UNITEXT for Physics

Wladimir-Georges Boskoff
Salvatore Capozziello

A Mathematical Journey to Relativity

Deriving Special and General Relativity with Basic Mathematics

Second Edition

Springer

UNITEXT for Physics

UNITEXT for Physics series publishes textbooks in physics and astronomy, characterized by a didactic style and comprehensiveness. The books are addressed to upper-undergraduate and graduate students, but also to scientists and researchers as important resources for their education, knowledge, and teaching.

Wladimir-Georges Boskoff · Salvatore Capozziello

A Mathematical Journey to Relativity

Deriving Special and General Relativity with Basic Mathematics

Second Edition

Springer

Wladimir-Georges Boskoff
Faculty of Mathematics and Informatics
Ovidius University
Constanţa, Romania

Salvatore Capozziello
Complesso Universitario Monte Sant'Angelo
University of Naples Federico II
Naples, Italy

ISSN 2198-7882 ISSN 2198-7890 (electronic)
UNITEXT for Physics
ISBN 978-3-031-54822-2 ISBN 978-3-031-54823-9 (eBook)
https://doi.org/10.1007/978-3-031-54823-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Wir müssen wissen, Wir werden wissen.

—David Hilbert

Dedicated to all those who do not like Mathematics with the hope that they will change their mind.

Preface to the Second Edition

Mathematics is the science which studies structures and relations among structures. The "truth" (whatever it means) in Mathematics is established by rigorous deductions made in an appropriately chosen frame described by axioms and definitions. Self-consistency is the first requirement which any mathematical result should satisfy. Sometimes, in Mathematics, theorems are formulated about space and changes of structures involving the space which becomes a language for models in Physics. The language of Relativity is the Differential Geometry, and our road to Relativity depends on it. In other words, while Geometry is often used to formulate theories of Physics, in General Relativity, Geometry coincides with Physics and a modern trend is that all Physics could coincide with Geometry. The aim of this Second Edition is to improve both the content and the style we present in this language. Therefore the two chapters dedicated to Differential Geometry are completely changed. We insist on the natural transfer of geometric structures between the ambient space and surfaces belonging to it. We present more examples and simplify the proofs of theorems. More exercises are solved considering students who need to understand the basic techniques of Differential Geometry in view of applications to Physics. The aim is taking into account the Einstein point of view: "As far as the laws of Mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality." Therefore the reader can realize that modeling Physics does not mean to describe the Reality in its deep meaning but only represent some features of Reality which can be studied by the self-consistency of the models as well as probed by experiments and observations validating or discarding such features.

The road passing through Differential Geometry allows us to step to Relativity after we proved the Bianchi second formula. It is why we succeeded in less than 80 pages to develop a chapter called "Differential Geometry at Work. Two Ways of Thinking the Gravity".... We find out how to simply deduce the Einstein field equations.

At the same time, the road we took allows to understand, in the next chapter, how models of Euclidean, Non-Euclidean and Elliptic geometries can be deduced from Geometry and Physics points of view. Yes, you read correctly, from Physics, the Physics related, for example, to Special Relativity.

There are changes inside almost each chapter of the First Edition, because we think to offer more information about FLRW universes and the dimensions of our observable Universe. Also if it is hard to believe, it is possible to prove that FLRW universes depend on the geometries of Gaussian constant curvature. Therefore such universes are created starting two dimensional metrics which describe Euclidean, Non-Euclidean and Elliptic geometries. In particular, we discuss some special exact solutions of Einstein field equations describing exotic space-times without matter, time and having the "geometric texture" shacked by planar gravitational waves. Finally, without claim of completeness, we give simple examples of black holes, wormholes or cosmic strings with the aim to show how Mathematics can be powerful in imagining the World. If in the First Edition we have presented the de Sitter universe, now we enlarge the approach presenting universes without matter and time. A detailed discussion is given for the Anti-de Sitter universe. We explained how the cosmological constant of de Sitter and Anti-de Sitter space-times depends on the affine radius of the Minkowski spheres which appear in the landscape description.

In order to improve the Differential Geometry tools, we presented differentiable manifolds, differential forms, structure equations, vector fields, affine connections, covariant derivatives induced by affine connections, torsion, curvature, parallel transport and geodesics. The aim is presenting concepts of Differential Geometry which can exist even without a metric. Setting a metric compatible with a connection means to consider only a specific connection, the so called Levi Civita connection.

In some sense, we return to the Differential Geometry presented in the previous chapters of the book, but we open the road to the Metric-Affine Gravity. An entire chapter is devoted to this topic and an entire section prepares the basic notations and notions necessary to understand it.

In this second edition, we added more figures and took care of observations and suggestions received from students and colleagues, trying to write an improved text suitable for all the audience.

Constanţa, Romania — Wladimir-Georges Boskoff
Naples, Italy — Salvatore Capozziello

Preface I to the First Edition

This book is an approach to Special and General Relativity from a full mathematical point of view. When Physics is studied, there is the need of understanding its language, that is Mathematics. Dirac's words describe very well what we want to do: *"God used beautiful Mathematics in creating the world,"* therefore we present a part of this *divine plan*, the beautiful Mathematics of Special and General Relativity. We wrote a textbook which, we believe, can be easily used by students in Mathematics, Physics, and Engineering studies. By teachers or by some other people who are interested in this subject. If someone already knows Mathematics, that is both basic Geometry and Differential Geometry, this person can neglect the first six chapters. She/he can start from Gravity in Newtonian Mechanics. People who study Physics should start from the very beginning in order to understand the development of Geometry. The improvement of mathematical language, in more than 2000 and 500 years, allowed to produce a common language for both Calculus and Linear Algebra: this approach ends up to a *dialect*, the Differential Geometry, which constitutes the basic tool of Relativity. Without the effort to understand the nature of the Non-Euclidean Geometry, the Differential Geometry could not occur. Without Differential Geometry, General Relativity could not exist. The first six chapters represent the adventure of Geometry from axioms until the Non-Euclidean Geometries through Differential Geometry. A lot of examples and solved exercises help the reader to understand the theory. Actually, the entire book, which is written in a unitary way, offers clear statements and proofs. About the proofs: it offers complete proofs and all computations are presented. In our opinion, this is the only way to understand the complicated computations which depend on the Differential Geometry language. Reading line by line, the reader can understand every single proof. The references which inspired us are mentioned at the beginning of each chapter, but also in the text. Some proofs and some approaches of the theory are completely original. If somebody is reading from the beginning to the end of this book, it becomes understandable why each subject presented is important for the topic. We hope that our humble efforts are useful, first of all, for learning people who this book is mainly dedicated. We thank our colleagues, our teachers, our friends and, first of all, our students whose questions, discussions, and remarks allowed us to enter the *perfect world* of Geometry towards its amazing realization which is General Relativity. We want to thank also Dr. Marina Forlizzi and the Springer staff for invaluable support in publishing this book.

As a final remark, we want to say that this book was conceived about 2 years ago during pleasant discussions on Mathematics and General Relativity in scientific congresses and meetings between the authors and was concluded during the severe period of the global Coronavirus disease. We hope that Science and its high values can be comforting even in difficult situations like the present one, as happened so many times in history.

Constanţa, Romania — Wladimir-Georges Boskoff
Naples, Italy — Salvatore Capozziello
March 2020

Preface II to the First Edition

What does a mathematical journey towards the general theory of Relativity look like? The authors propose an original itinerary moving from Euclidean and non-Euclidean Geometry created from axioms, to models of geometric Euclidean and non-Euclidean worlds. Differential Geometry of surfaces and then abstract Differential Geometry are special stops for two reasons:

1. To understand non-Euclidean Geometry models from this point of view;
2. To create the language by which we can describe the General Relativity and its consequences.

The physical world allows both Euclidean and non-Euclidean descriptions. To have an image of this physical world we need to continue the itinerary with supplementary stops: Newtonian and Lagrangian Mechanics, Special Relativity to reach, finally, General Relativity.

The content of the book is written to be self-contained. All the proofs are done with all the details presented for the reader. The problems are solved, or they have hints. Almost all the contents were presented to students at different university courses and, in our experience, they were well received.

In Chap. 1, we present, using a slightly modified Hilbert's axioms system, Euclidean and non-Euclidean geometries and what they mean. Here, the mathematical theory is built from a set of primary objects, which do not require definitions, together with a set of axioms. The collection of primary objects is chosen from the set theory. The axioms are stated in a formal form and the axiomatic theory is built as a collection of mathematically rigorous statements deduced from the axioms.

It exists a common part for Euclidean and Non-Euclidean Geometry, the so-called Absolute Geometry. Absolute Geometry consists in all the results that can be thought and proved using the axiomatic system before introducing a Parallelism Axiom. The main theorem in Absolute Geometry is the Legendre one, which states that the sum of measures of angles of a given triangle is less than or equal to two right angles.

The two consequences are:

1. The sum of angles in each triangle is equal to two right angles.
2. The sum of angles in each triangle is strictly less than two right angles.

A further axiom, the Parallelism Axiom, makes us to discover the Euclidean world, corresponding to the first case, i.e. the sum of angles is equal to two right angles.

The denial of the previous Parallelism Axiom leads us to the Non-Euclidean Geometry; here the sum of angles is strictly less than two right angles.

Euclidean Geometry and Non-Euclidean geometries are the frameworks where formulate Newtonian Mechanics and Relativity respectively, as we will see later.

Chapter 2 highlights how the Euclidean Geometry, previously introduced, can be constructed and viewed using algebra and trigonometry. All happens in a two-dimensional vector space endowed with an inner product invariant with respect to the group of Euclidean rotations. Basic facts on Euclidean Geometry are presented, the most important being Pythagoras Theorem and the generalized Pythagoras theorems.

Even if it seems there is no connection between Minkowski Plane Geometry and the geometries created from the axiomatic point of view, we present in detail the Minkowski Plane Geometry. The construction is related to the same two-dimensional vector space used to describe Euclidean Geometry, but instead of the Euclidean inner product, we have a Minkowski product. There exists a group of hyperbolic rotations which leaves invariant the Minkowski product. Minkowski Geometry is not as simple as Euclidean Geometry. There are space-like vectors, null vectors, and time-like vectors. Minkowski–Pythagoras Theorem has different statements with respect to the type of the involved side-vectors of the triangle. Even if in this chapter we construct an algebraic image of the Euclidean Geometry, we have not yet images about the Non-Euclidean Geometry. The next chapters deal with this issue.

Chapter 3 is dedicated to the tools we need to construct a first model of Non-Euclidean Geometry. This model is constructed in the interior disk of a given circle.

To be constructed, we need to understand the geometric inversion and basic facts about Projective Geometry. A projective invariant of a special projective map, attached to the previous given circle, allows us to construct a distance inside the disk of the circle. The Poincaré disk model is highlighted.

The “lines”, that is the geodesic lines of this distance, are orthogonal arcs of circles to the given circle. It is easy to see that there are more than two non-intersecting “lines” through a given point with respect to a given “line”. The sum of angles of a triangle in this Poincaré model is of course less than π.

Chapter 4 is related to the Differential Geometry of surfaces. In the first part, the surfaces are seen as subsets of a 3-dimensional Euclidean space. In this context, we understand how the Euclidean inner product of the Euclidean space induces a way to measure lengths and angles for vectors belonging to the tangent planes of the surface. We can also measure the length of curves that belong to surfaces, areas of regions, and all these using the metric attached to the surface.

The Differential Geometry of a surface continues by introducing a fundamental notion: the Gaussian Curvature of a surface at a point. If, at the beginning, the Gaussian Curvature seems to be dependent on the fact that the surface is seen in the

Euclidean ambient space, after we prove Gauss' formulas, we step into the intrinsic theory of surfaces. There, Gauss' equations and *Theorema Egregium* offer another perspective: each surface can be seen as a piece of a plane endowed with a metric, and this metric only determines the curvature.

The study continues in Minkowski three-spaces where we have to take care of the Minkowski-type nature of the normal vector to the surface. However, we have almost the same picture, the Minkowski product determines a non-Euclidean metric of a surface which allows us to conclude about the intrinsic Geometry of it. Therefore, in both cases the surface becomes irrelevant for our study. In fact, we study the Geometry of a metric and obtain relevant geometric aspects about the piece of plane endowed with that metric.

The covariant derivative introduced in the last part of the chapter allows us to define the parallel transport and the geodesics of surfaces. At the end of the chapter we introduce a short story about a person embedded in a surface with the aim to reveal how the person can develop a theory about its universe, which is the surface where he lives. The study is continuing in the next chapter, when we better understand the nature of geometric objects which appear in Differential Geometry.

Chapter 5 is fundamental for the book: the final image about the Non-Euclidean Geometries cannot be given without what we learn here. Basic Differential Geometry is about Differential Geometry when an extra dimension does not exist. In the previous chapter, we claimed that we did not know yet the mathematical nature of the multi-index quantities which appear in the Differential Geometry of surfaces. In this chapter we prove the tensor character of the metric coefficients, of the Riemann symbols, of the Ricci symbols, and also for the geodesics equations. All these multi-index quantities remain invariant when we deal with a change of coordinates. Why this is important? The substance of General Relativity is related to these changes of coordinates. A change of coordinates may reflect an acceleration field which is equivalent to a gravitational field and, in the context described, a nice example developed later is about the constant gravitational field.

The covariant derivative for contravariant vectors, which appears as a geometric property, allows us to think of a general definition for the covariant derivative of tensors. How we parallel transport vectors along curves, how geodesic lines appear, and some other important properties of parallel transport of vectors along geodesic lines are also studied. At the end of the chapter, the covariant derivation of Einstein's tensor allows the reader to have a first image on Einstein's field equations.

Chapter 6 is devoted to Non-Euclidean Geometry models and their physical interpretation. It is worth stressing that we are dealing with *models* and not only with *a model*. In fact, we imagine some other models of Non-Euclidean Geometry. There are some steps before providing these models. Differential Geometry gives us the possibility to see:

- how distances of the models produce the metrics;
- how the geodesics of the distances are also the geodesics of the metrics;
- how the models are related among them through their metrics according to convenient changes of coordinates.

It is important to stress that all these models are equivalent and contribute to the big picture. Specifically, the Poincaré disk model, the Poincaré half-plane model, the exterior disk model, the hemisphere model, and the hyperboloid models are studied and presented.

The first three models are connected among them by geometric inversions. The remaining models need appropriate changes of coordinates to connect them to the first three. In particular, the hyperboloid model is described by a Minkowski metric. At this point, we have the first connection between Non-Euclidean Geometry and Minkowski Geometry. Next, in the Eighth Chapter devoted to Special Relativity, a Minkowski-type metric appears giving a geometric image of the (so called) *physical reality* [1].

The physical example, developed at the end of the chapter, is due to Poincaré. The question is: Can we develop the Poincaré disk model starting from simple physical rules? The answer is yes and this is possible combining Physics and Geometry. Even if Poincaré developed the model stating that reality cannot be fully understood, after this example, it is easy to accept the fact that the Geometry is related to the Relativity description. In fact Non-Euclidean Geometry, seen through Differential Geometry, is needed to understand basic facts of General Relativity, as we see later.

Euclidean Geometry constitutes the framework of the Newtonian Mechanics. Chapter 7 is dedicated to understand how forces can explain what is happening in our surrounding world modelled into a three-dimensional Euclidean Space where only one clock gives the universal time. In this sense, the Newtonian Mechanics reveals an absolute space and an absolute time [2].

It is described the gravitational force together with the gravitational field. A mathematical artefact, the gravitational potential, is involved in two fundamental results: the vacuum field equation and the gravitational field equation. Looking at these equations and how difficult we mathematically obtained them, somebody can think that this is the maximum we can say about the gravitational field. But the tidal forces and the tidal acceleration equations offer another perspective. The vacuum field equation is encapsulated in the trace of the Hessian matrix involved in tidal acceleration equations.

If we try to obtain the geometric equivalent of these equations in a curved space, that is if we cancel out the Euclidean 3-dimensional space, the Hessian matrix is replaced by a curvature-dependent tensor whose trace is the Ricci symbol. In the future, we prove via Fermi coordinates, that this is a possible way to obtain Einstein's vacuum field equations. This is the first geometric change of the Euclidean frame when one studies forces.

This situation may arise another important change of perspective.

Suppose we have a force and the trajectories of a point subjected to this force. Is it possible to locally find a metric whose geodesics are the previous trajectories? The answer is yes. The Euler–Lagrange equations become the equations of the previous trajectories and the same Euler–Lagrange equations are the geodesic equations of a metric induced by another *mathematical artefact* called Lagrangian.

The study of the Lagrangian, starting from the mechanical one, is made through a functional called action. If the first-order variation of the action vanishes, we obtain the Euler–Lagrange equations. Later, we prove how another action, the Hilbert action, allows us to derive the Einstein field equations in General Relativity.

Kepler's laws are also studied in this chapter with the aim to prepare the reader understanding of planet trajectories in a metric, of course, later, in the part related to General Relativity.

Chapter 8 is devoted to Special Relativity. Reflection and refraction of light were explained in a satisfactory way by Newton who looks at light rays as trajectories of particles (after called photons).

In the middle of *XIXth* century, James Clark Maxwell offered another view: the light is an electromagnetic wave and it satisfies four equations, known as Maxwell's equations of the Electromagnetism. If we try to put them in accordance with Newton's theory, it appears the necessity of considering a medium in which the electromagnetic waves travel through space. It was called by physicists of that time the *ether*.

Ernst Mach's did not agree with the idea of ether and observed the necessity of the revision of all fundamental concepts of Physics [2].

Michelson–Morley experiment, who initially was designed to reveal the ether, had a result completely different with respect to the expectations. Albert Einstein explained the result of the experiment in a theory where he revised in a fundamental way the ideas of space and time, and no place for ether remained. The absolute space and the absolute time of the Newtonian mechanics were replaced by the specific space and the specific time of each observer. Different observers mean different inertial frames of coordinates, each one having its time axis and its space axes [3].

Einstein formulated the Special Relativity starting from two main postulates:

1. The laws of Physics has to be the same in all inertial reference frames.
2. The speed of light in vacuum, denoted by c, is the same for all the observers and it is the maximum speed reached by a moving object.

Presenting the theory, we preferred to balance it starting from two important works, the book by Callahan, [4], and a paper by Varićak [5], where we found the most possible geometric approach to Special Relativity.

Therefore we have adapted, in a new form, the basic ideas discussed there. The first idea is related to the consideration of two inertial frames of coordinates, one moving at a constant speed, another considered at rest, in which the two observers have to agree on the same laws of Physics [4]. In this way, the old Galilean transformations of coordinates are replaced by the Lorentz transformations. There are a lot of consequences: another formula for velocity addition, the time dilation, the length contraction, the covariance of Maxwell's law under Lorentz transformation, the rest-energy formula, the Doppler effect, and so on. The second one is related to the geometric understanding of these facts: we have a sort of equivalence between the so-called "geometric coordinates" and the "physical coordinates" [5]. The entity called "physical space-time" is understood through the geometric space-time where the results are easier to be viewed. This idea can be originally seen in [3].

Then, when we introduce a constant gravitational field via the accelerated frames, (see also [4] point of view) we can prove the bending of the light rays; the interpretation of the Doppler gravitational effect shows that accelerated frames are not inertial ones. Much more: a contradiction between the Minkowski flat space-time of Special Relativity and the gravitational Doppler effect occurs. A physical theory containing the old Mechanics, including gravity, and the modern electromagnetic waves theory needs to integrate the accelerated frames. In this way, we can add another argument towards the General Relativity.

Chapter 9 is devoted to General Relativity and Relativistic Cosmology. There is no other better description of the subject than the sentence by John Archibald Wheeler: *"Space-time tells matter how to move; matter tells space-time how to curve"* [6], [7]. How the space is curved appears from the Einstein field equations

$$R_{ij} - \frac{1}{2} R\, g_{ij} = \frac{8\pi G}{c^4} T_{ij}.$$

In the left-hand side we have "Geometry," a metric g_{ij} and its derivatives are involved; in the right-hand side we have a tensor depending on matter, the so-called energy-momentum tensor. The energy-momentum tensor establishes the metric, the metric produces geodesics described by the equations

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq} \frac{dx^p}{dt} \frac{dx^q}{dt} = 0.$$

They are trajectories of objects moving according to the Geometry of space-time. Therefore, the geodesic equations are the equivalent equations of curves which satisfy $\overrightarrow{F} = m\overrightarrow{a}$ from Mechanics.

The equations of geodesics of an initial metric switch accordingly to a change of coordinates into the equations of the geodesics of the new obtained metric. Changes of coordinates may provide a new state of a given frame, therefore the new state is described by a new metric provided by the old state metric via the coordinates change. The reader will understand how it works looking at the case of the constant gravitational field. However, after a change of coordinates, the metric of space-time remains the same being a tensor.

The chapter starts looking at the differences between the classical Newtonian Mechanics and Einstein's landscape of gravity described by Geometry. Einstein's field equations can be derived from the Hilbert action. A generalization of such an action, the so-called $f(R)$ gravity is also presented. The straightforward solution, in the case of vacuum field equations for spherical symmetry, is Schwarzschild's one. We present computations related to the orbits of planets and the bending of light rays.

Fermi's viewpoint on Einstein's vacuum field equations allows to obtain the classical counterparts of the relativistic equations in the case of the weak gravitational field.

After, we analyse the Einstein static universe and the basic considerations on the cosmological constant, as a part of the classical approach to the General Relativity.

A "metric for the Universe" is obtained for the cosmic expansion. It is the Friedman–Lemaître–Robertson–Walker metric. The way we obtain it is related to the energy-momentum tensor and the Cosmological Principle.

Black holes are an important prediction of General Relativity. We propose an introduction to their theory starting from the Rindler metric. The singularities which can be removed using Rindler's idea are geometric only. Schwarzschild metric is important in the study of black holes. The anomalies of black holes are explained via Kruskal–Szekeres coordinates and light cones inside and outside black holes are presented.

Another important prediction, the existence of gravitational waves, is discussed in this chapter to give a more complete picture of the physical landscape of General Relativity. Furthermore, cosmic strings are presented as a hypothetical structure considered, until now, only as a possible solution to Einstein's field equations. Another important solution of Einstein's field equations is the Gödel's, who succeeded to prove that a homogeneous universe without a global time coordinate can theoretically exist. We present the above solutions with all the details necessary to be easily understood at the undergraduate student level.

In Chap. 10, as a full geometric realization of Relativity, we present the so-called Affine Universe and the de Sitter space-time. From a cosmological point of view, this solution is fundamental to discuss, at an elementary level, the problems of primordial inflation and the late accelerated behaviour, often dubbed as dark energy. Starting from two different parameterizations, it is possible to describe the cosmological constant, the main ingredient of de Sitter solution. Essentially, it is possible to show that a curved universe can be achieved without a mass distribution. A possible explanation can be obtained starting from a Minkowski space-time where gravitational field (without masses) is considered. In this sense, this is a full geometric realization of Relativity.

Constanţa, Romania Wladimir-Georges Boskoff
Naples, Italy Salvatore Capozziello

Contents

Chapter 1
Euclidean and Non-Euclidean Geometries: How They Appear

Omnibus ex nihilo ducendis sufficit unum.

G. W. von Leibniz

We intend to construct these geometries using a slightly modified Hilbert's axioms system in the same way as it is done in [36, 37, 138, 185]. *An interesting thing is related to the fact that it exists as a common part for Euclidean and non-Euclidean Geometry, the so-called Absolute Geometry. Roughly speaking, the Absolute Geometry consists in all theorems that can be thought and proved using the axiomatic system before introducing a parallelism axiom.*

In our vision, the most important theorem in Absolute Geometry is the Legendre one:
"The sum of angles of a triangle is less than or equal two right angles."
It allows us to prove that only two situations hold:
"The sum of angles in each triangle is equal to two right angles."
or, the other situation:
"The sum of angles in each triangle is strictly less than two right angles"

Choosing an appropriate parallelism axiom we discover the Euclidean world, corresponding to the first case, i.e. the sum of angles is equal to two right angles. The denial of the previous parallelism axiom leads us to the non-Euclidean geometry; here the sum of angles is strictly less than two right angles. We have used few figures to illustrate these concepts, because the reader can remain with a false image about how lines look like. However, in Absolute Geometry, the reader can think and draw images as in the Euclidean geometry, because all the objects and all the theorems valid in absolute geometry are also valid in Euclidean Geometry. Here the lines are the ordinary straight lines of the plane. The images can be thought in a more complicated way if someone try to imagine them in a model for the non-Euclidean

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_1

geometry, because lines can look like arcs of circles, segments, etc. All the proofs are reported in such a way that they can be understood by reading them directly.

1.1 Absolute Geometry

The key idea of the axiomatic method is to build a theory from a set of primary objects, which do not require definitions, together with a set of axioms. Therefore the theory is built as a collection of mathematically rigorous statements deduced from the axioms.

The collection of primary objects of the geometry is given in the following, inherited from set theory. The objects of the first collection are called *points*, and they are denoted by capital letters A, B, C, ... The second collection contains *lines* , denoted by l, l', ... The third collection contains *planes*, denoted by Greek letters α, β, γ, ... Finally, the last collection contains only one element called *space,* denoted by S.

The first important group of axioms is related to the incidence of the objects described above. They describe who belongs to who, which set of objects can be included in which, how many objects are necessary to create another object, etc. Let us introduce the so-called "axioms of incidence".

The first axiom which helps us to construct a geometry establishes the existence and uniqueness of a line and its connection with two given distinct points.

Axiom $\mathbf{I_1}$: *For any two distinct points A and B there exists a unique line l which is incident with both A and B, i.e. $A \in l$ and $B \in l$.*

The unique line l of the previous axiom is often denoted by AB, indicating that it is the line that passes through the points A and B.

Axiom $\mathbf{I_2}$: *There exist at least two distinct points on any line. Moreover, there exist at least three distinct points which are not on the same line.*

In view of the axiom, it seems useful to be able to distinguish between points which are on a line from points which do not belong to the same line, therefore we introduce the following notion.

Definition 1.1.1 Any number of points are called *collinear* if there is a line which is incident to all of them. Otherwise, they are called *non-collinear* .

For example, axiom $\mathbf{I_1}$ asserts that every two distinct points are collinear, and axiom $\mathbf{I_2}$ guarantees the existence of at least three non-collinear points in the geometry we are constructing. The next two axioms establish the relationship between points and planes.

Axiom $\mathbf{I_3}$: *For any three arbitrary non-collinear distinct points A, B and C, there exists an unique plane α which contains A, B, and C.*

In general, such a plane is denoted by $\alpha := (ABC)$.

The following axiom establishes the relationship among points on a given line and a plane containing that line. This axiom plays a crucial role once we construct geometries with more number of points and lines.

Axiom $\mathbf{I_4}$ *: If two points* A *and* B*, which determine the line* l*, lie in the plane* α*, then every point of the line* l *lies in the plane* α*.*

In this case, we write $l \subset \alpha$ (regarded as a subset of points). The following axiom states that the minimum number of points in an intersection of two planes is two.

Axiom $\mathbf{I_5}$ *: If two planes* α *and* β *have a common point* A*, then they have another common point* B *distinct from* A*.*

An immediate consequence of axioms $\mathbf{I_4}$ and $\mathbf{I_5}$ is that if the planes α and β contain the two distinct points A and B, then they contain the whole line $l = AB$, and we write $\alpha \cap \beta = \{l\}$, again as an equality of sets of points.

The last axiom of incidence states the minimum number of points necessary to create the space.

Axiom $\mathbf{I_6}$ *: There exist at least four points which do not belong to the same plane.*

In the view of this last axiom $\mathbf{I_6}$, we give the following.

Definition 1.1.2 Any number of points are called *coplanar* if there is a plane which passes through all of them. Otherwise, they are called *non-coplanar*.

Axioms $\mathbf{I_1} - \mathbf{I_6}$ give rise to a simple model of a space created only with 4 points, 6 lines, and 4 planes.

The model described above can be written as follows. The distinct points are A, B, C, D, and the six lines are given by the following sets of points: $l_{AB} = \{A, B\}$, $l_{AC} = \{A, C\}$, $l_{BC} = \{B, C\}$, $l_{BD} = \{B, D\}$, $l_{CD} = \{C, D\}$, and $l_{AD} = \{A, D\}$. The four planes are $(ABC) = \{A, B, C\}$, $(ABD) = \{A, B, D\}$, $(ACD) = \{A, C, D\}$, $(BCD) = \{B, C, D\}$, and the space built by Axiom $\mathbf{I_6}$ is by definition $(ABCD)$.

We study below some immediate consequences of the group of six axioms of incidence. Notice that the results we prove below make sense even when applied to the simple model described above.

Theorem 1.1.3 *Two distinct lines have at most one common point.*

Proof Let l_1, l_2 be two distinct lines. We distinguish the following two cases. If $l_1 \cap l_2 = \emptyset$, then they have no point in common, therefore the conclusion of the theorem is true.

If $l_1 \cap l_2 \neq \emptyset$, then let $A \in l_1 \cap l_2$ be a point in their intersection. We assume, by contradiction, that there is another point $B \in l_1 \cap l_2$, $B \neq A$. In particular, $A, B \in l_1$, therefore $l_1 = AB$ (axiom $\mathbf{I}_1$). Similarly, $A, B \in l_2$, therefore $l_2 = AB$. Axiom $\mathbf{I_1}$ says then that $AB = l_1 = l_2$, in contradiction with the hypothesis that $l_1 \neq l_2$. Therefore, our assumption on the existence of a different point $B \in l_1 \cap l_2$ is false. In conclusion, A is the only common point of the two lines l_1 and l_2. □

The previous theorem motivates the following:

Definition 1.1.4 Two distinct lines that intersect in exactly one point are called *secant* lines.

The "Axioms of Order" deal with the undefined yet relation of *betweenness*, i.e. of a point lying between two other points. Once the axioms of order appear, the previous very simple model of geometry fails to exist. The axioms of order are formulated as follows.

Axiom $\mathbf{O_1}$ *: If a point B is between A and C, then A, B, C are three distinct collinear points on a line l, and B is between C and A.*

Imagine the line as a circle. The previous axiom tells us that such an image is not possible. The line l has no predefined "orientation". The only correct concept of order among points is defined to be "between".

Axiom $\mathbf{O_2}$ *: For every pair of distinct points A and B, there is at least another distinct point C such that B is between A and C.*

An immediate consequence of axiom $\mathbf{O_2}$, combined with axiom $\mathbf{I_2}$, is that a line contains at least three points. The axiom can be applied again to the pair $\{A, C\}$, so there exists another point D such that C is between A and D, etc.

Axiom $\mathbf{O_3}$: *Given three arbitrary points on a line, at most one of them is between the other two.*

Notice that the axiom $\mathbf{O_2}$ does not guarantee the existence of a point B between two given ones A and C. This will be proven below. Nevertheless, if we assume that there exists B between A and C, then the axiom $\mathbf{O}_3$ guarantees that A cannot be between B and C, and C cannot be between A and B. Another theorem will clarify the situation of three given points on a line.

Axiom $\mathbf{O_4}$ *(Pasch): Let A, B, C be three non-collinear points, and l a line situated in the plane (ABC) which does not pass through any of the points A, B, C. If the line l contains a point which is between A and B, then the line l contains either a point between A and C or a point between B and C.*

We denote by $\overline{ABC}$ when B is on the line AC and B is between A and C, and we will refer to it as the *order* ABC. Note that by axiom $\mathbf{O}_1$, the order $\overline{ABC}$ is the same as the order $\overline{CBA}$.

An immediate consequence of the axioms of order is the following

Theorem 1.1.5 *Given two points A and B on a line l, there is a point $M \in l$ such that we have the order $\overline{AMB}$.*

Proof There exists a point C not on the line AB (axiom $\mathbf{I_2}$). Then there exists a point D such that we have the order $\overline{ACD}$ (axiom $\mathbf{O_2}$).

Similarly, there exists the point E with respect to the order $\overline{DBE}$ (axiom $\mathbf{O_2}$). Then we apply axiom $\mathbf{O_4}$ for the points C, D, E and the line AB, so there exists a point M on the line AB such that we have order $\overline{AMB}$. □

The previous theorem suggests the following.

Definition 1.1.6 The set of points M on the line AB with the property that M is between A and B is called a *segment*, and it is denoted by $[AB]$.

Formally we can write

$$[AB] = \{M \in AB \mid \overline{AMB}\} \cup \{A, B\}.$$

The *interior* of the segment $[AB]$ is defined to be the set $[AB] - \{A, B\}$.

Note that the segment $[AB]$, seen as a set of points, is equal to the segment $[BA]$. Moreover, the order $\overline{AMB}$ is equivalent to $M \in [AB] - \{A, B\}$, so the previous theorem can be reformulated as follows: *the interior of every segment is non-empty*. We have also $[AA] = \{A\}$. Moreover, we can define now one of the most important object of any geometry: the triangle.

Definition 1.1.7 A configuration of three distinct non-collinear points A, B, C is called a *triangle*, and it is denoted by ΔABC. Moreover, the points A, B, C are called the *vertexes* of the triangle, and the segments determined by each pair of two vertexes are called the *sides* of the triangle.

The next theorem guarantees the existence and uniqueness of ordering for three collinear points.

Theorem 1.1.8 *Let A, B, C be three points on a line l. Then one and only one of the orders $\overline{ABC}$, $\overline{ACB}$, or $\overline{BAC}$ occurs.*

Proof We assume that we have neither the order $\overline{ACB}$, nor the order $\overline{BAC}$, and we prove that we must have the order $\overline{ABC}$. In our Euclidean intuition, we will prove that if B is not "to the left" of A and not "to the right" of C, then it must be between A and C.

There exists a point $D \notin AC$ (axiom $\mathbf{I_2}$). Then there exists a point $E \in DB$ with the order $\overline{EDB}$ (axiom $\mathbf{O_2}$). Looking at the triangle ΔBEC and the secant line AD, then there is a point F at the intersection of AD and EC, such that we have the order $\overline{EFC}$ (axiom $\mathbf{O_4}$). In the same way, there exists the point $\{G\} = CD \cap AE$, such that we have the order $\overline{AGE}$. The line CG is a secant line for the triangle ΔAEF, as we have the order $\overline{ADF}$. Moreover, considering the triangle ΔAFC and the secant line DE, it follows the order $\overline{ABC}$. □

The following theorems establish incidence relations between a line and a triangle. Historically they are attributed to Moritz Pasch, whose influential works have been one century ago in the centre of attention of many authors interested in foundations of geometry.

Theorem 1.1.9 *If a line l does not intersect two sides of a triangle ΔABC, then it cannot intersect the third one, either.*

Proof Without loss of generality, we can assume l does not intersect neither $[AC]$ nor $[BC]$. By contradiction, we assume l intersects $[AB]$, so l contains a point between A and B. Then the axiom $\mathbf{O_4}$ affirms that l must contain either a point between A and C, or a point between B and C, in contradiction with the hypothesis. □

Theorem 1.1.10 *If a line l intersects two sides of a triangle ΔABC, then it cannot intersect the third one.*

Proof Let us assume, by contradiction, that the line l intersects all sides $[BC]$, $[AC]$, and $[AB]$ of the triangle ΔABC in, respectively, D, E, and F. We can assume the order $\overline{EFD}$ on the line l. We consider the triangle ΔCDE and the secant line AB, which intersects $[DE]$ in F. It follows that AB intersects either $[DC]$ or $[EC]$ (axiom $\mathbf{O_4}$). In both cases, it follows that AB intersects either $[AC]$ or $[BC]$, respectively, in two points, which means that either $AB = BC$ or $AB = AC$ (axiom $\mathbf{I_1}$, in contradiction with the assumption that ΔABC is a triangle. □

In what follows, we introduce the notion of *half-line*. Let O be a fixed point on a line l and let $A, B \in l$ be two points such that we have the order $\overline{OAB}$. Then we call A and B to be on the same side of the point O. This defines a binary relation on the set of points of l.

Theorem 1.1.11 *The binary relation defined above is an equivalence relation on the set of points of a line l.*

Proof Reflexivity is obviously true, as for $A = B$, we have clearly the order $\overline{OAA}$. The symmetry follows from the fact that the order $\overline{OAB}$ is the same as the order $\overline{BAO}$ (axiom $\mathbf{O_1}$). For the transitivity, we observe: if we have $\overline{OAB}$ and $\overline{OBC}$, then it follows the order $\overline{OAC}$. □

In this context, we can define a half-line as follows.

Definition 1.1.12 The equivalence class of a point on a line l with respect to a fixed point $O \in l$ is called the *half-line* with vertex (origin) O.

An equivalent formulation would be as follows: given a pair of points A and B, the half-line starting at A and pointing in the direction of B consists of all points P so that we have either the order $\overline{ABP}$, or the order $\overline{APB}$. A half-line AB is often called a *ray* emanated from A towards B.

Theorem 1.1.13 *Let O and A be two points on a line l. The set of points $A' \in l$ such that we have the order $\overline{A'OA}$ forms a half-line with origin O.*

Proof Let A' be an arbitrary point such that $\overline{A'OA}$. Let B be a representative of the equivalence class defined by A with respect to O, i.e. A and B are on the same side of O. Thus we have the order $\overline{OAB}$. Let $B' \in l$ such that we have the order $\overline{B'OB}$. From the orders $\overline{BAO}$ and $\overline{B'OB}$ it follows the order $\overline{AOB'}$. But the orders $\overline{A'OA}$ and $\overline{B'OA}$ exclude the order $\overline{A'OB'}$ (try to prove this assertion). Therefore the points A' and B' are on the same side of O, which proves the conclusion of the theorem.□

The theorem above affirms that a point O on a line l divides the line into two half-lines. For any point $A \neq O$, we denote one half-line by (OA), and the other half-line by $(OA'$, also called the *complementary* half-line of (OA).

