\varphi}\dot{\theta} = 0.$$

One solution is obtained for $\varphi = \varphi_0 = const.$ Then we have:

$$\dot{\varphi} = \ddot{\varphi} = 0$$

and we find

$$\ddot{\theta} = 0 \quad \text{which implies} \quad \theta = \theta_0 + \nu t.$$

This describes a segment of a longitudinal great circle of the sphere. By symmetry, one can see that all geodesics are great circles.

Example 9.31 *Another interesting metric is that of $\mathbb{H}^2$, the hyperbolic space $(x_1 = x, x_2 = y > 0)$ with line element*

$$ds^2 = \frac{dx^2 + dy^2}{y^2},$$

or in matrix form

$$\begin{aligned}(g_{\mu\nu}) &= \begin{pmatrix} \frac{1}{y^2} & 0 \\ 0 & \frac{1}{y^2} \end{pmatrix} \\ (g^{\mu\nu}) &= \begin{pmatrix} y^2 & 0 \\ 0 & y^2 \end{pmatrix} \end{aligned} \qquad (20)$$

One way to obtain geodesics γ is to go back to the length functional l_γ and to remember that geodesics correspond to critical points of this functional. For practical computations, let us parameterize the curve by its x-coordinate as

$$(\gamma^\mu(x)) = (x, y(x)) \quad \text{which leads to} \quad (\dot{\gamma}^\mu(x)) = (1, y'(x))\,.$$

(Note that not all curves in $\mathbb{H}^2$ can be parameterized that way, but it turns out that all geodesics which are not vertical lines paralell to the y-axis can.) Then we have

$$l_\gamma = \int dx\;\frac{\sqrt{1 + (y')^2}}{y} \qquad \text{where } y = y(x)$$

and we get the Euler-Lagrange equations

$$\left(\frac{y'}{y\sqrt{1 + y'^2}}\right)' = -\frac{\sqrt{1 + y'^2}}{y^2}.$$

It can be seen that the solutions are arcs of circles in the x-y-plane with the centre on the x-axis, or vertical lines (exercise).

10 General Relativity

We now discuss how the techniques of the theory of manifolds can be used to formulate General Relativity. We will make the following basic assumptions:

1. the set M of all possible events in the universe is a manifold of dimension four,
2. on M there is a Lorentzian metric g_{ab},
3. any pointlike test-object with a mass $m > 0$ follows a timelike curve $\xi : I \to M$,
4. in the case of free fall, i.e. in the absence of forces other than gravity, ξ is a geodesic.

From now on we will restrict attention to manifolds of dimension four and for convenience we will also choose physical units such that $c = 1$.

We have dropped the assumption that $M = \mathbb{R}^4$, which is difficult to verify experimentally, but we still assume that M locally looks like $\mathbb{R}^4$, in the sense that it is a 4-dimensional manifold. Using this formulation, all charts are treated on an equal footing and there are no preferred choices of coordinates. In other words coordinates have no intrinsic physical meaning.

We have also replaced the rather rigid structure of η_{ab}, that was of such fundamental importance in Minkowski space, by a Lorentzian metric g_{ab}. This metric may vary from point to point and in some sense we may think of it as the gravitational field. It certainly determines the geodesics and the timelike geodesics are the same for any object, regardless of its mass (weak equivalence principle).

Of course we have yet to supply an equation of motion for the physical field g_{ab}. Because the gravitational field depends on the mass distribution of the matter in the universe, the equation of motion should also depend on this matter.

10.1 Kinematics of General Relativity

In this section we will supply some kinematical aspects of General Relativity, describing how observers can measure properties of particles. We will also investigate how the framework that we have developed so far can implement the strong equivalence principle, that the outcome of any local experiment in a freely falling laboratory is independent of the initial position and velocity of the laboratory. In addition we will see how the metric g_{ab} can encode gravity.

Let us now consider an object, e.g. a particle, following a timelike trajectory $\xi : I \to M$ in a Lorentzian manifold M. Our first purpose is to define notions of future and past on M. At each point $p \in M$, the tangent space T_pM is isomorphic to Minkowski space, so it has two light cones and we can choose one of these to be the future and the other the past. We can make such a choice for any $p \in M$, but it is not obvious that we can make this choice in a way that depends smoothly (or continuously) on the base point $p \in M$. Such a smooth dependence is desirable in order to avoid pathological situations, where a smooth timelike curve suddenly switches from being future pointing to past pointing. For this reason we will only consider spacetimes which are time-oriented in the following sense:

Definition 10.1 *A spacetime* (M, g_{ab}) *is called* time-orientable *when there exists a vector field* T^a *on* M *which is timelike at every point of* M*. Such a vector field determines a* time-orientation*: a timelike vector* $v^a \in T_pM$ *is called future pointing if it is in the same light cone as* T^a *(i.e. if* $v^aT_a < 0$*) and past pointing otherwise (*$v^aT_a > 0$*).*

A spacetime is called time-oriented *if it is time-orientable and a time-orientation has been chosen.*

Not every spacetime is time-orientable. However, when a spacetime is time-orientable, then there are many timelike vector fields T^a that determine the same time-orientation. Mathematically speaking, a time-orientation is an equivalence class of timelike vector fields, where two such vector fields T^a and T'^a are equivalent if their vectors lie in the same light cone at every point $p \in M$. Because M is connected and T^a and T'^a are both timelike, it suffices to verify this condition at one point. It follows that a time-orientable Lorentzian manifold has exactly two time-orientations. We will always assume that a choice of time-orientation can and has been made. Moreover, the trajectories of massive objects and observers are always assumed to be future pointing.

We return to our particle, following a timelike, future pointing trajectory $\xi : I \to M$ in the Lorentzian manifold M. In analogy to Special Relativity we can use the proper time along ξ as a parameter:

Theorem 10.2 *A time-like curve* ξ *in a (time-oriented) Lorentzian manifold* (M, g_{ab}) *can always be parameterised by proper time, i.e. such that* $g_{ab}(\xi(\tau))\dot{\xi}^a(\tau)\dot{\xi}^b(\tau) = -1$*, without changing its time-orientation.*

The proof is analogous to that of Theorem 6.5. We may define the velocity and acceleration in terms of the proper time by

$$v^a := \dot{\xi}^a, \qquad a^a := \dot{\xi}^b \nabla_b \dot{\xi}^a.$$

Note that both expressions are independent of any choice of coordinates, whereas the expression $\ddot{\xi}^\mu(\tau)$ does depend on this choice. The problem is that, to quantify the change of v^a at different values of the proper time τ, we need to compare tangent vectors at different points of the spacetime M. To do this in a coordinate independent way, we need to use the covariant derivative. Also note that the norm of v^a is constantly 1, just like in Special Relativity. It follows that $v_a a^a = 0$.

If the rest mass of the particle is m_0, then its energy-momentum and the force are defined as the dual vectors

$$P_a := m_0 v_a, \qquad F_a := v^b \nabla_b P_a.$$

Here F_a includes all forces other than gravity. One may easily verify that if m_0 is constant, $F_a = 0$ is equivalent to the geodesic equation $v^a \nabla_a v^b = 0$.

Let us now consider an observer, who follows another future pointing, timelike trajectory $s \mapsto \alpha(s)$, which we also assume to be parameterized by proper time. The velocity of the observer is $\dot{\alpha}^a$. If α and ξ both go through the point $p \in M$, then the observer may measure properties of the particle. Choosing an inertial frame in T_pM and comparing with Special Relativity we define the energy E and the (spatial) momentum $\mathbf{P}_a$, as measured by the observer at p, to be

$$E = -P_a \dot{\alpha}^a = -m_0 \dot{\xi}_a \dot{\alpha}^a, \qquad \mathbf{P}_a = P_a - E\dot{\alpha}_a.$$

To obtain $\mathbf{P}_a$ we simply projected out the component along $\dot{\alpha}^a$.

The kinematical concepts we introduced above make use of the fact that for any $p \in M$, the tangent space $T_p(M)$ equipped with the pseudo-inner product $g_{ab}(p)$ is isomorphic to Minkowski space. We now want to discuss how this identification between T_pM and Minkowski space can be made even more precise.

Theorem 10.3 (Riemannian normal coordinates) *At any $p \in M$ there is a diffeomorphism $\exp_p$ of an open neighbourhood $W \subset T_pM$ containing 0, onto an open neighbourhood $U \subset M$ containing $p \in U$, with the following properties:*

1. *For any $v^a \in W$ and $\lambda \in [0,1]$, $\lambda v^a \in W$,*

2. *For any $v^a \in W$, the curve $\gamma : [0,1] \to M$ defined by $\gamma(\lambda) := \exp_p(\lambda v^a)$ is a geodesic with $\gamma(0) = p$ and $\dot{\gamma}^a(0) = v^a$.*

The proof uses results on how solutions of differential equations depend on their initial data. These can be used to show that the map $\exp_p$ is well-defined and C^∞. The fact that it is a diffeomorphism on some neighborhood W of 0 follows from the inverse function theorem.

If we choose an orthonormal basis e_μ of T_pM, with e_0 timelike and with a dual basis $e^{*\mu}$, then we can construct a chart $\psi : U \to V$ by setting $\psi(x) := (e^{*1}(\exp_p(x)), \ldots, e^{*n}(\exp_p(x)))$. The coordinate system $x^{mu} : e^{*\mu} \circ \exp_p$ has the following properties:

$$x^\mu(p) = 0, \qquad g_{\mu\nu}(p) = \eta_{\mu\nu}, \qquad \partial_{x^\rho} g_{\mu\nu}(p) = 0.$$

The last equality is equivalent to the vanishing of the Christoffel symbol at $p \in M$ in Riemannian normal coordinates, which follows from the geodesic equation (19) for the curves $\lambda \mapsto \lambda x^\mu$. This shows that the identification of T_pM with Minkowski space can even be made up to first order derivatives of the metric.

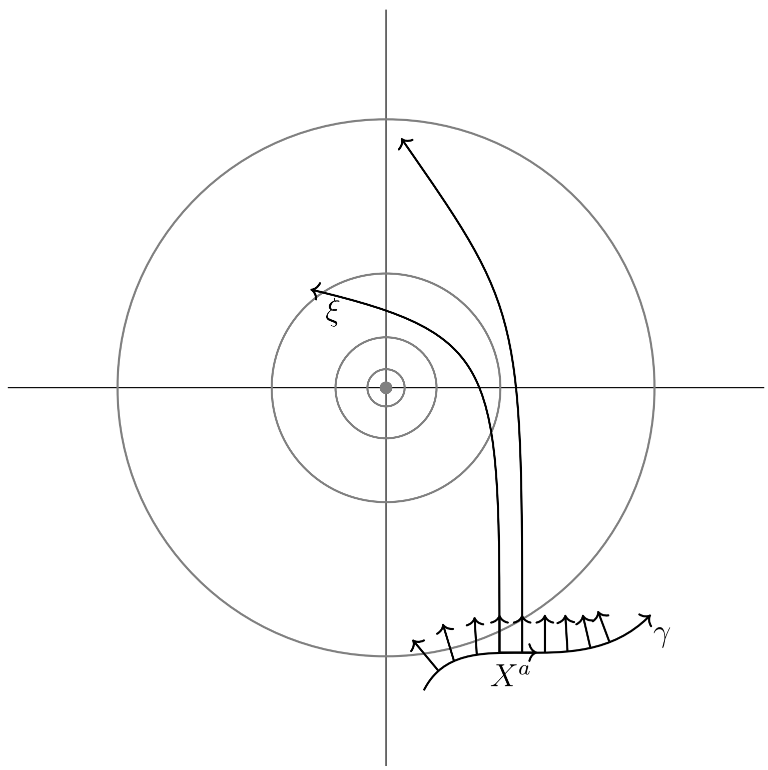
Note that the choice of coordinates heavily depends on the fact that $p \in M$ was fixed in advance, so in this coordinate system these special properties typically fail at any other point. However, it is possible to choose coordinates along any curve such that the metric takes the form $g_{\mu\nu} = \eta_{\mu\nu}$ and the derivatives $\partial_{x^\rho} g_{\mu\nu} = 0$ vanish all along the curve. (Such coordinates are called Fermi-Walker coordinates.) This explains how General Relativity satisfies the strong equivalence principle.

One cannot expect to do much better than Fermi-Walker coordinates, because the Riemann curvature tensor ${R_{abc}}^d$ is independent of the choice of coordinates, but its coordinate expression only depends on derivatives of the metric up to second order.

Definition 10.4 *A spacetime (M, g_{ab}) is called* flat*, when the Riemann curvature tensor vanishes, ${R_{abc}}^d = 0$.*

One can show that a flat spacetime can be covered by charts on which the metric takes the form $\eta_{\mu\nu}$ throughout the domain of the chart. I.e., we can cover M by open regions which look like open regions of Minkowski space (including the metric).

Our framework very nicely implements the strong equivalence principle, but how exactly does the metric g_{ab} encode gravity? The answer to this question is a bit subtle, because we like to think of gravity as a force which causes acceleration. In General Relativity, however, there is no longer any notion of absolute acceleration, because there is no preferred structure with respect to which something could be accelerated. (This is the ultimate consequence of treating all charts on an equal footing.) Nevertheless, we can investigate the relative acceleration of two freely falling objects, which will explain the relation between General Relativity and gravity.



Suppose that $\xi : (-a, a) \to M$ is a timelike, future pointing geodesic with $p := \xi(0)$ and let $X^a \in T_pM$ be a vector which is orthogonal to $\dot{\xi}^a$. We wish to displace $p = \xi(0)$ to an infinitesimally close point in the direction of X^a, track the displaced point as it follows its geodesic and then see how the displacement vector changes.

To make these ideas precise we construct a map $\gamma : (-a, a)^2 \to M$ as follows (for a suitably small $a > 0$): We define the geodesic $\gamma(s,0) := \exp_p(sX^a)$ and we extend $X^a \in T_pM$ to the vector field $X^a := \partial_s \gamma^a(s,0)$ along the curve. Note that X^a is parallely transported along the curve, because it is just the tangent vector field of a geodesic.

Next we let $T^a(\gamma(s,0))$ be the parallel transport of $\dot{\xi}^a(0) \in T_pM$ along the curve $\gamma(s,0)$. We then have $X^a T_a = 0$ along the curve $\gamma(s,0)$, because parallel transports preserve inner products. Finally we define $\gamma(s,t) := \exp_{\gamma(s,0)}(tT^a(\gamma(s,0)))$, i.e. we consider the geodesics generated by the vectors T^a on the curve $\gamma(s,0)$. We now extend T^a and X^a to the entire range of γ by setting $T^a := \partial_t \gamma^a(s,t)$ and $X^a := \partial_s \gamma^a(s,t)$. Note that T^a is a tangent to a geodesic $t \mapsto \gamma(s,t)$, so it is parallelly transported along this geodesic.

We have now constructed a one-parameter family of geodesics $t \mapsto \gamma(s,t)$ with $\gamma(0,t) = \xi(t)$. We may think of X^a on $\xi(t)$ as the relative position of, or the infinitesimal displacement to, a nearby geodesic. The relative velocity of the nearby geodesic is given by $V^b := T^a \nabla_a X^b$ and the relative acceleration by

$$A^c := T^a \nabla_a (T^b \nabla_b X^c).$$

Theorem 10.5 (Geodesic Deviation Equation) *In the notations above,*

$$A^a = -R_{bcd}{}^a T^b X^c T^d. \tag{21}$$

Proof: We first show that $T^a \nabla_a X^b = X^a \nabla_a T^b$. This follows from $\partial_t \partial_s = \partial_s \partial_t$ as follows. For any smooth function f on M we have $X(f) = \partial_s (f \circ \gamma)(s,t)$ and $T(f) = \partial_t (f \circ \gamma)(s,t)$. The right-hand sides can once again be interpreted as smooth functions on the range of γ, to which the vectors X and T can be applied. In this way we find $T(X(f)) = \partial_t \partial_s (f \circ \gamma) = X(T(f))$. Using the calculus of covariant derivatives, this means

$$\begin{aligned}
0 &= T^a \nabla_a (X^b \nabla_b f) - X^a \nabla_a (T^b \nabla_b f) \\
&= (T^a \nabla_a X^b) \nabla_b f - (X^a \nabla_a T^b) \nabla_b f + T^a X^b (\nabla_a \nabla_b f - \nabla_b \nabla_a f) \\
&= (T^a \nabla_a X^b) \nabla_b f - (X^a \nabla_a T^b) \nabla_b f.
\end{aligned}$$

Since f was arbitrary, we must have $T^a \nabla_a X^b = X^a \nabla_a T^b$.

Incidentally, note that $T^a T_a = 1$ on the range of γ and hence, by the geodesic equation for T^a,

$$\begin{aligned}
T^a \nabla_a (X^b T_b) &= T_b T^a \nabla_a (X^b) = T_b X^a \nabla_a (T^b) \\
&= \frac{1}{2} X^a \nabla_a (T_b T^b) = \frac{1}{2} X^a \nabla_a 1 = 0.
\end{aligned}$$

Since $X_b T^b = 0$ on $\gamma(s,0)$, it follows that $X_b T^b = 0$ on the entire range of γ.

Now we use the geodesic equation for T^a and the definition of the Riemann curvature tensor to compute:

$$\begin{aligned}
A^c &= T^a\nabla_a(T^b\nabla_b X^c) = T^a\nabla_a(X^b\nabla_b T^c)\\
&= (T^a\nabla_a X^b)\nabla_b T^c + X^b T^a \nabla_a\nabla_b T^c\\
&= (X^a\nabla_a T^b)\nabla_b T^c + X^b T^a\nabla_b\nabla_a T^c - X^b T^a R_{abd}{}^c T^d\\
&= X^a\nabla_a(T^b\nabla_b T^c) - R_{abd}{}^c T^a X^b T^d\\
&= -R_{abd}{}^c T^a X^b T^d.
\end{aligned}$$

Relabelling indices yields the result. □

Equation (21) shows that nearby geodesics can exhibit relative acceleration, which is described by the Riemann curvature of the metric. Roughly speaking, the gravitational field corresponds to the curvature of the metric. This effect goes under the name *gravitational tidal forces*.

10.2 Dynamics of General Relativity

In Newton's theory of gravity, the strength of the gravitational force on a test-mass is determined by the matter in the universe. In General Relativity, gravity is encoded by the metric g_{ab}, so it seems reasonable that g_{ab} should satisfy an equation of motion, Einstein's equation, which also involves the matter in the universe. This equation is the topic of the present section.

Because General Relativity should reduce to Newton's Theory of Gravity in a suitable limit, let us first recall how the dynamics of the latter theory is formulated. For this we revert to classical spacetime, with a classical inertial coordinate system $(t, \mathbf{x})$, where t is an absolute time coordinate. Let us consider a continuous distribution of matter, described by a mass density ρ. This mass density gives rise to a gravitational potential Φ, defined by

$$\Phi(\mathbf{x}) := \int \frac{-G_N\rho(\mathbf{x}')}{\|\mathbf{x}-\mathbf{x}'\|}\mathrm{d}\,\mathbf{x}',$$

so that the acceleration that the mass distribution causes on any test-mass m is

$$\mathbf{g}(\mathbf{x}) = -\nabla\Phi(\mathbf{x}) = \int -G_N\rho(\mathbf{x}')\frac{\mathbf{x}-\mathbf{x}'}{\|\mathbf{x}-\mathbf{x}'\|^3}\mathrm{d}\,\mathbf{x}'.$$

These are continuous versions of the usual point mass formula for Newton's law of gravitation, where the test-mass m drops out, due to the (weak) equivalence principle. Note that ρ can be recovered from Poisson's equation:

$$\Delta\Phi(\mathbf{x}) = -\nabla \cdot \mathbf{g}(\mathbf{x}) = 4\pi G_N \rho(\mathbf{x}), \tag{22}$$

where all derivatives are in the spatial directions. This differential equation tells us how the gravitational potential Φ depends on the mass density distribution ρ. Einstein's equation should be analogous to Poisson's equation, but formulated in a relativistic context.

We may compare the acceleration of a test-mass in Newtonian gravity to the geodesics of General Relativity:

$$\ddot{\xi}^i(t) = -\nabla^i\Phi(\xi(t)) \quad \longleftrightarrow \ddot{\xi}^\mu(t) = -\Gamma^\mu{}_{\nu\rho}\dot{\xi}^\nu(t)\dot{\xi}^\rho(t),$$

where $i = 1, 2, 3$ and $\mu = 0, 1, 2, 3$. Now suppose that we can choose coordinates x^μ such that x^0 is a time coordinate in which the variations of the metric are negligible, i.e. the gravitational field is constant. Also suppose that the velocity of the test-mass is small, so that $\dot{\xi}^\rho(t)$ is close to $(1, 0, 0, 0)$. The comparison above then suggests that

$$\nabla^i\Phi \quad \longleftrightarrow \quad \Gamma^i{}_{00} = -\frac{1}{2}g^{i\mu}\partial_\mu g_{00}$$

and Φ might correspond to $\Phi \simeq -\frac{1}{2} - \frac{1}{2}g_{00}$. (The term $-\frac{1}{2}$ is needed to find the Minkowski metric component $g_{00} = -1$ in the absence of masses.)

To find a suitable correspondence for $\Delta\Phi$ we need to consider second order derivatives. For this purpose we will investigate the gravitational tidal forces from the point of view of Newton's Theory of Gravitation. Given $\Phi(\mathbf{x})$ we can consider two freely falling particles at time t, at positions $\mathbf{x}$ and $\mathbf{x} + \mathbf{h}$. Their respective accelerations will be $-\nabla\Phi(\mathbf{x})$ and $-\nabla\Phi(\mathbf{x} + \mathbf{h})$, so the relative acceleration is

$$-\nabla\Phi(\mathbf{x} + \mathbf{h}) + \nabla\Phi(\mathbf{x}).$$

To find the relative acceleration of a particle which is infinitesimally close to the particle at $\mathbf{x}$ we take the derivative with respect to $\mathbf{h}$ at $\mathbf{h} = 0$ in the direction $\mathbf{X}$, which yields

$$-(\mathbf{X} \cdot \nabla)\nabla\Phi(\mathbf{x}).$$

This is the gravitational tidal force in the setting of Newtonian gravity and should be compared with the result in General Relativity, Theorem 10.5:

$$-R_{bcd}{}^{a}T^{b}T^{d} \quad \longleftrightarrow \quad -\nabla_{c}\nabla^{a}\Phi.$$

Here the timelike vector T^b can be compared to the classical flow of absolute time and, due to the antisymmetry properties of the Riemann curvature tensor, the indices a and b only range over the spatial coordinates. Contracting over a and b yields a suitable analogue of the term $\Box\Phi$ in Poisson's equation:

$$R_{bd}T^{b}T^{d} \quad \longleftrightarrow \quad \Delta\Phi. \tag{23}$$

Now we turn to a relativistic formulation of the mass density ρ. Keeping in mind the lesson of Special Relativity that mass is energy, it makes sense to compare ρ with an energy density. For a continuous (or C^{∞}) distribution of matter in a spacetime M, we define the energy density at a point $p \in M$, from the perspective of an observer at $p \in M$, to be the quotient of the mass in a spatial region B divided by the volume of B, in the limit where B shrinks to the point p. To see what this limit entails we consider an example, consisting of a matter distribution in Minkowski space, which we assume to consist of particles which are all at rest in some inertial frame. Now consider two inertial observers, related by a Lorentz boost, but both going through the same point $p \in M$ with velocity vectors T^a and T'^b, respectively. Suppose that the first observer is also at rest, so he measures an energy density which consists entirely of the rest masses of the particles, leading to a mass density ρ, say. To compute the result for the second observer, we apply a Lorentz transformation. We recall the following two relevant effects: the rest masses get multiplied by a factor $\gamma = T_a T'^a$ and the lengths in the direction of motion get multiplied by a factor γ^{-1}. The energy density for the second observer therefore becomes $\rho' = \gamma^2\rho$. The results so far can be formulated very neatly as follows:

$$\rho' = T_{ab}T'^{a}T'^{b}, \qquad T_{ab} := \rho T_a T_b.$$

Generalizing this idea leads to the following:

Tenet 10.6 *A continuous matter distribution in a spacetime (M, g_{ab}) can be described by a symmetric tensor $T_{ab} = T_{ba}$, the* stress-energy-momentum tensor *(which we will often abbreviate to* stress tensor*, for convenience). For an observer with velocity $T^a \in T_pM$, the components of T_{ab} are interpreted as follows:*

1. *$T_{ab}T^aT^b$ is the energy density at p,*

2. *$T_{ab}v^aT^b$ is the momentum in the direction v^a, if $v_aT^a = 0$,*

3. *$T_{ab}v^aw^b$ is the (v^a, w^b)-component of the stress tensor, if $v_aT^a = 0$ and $w_aT^a = 0$.*

For normal matter, the stress tensor is conserved, $\nabla^a T_{ab} = 0$, *and it satisfies the* weak energy condition*: $T_{ab}T^aT^b \geq 0$ for all timelike vectors T^a.*

In Special Relativity, the conservation equation $\nabla^a T_{ab} = 0$ is equivalent to the conservation of energy-momentum. In General Relativity, this interpretation only holds when gravity does not exert forces on the matter.

Thus, given a timelike vector T^a and a stress tensor T_{ab} we choose the quantity $T_{ab}T^aT^b$ to correspond with the mass density ρ for the observer described by T^a. In order to obtaine the appropriate analogue of Poisson's equation in General Relativity, we note that we may also choose instead the correspondence

$$2\left(T_{ab} - \frac{1}{2}g_{ab}T^c_{\ c}\right)T^aT^b \quad \longleftrightarrow \quad \rho. \tag{24}$$

To see this, we note that the main contribution to $T^c_{\ c}$ comes from the energy density $T_{ab}T^aT^b$, because the stress components in the stress tensor are typically comparably small in units where $c = 1$. The values of the left-hand side is therefore very close to $T_{ab}T^aT^b$, so the correspondence is still physically reasonable.

Together with the correspondence in Equation (23) we may now rewrite Poisson's equation as $R_{ab}T^aT^b = 8\pi G_N \left(T_{ab} - \frac{1}{2}g_{ab}T^c_{\ c}\right)T^aT^b$. Assuming that this holds for all timelike vectors T^a leads to

$$R_{ab} = 8\pi G_N \left(T_{ab} - \frac{1}{2}g_{ab}T^c_{\ c}\right).$$

In order to obtain an equation which just contains T_{ab} on the right-hand side we proceed as follows. Contracting over the indices a and b we find $R = -8\pi G_N T^c_{\ c}$. This can be used to eliminate the term $T^c_{\ c}$ in favour of R, which can be brought to the other side of the equation. The result is *Einstein's equation*:

$$G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R = 8\pi G_N T_{ab}. \tag{25}$$

Einstein's equation is a non-linear partial differential equation for the components of the metric g_{ab}, which occur in Einstein's tensor G_{ab} with derivatives up to second order. The metric also occurs in the equations of motion of any matter in the universe, which, in turn, determine the form of T_{ab}. One ought to solve all these equations of motion together.

From Bianchi's identity, $\nabla^a G_{ab} = 0$ (cf. Theorem 9.20) we immediately find $\nabla^a T_{ab} = 0$. This equation already imposes a strong, but physically justifiable restriction on the matter in the universe. Note that if we had chosen the correspondence $T_{ab} T^a T^b = \rho$ in our derivation of Einstein's equation, we would instead have found $R_{ab} = 8\pi G_N T_{ab}$ which would not have been consistent with the conservation of the stress tensor. In many cases, $\nabla^a T_{ab} = 0$ encodes the full equations of motion of this matter.

It is possible to formulate Einstein's equation as a Cauchy problem. This means that one can prescribe a manifold Σ of dimension 3 and initial values for the metric g_{ab} and its "normal" derivatives on Σ and then construct a unique, maximal manifold M with a Lorentzian metric g_{ab} which contains Σ as a spacelike hypersurface and whose metric has the correct data on Σ. We will not consider the rather technical aspects of this formulation.

10.3 Properties of General Relativity

With Einstein's equation in place, the content of General Relativity can be formulated concisely as follows (cf. [9]):

Tenet 10.7 (General Relativity) *Spacetime is a four-dimensional manifold with a Lorentzian metric, whose relation to the matter distribution in spacetime is given by Einstein's equation (25).*

Before we proceed to discuss a number of physical predictions and applications of this theory, let us pause to mention a few general properties of it.

If the Lorentzian manifold (M, g_{ab}) is a solution to Einstein's equation for some stress tensor T_{ab}, then we can immediately construct further solutions using embeddings:

Definition 10.8 *An* embedding *of manifolds M, M' of the same dimension, is a smooth injective map $\psi : M \to M'$ such that the range $\psi(M) \subset M'$ is open and the map $\psi^{-1} : \psi(M) \to M$ is also smooth.*

A diffeomorphism *is a surjective embedding $\psi : M \to M$. Because the inverse $\psi^{-1} : M \to M$ is also a diffeomorphism, the set of all diffeomorphisms of a manifold M forms a group.*

One particular kind of embedding is the *canonical inclusion map $U \to M$* of a subset $U \subset M$. Another kind is the charts that were used to define manifolds in the first place.

Given any embedding $\psi : M \to M'$, we can use ψ and ψ^{-1} to map tangent vectors and tensors from a point $p \in M$ to a point $\psi(p) \in M'$ and back. Assuming that all laws of physics can be formulated in terms of tensor fields (or similar geometric objects), we can use ψ and ψ^{-1} to transport all fields from M' to M, just as we did for charts. Now suppose that M' has a Lorentzian metric g'_{ab} and various tensor fields collectively denoted by ϕ, and let use denote by $g_{ab} := \psi^* g'_{ab}$ respectively $\phi = \psi^* \phi'$ the metric and tensor fields that we obtain on M via the embedding ψ. If (g'_{ab}, ϕ') satisfies Einstein's equation on M', then so does (g_{ab}, ϕ) on M. In formulae, if we write Einstein's equation as $E_{ab}(g, \phi) = 0$ (E_{ab} being the left minus right side), then E_{ab} is such that

$$E_{ab}(g, \phi) = \psi^* E_{ab}(g', \phi')$$

for any embedding $\psi : M \to M'$ and (g_{ab}, ϕ) related to (g'_{ab}, ϕ) in the manner described above. The above property of the Einstein field equations is often referred to as *general covariance*. One can show that general covariance implies, under certain reasonable technical assumptions, that E_{ab} must be a local differential operator, i.e. $E_{ab}|_p$ can be written, at each p as a contraction of $(g_{ab}, g^{ab}, \phi, \dots, \nabla_{(a_1} \dots \nabla_{a_r)} \phi, R_{abcd}, \dots, \nabla_{(a_1} \dots \nabla_{a_s)} R_{abcd})|_p$, where the orders r, s are locally bounded (Peetre's theorem, Thomas replacement theorem). Of course, we want E_{ab} to be given by the Einstein field equation with a local matter stress tensor, so general covariance along does not imply Einstein's equation but would hold for a broad class of field equations.

Intuitively speaking, general covariance means that the laws of physics cannot contain any non-dynamical quantities that refer to spacetime other than the metric g_{ab} (and possibly the time-orientation and orientation of M), and any other dynamical fields ϕ on a manifold M, and that the laws are described by partial differential equations (Einstein field equation). An example of a non-dynamical quantity would be a preferred coordinate system, which we can think of as a collection of four scalar fields $x^0, \dots, x^3 : M \to \mathbb{R}$ such that $(\nabla_a x^0, \dots, \nabla_a x^3)|_p$ spans $T_p^* M$ at each point p. Of course, there is no difficulty having a *dynamical* coordinate system. In this case, our four scalar fields would have to enter the stress tensor T_{ab} in a generally covariant way. Such theories are sometimes called "Einstein-ether"-theories, because the four scalar fields locally define a frame. From a general perspective, there is nothing special about such theories compared to theories with other kinds of matter fields.

Another consequence of general covariance is that the individual points $p \in M$ have no intrinsic physical meaning. In fact, we can only identify a point p by giving it a physical description, e.g. that it is the point where certain physical fields (including the metric g_{ab}) take certain prescribed values. Now, if we apply a diffeomorphism ψ to the Lorentzian manifold (M, g_{ab}) to obtain the new field configurations $(M, \psi^* g'_{ab})$, then a point $p \in M$ remains fixed, but the field configurations do not, so the physical description will now determine the point $\psi(p)$, rather than p.

Apart from being generally covariant in the above sense, which, as we have seen, is a feature of a broad class of theories, Einstein's equation are also local in another, more subtle sense. This property has to do with the fact that the Einstein field equations are basically *hyperbolic* partial differential equations. The prototype of such a hyperbolic equation is the wave equation for a scalar field $\phi : M \to \mathbb{R}$ on a fixed spacetime (M, g_{ab}),

$$\nabla_a \nabla^a \phi = j \ .$$

Suppose (M, g_{ab}) does not have any grossly pathological causal features, in the sense that there is a space like embedded hyper surface $\Sigma \subset M$ with the property that every inextendible causal curve $\gamma : (a, b) \to M$ (meaning $\dot{\gamma}^a$ timeline or null everywhere) intersects Σ precisely once. Then it can be shown that the wave equation has a well-posed initial value problem: For each $j \in C_0^\infty(M)$, and each choice of $p, q \in C_0^\infty(\Sigma)$, there is a smooth solution ϕ having $\phi|_\Sigma = q, n^a \nabla_a \phi|_\Sigma = p$, where n^a is the unit normal to Σ. Furthermore, if we change j in the causal future of Σ, then ϕ does not change in the causal past of Σ. Likewise, if we change p, q inside some compact set $K \subset \Sigma$, then ϕ does not change outside the domain of causal influence of K, i.e. outside the set of points p that can be connected to K by a causal curve. Since there are many ways of choosing Σ, j, p, q, this implies a kind of locality (i.e. local dependence on the source and initial conditions). There is a sense in which Einstein's equations are also hyperbolic and possess an initial value formulation of the above kind, although this feature is complicated by the fact that Einstein's equations are also generally covariant. A detailed discussion is outside the scope of our lectures and can be found e.g. in [8]. Let us only mention here that this feature is tightly related to the special form of E_{ab} and is not shared by most other generally covariant field equations.

10.4 Stress-Tensor and Energy Conditions

Let us list some stress tensors for various matter models.

1. Incoherent matter (a.k.a. dust): pointlike particles that follow timelike paths without influencing each other. The flow of the particles can be described by a future pointing, timelike velocity vector field u^a with $u^a u_a = 1$ with $u^a \nabla_a u^b = 0$ (so each particle follows a geodesic). The stress tensor is
$$T_{ab} = \rho u_a u_b,$$
where $\rho : M \to \mathbb{R}_{>0}$ is a positive mass density function that satisfies the continuity condition $\nabla_a(\rho u^a) = 0$.

2. Perfect fluids: analogous to dust, but now the particles exert a pressure P on each other, $P : M \to \mathbb{R}$. Using the same notations as before, the vector field u^a may no longer have geodesics as its integral curves. The stress tensor is
$$T_{ab} = \rho u_a u_b + P(u_a u_b + g_{ab}).$$
Note that the weak energy condition is satisfied, because $\gamma := u_a \dot{\alpha}^a \geq 1$ and hence $(\rho+P)\gamma^2 - P \geq \rho \geq 0$. The conservation $\nabla^a T_{ab}$ does give rise to conditions on ρ, P and u^a, which may be interpreted in terms of fluid dynamics. (In the non-relativistic limit, they reduce to conservation of mass and Euler's equation).

3. Electromagnetic fields: In Special Relativity, an electromagnetic field is described by a one-form A_a, modulo a gauge equivalence relation, $A_a \sim 0$ iff $A_a = \nabla_a \lambda$ for some function λ. From A_a one constructs a field-strength tensor $F_{ab} := 2\nabla_{[a} A_{b]}$, which is antisymmetric. Its six independent components encode the electric and magnetic fields, which, however, are observer dependent. Let $e_\mu \in T_p M$ be an orthonormal frame and consider an observer at $p \in M$ with velocity vector e_0 (which

must be timelike and future pointing). Then components of the electric and magnetic fields can be identified as follows, for this observer:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix}.$$

The energy density for the observer is $\frac{1}{2}(\|\mathbf{E}\|^2 + \|\mathbf{B}\|^2)$, which may be written in a coordinate independent way as $T_{ab}(e_0)^a(e_0)^b$ with

$$T_{ab} := F_a{}^c F_{cb} - \frac{1}{4} g_{ab} F^{cd} F_{cd}.$$

The fact that this stress tensor is conserved follows from Maxwell's equations in vacuum, which read

$$\nabla_{[a} F_{bc]} = 0, \qquad \nabla^a F_{ab} = 0.$$

4. A free scalar field: A free scalar field of mass $m \geq 0$ is a function $\phi : M \to \mathbb{R}$ satisfying the Klein-Gordon equation

$$\Box\phi + m^2\phi := \nabla^a \nabla_a \phi - m^2 \phi = 0.$$

It has a stress tensor given by

$$T_{ab} := \nabla_a \phi \nabla_b \phi - \frac{1}{2} g_{ab} (\nabla^c \phi \nabla_c \phi + m^2 \phi^2).$$

In the last two examples, the stress tensor is closely related to spacetime symmetries if any. We shall come back to this question later.

We have already mentioned that if γ is a timelike curve representing an observer with tangent $\dot{\gamma}^a = T^a$, then $\rho = T_{ab} T^a T^b$ has the interpretation of the energy density measured by this observer, whereas other components have the interpretation of the various pressures and stresses (see e.g. [2]). To discuss in general the relationship between pressures and energy density, it is sometimes useful to introduce an orthogonal tetrad adapted to γ. Such a tetrad by definition consists of 4 vectors $e^a_{(\mu)}, \mu = 0, 1, 2, 3$ such that $e^a_{(0)} = T^a$ and such that

$$g_{ab} e^a_{(\mu)} e^b_{(\nu)} = \eta_{\mu\nu}$$

where $\eta_{\mu\nu} = diag(-1,1,1,1)$ is the Minkowski metric. With this notation

$$\begin{aligned} T_{ab}e^a_{(0)}e^b_{(0)} = \rho \;\; &= \text{ energy density} \\ T_{ab}e^a_{(0)}e^b_{(i)} = P_i \;\; &= \text{ momentum density} \\ T_{ab}e^a_{(i)}e^b_{(j)} = \theta_{ij} &= \text{ stresses} \end{aligned}$$

in the "orthogonal frame" defined by the tetrad. Different tetrads defining the same space- and time orientations are related by a proper orthochronous Lorentz transformation, $e'^a_{(\mu)} = \Lambda_\mu{}^\nu e^a_{(\nu)}$, and these change the energy density, pressure density, and stress density in the usual way familiar from Special Relativity.

Energy conditions specify in general terms properties expected to hold from physical considerations. These conditions are traditionally grouped into the following categories.

a. Weak energy condition (WEC): For all timelike or null T^a we have

$$T_{ab}T^aT^b \geq 0.$$

For a perfect fluid with stress tensor $T_{ab} = \rho u_a u_b + P h_{ab}$ and fluid velocity u^a, this means that $\rho + P \geq 0$ and $\rho \geq 0$.

b. Null energy condition (NEC): We have

$$T_{ab}l^al^b \geq 0$$

for all null l^a. For a perfect fluid, it means $\rho + P \geq 0$. Note that the weak energy condition leads to the null energy condition.

c. Dominant energy condition (DEC): For all timelike T^a, the condition requires that the vector

$$f^b = -T^b{}_a T^a$$

is a timelike or null future pointing vector. Note that the dominant energy condition leads to the weak energy condition. The dominant energy condition and Einstein's equation lead to the positive mass theorem. For a perfect fluid, it means that $\rho \geq |P|$.

d. Strong energy condition (SEC): For all timelike T^a, this condition requires that

$$T_{ab}T^aT^b \geq \frac{1}{2}T^c{}_cT^aT_a.$$

Note that the strong energy condition implies the null energy condition but it does not imply the weak energy condition. The strong energy condition and Einstein's equation leads to focussing of geodesics (see Raychaudhuri equation).

It is instructive to illustrate the meaning of the various energy conditions for a perfect fluid with stress tensor $T_{ab} = \rho u_a u_b + P h_{ab}$.

Part III
Applications of General Relativity

The conceptual changes and theoretical improvements of Einstein's General Relativity are by no means purely academic. Many subtle consequences of his theory have been experimentally verified during the past century and the theory now forms the basis of our understanding of the universe at large scales (cosmology), introducing the notion of a big bang and of black holes at the centers of galaxies. Closer to home, General Relativity is needed for the proper working of the Global Positioning System (GPS), because the older, less precise descriptions of radio signals in the Earth's gravitational field lead to errors of many meters. The most direct prediction of General Relativity, however, is that gravity is a field theory, so it propagates in the form of waves, rather than acting at a distance. A direct observation of these gravitational waves is difficult, because their effects are very small, but at present several experiments are underway and it is hoped that the first gravitational waves will soon be observed, leading to novel ways to look up into the sky and study the universe we live in.