The set of points of a half-line is a total ordered set. Indeed, for two points A and B on a half-line, we have either A coincides with B, or we have one of the orders $\overline{OAB}$ or $\overline{OBA}$. If we have the order $\overline{OAB}$, we say A *precedes* B. Therefore, in view of this total ordering, for any two distinct points A and B on a half-line, either A precedes B or B precedes A.

In view of this remark, we can arrange any finite set of points on a line l in the order of their precedence. Moreover, if we denote the ordered points by $A_1, A_2, \ldots,$ then for any $i < j < k$ we have the order $\overline{A_i A_j A_k}$. This proves the following:

Theorem 1.1.14 *There is an order-preserving, one-to-one correspondence between any set of n points on a line l and the set of natural numbers* $\{1, 2, \ldots, n\}$.

Similarly as in the case of half-lines, one can introduce the following binary relation of the set of points in a plane.

Definition 1.1.15 If l is a line in a plane π and A, B are two points in π such that $[AB] \cap l = \emptyset$, then we say that the points A and B are *on the same side* of the plane π with respect to the line l.

This defines a binary relation on the set of points of the plane π.

As before, we prove the following:

Theorem 1.1.16 *The binary relation defined above is an equivalence relation.*

Proof Reflexivity and symmetry are obviously true. We have to prove the transitivity of this relation. Let A, B and B, C be on the same side of the plane π with respect to the line l. If follows that the intersections of l with $[AB]$, respectively $[BC]$, are empty. From a previous theorem it follows that $l \cap [AC] = \emptyset$, so the points A, C are on the same side of the plane with respect to the line l. □

In view of the theorem above, we give the following:

Definition 1.1.17 Let l be a fixed line in a plane π and a point $A \in \pi - l$. The equivalence class of A with respect to the line l is defined to be the half-plane determined by A and l. The line l is called the *border* of this half-plane.

Then we have the following.

Let l be a fixed line, and let $A \notin l$. Then the set of points A' with the property that the segment $[AA']$ intersects the line l forms a half-plane of border l.

Definition 1.1.18 This half-plane is called the *complementary* half-plane of the half-plane determined by l and A.

Note that every line l in a plane divides the plane into two half-planes, both with border l.

Definition 1.1.19 An *angle* is defined to be a pair of two half-lines h and k with the same origin O, denoted by $\angle(hk)$. The point O is called the *vertex* of the angle, and the half-lines h and k are called the *sides* of the angle.

If $h = (OA$ and $k = (OB$ are two half-lines defined by three non-collinear points O, A, and B (O is the vertex of the angle), then we will also denote the angle $\angle(hk)$ by $\angle AOB$.

Let us consider an angle $\angle(hk)$ in a plane π. Then, there are two distinguished half-planes: one is determined by the underlying line of the half-line h and the points of the half-line k, and, similarly, the other one is determined by the underlying line of the half-line k and the points of the half-line h.

Definition 1.1.20 We call the *interior* of the angle $\angle(hk)$, the intersection of the two half-planes above. The *exterior* of the angle $\angle(hk)$ consists of all the points in the plane which are neither in the interior, nor on the sides of the angle $\angle(hk)$.

In a similar fashion, one can define the interior of a triangle as follows.

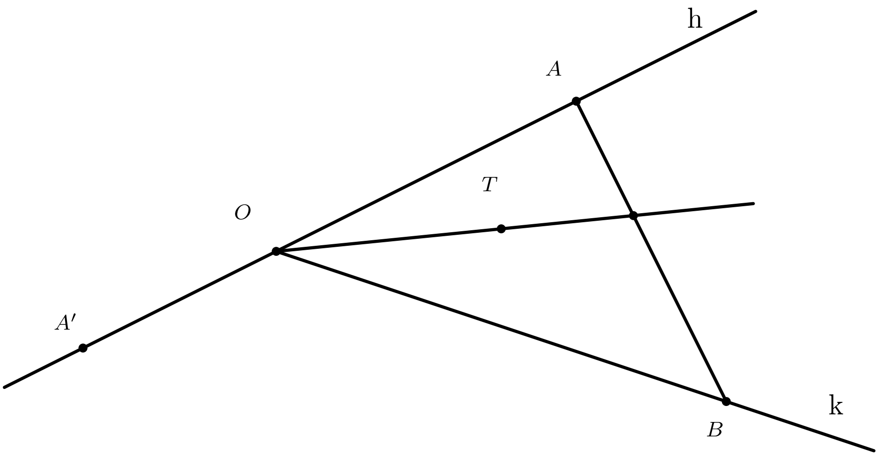


Fig. 1.1 Crossbar theorem

Definition 1.1.21 The *interior* of the triangle $\Delta ABC$is the intersection of the interiors of its angles.

Consider n half-lines with common vertex O and assume that there exists a line $l \not\ni O$ which intersects all of them. We can order all the intersection points ($\overline{A_1 A_2 A_3}$, etc.). This gives us the notion of a half-line being *between* two other half-lines, and implicitly an order on the set of half-lines.

The following theorem is usually known as the crossbar theorem, or, sometimes, as the transversal theorem . In the present approach, the proof relies on axiom $\mathbf{O_4}$, Pasch's axiom (Fig. 1.1).

Theorem 1.1.22 (Crossbar Theorem) *Let $\angle(hk)$ be an angle of vertex O. Let $A \in h$ and $B \in k$ be two points different than O, and T a point in the interior of the angle $\angle(hk)$. Then the half-line $(OT$ intersects the segments $[AB]$.*

Proof Denote by H_A the half-plane determined by OB and the point A. Consider a point A' on the complementary half-line of $(OA$, and $H_{A'}$ the half-plane determined by OB and the point A'. We apply Pasch's axiom $\mathbf{O_4}$ for the triangle $\Delta AA'B$ and the half-line $(OT$, which intersects $[AA']$ in O. Then $(OT$ should intersect either $[AB]$ or $[A'B]$. If $(OT$ doesn't intersect $[AB]$, it exists a point $L \in [A'B] \cap (OT$, in collision with the fact that all points of $(OT$ are in $H_{A'}$. □

As a final remark, we can observe that the complementary half-line of $(OT$, say $(OT'$ is included in the interior of the opposite angle of $\angle AOB$, say $\angle A'OB'$, therefore it cannot intersect neither $[A'B]$ nor $[AB]$, because they have empty intersection with the interior of $\angle A'OB'$.

Angles as $\angle AOT$ and $\angle TOB$ are called *adjacent angles*.

We introduce below the *axioms of congruence* and we study their immediate consequences. The congruence notion we introduce below is actually an equality notion, but it is called different just to make distinction between equality of real numbers and equality of geometric objects. The relationship between the set of real numbers and geometry is addressed later on.

The formulation of these axioms is after Arthur Rosenthal [185], which has considerably modified the original Hilbert's formulation of Axiom $\mathbf{E_4}$, by omitting the symmetry and transitivity properties of the congruence of angles. These properties can be actually proven from the axioms below.

The following axioms introduce the concept of *congruence* (equality) of segments and angles. The notion of congruence is written using the special symbol $\equiv$, in order to eliminate any confusion between this geometric notion with the equality notion from set or number theories. We will reserve the equality symbol $=$ for when we define the *values* of segments and angles.

Axiom $\boldsymbol{E_1}$: *If A and B are two points on a line l, and A' is a point on a line l', where l' is not necessarily distinct from l, then there exists a point B' on l' such that* $[AB] \equiv [A'B']$. For every segment $[AB] \equiv [BA]$.

As we can see from the previous axiom, the congruence $[AB] \equiv [A'B']$ is provided by the ability to construct the point B' on the line l' with the requested property.

Axiom $\boldsymbol{E_2}$: *If $[A'B'] \equiv [AB]$ and $[A''B''] \equiv [AB]$, then $[A'B'] \equiv [A''B'']$.*

Note that this axiom is not the transitivity property of congruence of segments. Transitivity will be proved in a theorem below. The next axiom establishes the additivity of the congruence of segments.

Axiom $\boldsymbol{E_3}$: *Let $[AB]$ and $[BC]$ be two segments of a line l, without common interior points, and let $[A'B']$ and $[B'C']$ be two segments without common interior points on a line l', where l' is not necessarily distinct from l. If $[AB] \equiv [A'B']$ and $[BC] \equiv [B'C']$, then $[AC] \equiv [A'C']$.*

The next axiom defines the congruence of angles in a plane.

Axiom $\boldsymbol{E_4}$: *Let $\angle(hk)$ be an angle in a plane π, and let l' be a line in a plane π', where π' is not necessarily distinct from π. Let h' be a half-line of l', where h' is not necessarily distinct from h. Then in one of the half-planes determined by l', there uniquely exists a half-line k', such that $\angle(hk) \equiv \angle(h'k')$. For every angle, $\angle(hk) \equiv \angle(hk)$ (reflexivity), and $\angle(hk) \equiv \angle(kh)$ (symmetry).*

As above, the congruence $\angle(hk) \equiv \angle(h'k')$ is provided by the ability to construct the angle $\angle(h'k')$ in one of the half-planes of π'.

Axiom $\boldsymbol{E_5}$: *For any angles, if $\angle(h'k') \equiv \angle(hk)$ and $\angle(h''k'') \equiv \angle(hk)$, then $\angle(h'k') \equiv \angle(h''k'')$.*

The next axiom is establishing conditions for congruences of angles of triangles. For an angle of a triangle $\triangle ABC$, say $\angle ABC$, we understand the angle determined by the half-lines $(BA$ and $(BC$.

Axiom $\mathbf{E_6}$: *Let* $\triangle ABC$ *and* $\triangle A'B'C'$ *be two triangles. If* $[AB] \equiv [A'B']$, $[AC] \equiv [A'C']$, *and* $\angle BAC \equiv \angle B'A'C'$, *then:*

$$\angle ABC \equiv \angle A'B'C' \qquad \angle ACB \equiv \angle A'C'B'.$$

The first two congruence axioms give the following result.

Theorem 1.1.23 *The congruence relation for segments is an equivalence relation.*

Proof We prove first the following statement: if we have two segments $[AB] \equiv [A'B']$, then $[AB] \equiv [B'A']$. Indeed, we have $[B'A'] \equiv [A'B']$ (axiom $\mathbf{E_1}$). Therefore $[AB] \equiv [A'B']$ and $[B'A'] \equiv [A'B']$, so, using axiom $\mathbf{E_2}$, it follows $[AB] \equiv [B'A']$.

Reflexivity now follows from axiom $\mathbf{E_1}$ ($[AB] \equiv [BA]$) and, from the statement above, it follows $[AB] \equiv [AB]$.

How to prove the symmetry? We have $[A'B'] \equiv [A'B']$, via the reflexivity proved above. Moreover, if $[AB] \equiv [A'B']$ it follows that $[A'B'] \equiv [AB]$, via Axiom $\mathbf{E_2}$. It is very important to notice that only from this point on, we have the right to assert that $[AB] \equiv [CD]$ is the same as $[CD] \equiv [AB]$.

For transitivity, we consider $[AB] \equiv [A'B']$, and $[A'B'] \equiv [A''B'']$. But the congruence $[A'B'] \equiv [A''B'']$ implies the congruence $[A''B''] \equiv [A'B']$ (symmetry). Then, from $[AB] \equiv [A'B']$ and $[A''B''] \equiv [A'B']$, it follows the congruence $[AB] \equiv [A''B'']$ (axiom $\mathbf{E_2}$). □

The congruence relation, being an equivalence relation, gives rise to a partition of the set of all segments in disjoint equivalence classes. This fact allows us to define all segments in an equivalence class to have the same *value*. We denote the value of a segment $[AB]$ by simply AB. Note that the same notation AB is also used for the line which passed through the points A and B. In general it is clear from the context if we refer to the line AB or to the value of the segment $[AB]$. Moreover, the congruence $[AB] \equiv [CD]$ can be also written as an equality of values, $AB = CD$, when there is no danger of confusion between equivalence classes and their representatives. In what follows, going back and forth between congruence of segments (or angles) and equality of their values, technically requires one to prove the independence of chosen representatives in a given equivalence class. For the simplicity of geometric arguments, we will omit these technical details.

Theorem 1.1.24 *Let* $(OA$ *be a half-line with origin* O. *If* C *and* C' *are two points on* $(OA$ *such that* $[OC] \equiv [OC']$, *then the points* C *and* C' *coincide.*

Proof Without loss of generality, we can assume the order $\overline{OCC'}$. Let I be a point which does not belong to the half-line $(OA$ (Axiom $\mathbf{I_3}$). Then, in the triangles ΔOCI and $\Delta OC'I$, we have $[OC] \equiv [OC']$, $[OI] \equiv [OI]$ and $\angle IOC \equiv \angle IOC'$. From

Axiom $\mathbf{E_6}$ it follows $\angle OIC \equiv \angle OIC'$, therefore the half-lines $(IC$ and $(IC'$ coincide as sets (Axiom $\mathbf{E_4}$). This implies $(IC \cap (OA = (IC' \cap (OA$, so C and C' coincide. □

Sometimes we write $C = C'$ whenever C and C' coincide. Notice that the equal sign which expresses the coincidence is not the same as the usual symbol $=$ of equality of numbers.

Note that Axiom $\mathbf{E_3}$ guarantees the additivity of the values of segments on same line. Indeed, if A, B, C and A', B', C' are points on the lines l and l', respectively, with orders $\overline{ABC}$ and $\overline{A'B'C'}$, respectively, such that $[AB] \equiv [A'B']$, $[BC] \equiv [B'C']$, then it follows directly from Axiom $\mathbf{E_3}$ that $[AC] \equiv [A'C']$. We can formally write the following equalities in terms of values of segments: $AC = AB + BC$ and $A'C' = A'B' + B'C'$.

Theorem 1.1.25 *The congruence relation for segments preserves the order relation.*

Proof Consider the points A, B, C on a line l, with the property that B is an interior point of the segment $[AC]$, i.e. we have the order $\overline{ABC}$. Moreover, let us consider the points A', B', C' on another line l', such that $[AB] \equiv [A'B']$, $[AC] \equiv [A'C']$, and B', C' are on the same half-line of vertex A'.

If we show that B' is interior to $[A'C']$, and $[B'C'] \equiv [BC]$, then it will follow the order $\overline{A'B'C'}$, which is the conclusion of our theorem. Indeed, assume the existence of another point $C'' \in l'$ with order $\overline{A'B'C''}$, such that $[B'C''] \equiv [BC]$. $[A'B'] \equiv [AB]$ and $[B'C''] \equiv [BC]$, so, by additivity, it follows $[A'C''] \equiv [AC]$. But $[A'C'] \equiv [AC]$, thus $[A'C''] \equiv [A'C']$, therefore it follows that $C' = C''$. Thus we have the desired order $\overline{A'B'C'}$. □

In view of the results above, one can define the *difference* operation among segments. Indeed, if $[AB]$ and $[AC]$ are two segments on a line l, such that they have order $\overline{ABC}$, then the difference of the values of $[AC]$ and $[AB]$ is the value of the segment $[BC]$, respecting the additivity property $AB + BC = AC$. Therefore we can write $AC - AB = BC$.

Definition 1.1.26 Two triangles $\triangle ABC$ and $\triangle A'B'C'$ are called *congruent*, and we denote by $\triangle ABC \equiv \triangle A'B'C'$, if they have congruent sides and congruent angles, respectively.

Concretely, $\triangle ABC \equiv \triangle A'B'C'$ if the following six congruences are respected:

$$[AB] \equiv [A'B'], \quad [BC] \equiv [B'C'], \quad [CA] \equiv [C'A'],$$

$$\angle BAC \equiv \angle B'A'C', \quad \angle ABC \equiv \angle A'B'C', \quad \angle BCA \equiv \angle B'C'A'.$$

When there is no danger of confusion, we denote by $\angle A$ the angle $\angle BAC$. The first result about congruence of triangles is the following.

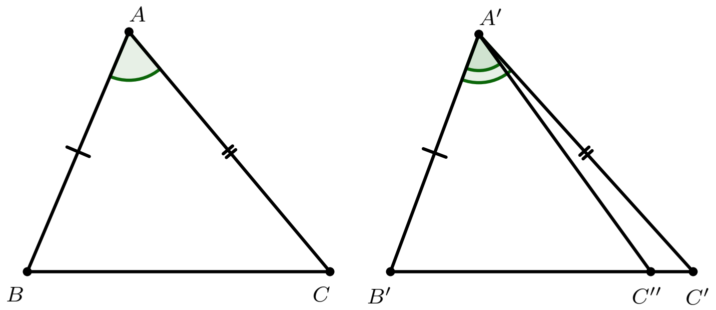


Fig. 1.2 Theorem SAS

Theorem 1.1.27 *If a triangle $\triangle ABC$ has two congruent sides, then it has two congruent angles, too. In this case, we call the triangle $\triangle ABC$ to be isosceles.*

Proof Without loss of generality, we can assume $[AB] \equiv [AC]$. Then the triangles $\triangle BAC$ and $\triangle CAB$ are in the conditions of Axiom $\mathbf{E_6}$, thus $\angle ABC \equiv \angle ACB$. □

The next theorem is the first important congruence case of triangles (Fig. 1.2).

Theorem 1.1.28 (SAS) *Let $\triangle ABC$ and $\triangle A'B'C'$ be two triangles, such that $[AB] \equiv [A'B']$, $[AC] \equiv [A'C']$, and $\angle BAC \equiv \angle B'A'C'$. Then $\triangle ABC \equiv \triangle A'B'C'$. (This congruence case is called Side-Angle-Side (SAS).)*

Proof Using axiom $\mathbf{E_6}$, we have $\angle ABC \equiv \angle A'B'C'$ and $\angle ACB \equiv \angle A'C'B'$. The only congruence left to show is $[BC] \equiv [B'C']$. Consider a point C'' on the half-line $(B'C'$ such that $[BC] \equiv [B'C'']$ (Axiom $\mathbf{E_1}$). Consider now the triangles $\triangle ABC$ and $\triangle A'B'C''$. From $[AB] \equiv [A'B']$, $[BC] \equiv [B'C'']$, and $\angle ABC \equiv \angle AB'C''$, it follows from axiom $\mathbf{E_6}$ that $\angle BAC \equiv \angle B'A'C''$. From the hypothesis, we have $\angle BAC \equiv \angle B'A'C'$. Then we have C' and C'' such that the angles $\angle C'A'B'$ and $\angle C''A'B'$ are congruent. Since C' and C'' are in the same half-plane with respect to the line $A'B'$, it follows from axiom $\mathbf{E_4}$ that $(A'C'$ and $(A'C''$ coincide, thus $C' = C''$. □

The next theorem establishes the second case of triangle congruence (Fig. 1.3).

Theorem 1.1.29 (ASA) *Let $\triangle ABC$ and $\triangle A'B'C'$ be two triangles, such that $[BC] \equiv [B'C']$, $\angle ABC \equiv \angle A'B'C'$, and $\angle ACB \equiv \angle A'C'B'$. Then $\triangle ABC \equiv \triangle A'B'C'$. (This congruence case is called Angle-Side-Angle (ASA).)*

Proof Let $A'' \in (B'A'$ such that $[BA] \equiv [B'A'']$. Consider the triangles $\triangle BAC$ and $\triangle B'A''C'$. Axiom $\mathbf{E_6}$ guarantees that $\angle BCA \equiv \angle B'C'A''$. Since A' and A'' are in the same half-plane with respect to $B'C'$, it follows that $(C'A'$ and $(C'A''$ coincide. Therefore, $A' = A''$. We apply Theorem SAS for the triangles $\triangle ABC$ and $\triangle A'B'C'$, where we now have $[AB] \equiv [A'B']$, $[BC] \equiv [B'C']$ and $\angle ABC \equiv \angle A'B'C'$. □

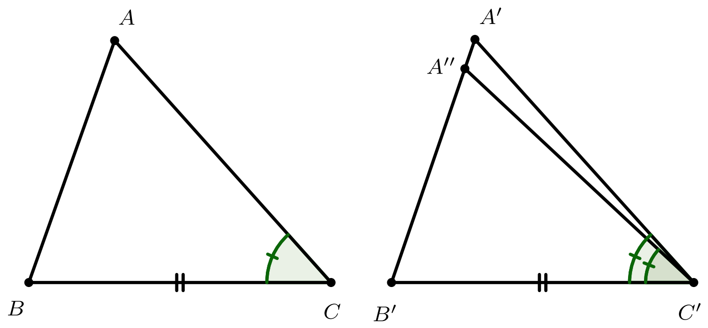


Fig. 1.3 Theorem ASA

Theorem 1.1.30 (Additivity of Angles) *If* $\angle(hl) \equiv \angle(h'l')$*, and* $\angle(lk) \equiv \angle(l'k')$*, where* l *and* l' *are half-lines interior to the angles* $\angle(hk)$ *and* $\angle(h'k')$*, then* $\angle(hk) \equiv \angle(h'k')$.

Proof Let H and K be two points such that $H \in h$ and $K \in k$. Using Crossbar Theorem, it follows that $l \cap [HK] \neq \emptyset$. Let $\{L\} = l \cap [HK]$. Now take $H' \in h'$ and $L' \in l'$ such that $[OH] \equiv [O'H']$ and $[OL] \equiv [O'L']$, and take K' on the half-line complement to $(L'H'$ such that $[L'K'] \equiv [LK]$. Notice that the congruence $\triangle OHL \equiv \triangle O'H'L'$ (case SAS) implies $[HL] \equiv [H'L']$ and $\angle OHL \equiv \angle O'H'L'$. But the segments $[HL]$, $[LK]$; $[H'L']$, $[L'K']$ satisfy the conditions of axiom $\mathbf{E_3}$, thus the triangles $\triangle OHK$ and $\triangle O'H'K'$ are congruent (case SAS). It follows that $\angle HOK \equiv \angle H'O'K'$, thus using axiom $\mathbf{E_4}$, it follows that the half-lines $(O'K'$ and k' coincide. □

Suppose we are in the same hypothesis as in Theorem of Additivity of Angles.

Theorem 1.1.31 *If* $\angle(hk) \equiv \angle(h'k')$*, and* $\angle(hl) \equiv \angle(h'l')$*, then* $\angle(lk) \equiv \angle(l'k')$.

Proof Consider the triangles $\triangle ABC$ and $\triangle A'BC$ such that A and A' are in different half-planes with respect to the line BC. If $[AB] \equiv [A'B]$ and $[AC] \equiv [A'C]$, then triangles $\triangle ABC$ and $\triangle A'BC$ have congruent angles, respectively. Considering the segments $[AA']$ and $[BC]$, we distinguish two cases: $[AA'] \cap [BC] \neq \emptyset$ or $[AA'] \cap [BC] = \emptyset$. In each one of these cases, we apply the theorem for isosceles triangles in the case of triangles $\triangle ABA'$ and $\triangle ABA'$, respectively. The conclusion of the theorem follows then immediately. □

Now we are in the right context to prove the following side-side-side (SSS) congruence theorem of triangles. Note that in the proof we do not use neither the symmetry, nor the transitivity of the equality relation for angles! These properties are an immediate corollary to the following theorem (Fig. 1.4).

Theorem 1.1.32 (SSS) *Let* $\triangle ABC$ *and* $\triangle A'B'C'$ *be two triangles, such that* $[AB] \equiv [A'B']$*,* $[BC] \equiv [B'C']$*, and* $[CA] \equiv [C'A']$*. Then* $\triangle ABC \equiv \triangle A'B'C'$*. This congruence case is called Side-Side-Side (SSS).*

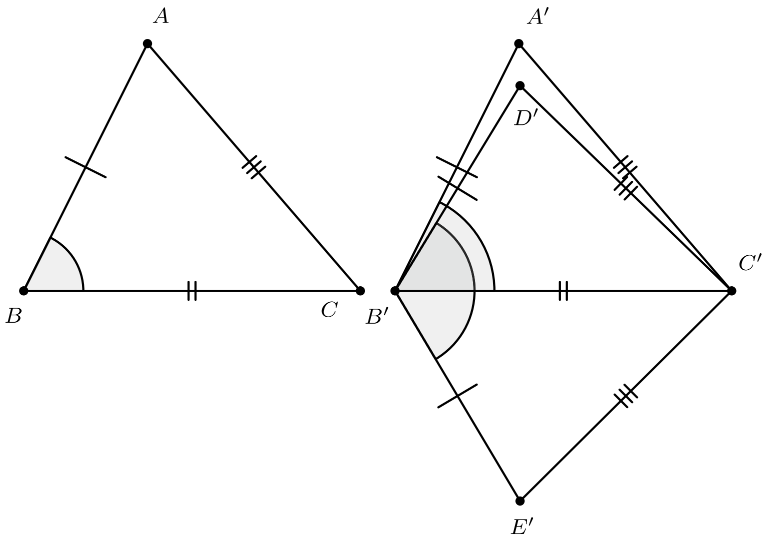


Fig. 1.4 Theorem SSS

Proof Consider the half-line $(B'D'$ such that $[B'D'] \equiv [AB]$ and $\angle D'B'C' \equiv \angle ABC$, D' in the half-plane determined by A' and $B'C'$.

Since $[BC] \equiv [B'C']$, $[BA] \equiv [B'D']$, and $\angle ABC \equiv \angle D'B'C'$, $\triangle ABC \equiv \triangle D'B'C'$ (case SAS). It follows that $[AC] \equiv [D'C']$. Let us construct a point E' in the complementary half-plane defined by the line $B'C'$ and the point A', such that $[B'E'] \equiv [B'D']$ and $\angle E'B'C' \equiv \angle C'B'D'$. It follows that $\triangle D'B'C' \equiv \triangle E'B'C'$ (case SAS), thus $[E'C'] \equiv [D'C'] \equiv [AC] \equiv [A'C']$. Similarly, $[E'B'] \equiv [B'D'] \equiv [AB] \equiv [A'B']$.

Then, using the fact that isosceles triangles have equal angles corresponding to equal sides, the triangles $\Delta A'B'C'$ and $\Delta E'B'C'$ are congruent (since we use the previous theorems with sum or difference of angles to prove that $\angle B'A'C' \equiv \angle B'E'C'$). Then $\angle A'B'C' \equiv \angle E'B'C'$, so in the half-plane determined by $B'C'$ and A' we have two half-lines $(B'D'$ and $(B'A'$, such that they determine $\angle A'B'C' \equiv \angle D'B'C'$. Therefore they are coincident and this means that the points A' and D' have to coincide. □

Corollary 1.1.33 *The congruence relation for triangles is an equivalence relation.*

Corollary 1.1.34 *The congruence relation for angles is an equivalence relation.*

The details are left for the reader, and here it is used $\mathbf{E_5}$. We have to mention that this equivalence relation allows us to define a value for all representatives of a class, which can be denoted by $v(\angle(hk)$, with the same remarks we did in the case of segments.

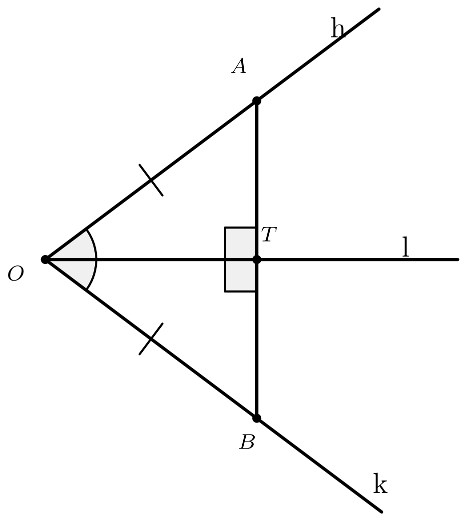


Fig. 1.5 Right angle existence

Definition 1.1.35 Let $\angle hk$ be an angle. The angle formed by a ray of angle $\angle hk$ and the complement of the other ray is called the supplementary angle to the angle $\angle hk$.

Definition 1.1.36 Two angles which have the same vertex and complementary sides are called opposite (or complementary) angles.

We propose two problems to the reader.

Problema 1.1.37 Supplementary angles of congruent angles are congruent.

Problema 1.1.38 Opposite angles are congruent.

Hint for both problems: choose points on rays such that congruent triangles occur (Fig. 1.5).

Definition 1.1.39 A right angle is an angle congruent to its supplementary angle. We denote by R the class of the right angles.

The following theorem establishes the existence of right angles in any geometry respecting all axioms introduced so far.

Theorem 1.1.40 *There exist right angles.*

Proof Consider the congruent angles $\angle hl$ and $\angle lk$ such that all rays have the common point O and l belongs to the interior of $\angle hk$. Choose $A \in h,\ B \in k$ such that $[OA] \equiv [OB]$. Crossbar theorem tells us that it exists $\{T\} = l \cap [AB]$. It is easy to see using congruent triangles that $\angle ATO \equiv \angle BTO$, i.e. $\angle ATO$ and $\angle BTO$ are both right angles. $\square$

The supplementary angle of an right angle is a right angle itself. The angle $\angle ATB$ can be seen as the sum of the right angles $\angle ATO$ and $\angle OTB$, therefore its class is $R + R$, i.e. $2R$.

Definition 1.1.41 The points A and B are called symmetric with respect the line l. The line AT is called perpendicular to l, T is called the foot of the perpendicular line to l passing through A.

Theorem 1.1.42 *All right angles are congruent.*

Proof By contradiction, we assume that there exist two right angles $\angle BAD$ and $\angle B'A'D'$ which are not congruent. We consider the supplementary angles $\angle CAD$ and $\angle C'A'D'$, respectively. Consider the half-line $(AE$ such that $\angle BAE \equiv \angle B'A'D'$ and observe the equality of angles $\angle CAE \equiv C'A'D'$. Therefore we have $\angle CAE \equiv \angle C'A'D' \equiv \angle B'A'D' \equiv \angle BAE$. Let $(AF$ such that $\angle CAE \equiv \angle BAE$. It results $\angle CAE \equiv \angle CAF$, in collision with $\mathbf{E_4}$. □

Theorem 1.1.43 *The perpendicular line from an exterior point to a given line is unique.*

Proof By contradiction, suppose AC and AC' are perpendicular lines to l, C, $C' \in l$. Consider the symmetric points B, B' of A with respect to l on each perpendicular line and choose $O \in l$ such that the order is $\overline{OC'C}$. It results $\triangle OCA \equiv \triangle OCB$ and $\triangle OC'A \equiv \triangle OC'B'$. We have $\angle BOC \equiv \angle AOC' \equiv \angle B'O'C'$ and $[OB] \equiv [OB'] \equiv [OA]$, i.e. B and B' coincide, therefore C coincides C'. □

Definition 1.1.44 Let $[AB]$ and $[A'B']$ be two segments. If there exists a point C in the interior of the segment $[AB]$ such that $[AC] \equiv [A'B']$, we say that the segment $[A'B']$ is less than the segment $[AB]$, and we denote by $[A'B'] < [AB]$.

In the same time we may say that the segment $[AB]$ is greater than the segment $[A'B']$ and we denote by $[AB] > [A'B']$. Note that the order $\overline{ABC}$ on a line determines the inequalities $[AB] < [AC]$ and $[BC] > [AC]$. We may also define $[AB] \leq [A'B']$, etc. The inequality relation $\leq$ is a partial order relation on the set of segment and more, if $[AB] > [A'B']$ and $[CD] > [C'D']$, then $[AB] + [CD] > [A'B'] + [C'D']$. The inequality can be transferred to values with the notations established there.

Definition 1.1.45 Let $\angle(h'k')$ and $\angle(hk)$ be two angles. If there is a line l in the interior of the angle $\angle(hk)$ such that $\angle(h'k') \equiv \angle(hl)$, then we can say that the $\angle(h'k')$ is less than the angle $\angle(hk)$, denoted by $\angle(h'k') \leq \angle(hk)$.

Or, we can say that the angle $\angle(hk)$ is greater than the angle $\angle(h'k')$, denoted by $\angle(hk) > \angle(h'k')$. We can easily define $\angle(hk) \geq \angle(h'k')$ or $\angle(h'k') \leq \angle(hk)$.

We do not insist and we left to the reader to prove that the inequality relations $\geq$ and $\leq$ are partial order relations on the set of angles.

Definition 1.1.46 Two lines which do not have any common point are called non-secant lines.

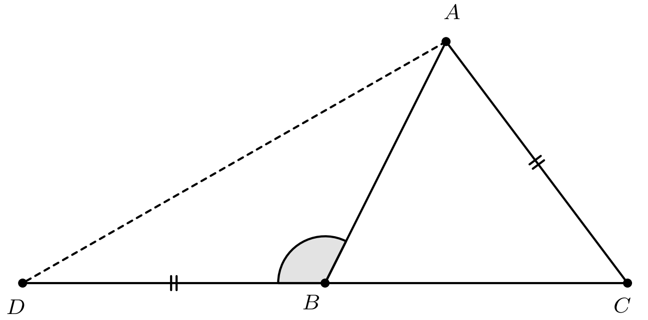


Fig. 1.6 The exterior angle theorem

Definition 1.1.47 Consider the triangle $\triangle ABC$. The angle formed by the half-line $(BA$ and the complement half-line of $(BC$, say $(BD$, is called the exterior angle of the triangle $\triangle ABC$ with respect to the vertex B.

Consider the angle formed by the half-line $(BC$ and the complement half-line of $(BA$, say $(BF$. This angle is also the exterior angle of the triangle $\triangle ABC$ with respect to the vertex B, and of course $\angle ABD \equiv \angle CBF$ as opposite angles. Having in mind the previous definition we can prove (Fig. 1.6).

Theorem 1.1.48 (Exterior Angle Theorem) *The exterior angle of a triangle with respect to a given vertex is greater than both the angles of the triangle which are not adjacent to it.*

Proof Let us fix the vertex to be B. We have to prove "the exterior angle of the triangle $\triangle ABC$ with respect the vertex B is greater than both the angles $\angle BAC$ and $\angle ACB$". Let D be a point on BC with the order $\overline{DBC}$ such that $[BD] \equiv [AC]$. We show that $\angle DBA > \angle BAC$. The other inequality results from $\angle ABD \equiv \angle CBF > \angle ACB$. We focus on the first inequality. By contradiction, let us suppose that $\angle ABD \equiv \angle BAC$. If we succeed to obtain a contradiction, the case $\angle ABD < \angle BAC$ is reduced to the previous case by considering $C_1 \in (BC)$ such that $\angle ABD \equiv \angle BAC_1$. Therefore, it remains to prove that $\angle ABD \equiv \angle BAC$ is impossible. In the given conditions it results $\triangle ABD \equiv \triangle CAB,\ (SAS)$, i.e. $\angle DAB \equiv \angle ABC$. Since $\angle CAD \equiv \angle CAB + \angle BAD \equiv \angle ABD + \angle ABC \equiv \angle CBD = 2R$, equivalent to $A \in BC$, in collision to the fact that ABC is a triangle. $\square$

Corollary 1.1.49 *The sum of two among the three angles of triangle is less than the sum of two right angles.*

Proof To simplify the writing denote by B_e the exterior angle with respect to the vertex B. The exterior angle theorem asserts that $B_e > A,\ B_e > C$. It results $B_e + B > A + B$, therefore $A + B < 2R$. $\square$

Definition 1.1.50 An angle of a triangle which is greater than a right angle is called an obtuse angle. An angle of a triangle which is less than a right angle is called an acute angle.

Corollary 1.1.51 *A triangle cannot have more than one obtuse angle.*

Proof Suppose there exists a triangle $\triangle ABC$ such that $A > R$ and $B > R$. Then $A + B > 2R$, in collision with the previous corollary. □

We left for the reader the following very nice problems:

Problema 1.1.52 Given a segment $[AB]$, there exists an unique point M (called the midpoint of the segment $[AB]$), such that $[AM] \equiv [MB]$.

Problema 1.1.53 Given an angle $\angle hk$, there is an unique half-line l in its interior such that $\angle hl \equiv \angle lk$ (the half-line l is called the bisector of the $\angle hk$).

The previous problems create an infinity of points in the interior of a given segment and an infinity of half-lines in the interior of an angle.

Definition 1.1.54 The perpendicular line from a vertex of a triangle on the line which contains the opposite side is called an altitude (or height) of the triangle.