11 Spacetime Symmetries

As in other branches of theoretical physics, in order to understand the implications of Einstein's equation, we have to study solutions describing real physical situations. Given the complexity of the equations, one must either

1. find solutions by numerical methods, or

2. make stringent symmetry assumptions and try to find analytic solutions.

The numerical analysis of Einstein's equations is a very important area of General Relativity, which is, however, beyond the scope of these notes. We therefore focus on the second option. It is surprising that solutions representing the two most important physical phenomena predicted by the theory can be found in this way:

1. cosmological solutions (Friedmann-Lemaître-Robertson-Walker metric) describing a dynamical cosmos,
2. black hole solutions (Schwarzschild, Kerr metric) describing objects from which no light can escape.

Before we come to these, it is useful and necessary to understand the nature of symmetries in General Relativity.

Since we describe spacetime geometry by a pair (M, g_{ab}) consisting of a manifold M and a Lorentzian metric g_{ab}, a symmetry is a map ψ preserving this structure. Thus, first of all ψ should be a diffeomorphism of M. In order to formulate that ψ preserves g_{ab}, we need the notions of pull-back and push-forward of tensors. The notion of push-forward is defined as follows. Let v^a be a vector at $p \in M$. Choose a curve $\gamma(s)$ in M whose derivative $\dot{\gamma}^a(0)$ equals v^a. The push forward of v^a, denoted $\psi_* v^a$, is the vector at $\psi(p)$ which is equal to the derivative $(d/dt)(\psi \circ \gamma)(0)$. The notion of pull-back applies to co-vectors and is defined by duality. Let w_a be a covector at $\psi(p)$. The pull back, denoted $\psi^* w_a$ is the covector at p whose action on v^a is related by

$$(\psi^* w_a) v^a \Big|_p = w_a \psi_* v^a \Big|_{\psi(p)}.$$

The notions of pull-back and push forward apply straightforwardly to tensors of higher rank. Coordinate expressions for the push-forward and pull-back are obtained as follows. Near p and $\psi(p)$ pick local coordinates x^μ and x'^μ. Let us write $\psi(x)^\mu = x'^\mu$ for the action of ψ in these local coordinates. (Note that there is slight abuse of notation here because ψ is really a map from M to M and not between coordinate vectors. What is meant here is the composition of ψ with the coordinate charts at p and $\psi(p)$). Then, if $t_{ab\ldots c}$ is a tensor field with coordinate components $t_{\mu\nu\cdots\sigma}$, the pull-back has coordinate components

$$(\psi^\star t)_{\mu\nu\cdots\sigma}(x) = t_{\alpha\beta\cdots\gamma}(x') \frac{\partial x'^\alpha}{\partial x^\mu} \cdots \frac{\partial x'^\gamma}{\partial x^\sigma} \tag{26}$$

where (x'^μ) is viewed as a function of (x^μ) in local coordinates (note the similarity with the transformation law for covariant tensors which of course is not accidental).

Definition 11.1 *A diffeomorphism ψ on M is called an isometry (or "symmetry") if $\psi^* g_{ab} = g_{ab}$.*

The isometries of a spacetime (M, g_{ab}) always form a Lie group because one can easily show that both the composition, as well as the inverse, of isometries are again isometries, and because these operations are continuous in a natural sense.

Example 11.2 (Minkowski space) *In the case* $M = \mathbb{R}^4, g_{ab} = \eta_{ab}$*, the isometries consist precisely of translations and Lorentz-transformations. The group of all isometries is isomorphic to* $O(3,1) \ltimes \mathbb{R}^4$*, the Poincare group.*

Just as the spacetime symmetries (isometries) form a Lie group, *infinitesimal symmetries* are associated with the Lie algebra of this group. Geometrically, infinitesimal symmetries are represented by vector fields ξ^a on M which, intuitively, describe "infinitesimal displacements". This can be made precise as follows. We call $\{\psi_t\}_{t\in\mathbb{R}}$ a 1-parameter group of diffeomorphisms if $\psi_t \circ \psi_s = \psi_{t+s}$ for all $s, t \in \mathbb{R}$ and if $\psi_{t=0} = id_M =$ identity on M. Given a 1-parameter group of diffeomorphisms, we can define a corresponding vector field ξ^a by the condition that

$$\frac{\partial}{\partial t} f(\psi_t(x))\Big|_{t=0} = \xi^a(x)\nabla_a f(x) \tag{27}$$

holds for all smooth real valued functions f on M. Given a covariant tensor field $t_{ab...c}$ on M, we call the operation

$$(\mathcal{L}_\xi t)_{ab...c} := \frac{\partial}{\partial t}(\psi_t^\star t)_{ab...c}\Big|_{t=0}$$

the *Lie-derivative*. The Lie derivative is independent of any choice of covariant derivative operator, but if we are given any covariant derivative operator, we may express the Lie derivative in terms of it:

$$(\mathcal{L}_\xi t)_{ab...c} = \xi^d \nabla_d t_{ab\cdots c} - (\nabla_a \xi^d) t_{db\cdots c} - (\nabla_b \xi^d) t_{ad\cdots c} \cdots - (\nabla_c \xi^d) t_{ab\cdots d}$$

In particular, we may choose $\nabla_a = \partial_a$ to be the flat derivative operator associated with some local coordinate system. This give a practical way to calculate the Lie derivative in coordinates, and to verify this formula using the coordinate expression for the pull-back, see (26).

Exercise 11.3 *Check that the right side is indeed independent of the choice of covariant derivative operator.*

Definition 11.4 *A vector field on M is called "Killing" for a spacetime metric g_{ab} on M if $(\mathcal{L}_\xi g)_{ab} = 0$. (Using the above expression for the Lie-derivative with ∇_a equal to the Levi Civita connection for g_{ab}, this is equivalent to*

$$\nabla_a \xi_b + \nabla_b \xi_a = 0 \quad (\textit{for } \xi_a = g_{ab}\xi^b)$$

since $\nabla_a g_{bc} = 0$.)

The relationship between Killing vectors and symmetries is the following. If $\{\psi_t\}_{t\in\mathbb{R}}$ is a 1-parameter group of symmetries of g_{ab}, then by definition, $\mathcal{L}_\xi g_{ab} = 0$. Conversely, if ξ^a is a Killing vector field, then there exist a 1-parameter group of symmetries $\{\psi_t\}_{t\in\mathbb{R}}$ such that (28) holds. Just as the isometries of a spacetime form a Lie group, the infinitesimal symmetries (Killing vector fields) form a Lie algebra whose commutator is given by the commutator of vector fields (see exercises).

Example 11.5 *For the symmetries in Example 11.2 we have:*

1. *For the translations in the z-direction, the associated infinitesmal generator (Killing vector field) is $(\xi^\mu) = (0, 0, 0, 1)$ in inertial coordinates $(x^\mu) = (t, x, y, z)$.*
2. *For rotations around the z-axis, $(\xi^\mu) = (0, 0, -z, y)$, and*
3. *for boosts along the x-axis, $(\xi^\mu) = (x, t, 0, 0)$.*

12 Cosmological Solutions to Einstein's Equation (Maximally Symmetric Universes)

Maximally symmetric universes are among the simplest, yet most important solutions to Einstein's equations. The form of their metric is restricted, up to one free function of time, by symmetry assumptions, whereas this function is determined by Einstein's equations. Let us first discuss the symmetries. We assume that $M = \mathbb{R} \times \Sigma$, and we assume that the line element is of the general form

$$ds^2 = -dt^2 + a(t)^2 \gamma_{ij}(x) dx^i dx^j$$

where $i, j = 1, 2, 3$ are coordinates on Σ. It is furthermore assumed that, for each fixed t, $\gamma_{ij} dx^i dx^j$ gives Σ the structure of a homogeneous, isotropic Riemannian manifold. Such a manifold by definition satisfies the following conditions:

1. Isotropy: For each $p \in \Sigma$ and any pair of unit vectors v^i, w^j at p there exists an isometry ψ of (Σ, γ_{ij}) such that $(\psi_\star v)^j = w^j$. For each orientation of Σ, there exists an isometry reversing the orientation.

2. Homogeneity: For each $p, p' \in \Sigma$ there exists an isometry ψ such that $\psi(p) = p'$.

The factor $a(t) > 0$ is called the *scale factor*. It is essentially the only non-trivial ingredient in the line element because it turns out that (Σ, γ_{ij}) must be maximally symmetric Riemannian spaces whose metric is basically fixed (locally) once we specify the (necessarily constant) curvature of (Σ, γ_{ij}). In order to see that (Σ, γ_{ij}) is basically "just symmetry", consider any fixed point $p \in \Sigma$, and let G be the isometry group. We may then define the "isotropy subgroup" G_0 leaving this point invariant. Any isometry $\psi \in G_0$ induces the linear isometric map $v^j \mapsto \psi_* v^j$ in the tangent space $T_p\Sigma$. Thus, G_0 is a subgroup of the group of linear isometric transformations of $T_p\Sigma$ (relative to the metric $\gamma_{ij}|_p$), and hence isomorphic to a subgroup of $O(n)$ (in n spatial dimensions). However, G_0 must actually be equal to $O(n)$. Otherwise, we could construct an invariant vector v^i (under G_0), or an invariant orientation. This is forbidden by our requirement of isotropy. If p' is any other point in Σ, there is an element $\psi \in G$ which carries p to p' by the homogeneity assumption, and this element must be unique up to $\psi \mapsto \psi \circ \psi'$, where $\psi' \in G_0$. Hence, we can put points in Σ into one-to-one correspondence with elements of G modulo elements of $G_0 \cong O(n)$ (acting by right multiplication on G). In other words, Σ must be a quotient $G/O(n)$ of dimension n. It is then possible to see that the only possibilities for G can be $G = O(n+1), E(n), O(n,1)$, i.e. Σ is a maximally symmetric space. Furthermore, using this information, one can derive the local form of γ_{ij} using methods from Lie group theory.

Here, we will derive the local form γ_{ij} by a different, more explicit "by hand" method. One can show that a maximally symmetric space of dimension n necessarily has ([5]) a Riemann tensor of the form

$$\mathcal{R}_{ijkl} \;\; = \;\; k\,(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk}),$$

where k in the last line is a real constant, and $\mathcal{R}_{ijkl}$ is the Riemann tensor of γ_{ij} (not to be confused with the $ijkl$-component of the Riemann tensor of the full spacetime metric!) [6]. In $n = 3$ spatial dimensions, the case of

[6] As an aside, we note that this condition can also be stated as $D_i \mathcal{R}_{jklm} = 0$, where D_i is the Levi-Civita connection of γ_{ij}

interest for cosmology, possible solutions to these equations will turn out to be locally isometric to

$$\begin{aligned} \Sigma \;\; &= \;\; \begin{cases} \mathbb{R}^3 & k=0, \text{ 3 dimensional flat space,} \\ \mathbb{H}^3 & k<0, \text{ 3 dimensional hyperbolic space,} \\ \mathbb{S}^3 & k>0, \text{ 3 dimensional sphere} \end{cases} \\ &\cong \;\; \begin{cases} {}^{E(3)}/_{O(3)} \\ {}^{O(3,1)}/_{O(3)} \\ {}^{O(4)}/_{O(3)} \end{cases} , \end{aligned}$$

as we had already argued before. It is clear that our condition on the Riemann tensor (28) is preserved if we take a quotient of these spaces by a suitable finite subgroup of the isometry group, so we cannot expect to do better than finding the general solution up to quotients. However, as it turns out, these then give the general solution. We now derive these results by working out the consequences of our equation (28) for the Riemann tensor in $n = 3$ dimensions. In $n = 3$ spatial dimensions, the Ricci tensor and the Riemann tensor each have six independent components, so nothing is gained or lost by taking the trace of (28) (i.e. contraction with γ^{jl}). This gives

$$\mathcal{R}_{ij} = 2k\gamma_{ij}.$$

By the isotropy assumption, γ_{ij} must be spherically symmetric about each point p. It can be shown that this implies the existence of a local coordinate system $(x^i) = (r, \theta, \phi)$ such that p is at $r = 0$, such that (θ, ϕ) parameterize the orbits of $O(3)$ near p (each is $\cong S^2$ and (θ, ϕ) are spherical coordinates), and such that the line element takes the form

$$\gamma_{ij}dx^i dx^j = e^{2\beta(r)}dr^2 + r^2(d\theta^2 + \sin^2\theta \; d\phi^2) \tag{28}$$

for some function $\beta(r)$ which needs to be determined. The expression $r^2(d\theta^2 + \sin^2\theta d\phi^2)$ is of course nothing but the line element of the round 2-sphere of radius r. A calculation shows that the non-zero Ricci tensor components of this line element are:

$$\begin{aligned} \mathcal{R}_{11} &= \frac{2}{r}\beta'(r) \\ \mathcal{R}_{22} &= e^{-2\beta(r)}(r\beta'(r) - 1) + 1 \\ \mathcal{R}_{33} &= \sin^2\theta \left\{ e^{-2\beta(r)} \left(r\beta'(r) - 1 \right) + 1 \right\} \end{aligned}$$

whereas, by definition,

$$\begin{aligned}\gamma_{11} &= e^{2\beta(r)},\\ \gamma_{22} &= r^2,\\ \gamma_{33} &= r^2\sin^2\theta\ .\end{aligned}$$

So the 11-component of (28) gives:

$$\begin{aligned}\frac{2}{r}\beta' &= 2ke^{2\beta} \Longrightarrow\\ \frac{1}{2}e^{-2\beta} &= -\frac{1}{2}kr^2 + \frac{c}{2} \qquad (c = \text{ integration constant}) \quad \Longrightarrow\\ \beta &= -\frac{1}{2}\log\left(c - kr^2\right).\end{aligned}$$

The 22-component of (28) gives:

$$\begin{aligned}2kr^2 &= \underbrace{e^{-2\beta(r)}}_{(c-kr^2)}\underbrace{(r\beta'(r)-1)}_{\left(\frac{kr^2}{c-kr^2}-1\right)}+1\\ &= kr^2 - c + kr^2 + 1 \qquad (29)\end{aligned}$$

from which it follows that $c = 1$. Consequently, the metric on Σ is locally given by:

$$\gamma_{ij}dx^idx^j = \frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)$$

The final result can be interpreted as follows depending on the sign of k:

$\underline{k=0}$: In this case $\mathcal{R}_{ijkl} = 0$. This should just be flat three dimensional Euclidean space. We can make this manifest by writing

$$\gamma_{ij}dx^idx^j = dr^2 + r^2\left(d\theta^2 + \sin^2\theta\ d\phi^2\right) = dx^2 + dy^2 + dz^2$$

under the coordinate change (polar coordinates)

$$\begin{aligned}x &= r\sin\theta\cos\phi\\ y &= r\sin\theta\sin\phi\\ z &= r\cos\theta\end{aligned}$$

as one immediately verifies. The isometry group of flat Euclidean space is $E(3)$. The subgroup $O(3)$ is the isotropy group of any point, which makes manifest again that we can view this space as the quotient $E(3)/O(3)$.

$\underline{k > 0}$: To get insight into the nature of this metric, we set $\sin^2\chi = kr^2$. Then we find

$$\gamma_{ij}dx^i dx^j = \frac{1}{k}\left(d\chi^2 + \sin^2\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right)$$

which is the round metric on $\mathbb{S}^3$ with the radius $\rho = 1/\sqrt{k}$, where (χ, θ, ϕ) are Euler angles. The isometry group of the round 3-dimensional sphere is $O(4)$. The isotropy group of any point is $O(3)$, which gives the representation of the three-sphere as $O(4)/O(3)$.

$\underline{k < 0}$: This time, we set $\sinh^2\chi = -kr^2$. Then we find

$$\gamma_{ij}dx^i dx^j = \frac{1}{-k}\left(d\chi^2 + \sinh^2\chi\left(d\theta^2 + \sin^2\theta d\phi^2\right)\right).$$

This is seen to be the metric on a 3-dimensional hyperboloid $\mathbb{H}^3$ with "radius" $\rho = 1/\sqrt{-k}$. The terminology "hyperboloid" becomes clear if we present this space as the following subset of $\mathbb{R}^4$:

$$\mathbb{H}^3 = \left\{(u, x, y, z)| - u^2 + x^2 + y^2 + z^2 = \frac{1}{k}\right\},$$

equipped with the metric induced from 4-dimensional Minkowski spacetime. The correspondence with the previous parameterization is

$$\begin{aligned} x &= \rho\sinh\chi\sin\theta\cos\phi \\ y &= \rho\sinh\chi\sin\theta\sin\phi \\ z &= \rho\sinh\chi\cos\theta \\ u &= \rho\cosh\chi. \end{aligned}$$

The parameterization in terms u, x, y, z also makes manifest that the isometry group of the hyperboloid is $O(3,1)$. The isotropy subgroup of a point on the hyperboloid is $O(3)$, which gives the representation as the quotient $O(3,1)/O(3)$.

We note again that the metrics are only determined locally by the differential equations, so the general solution will be a quotient Σ/Γ of the spaces just found by suitable discrete subgroups Γ of the isometry groups. Examples are

$$\mathbb{T}^3 = \frac{\mathbb{R}^3}{\mathbb{Z}^3} \quad \text{or} \quad \frac{\mathbb{S}^3}{\mathbb{Z}_p} = \mathbb{L}(p,q) \tag{30}$$

The discrete subgroups are, in those cases, $\Gamma = \mathbb{Z}^3 \subset E(3)$ (flat case), $\Gamma = \mathbb{Z}_p \subset O(4)$ (positive curvature). The space $L(p,q)$ is called a Lens-space, and the natural number q which is co-prime to p is related to the precise definition of the action of $\mathbb{Z}_p$ on the 3-sphere: If we identify $\mathbb{R}^4$ with $\mathbb{C}^2$ and $\mathbb{S}^3$ with points $(z_1, z_2) \in \mathbb{C}^2$ having $|z_1|^2 + |z_2|^2 = 1$ then any $U(2)$ matrix can be mapped, via this identification, with an orthogonal matrix in $O(4)$ acting on the 3-sphere. If n is a natural number mod p, then its action is $(z_1, z_2) \mapsto (e^{in/p} z_1, e^{inq/p} z_2)$. The investigation of such quotients is an interesting mathematical problem. Physically, quotients do not seem to play a role in cosmology so far, so we will not discuss them further.

To summarize, our 4-dimensional spacetime manifold is

$$M = \mathbb{R} \times \begin{cases} \mathbb{R}^3 & k = 0, \text{ "flat universe"} \\ \mathbb{S}^3 & k > 0, \text{ "closed universe"} \\ \mathbb{H}^3 & k < 0, \text{ "open universe"} \end{cases}$$

and our 4-dimensional spacetime metric is

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2 \gamma_{ij}(x) dx^i dx^j \\ &= -dt^2 + a(t)^2 \left\{ \frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2\theta \ d\phi^2) \right\}. \end{aligned}$$

The spacetime metric depends, up to the parameter[7] $k \in \mathbb{R}$, only on the unknown function $a(t)$. This must be determined using Einstein's equation. For this, we begin by writing down the Ricci tensor components in the coordinates $(x^\mu) = (t, r, \theta, \phi)$. A calculation (exercises) shows:

$$\begin{aligned} R_{00} &= -3\frac{\ddot{a}}{a} \\ R_{11} &= \frac{a\ddot{a} + 2\dot{a}^2 + 2k}{1 - kr^2} \\ R_{22} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2k) \\ R_{33} &= r^2 \sin^2\theta (a\ddot{a} + 2\dot{a}^2 + 2k) \end{aligned}$$

To set up Einstein's equation, we must specify a stress tensor satisfying $\nabla^a T_{ab} = 0$ which is compatible with the spacetime symmetries, i.e. for all

[7] To be precise, what matters is only the signum of k because we can always normalize k to $-1, 0$, or $+1$ by a change of coordinates $r \mapsto |k| r$ and a subsequent change of $a(t)$.

$\psi \in G$,

$$\psi^\star T_{ab} = T_{ab}.$$

It turns out that the most general such T_{ab} has the fluid form

$$T_{ab} = \rho u_a u_b + P(u_a u_b + g_{ab}) \ ,$$

where ρ, P are functions only of t and where $u_a = (dt)_a$ is the tangent field to the fluid lines. Equivalently, $(u^\mu) = (1,0,0,0)$ in our "comoving" (with the fluid) coordinates $(x^\mu) = (t, r, \theta, \phi)$, and

$$(T_\mu{}^\nu) = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & P & 0 & 0 \\ 0 & 0 & P & 0 \\ 0 & 0 & 0 & P \end{pmatrix} .$$

We complete our description of matter with an equation of state such as $P = w\rho$, where $w \in \mathbb{R}$ is constant. One usually requires that $|w| \leq 1$ (dominant energy condition) which includes the case of a "cosmological constant", $T_{ab} = -\rho g_{ab}$ with $\rho = Lambda > 0$ the cosmological constant.