Theorem 1.1.55 *At least one altitude among the three altitudes of a triangle lies in the interior of the triangle.*

Proof (Hint) Consider the altitude corresponding to the greatest angle of a triangle, say AD, $D \in BC$. B and C are mandatory acute angles. The order on BC has to be $\overline{BDC}$. □

Theorem 1.1.56 *In the triangle $\triangle ABC$, $[AC] > [AB]$ if and only if $\angle B > \angle C$.*

Proof Consider $D \in [AC]$ such that $[AB] \equiv [AD]$. It results $\angle B > \angle ABD \equiv \angle ADB > \angle C$. Conversely, assume $\angle B > \angle C$ and by contradiction, $[AC] \leq [AB]$. If $[AC] \equiv [AB]$, then $\angle B \equiv \angle C$, contradiction. If $[AC] < [AB]$, then $\angle B < \angle C$, contradiction. □

Theorem 1.1.57 (triangle inequality) *In every triangle $\triangle ABC$ the sum of two sides is bigger than the third side. For example $[BC] < [BA] + [AC]$.*

Proof Consider $D \in (BA$ such that the order is $\overline{BAD}$ and $[AD] \equiv [AC]$. It follows that $[BD] \equiv [BA] + [AD] \equiv [BA] + [AC]$. Since $\angle BDC \equiv \angle DCA < \angle DCB$ it follows that $[BD] > [BC]$, that is $[BA] + [AC] > [BC]$. □

Theorem 1.1.58 $[A_1A_n] \leq [A_1A_2] + [A_2A_3] + \cdots + [A_{n-1}A_n]$

Proof (Hint) $[A_1A_n] \leq [A_1A_2] + [A_2A_n] \leq [A_1A_2] + [A_2A_3] + [A_3A_n] \leq \cdots$. □

Theorem 1.1.59 *Consider the triangles $\triangle ABC$ and $\triangle A'B'C'$ such that $[AB] \equiv [A'B']$, $[AC] \equiv [A'C']$. If $\angle A > \angle A'$, then $[BC] > [B'C']$.*

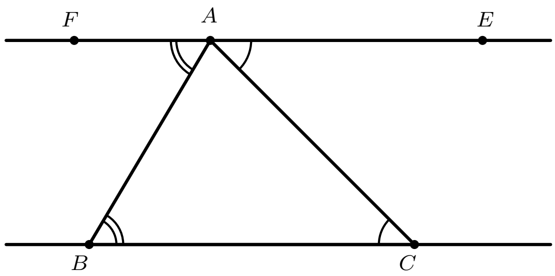


Fig. 1.7 The existence of a non-intersecting line

Proof (Hint) Consider the half-line $(AD$ such that $\angle BAD \equiv \angle B'A'C'$ and $[AD] \equiv [AC] \equiv [A'C']$. Observe that $(AD$ is included in the interior of the angle $\angle BAC$. We have $\triangle ABD \equiv \triangle A'B'C'$. The triangle $\triangle ACD$ is isosceles, $[AC] \equiv [AD]$, therefore $\angle DCB < \angle ADC < \angle BDC$, i.e. $[BC] > [BD] \equiv [B'C']$ (Fig. 1.7). □

Theorem 1.1.60 *From a point A exterior to a line d, one can construct at least one non-secant line to d.*

Proof Consider the points B, C on d and a half-line $(AE$ in the half-plane determined by A and d such that B and E are in opposite half-planes with respect to the line AC and $\angle EAC \equiv \angle BCA$. According to the exterior angle theorem we have $(AE \cap (BC = \emptyset$. The complementary half-line $(AF$ has the property $\angle BAF \equiv \angle ABC$. The same exterior angle theorem implies $(AF \cap (CB = \emptyset$. Therefore $FE \cap d = \emptyset$. □

Definition 1.1.61 The angles $\angle EAC$ and $\angle ACB$ are called interior alternate angles.

The angles $\angle FAB$ and $\angle ABC$ are interior alternate angles, too.

The reader observes that until now there is no a parallelism axiom involved in the construction we made. We are still in the absolute geometry area mentioned at the beginning of this chapter. The previous result is an important one. In the axiomatic frame created before it exists at least one non-secant line through a point with respect to a given line. There is only one or there are more? We left the answer for later.

The axioms before allow us to have infinitely many points on a line, but we don't know if a line can be "filled" with points or if it is "unbounded". Until now we can see that if we establish an origin O on a line and if we take a segment $[AB]$ we can construct on one half-line the points E_1, E_2, ..., E_n,... such that $[AB] \equiv [OE_1] \equiv [E_1E_2] \equiv [E_2E_3] \equiv \cdots \equiv [E_nE_{n+1}] \equiv ...$ and on the complementary half-line the points E_{-1}, E_{-2}, ..., E_{-n},... such that $[AB] \equiv [OE_{-1}] \equiv [E_{-1}E_{-2}] \equiv [E_{-2}E_{-3}] \equiv \cdots \equiv [E_{-n}E_{-(n+1)}] \equiv \cdots$, therefore we can associate for any integer number a point on the line l. Combining with a result before related about the existence of the midpoint of every segment, we can see on the line l all rational points having the form $\dfrac{n}{2^m}$.

So, not all the real numbers can be “seen” on l. And still the problem of unboundedness persists. Why? Since even if $[OE_n] = [OE_1] + [E_1E_2] + \cdots + [E_{n-1}E_n] < [OE_{n+1}]$ the following example of segments bigger and bigger is bounded in the segment [0, 1]. It is about the sequence of intervals $(0, 1 - \frac{1}{n})$, $n \in \mathbb{N}$. Can we make any connection between the set of real numbers and the points of a line? We need to introduce the *axioms of continuity* at this point.

Axiom $\mathbf{C_1}$ *(Axiom of Archimedes): Let* $[AB]$ *and* $[CD]$ *be two arbitrary segments such that* $[CD] < [AB]$. *Then, there exists a finite number of points* $A_1, A_2, ...A_n, ...$ *on the ray* $(AB$, *such that* $[CD] \equiv [AA_1] \equiv [A_1A_2] \equiv [A_2A_3] \equiv \cdots \equiv [A_{n-1}A_n]$, *the interiors of those segments have every two an empty intersection and finally, either* $B = A_n$ or $B \in (A_{n-1}A_n)$.

In view of the additivity property of segments we can write that it exists $n \in \mathbb{N}$ such that

$$[AA_1] + [A_1A_2] + [A_2A_3] + \cdots + [A_{n-1}A_n] \equiv n[CD] \geq [AB]$$

and the inequality may refer to values. The Axiom of Archimedes multiplies values by natural numbers and we expect to understand the value of a segment as a real positive number describing the length of the segment. Considering $n = 1$ in the previous inequality, we have the old inequality between segments, therefore $\mathbf{C}_1$ offer us the chance to understand the appropriate nature of values attached to the segments and the unboundedness of the set of natural numbers.

The next axiom is attributed to Cantor and it will be involved in “completing” the line with points we don’t know until now that they have to belong to a line.

Axiom $\boldsymbol{C_2}$ (Axiom of Cantor): *Let* $[A_1B_1], [A_2B_2], ...$ *be a sequence of segments on a given line* l, *such that every segment is included in the interior of the precedent one, i.e.* $[A_nB_n] \subset [A_{n-1}B_{n-1}]$ *for all* $n \geq 2$. *If we assume that no segment is included in the interior of all segments* $[A_nB_n]$, $n \in \mathbb{N}$, *then there is an unique point* M *on the line* l *such that* $\{M\} = [A_1B_1] \cap [A_2B_2] \cdots \cap [A_nB_n] \cap \cdots$

These two axioms of continuity allow us to use real positive numbers as the “values” of segments and angles. The results are a little bit more complicated and we try to suggest them without complete proofs.

Using the continuity axiom $\mathbf{C}_1$ we can assign the natural number 1 to the segment $[CD]$ and the number n to the value nCD. To every segment $[AB]$ we attach a system of coordinates on the line $d = AB$ such that A is the origin O and $1 = OA_1 = A_1A_2 = \cdots = A_{m-1}A_m = \cdots$. According to $\mathbf{C}_1$ it exists one integer $m \in \mathbb{N}$ such that $B \in [A_{n-1}A_n]$. If $B = A_{n-1}$ then $AB = n - 1$. If $B = A_n$ then $AB = n$. If $B = M_1$ is the midpoint of the segment $[A_{n-1}A_n]$ we assign to the value AB is the 2-adic number $n - 1$, 10000...0... In fact at each step from now when B is the midpoint of a segment, we associate 1 and the other decimals after are 0. Suppose $B \in [A_{n-1}M_1]$. We consider the value of AB as $\overline{n-1}$, 0 and we are looking after the next decimal observing where B is with respect to the midpoint M_2 of the segment $[A_{n-1}M_1]$. If $B \in [M_2M_1]$ the next decimal is 1, therefore the attached number is until now $\overline{n-1}$, 01 and we continue looking at the position of B with respect to the

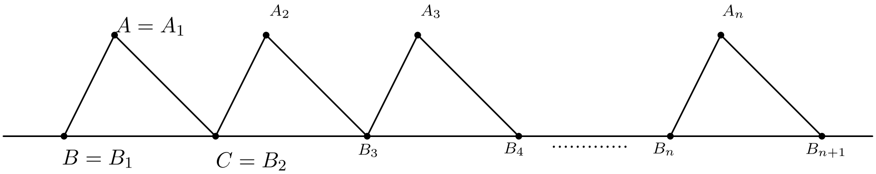


Fig. 1.8 Legendre's theorem

midpoint M_3 of the segment $[M_2 M_1]$. Imagine a little bit the position of B if the next three digits are 001 such that the value of AB is until now $\overline{n-1}, 01001$. We can continue to discover digits until B is a midpoint of a segment when we stop with a 1 followed by 0 only, or we never stop because the point stays in the intersection of all segments which are like in axiom $\mathbf{C}_2$. The real number

$$\overline{n-1}, a_1 a_2 a_3 ... a_n ...$$

with $a_i = 1$ if B is in the "at the right" segment, or with $a_i = 0$ if B is in the "at the left" segment, is the 2-adic number attached to the value of the segment AB.

Then we can show that to every real number we can assign a unique point on l.

The theory can be extended to angles with the following two theorems.

Theorem 1.1.62 *Let (a_1, b_1), (a_2, b_2), ... be a sequence of angles with common vertex O, with the property that the angle (a_{n+1}, b_{n+1}) is contained in the angle (a_n, b_n), for all $n \geq 1$. In the assumption that there is no angle contained in the interior of all angles in the sequence, then there is a unique half-line l in the intersection of the interior of all angles.*

Proof (Hint) Intersect all angles with a line l and denote the points of intersection with a_k by A_k and the points of intersection with b_k with B_k, etc. □

Theorem 1.1.63 *Let $\angle(hk)$ and $\angle(h'k')$ be two angles. There exists a natural number n such that $n\angle(hk) > \angle(h'k')$.*

Proof (Hint) Observe that the measure of the angle $\angle(h'k')$ is less than $2R$. If $\angle(hk) > R$ we take $n = 2$. If $R > \angle(hk) > \dfrac{R}{2}$ we take $n = 4$, etc. □

We have two statements for angles analogous with the axioms $\mathbf{C}_1$ and $\mathbf{C}_2$. We can develop a similar theory for defining measure of angles, restricting all the proofs in the case of segments to the interval $(0, 2R)$. Then, there is a one-to-one correspondence between the set of angles and the interval $(0, 2R)$ of the real numbers.

We prove now in the Absolute Geometry frame the most important result regarding the sum of the angles of a triangle (Fig. 1.8).

Theorem 1.1.64 (Legendre) *For any triangle, the sum of its angles is at most $2R$.*

Proof Consider the triangle $\triangle ABC$. We have to show $\angle A + \angle B + \angle C \leq 2R$. By contradiction, let us assume that $\angle A + \angle B + \angle C > 2R$. On the line BC we consider the points $B = B_1$, $C = B_2$, B_3, ..., B_n, B_{n+1} in this order such that $B_1B_2 = B_2B_3 = \cdots = B_nB_{n+1}$ and in the same half-plane the points $A = A_1$, A_2, A_3, ..., A_n, such that $\triangle A_1B_1B_2 \equiv \triangle A_2B_2B_3 \equiv \cdots \triangle A_nB_nB_{n+1}$. It is easy to see that the following triangles are congruent, $\triangle A_1B_2A_2 \equiv \triangle A_2B_3A_3 \equiv \cdots \triangle A_{n-1}B_nA_n$, therefore $A_1A_2 = A_2A_3 = \cdots = A_{n-1}A_n$. It is easy to deduce that $\angle A > \angle A_1B_2A_2$ and then $BC = B_1B_2 > A_1A_2$. The polygonal line $B_1A_1A_2..A_nB_{n+1}$ is bigger than the segment $B_1B_{n+1} = nBC$, that is

$$nBC < BA + (n-1)A_1A_2 + AC.$$

This one can be written in the form

$$(n-1)(BC - A_1A_2) < BA - BC + AC.$$

We know that $BC - A_1A_2 > 0,\ \ BA - BC + AC > 0$, i.e. there exists the segments $[ST], [MK]$ such that $BC - A_1A_2 = ST,\ \ BA - BC + AC = MK$ and $(n-1)ST < MK$. But in the last inequality the natural number n is arbitrary, in collision with $\mathbf{C}_1$. Therefore $\angle A + \angle B + \angle C$ cannot be greater than $2R$. It follows $\angle A + \angle B + \angle C \leq 2R$. □

The next definition takes care that the values of angles are real numbers.

Definition 1.1.65 For any triangle $\triangle ABC$ we define the defect of it, denoted $\mathcal{D}(ABC)$, to be $\mathcal{D}(ABC) = 2R - \angle A - \angle B - \angle C$.

Legendre's theorem states that $\mathcal{D}(ABC) \geq 0$ for any triangle. Let us investigate what other properties the defect of triangles may have.

Theorem 1.1.66 *If $P \in (BC)$, where $[BC]$ is a side of the triangle $\triangle ABC$, then $\mathcal{D}(APB) + \mathcal{D}(APC) = \mathcal{D}(ABC)$.*

Proof (Hint) Denote by $\angle A_1 = \angle BAP,\ \angle A_2 = \angle CAP,\ \angle P_1 = \angle APB,\ \angle P_2 = \angle APC$ and observe that $\angle A_1 + \angle A_2 = \angle A,\ \ \angle P_1 + \angle P_2 = 2R$. Then

$$\mathcal{D}(APB) + \mathcal{D}(APC) = 2R - \angle A_1 - \angle P_1 - \angle B + 2R - \angle A_2 - \angle P_2 - \angle C =$$

$$= 2R - \angle A - \angle B - \angle C = \mathcal{D}(ABC).$$

□

Theorem 1.1.67 *Consider a triangle $\triangle ABC$ and two points, $B_1 \in (AB),\ \ C_1 \in (AC)$. Then $\mathcal{D}(AB_1C_1) \leq \mathcal{D}(ABC)$.*

Proof (Hint) Consider the triangles $\triangle AB_1C_1,\ \ \triangle BB_1C_1,\ \ \triangle BCC_1$ and apply the previous theorem as follows: $\mathcal{D}(ABC) = \mathcal{D}(ABC_1) + \mathcal{D}(BCC_1) = \mathcal{D}(AB_1C_1) + \mathcal{D}(B_1C_1B) + \mathcal{D}(BCC_1) \geq \mathcal{D}(AB_1C_1)$. □

If "the big triangle" $\triangle ABC$ has $\mathcal{D}(ABC) = 0$, then, mandatory "the small triangle" $\triangle AB_1C_1$ has to fulfill $\mathcal{D}(AB_1C_1) = 0$.

Pay attention to the following construction.
Consider a right-angle triangle $\triangle BAC$, $\angle A = R$ with $\mathcal{D}(BAC) = 0$. In this case observe that $\angle B + \angle C = R$ and construct D in the opposite half-plane with respect to BC and A such that $\triangle ABC \equiv \triangle DBC$. It results in a quadrilateral $ABDC$ such that all angles are equal to R, and the opposite sides are equal, i.e. $[AB] \equiv [CD]$, $[AC] \equiv [BD]$. We may call this figure rectangle and it is easy to discover two more properties. The diagonals AD and BC are congruent and they cut in the middle of each one. Let us rename the rectangle $ABDC$ by $A_{00}A_{10}A_{11}A_{01}$.

We intend to pave the plane with tiles congruent to our created rectangle $A_{00}A_{10}A_{11}A_{01}$ for obtaining a so-called grid.

On the half-lines $A_{00}A_{10}$ we consider the points $A_{20}, A_{30}, ..., A_{n0}, ...$ such that the segment $[AB]$ is seen repeatedly as $[A_{00}A_{10}] \equiv [A_{10}A_{20}] \equiv [A_{20}A_{30}] \equiv \cdots \equiv [A_{(n-1)0}A_{n0}] \equiv \cdots$ and on the half-line $A_{00}A_{01}$ we consider the points A_{02}, $A_{03}, ..., A_{0n}, ...$ such that the segment $[AC]$ is seen repeatedly as $[A_{00}A_{01}] \equiv [A_{01}A_{02}] \equiv [A_{02}A_{03}] \equiv \cdots \equiv [A_{0(n-1)}A_{0n}] \equiv \cdots$.

The tiles we create and put on the first row are consecutively

$$A_{00}A_{10}A_{11}A_{01},\ A_{10}A_{20}A_{21}A_{11},\ A_{20}A_{30}A_{31}A_{21},$$

$$A_{30}A_{40}A_{41}A_{31}, ...,\ A_{n0}A_{(n+1)0}A_{(n+1)1}A_{n1}, ...,$$

on the second row

$$A_{01}A_{11}A_{12}A_{02},\ A_{11}A_{21}A_{22}A_{12},$$

$$A_{21}A_{31}A_{32}A_{22},\ A_{31}A_{41}A_{42}A_{32}, ...,\ A_{n1}A_{(n+1)1}A_{(n+1)2}A_{n2}, ..., etc.$$

A "general" tile in this pavement is $A_{kp}A_{(k+1)p}A_{(k+1)(p+1)}A_{k(p+1)}$.

It is easy to see that the points $A_{20},\ A_{11},\ A_{02}$ are collinear.

The same, the points $A_{30},\ A_{21},\ A_{12},\ A_{03}$ and in general

$$A_{n0},\ A_{(n-1)1},\ A_{(n-2)2}, ...,\ A_{2(n-2)},\ A_{1(n-1)},\ A_{0n}$$

are collinear points.

And it is also easy to observe that all triangles $A_{0n}A_{00}A_{n0}$ have the sum of angles equal to $2R$, i.e. $\mathcal{D}(A_{0n}A_{00}A_{n0}) = 0$. We are prepared to prove a very important theorem.

Theorem 1.1.68 *If there exists a right-angle triangle with defect* 0*, then all right-angle triangles have defect* 0*, i.e. the sum of their angles is* $2R$*.*

Proof Consider the right-angle triangle to be $\triangle BAC$, $\angle A = R$ and $\mathcal{D}(BAC) = 0$ and let a general right-angle triangle $\triangle EFG$, $\angle F = R$. According to Archimedes' axiom it exists $m \in \mathbb{N}$, $n \in \mathbb{N}$ such that $m \cdot AB > FE,\ n \cdot CA > FG$. Without loosing the generality we suppose $n > m$. Then the triangle $\triangle EFG$ can be "arranged" such as $F = A_{00},\ E \in (A_{00}A_{0n}),\ G \in (A_{00}A_{n0})$. According to the previous theory $0 \leq \mathcal{D}(EFG) \leq \mathcal{D}(A_{0n}A_{00}A_{n0}) = 0$, i.e. $\mathcal{D}(EFG) = 0$. □

Theorem 1.1.69 *If there exists a triangle with defect* 0*, then all triangles have defect* 0*, i.e. the sum of their angles is* $2R$*.*

Proof Consider a triangle with defect 0, say $\triangle ABC$. It exists an altitude which intersects the opposite side in its interior, say $AT,\ T \in (BC)$. The altitude and the sides of the triangle determine two right-angle triangles, $\triangle ABT$ and $\triangle ATC$, both with 0 defect, because the initial triangle is with 0 defect. The previous theorem asserts that all right-angle triangles have 0 defect. Now, consider an arbitrary triangle $\triangle DEF$. Suppose that the altitude which intersects the opposite side is $DP,\ P \in (EF)$. The two right-angle triangles $\triangle DFP$ and $\triangle DEP$ are with 0 defect, therefore the defect of the triangle $\triangle DEF$ is 0. □

There are only two situations that can happen. All the triangles have the sum of angles $2R$ or all the triangles have the sum of angles strictly less than $2R$. In the given context we cannot decide about the sum. The next axioms will clarify this aspect.

Definition 1.1.70 The collection of all properties and results deduced from all axioms of incidence, order, congruence, and continuity above is called an *absolute geometry.*

1.2 From Absolute Geometry to Euclidean Geometry Through Euclidean Parallelism Axiom

The question we asked before is how many non-secant lines can be constructed through an exterior point to a given line. We know that at least one can be constructed. Could there be two or more? The standard *Euclidean Parallelism Axiom* is stated as follows.

Axiom P: Given any line in a plane and given any point not incident to the given line, there exists at most one line that passes through the given and it is non-secant to the given line.

Definition 1.2.1 The collection of all properties and results deduced from all axioms of incidence, order, congruence, continuity, and Euclidean parallelism above is called as *Euclidean geometry.*

A direct consequence: In Euclidean geometry, i.e. in the axiomatic frame created by the axioms of incidence, order, congruence, continuity, and the Euclidean parallelism axiom there exists an unique line that passes through a given point A and it is

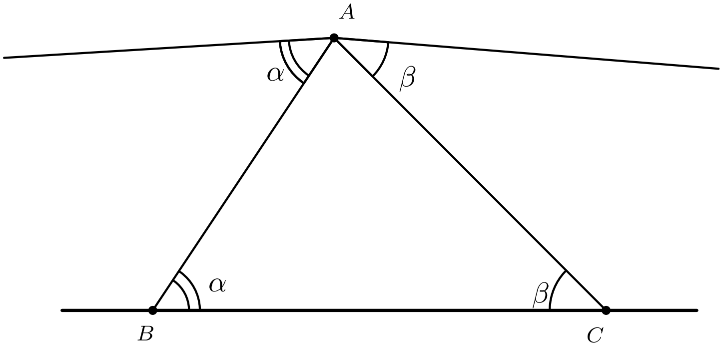


Fig. 1.9 Axiom of Euclidean parallelism

non-secant to the given line d. This unique non-secant line is called the *parallel line to the given line d* through the given point A (Fig. 1.9).

Theorem 1.2.2 *In Euclidean geometry, the sum of angles of any triangle is* $2R$.

Proof Consider a triangle $\triangle ABC$ and the unique parallel through A to BC. We have in mind the figure of Theorem 23 where the parallel was FE with $A \in FE$. The interior alternate angles $\angle EAC$ and $\angle ACB$ are equal. The same for the interior alternate angles $\angle FAB$ and $\angle ABC$. Therefore, if we look at the angle $\angle FAE$ we observe that it is equal to the sum of the angles of the triangle $\triangle ABC$ and, in the same time, it has the value $2R$. Since this particular triangle has the sum of angles equal to $2R$, all other triangles have the sum of angles equal to $2R$. □

In the case of the figure, $\angle A + \angle B + \angle C = \angle A + \alpha + \beta = 2R$.

Since $\angle A + \angle B + \angle C = 2R$, we deduce $2R - \angle A = \angle B + \angle C$. But $2R - \angle A$ is the value of the exterior angle A.

Corollary 1.2.3 (Exterior Angle Theorem in the Euclidean Geometry) *For every triangle, each exterior angle is the sum of the interior non-adjacent angles.*

Theorem 1.2.4 *If it exists a triangle with the sum of its angles equal to* $2R$ *the parallelism axiom is satisfied.*

Proof Assume that there exists a triangle $\triangle ABC$ with $\angle A + \angle B + \angle C = 2R$.

Therefore every triangle has the same property.

Let d be a line and M a point not on d. Let MN be the perpendicular line from M to d, $N \in d$. Let l be a perpendicular to MN in M. We know that l and d are non-secant lines. We have to prove that l is the only non-secant line through M to d.

Consider another line l' passing through M and denote by α the acute angle between MN and l'.

It makes sense to consider a triangle $M'N'P'$ such that

$$[MN] \equiv [M'N'],\ \angle M'N'P' = R,\ \angle N'M'P' = \alpha$$

and $\angle M'P'N' = R - \alpha$ (without to know that $R, \alpha, R - \alpha$ are the angles of a triangle, we do not know that only the angles R and α together with the side $M'N'$ determine a triangle).

Considering $P \in d$ with $[NP] \equiv [N'P']$, the triangles $\triangle MNP$ and $\triangle M'N'P'$ are congruent and one of the half-line of l' is coincident to MP. Therefore $l' \cap d = \{P\}$, i.e. l is the unique non-secant line through M to d. □

The story of Euclidean Geometry may continue with many theorems which can be proven only in this axiomatic frame. But we are interested in introducing non-Euclidean parallelism and models of non-Euclidean geometry. Therefore we remain in the axiomatic frame corresponding to the axioms of incidence, order, congruence, and continuity and, at this moment, we add another axiom, more specific the denial of the Axiom of Euclidean parallelism.

1.3 From Absolute Geometry to Non-Euclidean Geometry Through Non-Euclidean Parallelism Axiom

The Euclidean parallelism axiom, in the set theoretical language, can be written as

$$\forall d,\ \forall A \notin d,\ \#\{a \mid A \in a,\ a \cap d = \emptyset\} \leq 1,$$

where # denotes the number of elements of a set. In what follows we assume the negation of the previous axiom, and we call this the axiom of non-Euclidean parallelism. In set theory language, this translates to

$$\exists d_0,\ \exists A_0 \notin d_0,\ \#\{a \mid A_0 \in a,\ a \cap d_0 = \emptyset\} \geq 1.$$

Therefore the axiom of non-Euclidean parallelism is the following.

Axiom PN_0 : There exist both a line d_0 and a point A_0 exterior to d_0 with the property: at least two non-secant lines to d_0 passing through A_0 exist.

Definition 1.3.1 The collection of all properties and results deduced from all axioms of incidence, order, congruence, continuity, and non-Euclidean parallelism above is called a *non-Euclidean geometry.*

We study below some important results in the context of this geometry.

Theorem 1.3.2 *Axiom PN_0 acts as a global property, i.e. it holds for any line and any exterior point.*

Proof By contradiction, we assume that there is a point A and a line d which do not satisfy the property “there are at least two non-secant lines to d, passing through A”. Then, through A passes exactly one non-secant line to d. So, the Axiom P is satisfied for the pair (A, d). If we choose B and C on d, it is easy to see that the sum of angles of the triangle $\triangle ABC$ is $2R$, and this is equivalent as we saw before with Axiom P for all pairs $(M, l),\ M \notin l$, in collision with our assumption. □

We can restate the axiom of non-Euclidean geometry as follows.

Axiom PN: Given a line and a point exterior to the line, there exists at least two non-secant lines to the given line.

It is easy to prove

Theorem 1.3.3 *Let l be a given line in a plane and A be an exterior point to l. Let a_1 and a_2 be two lines in the same plane which pass through A and are non-secant to l. Then every line a passing through A and included in the interior of the angle $\angle a_1 a_2$ is non-secant to l.*

Proof (Hint) If a intersects l, then it does intersect a_1 or a_2, in collision with the fact that a is included in the interior of the angle $\angle a_1 a_2$. □

Corollary 1.3.4 *In non-Euclidean geometry there are an infinite number of non-secant lines to a given line through an exterior point.*

Theorem 1.3.5 *In a geometry which satisfies the groups of axioms of incidence, order, congruence, continuity, and the Axiom NP, the sum of angles of a triangle is less than $2R$.*

Proof (Hint) If it exists a triangle with the sum of angles equal to $2R$, then Axiom P is valid, contradiction. □

We conclude that the non-Euclidean geometry established by the absolute geometry together with the Axiom NP is completely different than the Euclidean geometry established by the absolute geometry together with Axiom P. More other interesting results may be found in both geometries, but in the following, we are interested in offering examples of models of Euclidean and non-Euclidean geometries.

Chapter 2
Basic Facts in Euclidean and Minkowski Plane Geometry

Entia non sunt multiplicanda praeter necessitatem.

W. Ockham

In Chap. 1, we found out that there exist different geometries in a plane. It depends on the axioms one chooses if Euclidean Geometry or Non-Euclidean Geometry is described. But how these geometries look like? In this chapter we present an algebraic model for Euclidean Geometry discussing some important theorems. We obtain a visual representation for the Euclidean Geometry of the plane. Making small changes in the algebraic construction of the Euclidean Geometry, it is possible to construct a Minkowski Geometry. This Geometry is deeply involved both with Physics and with Non-Euclidean Geometry. Later, we see how the models of Non-Euclidean Geometry are connected between them and how a Minkowski one is among them. The geometric objects in Minkowski Geometry seem to have a non-intuitive look, but the main theorems have a similar look with their Euclidean counterparts. Generally, Non-Euclidean models are more sophisticated and we need more mathematical tools in order to build them. This happens in the following chapters. One more comment: this chapter is not as formal as the previous one where we used the language style of an axiomatic theory. We can relax a little bit the mathematical language structure. The definitions appear often written as part of a mathematical algebraic description of geometric objects and italic letters are used to indicate them. The following notation is used: $A := B$. It means that the object A from the left side of the equality is described by definition through the object expression B from the right part of the equality. The word iff has the meaning of "if and only if".

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_2

2.1 Pythagoras Theorems in Euclidean Plane

The idea to consider a system of coordinates on a line was discussed in the previous chapter.

The coordinates are real numbers and their set, geometrically represented as a line, is denoted by $\mathbb{R}$. In the following we suppose known

- the set of natural numbers denoted by $\mathbb{N}$,
- the set of integers denoted by $\mathbb{Z}$,
- the set of rational numbers denoted by $\mathbb{Q}$, and
- the set of irrational numbers denoted by $\mathbb{R} - \mathbb{Q}$.

In the same time we have proved that the values of angles are real numbers.

Basic facts about matrix theory, groups, vector spaces, trigonometric, exponential, and logarithms functions are suppose known by the reader interested in the topic of this book.

When we are talking about a model of Euclidean Geometry in a plane, we have to start from the vector space $\mathbb{R}^2$ over the field $\mathbb{R}$. $x := (x_1, x_2)$, $y := (y_1, y_2)$ are called *vectors*. The vector space operations are $x + y := (x_1 + y_1, x_2 + y_2)$ and $\lambda x := (\lambda x_1, \lambda x_2)$.

The *Euclidean inner product of the vectors x and y* is defined by

$$\langle x, y \rangle := x_1 y_1 + x_2 y_2$$

and the *norm of* x, by $|x| := \sqrt{\langle x, x \rangle} = \sqrt{x_1^2 + x_2^2}$.

Two vectors are called *Euclidean orthogonal (or Euclidean perpendicular)* if their inner product is null.

According to the operations, the vector (x_1, x_2) can be thought as $x_1(1, 0) + x_2(0, 1)$, that is $(x_1, x_2) = x_1(1, 0) + x_2(0, 1) = x_1 e_1 + x_2 e_2$ so, the pair (x_1, x_2) can be seen also as a pair of coordinates of a point A of the plane.

The line determined by $xe_1,\ x \in \mathbb{R}$ is called the x-axis, and the line determined by $ye_2,\ y \in \mathbb{R}$ is called the *y-axis*.

Therefore, in the system of coordinates generated by the orthogonal vectors $e_1 = (1, 0)$; $e_2 = (0, 1)$, the geometric meaning of the vector $x = (x_1, x_2)$ is the oriented segment $\vec{OA}$, lying from the origin O with the coordinates $(0, 0)$ and the endpoint A with the coordinates (x_1, x_2).

Let us consider the 2×2 *rotation matrix* $A_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}$ in which the basic trigonometric function sine and cosine are involved as components.

We define $A_\alpha x := \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$A_\alpha x$, $A_\alpha y$ are two matrices with two lines and one column, and it makes sense to consider the inner product $\langle A_\alpha x, A_\alpha y \rangle$ by adding, after multiplying, the corresponding first and second components, i.e.

$$\langle A_\alpha x, A_\alpha y\rangle =$$

$$= (x_1 \cos\alpha - x_2 \sin\alpha)(y_1 \cos\alpha - y_2 \sin\alpha) + (x_1 \sin\alpha + x_2 \cos\alpha)(y_1 \sin\alpha + y_2 \cos\alpha).$$

Exercise 2.1.1 $\langle A_\alpha x, A_\alpha y\rangle = \langle x, y\rangle$.

Hint. Use $\sin^2\alpha + \cos^2\alpha = 1$.

Exercise 2.1.2 $|A_\alpha x| = |x|$.

Exercise 2.1.3 *If* $|x| = 1$, *then* $\langle A_\alpha x, x\rangle = \cos\alpha$.

If $|x| = 1$, then $|A_\alpha x| = |x| = 1$. Denote by u the unitary vector $A_\alpha x$. The previous relation for the unitary vectors u, x can be written in the form $\langle u, x\rangle = \cos\alpha$.

We can see the vector u as the rotation of the vector x, so the angle between these two vectors is α. For two arbitrary vectors a, b, the vectors $\frac{a}{|a|}, \frac{b}{|b|}$ are unitary and the previous relation becomes

$$\left\langle \frac{a}{|a|}, \frac{b}{|b|} \right\rangle = \cos\alpha,$$

i.e.

$$\frac{\langle a, b\rangle}{|a||b|} = \cos\alpha.$$

This last formula is known as the *Generalized Pythagoras Theorem*. Let us discuss why (Fig. 2.1).

Since we have the vectors $a = (a_1, a_2)$, $b = (b_1, b_2)$, we can think about the triangle OAB as the triangle determined by the points $O(0, 0)$, $A(a_1, a_2)$, $B(b_1, b_2)$.

Before continuing, we point out the meaning of the Euclidean Parallelism in this coordinate frame.

Let us consider $M(m_1, m_2)$ and $N(n_1, n_2)$.

Definition 2.1.4 The *lines* AB *and* MN *are Euclidean parallel* and we denote this by $MN||AB$, if the vectors

$$\vec{OD} = (m_1 - n_1, m_2 - n_2)$$

and

$$\vec{OC} = (a_1 - b_1, a_2 - b_2)$$

are collinear, i.e. $\exists\beta \neq 0$ such that $(m_1 - n_1, m_2 - n_2) = \beta(a_1 - b_1, a_2 - b_2)$.

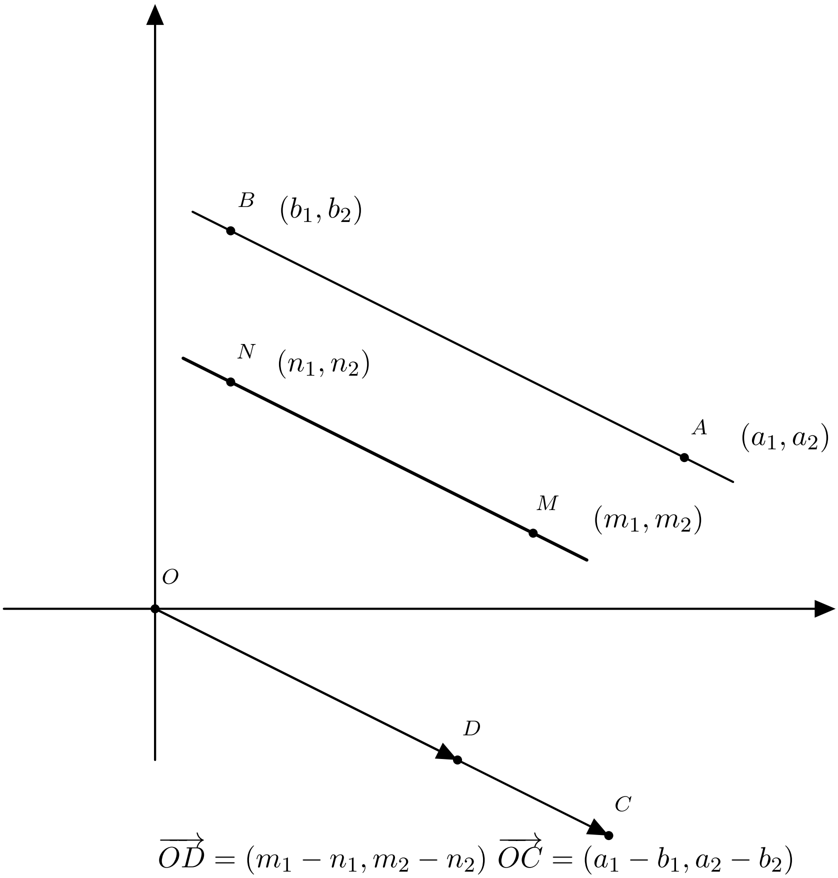


Fig. 2.1 Parallel lines seen through vector properties

The *Generalized Pythagoras Theorem* in AOB asserts

$$|AB|^2 = |OA|^2 + |OB|^2 - 2|OA||OB| \cos \alpha,$$

where $|OA| = |\overrightarrow{OA}| = |a|$,

$$|AB| = |\overrightarrow{AB}| = |OC| = |a - b| = \sqrt{\langle a - b, a - b\rangle} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2},$$

and $\angle AOB = \alpha$.

The formula explained and written above

$$|AB| := \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}$$

is called *the Euclidean distance between the two points* $A(a_1, a_2)$, $B(b_1, b_2)$ of the plane.

Theorem 2.1.5 *In the previous notations, it is*

$$|AB|^2 = |OA|^2 + |OB|^2 - 2|OA||OB|\cos\alpha,$$

iff

$$\frac{\langle a, b\rangle}{|a||b|} = \cos\alpha.$$

Proof We observe that we need to prove only that

$$2|OA||OB|\cos\alpha = |OA|^2 + |OB|^2 - |AB|^2$$

is the same as

$$2|a||b| = 2\langle a, b\rangle .$$

Or, this means

$$|OA|^2 + |OB|^2 - |AB|^2 = 2\left\langle \vec{OA}, \vec{OB}\right\rangle,$$

and, in coordinates, this becomes a quick computation for the reader, that is,

$$(a_1^2 + a_2^2) + (b_1^2 + b_2^2) - \left((a_1 - b_1)^2 + (a_2 - b_2)^2\right) = 2(a_1 b_1 + a_2 b_2)^2.$$

Corollary 2.1.6 *If* $\left\langle \vec{OA}, \vec{OB}\right\rangle = 0$*, i.e. the vectors a and b are orthogonal (Euclidean perpendicular), then we obtain the standard Pythagoras' theorem.*

The side AB is called a *hypotenuse*, and OA, OB are called *legs* of the triangle OAB.

Theorem 2.1.7 (Pythagoras' Theorem) *In the previous notations, it is*

$$|AB|^2 = |OA|^2 + |OB|^2.$$

The angle corresponding to orthogonal vectors is described by the condition $\cos\alpha = 0$, that is, its measure is $\alpha = \frac{\pi}{2}$.

Therefore $\frac{\pi}{2}$ is the value of the right angle R. The sum of angles of a triangle in Euclidean Geometry becomes $\angle A + \angle B + \angle C = \pi$.

Theorem 2.1.8 (Thales Theorem) *Let us consider* $O(0, 0)$, $A(x_1, x_2)$, $B(y_1, y_2)$, $A_1(\mu x_1, \mu x_2)$, $B_1(\lambda y_1, \lambda y_2)$. *Then,* $AB || A_1 B_1$ *iff* $\lambda = \mu$.

Proof In coordinates $\vec{AB} = (y_1 - x_1, y_2 - x_2)$ and $\vec{A_1B_1} = (\lambda y_1 - \mu x_1, \lambda y_2 - \mu x_2)$. The parallelism between AB and A_1B_1 is equivalent to: $\exists \beta$ such that $(\lambda y_1 - \mu x_1, \lambda y_2 - \mu x_2) = \beta(y_1 - x_1, y_2 - x_2)$. Therefore

$$(\lambda - \beta)y_1 - (\mu - \beta)x_1 = 0,$$

$$(\lambda - \beta)y_2 - (\mu - \beta)x_2 = 0$$

for arbitrary x_1, x_2, y_1, y_2, that is, $AB||A_1B_1$ iff $\lambda = \mu$.

Thales theorem can be written in the form:
Consider the triangle OAB, $A_1 \in OA$, $B_1 \in OB$.
Then $AB||A_1B_1$ iff $\dfrac{|OA|}{|OA_1|} = \dfrac{|OB|}{|OB_1|}$.

Problema 2.1.9 *Consider the triangle $OAB, A_1 \in OA$, $B_1 \in OB$. Then $AB||A_1B_1$ iff* $\dfrac{|OA|}{|OA_1|} = \dfrac{|OB|}{|OB_1|} = \dfrac{|AB|}{|A_1B_1|}$.

Hint. Construct a parallel from B to OA, denote by X the point of intersection between the parallel and A_1B_1, apply Thales Theorem in the form $\dfrac{|B_1X|}{|B_1A_1|} = \dfrac{|B_1B|}{|B_1O|}$ and use the properties of proportions. □

It is not very difficult to express line equations in the Euclidean plane.

If the line d passes through $A(a_1, a_2)$; $B(b_1, b_2)$ the *equation of d* is

$$y - a_2 = \frac{a_2 - b_2}{a_1 - b_1}(x - a_1).$$

The ratio denoted by m, $m := \dfrac{a_2 - b_2}{a_1 - b_1}$ is called a *slope* for the line d. The slope m has the value $m = \tan\alpha$, where α is the angle between the Ox and d in this order.

Exercise 2.1.10 *Show that two lines d_1 and d_2 are Euclidean perpendicular iff $m_1m_2 = -1$.*

Hint. Use Euclidean exterior angle theorem and $\tan(\alpha + \beta)$ formula.

The *equation of a circle centred at (a_1, a_2) with radius r* is expressed with respect to the Euclidean distance between the centre and a point (x, y) on the circle:

$$(x - a_1)^2 + (y - a_2)^2 = r^2.$$

The *interior of a circle C* is denoted by $intC$ and, between the two regions in which a circle divides the plane, it is the region containing its centre. The Euclidean distance between the centre and a point belonging to this region is less than the radius. The complementary region is called the *exterior of the circle*. The Euclidean distance between the centre and a point belonging to this region is greater than the radius.

There are many properties related to circles and lines attached to triangles in the Euclidean Geometry. Some of them will be studied in the next chapter. The Euclidean plane is denoted by E^2.

2.2 Space-Like, Time-Like, and Null Vectors in Minkowski Plane

When we are talking about a model of Minkowski Geometry in a plane, we have to start from the same vector space $\mathbb{R}^2$ over the field $\mathbb{R}$. Here, $x = (x_1, x_2),\ y = (y_1, y_2)$ are called vectors, as in the Euclidean case.

The vector space operations are the same $x + y := (x_1 + y_1, x_2 + y_2)$ and $\lambda x := (\lambda x_1, \lambda x_2)$.

The *Minkowski product of the vectors* x *and* y is defined by

$$\langle x, y\rangle_M := x_1 y_1 - x_2 y_2$$

and the *Minkowski norm of* x by $|x|_M := \sqrt{|\langle x, x\rangle|_M} = \sqrt{|x_1^2 - x_2^2|}$.

Two vectors are called *Minkowski orthogonal* if their Minkowski product is null. In a system of coordinates generated by the Minkowski orthogonal vectors $e_1 = (1, 0)$; $e_2 = (0, 1)$, the geometric meaning of the vector $x = (x_1, x_2)$ is the oriented segment $\vec{OA}$, lying from the origin O with the coordinates $(0, 0)$ and the endpoint A with the coordinates (x_1, x_2). This is exact as in the Euclidean case.

Even the parallelism is like in the Euclidean case. Consider $M(m_1, m_2)$ and $N(n_1, n_2)$.

Definition 2.2.1 The lines AB and MN are parallel and we denote this by $MN||AB$, if the vectors $\vec{OD}= (m_1 - n_1, m_2 - n_2)$ and $\vec{OC}= (a_1 - b_1, a_2 - b_2)$ are collinear, i.e. $\exists\beta \neq 0$ such that $(m_1 - n_1, m_2 - n_2) = \beta(a_1 - b_1, a_2 - b_2)$.

However, in a Minkowski space, we have three different kind of vectors $\vec{OA}$. Let us explain. There are space-like vectors, time-like vectors, and null vectors.

A vector x is a *space-like vector* if $\langle x, x\rangle_M < 0$.

Examples are $b = (-1, 2)$, $e = (2, -3)$, or in general $a = (a_1, a_2)$ with $|a_1| < |a_2|$.

A vector x is a *time-like vector* if $\langle x, x\rangle_M > 0$.

Examples are $b = (3, 2)$, $e = (-4, -3)$, or in general $a = (a_1, a_2)$ with $|a_1| > |a_2|$.

A vector x is a *null vector* if $\langle x, x\rangle_M = 0$.

Examples are $b = (-1, 1), e = (2, 2)$, or in general $a = (a_1, a_2)$ with $|a_1| = |a_2|$.

The reader can observe that Minkowski orthogonal vectors have to be pairs, one space-like and one time-like. An example: $x = (x_1, x_2)$; $v = (kx_2, kx_1)$.

Consider the 2×2 "hyperbolic rotation" matrix $A_\alpha = \begin{pmatrix} \cosh \alpha & \sinh \alpha \\ \sinh \alpha & \cosh \alpha \end{pmatrix}$ in which the basic hyperbolic trigonometric functions sine and cosine are involved as components.

$$\sinh \alpha = \frac{e^\alpha - e^{-\alpha}}{2}, \ \cosh \alpha = \frac{e^\alpha + e^{-\alpha}}{2}.$$

This matrix is called a *hyperbolic rotation* and this name is legitimate by the next quick exercises.

As in the Euclidean case, $A_\alpha x$, $A_\alpha y$ are two matrices with two lines and one column, and it makes sense to consider the Minkowski product $\langle A_\alpha x, A_\alpha y\rangle_M$.

Exercise 2.2.2 $\langle A_\alpha x, A_\alpha y\rangle_M = \langle x, y\rangle_M$.

Hint. Use $\cosh^2 \alpha - \sinh^2 \alpha = 1$. Therefore

$$\langle A_\alpha x, A_\alpha y\rangle_M = (x_1 \cosh \alpha + x_2 \sinh \alpha)(y_1 \cosh \alpha + y_2 \sinh \alpha) -$$

$$- (x_1 \sinh \alpha + x_2 \cosh \alpha)(y_1 \sinh \alpha + y_2 \cosh \alpha) =$$

$$= x_1 y_1 - x_2 y_2 = \langle x, y\rangle_M$$

We leave the reader to prove that it does not exist α such that $A_\alpha e_1 = e_2$.

Or, more general, after rotating a time-like (space-like) vector we cannot obtain a space-like (time-like) vector. As we will see below, this property is related to causality in Relativity.

Exercise 2.2.3 $|A_\alpha x|_M = |x|_M$.

So, the matrices A_α preserve the Minkowski type and the Minkowski length of vectors.

Definition 2.2.4 If $u = A_\alpha v$ we say that α is the *oriented hyperbolic angle* between v and u. Obviously, $-\alpha$ is the oriented hyperbolic angle between u and v.

Next, we discuss about time-like vector properties.

A *future-pointing time-like vector* v fulfills the property $\langle v, e_1\rangle_M > 0$. An example is $v = (3, 2)$.

Otherwise the vector v is a *past-pointing space-like vector*. $v = (-3, -2)$ is an example, and the reader can observe that if we consider the lines $d_1 : \ x_2 = x_1$ and $d_2 : \ x_2 = -x_1$ which describe the *null cone*, a future-pointing time-like vector is a vector $v = \overrightarrow{OA}$ with $A = (a_1, a_2)$ included in the interior of the angle $\angle d_1 d_2$ (i.e. $|a_1| > |a_2|$) such that $a_1 > 0$.

Exercise 2.2.5 *If v is a future-pointing time-like vector, then $A_\alpha v$ is a future-pointing time-like vector.*

Hint. Since we have proved that the time-like property is kept after a hyperbolic rotation, it remains to prove that the future-pointing property is preserved.

Or $\langle A_\alpha v, e_1\rangle_M = a_1 \cosh\alpha + a_2 \sinh\alpha$.

We have $|a_1| > |a_2|$ and $|\sinh\alpha| < \cosh\alpha$, i.e. $|a_1\cosh\alpha| > |a_2\sinh\alpha|$.

It remains to observe that there are triangles in this Minkowski Geometry in which the meaning of angle does not exist. The triangles in which we can discuss about angles are called pure triangles, i.e. in such triangles all the sides are time vectors, all pointing towards the future (or, all pointing towards the past). How we can create such a triangle? We start with two, say, future-pointing time-like vectors, $x = (x_1, x_2)$, $y = (ky_1, ky_2)$ and we choose $k > 0$ such that $y - x = (ky_1 - x_1, ky_2 - x_2)$ is future-pointing time-like vector.

Exercise 2.2.6 *If x and y are future-pointing time-like vectors then*
1. $\langle x, y\rangle_M > 0$;
2. $x + y$ *is a future-pointing time-like vector;*
3. $\langle x, y\rangle_M \le |x|_M|y|_M$, *where the equality happens iff* $y = kx$;
4. $|x + y|_M \le |x|_M + |y|_M$, *the equality happens iff* $y = kx$.

2.3 Minkowski–Pythagoras Theorems

Let us start with a simple exercise.

Exercise 2.3.1 *If x is a space-like vector such that $|x|_M = 1$ then $\langle A_\alpha x, x\rangle_M = -\cosh\alpha$.*

Hint. $\langle A_\alpha x, x\rangle_M = (x_1\cosh\alpha + x_2\sinh\alpha)x_1 - (x_1\sinh\alpha + x_2\cosh\alpha)x_2 = (x_1^2 - x_2^2)\cosh\alpha = -\cosh\alpha$.

Denote by u the unitary vector $A_\alpha x$. The previous relation for the unitary vectors u, x can be written in the form $\langle u, x\rangle_M = -\cosh\alpha$.

For two arbitrary future-pointing space-like vectors a, b, the vectors $\dfrac{a}{|a|_M}, \dfrac{b}{|b|_M}$ are unitary and the previous relation becomes

$$\left\langle \frac{a}{|a|_M}, \frac{b}{|b|_M} \right\rangle_M = -\cosh\alpha,$$

i.e.

$$\frac{\langle a, b\rangle_M}{|a|_M|b|_M} = -\cosh\alpha.$$

According to the Euclidean case, this last formula can be called the *Generalized Minkowski–Pythagoras theorem.*

Consider a *Minkowski right triangle* OAB, i.e. a triangle such that the vectors $\vec{OA}$ and $\vec{OB}$ are Minkowski orthogonal, that is, $\left\langle \vec{OA}, \vec{OB} \right\rangle_M = 0$. The side AB is called a *Minkowski hypotenuse*, and OA, OB are called *Minkowski legs* of the triangle OAB.

An example is given by $\vec{OA}= (0, a),\ \ a > 0;\ \ \vec{OB}= (b, 0),\ \ b > 0$. When we consider the vector $\vec{AB}= (b, -a)$ it depends on the absolute values $|a|, b$ if this vector is a time-like vector, a space-like vector, or a null vector.

So, the Minkowski–Pythagoras theorem asserts that "*in a Minkowski right triangle, the square of the Minkowski hypotenuse is the difference of the square of Minkowski legs*."

The endpoints of unitary space-like vectors determine a Minkowski circle. The equation of this circle is $x^2 - y^2 = -1$. From the Euclidean point of view this is a hyperbola equation.

The endpoints of unitary time-like vectors determine a Minkowski circle, too. The equation of this circle is $x^2 - y^2 = 1$.

Exercise 2.3.2 *What kind of triangle is determined by three arbitrary points of the Minkowski circle* $x^2 - y^2 = -1$*?*

The answer is: a *pure time-like triangle*, i.e. a triangle in which each side is a time-like vector pointing the future (or all three pointing the past).

There are a lot of nice geometric properties for Minkowski circles, some of them similar to Euclidean properties. For our purpose the facts highlighted above are enough to continue. The Minkowski plane is denoted by M^2.

Chapter 3
From Projective Geometry to Poincaré Disk. How to Carry Out a Non-Euclidean Geometry Model

Virtus unita fortior agit.

This chapter is devoted to a first model of Non-Euclidean Geometry. To construct this model, we need to deal with one of the most important transformations of the Euclidean plane, the geometric inversion. We still need some other acquirements, therefore we meet the Projective Geometry. An invariant described by a special projective map of a circle allows us to construct a non-Euclidean distance inside the disk. Elaborating the previous model we highlight the Poincaré disk model.

3.1 Geometric Inversion and Its Properties

The geometric inversion is a classical transformation of elementary Euclidean Plane Geometry. To describe it, let us consider a circle centred at O and radius R, denoted $C(O, R)$.

A *geometric inversion of centre* $O \in E^2$ *and radius* R maps each point $M \in E^2$, $M \neq O$ to the point N on the radius OM such that $|OM| \cdot |ON| = R^2$, where $|OM|$ is the Euclidean length between the points O and M.

The circle $C(O, R)$ is called *circle of inversion*.

The points M *and* N *are called homologous inverse points* with respect to the previous geometric inversion determined by the circle of inversion $C(O, R)$. R^2 is called *power of inversion*.

We prefer to use "inversion of centre $O \in E^2$ and power R^2" instead of "inversion of centre $O \in E^2$ and radius R".

Suppose we know the homologous inverse M and N with respect to a geometric inversion having O as a centre and R^2 as a power and the order of points on the radius

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_3

OM is O, M, N, or O, N, M, i.e. $O \notin (MN)$. This is a *direct inversion*, when the oriented segments OM and ON have the same direction.

Now choose N' the symmetric of N with respect to O, i.e. $|ON'| = |ON|$. We have N' on the radius OM, $|OM| \cdot |ON'| = R^2$ and $O \in (MN')$.

Therefore, for the inverse N of a point M, with respect to a given inversion, we have two possibilities:

(1) O does not belong to the segment (MN),
(2) O belongs to the segment (MN).

To conclude, when we are talking about an inversion and the inverse N belongs to the radius OM, we have to specify if it is the *direct geometric inversion* i.e. we are talking about the map

$$T_{O;R^2} : E^2 - \{O\} \to E^2 - \{O\},\ T_{O;R^2}(M) = N,\ O \notin (MN),\ |OM| \cdot |ON| = R^2,$$

or we are talking about the map

$$T^s_{O;R^2} : E^2 - \{O\} \to E^2 - \{O\},\ T^s_{O;R^2}(M) = N,\ O \in (MN),\ |OM| \cdot |ON| = R^2,$$

which can be called *symmetric geometric inversion*.

All next results are done for the direct geometric inversion, that is for the map $T_{O;R^2}$. All the properties obtained can be easily transferred by symmetry with respect O for the symmetric geometric inversion. When in a problem we use an inversion, the reader finds the information if it is a direct one or symmetric one looking at the map involved, i.e. $T_{O;R^2}$ or $T^s_{O;R^2}$.

The main *properties of the direct geometric inversion* are:

1. If $T_{O;R^2}(M) = N$, then $T_{O;R^2}(N) = M$.

In simple words, if N is the inverse of M, then M is the inverse of N.

That is,

$$T_{O;R^2}(T_{O;R^2}(M)) = M.$$

This property can be written in a simpler form as

$$T^2 = id_{E^2-\{O\}}$$

and it highlights that the geometric inversion is an idempotent transformation.

2. $T_{O;R^2}(C(O, R)) = C(O, R)$, that is the circle of inversion is invariant under the inversion it generates.

3. A line d which passes through the pole of inversion is invariant under inversion, i.e.

$$T_{O;R^2}(d - \{O\}) = d - \{O\}.$$

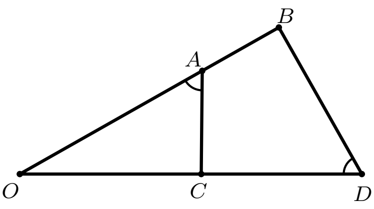


Fig. 3.1 Inversion main figure

Before continuing, some notions are needed. A *cyclic quadrilateral* $ABCD$ is a quadrilateral which vertexes A, B, C, D belong to a circle Γ, called the *circumcircle* of the quadrilateral (Fig. 3.1).

4. If $T_{O;R^2}(A) = B$ and $T_{O;R^2}(C) = D$, then $ABDC$ is a cyclic quadrilateral.

Hint: From $|OA| \cdot |OB| = |OC| \cdot |OD| = R^2$ it results $\dfrac{|OA|}{|OD|} = \dfrac{|OC|}{|OB|}$. Then, triangles $\triangle OAC$ and $\triangle ODB$ are similar, i.e. $\angle OAC = \angle CDB$, that is the quadrilateral $ABDC$ is a cyclic one.

Why the circle of inversion is important? Because it allows us to construct the inverse of a point.

5. Construction of the inverse of a given point.

Suppose M belongs to $intC(O, R)$ (Fig. 3.2).

We consider the radius OM and the perpendicular line to OM in M which intersects the circle at the points S and S'. Next, we refer to S. The tangent at S intersects the radius OM in N.

If we look at the right triangle $\triangle OSN$, $R^2 = |OS|^2 = |OM| \cdot |ON|$, i.e. N is the (direct) inverse of M.

Suppose M is outside the circle of inversion (Fig. 3.3).

We construct the radius OM and one of the tangent to the circle, MS, $S \in C(O, R)$. The perpendicular from S to OM intersects OM in N. In the right triangle $\triangle OSM$ we have $R^2 = |OS|^2 = |OM| \cdot |ON|$, i.e. N is the inverse of M.

Let us observe that in the above two situations the circle passes through $SMS'N$ is orthogonal to the circle of inversion. If $M \in C(O, R)$, $N = M$, that is the inverse of M is M itself.

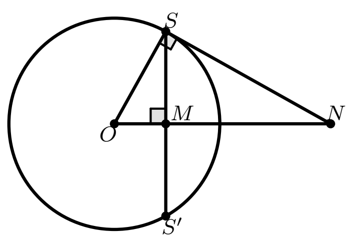


Fig. 3.2 The inverse point of an interior point M

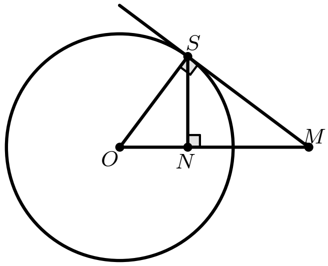


Fig. 3.3 The inverse point of an exterior point M

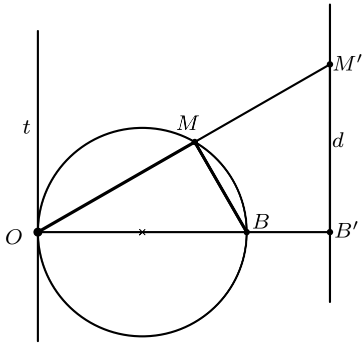


Fig. 3.4 The inverse of a line d such that $O \notin d$

6. Consider a line $d \subset E^2,\ O \notin d$. Then, $T_{O;R^2}(d) = C - \{O\}$, i.e. the inverse of the line d is a circle $C - \{O\}$, such that the tangent line in O to the circle C is parallel to d (Fig. 3.4).

Proof Denote by B' the intersection between d and the perpendicular line from O to d. The inverse of B' is B. Consider a point $M' \in d$ and its inverse M. The quadrilateral $B'BMM'$ is cyclic, therefore $\angle OMB$ is a right one, i.e. when M' belongs to d, M belongs to the circle having $(OB]$ as a diameter. Since the diameter BO is perpendicular to the tangent denoted by t in O to the circle, it results $d \parallel t$. □

7. The inverse of a circle C passing through O is line $d,\ O \notin d, d \parallel t$, where t is the tangent at O to the circle C.

Proof The inversion $T_{O;R^2}$ is an idempotent transformation. If we are looking backward at the previous property of inversion the result is obvious. □

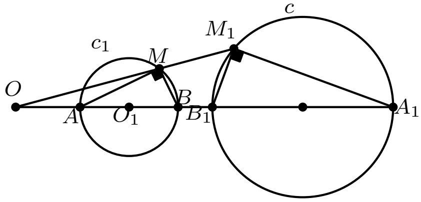
8. The inverse of a circle $C_1,\ O \notin C_1$ is a circle $C_2,\ O \notin C_2$, i.e. $T_{O;R^2}(C_1) = C_2$.

Proof Consider the radius OO_1 where O_1 is the centre of the circle C_1. Denote $\{A, B\} := OO_1 \cap C_1$ and suppose the order of points is O, A, O_1, B (Fig. 3.5).

Consider A_1, B_1 the inverses of A, B respectively. Since $|OA| < |OB|$ and $|OA| \cdot |OA_1| = |OB| \cdot |OB_1| = R^2$ it results $|OA_1| > |OB_1|$. Without losing the generality, we can suppose the order of points on the radius OO_1 is O, A, O_1, B, B_1, A_1. Consider $M \in C_1$ and its inverse M_1.

Using the cyclic quadrilaterals AA_1M_1M and BB_1M_1M it results both $\angle OAM = \angle MM_1A_1$ and $\angle ABM = \angle MM_1B_1$.

Fig. 3.5 Inversion of circles



Since $\angle OAM = \frac{\pi}{2} + \angle MBA = \angle MM_1A_1 = \angle MM_1B_1 + \angle B_1M_1A_1$, we have

$$\angle B_1M_1A_1 = \frac{\pi}{2},$$

that is M_1 belongs to a circle of diameter B_1A_1. □

9. Consider $T_{O,R^2}(A) = A_1,\ T_{O,R^2}(B) = B_1$. Then $|A_1B_1| = R^2 \cdot \frac{|AB|}{|OA| \cdot |OB|}$.

Proof The triangles $\triangle OAB$ and $\triangle OB_1A_1$ are similar, therefore $\frac{|A_1B_1|}{|AB|} = \frac{|OA_1|}{|OB|}$.

It results

$$\frac{|A_1B_1|}{|AB|} = \frac{|OA_1|}{|OB|} \cdot \frac{|OA|}{|OA|},$$

that is $|A_1B_1| = R^2 \cdot \frac{|AB|}{|OA| \cdot |OB|}$. □

10. Orthogonal circles to the circle of inversion are preserved by inversion (Fig. 3.6).

Proof Denote by S, S' the intersection points between the circle of inversion $C(O, R)$ and the orthogonal circle γ. Consider $M, N \in \gamma$ such that O, M, N are collinear points. Since $|OM| \cdot |ON| = |OS|^2 = R^2$, it results $T_{O,R^2}(M) = N$, i.e. $T_{O,R^2}(\gamma) = \gamma$. □

11. The inversion is a *conformal map*, i.e. it preserves the angles between curves.

Fig. 3.6 The inverse of a circle orthogonal to the fundamental circle of inversion

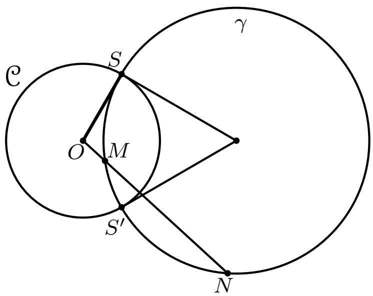
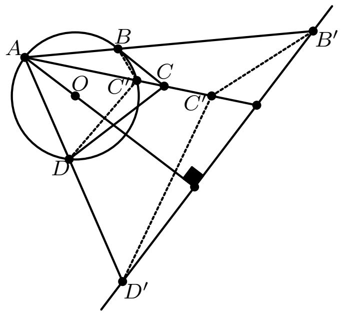


Fig. 3.7 Ptolemy's theorem



Proof The angle between two curves at their point of intersection, S, is the angle between the tangent lines at S to the curves. Let $\Gamma_1, \Gamma_2; \Gamma_1^1, \Gamma_2^2$ be four curves such that $T_{O,R^2}(\Gamma_1) = \Gamma_1^1, T_{O,R^2}(\Gamma_2) = \Gamma_2^2,\ O \notin \Gamma_1 \cup \Gamma_2 \cup \Gamma_1^1 \cup \Gamma_2^2,\ S \in \Gamma_1 \cap \Gamma_2,\ T_{O,R^2}(S) = S',\ T_{O,R^2}(M_1) = M_1^1,\ T_{O,R^2}(N_1) = N_1^1$. The quadrilateral $SM_1M_1^1S'$ and $SM_2M_2^2S'$ are cyclic therefore $\angle M_1SM_2 = \angle M_1^1S'M_2^2$. When radius OM approaches OS the previous angles are still equal. The limit position highlights the previous angles as angles between tangent lines to the curves. □

Examples of problems solved by inversion (Fig. 3.7).

Problema 3.1.1 (*Ptolemy's Theorem*) The products of the lengths of two diagonals of a quadrilateral is less than or equal to the sum of the products of opposite sides and the equality holds if and only if the quadrilateral is a cyclic one.

Solution. (Hint) Consider the inversion of centre A and arbitrary power $k > 0$ and denote by B', C', D' the inverses of the points B, C, D. We have

$$|B'D'| = k \cdot \frac{|BD|}{|AB| \cdot |AD|},\ |B'C'| = k \cdot \frac{|BC|}{|AB| \cdot |AC|},\ |C'D'| = k \cdot \frac{|CD|}{|AC| \cdot |AD|}.$$

Replacing in $|B'D| \le |B'C'| + |C'D'|$, and taking into account that the equality happens when B', C', D' are collinear it results the statement. □

Problema 3.1.2 Consider two pairs of circles, $\gamma_x, \Gamma_x; \gamma_y, \Gamma_y$ which pass through the same point O having the centres on perpendicular axes Ox, Oy. Then the four points of mutual intersection are cyclic.

Solution. (Hint) Consider an inversion of centre O and power $k > 0$, $T_{O,k}$. The circles $\gamma_x, \Gamma_x; \gamma_y, \Gamma_y$ which pass through O are mapped into a rectangle $A'B'C'D'$ whose vertexes comes from A, B, C, D respectively. Since a rectangle allows a circumcircle, by inversion, this circumcircle comes from the circle containing the initial points A, B, C, D. □

Problema 3.1.3 Two circles intersect at A and B. The tangent lines at A to the circles intersect the circles at M and N. Let B_1 be the symmetric of A with respect to B. Prove that the quadrilateral AMB_1N is a cyclic one.

Solution. Consider an inversion of centre A and power $k > 0$, $T_{A,k}$. The three lines AM, AN and AB are transformed after the rule: $T_{A,k}(AM) = AM$, $T_{A,k}(AN) = AN$, $T_{A,k}(AB) = AB$, $T_{A,k}(M) = M_1$, $T_{A,k}(N) = N_1$, $T_{A,k}(B) = B'$. Since the circles passing through A are transformed into lines parallel to the tangents AM and AN it is easy to deduce that the quadrilateral $AM_1B'N_1$ is a parallelogram. The point B_1 is mapped by inversion into B_1' such that $|AB_1| \cdot |AB_1'| = k = 2|AB| \cdot \frac{1}{2}|AB'|$, i.e. B_1' is the centre of the previous parallelogram. Therefore the diagonal M_1N_1 which contains B_1' comes from the inversion of a circle containing the points A, M, B_1, N. □

For the next problem the reader has to know what is an inscribed circle for a given triangle, and the fact that "the lines which connect the vertexes to the opposite tangent points (of the circle with the sides) are concurrent lines". The point of concurrence is called *Gergonne's point*.

Problema 3.1.4 Denote by $C(O, R)$ the circumcircle of the triangle $\triangle ABC$, A_1, B_1, C_1 the midpoints of the sides $[BC]$, $[CA]$, $[AB]$ respectively.

Prove that the circles Γ_{AOA_1}, Γ_{BOB_1}, Γ_{COC_1} have a common point E, $E \neq O$.

Solution. An inversion of centre O and power R^2 preserves A, B, C and $C(O, R)$.

The circles Γ_{AOA_1}, Γ_{BOB_1}, Γ_{COC_1} are mapped into lines passing through A, B, C respectively.

$T_{O,R^2}(A_1) = A_2$ such that $|OA_1| \cdot |OA_2| = |OB|^2 = |OC|^2 = R^2$, i.e. A_2 is the intersection between the tangents at B and C to $C(O, R)$. In the same way we obtain the points B_2 and C_2. If we look at the triangle $\triangle A_1B_1C_1$ which has as inscribed circle $C(O, R)$, the lines A_1A, B_1B, C_1C intersects at Gergonne's point. The inverse of Gergonne's point is E. □

3.2 Cross Ratio and Projective Geometry

Consider four distinct collinear points A, B, C, D on the line d. Attach them the coordinates x_A, x_B, x_C, x_D, respectively. Choose two possible ordered pairs, say (A, B); (C, D), that is, consider the ordered pairs of coordinates (x_A, x_B); (x_C, x_D).

Definition 3.2.1 The cross ratio of four ordered points is the real number

$$[AB; CD] := \frac{x_C - x_A}{x_C - x_B} : \frac{x_D - x_A}{x_D - x_B}.$$

One can see that the definition can be written in the form

$$[AB; CD] := \frac{CA}{CB} : \frac{DA}{DB},$$

but in this case we have to point that if the order on d for the points A and C is given by $x_A < x_C$, then the meaning of CA is $|CA|$, and if $x_A > x_C$ we have $CA = -|CA|$.

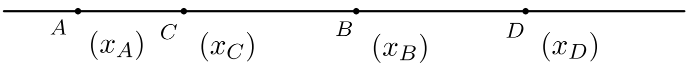


Fig. 3.8 Cross ratio

Exercise 3.2.2 If the order of points A, B, C, D on d is given by $x_A < x_B < x_C < x_D$, then $[AB; CD] > 0$.

Exercise 3.2.3 If the order of points A, B, C, D on d is given by $x_A < x_C < x_B < x_D$, then $[AB; CD] < 0$ (Fig. 3.8).

Exercise 3.2.4 If the order of points A, B, C, D on d is given by $x_A < x_B < x_C < x_D$, then

$$[AD; CB] = \frac{x_C - x_A}{x_C - x_D} : \frac{x_B - x_A}{x_B - x_D} > 0.$$

Observe that in this last case the ordered pairs are (A, D); (C, B) and the cross ratio can be written in the equivalent form $[AD; CB] := \frac{CA}{CD} : \frac{BA}{BD}$ with the meaning explained above.

Exercise 3.2.5 $[AD; BC] + [AB; DC] = 1$ if and only if the order of points on the line d is A, B, C, D.

Hint.

$$[AD; BC] + [AB; DC] = \frac{BA \cdot CD}{BD \cdot CA} + \frac{DA \cdot CB}{DB \cdot CA} = \frac{BA \cdot CD + DA \cdot CB}{BD \cdot CA} = 1.$$

If $A(x_A)$, $B(x_B)$, ... etc.,

$$(x_B - x_A)(x_D - x_C) + (x_D - x_A)(x_C - x_B) = (x_C - x_A)(x_D - x_B)$$

iff the order is as in the statement before.

Exercise 3.2.6 $[AD; BC] = [DA; CB]$.

Exercise 3.2.7 Consider $A(-1)$, $B(0)$, $C(1)$, $D(x)$. Determine x such that

$$[AC; BD] = -1.$$

Hint. If we write the given condition, it results $x + 1 = x - 1$. There is no real x.

To maintain the possibility to have four distinct points with a given cross ratio, as well as for a given cross ratio and three distinct points to exist a fourth point such that the cross ratio is a given one, we have to accept that for each line d it exists an abstract point, denoted ∞, such that for $A \neq B$, $\frac{\infty A}{\infty B} = 1$. This point is called *point at infinity for the line d*.

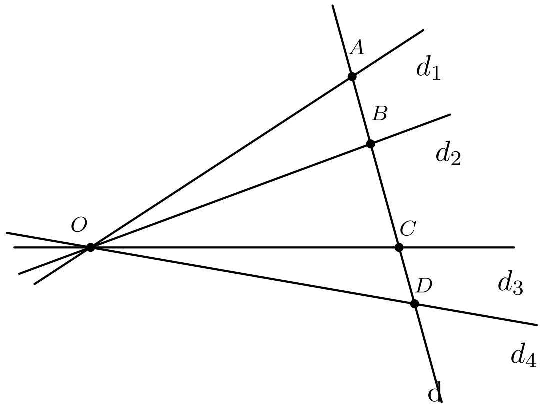


Fig. 3.9 Pappus' theorem

The cross ratio of collinear points can be extended to *pencils of lines*. Consider the lines d_1, d_2, d_3, d_4 and $\{O\} = d_1 \cap d_2 \cap d_3 \cap d_4$. Let d be an arbitrary line and $\{A\} = d \cap d_1$; $\{B\} = d \cap d_2$; $\{C\} = d \cap d_3$; $\{D\} = d \cap d_4$. Choose two ordered pairs of lines, say (d_1, d_2); (d_3, d_4).

By definition $[d_1 d_2; d_3, d_4] := [AB; CD]$.

If we look at this definition it seems that it depends on the line d we choose. Therefore, we have to prove that if we choose another line d' and $\{A'\} = d' \cap d_1$; $\{B'\} = d' \cap d_2$; $\{C'\} = d' \cap d_3$; $\{D'\} = d' \cap d_4$, then $[AB; CD] = [A'B'; C'D']$ (Fig. 3.9).

Theorem 3.2.8 (Pappus' Theorem) *The cross ratio of four lines in a pencil depends only by the angles of the pencil.*

Proof We are in the case: the pencil of lines d_1, d_2, d_3, d_4 with $\{O\} := d_1 \cap d_2 \cap d_3 \cap d_4$, the arbitrary line d and $\{A\} = d \cap d_1$; $\{B\} = d \cap d_2$; $\{C\} = d \cap d_3$; $\{D\} = d \cap d_4$; suppose the order on d being A, B, C, D and use four times sine theorem:

$$\frac{CA}{\sin \angle COA} = \frac{OC}{\sin \angle OAC};$$

$$\frac{CB}{\sin \angle COB} = \frac{OC}{\sin \angle OBD};$$

$$\frac{DA}{\sin \angle DOA} = \frac{OD}{\sin \angle OAD};$$

$$\frac{DB}{\sin \angle DOB} = \frac{OD}{\sin \angle OBD};$$

therefore $[AB; CD] := \dfrac{CA}{CB} : \dfrac{DA}{DB} = \dfrac{\sin \angle COA}{\sin \angle COB} : \dfrac{\sin \angle BOA}{\sin \angle BOD}$. □

Observe that in fact the cross ratio depends on the sine of angles.

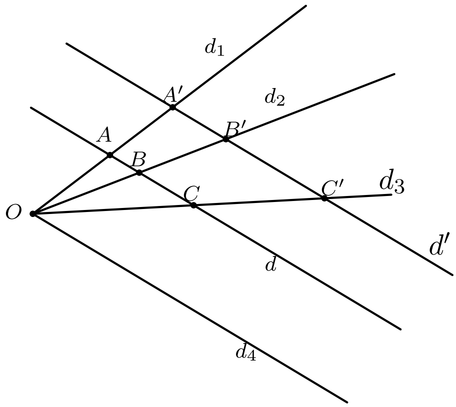


Fig. 3.10 Existence of poits at infinity

Another approach can be

Theorem 3.2.9 *If* O *is the origin and the lines of the pencil are* $d_k : y = m_k \cdot x,\ k \in \{1, 2, 3, 4\}$, *then*

$$[d_1 d_2; d_3 d_4] = \frac{m_3 - m_1}{m_3 - m_2} : \frac{m_4 - m_1}{m_4 - m_2}.$$

Proof Consider d having the equation $x = 1$. The points A, B, C, D on d have the coordinates $(1, m_1), (1, m_2), (1, m_3), (1, m_4)$. □

The pencils of lines allow us to better understand the points at infinity of lines. As above, consider the lines d_1, d_2, d_3, d_4 having the property $\{O\} = d_1 \cap d_2 \cap d_3 \cap d_4$. Let d be parallel to d_4 and $\{A\} = d \cap d_1;\ \{B\} = d \cap d_2;\ \{C\} = d \cap d_3$ (Fig. 3.10).