We can now solve Einstein's equation, which we will use in the trace reversed form $R_{ab} = 8\pi G_N \left(T_{ab} - {}^1\!/\!{}_2 g_{ab} T\right)$, where $T = g_{ab}T^{ab} = -\rho + 3P$ in our case. However, before investigating this equation directly, it is useful to look at the consequences of the equation $\nabla_a T^{ab} = 0$ (Bianchi-identity). This gives

$$0 = -\dot{\rho} - \frac{3\dot{a}}{a}(\underbrace{\rho + P}_{(1+w)\rho})$$

which for $\rho > 0$ can be solved by

$$\begin{aligned} \frac{\dot{\rho}}{\rho} &= -3(1+w)\frac{\dot{a}}{a} \\ \Leftrightarrow \qquad \frac{d}{dt}\log\rho &= \frac{d}{dt}(\log a^{-3(1+w)}) \end{aligned} \tag{31}$$

Hence

$$\rho \propto a^{-3(1+w)} \ . \tag{32}$$

For some conventional forms of matter this gives

w	matter	
0	pressureless dust	$\rho \propto a^{-3}$
$\frac{1}{3}$	photons	$\rho \propto a^{-4}$
-1	cosmological constant	$\rho \propto a^{0}$

We now look at Einstein's equation directly. The 00-component gives

$$-3\frac{\ddot{a}}{a} = 4\pi G_N(\rho + 3P)$$

whereas any of the ii-components gives

$$\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + \frac{2k}{a^2} = 4\pi G_N(\rho - P).$$

One can eliminate $\frac{\ddot{a}}{a}$ to get

$$\boxed{\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G_N}{3}\rho - \frac{k}{a^2}} \qquad \text{(first Friedmann equation)}$$

and

$$\boxed{\frac{\ddot{a}}{a} = -\frac{4\pi G_N}{3}(\rho + 3P)} \qquad \text{(second Friedmann equation)}$$

These two equations, first derived by Friedmann and independently by Lemaitre, encode the full dynamics of homogeneous, isotropic spacetimes. We point out the following very important

Qualitative features:

$\underline{k < 0}$: Then, $\dot{a}$ cannot become zero, which leads to a universe that is forever expanding.

$\underline{k = 0}$: For any matter having $P \geq 0$, ρ must decrease as a increases at least as rapidly as $\rho \approx a^{-3}$ (from the t-component of $\nabla_a T^{ab} = 0$). This causes ρa^2 to become zero. Thus, if $k = 0$, the expansion velocity $\dot{a}$ approaches zero as $t \to \infty$. (By contrast, for $k < 0$, $\dot{a} \to \sqrt{|k|}$ as $t \to \infty$.)

$\underline{k > 0}$: Then, the universe cannot expand forever, because the first term ρa^2 decreases faster than the k-term. There exists a critical $a = a_c$ so that $a \leq a_c$ for all time. Also, a cannot asymptotically approach a_c, as $-\ddot{a}$ is bounded below, which means that the universe 'bounces'.

We also note that, for any sign of k, the terms in the first Friedmann equation cannot balance each other for generic values of w. Thus, in view of (32), for any $w \geq -1$, $\dot{a}$ must diverge at early times, i.e. we are lead to the inevitable conclusion that under generic conditions, the universe must have started with a singular state having $\dot{a} = \infty$, i.e. a "big bang". This conclusion is one of the most important predictions of general relativity.

A table of some solutions is shown below. In this table, η is the "conformal" time coordinate, related to the time coordinate t by $a(\eta)d\eta = dt$. Furthermore, $C = \frac{1}{3}8\pi G_N \rho a^3$ and $C' = \frac{1}{3}8\pi G_N \rho a^4$ are constants

Geometry	Dust ($P = 0$)	Radiation ($P = \frac{\rho}{3}$)
$k = +1 \quad (\mathbb{S}^3)$	$\begin{aligned} a &= \frac{1}{2}C(1-\cos\eta) \\ t &= \frac{1}{2}C(\eta - \sin\eta) \end{aligned}$	$a = \sqrt{C'}\sqrt{1-\left(1-\frac{t}{\sqrt{C'}}\right)^2}$
$k = 0 \quad (\mathbb{R}^3)$	$a = \sqrt[3]{\frac{9C}{4}}t^{\frac{2}{3}}$	$a = (4C')^{\frac{1}{4}}t^{\frac{1}{2}}$
$k = -1 \quad (\mathbb{H}^3)$	$\begin{aligned} a &= \frac{1}{2}C(\cosh\eta - 1) \\ t &= \frac{1}{2}C(\sinh\eta - \eta) \end{aligned}$	$a = \sqrt{C'}\sqrt{\left(1-\frac{t}{\sqrt{C'}}\right)^2 - 1}$

It is also useful to record that, for a flat universe ($k = 0$) and a general equation of state $P = w\rho$, with $w > -1$, we get (exercise)

$$a(t) \propto t^{\frac{2}{3(1+w)}} .$$

The limiting case $w = -1$ is also interesting and corresponds to a cosmological constant. In that case, an exponential expansion is found (exercise)

$$a(t) \propto e^{tH} \ , \tag{33}$$

where the constant $H = \dot{a}/a$ is called "Hubble constant".

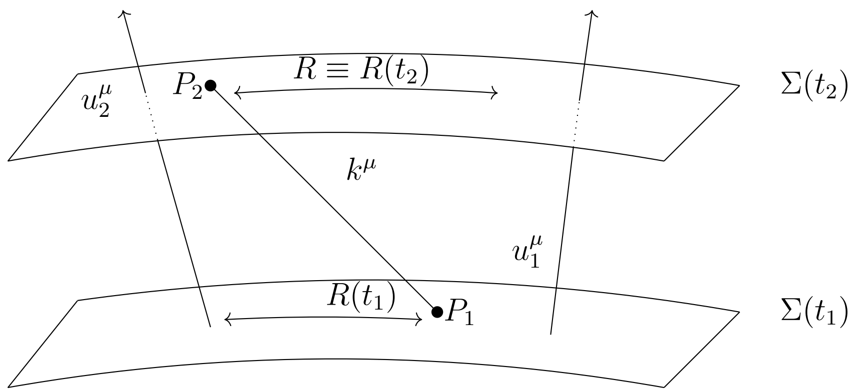
One may object that the homogeneous and isotropic spacetimes we have discussed could just be a mathematical exercise without much physical significance because our symmetry assumptions are clearly not satisfied in reality. Indeed, on small scales (e.g. molecular, solar system, or galactic), the stress energy is very far from being distributed in a homogeous and isotropic manner. Consequently, on such scales also spacetime metric is far from being homogeneous and isotropic (although less so, because the metric is "two derivatives down" from the stress tensor according to Poisson's/Einstein's equation). The hypothesis of homogeneity and isotropy seems, however, a very good approximation on cosmological scales, and the geometries we have described therefore lead to a good description of our universe in the very large. Current measurements indicate that $k = 0$ (or very nearly so) and that the scale factor appears to have been of the exponential type in the early cosmos (big bang, inflation), followed by a power law epoch (radiation dominated, then dust), and then followed again by an exponential type epoch. This behavior of the scale factor is incompatible with Einstein's equations if only conventional, baryonic, matter is assumed to be present. In particular, to get an exponential type expansion, we would need matter that is at least very similar to a "cosmological constant", as we have seen. The origin of such a hypothetical kind of matter component is still under discussion. At any rate, the prediction of a dynamical cosmos (for any type of matter) is one of the, if not the, most remarkable predictions of General Relativity.

12.1 Redshift in cosmological spacetimes

A simple, but very important effect in an expanding universe is the phenomenon of redshift. Let us consider a light ray, following an affinely parameterized null geodesic with 'wave vector' k^a. The frequency ω measured by a comoving (with the "Hubble flow" described by u^a) observer following the curve γ with velocity $u^a = \dot{\gamma}^a = (\partial/\partial t)^a$ is

$$\omega = -k_a u^a.$$

We now wish to compute the change in frequency when two observers measure the same light ray.



This calculation is aided by the presence of Killing vector fields in our examples. If ξ^a is a spacelike Killing vector field, then one can show that $k_a \xi^a = f$ is constant along the light ray (exercises). Assuming that k^a has a projection into Σ tangent to the Killing field ξ^a, we find (using the shorthand $|\xi|^2 = \xi^a \xi_a$)

$$\begin{aligned}
\frac{a(t_1)^2}{a(t_2)^2} &= \frac{|\xi|^2_{p_1}}{|\xi|^2_{p_2}} \\
k_a u_1^a &= -\left.\frac{k_a \xi^a}{|\xi|}\right|_{p_1} \\
k_a u_2^a &= -\left.\frac{k_a \xi^a}{|\xi|}\right|_{p_2} \\
\Rightarrow \quad \frac{\omega_1}{\omega_2} &= \frac{a(t_2)}{a(t_1)}.
\end{aligned}$$

The redshift factor is

$$\begin{aligned}
z &= \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{\omega_1}{\omega_2} - 1 = \frac{a_2}{a_1} - 1 \\
&\approx HR
\end{aligned} \tag{34}$$

where R is the distance which is $\approx t_2 - t_1$ for nearby galaxies. For a flat universe ($k = 0$) this is

$$R = a(t)\sqrt{x^2 + y^2 + z^2} = \text{ geodesic distance.}$$

Another way to state (34) is

$$\frac{a_1}{a_2} = \frac{a(\text{emitted})}{a(\text{observed})} = \frac{1}{1+z}.$$

The 'physical'= geodesic distance R on a given slice between two points is of little empirical interest, since we can only observe objects in our past lightcone. A more useful notion of distance is the "luminosity distance", d_L, defined as

$$d_L^2 = \frac{L}{4\pi F}. \tag{35}$$

Here L is the absolute luminosity and F is the flux seen by the observer. For Minkowski space, $d_L = R$ since in that case the area of a sphere of the physical radius R is $4\pi R^2 = A$, that is $\frac{F}{L} = \frac{1}{A}$ in a flat space ($a(t) \equiv 1, k = 0$). On the other hand, in an expanding ($k = 0$) Friedmann-Lemaître-Robertson-Walker universe, we have instead

$$\frac{F}{L} = \frac{1}{(1+z)^2 A}$$

since the photons arriving at t_2 are $(1+z)^{-1}$ times less energetic ($E_{\text{ph}} = h\nu = \hbar\omega$) and the rate of emission also goes down by the same factor. Thus,

$$d_L = (1+z)R.$$

Because $(k^\mu) = (1, \dot{R}, 0, 0)$ is null we must have $1 = a^2\dot{R}^2$ and hence $\dot{R} = \frac{1}{a}$. Then,

$$\begin{aligned}
R = R(t_2) &= a(t)\sqrt{x^2 + y^2 + z^2} \\
&= \int_{t_1}^{t_2} \dot{R}(t)dt \\
&= \int_{t_1}^{t_2} \frac{dt}{a(t)} \\
&= \int_{a_1}^{a_2} \frac{da}{a^2 H(a)},
\end{aligned}$$

or, with $dz = \frac{-da}{a^2}$

$$
\begin{aligned}
R &= \int_{z_2}^{z_1} \frac{dz}{H(z)} \\
&= \int_0^{z} \frac{dz'}{H(z')} \\
\Rightarrow \quad d_L &= (1+z)\int_0^{z} \frac{dz'}{H(z')}.
\end{aligned}
$$

It is common to use the first Friedmann equation – or rather its obvious generalization to several species of "particles" – in order to replace H^2 by

$$
H^2 = \frac{8\pi G_N}{3} \sum_{\text{species } i} \rho_i .
$$

Assuming that each species evolves according to a power law, we find

$$
\rho_i = \rho_{i,\text{today}} a^{-n_i}
$$

where $\rho_{i,\text{today}}$ is the matter density at t_2 (today). Assuming without loss of generality that $a_2 = 1$ we find

$$
\begin{aligned}
H(z) &= \sqrt{\frac{8\pi G_N}{3}} \left(\sum_i \rho_{i,\text{today}} (1+z)^{n_i} \right)^{\frac{1}{2}} \\
\Rightarrow \quad \frac{H(z)}{H_{\text{today}}} &= \left(\sum_i \Omega_{i,\text{today}} (1+z)^{n_i} \right)^{\frac{1}{2}}
\end{aligned}
$$

where $\Omega_i(t) \equiv \frac{8\pi G_N}{3} \frac{\rho_i(t)}{H(t)^2}$. Then we get

$$
\boxed{d_L(z) = \frac{1+z}{H_{\text{today}}} \int_0^{z} \frac{dz'}{\left(\sum_i \Omega_{i,\text{today}} (1+z')^{n_i}\right)^{\frac{1}{2}}}}
$$

In practice, we measure $d_L(z)$ for large redshifts z and extract $\Omega_{i,\text{today}}$ and H_{today}. For that, we need objects with known intrinsic luminosity L, such as type IIa supernovae.

12.2 Particle Horizons

Another important feature of expanding universes is the possible existence of "particle horizons". This is most easily demonstrated for flat universes, $k = 0$, where the metric is

$$g = -dt^2 + a(t)^2(dx^2 + dy^2 + dz^2). \tag{36}$$

Let us introduce again the *conformal time parameter* η by

$$\eta = \int_{t_0}^{t} \frac{dt'}{a(t')} \qquad \left(\frac{d\eta}{dt} = \frac{1}{a(t)} \Leftrightarrow a d\eta = dt\right) . \tag{37}$$

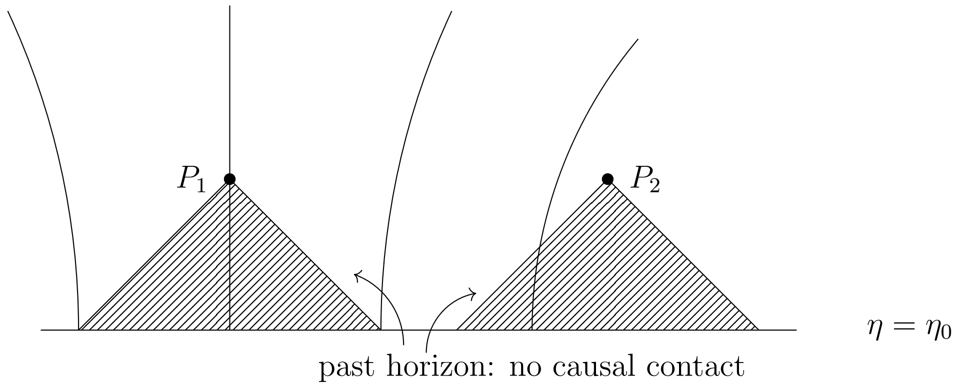
from which it follows that

$$ds^2 = a(\eta)^2 \Big\{ \underbrace{-d\eta^2 + dx^2 + dy^2 + dz^2}_{\text{Minkowski-space}} \Big\} \tag{38}$$

We see that writing the metric in terms of η has the advantage that its relationship with Minkowski space becomes manifest: It is "conformal" to Minkowski spacetime, or possible a subset thereof.

Due to the conformal factor $a(\eta)^2$, geodesics in Minkowski space in general do not in general coincide with the geodesics of g_{ab}. However, the conformal factor preserves the causal character of a curve, and therefore the causal relationships in spacetime. In particular, we may ask whether it is possible for two points p, p' in spacetime to be such that their causal pasts (i.e. the set of all points that can be reached by past directed timelike or null curves) are disjoint. It is clear form the following picture that this will be the case if and only if the parameter η has a finite range for negative values, which in turn will be the case if and only if $\int_t^{t_1} \frac{dt'}{a(t')}$ converges to a finite value for $t \to t_0$.

Whether or not this is the case therefore depends on the behavior of $a(t)$ near $t = t_0$, which is in turn determined by the equation of state. Indeed, recall that for $P = w\rho$, we had $a(t) \propto t^{\frac{2}{3(w+1)}}$ (choosing $t_0 = 0$) so the integral is finite in particular for all $w \geq 0$ (e.g. dust, $a \propto t^{2/3}$, or radiation, $a \propto t^{1/2}$.) If there are points p, p' with disjoint causal past, then a particle at p could never have been in contact with a particle at p' – one says that "there are particle horizons". Thus, we get regions in spacetime which are causally disjoint, and for this reason, will not have had the opportunity to equilibrate with each other. Current observations seem to exclude the



presence of such horizons, meaning for instance that the scale factor $a(t)$ could not have behaved like that of radiation or dust all the way to the big bang ($t = 0$). On the other hand, particle horizons are not present e.g. for an exponential scale factor $a(t) \propto e^{tH}$ because the integral $\int_t^{t_1} dt'/a(t')$ then clearly diverges at the lower end $t \to -\infty$. Therefore such an "inflationary phase" is consistent with the absence of horizons. It is indeed currently believed that our universe underwent such a phase shortly after the big bang.

13 Black Holes

In the previous subsection, we have obtained solutions to Einstein's equation with a non-zero stress tensor representing varous types of fluid matter. However, in General Relativity, one can have interesting non-trivial solutions even for a vanishing stress tensor. Such solutions are called "vacuum solutions" and obey

$$R_{ab} = 0 \ . \tag{39}$$

In particular, unlike in the case of Newtonian Gravity [8], one can obtain *static* vacuum solutions. It turns out that these solutions describe objects that have properties unlike any other objects known before: black holes.

[8] In Newtonian Gravity, the only globally defined solution to Poisson's equation with vanishing ρ which is decaying at spatial infinity is the trivial one. The possibility of non-trivial static vacuum solutions in General Relativity can be ascribed to the fact that Einstein equations are non-linear, meaning that gravity can act as its own source in a sense.

13.1 Derivation of the Schwarzschild Solution

Given the complexity of Einstein's equations, it is somewhat surprising that this family of static solutions, known as the "Schwarzschild solution", is actually rather easy to derive. To get started one assumes, as seems evidently reasonable, that

1. g_{ab} has as its isometry group the group of rotations $SO(3)$ and, of course, that

2. g_{ab} is static. This means that there are 'time-shift' isometries whose orbits are orthogonal to a spacelike surface Σ.

[The first assumption can actually be shown to be a consequence of the second one – this is known as "Israels theorem".] One can show that time shifts and rotations commute, and that there exists a local coordinate system $(x^\mu) = (t, r, \theta, \phi)$ such that

$$g = -\underbrace{f(r)\,dt^2 + h(r)}_{\text{free functions}}\,dr^2 + \underbrace{r^2\left(d\theta^2 + \sin^2\theta\ d\phi^2\right)}_{\text{metric on } \mathbb{S}^2_r}$$

with a time shift acting by $t \to t+$const and rotations acting on the spherical polar coordinates ϕ and θ in the usual way.

It is not hard, although tedious, to compute the Ricci tensor of this metric in the coordinates $(x^\mu) = (t, r, \theta, \phi)$. The off-diagonal components vanish identically, whereas the diagonal components give the equations

$$0 = \frac{1}{2}\frac{1}{\sqrt{fh}}\left[\frac{f'}{\sqrt{fh}}\right]' + (rfh)^{-1}f' \tag{40}$$

$$0 = -\frac{1}{2}\frac{1}{\sqrt{fh}}\left[\frac{f'}{\sqrt{fh}}\right]' + (rh^2)^{-1}h' \tag{41}$$

$$0 = -\frac{f'}{2fh} + (2h^2)^{-1}h' + \frac{1-\frac{1}{h}}{r} \tag{42}$$

Equations (40) and (41) give

$$\frac{f'}{f} + \frac{h'}{h} = 0 \quad \Leftrightarrow \quad f = \frac{k}{h} \qquad (k > 0 \text{ a constant,})$$

and by rescaling $t \to \sqrt{k}t$ we may set $k = 1$. Then (42) gives

$$\begin{aligned} -f' + \frac{1-f}{r} &= 0 \\ \Leftrightarrow \quad (rf)' &= 1 \\ \Leftrightarrow \quad f(r) &= 1 + \frac{C}{r}, \end{aligned}$$

where C is some constant. The desired vacuum solution is thus:

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

Before embarking on a more detailed analysis of this metric, we make the following crude observations:

1. As $r \to \infty$

$$ds^2 \to -dt^2 + \underbrace{dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)}_{dx^2+dy^2+dz^2} \quad = \text{Minkowski space}$$

   So the metric is "asymptotically flat".