In this case we have to consider the point at infinity to define $[d_1 d_2; d_3 d_4]$;

$[d_1 d_2; d_3 d_4] = [AB; C\infty] = \dfrac{|CA|}{|CB|}$. If we consider another line, say d', such that $d' \parallel d$ and if we denote by $\{A'\} = d' \cap d_1,\ \{B'\} = d' \cap d_2,\ \{C'\} = d \cap d_3$, then $[d_1 d_2; d_3, d_4] := [A'B'; C\infty]$. The lines d, d', d_4 have empty intersection in E^2. This abstract point who doesn't belong to the Euclidean plane, say ∞, can be taught as the intersection of parallel lines d, d' with d_4.

We define for all parallel lines the same abstract point ∞.

If, in a system of coordinates, all the parallel lines have the slope m, we may think that this point at infinity is attached to this slope. We can even denote this point by ∞_m. An interesting question can be asked: which geometrical structure will be assigned to $\{\infty_m,\ m \in \mathbb{R}\}$? It can be taught as an abstract line? Or it is more intuitive to be taught as an abstract circle? Or it is something else? We see the answer a little bit later.

The cross ratio can be extended to four points distinct points A, B, C, D on a circle Γ. Choose $M \in \Gamma$ and the pencil determined by the rays $d_1 = MA, d_2 = MB, d_3 = MC, d_4 = MD$.

By definition, $[AB; CD]_\Gamma := [d_1 d_2; d_3 d_4]$.

Pappus' theorem shows that this definition is independent of the choice of M. Here, it is important our observation related to the fact that the cross ratio of pencils depends on sine of angles. Since $\sin\alpha = \sin(\pi - \alpha)$, the point M can be chosen even between two consecutive points, that is we can have, for a given sense on our circle, even the order A, B, M, C, D. The cross ratio is the same as for the order, say M, A, B, C, D.

Next theorem shows that the previous cross ratio $[AC; BD]_\Gamma$ can be transferred to the segment lines BA, BC, DA, DC determined by the four distinct points on the circle. We keep our notation generated by the order of points, now on the circle. To have a clear statement, for a chosen sense on our circle, let us consider the points M, A, B, C, D in this order. Denote the angles of the pencil created by $\angle AMB = \alpha$, $\angle BMC = \beta$, $\angle CMD = \gamma$.

Theorem 3.2.10 $[AC; BD]_\Gamma = \dfrac{BA}{BC} : \dfrac{DA}{DC}$.

Proof Consider the segment line $[AD]$ and its intersection with MB, MC denoted by B_1, C_1 respectively. The order on the segment line $[AD]$ is then A, B_1, C_1, D. If we denote by R the radius of Γ we have $|AB| = 2R\sin\alpha$, $|BC| = 2R\sin\beta$, $|AD| = 2R\sin(\alpha+\beta+\gamma)$, $|CD| = 2R\sin\gamma$. Taking the order into consideration, we can write

$$[AC; BD]_\Gamma = [d_1 d_3; d_2 d_4] = \frac{\sin\alpha}{\sin\beta} : \frac{\sin(\alpha+\beta+\gamma)}{\sin\gamma} = \frac{BA}{BC} : \frac{DA}{DC}$$

□

Exercise 3.2.11 $[AD; BC]_\Gamma + [AB; DC]_\Gamma = 1$ if and only if the order of points on the circle Γ is A, B, C, D.

Solution. (Hint) We use the previous theorem, i.e. we express the

$$[AD; BC]_\Gamma + [AB; DC]_\Gamma = \frac{BA \cdot CD}{BD \cdot CA} + \frac{DA \cdot CB}{DB \cdot CA} = \frac{BA \cdot CD + DA \cdot CB}{BD \cdot CA} = 1$$

iff $BA \cdot CD + DA \cdot CB = BD \cdot CA$, that is Ptolomy's equality must happen. □

Theorem 3.2.12 *Let I be an interior point of $C(O, r)$. Consider the chords AA', BB', CC', DD' such that $\{I\} = AA' \cap BB' \cap CC' \cap DD'$ and the order is $A, B, C, D, A', B', C', D'$.*

Then $[A'C'; B'D']_\Gamma = [AC; BD]_\Gamma$.

Proof Let us observe that the symmetric inversion T^s_{I,R^2-OI^2} maps the circle $C(O, R)$ in itself, since $R^2 - OI^2$ is the power of I with respect to the circle, and $|IA| \cdot |IA'| = |IB| \cdot |IB'| = |IC| \cdot |IC'| = |ID| \cdot |ID'| = R^2 - OI^2$, that is $T^s_{I,R^2-OI^2}(A) = A'$, $T^s_{I,R^2-OI^2}(B) = B'$, $T^s_{I,R^2-OI^2}(C) = C'$, $T^s_{I,R^2-OI^2}(D) = D'$. It results

$$|B'A'| = (R^2 - OI^2) \cdot \frac{|BA|}{|IB| \cdot |IA|}, \quad |B'C'| = (R^2 - OI^2) \cdot \frac{|BC|}{|IB| \cdot |IC|},$$

$$|D'A'| = (R^2 - OI^2) \cdot \frac{|DA|}{|ID| \cdot |IA|}, \quad |D'C'| = (R^2 - OI^2) \cdot \frac{|DC|}{|ID| \cdot |IC|}.$$

Taking into consideration the established order and the theorem which transfers the cross ratio from circle to segment lines, we obtain $[A'C'; B'D']_\Gamma = [AC; BD]_\Gamma$. □

We obtain a similar result for a point J outside the circle using a direct inversion T_{J,OJ^2-R^2} and $T_{J,OJ^2-R^2}(A) = A'$, *etc.* $[A'C'; B'D']_\Gamma = [AC; BD]_\Gamma$.

More general, *an inversion $T_{O,k}$ leaves unchanged the cross ratio of four collinear points or the cross ratio of four cyclic points.* It doesn't matter if the four collinear points are mapped into cyclic points, or the cyclic points are mapped into cyclic (or collinear) points. This result is a fundamental one.

Definition 3.2.13 A projective map of a circle $C(O, R)$ is a one to one function $f : C(O, R) \to C(O, R)$ such that for any four points $A_i,\ i \in \{1, 2, 3, 4\}$ and their images $B_i = f(A_i)$, it happens

$$[A_1A_2; A_3A_4]_{C(O,R)} = [B_1B_2; B_3B_4]_{C(O,R)}.$$

Definition 3.2.14 The points which correspond in a projective map f, e.g. A_i and $f(A_i)$, are called homologous points.

According to a previous result, let observe that the symmetric inversion $T^s_{I,R^2-OI^2},\ I \in intC(O, R)$ is a projective map of the circle $C(O, R)$.

More, $T^s_{R^2-OI^2}$ can be identified with the simpler map determined by the point I denoted

$$I : C(O, R) \to C(O, R),\ I(A) = A',$$

where $A' \neq A$ is the other intersection between AI and the circle $C(O, R)$.

The same for the direct inversion $T_{I,OJ^2-R^2},\ J \in extC(O, R)$. This is a projective map and can be identified with $J : C(O, R) \to C(O, R),\ J(A) = A'$, where $A' \neq A$ is the other intersection between JA and the circle $C(O, R)$.

If we consider $I_1, I_2 \in intC(O, R)$ it can be obtained that $T^s_{I_1,R^2-OI_1^2} \circ T^s_{I_2,R^2-OI_2^2} : C(O, R) \to C(O, R)$ is a projective map.

Definition 3.2.15 A projective map between two lines d_1 and d_2 is a one to one function $f : d_1 \to d_2$ such that for any four points $A_i \in d_1,\ i \in \{1, 2, 3, 4\}$ and their images $B_i = f(A_i) \in d_2$, it happens

$$[A_1A_2; A_3A_4] = [B_1B_2; B_3B_4].$$

Since the previous definition has also sense for $f : d \to d$, we may talk about *projective maps on d*.

Theorem 3.2.16 *A projective map between two lines is determined by three pairs of homologous points.*

Proof Denote the homologous points in the form $A \to B$ instead of $B = f(A)$, because this notation will help us later.

Then, we know the three pairs of homologous points $A_0 \to B_0$, $A_1 \to B_1$, $A_2 \to B_2$.

We have to show that for any four arbitrary points A_i, A_j, A_k, A_l and their homologous B_i, B_j, B_k, B_l the relation $[A_iA_j;\, A_kA_l] = [B_iB_j;\, B_kB_l]$ is deduced from $[A_0A_1;\, A_2A_i] = [B_0B_1;\, B_2B_i]$ using successively the indexes i, j, k and l.

The idea is to find somehow a procedure of replacement of the homologous points initially given.

We also have $[A_1A_2;\, A_0A_j] = [B_1B_2;\, B_0B_j]$ and $[A_1A_2;\, A_iA_0] = [B_1B_2;\, B_iB_0]$. It results

$$[A_1A_2;\, A_0A_j] \cdot [A_1A_2;\, A_jA_0] = [B_1B_2;\, B_0B_j] \cdot [B_1B_2;\, B_iB_0],$$

that is

$$[A_1A_2;\, A_iA_j] = [B_1B_2;\, B_iB_j].$$

We succeeded to replace the pair of homologous points $A_0 \to B_0$.

Then, the from previous relation and $[A_1A_k;\, A_jA_i] = [B_1B_k;\, B_iB_j]$ we have

$$[A_1A_2;\, A_iA_j] \cdot [A_1A_k;\, A_jA_i] = [B_1B_2;\, B_iB_j] \cdot [B_1B_k;\, B_iB_j],$$

i.e.

$$[A_2A_k;\, A_jA_i] = [B_2B_k;\, B_jB_i].$$

Finally, taking into consideration the previous result and $[A_lA_2;\, A_jA_i] = [B_lB_2;\, B_jB_i]$ it results

$$[A_2A_k;\, A_iA_j] \cdot [A_lA_2;\, A_jA_i] = [B_2B_k;\, B_jB_i] \cdot [B_lB_2;\, B_jB_i],$$

that is

$$[A_lA_k;\, A_jA_i] = [B_lB_k;\, B_jB_i].$$

□

Consequences: The way the previous theorem was proved makes it to hold for projective maps: we can imagine between two circles, between a line and a circle, on the same line or on the same circle.

This theorem can be easily extended to projective pencils: they are determined by three pairs of homologous rays. Generally speaking, a projective map is determined by the knowledge of three pairs of homologous points.

If two projective maps f and f_1 has the same three pairs of homologous points, then $f = f_1$.

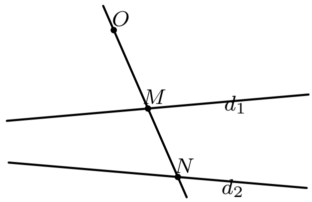


Fig. 3.11 Projectivity between two lines

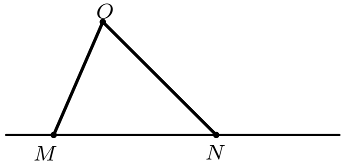


Fig. 3.12 Projectivity determined by an angle on a line

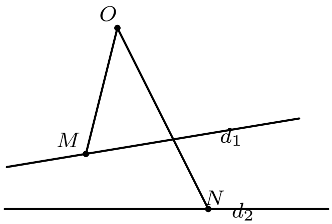


Fig. 3.13 Projectivity determined by an angle between two lines

Other examples of projective maps:

1. Consider two distinct lines d_1, d_2 and a point O in E^2 who doesn't belong to $d_1 \cup d_2$. A moving ray through O intersects d_1 in M and d_2 in N. Then, using Pappus' theorem, $M \to N$ is a projective map between d_1 and d_2 (Figs. 3.11 and 3.12).

2. Two points moving with the same speed on two distinct lines, or on a same line, determine a projective map.

3. Consider a point $O \notin d$ and a constant angle given angle with its vertex in O rotating around O. The first side of the angle intersect d in M and the second side in N. Again, using Pappus' theorem, $M \to N$ is a projective map on d (Fig. 3.13).

4. The example above may be extended. Consider two lines d_1 and d_2. The given constant angle intersects the lines such that M is on d_1 and N is on d_2. Using Pappus' theorem, $M \to N$ is a projective map between d_1 and d_2.

For a projective map on d, denote the coordinate of M by x and the coordinate of N by y, where $M \to N$ are homologous points.

Theorem 3.2.17 *A projective map on d determines a function* $h(x) = \dfrac{Ax + B}{Cx + d}$, *$A, B, C, D$ being real constants such that $AD - BC \neq 0$.*

Proof (Hint) Suppose the three given homologous points are $0 \to y_0,\ 1 \to y_1,\ x_2 \to y_2$. The condition $[xx_2; 01] = [yy_2; y_0y_1]$ becomes

$$\frac{x}{x_2} : \frac{x-1}{x_2-1} = \frac{y-y_0}{y_2-y_0} : \frac{y-y_1}{y_2-y_1}.$$

After computations it results the desired formula. After another computation the coefficients verify $AD - BC \neq 0$. □

Theorem 3.2.18 *The function* $h(x) = \dfrac{Ax+B}{Cx+d}$, A, B, C, D *being constants such that* $AD - BC \neq 0$ *describes a projective map on* d.

Proof (Hint) Replace $y_k = h(x_k)$ by $\dfrac{Ax_k+B}{Cx_k+D}$ in $[y_1y_2; y_3y_4]$ and use $AD - BC \neq 0$ to simplify. A straightforward computation shows that $[x_1x_2; x_3x_4] = [y_1y_2; y_3y_4]$. □

Definition 3.2.19 A projective map on d which interchanges a pair of homologous points is called geometric involution, or simple, involution (Fig. 3.14).

Theorem 3.2.20 *All pairs of homologous points interchange in an involution.*

Proof Consider $A \to B$, $B \to A$, $M \to N$ and $N \to X$. We wish to prove that $X = M$. Since $[AB; MN] = [BA; NX]$ it results $[AB; MN] = [AB; XN]$ i.e. $X = M$. □

The same considerations hold for involutions of a circle. They have a pair of homologous points which can be interchanged. In fact all pairs of homologous points can be interchanged. Therefore $M \to N$ implies $N \to M$, too.

We saw above two examples of involutions of a circle: T^s_{I,R^2-OI^2} is an interior involution, i.e. it is described by the point $I \in intC(O, R)$. T_{J,OJ^2-R^2} is an exterior involution, i.e. it is described by the point $J \in extC(O, R)$.

Consider a projective map between two lines, $f : d_1 \to d_2$. Denote $\{O\} := d_1 \cap d_2$ and suppose that $f(O) = O$. Such a point is called a *self-homologous point*. A projective map between two lines as above with a self-homologous point is called a *perspective map*.

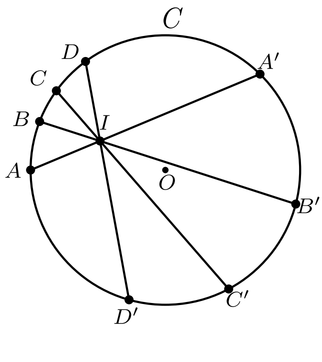


Fig. 3.14 Interior involution of a circle

Theorem 3.2.21 *For a perspective map between d_1 and d_2, the lines which connect homologous points have a common point.*

Proof The perspective map is determined by $O \to O,\ A_1 \to A_2,\ B_1 \to B_2$. Denote $\{I\} = A_1A_2 \cap B_1B_2$ and consider the pencil of lines IO, IA_1, IB_1, IM where $A_1, B_1, M \in d_1$. Suppose that $f(M) = N,\ N \in d_2$ and denote by $\{N'\} = IM \cap d_2$. The perspective map f implies $[OA_1;\ B_1M] = [OA_2;\ B_2N]$, and Pappus' theorem implies $[OA_1;\ B_1M] = [OA_2;\ B_2N']$. Therefore $N = N'$ and the arbitrary line MN which connects homologous points contains I. □

Consider a set of arbitrary indexes denoted by $\mathbb{I}$, $O_1, O_2 \in E^2$. Also consider both the lines passing through O_1, denoted by $\alpha_i,\ i \in \mathbb{I}$ and the lines passing through O_2, denoted by $\beta_i,\ i \in \mathbb{I}$. Let denote by $O_1(\alpha),\ \ O_2(\beta)$ the two pencils of lines. The next definition makes sense even if $O_1 = O_2$.

Definition 3.2.22 Two pencils of lines are projective if there exists a one to one map $f : O_1(\alpha) \to O_2(\beta)$ such that for any four rays of the first pencil, say $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and their images $\beta_1, \beta_2, \beta_3, \beta_4$, we have $[\alpha_1\alpha_2;\ \alpha_3\alpha_4] = [\beta_1\beta_2;\ \beta_3\beta_4]$. α_1 and β_1 are called homologous rays.

Example 3.2.23 Consider a line d and a projective map $f : d \to d$. Choose four arbitrary points A_1, A_2, A_3, A_4 on d and their images via f, B_1, B_2, B_3, B_4. We have $[A_1A_2;\ A_3A_4] = [B_1B_2;\ B_3B_4]$. Therefore $[OA_1OA_2;\ OA_3OA_4] = [OB_1OB_2;\ OB_3OB_4]$. It results that $O(OA)$ and $O(Of(A))$ are projective pencils of lines.

Example 3.2.24 Consider a line d and a projective map $f : d \to d$. Choose four arbitrary points A_1, A_2, A_3, A_4 on d and their images via f, B_1, B_2, B_3, B_4. We have $[A_1A_2;\ A_3A_4] = [B_1B_2;\ B_3B_4]$. Therefore $[O_1A_1O_1A_2;\ O_1A_3O_1A_4] = [O_2B_1O_2B_2;\ O_2B_3O_2B_4]$. It results that $O_1(O_1A)$ and $O_2(O_2f(A))$ are projective pencils of lines.

Example 3.2.25 Consider the lines d, d' and a projective map $f : d \to d'$. Choose four arbitrary points A_1, A_2, A_3, A_4 on d and their images via f, B_1, B_2, B_3, B_4 on d'. We have $[A_1A_2;\ A_3A_4]_d = [B_1B_2;\ B_3B_4]_{d'}$. Therefore $[O_1A_1O_1A_2;\ O_1A_3O_1A_4] = [O_2B_1O_2B_2;\ O_2B_3O_2B_4]$. It results that $O_1(O_1A)$ and $O_2(O_2f(A))$ are projective pencils of lines.

We left to the reader to prove: "A projective map between two pencils of lines in determined by three pairs of homologous rays".

Theorem 3.2.26 (Steiner) *Consider two projective pencils of lines. Their homologous rays intersect on a conic.*

Proof To simplify the proof consider a projective map f on the line of equation $y = 1,\ O_1(0,0),\ O_2(1,0)$. If $M(x, 1)$ and $N\left(\dfrac{Ax+B}{Cx+D}, 1\right)$ are the homologous points, the equations of the lines O_1M and O_2N are $Y = \dfrac{1}{x}X$ and $Y = \dfrac{1}{\dfrac{Ax+B}{Cx+D} - 1}(X -$

1). It is not necessary to compute the coordinates of the intersection $O_1M \cap O_2N$. We can substitute x from the first equation, i.e. $x = \frac{X}{Y}$, and replace in the second. It results

$$Y = \frac{1}{\dfrac{A\frac{X}{Y} + B}{C\frac{X}{Y} + D} - 1}(X - 1),$$

therefore, the coordinates of the intersection point lie on the conic of equation

$$-CX^2 + (A - C - D)XY + (B - D)Y^2 + CX + DY = 0.$$

The reader has to observe that Steiner's conic contains O_1 and O_2. □

Problema 3.2.27 Consider M and N on the hypotenuse BC of an isosceles rectangle triangle ABC such that $MN^2 = BM^2 + CN^2$. Prove that $\angle MAN = \frac{\pi}{4}$.

Solution. Consider a system of coordinates such that $A(0, a)$, $B(-a, 0)$, $C(a, 0)$, $M(x, 0)$, $N(y, 0)$. $MN^2 = BM^2 + CN^2$ can be written in the form

$$(x + a)^2 + (a - y)^2 = (y - x)^2,$$

that is $y = \frac{ax + a^2}{-x + a}$. This is a projective map on BC.

Since $x = -a$ implies $y = 0$, it results $B \to O$. $x = 0$ implies $y = a$, therefore $O \to C$. For $x = a$ it results $y = \infty$, i.e. $C \to \infty$.

So, this projective map on BC is determined by $B \to O,\ O \to C,\ C \to \infty$.

If we consider a rotating angle $\angle MAN = \frac{\pi}{4}$ we observe three important positions

$$M = B, N = O;\quad M = O, N = C;\quad M = C, N = \infty$$

such that $\angle BAO = \angle OAC = \angle CO\infty = \frac{\pi}{4}$. Therefore the rotating angle leads to a projective map on BC determined by $B \to O,\ O \to C,\ C \to \infty$. The two projective maps are coincident, therefore always $\angle MAN = \frac{\pi}{4}$. □

Problema 3.2.28 (*Karya's point*) Let I be the incentre of the triangle $\triangle ABC$ and D', E', F' be the symmetric of I with respect the sides BC, CA, AB.

Then, $AD' \cap BE' \cap CE' \neq \emptyset$.

Solution. (D. Barbilian) Denote by D, E, F the contacts of the incircle with the sides BC, CA, AB respectively. Denote also by D_1, D_2 the intersection points of ID with AC and AB, respectively. The same, $\{E_1\} = IE \cap BC$; $\{E_2\} = IE \cap BA$. Consider first two moving points $M \in ID,\ N \in IE$ who start to move with the same speed from I in the direction of D, respectively E (Fig. 3.15).

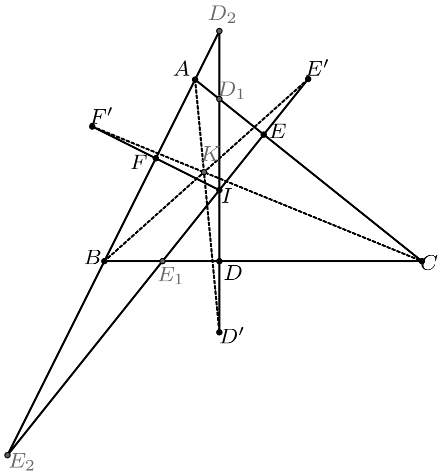


Fig. 3.15 Karya's Point

We know that $M \to N$ is a projective map between the lines ID and IE. It results a projective map between the pencils $A(AM)$ and $B(BN)$. The intersection point between AM and BN lies on a conic. A conic is determined by the knowledge of five distinct points of it. The initial moment $M = N = I$ implies that I belongs to the conic. When $M = D$ it results $N = E$, therefore Gergonne's point of the triangle ABC, G_e, belongs to the conic. Now consider M, N moving from I to D_1, E_1 respectively. Since the triangles IED_1 and IED_1 are congruent, when $M = D_1$ it results $N = E_1$, therefore $AM \cap BN = \{C\}$. According to Steiner's theorem the conic contains A and B. But it is easy to observe for this projective map why. When $M = D_2$, $AD_2 \cap BN = \{B\}$, and when $N = E_2$, $AM \cap BE_2 = \{A\}$. Therefore Steiner conic for the projective pencils $A(AM),\ B(BN)$ is determined by A, B, C, I, G_e. The same for the Steiner conics determined by the projective pencils $A(AM), C(CP)$ and $B(BN), C(CP)$. Therefore the three Steiner conics are coincident, i.e. $AM \cap BN \cap CP \neq \emptyset$ when $|IM| = |IN| = |IP|$. In the particular case of the statement, a particular point belongs to Steiner's conic. We are talking about Karya's point. If we choose the points at infinity of AM, BN, CP we can see that the orthocenter H belongs to the Steiner conic. As you can see, this projective solution allows us to highlight many other points of intersection among AM, BN and CP described by the condition $|IM| = |IN| = |IP|$. □

There is a special case when Steiner's conic is a line only.

Definition 3.2.29 Two projective pencils of lines, $O_1(\alpha)$ and $O_2(\beta),\ O_1 \neq O_2$, are called perspective pencils if the ray O_1O_2 is self-homologous, i.e. $O_1O_2 \to O_1O_2$.

Theorem 3.2.30 *If $O_1(\alpha)$ and $O_2(\beta),\ O_1 \neq O_2$ are perspective pencils, then the homologous rays intersection lies on a line.*

Proof Denote $p = O_1O_2$. The perspective map is determined by $p \to p,\ \alpha_1 \to \beta_1,\ ,\alpha_2 \to \beta_2$. Denote $\{T_1\} = \alpha_1 \cap \beta_1,\ ,\{T_2\} = \alpha_2 \cap \beta_2,\ $. If $\alpha \to \beta$, and $\{T\} =$

$\alpha \cap T_1T_2$ we observe that $\alpha \to O_2T$ belongs to the previous projective map. Therefore, $O_2T = \beta$, and $\alpha \cap \beta$ always belongs to T_1T_2. This line is called the perspective axis of the perspective pencils of lines $O_1(\alpha)$ and $O_2(\beta)$. $\square$

Let us answer to the question: which is the geometrical structure of $\{\infty_m,\ m \in \mathbb{R}\}$?

First, it is easy to see that for a given line $d \subset E^2$ it exist two perspective pencils of lines, $O_1(\alpha)$ and $O_2(\beta)$, such that d is the perspective axis of the previous perspective pencils. If we have two perspective pencils there is one case in which the perspective axis doesn't exist: when the homologous rays of the perspective pencils are parallel. Exactly as in the case of the abstract infinity point of a line added to preserve a geometric rule, we do the same thing. In the case of parallel perspective pencils the perspective axis is an abstract line, called the *line at infinity of* E^2. Therefore we may denote $d_\infty := \{\infty_m,\ m \in \mathbb{R}\}$.

Perspective pencils allow us to construct a special line assigned to any projective map f on a circle: the axis of f. This line plays a crucial role in the construction of Poincaré disk model.

Consider for a projective map f on a circle Γ the homologous points $M, M',\ M \to M'$ which describe the projective map f. If we choose two particular pairs of homologous points, say $A \to A',\ B \to B'$, the point $\{P\} = AB' \cap A'B$ allows us to create a function $g : \Gamma \to \Gamma,\ g(N) = N',\ \{N'\} := NP \cap \Gamma$. It is obvious to observe that g is an involution of Γ.

Theorem 3.2.31 *(i) The map* $f \circ g_P : \Gamma \to \Gamma$ *is an involution of* Γ.

(ii) The locus of points $I_f \in E^2$ *such that* $f \circ g_{I_f}$ *is an involution of* Γ *is a line.*

Proof (i) f and g_P are projective maps on Γ, then $f \circ g_P$ is a projective map on Γ. It remains to prove that $f \circ g_P$ has a pair of homologous points which interchanges. We show that $f \circ g_P(A') = B'$ and $f \circ g_P(B') = A'$. Let's compute $f \circ g_P(A') = f(g_P(A')) = f(B) = B'$. In the same way $f \circ g_P(B') = f(g_P(B')) = f(A) = A'$, therefore $f \circ g_P$ is an involution of Γ.

(ii) Consider $A \to A'$ a given pair of homologous point of f and $M \to M'$ the general pair of homologous points of f. Therefore $A \to A'$ is a particular pair obtained from the general pair by replacing M by A. The pencils $A(AM')$ and $A'(A'M)$ are perspective, the self-homologous ray being AA'. The homologous rays intersection, i.e. $\{I_f\} = AM' \cap A'M$, lies on the perspective axis, therefore the locus is a line. $\square$

This line is called the *axis of the projective map* f. The previous theorem shows that this line is $\{I_f \mid \{I_f\} = AM' \cap MA',\ M \in \Gamma\}$. A direct consequence appears.

Theorem 3.2.32 *(i) If* $A \to A',\ B \to B',\ C \to C'$ *are homologous points of* f *on* Γ, *then the points* $\{U\} = AB' \cap BA',\ \{V\} = AC' \cap CA',\ \{W\} = BC' \cap CB'$ *are collinear.*

(ii) All projective maps of a circle can be written as a product of involutions (in a non-unique way).

(iii) Two interior involutions of $\Gamma,\ I, J$ *determine in an unique way the projective map of* $\Gamma,\ f = I \circ J$ *such that the axis of* f *is* IJ.

(iv) Denote $\{s, S\} = IJ \cap \Gamma$ *such that the order is* s, I, J, S. *Then* $f(s) = s$, $f(S) = S$ *and* $[IJ; Ss] > 1$.

(v) If $M \in \Gamma$, $J(M) = M'$, $M' \in \Gamma$, $I(M') = N$, $N \in \Gamma$, *for an arbitrary* $X \in IJ$ *there is an unique* $Y \in IJ$ *such that* $XN \cap MY \in \Gamma$.

(vi) $X \to Y$ *is a projective map on the line* IJ *(This map is called the axial decomposition of* f*).*

(vii) $[SsIJ] = [SsXY]$.

Proof (i) U, V, W are three particular points of $\{I_f \mid \{I_f\} = AM' \cap MA', M \in \Gamma\}$.

(ii) If we choose the point, say $\{U\} = AB' \cap BA'$, then $f \circ U = L$ where $\{L\}$ is the intersection between the axis of f and $A'B'$. Therefore $f = L \circ I$.

(iii) (iv) and (v) are obvious.

(vi) Consider the projective map on IJ determined by particular positions of X and Y, $s \to s$, $S \to S$, $I \to J$. If remains to prove $[sSIX] = [sSJY]$. Consider the pencils $M(Ms, MS, MJ, MY)$ and $N(Ns, NS, NI, NX)$. Since the angles involved are equal it results

$$[MsMS; MJMY] = [NsNS; NINX],$$

i.e. $[sSIX] = [sSJY]$.

(vii) From $[sSIX] = [sSJY]$ it results $[SsIX] = [SsJY]$. If you write the last one equality it results

$$\frac{IS}{Is} : \frac{XS}{Xs} = \frac{JS}{js} : \frac{YS}{Ys}.$$

This one can be thought as

$$\frac{IS}{Is} : \frac{JS}{ss} = \frac{XS}{Xs} : \frac{YS}{Ys},$$

which means $[SsIJ] = [SsXY]$, or equivalently $[sSIJ] = [sSXY]$. □

Before continuing, let us conclude in the following way.

For I and J belonging to the interior of our circle, we construct the projectivity of the circle $f := I \circ J$ determined by the product of the given interior involutions. Suppose $M \to N$ in this projectivity. For each $X \in IJ$, we construct $Y \in IJ$ as in the previous theorem. The projectivity $X \to Y$ on IJ is determined by f. It is called the axial decomposition of the projectivity f (of the circle) on the line IJ. This was the most important step towards the construction of a non-Euclidean distance in the interior of the circle.

The next steps are the Theorems 3.3.1, 3.3.2 and 3.3.3 in the next section.

Let us observe: Consider $S, S' \in \Gamma$ such that the chord SS' is not a diameter; Then, the tangents at S, S' meet at the centre of an orthogonal circle to Γ.

If $A \in int\Gamma$ and $S \in \Gamma$, we know that the orthogonal circle to Γ passing through A and S is constructed in the following way: the tangent at S meet the perpendicular bisector of the segment AS at the centre of the orthogonal circle.

If $A, B \in int\Gamma$ the orthogonal circle to Γ passing through A and B is constructed in the following way: we construct A', the inverse of A in the direct inversion T_{O,R^2}, where O, R are the centre, respectively the radius of Γ. The perpendicular bisectors of the triangle ABA' meet at the centre of the orthogonal circle we are looking for. Observe that B', the inverse of B in the same inversion belongs to this circle.

Another more important observation is:

Proposition 3.2.33 *If $A, B, C, D \in \Gamma$ such that $\{L\} = AB \cap CD$, $L \in int\Gamma$, then the orthogonal circles determined by the chords AB, CD denoted by γ_{AB}, γ_{CD} respectively, meet in X, X' such that O, X, L, X' are collinear.*

Proof O and L have equal powers with respect γ_{AB}, γ_{CD}.

The powers are R^2 for O, $u := |LA| \cdot |LB| = |LC| \cdot |LD|$ for L respectively, therefore they belong to the radical axis of the two circles. But the radical axis passes through the points of intersection of the two orthogonal circles, i.e. O, X, L, X' are collinear. Extra, $|OX| \cdot |OX'| = R^2$, that is X and X' are inverse in T_{O,R^2}. □

3.3 Poincaré Disk Model

We underline some results proved above, results which are necessary to introduce the Poincaré disk model. If $I, J \in int\Gamma$, $f := I \circ J$ is a projective map on Γ such that IJ is its axis.

If $X \to Y$ are the homologous points in the axial decomposition of f on $a := IJ$, and $\{s, S\} = a \cap \Gamma$ such that the order is s, I, J, S; s, X, Y, S respectively, then $[IJSs] = [XYSs] = k > 1$. Therefore $[SsIJ] = k$ is an invariant of the axial decomposition of f. In fact k depends on I and J, that is $k = k_{IJ}$ is an invariant attached to the involutions I and J on the axis of the projective map $f = I \circ J$.

If we consider the orthogonal circle to Γ through s and S, denoted g, on the arc $g := g_{sS}$ from the $int\Gamma$ we can consider two special points I', J', $\{I'\} := OI \cap g_{sS}$, $\{J'\} := OJ \cap g_{sS}$.

A very important result will be proved:

$$[IJSs] = [I'J'Ss]_g^2.$$

Let us describe again the context. Consider the circle Γ centred at O and $int\Gamma$ the disk enclosed by Γ. Let I and J be in $int\Gamma$ and denote by s and S the intersections of the line IJ with Γ. Suppose the order is s, I, J, S. Denote by g the orthogonal arc to Γ passing through s and S, and let I' and J' be the intersections of g with OI and OJ, respectively (Fig. 3.16).

We have to consider the direct inversion of pole S and power $\mu = (sS)^2$.

The point s is fixed by this transformation. The circle Γ, which passes through the pole of inversion, is transformed into the line $i(\Gamma)$ which passes through s. The arc g is transformed into the line $i(g)$, and $i(g) \perp i(\Gamma)$. Let $d_1 := OI$ and $d_2 := OJ$.

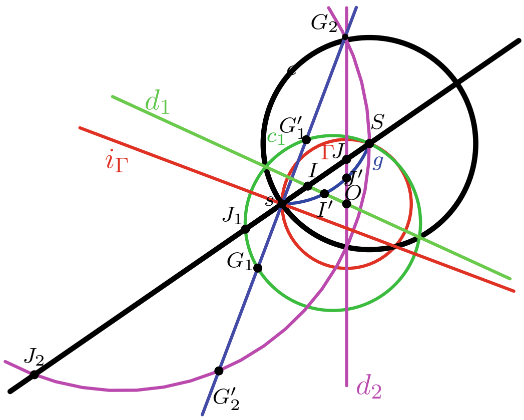


Fig. 3.16 $[IJSs] = [I'J'Ss]_g^2$

The line d_1, which doesn't pass through the pole of inversion, is transformed into the circle c_1 passing through S. Furthermore, c_1 contains the images of I' and I, denoted by G_1 and J_1, respectively. In fact, since $d_1 \perp \Gamma$, then c_1 and $i(\Gamma)$ must also be orthogonal, which means that c_1 has the line $i(\Gamma)$ as a diameter. A similar reasoning can be done for the line d_2.

We introduce the following notations: $\{G_1, G'_1\} = c_1 \cap i(g)$, and $\{G_2, G'_2\} = c_2 \cap i(g)$.

Finally, we remark that S is mapped by this inversion into ∞.

Then we have the following:

$$[IJSs] = [J_1J_2\infty s] = \frac{|sJ_2|}{|sJ_1|},$$

$$[I'J'Ss]_g = [G_1G_2\infty S] = \frac{|sG_2|}{|sG_1|}.$$

The power of the point s with respect to c_1 yields:

$$|Ss| \cdot |sJ_1| = |sG_1| \cdot |sG'_1| = |sG_1|^2.$$

Similarly, the power of s with respect to c_2 yields:

$$|Ss| \cdot |sJ_2| = |sG_2| \cdot |sG'_2| = |sG_2|^2.$$

Therefore, we have

$$\frac{|sJ_1|}{|sJ_2|} = \left(\frac{|sG_1|}{|sG_2|}\right)^2.$$

This result actually means we proved the following

Theorem 3.3.1 $[IJSs] = [I'J'Ss]_g^2$.

It results that the points $I, J \in a$ generate $I', J' \in g_{sS}$ such that the invariant $k_{IJ} = [IJSs]$ generates the invariant $K_{I'J'} = [I'J'Ss]_g^2$ and $[IJSs] = [I'J'Ss]_g^2$.

Therefore we move points and invariants from the axis of a projective map of a circle Γ to an orthogonal arc g to Γ.

On the initial configuration, we apply a symmetric inversion of pole J' and power μ', where μ' is the power of J' with respect to the circle Γ.

Consequently, the circle Γ is mapped into Γ itself by this transformation.

The arc g becomes the line $i(g)$, which is a diameter in Γ.

The point I' is transformed into F_1', which lies on $i(g)$, such that $|J'I'| \cdot |J'F_1'| = \mu$.

The pole J' is mapped into ∞.

The point $P \in \Gamma$ is transformed into $P' \in \Gamma$, such that $|J'P| \cdot |J'P'| = \mu'$ and P' is the second intersection of Γ with the line $J'P$.

Denote $i(s)$ and $i(S)$ the images of s and S through the previously described inversion. We have

$$[I'J'Ss]_g = [F_1'\infty i(S)i(s)] = \frac{|i(S)F_1'|}{|i(s)F_1'|} = \frac{\max_{P'\in\Gamma}|P'F_1'|}{\min_{P'\in\Gamma}|P'F_1'|}.$$

Furthermore,

$$|P'F_1'| = \mu' \cdot \frac{|PI'|}{|J'P| \cdot |I'J'|} = \frac{\mu'}{|I'J'|} \cdot \frac{|PI'|}{|PJ'|}.$$

This shows us that $|P'F_1'|$ reaches its maximum and minimum in the same time as the ratio $\frac{|PI'|}{|PJ'|}$ (Fig. 3.17).

Therefore, we have proved.

Theorem 3.3.2

$$[I'J'Ss]_g = \frac{\max_{P\in\Gamma}\frac{|PI'|}{|PJ'|}}{\min_{P\in\Gamma}\frac{|PI'|}{|PJ'|}}.$$

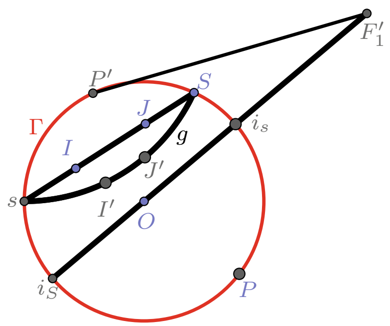


Fig. 3.17 Poincaré modified distance

Let see how these algebraic invariants generate distances in the interior of Γ. Denote

$$d_a(I, J) := \frac{1}{2} ln[IJSs],$$

$$d_g(I', J') := ln[I'J'Ss]_g,$$

$$d(I', J') := \ln \frac{\max_{P\in\Gamma} \frac{|PI'|}{|PJ'|}}{\min_{P\in\Gamma} \frac{|PI'|}{|PJ'|}}.$$

The previously proven facts allow us to assert

Theorem 3.3.3 $d_a(I, J) = d_g(I', J') = d(I', J')$.

After Definition 3.3.4 we will prove that all three formulas are distances for the interior of the circle Γ. Having this in mind, let us say that the first distance, d_a is related to the Cayley-Klein model of a Non-Euclidean geometry. It is easy to see that if we choose a point K on IJ line such that the order of points is I, J, K, we obtain the following equality related to the Cayley-Klein distance,

$$d_a(I, K) = d_a(I, J) + d_a(J, K).$$

Therefore the chord IJ becomes a line of this geometry. It is obvious that from a point L which doesn't belong to the line IJ, one can construct at least two chords, that is two lines of this distance, passing through L which do not intersect IJ. This is the first model of Non-Euclidean geometry. A Differential Geometry treatment of this model can be seen later when we discuss the hyperboloid models of the Non-Euclidean geometry.

The second and the third distances are related to the Poincaré model.

Consider two arbitrary sets K and U.

Definition 3.3.4 The function $f : K \times U \to \mathbb{R}_+^*$ is called an *influence of the set K over U* if for any $A, B \in U$ the ratio $g_{AB}(P) = \dfrac{f(P, A)}{f(P, B)}$ has a maximum $M_{AB} \in \mathbb{R}$ when $P \in K$.

Note that $g_{AB} : K \to \mathbb{R}_+^*$. If we assume the existence of max $g_{AB}(P)$, when $P \in K$, then there also exists $m_{AB} = \min_{P\in K} g_{AB}(P) = \dfrac{1}{M_{BA}}$.

Consider $d : U \times U \to \mathbb{R}_+$ given by

$$d(A, B) = \ln \frac{\max_{P\in K} g_{AB}(P)}{\min_{P\in K} g_{AB}(P)}.$$

It is easy to prove that the previous formula leads to a semi-distance, i.e.:

(1) if $A = B$ then $d(A, B) = 0$; (2) d is symmetric; (3) d satisfies triangle inequality.

(1) and (2) are obvious. For (3) let A, B, C distinct points in J and the pair of points $S_0, s_0 \in K$, $S_1, s_1 \in K$, $S_2, s_2 \in K$ such that

$$\max_{P\in K} g_{AB}(P) = \frac{f(S_0, A)}{f(S_0, B)}, \quad \min_{P\in K} g_{AB}(P) = \frac{f(s_0, A)}{f(s_0, B)}$$

$$\max_{P\in K} g_{AC}(P) = \frac{f(S_1, A)}{f(S_1, C)}, \quad \min_{P\in K} g_{AC}(P) = \frac{f(s_1, A)}{f(s_1, C)},$$

$$\max_{P\in K} g_{BC}(P) = \frac{f(S_2, B)}{f(S_2, C)}, \quad \min_{P\in K} g_{BC}(P) = \frac{f(s_2, B)}{f(s_2, C)}.$$

If S_0, S_2 are replaced by S_1 and s_0, s_2 are replaced by s_1 we obtain

$$d(A, B) + d(B, C) = \ln\left[\left(\frac{f(S_0, A)}{f(S_0, B)} : \frac{f(s_0, A)}{f(s_0, B)}\right) \cdot \left(\frac{f(S_2, B)}{f(S_2, C)} : \frac{f(s_2, B)}{f(s_2, C)}\right)\right] \geq$$

$$\geq \ln\left(\frac{f(S_1, A)}{f(S_1, C)} : \frac{f(s_1, A)}{f(s_1, C)}\right) = d(A, C).$$

In particular for $f(P, A) = |PA|$, $K = \Gamma$ is a circle and $U := int\Gamma$ its interior, we obtain that our last formula among the previous three is a semi-distance on $int\Gamma$. But there is no pair $(A, B) \in U \times U$, $A \neq B$, such that the ratio $g_{AB}(P) = \dfrac{f(P, A)}{f(P, B)}$ is constant for all $P \in K$ (in the case when $K = \Gamma$ is a circle and $U := int\Gamma$), that is if $d(A, B) = 0$ it results $A = B$, i.e. all three equal formulas $d_a(I, J) = d_g(I', J') = d(I', J')$ are distances.

Definition 3.3.5

$$d(I', J') = \ln \frac{\max_{P\in\Gamma} \frac{|PI'|}{|PJ'|}}{\min_{P\in\Gamma} \frac{|PI'|}{|PJ'|}}.$$

is called a *Poincaré distance between the points I' and J' of the disk.*

We prefer to consider this general form of the distance d, because if we change K and U, we can obtain available distances on U which come only from the existence of the asked maximum. The reader will see this in the cases of the "semi-plane" and "exterior of the disk" models for Non-Euclidean Geometry.

All these beautiful geometric facts were possible because of the axial projective map derived from a projective map of a circle.

Problema 3.3.6 Show that for three points A, B, C in this order on the orthogonal arc g to the circle Γ, $A, B, C \in int\Gamma$, we have $d(A, C) = d(A, B) + d(B, C)$.

Solution. Suppose the order is s, A, B, C, S where s, S are the "ends" of the arc g belonging to the centre. In fact, the ratios $\frac{|PA|}{PC|}, \frac{|PA|}{PB|}, \frac{|PB|}{PC|}$ have their maximum when $P = S$ and the minimum when $P = s$. And now just add. □

When the orthogonal arc is a diameter and $s(-1), I(0), J(x), S(1),\ x > 0$, then

$$d(I, J) = ln\frac{1+x}{1-x}.$$

We can observe that when $J \to S,\ i.e.\, x \to 1$ then $d(I, J) \to \infty$. The disk becomes unbounded with respect to this distance.

What kind of Geometry do we have inside the disk? Next, we prove that it is a Non-Euclidean one. We can expect at this result taking into account the Cayley–Klein model presented above.

A point $I \in int\Gamma$ is called an *n-point* in our Geometry. The points of the circle Γ are called ∞*-points*.

An orthogonal arc of a circle to Γ is called an *n-line*. Such an n-line is uniquely determined by two n-points, by two ∞-points, or by an ∞-point and an n-point. Two n-lines intersect at most at an n-point. Three n-points are called *n-collinear* if they belong to an n-line.

It is easy to show that there exist non-intersecting n-lines. If two chords Ss and $S's'$ do not intersect in the interior of the disk, then the orthogonal to Γ arcs of circles having the same endpoints are n-lines with empty intersection, that is non-intersecting n-lines. Through an n-point which doesn't belong to a given n-line we can construct at least two non-intersecting n-lines with respect to the given n-line.

In fact, if the given n-line is the orthogonal arc γ_{sS} and $I \notin \gamma_{sS}$, among the infinitely many non-intersecting n-lines there exist two special ones, γ_{sI}, γ_{SI} which are called n-parallels to γ_{sS}.

The *angle between two n-lines* is, by definition, the Euclidean angle between the tangents to the arcs at the common point.

An *n-triangle* is determined by three non-n-collinear points. The sides of an n-triangle are n-lines.

What about the sum of the angles in an n-triangle?

According to the theory described in the previous chapter, it is enough to study what happens in the case of one given triangle. We can choose a triangle with one vertex at the centre O of Γ and two other n-points, A and B. Consider the Euclidean triangle AOB. The sum of the angles of the Euclidean triangle is π. The angle at O, i.e. AOB is common to both triangles, but each other n-angle is less than the corresponding Euclidean angle. Therefore, the sum of the angles of the n-triangle is less than π.

More about this model of Non-Euclidean Geometry and some other models connected to this one can be understood only after we study Differential Geometry.

Chapter 4
Revisiting the Differential Geometry of Surfaces in 3D-Spaces

If you want to understand the infinity you first feel the taste of stars ...

The physicist in preparing for his work needs three things: Mathematics, mathematics, and mathematics.

Wilhelm Röntgen

We intend to present some basic facts related to the Differential Geometry of a surface in a $3D$*-space in the simplest form we imagined. All readers must know basic Calculus and they have to accept from the beginning that we work with functions which are smooth, that is they are indefinitely differentiable functions in one or several variables at each point of their domain of definition.*

The $3D$*-spaces in which we are looking at surfaces are endowed with Euclidean or Minkowski-type metrics induced by quadratic forms attached to bilinear products. We have to remember how the "objects" revealed by the application of Calculus in Geometry offer the new landscape that allows us to correctly understand the foundations of Geometry and to step to Relativity. Simple computations highlight all we need to know about the geometry of surfaces. The change of coordinates preserves the results described by theorems and preserves the nature of geometric "objects" we refer, changing only the space of the geometry studied. Looking at all these surfaces, we intend to obtain models for Euclidean, Non-Euclidean, and Elliptic geometries. This is possible because, in our journey through the Differential Geometry of surfaces, we will understand the crucial role of Theorema Egregium by Gauss.*

This is the approach when the ambient space of the set in which we intend to create a geometry becomes unnecessary. Only the set of coordinates, together with a metric, are necessary to describe the geometry. This point of view is continued in the next short chapter dedicated to basic Differential Geometry. The two chapters

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_4

are both made to simplify, to continue, and to deepen the ideas developed from the first edition of this book, A Mathematical Journey to Relativity (see [34]*).*

4.1 Basic Notations and Definitions of the Geometry of Surfaces

The complexity of formulas in Differential Geometry needs some simplifications. Therefore, before starting with the definitions, we introduce first some necessary notations:

- $a_i b^i$ means $\sum_{i=1}^{n} a_i b^i$. This is the *Einstein summation convention*, or simply *Einstein notation*.
- The index i from the previous formula is called a *dummy index* because we can replace it by some other letter, say s, without changing the meaning. That is $a_s b^s$ means $\sum_{s=1}^{n} a_s b^s$.
- It can be extended for double or triple sums, that is $a_{ij}x^i y^j = \sum_{i,j=1}^{n} a_{ij}x^i y^j$ or $a_{ijk}x^i y^j z^k = \sum_{i,j,k=1}^{n} a_{ijk}x^i y^j z^k$. One can adopt this convention for multiple sums, the sums being thought before the indexes up and down or down and up denoted by the same letter.

  If below one reads something like $\Gamma^i_{sj}\Gamma^s_{kl}$, this means $\sum_{s=1}^{n} \Gamma^i_{sj}\Gamma^s_{kl}$.
- The number n is related to the dimension of the set endowed with a coordinate system, set in which we develop Differential Geometry concepts. In the case of surfaces, $n = 2$.
- The Euclidean three-dimensional space, denoted by E^3, can be thought as the vector space $\mathbb{R}^3$ over the field $\mathbb{R}$ endowed with the *Euclidean inner product*

$$\langle a, b\rangle := a_0 b_0 + a_1 b_1 + a_2 b_2,$$

  where $a = (a_0, a_1, a_2)$, $b = (b_0, b_1, b_2)$. Often we refer to this space as the Euclidean 3D space.
- In a frame generated by the unit vectors $\vec{i} = (1, 0, 0)$, $\vec{j} = (0, 1, 0)$ and $\vec{k} = (0, 0, 1)$, the components of a vector $\vec{a}$ with respect to this basis become coordinates in the new frame, that is, we can assign them to a point A. We can write $A(a_0, a_1, a_2)$ and this point can be seen as the endpoint of the vector $\vec{a}$ whose origin is in the point $(0, 0, 0)$. To simplify, we often write directly a to denote the vector $\vec{a}$.

- *Euclidean perpendicular vectors* correspond to null inner product, i.e. a and b are perpendicular (or orthogonal) if $\langle a, b\rangle = 0$. With respect to the Euclidean inner product, the previous basis is an orthogonal one.
- The *length of the vector* a is, by definition, $||a|| := \sqrt{\langle a, a\rangle} = \sqrt{a_0^2 + a_1^2 + a_2^2}$.
- The Cauchy–Schwarz inequality for the triples (a_0, a_1, a_2), (b_0, b_1, b_2) is

$$(a_0b_0 + a_1b_1 + a_2b_2)^2 \leq (a_0^2 + a_1^2 + a_2^2)(b_0^2 + b_1^2 + b_2^2),$$

  that is, for vectors, the inequality can be written in terms of inner product and norm in the form $\langle a, b\rangle^2 \leq ||a||^2 \cdot ||b||^2$. The equality happens when the triples are proportional: this fact corresponds to collinear vectors.
- If two vectors a and b are not collinear, they determine a plane.
  In this plane it makes sense to define the *angle* α *between the non-zero vectors* a *and* b by the formula

$$\cos\alpha := \frac{\langle a, b\rangle}{||a|| \cdot ||b||}.$$

  Written in the form

$$\langle a, b\rangle = ||a|| \cdot ||b|| \cos\alpha$$

  this formula is known as Pythagoras generalized theorem for the triangle OAB determined by the vectors a and b.
- The *Euclidean distance* between two points $A(a_0, a_1, a_2)$, $B(b_0, b_1, b_2)$ is given by the formula

$$d(A, B) := ||a - b|| = \sqrt{\langle a - b, a - b\rangle} = \sqrt{(a_0 - b_0)^2 + (a_1 - b_1)^2 + (a_2 - b_2)^2}$$

  The length of a vector a becomes the distance between the origin $O(0, 0, 0)$ and the point $A(a_0, a_1, a_2)$. The Euclidean distance is denoted by $||OA||$ or by $|OA|$. We prefer this last notation and we keep in mind that $||AB|| = |AB|$. Looking again to the Pythagoras generalized theorem, according to the new notations, we have

$$|AB|^2 = |OA|^2 + |OB|^2 + 2|OA||OB|\cos\alpha.$$

- The *crossproduct of two vectors* is the vector given by the formula

$$a \times b = (a_1b_2 - a_2b_1, -a_0b_2 + a_2b_0, a_0b_1 - a_1b_0).$$

It is easier to remember it from the formal developing of the following determinant,

$$\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \end{vmatrix}.$$

Since $\langle a \times b, a \rangle = 0$ and $\langle a \times b, b \rangle = 0$ the vector $a \times b$ is orthogonal to the plane determined by the vectors a and b. The following formula holds, $||a \times b|| = ||a|| \cdot ||b|| \cdot \sin \alpha$.

4.2 Surfaces, Tangent Planes and Gauss Frames

After these preliminaries, we can define a surface. The algebraic point of view is related to equations. $X^2 + Y^2 + Z^2 = 1$ is the algebraic definition of a sphere centred at the origin with radius 1. Or $2X + 3Y - Z = 6$ is the equation of a plane, etc.

In Differential Geometry, we deal with smooth functions describing a surface.

The previous sphere can be seen as the smooth function $f : (0, \pi) \times (0, 2\pi) \to E^3$,

$$f(x^1, x^2) = (\sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1), \;\; x^1 \in (0, \pi), \;\; x^2 \in (0, 2\pi).$$

Or, the previous plane can be seen as the smooth function $f : \mathbb{R} \times \mathbb{R} \to E^3$,

$$f(x^1, x^2) = (x^1, x^2, 2x^1 + 3x^2 - 6), \;\; x^1 \in \mathbb{R}, \;\; x^2 \in \mathbb{R}.$$

We will see that the vectors $\dfrac{\partial f}{\partial x^1}$ and $\dfrac{\partial f}{\partial x^2}$ are the key for describing the geometry of the surface f.

- A *surface* in the Euclidean three-dimensional space E^3 is a smooth mapping f of an open set $U \subset \mathbb{R}^2$ into E^3 with an extra property: at each point $f(x)$ of the surface, there must be a tangent plane (Figs. 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7 and 4.8).

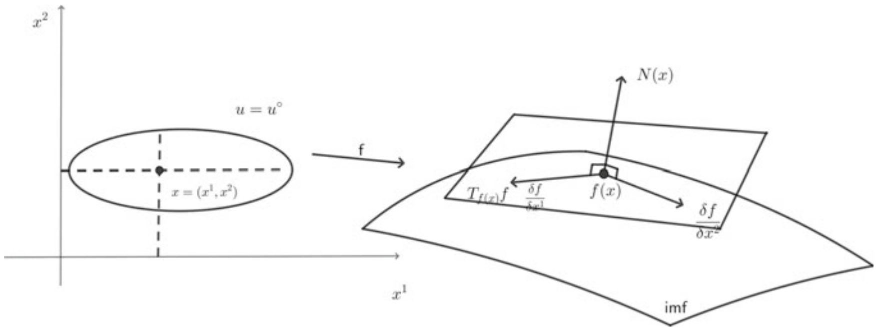


Fig. 4.1 Surface and tangent space

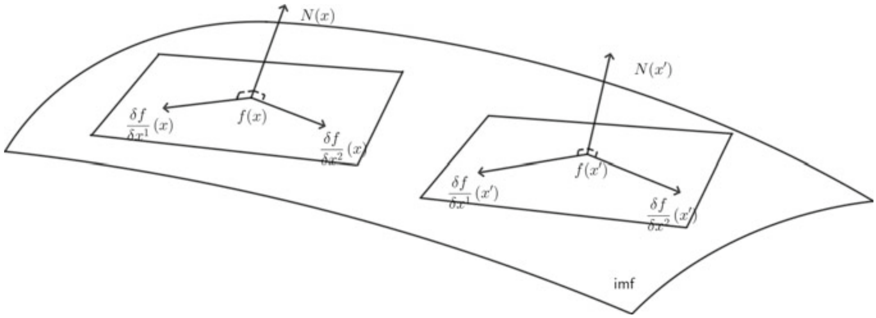


Fig. 4.2 The Gauss frames at $f(x)$ and $f(x')$

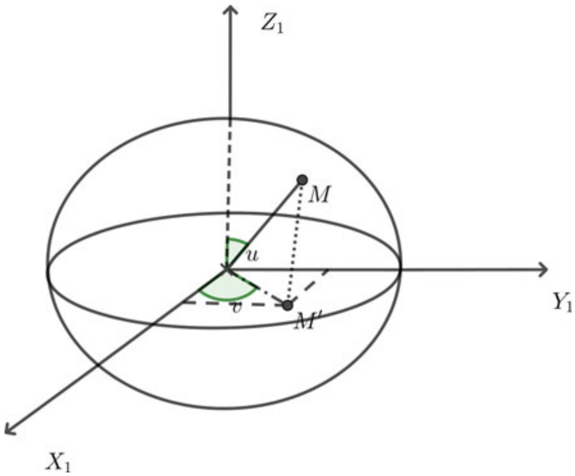


Fig. 4.3 Sphere

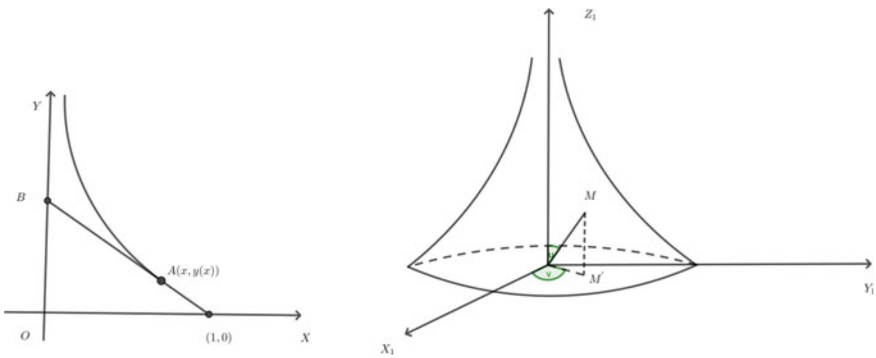


Fig. 4.4 Tractrix and Pseudosphere

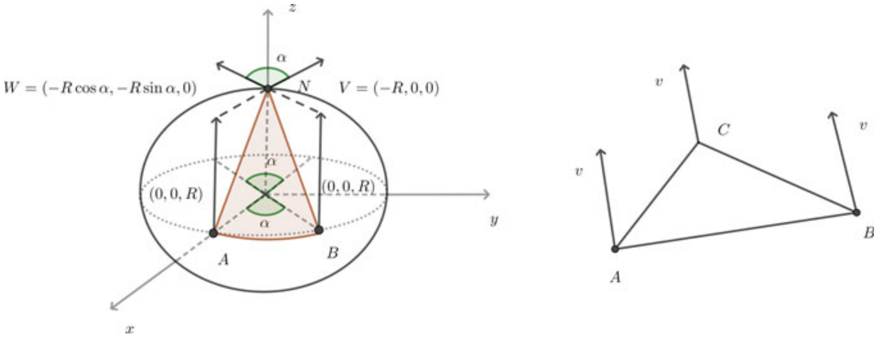


Fig. 4.5 Parallel transport on a sphere and on a plane

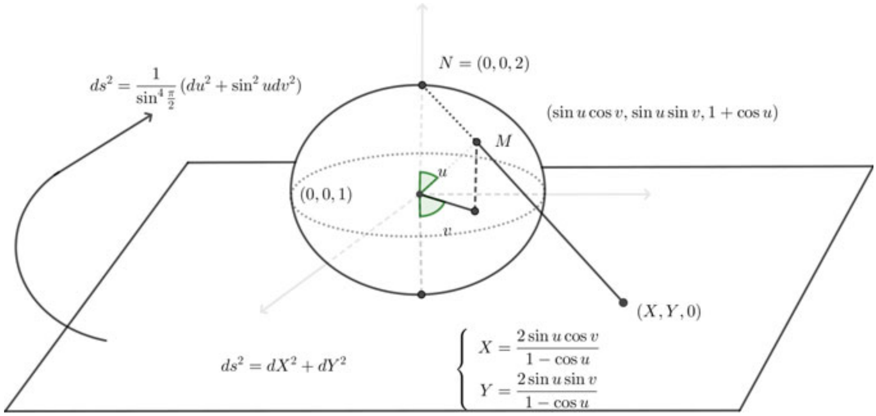


Fig. 4.6 Metric of the sphere induced by $ds^2 = dX^2 + dY^2$

Which is the meaning of this definition?

$f : U \longrightarrow \mathbb{R}^3$ is written as $f(x) = \left(f^1(x), f^2(x), f^3(x)\right)$, where $x = (x^1, x^2)$. The two vectors

$$\frac{\partial f}{\partial x^1}(x) = \left(\frac{\partial f^1}{\partial x^1}(x), \frac{\partial f^2}{\partial x^1}(x), \frac{\partial f^3}{\partial x^1}(x)\right),$$

$$\frac{\partial f}{\partial x^2}(x) = \left(\frac{\partial f^1}{\partial x^2}(x), \frac{\partial f^2}{\partial x^2}(x), \frac{\partial f^3}{\partial x^2}(x)\right)$$

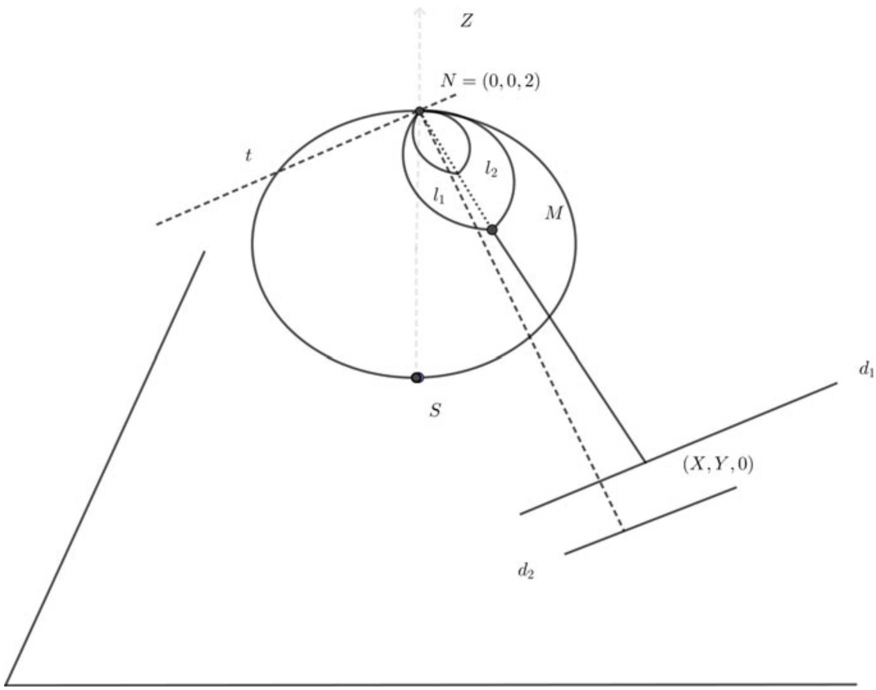


Fig. 4.7 Euclidean parallel lines on a sphere induced by Euclidean parallel lines in a plane

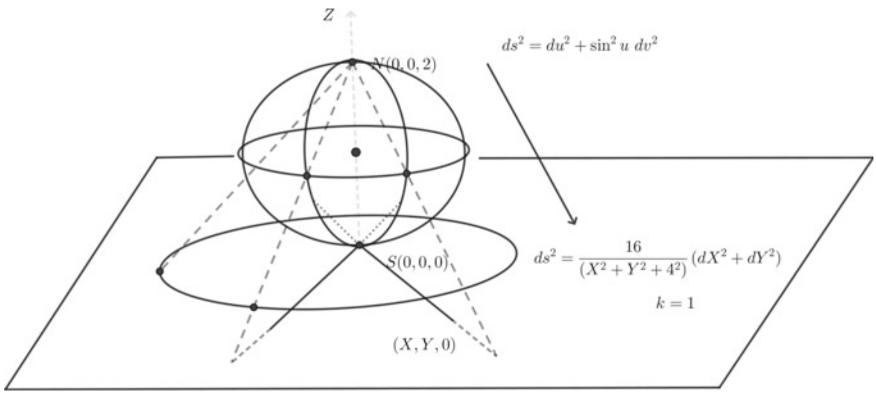


Fig. 4.8 Metric of the plane induced by $ds^2 = du^2 + \sin^2 u \, dv^2$

form the matrix

$$\begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x) & \frac{\partial f^1}{\partial x^2}(x) \\ \frac{\partial f^2}{\partial x^1}(x) & \frac{\partial f^2}{\partial x^2}(x) \\ \frac{\partial f^3}{\partial x^1}(x) & \frac{\partial f^3}{\partial x^2}(x) \end{pmatrix}.$$

If the previous matrix has rank 2, then $\frac{\partial f}{\partial x^1}(x)$ and $\frac{\partial f}{\partial x^2}(x)$ are linear independent vectors and the tangent plane at $f(x)$ exists and it has the equation

$$\begin{vmatrix} X - f^1(x) & Y - f^2(x) & Z - f^3(x) \\ \frac{\partial f^1}{\partial x^1}(x) & \frac{\partial f^2}{\partial x^1}(x) & \frac{\partial f^3}{\partial x^1}(x) \\ \frac{\partial f^1}{\partial x^2}(x) & \frac{\partial f^2}{\partial x^2}(x) & \frac{\partial f^3}{\partial x^2}(x) \end{vmatrix} = 0.$$

- The *tangent plane* is denoted by $T_{f(x)}f$; the linear independent vectors $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\}$ determine a *basis for the tangent plane* $T_{f(x)}f$.
- Any vector $X(x)$ which belongs to $T_{f(x)}f$ can be written in the form

$$X(x) = X^1(x)\frac{\partial f}{\partial x^1}(x) + X^2(x)\frac{\partial f}{\partial x^2}(x)$$

  where the coefficient $X^1, X^2 : U \longrightarrow \mathbb{R}$ are smooth maps.
- The vector $\frac{\partial f}{\partial x^1} \times \frac{\partial f}{\partial x^2}$ generates the normal unitary vector $N(x) := \frac{\frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x)}{\left\| \frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) \right\|}$. $N(x)$ is called the *Gauss map* of the surface f at the point $f(x)$ or simply, the *Gauss vector*. The name of this map is related to the German mathematician Karl Friedrich Gauss, the founder of the Differential Geometry of surfaces. Almost all the results presented in this chapter were discovered by Gauss.
- The frame $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x), N(x)\right\}$ is called a *Gauss frame* attached to the surface f at $f(x)$. At each point of a surface this frame is a vector basis in E^3.

A further comment is necessary about the tangent vectors $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\}$.

Denote by T_xU the two-dimensional vector space having the origin at $x \in U$, determined by the vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$. Since the surface f is the map $f : U \longrightarrow E^3$, it makes sense to consider the linear map $df_x : T_xU \longrightarrow E^3$,

$$df_x = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x) & \frac{\partial f^1}{\partial x^2}(x) \\ \frac{\partial f^2}{\partial x^1}(x) & \frac{\partial f^2}{\partial x^2}(x) \\ \frac{\partial f^3}{\partial x^1}(x) & \frac{\partial f^3}{\partial x^2}(x) \end{pmatrix}.$$

It is easy to see that $df_x(e_1) = \frac{\partial f}{\partial x^1}(x)$, $df_x(e_1) = \frac{\partial f}{\partial x^2}(x)$.

Therefore a vector $X(x) = X^1(x)e_1 + X^2(x)e_2 \in T_xU$ is mapped into the vector

$$df_x(X) = df_x[X^1(x)e_1 + X^2(x)e_2] = X^1(x)\frac{\partial f}{\partial x^1}(x) + X^2(x)\frac{\partial f}{\partial x^2}(x) \in T_{f(x)}f.$$

4.3 The Metric of a Surface

We have now all the ingredients to define the coefficients of the metric of a surface f.

- Denote by $g_{ij}(x) := \left\langle \frac{\partial f}{\partial x^i}(x), \frac{\partial f}{\partial x^j}(x) \right\rangle$.
- It is easy to observe the symmetry of the g_{ij} coefficients, i.e. $g_{ij}(x) = g_{ji}(x)$. Where do these coefficients come from?

  If $X(x) = X^1(x)e_1 + X^2(x)e_2$, $Y(x) = Y^1(x)e_1 + Y^2(x)e_2 \in T_xU$, then the inner product $\langle X(x), Y(x)\rangle$ in T_xU leads to the standard Euclidean 2D inner product

  $$\langle X(x), Y(x)\rangle = \left\langle X^1(x)e_1 + X^2(x)e_2, Y^1(x)e_1 + Y^2(x)e_2 \right\rangle = X^1(x)Y^1(x) + X^2(x)Y^2(x),$$

  because $\langle e_i, e_j\rangle = \delta_{ij}$.

  Now, if we compute $\langle df_xX, df_xY\rangle$ in $T_{f(x)}f$, we obtain

  $$\langle df_xX, df_xY\rangle = \left\langle X^1(x)\frac{\partial f}{\partial x^1}(x) + X^2(x)\frac{\partial f}{\partial x^2}(x), Y^1(x)\frac{\partial f}{\partial x^1}(x) + Y^2(x)\frac{\partial f}{\partial x^2}(x) \right\rangle =$$

  $$= g_{11}(x)X^1(x)Y^1(x) + g_{12}(x)X^1(x)Y^2(x) + g_{21}(x)X^2(x)Y^1(x) + g_{22}(x)X^2(x)Y^2(x),$$

therefore we highlight the g_{ij} coefficients. The Einstein summation convention simplifies the last formula,

$$\langle df_x X, df_x Y\rangle = g_{ij}(x)X^i(x)Y^j(x).$$

A direct consequence of the definition is

$$\langle df_x X, df_x X\rangle = g_{ij}(x)X^i(x)X^j(x) \geqslant 0.$$

- The last formula highlights a quadratic form denoted by ds^2 which acts after the rule

$$ds^2(X, X) = g_{11}\cdot (X^1)^2 + g_{12}\cdot X^1X^2 + g_{21}\cdot X^2X^1 + g_{22}\cdot \left(X^2\right)^2,$$

if the vector X is thought as $X = (X^1, X^2)$. We cancelled the variable x to give the image usually seen in textbooks.
Taking into account that $dx^i(X) = X^i$, the previous formula of the quadratic form can be written in the simplified form

$$ds^2 = g_{11}\cdot \left(dx^1\right)^2 + g_{12}\cdot dx^1dx^2 + g_{21}\cdot dx^2dx^1 + g_{22}\cdot \left(dx^2\right)^2$$

or, using the Einstein notation,

$$ds^2 = g_{ij}dx^idx^j.$$

This quadratic form is called *the metric for the surface* f.
- The coefficients of the metric satisfy

$$det(g_{ij}(x)) = g_{11}(x)g_{22}(x) - g_{12}(x)g_{21}(x) > 0$$

and

$$\left|\left|\frac{\partial f}{\partial x^1}(x)\times\frac{\partial f}{\partial x^2}(x)\right|\right| = \sqrt{det(g_{ij}(x))}.$$

Both assertions are simple to prove. For the first, one takes into account the Cauchy–Schwartz inequality for the non-collinear vectors $\frac{\partial f}{\partial x^1}(x)$ and $\frac{\partial f}{\partial x^2}(x)$. It results

$$det g_{ij}(x) =$$

$$\left\langle\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^1}(x)\right\rangle\left\langle\frac{\partial f}{\partial x^2}(x), \frac{\partial f}{\partial x^2}(x)\right\rangle - \left\langle\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\rangle\left\langle\frac{\partial f}{\partial x^2}(x), \frac{\partial f}{\partial x^1}(x)\right\rangle$$

$$= \left|\left|\frac{\partial f}{\partial x^1}(x)\right|\right|^2\left|\left|\frac{\partial f}{\partial x^2}(x)\right|\right|^2 - \left\langle\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\rangle^2 > 0.$$

The second equality is provided by the previous formula written in the form

$$||a \times b||^2 = ||a||^2 \cdot ||b||^2 \cdot (1 - \cos^2 \alpha)$$

for $a = \dfrac{\partial f}{\partial x^1}(x)$ and $b = \dfrac{\partial f}{\partial x^2}(x)$.

Example 4.3.1 (*Metric of the plane $z = 0$*) Let us consider a plane having the algebraic equation $z = 0$. When we study surfaces from the Differential Geometry point of view, we prefer to write the previous plane as a function. It is $f(x^1, x^2) = (x^1,\ x^2,\ 0),\ x^1 \in \mathbb{R},\ x^2 \in \mathbb{R}$. The metric coefficients

$$g_{ij}(x) = \left\langle \frac{\partial f}{\partial x^i}(x), \frac{\partial f}{\partial x^j}(x) \right\rangle$$

are determined after we compute

$$\frac{\partial f}{\partial x^1}(x) = (1,\ 0,\ 0)$$

$$\frac{\partial f}{\partial x^2}(x) = (0,\ 1, 0),$$

therefore the metric of the plane is

$$ds^2 = (dx^1)^2 + (dx^2)^2$$

as we expected because it is related to the 2D restriction of the quadratic form attached to the 3d Euclidean inner product.

Example 4.3.2 (*Metric of the sphere*) Consider the unit sphere as it appears at the beginning of Sect. 1.2:

$$f(x^1, x^2) = (\sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1),\ x^1 \in (0, \pi),\ x^2 \in (0, 2\pi).$$

In order to compute the metric coefficients $g_{ij}(x)$, we first compute

$$\frac{\partial f}{\partial x^1}(x) = (\cos x^1 \cos x^2, \cos x^1 \sin x^2, -\sin x^1)$$

$$\frac{\partial f}{\partial x^2}(x) = (-\sin x^1 \sin x^2, \sin x^1 \cos x^2, 0).$$

Then we use the formula $g_{ij}(x) = \left\langle \dfrac{\partial f}{\partial x^i}(x), \dfrac{\partial f}{\partial x^j}(x) \right\rangle$ taking into account the Euclidean inner product, i.e. $\langle a, b \rangle := a_0 b_0 + a_1 b_1 + a_2 b_2$.