2. From our discussion surrounding the derivation of Einstein's equations, we have

$$g_{00} \cong -1 - 2\Phi \tag{43}$$

   where Φ is the Newtonian potential. This leads us to identify the constant C as (restoring the speed of light c temporarily in our formulas)

$$-\frac{C}{2} \cong \frac{G_N M}{c^2}, \tag{44}$$

   implying in particular that we should take $C \leq 0$. In Newtonian gravity, the radius $r_S = \frac{2G_N M}{c^2}$ is precisely the surface of a spherical object of mass M such that the escape velocity for a particle is equal to the speed of light (Laplace). This crudely suggests that the metric might have something to do with a "black hole", as we will confirm below. With the notation $r_S =$ "Schwarzschild radius,"

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

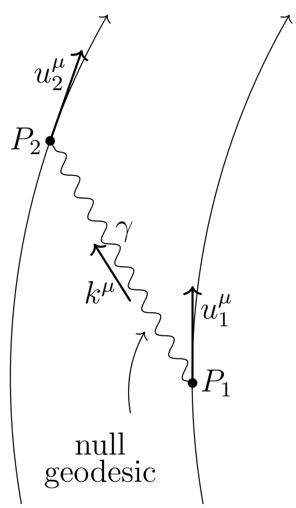
One gets a better intuition about the size of the constants if one writes for instance

$$\begin{aligned} r_S &= \frac{2G_N M}{c^2} \approx 3\left(\frac{M}{M_\odot}\right)\text{km} \qquad \text{with} \\ M_\odot &= \text{mass of the sun} = 2\times 10^{33}\text{g} \ , \end{aligned}$$

i.e. the sun is vastly bigger than its Schwarzschild radius.

13.2 The redshift effect

Consider two 'static' observers, each following a curve of constant r, θ, ϕ, exchanging a light signal. The tangents are denoted u_1^a, u_2^a, respectively.



The locally measured frequencies are (compare the corresponding discussion in cosmological spacetimes) at two points p_1, p_2 are:

$$\begin{aligned} \omega_1 &= -k_a u_1^a\big|_{p_1} \\ \omega_2 &= -k_a u_2^a\big|_{p_2} \end{aligned}$$

We have $u_1^a u_{1a} = -1 = u_2^a u_{2a}$, and the static observers are tangent to the Killing vector field $\xi^a = \left(\frac{\partial}{\partial t}\right)^a$. Note that $\xi_a k_a$ does not change along the null geodesic representing the signal, since ξ^a is Killing, and since k^a is geodesic.

We may write

$$\begin{aligned} u_1^a &= \left.\frac{\xi^a}{|\xi|}\right|_{p_1} \\ u_2^a &= \left.\frac{\xi^a}{|\xi|}\right|_{p_2} \end{aligned}$$

where (setting $G_N = 1 = c$ in the follwing)

$$\begin{aligned} |\xi|^2 &= -g_{ab}\xi^a\xi^b \\ &= 1 - \frac{2M}{r}. \end{aligned}$$

So we find

$$\begin{aligned} \frac{\omega_1}{\omega_2} &= \frac{k_a u_1^a}{k_a u_2^a} = \frac{\left.{}^{(k_a\xi^a)}/_{|\xi|}\right|_{p_1}}{\left.{}^{(k_a\xi^a)}/_{|\xi|}\right|_{p_2}} \\ &= \sqrt{\frac{1-\frac{2M}{r_2}}{1-\frac{2M}{r_1}}}\ . \end{aligned}$$

If the emitter (1) is closer to the 'center' than the receiver (2), $r_1 < r_2$, the frequency will decrease ($\omega_2 < \omega_1$) and hence, by $E = \hbar\omega$, the energy of a photon climbing out the gravitational well is decreased. If $\frac{G_N M}{c^2} \ll r_1, r_2$ then

$$\begin{aligned} \frac{\Delta\omega}{\omega} &\approx -\frac{G_N M}{c^2 r_1} + \frac{G_N M}{c^2 r_2} \\ \text{or} \quad \Delta E &= \hbar\Delta\omega \approx \frac{\hbar\omega}{c^2}\left[-\frac{G_N M}{r_1} + \frac{G_N M}{r_2}\right]. \end{aligned}$$

Since a photon with reduced energy is redder according to Einstein's formula $E = h\nu$, the effect is called "gravitational redshift".

13.3 Geodesics

To determine the orbits of material particles and light rays in the Schwarzschild geometry, we need to study geodesics. We denote the geodesic by $\gamma : \mathbb{R} \to M, s \mapsto \gamma(s)$ and by $\dot\gamma^a$ the tangent vector to the geodesic. In coordinates $(x^\mu) = (t, r, \theta, \phi)$

$$\frac{d\gamma^\mu}{ds} = \dot\gamma^\mu = (\dot t, \dot r, \dot\theta, \dot\phi)\ .$$

We assume first that the geodesic is timelike, and choose s to be proper time. It follows that there holds

$$g_{ab}\dot{\gamma}^a\dot{\gamma}^b = -1$$

along the curve. (Proof: act with $\dot{\gamma}^c\nabla_c \cong \frac{\partial}{\partial s}$ on $g_{ab}\dot{\gamma}^a\dot{\gamma}^b$ and use $\dot{\gamma}^a\nabla_a\dot{\gamma}^b = 0$.) Additionally, the quantities

$$\left.\begin{array}{rl} E = & -\dot{\gamma}_a \left(\frac{\partial}{\partial t}\right)^a \\ L = & \quad\dot{\gamma}_a \left(\frac{\partial}{\partial \phi}\right)^a \end{array}\right\} \tag{45}$$

are also constant along the curve, i.e. independent of s, because $(\partial/\partial t)^a$ and $(\partial/\partial\phi)^a$ are Killing fields. (This follows e.g. from eq.??, because the line element is independent of t, ϕ.) Furthermore, since $\theta \mapsto \pi - \theta$ is an isometry of Schwarzschild, it is consistent to assume that $\theta(s) = \frac{\pi}{2}$, so $\dot{\theta} = 0$ along the geodesic (exercise). Without loss of generality, we may choose γ to be in such an equatorial plane. E has the interpretation of the energy of the particle, and L that of the angular momentum in the equatorial plane. For null geodesics, we have the same constants of motion, but the normalization condition is now $g_{ab}\dot{\gamma}^a\dot{\gamma}^b = 0$.

Substitution of the constants of motion into the normalisation condition yields, for timelike geodesics,

$$\begin{aligned} \frac{1}{2}\dot{r}^2 + V(r) &= \frac{1}{2}E^2 \\ \text{where} \qquad V(r) &= \frac{1}{2} - \underbrace{\frac{M}{r}}_{\substack{\text{Newtonian}\\ \text{term}}} + \underbrace{\frac{L^2}{2r^2}}_{\substack{\text{Angular}\\ \text{momentum}\\ \text{barrier}\\ \text{(centrifugal}\\ \text{force)}}} - \underbrace{\frac{ML^2}{r^3}}_{\text{new}}. \end{aligned}$$

$V(r)$ can be viewed as the 'effective potential' seen by the geodesic. In the null case, it is given instead by

$$V(r) = \frac{L^2}{2r^2} - \frac{ML^2}{r^3}$$

The radial motion is thus the same as that of a particle in a potential ($V(r)$ in either case, although the form of the potential is different in the null case. Once the radial motion has been determined, the angular motion is found in either case by solving

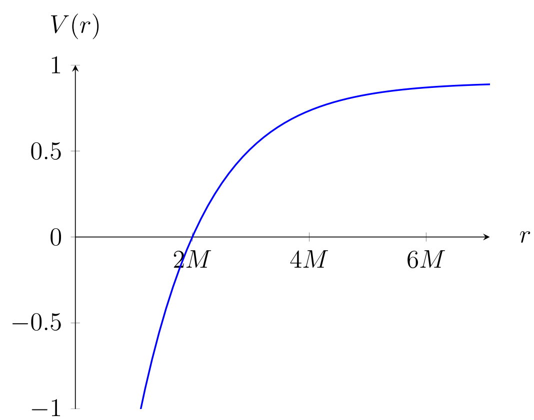
$$L = r^2\dot{\phi} \quad \text{and} \quad E = \left(1 - \frac{2M}{r}\right)\dot{t}.$$

Looking at the potential $V(r)$ in the timelike case, we have the familiar terms from the Kepler problem corresponding to gravitational attraction and centrifugal force. In addition to the familiar terms, we have the new term $-\frac{ML^2}{r^3}$, which is of the same sign as the Newtonian term, but wins at small distances r (and is insignificant at large distances). It is thus plausible that the behavior of timelike geodesics will differ from the familiar motion of particles in a central $1/r$-potential for small r.

We are particularly interested in (quasi-) periodic orbits, corresponding to radial oscillations around the minima of the effective potential $V(r)$. To find the extrema of $V(r)$ in the timelike case, we compute

$$\begin{aligned} 0 &= V'(r) = \frac{Mr^2 - L^2 r + 3ML^2}{r^4} \\ \Rightarrow \qquad R_\pm &= \frac{L^2}{2M} \pm \left(\left(\frac{L^2}{2M}\right)^2 - 3L^2\right)^{\frac{1}{2}}. \end{aligned}$$

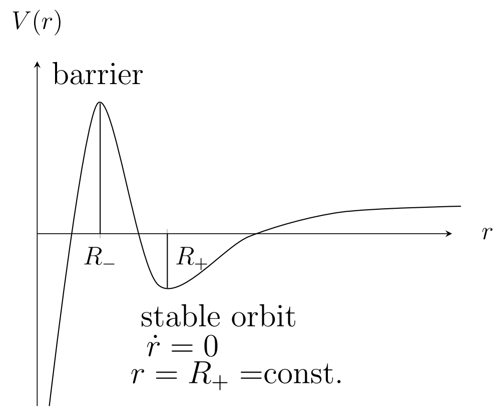
So if $L^2 < 12M^2$, then there are no extrema and the potential $V(r)$ looks like



Hence, there are no stable bound orbits, and a particle having $\dot{r} < 0$ initially will fall right into the "singularity[9]" $r = r_S$", and further into $r = 0$.

[9]We will later clarify the true nature of this, only apparent, "singularity".

On the other hand, if $L^2 \geq 12M^2$, it is easy to check that there is a minimum R_+ and a maximum R_-, and the potential $V(r)$ is as in the following figure.



We conclude that

$$\begin{aligned} R_+ &= r \quad \text{corresponds to a stable circular orbit,} \\ R_- &= r \quad \text{corresponds to an unstable circular orbit.} \end{aligned}$$

For $L \gg M$, the stable orbit is approximately at $R_+ \approx \frac{L^2}{M}$, which gives the Newtonian formula for the orbit of a mass with angular momentum L orbiting a central point mass M. (This is another way of seeing that the identification $C \to -2M$ is physically correct.) Note that the minimum value of R_+ such that stable circular orbit exist is attained when $R_+ = R_- \Rightarrow L^2 = 12M^2 \Rightarrow R_+ \geq 6M$. That shows that no stable circular orbits exist for sufficiently small r-values in General Relativity, because the new term in $V(r)$ wins for $r < 6M$. The energy E of a particle in the stable circular orbit $r = R_+$ is

$$\begin{aligned} \frac{1}{2}E^2 &= V(R_+) = \frac{1}{2} - \frac{M}{R_+} + \frac{L^2}{2R_+^2} - \frac{ML^2}{R_+^3} \\ &= \frac{1}{2}\frac{(R_+ - 2M)^2}{R_+(R_+ - 3M)} \\ V'(R_+) &= 0, \\ \text{so} \qquad E &= \frac{R_+ - 2M}{R_+^{\frac{1}{2}}(R_+ - 3M)^{\frac{1}{2}}} \longrightarrow 1 \text{ as } R_+ \to \infty \end{aligned}$$

Therefore, a particle in an unstable circular orbit in the range $3M < R < 4M$ having an energy which is bigger than that of $E(\infty)$ escapes to infinity. The binding energy $E_B = E(\infty) - E = 1 - E$ for the smallest stable circular orbit ($R_+ = 6M$) is given by

$$E_B = 1 - \sqrt{\frac{8}{9}} \approx 0.06 = 6\%.$$

Due to gravitational radiation (covered later), a body starting in a circular stable orbit will loose some of its energy and therefore gradually decrease r down to $r = R_{\min} = 6M$. The total energy lost (and hence emitted by gravitational radiation) is thus at most about 6% of its total energy. For a body rotating around an ultra-spinning Kerr black hole (a rotating analogoue of the Schwarzschild solution which is beyond the scope of these notes), the ratio is even as big as $\approx 40\%$. Thereby, a substantial portion of energy can be converted into gravitational radiation.

In order to find the oscillations of r around the minimum of $V(r)$, we carry out a Taylor exansion around $r = R_+$,

$$V(r) \approx V(R_+) + \frac{1}{2}\omega_r^2 (r - R_+)^2 + \mathcal{O}\left[(r - R_+)^3\right]$$

The "oscillation frequency" in the radial direction is given by

$$\omega_r^2 = V''(R_+) = \frac{M(R_+ - 6M)}{R_+^3 (R_+ - 3M)}.$$

On the other hand, the angular frequency of the geodesic is $\omega_\phi = \dot{\phi}$ and therefore

$$\omega_\phi^2 = \frac{L^2}{R_+^4} = \frac{M}{R_+^2 (R_+ - 3M)}.$$

Hence, for $R_+ \gg M$ (Newtonian limit), $\omega_\phi \approx \omega_r$ and the particle returns to the original r value after each orbit. In full General Relativity (without taking the Newtonian limit), there is instead a precession of the perhelion with frequency

$$\omega_p = \omega_\phi - \omega_r = -\left(\sqrt{1 - \frac{6M}{R_+}} - 1\right)\omega_\phi.$$

For $R_+ \gg M$ we get to the lowest non-vanishing order

$$\omega_p \approx \frac{3M\omega_\phi}{R_+} \approx \frac{3M}{R_+}\left(\frac{M}{R_+^3}\right)^{\frac{1}{2}} = \frac{3M^{\frac{3}{2}}}{R_+^{\frac{5}{2}}},$$

and restoring c, G_N:

$$\omega_p \approx \frac{3(G_N M)^{\frac{3}{2}}}{c^2 R_+^{\frac{5}{2}}}.$$

For Mercury, this gives (taking also the eccentricity of the orbit into account) $\omega_p(☿) = \frac{43''}{100\mathrm{a}}$. It is surprising – and a lucky coincidence – that this minute precession had been observed around the time General Relativity was conceived. Newtonian calculations, including even the perturbations by the other planets, which could be carried out thanks to the enormous advances in Celestial Mechanics in the end of the 19th century, could not account for this effect, whereas General Relativity could. Hence, the prediction of the perihelion precession of Mercury was historically the first test of General Relativity.

A similar analysis can be carried out for null-geodesics "skimming the surface at the Schwarzschild radius", and leads to the prediction of light bending – another early test of General Relativity (exercises).

13.4 Kruskal extension

In order to gain insight into the nature of the apparent singularity of the Schwarzschild metric at $r = r_S$, we next consider radially outgoing null-geodesics, which exist due to the symmetries of the line element. As before, our geodesics are denoted by $(\gamma^\mu(s)) = (t(s), r(s), \theta(s), \phi(s))$, and for radial geodesics $(\dot\gamma^\mu) = (\dot t, \dot r, 0, 0)$. The null-condition leads to

$$\begin{aligned} 0 &= g_{ab}\dot\gamma^a\dot\gamma^b \\ &= -\left(1 - \frac{2M}{r}\right)\dot t^2 + \frac{\dot r^2}{1 - \frac{2M}{r}} \end{aligned}$$

or

$$\frac{dt}{dr} = \pm\left(\frac{r}{r - 2M}\right)$$

or

$$\begin{aligned} t &= \pm r_* + \text{const.} \\ \text{with} \qquad r_* &= r + 2M \log\left(\frac{r}{2M} - 1\right) \Rightarrow dr_* = \frac{dr}{1 - \frac{2M}{r}}. \end{aligned}$$

It seems a good idea to try to rewrite the line element in terms of coordinates obtained from the affine parameters along radial null geodesics. Since we have just seen that the latter are defined by the conditions $\pm r_* + t =$ constant and $\phi, \theta =$ constant, the appropriate coordinates are $u = t - r_*, v = t + r_*$ (and θ, ϕ), in terms of which the line element becomes

$$ds^2 = -\left(1 - \frac{2M}{r}\right) du dv + r^2(d\theta^2 + \sin^2\theta\ d\phi^2)\ .$$

Here r is now viewed as a function of u and v; explicitly

$$r + 2M \log\left(\frac{r}{2M} - 1\right) = r_* = \frac{v - u}{2}.$$

We now make further transformations

$$\begin{aligned} U &= -\exp\left(\frac{-u}{4M}\right) \equiv T - X\ , \\ V &= \exp\left(\frac{v}{4M}\right) \equiv T + X\ . \end{aligned}$$

The relationship between the original coordinates (t, r) and the new ones (T, X) is summarized in the following equations

$$\left(\frac{r}{2M} - 1\right) \exp\left(\frac{r}{2M}\right) = X^2 - T^2\ , \tag{46}$$

$$\frac{t}{2M} = \log\left(\frac{T + X}{X - T}\right) = 2\,\text{arctanh}\,\frac{T}{X}\ . \tag{47}$$

Changing to the coordinates (T, X) leads to the "Kruskal form" of the line element

$$ds^2 = \frac{32M^3}{r} \exp\left(-\frac{r}{2M}\right)(-dT^2 + dX^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{48}$$

The Kruskal form of the line element shows that $r = r_S$ is not a singularity, because we can clearly extend the metric analytically across this value, at least until $r = 0$. Thus, the true geometry is the analytically extended manifold labelled by the coordinates X, T, and the coordinates on $\mathbb{S}^2$, consistent with $r > 0$. We see:

1. The allowed range of T and X consistent with $r > 0$ is $X^2 - T^2 > -1$. The value $X^2 - T^2 = -1$ corresponds to $r = 0$. This is seen to be a true singularity, e.g. by evaluating the "Kretschmann invariant": $R_{abcd}R^{abcd} \to \infty$.

2. By contrast, $r = 2M$ $(r = r_S)$ corresponds to $T = \pm X$, which is not a singularity.

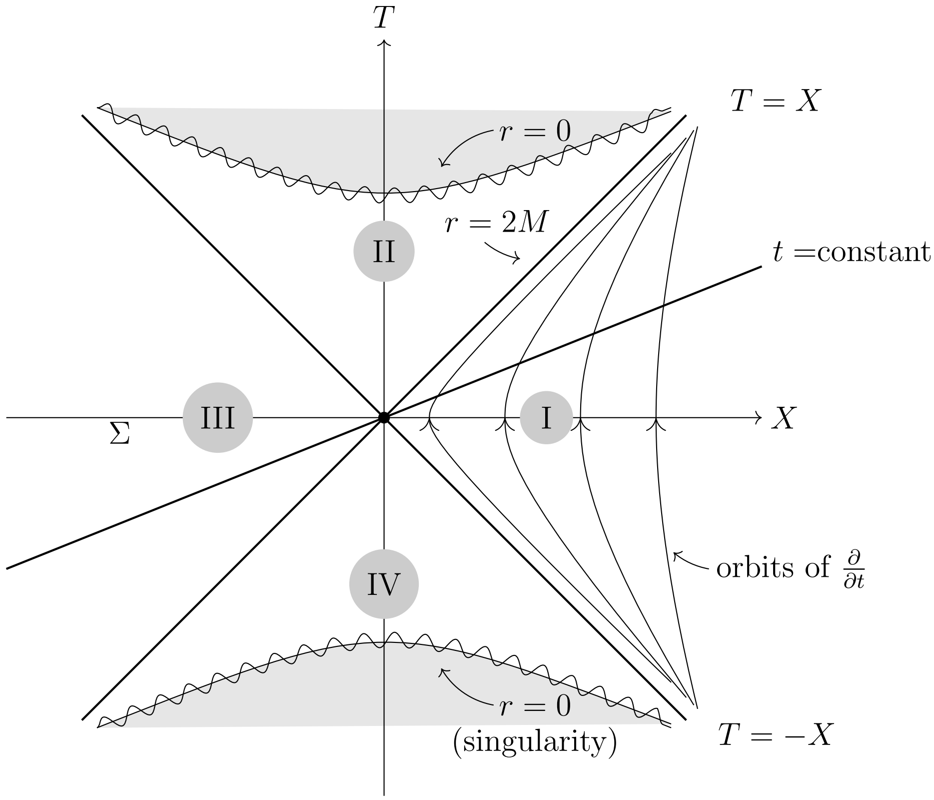
3. The surfaces of constant t corresponds to $\frac{T}{X} =$ constant.