It results $g_{11}(x) = 1,\ g_{12}(x) = g_{21}(x) = 0,\ g_{22}(x) = \sin^2 x^1$, therefore the metric of the unit sphere is

$$ds^2 = (dx^1)^2 + \sin^2 x^1\ (dx^2)^2.$$

Often, textbooks present such formulas without the upper indexes, that is the sphere is parameterized in the form

$$f(u, v) = (\sin u \cos v, \sin u \sin v, \cos u),\ u \in (0, \pi),\ v \in (0, 2\pi)$$

and the metric is visualized in the simpler form

$$ds^2 = (du)^2 + \sin^2 u\ (dv)^2.$$

At the end of computations, we can say that "using $x^1 = u$ and $x^2 = v$ the formulas obtained are written in the following form..." and the formulas obtained have a more agreeable aspect. One more comment: if the sphere has radius R, i.e.

$$f(u, v) = (R \sin u \cos v, R \sin u \sin v, R \cos u),\ u \in (0, \pi),\ v \in (0, 2\pi)$$

the metric is

$$ds^2 = R^2 (du)^2 + R^2 \sin^2 u\ (dv)^2.$$

Example 4.3.3 (*Metric of the pseudosphere*) Another important example of surface is the pseudosphere proposed by Eugenio Beltrami. The pseudosphere is obtained from the rotation of a *tractrix* curve around an axis. The tractrix is imagined as "a curve whose tangents are all of equal length" in the sense we explain below. At a given point A of the tractrix, we consider the tangent. The tangent intersects the tractrix asymptote at a second point denoted by B. AB is the segment of constant length we refer above. If the initial point is $(1, 0)$ and the asymptote is the $y-$axis, the constant length becomes 1.

Identifying the tractrix equation $y = y(x)$ means to select the point where the tangent line equation

$$Y - y(x) = y'(x) \cdot (X - x)$$

intersects the line $X = 0$. It results $Y = y(x) - x \cdot y'(x)$. The constant length, from the definition

$$(Y - y(x))^2 + x^2 = 1,\ x \in (0, 1)\,,$$

leads to the tractrix differential equation

$$x^2 \cdot (y'(x))^2 + x^2 = 1,\ x \in (0, 1),\ y'(x) < 0.$$

Therefore we have to solve the differential equation

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

The equivalent form $-\dfrac{dy}{dx} = \dfrac{\sqrt{1-x^2}}{x}$ leads to

$$\int dy = -\int \frac{\sqrt{1-x^2}}{x} dx.$$

For $t := \sqrt{1-x^2}$ we have

$$\int \frac{t^2}{1-t^2} dt = -t - \frac{1}{2} \ln \frac{1-t}{1+t},$$

i.e.

$$y(x) = -\int \frac{\sqrt{1-x^2}}{x} dx = -\sqrt{1-x^2} - \ln \frac{1-\sqrt{1-x^2}}{x} + C.$$

The constant C is determined by the condition $y(1) = 0$, that is $C = 0$.

Finally we can consider the equation of the symmetric tractrix with respect to the $x-$axis

$$y(x) = \sqrt{1-x^2} + \ln \frac{1-\sqrt{1-x^2}}{x}.$$

The pseudosphere is obtained when the tractrix is rotated around the $y-$axis and its equation is

$$f(x, y) = \left(x, y, \sqrt{1-(x^2+y^2)} + \ln \frac{1-\sqrt{1-(x^2+y^2)}}{\sqrt{x^2+y^2}} \right), \; x, y \in (-1, 1).$$

We prefer the parameterization: $x^1 \in \left(0, \dfrac{\pi}{2}\right), \; x^2 \in (0, 2\pi)$

$$f(x^1, x^2) = \left(\sin x^1 \cos x^2, \; \sin x^1 \sin x^2, \; \cos x^1 + \ln \left(\tan \left(\frac{x^1}{2} \right) \right) \right),$$

which provides the metric

$$ds^2 = \cot^2 x^1 \left(dx^1\right)^2 + \sin^2 x^1 \left(dx^2\right)^2.$$

Computations start from

$$\frac{\partial f}{\partial x^1}(x) = \left(\cos x^1 \cos x^2, \cos x^1 \sin x^2, \frac{\cos^2 x^1}{\sin x^1}\right)$$

$$\frac{\partial f}{\partial x^2}(x) = (-\sin x^1 \sin x^2, \sin x^1 \cos x^2, 0)\,.$$

The reader is invited to complete the exercise. Again, the form

$$ds^2 = \cot^2 u\,(du)^2 + \sin^2 u\,(dv)^2$$

is more elegant, but not useful in the case when we are interested to find the Gaussian curvature, as we will see below.

4.4 How Metric is Changing with Respect to Changes of Coordinates and Isometries

- If we consider a change of coordinates $\varphi : \overline{U} \longrightarrow U$, our surface f in the new coordinates becomes $\bar{f} = f \circ \varphi : \overline{U} \longrightarrow E^3$. The metric coefficients are preserved by a change of coordinates in a sense we will understand after some computations. If $\bar{x} \in \overline{U}$, $x = \varphi(\bar{x})$, $\overline{X} \in T_x\overline{U}$ and $X = d\varphi_{\bar{x}}\overline{X} \in T_xU$, we have

$$\langle d\bar{f}_{\bar{x}}\overline{X}, d\bar{f}_{\bar{x}}\overline{X}\rangle = \langle d(f \circ \varphi)_{\bar{x}}\overline{X}, d(f \circ \varphi)_{\bar{x}}\overline{X}\rangle = \langle df_x(d\varphi_{\bar{x}}\overline{X}), df_x(d\varphi_{\bar{x}}\overline{X})\rangle = \langle df_x(X), df_x(X)\rangle$$

- An isometry of the Euclidean three-dimensional space E^3 is a map $B : E^3 \longrightarrow E^3$ which preserves distances. The initial surface f is transformed by an isometry into another surface $\tilde{f} = B \circ f : U \longrightarrow E^3$. A vector is transformed by an isometry into another vector having the same length. Two vectors with the same application point M_0 are transformed into two vectors having as application point $B(M_0)$. The transformed vectors have their lengths preserved. It is easy to observe that their initial angle between them is also preserved.
  Taking into consideration all these observations, the metric coefficients g_{ij} are preserved by isometries of the Euclidean space.
- A smooth function $a : I \subset \mathbb{R} \longrightarrow E^3$, $\quad a(t) = (a_1(t), a_2(t), a_3(t))$ is called a curve of the Euclidean three-dimensional space E^3.

- If $x = x(t) = (x^1(t), x^2(t))$ in U, we obtain $f(x(t)) = \big(f^1(x(t)), f^2(x(t)), f^3(x(t))\big)$, that is a one-parameter function with the image contained in the image of our surface. It makes sense to define for $x : I \subset \mathbb{R} \longrightarrow U \subset \mathbb{R}^2$, the map $c := f \circ x : I \subset \mathbb{R} \longrightarrow \mathbb{R}^3$.
  The map c is called a *curve on the surface* $f : U \longrightarrow \mathbb{R}^3$.

- Two properties of a curve on a surface are important.
  The tangent vector $\dot{c}(t)$ belongs to the tangent plane to the surface $T_{c(t)} f$.
  The length of this tangent vector depends on the coefficients g_{ij} of the metric.
  The first one comes from the computation:
  $\dot{c}\,(t) = \frac{d}{dt}\,(f \circ x)\,(t) = \frac{\partial f}{\partial x^1}(x(t)) \cdot \dot{x}^1\,(t) + \frac{\partial f}{\partial x^2}(x(t)) \cdot \dot{x}^2\,(t) \in T_{f(x(t))} f.$
  The second assertion is a consequence of

$$\|\dot{c}\,(t)\|^2 = \langle \dot{c}\,(t)\,, \dot{c}\,(t)\rangle =$$

$$= \left\langle \frac{\partial f}{\partial x^1}(x(t)) \cdot \dot{x}^1\,(t) + \frac{\partial f}{\partial x^2}(x(t)) \cdot \dot{x}^2\,(t)\,, \frac{\partial f}{\partial x^1}(x(t)) \cdot \dot{x}^1\,(t) + \frac{\partial f}{\partial x^2}(x(t)) \cdot \dot{x}^2\,(t)\right\rangle.$$

- Let us observe how the metric of the Euclidean 3D space produces the metric of a surface contained in the 3D Euclidean space. This will allow us to determine the metric of surfaces in a very simple way which will be present in examples.
  If $x = (x^1, x^2)$, the following equality holds

$$\begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x) & \frac{\partial f^2}{\partial x^1}(x) & \frac{\partial f^3}{\partial x^1}(x) \\ \frac{\partial f^1}{\partial x^2}(x) & \frac{\partial f^2}{\partial x^2}(x) & \frac{\partial f^3}{\partial x^2}(x) \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(x) & \frac{\partial f^1}{\partial x^2}(x) \\ \frac{\partial f^2}{\partial x^1}(x) & \frac{\partial f^2}{\partial x^2}(x) \\ \frac{\partial f^3}{\partial x^1}(x) & \frac{\partial f^3}{\partial x^2}(x) \end{pmatrix} =$$

$$= \begin{pmatrix} g_{11}(x) & g_{12}(x) \\ g_{21}(x) & g_{22}(x) \end{pmatrix}.$$

  In fact, the above formula can be written as

$$df_x^T \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot df_x = \begin{pmatrix} g_{11}(x) & g_{12}(x) \\ g_{21}(x) & g_{22}(x) \end{pmatrix}$$

  and represents a metric change at the level of each tangent plane. To continue, it is simple to observe that the Euclidean 3D inner product determines a metric and this metric has the form $ds^2 = dX_1^2 + dY_1^2 + dZ_1^2$. The Euclidean metric

$$ds^2 = dX_1^2 + dY_1^2 + dZ_1^2$$

  endows the surface with a metric induced by the coordinates

$$X_1 = f^1(x^1, x^2);\;\; Y_1 = f^2(x^1, x^2);\;\; Z_1 = f^3(x^1, x^2),$$

  that is with the metric

$$ds^2 = g_{ij}(x)dx^i dx^j.$$

How it works the previous formula when we intend to compute effectively the metric of surfaces?

Example 4.4.1 (*Metric of the sphere computed from the direct transfer of the ambient Euclidean metric*) Let us see the example with the metric of the unit sphere. The metric of the ambient Euclidean 3D space is

$$ds^2 = dX_1^2 + dY_1^2 + dZ_1^2,$$

while

$$X_1 = \sin x^1 \cos x^2,\ Y_1 = \sin x^1 \sin x^2,\ Z_1 = \cos x^1.$$

Therefore

$$dX_1 = \cos x^1 \cos x^2 dx^1 - \sin x^1 \sin x^2 dx^2$$

$$dY_1 = \cos x^1 \sin x^2 dx^1 + \sin x^1 \cos x^2 dx^2,$$

$$dZ_1 = -\sin x^1 dx^1.$$

Computing

$$dX_1^2 + dY_1^2 + dZ_1^2$$

it results the metric of the unit sphere in the form

$$ds^2 = (dx^1)^2 + \sin^2 x^1\ (dx^2)^2.$$

Let see how it works the above approach in the case of the pseudosphere.

Example 4.4.2 (*Metric of the pseudosphere computed from direct transfer of the ambient Euclidean metric*) The surface form

$$f(x^1, x^2) = \left(\sin x^1 \cos x^2,\ \sin x^1 \sin x^2,\ \cos x^1 + \ln\left(\tan\left(\frac{x^1}{2}\right)\right)\right),\ x^1 \in \left(0, \frac{\pi}{2}\right),\ x^2 \in (0, 2\pi),$$

makes us to consider the parameterization

$$X_1 = \sin x^1 \cos x^2,\ Y_1 = \sin x^1 \sin x^2,\ Z_1 = \cos x^1 + \ln\left(\tan\left(\frac{x^1}{2}\right)\right).$$

Therefore

$$dX_1 = \cos x^1 \cos x^2 dx^1 - \sin x^1 \sin x^2 dx^2$$

$$dY_1 = \cos x^1 \sin x^2 dx^1 + \sin x^1 \cos x^2 dx^2,$$

$$dZ_1 = \frac{\cos^2 x^1}{\sin x^1} dx^1.$$

Computing

$$dX_1^2 + dY_1^2 + dZ_1^2$$

it results the metric of the pseudosphere in the form

$$ds^2 = \cot^2 x^1 \left(dx^1\right)^2 + \sin^2 x^1 \left(dx^2\right)^2.$$

So, it becomes clear that the metric of any surface can be easily determined after we compute the differentials $dX_1,\ dY_1$ and dZ_1 of the given parameterization of the surface and we introduce them into the metric of the ambient space, here the Euclidean 3D one: $ds^2 = dX_1^2 + dY_1^2 + dZ_1^2$.

To anticipate, the metric can be obtained with this very simple procedure even if the ambient space is not the Euclidean one. Supposing the metric of the ambient space is $ds^2 = dX_1^2 - dY_1^2 - dZ_1^2$ and replacing there the differentials $dX_1,\ dY_1$ and dZ_1 of the given parameterization of the surface, we obtain the metric of the surface corresponding to the new ambient space.

4.5 Intrinsic Properties of Surfaces

- The *length of a curve* $c = f \circ x : I \longrightarrow E^3$ on the surface f between the points $c(a)$ and $c(b)$ is given by the formula

$$L_c = \int_a^b \|\dot{c}\,(t)\|\, dt.$$

  It can be expressed in terms of the metric in the form

$$L_c = \int_a^b \sqrt{g_{ij}\,(x\,(t)) \cdot \dot{x}^i\,(t) \cdot \dot{x}^j\,(t)}dt.$$

- For two curves $c = f \circ x : I \longrightarrow E^3$ and $\bar{c} : f \circ \bar{x} : \bar{I} \longrightarrow R^3$ on the surface f, the angle between them at the common point $\bar{c}\left(\bar{t}_0\right) = c\,(t_0)$, is the acute angle between the two tangents to the curves at the common point.
  The angle β of the curves c and $\bar{c}$ at their common point $c\,(t_0) = \bar{c}\left(\bar{t}_0\right)$ can be computed by the formula

$$\cos\beta = \frac{\left\langle \dot{c}\left(t_0\right), \dot{\bar{c}}\left(\bar{t}_0\right)\right\rangle}{\left\|\dot{c}\left(t_0\right)\right\| \cdot \left\|\dot{\bar{c}}\left(\bar{t}_0\right)\right\|}$$

that is, it can be expressed in terms of the metric by the formula

$$\cos\alpha = \frac{g_{ij}\left(x\left(t_0\right)\right) \cdot \dot{x}^i\left(t_0\right) \cdot \dot{\bar{x}}^j\left(\bar{t}_0\right)}{\sqrt{g_{rs}\left(x\left(t_0\right)\right) \cdot \dot{x}^r\left(t_0\right) \cdot \dot{x}^s\left(t_0\right)} \cdot \sqrt{g_{pq}\left(\bar{x}\left(\bar{t}_0\right)\right) \cdot \dot{\bar{x}}^p\left(\bar{t}_0\right) \cdot \dot{\bar{x}}^q\left(\bar{t}_0\right)}}.$$

Therefore the lengths of tangent vectors to curves on a surface depend on the coefficients of the metric of the surface; the length of curves on a surface depends on the metric of the surface; the angle between two tangent vectors and, as a consequence, the angle between two curves depends on the metric of the surface. It can be proven that the area of a region on a surface depends on the metric of the surface. The formula for the *area of a region* $f(D),\ D \subset U$ is

$$\sigma_{f(D)} = \iint_D \sqrt{\det\left(g_{ij}\left(x\right)\right)}dx^1dx^2.$$

All the possible geometric properties, depending on the metric of the surface, are called *intrinsic geometric properties* of a surface. Therefore, we can say that

– the length of a curve,
– the angle between two curves,
– the area of a region,

are concepts belonging to the intrinsic geometry of the surface.
The change of coordinates and the isometries preserve the intrinsic nature of geometric properties.

Example 4.5.1 (*Examples of the previous notions in the case of unit sphere*) We start from

$$f(x^1, x^2) = (\sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1),\ x^1 \in (0, \pi),\ x^2 \in (0, 2\pi).$$

Consider the "equator" of the sphere which is determined by $x^1 = \dfrac{\pi}{2}$,

$$c_e(x^2) = (\cos x^2, \sin x^2, 0),\ x^2 \in (0, 2\pi).$$

The tangent vector at each point of it is

$$\dot{c}_e(x^2) = (-\sin x^2, \cos x^2, 0),\ x^2 \in (0, 2\pi).$$

The length L of this curve is computed by the formula

$$L = \int_0^{2\pi} \|\dot{c}_e(t)\| \, dt = \int_0^{2\pi} dt = 2\pi.$$

Same, consider the "meridian" obtained choosing $x^2 = \alpha,\ \alpha \in (0, \pi)$,

$$c_m(x^1) = (\sin x^1 \cos\alpha, \sin x^1 \sin\alpha, \cos x^1),\ x^1 \in (0, \pi).$$

In this case the tangent vector is

$$\dot{c}_m(x^1) = (\cos x^1 \cos\alpha, \cos x^1 \sin\alpha, -\sin x^1),\ x^1 \in (0, \pi).$$

First, let us observe that the two curves meet at $(\cos\alpha, \sin\alpha, 0)$ that is at $c_e(\alpha) = c_m\left(\frac{\pi}{2}\right)$ and

$$\|\dot{c}_e(\alpha)\| = \left\|\dot{c}_m\left(\frac{\pi}{2}\right)\right\| = 1.$$

The angle between the two curves is determined by the formula

$$\cos\beta = \frac{\left\langle \dot{c}_e(\alpha), \dot{c}_m\left(\frac{\pi}{2}\right)\right\rangle}{\|\dot{c}_e(\alpha)\| \cdot \left\|\dot{c}_m\left(\frac{\pi}{2}\right)\right\|},$$

i.e. $\cos\beta = 0$ which implies $\beta = \frac{\pi}{2}$ as we expected.

Let us consider also the "meridian" corresponding to $\alpha = 0$, here denoted by c_0. The two meridians determine a surface on the unit sphere f parameterized with respect to the variables $x^1 \in (0, \pi),\ x^2 \in (0, \alpha)$. Therefore $D := (0, \pi) \times (0, \alpha)$.

The area of this surface is computed by the formula

$$\sigma_{f(D)} = \iint\limits_D \sqrt{\det\left(g_{ij}(x)\right)} dx^1 dx^2 = \int_0^{\pi}\int_0^{\alpha} \sin x^1 dx^1 dx^2 = \alpha \int_0^{\pi} \sin x^1 dx^1 = 2\alpha.$$

4.6 Extrinsic Properties of Surfaces. The Weingarten Equations

The Gauss frame contains also the Gauss map $N(x)$. The concepts involving Gauss map are related to the extra dimension of the surface f. These concepts will be called *extrinsic*,, and the corresponding geometric properties will be called extrinsic geometric properties.

- A first very interesting property we present is related to the partial derivatives of the Gauss map $N(x)$. It can be shown that they belong to the tangent plane $T_{f(x)}f$. Indeed, since the length of Gauss map $N(x)$ is 1, if one considers the partial derivatives of

$$\langle N(x), N(x)\rangle = 1,$$

it results both $\left\langle \frac{\partial N}{\partial x^1}(x), N(x)\right\rangle = 0$ and $\left\langle \frac{\partial N}{\partial x^2}(x), N(x)\right\rangle = 0$.
Therefore the vectors $\frac{\partial N}{\partial x^1}(x)$ and $\frac{\partial N}{\partial x^2}(x)$ are orthogonal to the Gauss vector N at each point $f(x)$ on the surface, i.e. $\left\{\frac{\partial N}{\partial x^1}(x), \frac{\partial N}{\partial x^2}(x)\right\} \subset T_{f(x)}f$.

- Denote $h_{ij}(x) := -\left\langle \frac{\partial N}{\partial x^i}(x), \frac{\partial f}{\partial x^j}(x)\right\rangle$.

- It is relatively easy to show that $h_{12}(x) = h_{21}(x)$. Starting from the relations

$$\left\langle N(x), \frac{\partial f}{\partial x^1}(x)\right\rangle = 0 \text{ and } \left\langle N(x), \frac{\partial f}{\partial x^2}(x)\right\rangle = 0,$$

it results

$$\left\langle \frac{\partial N}{\partial x^2}(x), \frac{\partial f}{\partial x^1}(x)\right\rangle + \left\langle N(x), \frac{\partial^2 f}{\partial x^2 \partial x^1}(x)\right\rangle = 0$$

and

$$\left\langle \frac{\partial N}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\rangle + \left\langle N(x), \frac{\partial^2 f}{\partial x^1 \partial x^2}(x)\right\rangle = 0,$$

that is

$$h_{12}(x) = \left\langle N(x), \frac{\partial^2 f}{\partial x^1 \partial x^2}(x)\right\rangle = \left\langle N(x), \frac{\partial^2 f}{\partial x^2 \partial x^1}(x)\right\rangle = h_{21}(x).$$

- The coefficients $h_{ij}(x)$ are preserved by changes of coordinates and isometries of the Euclidean three-dimensional space, the arguments being the same as for the coefficients g_{ij} of the metric.

- Let us discover some very important relations called Weingarten's relations. They express the fact that the partial derivatives $\frac{\partial N}{\partial x^i}, \frac{\partial N}{\partial x^j}$ belong to the tangent plane $T_{f(x)}f$.
It implies the existence of some coefficients $h^i_j(x)$ $i, j \in 1, 2$, such that

$$-\frac{\partial N}{\partial x^i}(x) = h^s_i(x)\frac{\partial f}{\partial x^s}(x).$$

The two previous formulas (for $i = 1$ and for $i = 2$) are written in the Einstein notation, the dummy index being s. The inner product of the previous relation(s) by $\frac{\partial f}{\partial x^j}(x)$, $j = 1$ and $j = 2$, leads to

$$h_{ij}(x) = \left\langle -\frac{\partial N}{\partial x^i}(x), \frac{\partial f}{\partial x^j}(x) \right\rangle = h_i^s(x) g_{sj}(x).$$

These formulas, i.e. the four relations, are called *Weingarten's formulas* and they can be also written in a matrix form:

$$\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = \begin{pmatrix} h_1^1 & h_1^2 \\ h_2^1 & h_2^2 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}.$$

- The Weingarten matrix $h_i^j(x)$ can also be denoted by W.
- From $det(AB) = det A \cdot det B$, it results $det(h_i^j(x)) = \frac{det(h_{ij}(x))}{det(g_{ij}(x))}$.

Example 4.6.1 (*The Gauss map and Weingarten matrix in the case of plane z = 0*) If $f(x^1, x^2) = (x^1, x^2, 0)$, the Gauss map

$$N(x) := \frac{\frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x)}{\left\| \frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) \right\|}$$

is determined by

$$\frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = (0, 0, 1),$$

i.e. $N(x) = (0, 0, 1)$. The matrix of the metric coefficients is

$$g_{ij}(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the matrix h_{ij} has the form

$$h_{ij}(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The Weingarten matrix is determined by the equality

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} h_1^1 & h_1^2 \\ h_2^1 & h_2^2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

therefore it is

$$h^i_j(x) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Example 4.6.2 (*The Gauss map and Weingarten matrix in the case of the sphere of radius R*) Let us consider the surface

$$f(x^1, x^2) = (R\sin x^1 \cos x^2, R\sin x^1 \sin x^2, R\cos x^1),\ \ x^1 \in (0,\pi),\ \ x^2 \in (0, 2\pi).$$

It allows the partial derivatives

$$\frac{\partial f}{\partial x^1}(x) = (R\cos x^1 \cos x^2, R\cos x^1 \sin x^2, -R\sin x^1),$$

$$\frac{\partial f}{\partial x^2}(x) = (-R\sin x^1 \sin x^2, R\sin x^1 \cos x^2, 0).$$

The matrix of the metric coefficients is

$$g_{ij}(x) = \begin{pmatrix} R^2 & 0 \\ 0 & R^2\sin^2 x^1 \end{pmatrix}.$$

The Gauss map

$$N(x) := \frac{\dfrac{\partial f}{\partial x^1}(x) \times \dfrac{\partial f}{\partial x^2}(x)}{\left\| \dfrac{\partial f}{\partial x^1}(x) \times \dfrac{\partial f}{\partial x^2}(x) \right\|}$$

is now determined by

$$\frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ R\cos x^1 \cos x^2 & R\cos x^1 \sin x^2 & -R\sin x^1 \\ -R\sin x^1 \sin x^2 & R\sin x^1 \cos x^2 & 0 \end{vmatrix} =$$

$$= (R^2 \sin^2 x^1 \cos x^2, R^2\sin^2 x^1 \sin x^2, R^2 \sin x^1 \cos x^1)$$

and $\left\| \dfrac{\partial f}{\partial x^1}(x) \times \dfrac{\partial f}{\partial x^2}(x) \right\| = R^2 \sin x^1$, therefore

$$N(x) = (\sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1).$$

Since $\dfrac{\partial N}{\partial x^i}(x) = \dfrac{1}{R}\dfrac{\partial f}{\partial x^i}(x)$, the matrix h_{ij} has the form

$$h_{ij}(x) = \begin{pmatrix} -R & 0 \\ 0 & -R\sin^2 x^1 \end{pmatrix}.$$

The equality

$$\begin{pmatrix} -R & 0 \\ 0 & -R\sin^2 x^1 \end{pmatrix} = \begin{pmatrix} h_1^1 & h_1^2 \\ h_2^1 & h_2^2 \end{pmatrix} \begin{pmatrix} R^2 & 0 \\ 0 & R^2 \sin^2 x^1 \end{pmatrix}$$

determines the Weingarten matrix

$$h^i_j(x) = \begin{pmatrix} -1/R & 0 \\ 0 & -1/R \end{pmatrix}.$$

4.7 The Gaussian Curvature of Surfaces

When we express two vectors $\{v_1, v_2\}$ of a two-dimensional vector space with respect to a basis $\{e_1, e_2\}$, it is highlighted a 2×2 matrix.

The Weingarten matrix presented above is, as we mentioned, such a matrix which allows to express the vectors $\left\{-\frac{\partial N}{\partial x^1}(x), \frac{\partial N}{\partial x^2}(x)\right\}$ which belong to the two-dimensional vector space $T_{f(x)}f$ with respect to its basis $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\}$.

We present now the definition of one of the most important concepts of the Differential Geometry, the *Gaussian curvature* of a surface f at a point $f(x)$. Denoted by $K(x)$, it is expressed by the formula

$$K(x) := det(h^i_j(x)).$$

The determinant of the Weingarten matrix plays the role of a quantity which has to express the "bending" of a surface at a point.

- A direct consequence is the formula

$$K(x) = \frac{det(h_{ij}(x))}{det(g_{ij}(x))}$$

which allows the quick computation of the curvature of surfaces at a point.

The previous examples show us that

- **The Gaussian curvature of the plane** $z = 0$ **at each point is** $K(x) = 0$**, because** $\det h_{ij}(x) = 0$.
- **The Gaussian curvature of a sphere of radius** R **at each point is** $K(x) = \frac{1}{R^2}$, **which can be computed directly using the Weingarten matrix formula.**

Besides, one can compute it using $\det h_{ij}(x) = R^2 \sin^2 x^1$ **and** $\det g_{ij}(x) = R^4 \sin^2 x^1$.
We leave to the reader two problems related to the computation of the Gaussian curvature. It is important not only to compute it but also to understand how we establish the parameterizations for cylinder and cone surfaces below. The reader has to be inspired from how we established the parameterizations for plane, sphere, and pseudosphere.

Problem 4.7.1 Compute the Gaussian curvature at a point of a circular cylinder.

Hint: $f(x^1, x^2) = (R\cos x^1, R\sin x^1, x^2)$; $K(x) = 0$;

Problem 4.7.2 Compute the Gaussian curvature at a point of a circular cone.

Hint. $f(x^1, x^2) = (x^2\cos x^1, x^2\sin x^1, x^2)$; $K(x) = 0$;

The result in both cases is $K = 0$. We can understand now the word "bending" used above. There is a way of bending a plane, at least locally, to obtain a cylinder or a cone. The curvature $K = 0$ allows this. The question is if there is a similar explanation for surfaces having non-zero Gaussian curvature.
The answer will be obtained after understanding the nature of Gaussian curvature.

- First, let us observe that Gaussian curvature of a surface f at a point $f(x)$ remains invariant under a change of coordinates. This happens because the metric coefficients g_{ij} and the matrix h_{ij} coefficients are preserved at the corresponding points by a change of coordinates.

- The Gaussian curvature is related to the fact that the surface "lives" in the ambient 3D Euclidean space E^3. So, the curvature seems to be an "extrinsic" property of a surface. In order to give more details, the Gaussian curvature depends on h_{ij} matrix whose elements are described with respect to the Gauss map $N(x)$. If we look carefully, there is a determinant involving h_{ij} which is an area of a well chosen parallelogram.
We can prove that this area can be written with respect to the metric components g_{ij}, i.e. the Gaussian curvature is part of the intrinsic geometry of a surface.
Then, only the metric will be important because we can find a procedure to transfer it from a set to another. Say, if we transfer the metric of a plane to a sphere, our intuition regarding the "shape" of surfaces will be no longer important. Because the sphere has its Gaussian curvature depending on the transferred metric from the plane, it is $K = 0$.

In the same way, we can transfer the metric of a given sphere to a plane, offering the Gaussian curvature $K(x) = 1/R^2$ to the plane. The 0 curvature intuition for the plane will disappear.

4.8 The Geometric Interpretation of Gaussian Curvature

Let us first understand the geometric meaning of Gaussian curvature in terms of the ratio of areas.

The vectors $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x), -\frac{\partial N}{\partial x^1}(x), -\frac{\partial N}{\partial x^2}(x)\right\}$, belonging to the tangent plane $T_{f(x)}f$, are related by the Weingarten formulas

$$-\frac{\partial N}{\partial x^i}(x) = h_i^s(x)\frac{\partial f}{\partial x^s}(x), \ i \in \{1, 2\}.$$

Denote

$$e_{ij}(x) := \left\langle \frac{\partial N}{\partial x^1}(x), \frac{\partial N}{\partial x^2}(x) \right\rangle$$

and recall the equality

$$\left\| \frac{\partial N}{\partial x^1}(x) \times \frac{\partial N}{\partial x^2}(x) \right\| = \sqrt{\det(e_{ij}(x))}.$$

Since

$$e_{ij}(x) = \left\langle \frac{\partial N}{\partial x^1}(x), \frac{\partial N}{\partial x^2}(x) \right\rangle = \left\langle h_i^s(x)\frac{\partial f}{\partial x^s}(x), h_j^r(x)\frac{\partial f}{\partial x^r}(x) \right\rangle = h_i^s(x)h_j^r(x)g_{rs}(x)$$

it results

$$\left\| \frac{\partial N}{\partial x^1}(x) \times \frac{\partial N}{\partial x^2}(x) \right\|^2 = \det(e_{ij}(x)) = \det h_i^s(x) \cdot \det h_j^r(x) \cdot \det(g_{rs}(x)),$$

therefore

$$\left\| \frac{\partial N}{\partial x^1}(x) \times \frac{\partial N}{\partial x^2}(x) \right\|^2 = [\det(h_i^j(x))]^2 \cdot \left\| \frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) \right\|^2.$$

The absolute value of the Gaussian curvature of a surface f at the point $f(x)$ is given by the ratio of the areas determined by the vectors $\left\{-\frac{\partial N}{\partial x^1}(x), -\frac{\partial N}{\partial x^2}(x)\right\}$ respectively $\left\{\frac{\partial f}{\partial x^1}(x), \frac{\partial f}{\partial x^2}(x)\right\}$,

$$|K(x)| = \frac{\left\| \frac{\partial N}{\partial x^1}(x) \times \frac{\partial N}{\partial x^2}(x) \right\|}{\left\| \frac{\partial f}{\partial x^1}(x) \times \frac{\partial f}{\partial x^2}(x) \right\|}.$$

Therefore the absolute value of Gaussian curvature is the ratio of two areas of parallelograms determined by the vectors involved in Weingarten's relations.

A consequence of this result is: Gaussian curvature is preserved by isometries of the Euclidean space E^3.

Why? Isometries are maps which preserve distances. And angles, too. Therefore areas of triangles are also preserved. So, the areas of parallelograms are preserved.

4.9 Christoffel Symbols, Riemann Symbols and Gauss Formulas

All the following results, i.e. Gauss formulas, Gauss equations, *Theorema Egregium*, were obtained by Gauss and they look different with respect to the modern view. Elwin Bruno Christoffel and Bernhard Riemann succeeded to simplify the form of Gauss results after they introduced the so-called Christoffel symbols and Riemann symbols we are going to study below.

We define the *Christoffel symbols of first kind*,

$$\Gamma_{ij,k}(x) := \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j}(x) + \frac{\partial g_{jk}}{\partial x^i}(x) - \frac{\partial g_{ij}}{\partial x^k}(x)\right)$$

and the *Christoffel symbols of second kind*, as

$$\Gamma^i_{jk}(x) := g^{is}(x)\Gamma_{jk,s}(x) = \frac{1}{2}\, g^{is}(x)\left(\frac{\partial g_{js}}{\partial x^k}(x) + \frac{\partial g_{ks}}{\partial x^j}(x) - \frac{\partial g_{jk}}{\partial x^s}(x)\right),$$

where $g^{ij}(x)$ is the inverse of the matrix of the coefficients $g_{ij}(x)$ of the metric.

An important observation is the fact that these matrices, which are each other inverse, follow the formula

$$g^{sj}(x)g_{is}(x) = g_{is}(x)g^{sj}(x) = \delta_i^j$$

which is expressed using the Einstein notation.

Of course, one can calculate each g^{ij} exactly as one did it when he studied the inverse of a matrix, that is $g^{11}(x) = \dfrac{g_{22}(x)}{det(g_{ij})}$, etc.

In order to simplify the notation, we cancel x in all the formulas below.

The vector $\dfrac{\partial^2 f}{\partial x^i \partial x^k}$ can be expressed as a linear combination of the Gauss frame vectors $\left\{\dfrac{\partial f}{\partial x^1}, \dfrac{\partial f}{\partial x^2}, N\right\}$, that is

$$\frac{\partial^2 f}{\partial x^i \partial x^k} = A_{ik}^s \cdot \frac{\partial f}{\partial x^s} + a_{ik} \cdot N.$$

The inner product of both members of N leads to $a_{ik} = h_{ik}$, where h_{ik} are the components of the matrix which appeared when we studied the Weingarten matrix. We obtain

$$\frac{\partial^2 f}{\partial x^i \partial x^k} = A_{ik}^s \cdot \frac{\partial f}{\partial x^s} + h_{ik} \cdot N.$$

Now, the inner product of both members of $\dfrac{\partial f}{\partial x^j}$ leads to

$$\left\langle \frac{\partial^2 f}{\partial x^i \partial x^k}, \frac{\partial f}{\partial x^j} \right\rangle = A_{ik}^s \cdot g_{sj},$$

which implies $A_{ik}^s = A_{ki}^s$. On the other hand, if we apply the partial derivative with respect to x^k to the equality $g_{ij} = \left\langle \dfrac{\partial f}{\partial x^i}, \dfrac{\partial f}{\partial x^j} \right\rangle$, we have

$$\frac{\partial g_{ij}}{\partial x^k} = \left\langle \frac{\partial^2 f}{\partial x^k \partial x^i}, \frac{\partial f}{\partial x^j} \right\rangle + \left\langle \frac{\partial f}{\partial x^i}, \frac{\partial^2 f}{\partial x^k \partial x^j} \right\rangle,$$

and this can be written as

$$\frac{\partial g_{ij}}{\partial x^k} = A_{ik}^s \cdot g_{sj} + A_{jk}^s \cdot g_{si}.$$

If $i \to j \to k \to i$ we obtain two further relations

$$\frac{\partial g_{jk}}{\partial x^i} = A_{ji}^s \cdot g_{sk} + A_{ki}^s \cdot g_{sj}$$

$$\frac{\partial g_{ki}}{\partial x^j} = A_{kj}^s \cdot g_{si} + A_{ij}^s \cdot g_{sk}$$

Adding the first two and subtracting the last one, we have

$$\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} = 2A_{ij}^s \cdot g_{sk},$$

and then

$$A_{ij}^r = \Gamma_{ij}^r.$$

Therefore, the Gauss formulas are

$$\frac{\partial^2 f}{\partial x^i \partial x^k} = \Gamma^s_{ik} \cdot \frac{\partial f}{\partial x^s} + h_{ik} \cdot N.$$

The above formulas

$$\frac{\partial g_{ij}}{\partial x^k} = A^s_{ik} \cdot g_{sj} + A^s_{jk} \cdot g_{si}$$

are called *Ricci's equations* and, according to the Christoffel symbols, they are

$$\frac{\partial g_{ij}}{\partial x^k} = \Gamma^s_{ik} \cdot g_{sj} + \Gamma^s_{jk} \cdot g_{si}.$$

We can define

- the *Riemann symbols of second kind* by

$$R^h_{ijk} := \frac{\partial \Gamma^h_{ik}}{\partial x^j} - \frac{\partial \Gamma^h_{ij}}{\partial x^k} + \Gamma^h_{mj}\Gamma^m_{ik} - \Gamma^h_{mk}\Gamma^m_{ij}$$

- the *Riemann symbols of first kind* by $R_{ijkl} := g_{is} R^s_{jkl}$, and
- the *Ricci symbols* by: $R_{ij} := R^s_{isj}$.