4. At $T = 0 = X$ we have, from equation (46), that

$$dr\Big|_{T=X=0} = \frac{4M}{e}(XdX - TdT)\Big|_{X=T=0} = 0$$

   and that $g^{ab}(dr)_b$ and $\left(\frac{\partial}{\partial t}\right)^a$ (when expressed in X, T-coordinates) become co-linear at $T = \pm X$. It follows that $\frac{\partial(r,t)}{\partial(X,T)}$ becomes singular there, too, showing that (r, t) are 'bad coordinates'. The apparent singularity at $r = r_S$ in the original coordinates is hence due to a bad choice of coordinates.

The causal structure of Schwarzschild following from equation (48) is best illustrated in a diagram in which the (ϕ, θ)-coordinates are suppressed.



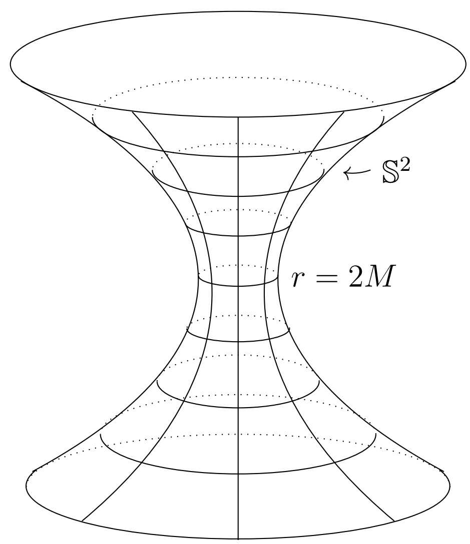
1. Region I (excluding $r = 2M$-lines) corresponds to the original coordinate range $r > 2M$ ('exterior ').

2. Region II has the property that no lightlike-future-directed curves can enter the exterior region I ('black hole').

3. Region III has the property that no lightlike-future-directed curves can stay within that region forever ('white hole').

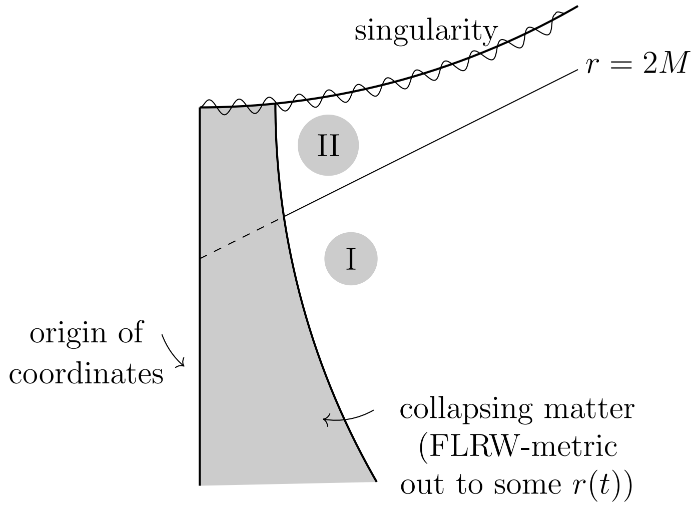
4. Region IV has properties identical to those of region I. If we consider the metric $h = h_{ij}dx^i dx^j$ with $(x^i) = (X, \theta, \phi)$ induced on $\Sigma = \{T = 0\}$, we obtain a Riemannian 3-manifold streching between regions I and IV. Its geometry is illustrated in the following figure (for a $2+1$ version of Schwarzschild for the sake of visualization). The embedding into flat space is intended to be such that the induced metric corresponds to

h_{ij}. We hence see the appearance of a "throat" connecting the exterior (region I) with a "parallel universe" (region IV).



5. The black hole/white hole regions are separated from the exterior region by a pair of null surfaces called 'event horizons '. In the old coordinates, these surfaces are both located at $r = r_S$. Thus, rather than being a singularity, the Schwarzschild radius describes the location of the event horizons.

The 'parallel universe' has attracted considerable attention in Science Fiction, but it is unrealistic that it could be formed in the real world. A more realistic spacetime diagram illustrating the formation of a black hole from collapsing matter is as follows. In this diagram (describing a solution suitably patched together from a part of Schwarzschild and a suitable spherically symmetric solution of the Einstein equation with a stress tensor in perfect fluid form), the parallel universe is covered up by the collapsing star.



14 Linearized Gravity and Gravitational Radiation

In the preceding sections we have described (i) a dynamical solution to Einstein's equations with no "spatial excitations" (cosmological solutions) and (ii) a static solution with non-trivial spatial dependence (Schwarzschild black hole). A generic solution has spatial excitations that evolve in time. It turns out that the evolution equations implied by Einstein's equations have the character of a "quasi-linear" wave equation. The analysis of such equations is beyond the scope of these notes, see for example [8] for a detailed exposition. However, it turns out that one can derive, without major difficulties, the solutions describing small "linear" perturbations of Minkowski spacetime. These solutions describe gravitational waves. They have no counterpart in Newtonian Gravity and are, in this sense, a genuinely new prediction of General Relativity. This prediction has recently been confirmed by the LIGO gravitational wave detector which has observed the gravitational waves produced by the merger of two black holes.

14.1 Gravitational waves in empty space

We first describe how to obtain the linearized Einstein equations. Mathematically, the best way to proceed is to suppose that one has a (differentiable) 1-parameter family $\{g_{ab}(s)\}_{s\in\mathbb{R}}$ of solutions to the Einstein equations, e.g. in vacuum,

$$R_{ab}(s) = 0$$

where we mean the Ricci tensor of the metric $g_{ab}(s)$. We think of $g_{ab} \equiv g_{ab}(0)$ as the "background", and we think of $g_{ab}(s), |s| \ll 1$ as small "deviations" or "perturbations" of this background. The first order deviation is just the derivative with respect to s at $s = 0$,

$$\gamma_{ab} \equiv \frac{\partial}{\partial s} g_{ab}(s)\bigg|_{s=0} .$$

γ_{ab} is referred to as the "linear perturbation". In General Relativity, the observable is not the metric, but its "gauge equivalence class", i.e. the set of all metrics related to g_{ab} by a diffeomorphism, $\psi^* g_{ab}$. Consequently, if $\{\psi(s)\}_{s\in\mathbb{R}}$ is a (differentiable) 1-parameter family of diffeomorphisms of M, then $\{g_{ab}(s)\}_{s\in\mathbb{R}}$ and $\{\psi(s)^* g_{ab}(s)\}_{s\in\mathbb{R}}$ should be viewed as physically describing the same families of spacetimes. Thus, at the linearized level (i.e. differentiating with respect to s and making use of the notion of Lie-derivatives), we find that γ_{ab} and $\gamma_{ab} + \mathcal{L}_X g_{ab}$ physically describe the same perturbation, where X^a is the generator of the family of diffeomorphisms at $s = 0$. One also says that the "gauge-invariance" at the linearized level is

$$\gamma_{ab} \to \gamma_{ab} + \mathcal{L}_X g_{ab} = \gamma_{ab} + \nabla_a X_b + \nabla_b X_a \ .$$

The linearized (vacuum) Einstein equations are obtained by simply differentiating the Einstein equations with respect to s at $s = 0$, namely

$$\dot{R}_{ab} \equiv \frac{\partial}{\partial s} R_{ab}(s)\bigg|_{s=0} = 0 \ .$$

The left side of this equation is readily calculated in terms of g_{ab} and γ_{ab}, but we will not give the general expression here. Rather we will specialize directly to the case of interest for us in which $g_{ab} = \eta_{ab}$ is Minkowski space. With $g_{ab}(s) = \eta_{ab} + s\gamma_{ab} + \mathcal{O}(s^2)$, we get:

$$\dot{g}^{ab} = -\gamma^{ab} \tag{49}$$

$$\dot{\Gamma}^c_{ab} = \frac{1}{2}\eta^{cd}\left(\partial_a\gamma_{bd} + \partial_b\gamma_{ad} - \partial_d\gamma_{ab}\right) \ , \tag{50}$$

where here and in the following, we adopt the convention that indices on expressions related to γ_{ab} are raised and lowered with η^{cd}, and where an overdot means a derivative with respect to s at $s = 0$. With these expressions at hand, we find that the linearized Riemann tensor is

$$\dot{R}_{eab}{}^{c} = -\frac{1}{2}\eta^{cd}\partial_e\left(\partial_a\gamma_{bd} + \partial_b\gamma_{ad} - \partial_d\gamma_{ab}\right) - (a \leftrightarrow e) \ ,$$

and we find that the linearized Ricci tensor is

$$\dot{R}_{ab} = \partial^c\partial_{(b}\gamma_{a)c} - \frac{1}{2}\partial^c\partial_c\gamma_{ab} - \frac{1}{2}\partial_a\partial_b\gamma \ ,$$

where $\gamma = \eta^{ab}\gamma_{ab}$ and where parenthesis denote symmetrization of the respective tensor indices. We then get for the linearized Einstein tensor:

$$\begin{aligned} \dot{G}_{ab} &= \dot{R}_{ab} - \frac{1}{2}\eta_{ab}\dot{R}_{cd}\eta^{cd} \\ &= \partial^c\partial_{(b}\gamma_{a)c} - \frac{1}{2}\partial^c\partial_c\gamma_{ab} - \frac{1}{2}\partial_a\partial_b\gamma - \frac{1}{2}\eta_{ab}\left(\partial^c\partial^d\gamma_{cd} - \partial^c\partial_c\gamma\right) \ . \end{aligned} \tag{51}$$

Using the formulae for Lie derivatives, the linearized gauge transformation on a Minkowski background is

$$\gamma_{ab} \to \gamma_{ab} + \partial_a X_b + \partial_b X_a \ ,$$

where X_a is an arbitrary (smooth) tensor. To investigate how these quantities change under a linearized gauge transformation we note that, for instance, $R_{abcd}[\psi^* g] = \psi^* R_{abcd}[g]$ for any metric g_{ef} and any diffeomorphism M (this of course expresses the "general covariance" of quantities like the Riemann tensor.) Applying this formula to a 1-parameter family $\{g_{ab}(s)\}$ of metrics and a 1-parameter family $\{\psi(s)\}$ of diffeomorphisms, taking a derivative with respect to s at $s = 0$, and using the formulas for the Lie-derivate, we get the transformation formulae

$$\begin{aligned} \dot{R}_{abcd} &\to \dot{R}_{abcd} + \mathcal{L}_X R_{abcd} \ , \\ \dot{G}_{ab} &\to \dot{G}_{ab} + \mathcal{L}_X G_{ab} \ , \end{aligned} \tag{52}$$

which hold for any linear perturbation of any background. Similar formulae also hold, by the same argument, for any other tensor field that is locally and constructed out of g_{ab}, ∇_a, g^{ab}. We can conclude from such formulae that any linearized quantity whose counterpart vanishes in the background, is automatically gauge invariant. For instance, if $G_{ab} = 0$ in the background, then the linearized Einstein tensor $\dot{G}_{ab}$ is gauge invariant. In Minkowski spacetime $R_{abcd} = 0$, so even the linearized Riemann tensor $\dot{R}_{abcd}$ is gauge invariant. In a FLRW (homogeneous, isotropic) spacetime $C_{abcd} = 0$, so the linearized Weyl-tensor $\dot{C}_{abcd}$ in such a spacetime is gauge invariant.

We will now derive a wave equation for the linearized Riemann tensor on Minkowski spacetime. For this, we first look at the linearized Bianchi identities on Minkowski spacetime. They read

$$\partial_{[a}\dot{R}_{bc]de} = 0 \tag{53}$$

$$\partial^d \dot{R}_{abcd} = 0 \ . \tag{54}$$

We now apply ∂^a to the first equation and use the second equation as well as $\dot{R}_{ab} = 0$. Then we get, indeed,

$$\partial^a \partial_a \dot{R}_{bcde} = 0 \ .$$

To analyze the effect of metric perturbations on the motion of test-observers, we choose the wave vector ωk_a, where

$$k^a = \frac{1}{\sqrt{2}}\left[\left(\frac{\partial}{\partial t}\right)^a - \left(\frac{\partial}{\partial z}\right)^a\right]$$

and consider a corresponding plane fronted wave-like perturbation γ_{ab} moving in the z-direction with spacetime dependence $\sin[\omega(t+z)]$. The corresponding linearized Riemann tensor is then a solution to the wave equation. To derive the motion of test-observers on this background, it is convenient to introduce another null vector l^a by

$$l^a = \frac{1}{\sqrt{2}}\left[\left(\frac{\partial}{\partial t}\right)^a + \left(\frac{\partial}{\partial z}\right)^a\right] \ ,$$

and define the symmetric tensor

$$\Omega_{ab} \equiv \frac{1}{2}\ \dot{R}_{acbd} l^c l^d \qquad \Rightarrow \qquad \partial^c \partial_c \Omega_{ab} = 0 \ .$$

The quantity Ω_{ab} has the following properties:

1. Ω_{ab} is invariant under linear gauge transformations.

2. $\Omega_{ab}k^b = 0$ (from the second linearized Bianchi identity and the dependence $\dot{R}_{abcd} \propto \sin[\omega(k^e x_e)]$), and $\Omega_{ab}l^b = 0$ (from $\dot{R}_{(ab)cd} = R_{ab(cd)} = 0$).

3. $\Omega_{ab}\eta^{ab} = 0$ (from $\dot{R}_{ab} = 0$), and $\Omega_{ab} = \Omega_{ba}$ (from $\dot{R}_{abcd} = \dot{R}_{cdab}$).

It follows that Ω_{ab} is a trace-free, symmetric tensor having components only in the x, y-directions. We can therefore write

$$\Omega_{ab} = \omega^2 \left\{ h_+ \cdot \epsilon^+_{ab} + h_\times \cdot \epsilon^\times_{ab} \right\} \sin[\omega(t+z)] \ ,$$

where $\{\epsilon^+_{ab}, \epsilon^\times_{ab}\}$ forms a basis of such tensors; in coordinates

$$\epsilon^+_{ab} = (dx)_a(dx)_b - (dy)_a(dy)_b \ , \qquad \epsilon^\times_{ab} = (dx)_a(dy)_b + (dx)_b(dy)_a \ ,$$

and where $h_+, h_\times \in \mathbb{R}$ are the amplitues of the "polarizations" $+, \times$. In inertial coordinates the polarization tensors for a wave moving in the z-direction are

$$(\epsilon^+_{\mu\nu}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (\epsilon^\times_{\mu\nu}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \ .$$

We now consider the effect of a plane gravitational wave moving in the z-direction on the motion of test-observers. The motion of the test-observers is described by the geodesic deviation equation. We need to understand what this equation tells us at the order of approximation considered here. For this, it is best again to think about a family of metrics $g_{ab}(s) = \eta_{ab} + s\gamma_{ab} + \mathcal{O}(s^2)$. We consider geodesics starting on a slice $\Sigma = \{t = 0\}$ which are initially parallel with tangent vector $T^a = (\partial/\partial t)^a$. The evolution of these geodesics takes place in the s-dependent spacetime $g_{ab}(s)$. For $s = 0$, the metric is flat space and the geodesic deviation vector vanishes identically. Consequently, the geodesic deviation vector is Taylor expanded to *second* order in s as

$$X^a(s,t) = s\dot{X}^a(t) + \frac{1}{2}s^2\ddot{X}^a(t) + \mathcal{O}(s^3),$$

where an overdot again stands for a derivative with respect to the parameter s (and *not* the time parameter, t). We next differentiate the geodesic deviation equation several times with respect to s at $s = 0$. In principle, all quantities in the geodesic deviation equation depend upon the parameter s, including also $T^a = T^a(s,t)$, the tangent to the geodesics, because the metric $g_{ab}(s)$ depends upon s. Taking one s-derivative of the geodesic deviation equation gives

$$\frac{\partial^2}{\partial t^2}\dot{X}^a(t) = 0 \ ,$$

because the zeroth order deviation vector and Riemann tensor vanish. Since the geodesics are not initially diverging, it follows that $\dot{X}^a(t) = \dot{X}^a(0)$ is independent of t. Taking two s-derivatives gives

$$\begin{aligned} \frac{\partial^2}{\partial t^2}\ddot{X}^a(t) &= \dot{R}^a{}_{bcd}(t)T^b(0)T^d(0)\dot{X}^c(t) \ , \qquad (55) \\ &= \Omega^a{}_b(t)\dot{X}^b(0) \ , \qquad (56) \end{aligned}$$

because at $s=0$, $T^a(0,t) = (\partial/\partial t)^a \equiv T^a$ is constant, and because we have already seen that $\dot{X}^a(t) = \dot{X}^a(0)$ is constant, too. In the second line, we have also used the definition and properties of Ω_{ab}.

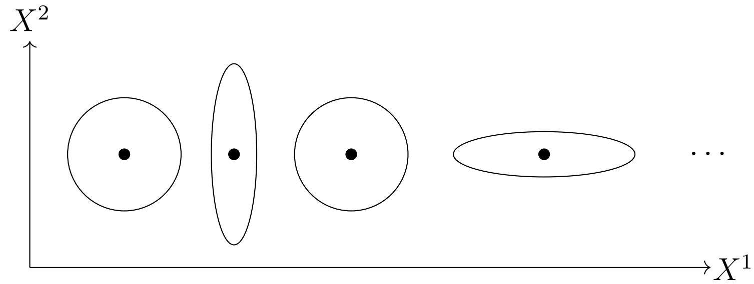
This equation may now be integrated. We can write $(X^\mu) = (0, X^1, X^2, 0)$ up to second order in the expansion in s, because the z-component must vanish, and the t-component vanishes by construction of the congruence. Integration gives for $h_\times = 0$

$$\left.\begin{aligned} X^1(t) &\approx X^1(0) + \tfrac{1}{2}h_+ \sin(\omega t)\ X^1(0), \\ X^2(t) &\approx X^2(0) - \tfrac{1}{2}h_+ \sin(\omega t))\ X^2(0). \end{aligned}\right\} \qquad (57)$$

(Taylor expansion up to and including order s^2-terms with $s=1$), whereas for $h_+ = 0$, we obtain

$$\left.\begin{aligned} X^1(t) &\approx X^1(0) + \tfrac{1}{2}h_\times \sin(\omega t)\ X^2(0), \\ X^2(t) &\approx X^2(0) + \tfrac{1}{2}h_\times \sin(\omega t)\ X^1(0). \end{aligned}\right\} \qquad (58)$$

These displacements correspond to oscillations of a ring of test-masses (in the rest frame defined by T^a) in the (x,y)-plane as shown in the following figure.



We summarize our discussion as follows. At the linearized level, perturbations of Minkowski space (or rather, their corresponding Riemann tensor) obey a homogeneous wave equation. A plane wave moving in the z-direction gives rise to an oscillation of a ring of test-masses in the (x, y)-plane. The oscillation pattern depends on the polarization, $+$, respectively $\times$. Since the physical degrees of freedom of the gravitational field (i.e. gauge invariant information) can only manifest themselves via their influence on test-masses, we can say that a gravitational wave has "two degrees of freedom" (per wave vector), namely $+, \times$.

14.2 Sources of gravitational waves

We next discuss the production of gravitational waves. For this, we need to study the linearized Einstein equations with a non-trivial stress tensor T_{ab} representing the source. It is convenient at this stage to introduce the "trace reversed" variable

$$h_{ab} = \gamma_{ab} - \frac{1}{2}\eta_{ab}\gamma^c_c \qquad \Leftrightarrow \qquad \gamma_{ab} = h_{ab} - \frac{1}{2}\eta_{ab}h^c_c \ .$$

In terms of this variable, the linearized Einstein tensor takes the form

$$\begin{aligned} \dot{G}_{ab} &= -\frac{1}{2}\partial^c\partial_c h_{ab} + \partial^c\partial_{(a}h_{b)c} - \frac{1}{2}\eta_{ab}\partial^c\partial^d h_{cd} \\ &= 8\pi G_N \ T_{ab} \ . \end{aligned}$$

where we now have a stress energy tensor on the right side. To be consistent at the linearized level, the stress energy of the source should satisfy $\partial^a T_{ab} = 0$ (flat covariant derivative). Recall that the gauge invariance of General Relativity at the linearized level is

$$\gamma_{ab} \to \gamma_{ab} + \partial_a X_b + \partial_b X_a \ .$$

where the expression for the Lie derivative of the Minkowski metric has been used, and recall that $\dot{G}_{ab}$ is gauge invariant. We may use the gauge invariance to fix a particularly useful representer in the gauge equivalence class of γ_{ab} (and thus h_{ab}). For this, we note that under a gauge transformation

$$\partial^c h_{ac} \to \partial^c h_{ac} + \partial^c \partial_c X_a \ .$$

Because $\partial^c \partial_c$ is the wave operator in Minkowski spacetime, we can find a solution X_a to the equation $\partial^c \partial_c X_a = -\partial^c h_{ca}$. Using any X_a satisfying this equation in the gauge transformation, it follows that the gauge-transformed linear perturbation has $\partial^c h_{ac} = 0$. This gauge is called the "Lorentz gauge", by analogy with Maxwell's equations. In the Lorentz gauge, the linearized Einstein equation simply becomes

$$\partial^c \partial_c h_{ab} = -16\pi G_N \ T_{ab}$$

We see again that the evolution equation for linear perturbations is a *wave equation*. It can be shown that the residual gauge freedom can be used up to impose even more stringent gauge conditions such as $\gamma = 0 = \gamma_{ab} T^b$ (in the source free region where $T_{ab} = 0$), see [Wald 1984] for a detailed discussion. For a plane gravitational wave in empty space with wave-vector ωk_a as in the preceding section, we then get the conditions $\gamma = \gamma_{ab} T^b = \gamma_{ab} k^b = 0$. This reduces the number of independent components of γ_{ab} from 10 down to 2 (corresponding to the fact that there are 8 independent gauge conditions). Thus, we see again that the gravitational field has 2 degrees of freedom per wave vector ωk_a, corresponding to $+, \times$ polarized waves. We next discuss the

Production of gravitational waves: Significant amounts of gravitational waves are produced in Nature by binary systems with large masses and large orbital frequencies, for instance by binaries of neutron stars or black holes (especially during their merger phase), but also in the Early Universe. Here we imagine a localized source such as a binary. The (in principle very complicated) structure of the source is supposed to be encoded in the matter stress tensor, T_{ab}. We imagine that T_{ab} has compact support in a spacetime region where for instance a collapse or merger takes place. We should solve $\partial^c \partial_c h_{ab} = -16\pi G_N \ T_{ab}$ with "retarded boundary conditions" (the effects of the source propagate to the future), which corresponds to

$$h_{ab} = -16\pi G_N \ \Delta^{\mathrm{ret}} * T_{ab} \ .$$

Here, the star is convolution, and Δ^{ret} is the retarded propagator

$$\Delta^{\text{ret}}(t, \mathbf{x}) = -\frac{1}{2\pi}\Theta(-t)\delta(t^2 - \|\mathbf{x}\|^2),$$

where Θ is the heaviside step function. From this, we get (with $G_N = 1$)

$$\begin{aligned} h_{ab}(t, \mathbf{x}) &= 8\int dt' d^3x' \; \Theta(t-t')\delta\left((t-t')^2 - \|\mathbf{x}-\mathbf{x}'\|^2\right) T_{ab}(t', \mathbf{x}') \qquad (59) \\ &= 4\int d^3x' \; \frac{T_{ab}(t', \mathbf{x}')}{\|\mathbf{x}-\mathbf{x}'\|}\Bigg|_{t-t'=\|\mathbf{x}-\mathbf{x}'\|} \\ &= 4\int\limits_{\dot{V}-(x)} \frac{T_{ab}(x')}{\|\mathbf{x}-\mathbf{x}'\|}\underbrace{ds(x')}_{=r^2 dr d\Omega}, \end{aligned}$$

where $ds(x')$ is the invariant line element on the past lightcone centered at $x = (\mathbf{x}, t)$, see the following figure. By taking a Fourier-transformation in t

$$\hat{h}_{ab}(\omega, \mathbf{x}) = \frac{1}{\sqrt{2\pi}}\int e^{i\omega t} h_{ab}(t, \mathbf{x}) dt \ ,$$

we obtain

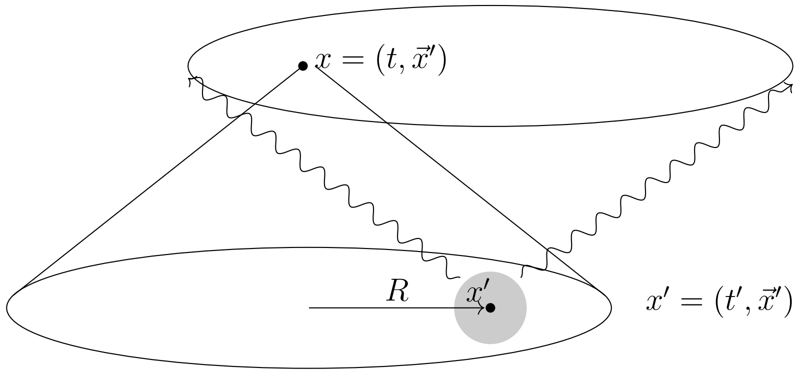
$$\hat{h}_{ab}(\omega, \mathbf{x}) = 4\int \frac{\hat{T}_{ab}(\omega, \mathbf{x}')}{\|\mathbf{x}-\mathbf{x}'\|} e^{i\omega\|\mathbf{x}-\mathbf{x}'\|} d^3x'$$

The divergence-free condition $\partial^\mu h_{\mu\nu} = 0$ gives in Fourier-space

$$-i\omega\hat{h}_{t\nu} = \partial^j \hat{h}_{j\nu}$$

Now we assume that $R \gg \frac{1}{\omega}$ and $e^{i\omega\|\mathbf{x}-\mathbf{x}'\|}$ are nearly constant over the source, as seems physically reasonable (see the following figure). Then we can say that

$$\frac{e^{i\omega\|\mathbf{x}-\mathbf{x}'\|}}{\|\mathbf{x}-\mathbf{x}'\|} \approx \frac{e^{i\omega R}}{R} \ .$$



The following elementary manipulations are valid under the assumption that T_{ab} decays sufficiently rapidly away from the source (for instance compact support) so that we can perform partial integrations with vanishing boundary terms:

$$
\begin{aligned}
\int \hat{T}^{ij} d^3x &= \int \Big[\overbrace{\partial_k \left(\hat{T}^{kj} x^i \right)}^{=0 \text{ by Gauss theorem}} - \partial_k \hat{T}^{kj} x^i \Big] d^3x \\
&= -i\omega \int \hat{T}^{tj} x^i d^3x \\
&= -\frac{i\omega}{2} \int \left(\hat{T}^{tj} x^i + \hat{T}^{ti} x^j \right) d^3x \\
&= -\frac{i\omega}{2} \int \Big[\overbrace{\partial_k \left(\hat{T}^{tk} x^j x^i \right)}^{=0 \text{ by Gauss theorem}} - \partial_k \hat{T}^{tk} x^i x^j \Big] d^3x \\
&= -\frac{\omega^2}{2} \underbrace{\int \hat{T}^{tt} x^i x^j d^3x}_{=\frac{1}{3}\hat{q}^{ij}(\omega) \text{ (quadrupole-tensor)}}
\end{aligned}
$$

So we get

$$\hat{h}^{ij}(\omega, \mathbf{x}) \approx -\frac{2\omega^2}{3} \frac{e^{i\omega R}}{R} \hat{q}^{ij}(\omega)$$

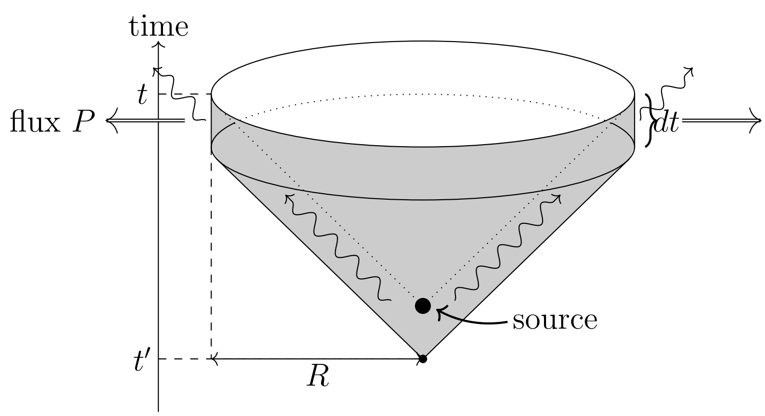
and after an inverse Fourier-transformation

$$h_{ij}(t, \vec{x}) \approx \frac{2}{3R} \frac{d^2}{dt^2} q_{ij} \underbrace{(t - R)}_{=t'}$$

where the approximation is valid in the slow motion and large-distance approximation.

One would like to calculate the energy flux of gravitational radiation, i.e. the energy emitted by the source per unit of time. In General Relativity, the notion of the total energy of a spacetime or parts thereof is actually not so easy to define, mainly due to the invariance of the theory under diffeomorphisms. We shall not discuss this complicated issue further, but note that, in the case of *linearized* gravity, a satisfactory notion of energy $E(t)$ associated with suitable time slices $\Sigma(t)$ can be defined. With this notion, the flux, $P(t)$ is then defined as

$$\underbrace{P(t)}_{\text{Flux}}\,dt = d\,\underbrace{E(t)}_{\text{Energy of slice }\Sigma(t)}$$



One way to define this energy E for linear perturbations is as follows: Set, for any pair of linearized perturbations:

$$w^a = \eta^{abcdef}\left(\gamma^{(1)}_{bc}\partial_d\gamma^{(2)}_{ef} - \gamma^{(2)}_{bc}\partial_d\gamma^{(1)}_{ef}\right)$$

where

$$\eta^{abcdef} = \eta^{ae}\eta^{fb}\eta^{cd} - \frac{1}{2}\eta^{ad}\eta^{be}\eta^{cf} - \frac{1}{2}\eta^{bc}\eta^{ae}\eta^{fd} - \frac{1}{2}\eta^{ab}\eta^{cd}\eta^{ef} + \frac{1}{2}\eta^{bc}\eta^{ad}\eta^{ef}\ .$$

(This quantity is also called the "symplectic current".) A calculation using the linearized equations of motion $\dot{R}_{ab} = 0$ (for both perturbations) shows that $\partial^a w_a = 0$ in the source free region. Furthermore, define $j^a = w^a(\gamma, \partial_t\gamma)$. Then it can be shown that

1. $\partial_a j^a = 0$ in the source free region (this immediately follows from $\partial^a w_a = 0$).

2. If $\Sigma(t)$ is a surface in the source free region, then

$$E(t) = \frac{1}{8\pi G_N} \int_{\Sigma(t)} j_a n^a dS$$

is gauge invariant (here n^a is the unit normal to the surface and dS the integration element).

3. $E(t)$ remains unchanged if we deform any compact subset of the surface $\Sigma(t)$.

4. $E(t)$ is decreasing with time in the source free region.

The proofs of these claims follow from the arguments in [7]. These properties suggest that $E(t)$ should be viewed as the energy of the linear perturbation γ_{ab} at "time t" (in the source free region) if we define $\Sigma(t)$ as a suitably "asymptotically hyperboloidal" slice approaching a lightcone at "retarded time", t. A possible choice is

$$\Sigma(t) = \{(x^\mu) \mid (x^0 - t)^2 - \sum_{j=1}^{3} (x^j)^2 = 1\} \ .$$

The corresponding flux $P(t)$ emitted by a gravitational wave in the far region may then be calculated. After a rather lengthy calculation not recorded here, it is found that

$$P(t) = \frac{d}{dt} E(t) \approx \frac{G_N}{45} \sum_{i,j=1}^{3} \left(\dddot{Q}_{ij}(t - R) \right)^2$$

where $Q_{ij} = q_{ij} - \frac{1}{3}\delta_{ij} q$ is the traceless part of the quadrupole tensor. This relation is known as the "quadrupole formula". (Note that the corresponding formula in electromagnetism involves only the dipole moment. This difference can be traced back to the difference in the tensor character of both fields.)

To get an impression of the order of magnitude of gravitational radiation, we consider an

Example: We consider a rigidly rotating rod of length $2L$ and orbital frequency Ω. In order to determine the stress tensor representing the rod, we first choose cylindrical coordinates (t, r, z, ϕ) adapted to the rod, assuming that the rotation is around the z-axis. The line element of Minkowski spacetime becomes

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}r^2 + r^2\,\mathrm{d}\phi^2 + \mathrm{d}z^2 \tag{60}$$

in those coordinates. In order to make the formulas look more intuitive, it is preferable to further switch, at an intermediate stage, to a body-fixed cylindrical coordinate system defined by $\hat{t} = t$, $\hat{r} = r$, $\hat{z} = z$, $\hat{\phi} = \phi + \Omega t$, in which the line element reads

$$\mathrm{d}s^2 = -\left(1 - \Omega^2\hat{r}^2\right)\mathrm{d}\hat{t}^2 - 2\hat{r}^2\Omega\,\mathrm{d}\hat{\phi}\,\mathrm{d}\hat{t} + \mathrm{d}\hat{r}^2 + \hat{r}^2\,\mathrm{d}\hat{\phi}^2 + \mathrm{d}\hat{z}^2 \ . \tag{61}$$

Going over from here to body-fixed 'Cartesian' coordinates $\hat{x} = \hat{r}\cos\hat{\phi}$, $\hat{y} = \hat{r}\sin\hat{\phi}$ (rest frame of the rod) we have

$$\mathrm{d}s^2 = -\left(1 - \Omega^2\hat{r}^2\right)\mathrm{d}\hat{t}^2 - 2\Omega\left(\hat{x}\,\mathrm{d}\hat{y} - \hat{y}\,\mathrm{d}\hat{x}\right)\mathrm{d}\hat{t} + \mathrm{d}\hat{x}^2 + \mathrm{d}\hat{y}^2 + \mathrm{d}\hat{z}^2 \ . \tag{62}$$

Without loss of generality, we can assume that the rod is in the $\hat{x}$-direction in its rest frame and we assume that it is reflection symmetric. Thus, its energy momentum tensor in the coordinates $(\hat{t}, \hat{x}, \hat{y}, \hat{z})$ must be of the general form

$$T^{\hat{t}\hat{t}} = \rho(|\hat{x}|)\Theta(L - |\hat{x}|)\delta(\hat{y})\delta(\hat{z})\,, \tag{63a}$$
$$T^{\hat{x}\hat{x}} = -\sigma(|\hat{x}|)\Theta(L - |\hat{x}|)\delta(\hat{y})\delta(\hat{z})\,, \tag{63b}$$

and all other components are zero. ρ is the energy density per unit of length of the rod, and σ the stress along the rod per unit length. Transforming back to non-corotating cylindrical coordinates, we get

$$T^{tt} = \rho(r)\Theta(L - r)\frac{\delta(\sin(\phi + \Omega t))}{r}\delta(z)\,, \tag{64a}$$
$$T^{t\phi} = -\Omega\rho(r)\Theta(L - r)\frac{\delta(\sin(\phi + \Omega t))}{r}\delta(z)\,, \tag{64b}$$
$$T^{rr} = -\sigma(r)\Theta(L - r)\frac{\delta(\sin(\phi + \Omega t))}{r}\delta(z)\,, \tag{64c}$$
$$T^{\phi\phi} = \Omega^2\rho(r)\Theta(L - r)\frac{\delta(\sin(\phi + \Omega t))}{r}\delta(z)\,. \tag{64d}$$

where

$$\delta(\phi \mod \pi) = \sum_{k=-\infty}^{\infty}\delta(\phi + k\pi) = \delta(\sin\phi)\,. \tag{65}$$

Covariant conservation $\partial_a T^{ab} = 0$ gives the single condition (exercises)

$$\sigma(r) = \Omega^2 \int_r^L s\rho(s)\,\mathrm{d}s\,, \tag{66}$$

which actually could have been guessed beforehand. We now specialize to a density ρ representing two point masses A and B of equal mass M separated by $2L$. Transforming the above general formula to the non-rotating inertial coordinates (t,x,y,z) gives the tt-component

$$\begin{aligned} T^{tt}(t,x,y,z) &= M\delta(z)\Big\{\underbrace{\delta\left(x - L\cos\Omega t\right)\delta\left(y - L\sin\Omega t\right)}_{A} \\ &+ \underbrace{\delta\left(x + L\cos\Omega t\right)\delta\left(y + L\sin\Omega t\right)}_{B}\Big\}. \end{aligned}$$

Based on this formula, we can now determine the power of gravitational radiation given off by the rotating rod according to the quadrupole formula. We first need to calculate the reduced quadrupole (with $q_{ij} = 3\int T^{tt} x_i x_j d^3x$ and $Q_{ij} = q_{ij} - \frac{1}{3}\delta_{ij} q$ as before), which turns out to be

$$(Q_{ij}) = \frac{ML^2}{3}\begin{pmatrix} 1+3\cos 2\Omega t & 3\sin 2\Omega t & 0 \\ 3\sin 2\Omega t & 1-3\cos 2\Omega t & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

Taking three time derivatives, substituting this into the quadrupole formula, and reinstating c, this is seen to result in the flux of the binary:

$$P_{\text{binary}} = -\frac{128}{45c^5} G_N M^2 L^4 \Omega^6\ .$$

The factor Ω^6 stems from six time derivatives in the quadrupole formula and effectively turns P into an astronomically small number for typical systems: For instance, for masses/lengths of the order of the Earth-Sun system, one finds a flux of about just 100 Watts per second. To get an appreciable flux, one needs sources that are very massive and spinning very fast, and one needs sufficiently long observation times. Such sources are provided for example by pulsars, consisting of a pair of orbiting neutron stars. The orbital frequency Ω is observable to a high precision due to the "lighthouse effect", whereas the gravitational radiation produced is sufficient to give appreciable energy loss on human timescales. This energy loss affects the orbital frequency and can hence be observed indirectly, in principle. Such a change in the orbital frequency has now been observed in several pulsars. The quantitative results seem to confirm the quadrupole formula, and hence are a stringent test of General Relativity.

15 The Global Positioning System

The motivation behind the Global Positioning System is to accurately determine positions and times for any events near Earth. To do this, we need to take certain relativistic effects into account. The implementation of the GPS seems to be the first application of General Relativity on a large scale which is relevant to a general public because of commercial applications like car navigation equipment.

The results and presentation in section are mostly based on [1].

15.1 Introduction

Let us ignore gravity for the moment and suppose that we can describe all events near Earth using inertial coordinates $(ct, \mathbf{x})$ in the sense of Special Relativity. Suppose that we (or an observer O) are at some unknown event $(ct_O, \mathbf{x}_O)$, but we receive radio signals from four sources and information about the four events $(ct_j, \mathbf{x}_j)$, $j = 1, 2, 3, 4$, where these signals originated. In general we can then determine the coordinates $(ct_O, \mathbf{x}_O)$ using a simple triangulation: because the radio signals travel at the speed of light, the four equations

$$t_O - t_j = c^{-1} \|\mathbf{x}_O - \mathbf{x}_j\|, \qquad j = 1, 2, 3, 4 \tag{67}$$

must be satisfied. These four equations allow us to determine the four unknown coordinates, in general.

A few remarks are in order:

1. If the signals are not received at exactly the same time, but at times $t_O + \delta_j$ and at positions $\mathbf{x}_O + \mathbf{d}_j$, we can use the equations

$$t_O + \delta_j - t_j = c^{-1} \|\mathbf{x}_O + \mathbf{d}_j - \mathbf{x}_j\|, \qquad j = 1, 2, 3, 4.$$

   Note that the values for δ_j can be measured by the local observer (setting e.g. $\delta_1 = 0$ to eliminate a free constant). In addition we can make reasonable estimates for the vectors $\mathbf{d}_j$, because in most applications these are mostly due to the rotation of Earth. (The rotation speed depends on the latitude, but it is typically much larger than the speed of the observer with respect to Earth.) In this way we reduce the problem again to four unknowns, which may be obtained from four equations.

2. In the GPS, the sources are artificial satellites in orbit around Earth. There are 24 GPS satellites put into orbit in such a way that at least four of them are visible at almost every place and time on Earth. If the orbits of the satellites are known rather accurately as a function of time, it only remains for the satellite to determine the time, using an on-board clock.

3. In order to determine positions with an accuracy of 1m, we see from equation (67) that we need to determine times with an accuracy of $\frac{1\mathrm{m}}{c\frac{\mathrm{m}}{\mathrm{s}}} \sim 3.33 \cdot 10^{-9}\mathrm{s} = 3.33\mathrm{ns}$.

High precision time measurement is one of the main challenges of the GPS system, and it is affected by the gravitational field of Earth and the motion of the satellites (and the observer O who wishes to determine its event). Some basic questions that are raised by the GPS and that involve relativity theory are:

1. How do we model the gravitational field of Earth?

2. What coordinates do we use to describe events?

3. How do we accurately measure the time coordinate using a clock on a satellite in orbit?

4. How do we accurately synchronise the clocks on various GPS satellites?

5. How do we communicate between satellites and the user, without losing accuracy?

The next subsections will address these issues.

15.2 Modelling the gravitational field of Earth

The gravitational field of Earth is rather complicated due to the details of its shape (not quite a sphere), its mass distribution (core vs. surface, e.g.) and its motion (rotation around an axis which itself is precessing, orbit around the sun). We will need to make some simplifying assumptions to deal with those issues, but if our assumptions are too strong, they will decrease the accuracy of our GPS.

We will neglect the dynamics of Earth's gravitational field, assume that the gravitational field is weak and that Earth has at least some symmetry, but not as much as spherical symmetry. In the weak field approximation to General Relativity, around a Minkowski background in spherical coordinates, we can write the metric as

$$ds^2 = \left(1 + \frac{2V}{c^2}\right) c^2 dt^2 - \left(1 - \frac{2V}{c^2}\right) (dr^2 + r^2 d\theta^2 + r^2 \sin^2(\theta) d\varphi^2), \tag{68}$$

where we can take V to be the Newtonian gravitational potential of Earth. This will be an approximate solution to Einstein's equation (it may be compared e.g. to the Schwarzschild metric).

If the mass distribution of Earth is given by a function $\rho(\mathbf{x})$, then the Newtonian gravitational potential is given by

$$V(\mathbf{x}) = -G_N \int \frac{\rho(\mathbf{y})}{\|\mathbf{x} - \mathbf{y}\|} d\mathbf{y}.$$

We take the coordinates to be centered on the centre of Earth and we use the Taylor expansion

$$\frac{1}{\|\mathbf{x} - \mathbf{y}\|} = \frac{1}{\|\mathbf{x}\|} + \sum_{i=1}^{3} y^i \frac{x^i}{\|\mathbf{x}\|^3} + \frac{1}{2} \sum_{i,j=1}^{3} y^i y^j \frac{3x^i x^j - \|\mathbf{x}\|^2 \delta^{ij}}{\|\mathbf{x}\|^5} + \ldots$$

to make a multi-pole expansion of V, namely

$$\begin{aligned} V(\mathbf{x}) &= \frac{-G_N M}{\|\mathbf{x}\|} - G_N \sum_{i=1}^{3} N^i \frac{x^i}{\|\mathbf{x}\|^3} - \frac{G_N}{2} \sum_{i,j=1}^{3} \frac{q^{ij}}{3} \frac{3x^i x^j - \|\mathbf{x}\|^2 \delta^{ij}}{\|\mathbf{x}\|^5} + \ldots \\ M &:= \int \rho(\mathbf{y}) d\mathbf{y}, \\ N^i &:= \int y^i \rho(\mathbf{y}) d\mathbf{y}, \\ q^{ij} &:= 3 \int y^i y^j \rho(\mathbf{y}) d\mathbf{y}. \end{aligned}$$

Here M is the total mass of the Earth. To find the other multi-pole coefficients we make some assumptions on the mass distribution ρ, namely that it has cylindrical symmetry around the x^3-axis, which we take to coincide with Earth's rotational axis, and that it has a reflection symmetry in the equatorial plane $x^3 = 0$. This means in particular that ρ is invariant if we change the sign of one of the coordinates y^i. It then easily follows that $N^i = 0$ and $q^{ij} = 0$ if $i \neq j$. Moreover, by the spherical symmetry, $q^{11} = q^{22}$. Using the fact that

$$\sum_{i=1}^{3} q^{11} \frac{3x^i x^i - \|\mathbf{x}\|^2 \delta^{ii}}{\|\mathbf{x}\|^5} = 0$$

we then find

$$V(\mathbf{x}) \simeq \frac{-G_N M}{r} \left(1 - J_2 \frac{a^2}{r^2} \frac{3\cos^2(\theta) - 1}{2} \right), \tag{69}$$

in spherical coordinates (r, θ, ϕ), where $a \sim 6.38 \cdot 10^6$m is Earth's radius at the equator and

$$J_2 = \frac{1}{3a^2 M}(q^{11} - q^{33}) \sim 1.08 \cdot 10^{-3}$$

is Earth's quadrupole moment coefficient. Higher multi-pole moments are not needed for GPS at the present level of accuracy, so our model for Earth's gravitational field consists of the metric (68) with V given by Equation (69).

15.3 Choice of coordinates

The form of the metric (68) already uses a set of coordinates (t, r, θ, ϕ) with a number of nice properties:

1. (t, r, θ, ϕ) are approximately inertial coordinates (written in spherical coordinates). They would be exactly inertial if we would ignore the gravitational field, setting $V = 0$. Because they are centered on the centre of Earth, they are almost inertial coordinates near Earth's geodesic world-line (compare to Fermi-Walker coordinates along a geodesic).

2. (r, θ, ϕ) nicely reflect the assumed symmetries: the rotation of Earth is described by a varying angular coordinate ϕ. Note that our coordinate system does not rotate along with Earth. (This would violate the approximately inertial property.)

3. t describes the proper time of a static observer at infinity, i.e. far away from Earth, where the potential V can be neglected altogether, because $V \sim 0$ when r is large.

These coordinates are very useful for describing e.g. the motions of the GPS satellites. However, they are less useful for the GPS users, who essentially rotate along with Earth. We therefore introduce an additional, rotating coordinate system:

$$t' = t, \qquad r' = r, \qquad \theta' = \theta, \qquad \phi' = \phi - \omega t, \tag{70}$$

where $\omega \sim 7.29 \cdot 10^{-5} \frac{\text{rad}}{\text{s}} \sim \frac{2\pi \text{ rad}}{24 \text{ hrs}}$ is the angular frequency of Earth's rotation. In these coordinates we have, up to order c^{-2}:

$$\begin{aligned} ds^2 &= \left(1 + \frac{2\Phi}{c^2}\right) c^2 dt'^2 - \frac{2\omega(r' \sin(\theta'))^2}{c} d\phi' \, cdt' \\ &\quad - \left(1 - \frac{2V}{c^2}\right)(dr'^2 + r'^2 d\theta'^2 + r'^2 \sin^2(\theta') d\varphi'^2), \end{aligned} \tag{71}$$

where $\Phi := V - \frac{(\omega r' \sin(\theta'))^2}{2}$ is an effective gravitational potential, which includes Earth's rotation as a centripetal potential term. Using Equation (69) for V we find

$$\Phi = \frac{-G_N M}{r'} + \frac{G_N M J_2 a^2}{r'^3} \frac{3\cos(\theta')^2 - 1}{2} - \frac{\omega^2 r'^2 \sin^2(\theta')}{2}. \tag{72}$$

To see how Earth's rotation influences time measurements we consider a clock at a fixed position on Earth, so r', θ', ϕ' are constant. The proper time τ is then related to the coordinate time t' by

$$d\tau = \left(1 + \frac{2\Phi}{c^2}\right)^{\frac{1}{2}} dt' \sim \left(1 + \frac{\Phi}{c^2}\right) dt', \tag{73}$$

up to order c^{-2}. At the equator we have $r' = a$, $\theta' = \frac{\pi}{2}$ and for $\Phi_0 := \Phi|_{\text{equator}}$ we have

$$c^{-2}\Phi_0 = \frac{-G_N M}{ac^2} - \frac{G_N M J_2}{2ac^2} - \frac{\omega^2 a^2}{2c^2} \sim -6.95 \cdot 10^{-10} - 3.76 \cdot 10^{-13} - 1.2 \cdot 10^{-12}.$$

The conclusion is that according to Equation (72), the (proper) time measured by a clock at the equator differs from the coordinate time t' by a change of rate or the order $\sim 7 \cdot 10^{-10}$.

Let us conclude this subsection with three remarks:

1. Within a matter of seconds, the error between $d\tau$ and dt' would reach our desired time accuracy $\sim 3{\cdot}10^{-9}$s. Note, however, that we ultimately want to compare time measurement on Earth with that on a satellite, not at infinity.

2. Two clocks which are fixed on Earth differ by a rate change that is determined by the values of Φ at their respective locations. These effects have to be taken into account when comparing clocks, e.g. for setting international time standards.

3. All clocks fixed on Earth at points where $\Phi = \Phi_0$ run at the same rate. These points form a surface, called the geoid of Earth.

4. For the time difference between $d\tau$ and dt' to add up to 1s, we need to wait $\sim 1.4 \cdot 10^9$s$\sim$ 44 years. One should expect the differences between the proper times ate various places on Earth to be smaller than that, so for practical purposes they are negligible. (There is an effect as in the twin-paradox when one person lives on the equator and the other on the North pole, say, but it is very small.)

15.4 Time measurement on a satellite

To compare time measurement on Earth with that on a satellite it is convenient to replace the time coordinate $t = t'$ by

$$t'' := \left(1 + \frac{\Phi_0}{c^2}\right) t' = \left(1 + \frac{\Phi_0}{c^2}\right) t,$$

which is the proper time measured by a clock fixed on Earth at the equator (or any other point on the geoid). The metric then takes the form

$$ds^2 = \left(1 + \frac{2(V - \Phi_0)}{c^2}\right) c^2 dt''^2 - \left(1 - \frac{2V}{c^2}\right)(dr^2 + r^2 d\theta^2 + r^2 \sin^2(\theta) d\varphi^2) \tag{74}$$

in non-rotating coordinates $(t'', r, \theta, \varphi)$.

Like any massive body, we can model the orbit of a satellite as a time-like curve $\gamma(t'') = (t'', r(t''), \theta(t''), \varphi(t''))$ with (coordinate dependent) velocity

$$v := \left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2(\theta)\dot{\varphi}^2\right)^{\frac{1}{2}},$$

where $\dot{}$ denotes a derivative w.r.t. the time coordinate t''. The proper time along γ satisfies, up to order c^{-2},

$$\begin{aligned} c^2 d\tau^2 &= \left(1+\frac{2(V-\Phi_0)}{c^2}\right)c^2dt''^2 - \left(1-\frac{2V}{c^2}\right)\frac{v^2}{c^2}c^2dt''^2 \\ &\sim \left(1+\frac{2(V-\Phi_0)}{c^2}-\frac{v^2}{c^2}\right)c^2dt''^2 \\ d\tau &\sim \left(1+\frac{V-\Phi_0}{c^2}-\frac{v^2}{2c^2}\right)dt''. \end{aligned}$$

This formula shows that the proper time along γ is affected by the gravitational field through V and by the motion through v. (The rate change of a clock due to its motion w.r.t. the coordinate system is known as the second order (relativistic) Doppler effect, because it depends in second order on $\frac{v}{c}$.)

Let us now describe the world-lines of the satellites in more detail, in order to find out the rates of their clocks in comparison to t''. The GPS satellites have orbits iwth an average altitude $r \sim 2.02 \cdot 10^7$m$\sim 3.2a$ (where a is again Earth's radius). At this altitude we can approximate $V \sim \frac{-G_N M}{r}$, because the quadrupole term falls of rapidly enough with the distance r. The satellite's motion is then accurately described by Newtonian gravity, leading to an elliptic orbit.

The distance r and the velocity v of the satellite change as it moves along its orbit. Because the elliptic orbits are simple enough, we can at least eliminate v from the problem as follows. We first note that the orbit takes place in a plane, where it can be described by the distance r and an angle ϕ, both depending on t''. Because Newtonian gravity is a conservative force, the total energy E (per unit mass of the satellite) is conserved, as is the angular momentum $L = r^2\dot{\phi}$. Adding kinetic and potential energy, E can be written as

$$E = \frac{v^2}{2} - \frac{G_N M}{r} = \frac{1}{2}\dot{r}^2 + \frac{L^2}{2r^2} - \frac{G_N M}{r}.$$

At the perigee (point farthest from Earth) and apogee (point closest to Earth), we have $\dot{r} = 0$ and $r = r_1$, respectively $r = r_0$, so that

$$E = \frac{L^2}{2r_1^2} - \frac{G_N M}{r_1} = \frac{L^2}{2r_0^2} - \frac{G_N M}{r_0}.$$

Eliminating L from these equations and using the fact that $r_0 + r + 1 = 2s$, where s is the semi-major axis of the elliptic orbit, we find

$$E = \frac{v^2}{2} - \frac{G_N M}{r} = -\frac{G_N M}{2s}.$$

This leads to

$$d\tau \sim \left(1 - \frac{\Phi_0}{c^2} - \frac{3G_N M}{2c^2 s} + \frac{2G_N M}{c^2}\left(\frac{1}{s} - \frac{1}{r}\right)\right) dt''.$$

Only the last term in brackets varies along the orbit, and it remains small when the orbit is close to being circular. The constant rate corrections

$$-\frac{\Phi_0}{c^2} - \frac{3G_N M}{2sc^2} \sim 6.97 \cdot 10^{-1} - 2.5 \cdot 10^{-10} \sim 4.46 \cdot 10^{-1}$$

can be implemented in the atomic clock before launch of the satellite (and after choosing the semi-major axis s).

Let us close this subsection with some remarks on the last result:

1. Equation (75) shows the change of rate between a clock on Earth and on a satellite, taking several relativistic effects into account, some of which have opposite effects. One way of grouping these effects is as follows:
$$\frac{V - \Phi_0}{c^2} - \frac{v^2}{2c^2} = \frac{V - V_0}{c^2} + \frac{(\omega a)^2 - v^2}{2c^2},$$
where $V_o := V|_{\text{equator}}$. Here the first term on the right-hand side describes a gravitational blue-shift effect: the term is positive, indicating that clocks in orbit beat too fast. The second term describes the second order Doppler effects due to the motion of the satellite and the rotation of Earth. The satellites move faster than the rotating Earth (they circle Earth twice a day and their orbits are longer than the circumference of Earth), so this term is negative, indicating that clocks in orbit beat too slow.

2. Another way of grouping the various effects is as follows:

$$\begin{aligned} \frac{-3G_N M}{2c^2 s} - \frac{\Phi_0}{c^2} &= -\frac{G_N M}{c^2}\left(\frac{3}{2s} - \frac{1}{a}\right) + \frac{G_N M J_2}{2c^2 a} + \frac{(\omega a)^2}{2c^2} \\ &\sim 4.45 \cdot 10^{-10} + 3.76 \cdot 10^{-13} + 1.2 \cdot 10^{-12}, \end{aligned}$$

where we have once again eliminated the velocity of the satellite in favour of its distance. Here the first term shows the effect of Earth's mass, which at our desired accuracy becomes relevant after ~ 7.5s. The second term shows the effect of Earth's shape, which becomes relevant after $\sim 8.9 \cdot 10^3$s~ 2.5hrs. The last term shows the effect of Earth's rotation, which becomes relevant after $\sim 2.8 \cdot 10^3$s~ 46min.

3. The variable rate correction $\frac{2G_N M}{c^2}\left(\frac{1}{s} - \frac{1}{r}\right)$ can add up to relevant contributions. It could be corrected by the software on the satellite before broadcasting the coordinates, but in the GPS this correction is left to the receiver.

15.5 How to synchronise clocks on different satellites.

Suppose that we have two GPS satellites in orbit, who measure proper times τ_1 and τ_2, respectively. So far we have only discussed how to adjust their clock rates to determine time differences in terms of t''. However, even after correcting the clock rate, there may still be a constant shift in the time coordinates determined by each satellite. To compensate for this shift we need to synchronise their clocks. Recall that the synchronisation of clocks which are located at different places is a non-trivial issue in relativity theory.

Let us suppose that satellite 1 measures the time t'', after correction of its clock rate. Then suppose that satellite 1 sends a signal at time t''_s, which arrives at time t''_a at satellite 2. According to the clock on satellite 2 the signal arrives at some time $\tilde{t}''_a$, which differs from t''_a by a constant, $\tilde{t}''_a - t''_a$. To synchronise the clock on satellite 2 with that on satellite 1 we need to adjust for this constant, i.e. we need to find out t''_a.

Because satellite 1 is a GPS satellite, it also broadcasts the coordinates of the event where the signal originates, so satellite 2 can act as a GPS receiver to find out the value of t''_s and the position coordinates $\mathbf{x}_s$ of the event when the signal was sent. In order to find t''_a we only need to know the distance l that the signal has travelled, because it travels at the constant speed of light c, so $l = c(t''_a - t''_s)$. When the signal arrives, satellite 2 is at position $\mathbf{x}_a$, so

$$l \sim \|\mathbf{x}_a - \mathbf{x}_s\|,$$

where we used the Minkowski metric as the lowest order in a weak field approximation to the spacetime metric. Higher order terms constribute corrections of the order c^{-1}. Thus,

$$t''_a = t''_s + c^{-1}\|\mathbf{x}_a - \mathbf{x}_s\|$$

up to order c^{-1}.

15.6 Communication with users on Earth

The satellites send out signals, which are electromagnetic waves of a certain frequency. For users on Earth it is often useful to measure this frequency, e.g. for determining velocities by the Doppler effect.

When expressed in the almost inertial coordinates (t'', r, θ, ϕ), the signal does not alter its frequency along its light-like geodesic from the satellite to the user, because the metric is static. However, we do need to take the relativistic Doppler effect into account which is caused by the motion of the satellite and the user w.r.t. the almost inertial coordinates.

The relativistic Doppler effect describes how the frequency of a signal changes under a change of (inertial) coordinate system. It can be expressed in terms of the velocity $\beta = \frac{v}{c}$ between the two coordinate systems. Making an expansion in terms of β we find no effect at order zero. The classical effect appears at first order and the relativistic effects occur at order two or higher. For our purposes we can distinguish:

1. A transversal effect:
   (The term transversal requires the choice of a coordinate system.) This effect is of second order β, so it is absent in the classical Doppler effect. In our case this effect has been accounted for already in the rate change of the satellite's clock (by the $-\frac{v^2}{2c^2}$-term).

2. A longitudinal effect:
   At first order in β this is the classical Doppler effect. Including its relativistic corrections, it states that the frequency changes according to

   $$f = f'\sqrt{\frac{1-\beta}{1+\beta}} \simeq f'(1 - \beta + \ldots),$$

   where f and f' are the frequencies in the two relevant coordinate systems and β is the longitudinal component of $\frac{v}{c}$. This correction has to be applied for the change of coordinates at the satellite (from inertial coordinates in which the satellite is at rest, to the given coordinates) and at the user. The received frequency f_R is then related to the broadcast frequency f_0 by

$$f_R \simeq f_0 \frac{1 - c^{-1}\mathbf{N} \cdot \mathbf{v}_R}{1 - c^{-1}\mathbf{N} \cdot \mathbf{v}},$$

where $\mathbf{N}$ denotes the direction of the signal's light-like geodesic, $\mathbf{v}_R$ is the velocity of the receiver and $\mathbf{v}$ that of the satellite.

These motion dependent effects are of order 10^{-5}, but they cannot be corrected in advance. Instead, by measuring the received frequency f_R, the user can reconstruct his velocity $\mathbf{v}_R$ using the formulae for the Doppler effect.

15.7 Conclusions

All in all, GPS consists of three so-called "segments":

1. Control segment:
   This consists of a number of monitoring stations, which gather information from the satellites, compute their orbits and (position dependent) frequency corrections for the next few hours. This information is then uploaded to the satellites, to pass it on to the users. (Additional information which is monitored and passed on includes e.g. the "weather" in the ionosphere, which can affect the speed of light in that part of Earth's atmosphere significantly.)

2. Space segment:
   This consists of 24 satellites, carrying atomic clocks, with additional spare clocks and spare satellites. The satellites trasmit (a) timing signals, and (b) corresponding messages, specifying the coordinates of the timing signal's source event, as well as additional data needed to determine event coordinates.

3. User segment:
   This consists of all users which receive the satellite signals and use them to determine their position, time and velocity. Here we can distinguish two kinds of users: the commercial users and the U.S. military. Whereas the military receives the satellite signals on a restricted radio frequency, other users receive it on a frequency which can be received by

commercially available receivers. However, the signal which is broadcast at the latter frequency is first distorted by small random noise, to reduce the accuracy. This is to prevent the use of GPS for unwanted military purposes by others than the U.S. military.

References

[1] N. Ashby, *Relativity in the Global Positioning System*, Living Rev. Relativity **6**, (2003), 1

[2] S. Carroll, *Spacetime and Geometry*, Addison-Wesley (2003)

[3] A. Einstein, *Relativity*, (Routledge Classics) Taylor and Francis, Hoboken (2012)

[4] A. Einstein, *Über spezielle und allgemeine Relativitätstheorie*, Akad.-Verlag, Berlin (1969)

[5] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces*, AMS (2001)

[6] E. Herlt, N. Salié, *Spezielle Relativitätstheorie*, (Wissenschaftliche Taschenbuecher) Akademie-Verlag, Berlin (1978)

[7] S. Hollands and R.M. Wald, *Stability of Black Holes and Black Branes*, Commun. Math. Phys. **321**, no. 3, 629–680

[8] H. Ringstrom, *The Cauchy Problem in General Relativity*, EMS/ESI Lecture Notes in Mathematical Physics (2009)

[9] R. M. Wald, *General Relativity*, University of Chicago Press, Chicago (1984)

[10] N.M.J. Woodhouse, *Special Relativity*, Springer, Berlin etc. (1992)