  All these symbols depend only on g_{ij}, i.e. they belong to the intrinsic geometry of surfaces.

  Let us first observe that the metric coefficients g_{ij} allow us to lower indexes as in the formula

$$R_{ijkl} = g_{is} R^s_{jkl}.$$

  The components of the inverse matrix of the metric coefficients allow us to rise indexes, that is

$$R^i_{jkl} = g^{is} R_{sjkl}.$$

  The last formula can be derived if we multiply by g_{mi} and we consider the sum after the dummy index i. It results $g_{mi} R^i_{jkl} = g_{mi} g^{is} R_{sjkl}$, that is the equality $R_{mjkl} = R_{mjkl}$ holds.

  If we have a multi-index quantity, say T^{ij}_{lmn}, we can derive $T^j_{\alpha lmn}$ lowering the index i by the rule

$$T^j_{\alpha lmn} := g_{\alpha i} T^{ij}_{lmn}.$$

  The same from $T^j_{\alpha lmn}$: we can obtain T^{ij}_{lmn} rising the index α by the rule

$$T^{ij}_{lmn} := g^{i\alpha} T^j_{\alpha lmn},$$

  etc.

4.10 The Gauss Equations and the *Theorema Egregium*

The result we wish to prove, i.e. the celebrated *Theorema Egregium*, clarifies an important aspect of the Differential Geometry of surfaces. The Gaussian curvature of the surface depends only on the metric. Gauss called this important result *Theorema Egregium*, i.e. the Remarkable Theorem. The result appeared in 1827 in a paper entitled "General Investigations of Curved Surfaces". The original title was in Latin, *Disquisitiones Generales circa Superficies Curvas* [105].

Theorem 4.10.1 *For any surface the following equality holds*

$$R_{ijkl} = h_{ik} \cdot h_{jl} - h_{il} \cdot h_{jk}$$

Proof We consider the partial derivative with respect to x^i of the Gauss formulas

$$\frac{\partial^2 f}{\partial x^j \partial x^k} = \Gamma^s_{jk} \cdot \frac{\partial f}{\partial x^s} + N \cdot h_{jk}.$$

We obtain

$$\frac{\partial^3 f}{\partial x^i \partial x^j \partial x^k} = \frac{\partial \Gamma^s_{jk}}{\partial x^i} \cdot \frac{\partial f}{\partial x^s} + \Gamma^s_{jk} \cdot \frac{\partial^2 f}{\partial x^i \partial x^s} + \frac{\partial N}{\partial x^i} \cdot h_{jk} + N \cdot \frac{\partial h_{jk}}{\partial x^i}.$$

Let us take into consideration the Gauss and Weingarten formulas, in particular

$$\frac{\partial N}{\partial x^i} = -h^r_i \frac{\partial f}{\partial x^r}\,.$$

It results

$$\frac{\partial^3 f}{\partial x^i \partial x^j \partial x^k} = \left(\frac{\partial \Gamma^r_{jk}}{\partial x^i} + \Gamma^r_{is} \cdot \Gamma^s_{jk} - h_{jk} \cdot h^r_i \right) \cdot \frac{\partial f}{\partial x^r} + \left(\frac{\partial h_{jk}}{\partial x^i} + \Gamma^s_{jk} \cdot h_{si} \right) \cdot N.$$

Using $i \to j \to k \to i$, we obtain the formula

$$\frac{\partial^3 f}{\partial x^j \partial x^k \partial x^i} = \left(\frac{\partial \Gamma^r_{ki}}{\partial x^j} + \Gamma^r_{js} \cdot \Gamma^s_{ki} - h_{ki} \cdot h^r_j \right) \cdot \frac{\partial f}{\partial x^r} + \left(\frac{\partial h_{ki}}{\partial x^j} + \Gamma^s_{ki} \cdot h_{sj} \right) \cdot N.$$

Comparing the coefficients of $\dfrac{\partial f}{\partial x^r}$ and N, the following equality holds:

$$\frac{\partial h_{jk}}{\partial x^i} + \Gamma^s_{jk} \cdot h_{si} = \frac{\partial h_{ki}}{\partial x^j} + \Gamma^s_{ki} \cdot h_{sj}$$

for the coefficients of N, and

$$\frac{\partial \Gamma^r_{jk}}{\partial x^i} + \Gamma^r_{is} \cdot \Gamma^s_{jk} - h_{jk} \cdot h^r_i = \frac{\partial \Gamma^r_{ki}}{\partial x^j} + \Gamma^r_{js} \cdot \Gamma^s_{ki} - h_{ki} \cdot h^r_j$$

for the coefficients of $\dfrac{\partial f}{\partial x^r}$. The first equality gives the *Codazzi–Mainardi equations*. The second one can be rearranged in the form

$$\frac{\partial \Gamma^r_{kj}}{\partial x^i} - \frac{\partial \Gamma^r_{ki}}{\partial x^j} + \Gamma^r_{is} \cdot \Gamma^s_{kj} - \Gamma^r_{js} \cdot \Gamma^s_{ki} = h_{jk} \cdot h^r_i - h_{ki} \cdot h^r_j,$$

i.e.

$$R^r_{kij} = h_{jk} \cdot h^r_i - h_{ki} \cdot h^r_j.$$

Multiplying by g_{lr} we obtain the *Gauss equations*

$$R_{lkij} = g_{lr} \cdot R^r_{kij} = g_{lr} \cdot h^r_i \cdot h_{jk} - g_{lr} \cdot h^r_j \cdot h_{ki} = h_{li} \cdot h_{jk} - h_{lj} \cdot h_{ki} = h_{li} h_{kj} - h_{lj} h_{ki}.$$

Therefore, the Gauss equations can be written in the form

$$R_{ijkl} = h_{ik} \cdot h_{jl} - h_{il} \cdot h_{jk}.$$

□

Corollary 4.10.2 *The Riemann symbols R_{ijkl} have the properties*

$$R_{ijkl} = -R_{ijlk};$$

$$R_{ijkl} = -R_{jikl};$$

$$R_{ijkl} = R_{jilk};$$

$$R_{ijkl} = R_{klij};$$

$$R_{ijkl} + R_{iklj} + R_{iljk} = 0 \ (the\ Bianchi\ first\ identity).$$

Proof Let us use the previous Gauss equations

$$R_{ijkl} = h_{ik} \cdot h_{jl} - h_{il} \cdot h_{jk}$$

and some replacements of indexes. □

A consequence of the first relation is $R_{2111} = -R_{2111}$, that is $R_{2111} = 0$. In the same way $R_{1222} = 0$, or in general, if three indexes coincide, then $R_{jiii} = 0$. We may also observe the following relations

$$R_{1212} = -R_{2112} = -R_{1221} = R_{2121}.$$

In the same way $R_{iikl} = -R_{iikl}$, i.e. $R_{iikl} = 0$.

Theorem 4.10.3 (Theorema Egregium) *The Gaussian curvature of a surface depends only on the metric coefficients.*

Proof This is a one line proof:

$$K(x) = \frac{\det\left(h_{ij}(x)\right)}{\det\left(g_{ij}(x)\right)} = \frac{h_{11}(x) \cdot h_{22}(x) - h_{12}^2(x)}{\det\left(g_{ij}(x)\right)} = \frac{R_{1212}(x)}{\det\left(g_{ij}(x)\right)}.$$

□

The previous theorem shows that the Gauss curvature belongs to the intrinsic geometry of a surface.

As we discussed earlier, this important result allows us to think about Differential Geometry in a more general frame, for example, considering sets of coordinates which are not necessarily embedded in a geometric structured space with an extra dimension. The Differential Geometry of such a set will be described only by a "metric tensor", i.e. the matrix g_{ij}, which gives the metric

$$ds^2 = g_{ij}(x)dx^i dx^j.$$

The metric is the only ingredient we need to develop the Differential Geometry on sets of coordinates without extra dimensions.

Example 4.10.4 (*Gaussian curvature of pseudosphere computed by Theorema Egregium*) Let us revisit the surface proposed by Eugenio Beltrami in his 1868 paper [179] on models of hyperbolic geometries, the *pseudosphere*.
Its metric

$$ds^2 = \cot^2 x^1 \left(dx^1\right)^2 + \sin^2 x^1 \left(dx^2\right)^2$$

was obtained before when we studied the tractrix. The aim is to compute its curvature using the Theorema Egregium. The computations start with the Christoffel symbols of the first kind:

$$\Gamma_{22,2} = \Gamma_{12,1} = \Gamma_{21,1} = \Gamma_{11,2} = 0$$

$$\Gamma_{12,2} = \Gamma_{21,2} = -\Gamma_{22,1} = \sin x^1 \cos x^1; \;\; \Gamma_{11,1} = -\cot x^1 \cdot \frac{1}{\sin^2 x^1}.$$

Then, the Christoffel symbols of the second kind:

$$\Gamma^2_{22} = \Gamma^1_{12} = \Gamma^1_{21} = \Gamma^2_{11} = 0$$

$$\Gamma^2_{12} = \Gamma^2_{21} = \cot x^1;\ \Gamma^1_{22} = -\frac{\sin^3 x^1}{\cos x^1};\ \Gamma^1_{11} = -\frac{1}{\sin x^1 \cos x^1}.$$

It results

$$R^1_{212} = \frac{\partial \Gamma^1_{22}}{\partial x^1} - \frac{\partial \Gamma^1_{21}}{\partial x^2} + \Gamma^1_{11}\Gamma^1_{22} + \Gamma^1_{21}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^1_{21} - \Gamma^1_{22}\Gamma^2_{21} = -\sin^2 x^1$$

and

$$R_{1212} = g_{1s}R^s_{212} = g_{11}R^1_{212} = -\cos^2 x^1;\ det(g_{ij}) = \cos^2 x^1,$$

that is $K(x) = -1$.

Example 4.10.5 (*Gaussian curvature of a sphere computed by Theorema Egregium*) In the section devoted to extrinsic geometric properties of a surface, we computed the Gaussian curvature of sphere of radius R and we found that $K(x) = \frac{1}{R^2}$. We intend to compute the same Gaussian curvature, now using the Theorema Egregium. We need only the metric of the sphere,

$$ds^2 = R^2(dx^1)^2 + R^2 \sin^2 x^1\ (dx^2)^2.$$

Having the previous example in mind, now we can directly observe that the only non-zero Christoffel symbols are $\Gamma^1_{22} = -\sin x^1 \cos x^1$ and $\Gamma^2_{12} = \Gamma^2_{21} = \cot x^1$. It results

$$R^1_{212} = \frac{\partial \Gamma^1_{22}}{\partial x^1} - \frac{\partial \Gamma^1_{21}}{\partial x^2} + \Gamma^1_{11}\Gamma^1_{22} + \Gamma^1_{21}\Gamma^2_{22} - \Gamma^1_{12}\Gamma^1_{21} - \Gamma^1_{22}\Gamma^2_{21} = \sin^2 x^1$$

and

$$R_{1212} = g_{1s}R^s_{212} = g_{11}R^1_{212} = R^2 \sin^2 x^1;\ det(g_{ij}) = R^4 \sin^2 x^1,$$

that is $K(x) = \frac{1}{R^2}$.

Let us conclude. In some examples we noticed the existence of surfaces with null Gaussian curvature as the plane, the circular cylinder, and the circular cone. For these surfaces, the Theorema Egregium works easy: the null Gaussian curvature happens because all $\Gamma^i_{jk} = 0$. We also highlighted spheres as surfaces of positive constant Gaussian curvature. On the other hand, the pseudosphere was a first example of a surface with constant negative curvature. Later in this book, we will study the connection between surfaces of constant Gaussian curvature and classical geometries, i.e. Euclidean, Spherical, and Non-Euclidean.

4.11 The Einstein Theorem

Consider the *Ricci symbols* defined by the "contraction" of the Ricci symbol R^i_{jkl}, i.e.

$$R_{ij} := R^s_{isj} = R^1_{i1j} + R^2_{i2j}$$

The name of these symbols is related to the Italian mathematician Gregorio Ricci Curbastro whose work, together with his student Tullio Levi-Civita in tensor calculus, made possible the development of General Relativity language.

Theorem 4.11.1 (Einstein) *For a surface f, the Ricci tensor is proportional to the metric tensor via the Gaussian curvature, i.e.*

$$R_{ij}(x) = K(x) \cdot g_{ij}(x).$$

Proof Let us drop x in the notation and compute the four Ricci symbols.

$$R_{11} = R^s_{1s1} = R^1_{111} + R^2_{121} = 0 + R^2_{121} = R^2_{121}.$$

Since

$$R^2_{121} = g^{2s} \cdot R_{s121},$$

it results

$$R_{11} = g^{21} \cdot R_{1121} + g^{22} \cdot R_{2121} = 0 + g^{22} \cdot R_{2121} =$$

$$= g^{22} \cdot R_{2121} = \frac{g_{11}}{\det\left(g_{ij}\right)} \cdot R_{1212} = \frac{R_{1212}}{\det\left(g_{ij}\right)} \cdot g_{11} = K \cdot g_{11}.$$

In a similar way, starting from the Ricci symbol R_{22}, we have

$$R_{22} = R^s_{2s2} = R^1_{212} + R^2_{222} = R^1_{212} + 0 = R^1_{212}.$$

Then

$$R^1_{212} = g^{1s} \cdot R_{s212}$$

implies

$$R_{22} = g^{11} \cdot R_{1212} + g^{12} \cdot R_{2212} = g^{11} \cdot R_{1212} + 0 = \frac{g_{22}}{\det\left(g_{ij}\right)} \cdot R_{1212} = K \cdot g_{22}.$$

A little bit more complicated is for R_{12}.

$$R_{12} = R^s_{1s2} = R^1_{112} + R^2_{122} = g^{1s} \cdot R_{s112} + g^{2s} \cdot R_{s122} =$$

$$= g^{11} \cdot R_{1112} + g^{12} \cdot R_{2112} + g^{21} \cdot R_{1122} + g^{22} \cdot R_{2122} =$$

$$= 0 + g^{12} \cdot R_{2112} + 0 + 0 = -g^{12} R_{1212} = -\frac{-g_{21}}{\det(g_{ij})} \cdot R_{1212} = \frac{R_{1212}}{\det(g_{ij})} \cdot g_{21} = K \cdot g_{12}.$$

In the same way, we can prove that $R_{21} = K \cdot g_{21}$. □

A first consequence of the Einstein theorem is related to the symmetry of Ricci's symbols for surfaces, so, for a given surface, it results $R_{ij} = R_{ji}$. This result is obtained in the case of two variables. A general result about the symmetry of the Ricci symbols and their geometric nature is presented in the next chapter.
Why this result is related to Einstein? Let us rise one of the indexes,

$$R_j^i = g^{is} R_{sj}$$

and define the *scalar curvature* R or the *Ricci curvature* R by the formula

$$R := R_i^i = R_1^1 + R_2^2.$$

It is easy to see that

$$R_1^1 = R_2^2 = K, \ i.e. \ R = 2K.$$

Therefore the Einstein theorem can be written un the form

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} = 0.$$

The left member of this equality is the *Einstein tensor* and it appears in the Einstein fields equations we will derive latter.

4.12 Covariant Derivative, Parallel Transport and Geodesics

Let $c = f \circ x : I \longrightarrow E^3$ be a curve on the surface f and let $X : I \longrightarrow E^3$ be a smooth map along the curve such that

$$X(t) = X^k(t) \cdot \frac{\partial f}{\partial x^k}(x(t)) \in T_{c(t)} f.$$

Such a smooth map is called a *vector field along the curve* c. Therefore $\frac{dX}{dt}(t)$ is a vector field along the curve c, which, in general, does not belong to $T_{c(t)} f$. Indeed,

$$\frac{dX(t)}{dt} = \dot{X}^k(t) \cdot \frac{\partial f}{\partial x^k}(x(t)) + X^k(t) \cdot \frac{\partial^2 f}{\partial x^j \partial x^k}(x(t)) \cdot \dot{x}^j(t).$$

Using Gauss' formulas

$$\frac{\partial^2 f}{\partial x^j \partial x^k} = \Gamma^i_{jk} \cdot \frac{\partial f}{\partial x^i} + h_{jk} \cdot N$$

after arranging the dummy indexes, we obtain

$$\frac{dX(t)}{dt} = \left[\dot{X}^k(t) + \Gamma^k_{ij}(x(t)) \cdot X^i(t) \cdot \dot{x}^j(t)\right] \frac{\partial f}{\partial x^k}(x(t)) +$$

$$+ X^k(t) \cdot x^j(t) \cdot h_{kj}(x(t)) \cdot N(x(t)).$$

Let us project the vector $\frac{dX(t)}{dt}$ onto $T_{c(t)} f$. This projection makes the normal component to vanish, therefore the vector field, obtained along the curve c, is

$$pr_N \frac{dX(t)}{dt} = \left[\dot{X}^k(t) + \Gamma^k_{ij}(x(t)) \cdot X^i(t) \cdot \dot{x}^j(t)\right] \frac{\partial f}{\partial x^k}(x(t)).$$

Let us denote the vector field obtained in $T_{c(t)}$ as

$$\frac{\nabla X(t)}{dt} := pr_N \frac{dX}{dt}(t).$$

This result means that, at each point $c(t)$ of the curve c, in each $T_{c(t)} f$, the vector $\frac{\nabla X(t)}{dt}$ has the form given by the previous formula. We call this vector field the *covariant derivative of the initial vector field X*.

The covariant derivative of the vector field X is

$$\frac{\nabla X(t)}{dt} = \left[\dot{X}^k(t) + \Gamma^k_{ij}(x(t)) \cdot X^i(t) \cdot \dot{x}^j(t)\right] \frac{\partial f}{\partial x^k}(x(t)).$$

One of the most important definitions of Differential Geometry of surfaces is the *parallel transport of a vector field along a curve*. The parallel transport along a curve c of the vector field X is given by the condition

$$\frac{\nabla X(t)}{dt} = \vec{0}.$$

Therefore, the *equations of the parallel transport* are

$$\dot{X}^k(t) + \Gamma^k_{ij}(x(t)) \cdot X^i(t) \cdot \dot{x}^j(t) = 0, \ k \in \{1, 2\}.$$

The parallel transport equations can be completely determined if we consider an initial condition. It is enough having a point p of the curve and the initial vector V_p

at p. Then, the system of equations has, as a unique solution, the vector field X such that, at $p = c(t_0)$, it is $X(t_0) = V_p$.

Let us underline the following point. The system of differential equations

$$\dot{X}^k (t) + \Gamma_{ij}^k (x (t)) \cdot X^i (t) \cdot \dot{x}^j (t) = 0, \quad k \in \{1, 2\}$$

describing the parallel transport shows that the vector field

$$X (t) = X^i (t) \cdot \frac{\partial f}{\partial x^i} (x (t))$$

is completely determined if we know X at a given (known) point of the curve $c (t) = (f \circ x) (t)$.

Example 4.12.1 The case of parallel transport along any curve of the plane We may consider, without loss of generality, that the algebraic equation of the plane is $z = 0$. Then for the surface f is $f(x^1, x^2) = (x^1, x^2, 0)$, the metric is

$$ds^2 = (dx^1)^2 + (dx^2)^2,$$

all $\Gamma_{ij,k} = 0$, all $\Gamma_{ij}^k = 0$, and the equations of the parallel transport are

$$\dot{X}^k (t) = 0, \; k \in \{1, 2\}.$$

It results, by integration, that the vector field X is a constant one, i.e. $X(t) = (a, b)$. If we are looking at the support lines of this vector field along a given line, we see Euclidean parallel lines, therefore we understand the meaning of the parallel transport above.

Example 4.12.2 Parallel transport on sphere Consider the sphere

$$ds^2 = R^2(dx^1)^2 + R^2 \sin^2 x^1 \, (dx^2)^2$$

and its non-zero Christoffel symbols $\Gamma_{22}^1 = -\sin x^1 \cos x^1$ and $\Gamma_{12}^2 = \Gamma_{21}^2 = \cot x^1$. The equations of the parallel transport along the curve $c(t) = f(x(t))$, where $x(t) = (x^1(t), x^2(t)) \subset U$, are

$$\dot{X}^1 (t) - \sin x^1(t) \cos x^1(t) \cdot X^2 (t) \cdot \dot{x}^2 (t) = 0$$

$$\dot{X}^2 (t) + \cot x^1(t) \cdot X^1 (t) \cdot \dot{x}^2 (t) + \cot x^1(t) \cdot X^2 (t) \cdot \dot{x}^1 (t) = 0.$$

Let us choose the equator of the sphere obtained for $x(t) = \left(\frac{\pi}{2}, \; t\right)$. Since $x^1(t) = \frac{\pi}{2}$, the equations of the parallel transport become

$$\dot{X}^k (t) = 0, \; k \in \{1, 2\}$$

with the solutions $X^1(t) = a$ and $X^2(t) = b$. Therefore, taking into account that along the equator the vectors $\dfrac{\partial f}{\partial x^i}(x(t)), i \in \{1, 2\}$ are

$$\frac{\partial f}{\partial x^1}\left(\frac{\pi}{2}, t\right) = (0, 0, -R),$$

$$\frac{\partial f}{\partial x^2}\left(\frac{\pi}{2}, t\right) = (-R \sin t, R \cos t, 0),$$

the parallel transported vector along the equator has the form

$$X(t) = a \cdot \frac{\partial f}{\partial x^1}\left(\frac{\pi}{2}, t\right) + b \cdot \frac{\partial f}{\partial x^2}\left(\frac{\pi}{2}, t\right) = (-bR \sin t, bR \cos t, -aR).$$

A first example is the vector $(0, 0, R)$ which is obtained when $a = 0$ and $b = -1$. This vector is always parallel to z-axis and it is parallel transported at each point along the equator. Another important example is the tangent vector to the equator, $\dot{c}_e(t) = (-\sin t, \cos t, 0)$ which is obtained when $a = \dfrac{1}{R}$ and $b = 0$. The tangent vector is parallel transported at each point along the equator.

In the particular case when the tangent field $\dot{c}(t)$ to the curve c is parallel transported along the curve, then the curve c, by definition, is called *a geodesic of the surface f*.

The previous example shows that the equator of the sphere is a geodesic of the sphere because the tangent vector to equator is parallel transported. In an example below we will revisit the subject in the new formalism we are going to present.

The equations of a geodesic are $\dfrac{\nabla \dot{c}(t)}{dt} = \vec{0}$. In fact, as above, there are two equations and we highlighted this by using the arrow to define the null vector. Since $\dot{c}(t) = \dot{\widehat{(f \circ x)}}(t) = \dot{x}^i(t) \dfrac{\partial f}{\partial x^i}(x(t))$, the equations are

$$\ddot{x}^k(t) + \Gamma^k_{ij}(x(t)) \cdot \dot{x}^i(t) \cdot \dot{x}^j(t) = 0, \quad k \in \{1, 2\}.$$

The system of equations for geodesics is completely determined if we consider an initial condition which is similar to the one presented in the case of the parallel transport:

- a point $c(t_0)$ of the geodesic;
- the initial vector $v_{c(t_0)} = (\dot{x}^1(t_0), \dot{x}^2(t_0))$.

Example 4.12.3 Geodesic of the plane Let us show that the geodesics of the plane $z = x^3 = 0$ are lines.

Since $f(x^1, x^2) = (x^1, x^2, 0)$, all $\Gamma^k_{ij} = 0$ and the equations of the geodesics are $\ddot{x}^k(t) = 0, k \in \{1, 2\}$.

It results, by integration, the curve $c(t) = (v_1 t + x_0^1, v_2 t + x_0^2, 0)$ which is a line in the Euclidean plane $z = 0$. Along the curve, the speed is constant, i.e. $||\dot{c}(t)|| = \sqrt{(v_1)^2 + (v_2)^2}$.

Example 4.12.4 Geodesic of the cylinder Find the geodesics of the cylinder $f(x^1, x^2) = (\cos x^1, \sin x^1, x^2)$.

Following the same way as in the previous example, the geodesics are helices of the cylinder f,

$$c(t) = (\cos(v_1 t + \alpha_0), \sin(v_1 t + \alpha_0), v_2 t + \beta_0),$$

and the speed along the geodesic is constant, $||\dot{c}(t)|| = \sqrt{(v_1)^2 + (v_2)^2}$.

Example 4.12.5 Geodesic of the sphere In this case, we have to study the geodesics of the metric

$$ds^2 = R^2 dx^2 + R^2 \sin^2 x dy^2.$$

In this example, we write the metric without upper indexes for variables. Clearly, this metric is the one of a sphere with radius R.

In order to perform the calculations, let us consider $x = x^1,\ y = x^2$, i.e. $g_{11} = R^2,\ g_{22} = R^2 \sin^2 x_1,\ g_{12} = g_{21} = 0,\ g^{11} = \dfrac{1}{R^2},\ g^{22} = \dfrac{1}{R^2 \sin^2 x_1},\ g^{12} = g^{21} = 0.$

The Christoffel symbols are

$$\Gamma_{11,1} = \Gamma_{11,2} = \Gamma_{12,1} = \Gamma_{21,1} = \Gamma_{22,2} = 0,\ \Gamma_{12,2} = \Gamma_{21,2} = R^2 \sin x_1 \cos x_1 = -\Gamma_{22,1}$$

$$\Gamma_{11}^1 = \Gamma_{11}^2 = \Gamma_{12}^1 = \Gamma_{21}^1 = \Gamma_{22}^2 = 0,\ \Gamma_{12}^2 = \Gamma_{21}^2 = \cot x_1,\ \Gamma_{22}^1 = -\sin x_1 \cos x_1.$$

Now we can replace the variables with the initial ones. The geodesic equation, written for the first variable, is

$$\ddot{x} - \sin x \cos x \cdot (\dot{y})^2 = 0.$$

For the second variable y, the geodesic equation is

$$\ddot{y} + 2 \cot x \cdot \dot{x}\dot{y} = 0.$$

We may observe that $x = \dfrac{\pi}{2},\ \ y = s$ is a solution, therefore $c(s) = (R \cos s, R \sin s, 0)$ is a geodesic of the sphere. It is the great circle obtained at the intersection of the plane $z = 0$ with the sphere. We found this geodesic and presented it in the example related to the parallel transport. If we rotate the sphere around its centre, another great circle replaces the equator and therefore becomes a geodesic. So the geodesics of the spheres are the great circles of the sphere. The terminology is related to the fact that among all circles obtained at the intersection of a sphere by planes, the maximum radius is for the circles obtained at the intersection of planes passing through the centre of the sphere.

A comment is important at this point: a rotation of the sphere around the origin means a change of coordinates of the sphere. The above result can be reformulated as "the change of coordinates mapping a geodesic into another geodesic". This result is proved in the next chapter but we mentioned it now to show that all great circles are geodesics of the sphere.

In all examples, the tangent vector to geodesic has constant length. This is a general property for geodesics and the proof of this fact is also provided in the next chapter.

Before continuing, it is better to add something about curves in a $3D$-Euclidean space. The differential properties of curves are described by a special frame, called the *Frenet frame*. Such a frame depends on the vectors $\dot{c}(t)$ and $\ddot{c}(t)$ which describe the osculating plane. Using these two vectors one can construct another two vectors, $\{e_1(t), e_2(t)\}$, which form an orthonormal frame in the osculating plane. Considering $e_3 := e_1 \times e_2$ the set $\{e_1(t), e_2(t), e_3(t)\}$ becomes the Frenet frame at each point $c(t)$ of the curve. The Frenet frame describes the curvature $K_1(t)$ and the torsion $K_2(t)$ at each point of the curve. Curves included in a plane are characterized by null torsion. In this case $e_3(t)$ is a constant vector along c and e_1 and e_2 are obviously included in the plane where the curve lies.

If the curve lies on a surface f, we can define the *Darboux frame*. This is formed by the previous unit speed vector $e_1 = \dfrac{\dot{c}}{||\dot{c}||}$ and another two vectors. One of them is the Gauss vector $N(t)$ which is part of the Gauss frame and $e(t) := N(t) \times e_1(t)$. The frame $\{e_1(t), N(t), e(t)\}$ coexists together with the Frenet frame if a curve is part of a surface. A very nice result which can be proved is the following:

> The curve c is a geodesic of f if and only if $e_2(t)$ and $N(t)$ are collinear vectors.

To prove this statement, let us consider a sphere with a great circle. This circle is included in a plane, therefore the torsion of it is 0 at each point $c(t)$. It results that e_1 is tangent to the circle and e_2 is included in the plane of the curve such that the line containing e_2 passes through the centre of the sphere. The same line contains the Gauss vector N, therefore N and e_2 are collinear. The great circle is a geodesic.

The theory presented above is true in the Minkowski spaces, therefore planes containing the origin intersects Minkowski spheres after geodesic curves. More properties about the Minkowski spaces will be discussed below, when we will consider the possibility that the ambient $3D$ space is not an Euclidean one.

Example 4.12.6 Gaussian curvature and parallel transport Consider the previous sphere, its equator and two meridians. The meridians are part of great circles, therefore they are geodesics. The same holds for the equator.

Denote by N the "North Pole" of the sphere and by A and B the intersections between the two meridians and the equator. In this way, we obtained the spherical triangle ABN.

We can choose the coordinates of the points N, A and B as $(0, 0, R)$, $(R, 0, 0)$ and $(\cos\alpha, \sin\alpha, 0)$ respectively.

Consider the vector $(0, 0, R)$. This vector is a tangent vector to the first meridian at A.

If it is parallel transported to the North Pole N, it becomes $V = (-R, 0, 0)$. If now we parallel transport it from A to B, it remains $(0, 0, R)$. Finally, if we parallel transport it from B to N, it becomes $W = (R\cos(\pi+\alpha), R\sin(\pi+\alpha), 0)$. The angle between V and W is α.

Let us observe that if we act in the Euclidean plane and we parallel transport a given vector v along the contour of the ABC triangle in the same way, that is from A to B, respectively, from A to C then from C to B, it results in the same vector v, therefore the angle between the two final vectors is 0. The null Gaussian curvature of the plane allows this result.

The angle α between V and W obtained in the case of the sphere is related to the curvature of the sphere.

4.13 Changes of Coordinates

Let

$$ds^2 = g_{ij}(x)dx^i dx^j$$

be a metric written with the Einstein notation and $x = (x^1, x^2)$ be a point which belongs to $U \subset \mathbb{R}^2$. Suppose we wish to change the coordinates, i.e. we consider the transform

$$\begin{cases} x^1 = x^1(\overline{x}^1, \overline{x}^2) \\ x^2 = x^2(\overline{x}^1, \overline{x}^2), \end{cases}$$

where $\overline{x} = (\overline{x}^1, \overline{x}^2)$ is a point belonging to $\overline{U} \subset \mathbb{R}^2$. The transformation can be described by $x = x(\overline{x})$ or, another possible notation is $x^h = x^h(x^r)$, $h \in \{1, 2\}$, $r \in \{1, 2\}$.

How the metric coefficients change with respect to the new coordinates?

The old g_{ij} depending on $x = (x^1, x^2)$ are replaced by new coefficients $\overline{g}_{kl}$ depending on $\overline{x} = (\overline{x}^1, \overline{x}^2)$. Therefore we are interested to find the new metric form

$$d\overline{s}^2 = \overline{g}_{kl} d\overline{x}^k \overline{x}^l$$

corresponding to the coordinate transformation.

To do this it is enough to observe that $dx^i = \dfrac{\partial x^i}{\partial \overline{x}^k} d\overline{x}^k$ and $dx^j = \dfrac{\partial x^j}{\partial \overline{x}^l} d\overline{x}^l$, i.e.

$$g_{ij}(x)dx^i dx^j = g_{ij}(x(\overline{x}))\frac{\partial x^i}{\partial \overline{x}^k}\frac{\partial x^j}{\partial \overline{x}^l} d\overline{x}^k d\overline{x}^l = \overline{g}_{kl} d\overline{x}^k \overline{x}^l,$$

so the coefficients are changed according to the rule $\overline{g}_{kl} = g_{ij}\dfrac{\partial x^i}{\partial \bar{x}^k}\dfrac{\partial x^j}{\partial \bar{x}^l}$.

First of all, let us observe that the number of variables does not affect the previous considerations. If the change of coordinates is $x^h = x^h(x^r),\ h \in \{1, 2, 3, ..., n\},\ r \in \{1, 2, 3, ..., n\}$, the formulas $dx^i = \dfrac{\partial x^i}{\partial \overline{x}^k} d\overline{x}^k$ and $dx^j = \dfrac{\partial x^j}{\partial \overline{x}^l} d\overline{x}^l$ hold. Therefore the same happens with the equality

$$g_{ij}(x)dx^i dx^j = g_{ij}(x(\overline{x}))\frac{\partial x^i}{\partial \overline{x}^k}\frac{\partial x^j}{\partial \overline{x}^l} d\overline{x}^k d\overline{x}^l = \overline{g}_{kl} d\overline{x}^k \overline{x}^l.$$

The transformation formula $\overline{g}_{kl} = g_{ij}\dfrac{\partial x^i}{\partial \bar{x}^k}\dfrac{\partial x^j}{\partial \bar{x}^l}$ is preserved for any number of variables.

Let us give some examples.

Example 4.13.1 Euclidean metric in polar coordinates Let see how the Euclidean metric

$$ds^2 = dx^2 + dy^2$$

transforms if

$$\begin{cases} x = x(r, \theta) = r\cos\theta \\ y = y(r, \theta) = r\sin\theta. \end{cases}$$

We have

$$\begin{cases} dx = dr\cos\theta - r\sin\theta d\theta \\ dy = dr\sin\theta + r\cos\theta d\theta, \end{cases}$$

therefore

$$d\overline{s}^2 = dr^2 + r^2 d\theta^2.$$

This is the Euclidean metric in polar coordinates. Some textbooks use ds^2 instead of $d\overline{s}^2$ because, if you look at the proof above, you can see explicitly the equality between the two forms.

This is also an argument for the statement: a change of coordinates preserves the Gaussian curvature at corresponding points.

Example 4.13.2 Transforming the metric of the pseudosphere into the metric of the Poincaré half-plane Let us consider the metric of the pseudosphere

$$ds^2 = \cot^2 u\, du^2 + \sin^2 u\, dv^2,$$

where $u \in (0, \frac{\pi}{2}),\ v \in (0, 2\pi)$. The transformation of coordinates

$$\begin{cases} u = u(x, y) = \arcsin \frac{1}{y} \\ v = v(x, y) = x,\ x \in \mathbb{R},\ y > 0 \end{cases}$$

leads to

$$\begin{cases} du = -\frac{1}{y\sqrt{y^2 - 1}} dy \\ dv = dx \end{cases}$$

and the two equalities $\sin^2 u = \frac{1}{y^2};\ \cot^2 u = y^2 - 1$. It results

$$d\overline{s}^2 = \frac{dx^2 + dy^2}{y^2},$$

which, as we will see later, is the metric of the Poincaré half-plane involved in the description of a Non-Euclidean geometry model. Therefore the pseudosphere is related to the Non-Euclidean geometry.

Let us see how the Christoffel symbols change. We start from the same observation: the obtained formulas are the same even if our change of coordinates is $x^r = x^r(\overline{x}^h)$, $r \in \{0, 1, ..., n\}$, $h \in \{0, 1, ..., n\}$.

The Christoffel symbols of first kind

$$\Gamma_{ij,k} = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right)$$

transform under the rule

$$\overline{\Gamma}_{ij,k} = \Gamma_{rs,p} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \frac{\partial x^p}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k}.$$

This formula is not difficult to obtain. See also [34]. It is important because it offers the possibility to obtain the transformation of Christoffel symbols of the second kind, $\Gamma^i_{jk} = g^{is}\Gamma_{jk,s}$. In the same Ref. [34], it is proven that Christoffel symbols of the second kind transform under the rule

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\Gamma^k_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\Gamma}^r_{ij} \frac{\partial x^k}{\partial \overline{x}^r}.$$

The last two formulas are necessary to know how Riemann and Ricci symbols transform under a change of coordinates. They are

$$\begin{cases} \overline{R}^a_{bgd}\frac{\partial x^i}{\partial \overline{x}^a} = R^i_{jkl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d} \\ \overline{R}_{ebgd} = R_{rjkl}\frac{\partial x^r}{\partial \overline{x}^e}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d} \\ \overline{R}_{bg} = R_{jl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^l}{\partial \overline{x}^g} \end{cases}$$

and a complete proof can be seen in [34] or in Chap. 13 when we discuss the Levi-Civita connection. Here, we propose an argument for the coordinates transformation in the case of Ricci symbols.

Let us start from the definition $\overline{R}_{bg} = \overline{R}^a_{bad}$. Then $\overline{R}_{bg}\frac{\partial x^i}{\partial \overline{x}^a} = \overline{R}^a_{bad}\frac{\partial x^i}{\partial \overline{x}^a} = R^i_{jil}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^i}{\partial \overline{x}^a}\frac{\partial x^l}{\partial \overline{x}^g}$, therefore

$$\overline{R}_{bg} = R_{jl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^l}{\partial \overline{x}^g}.$$

It is worth observing that for

- the metric coefficients g_{ij}
- the Riemann symbols of second kind R^i_{jkl}
- the Riemann symbols of first kind R_{ijkl}
- the Ricci symbols R_{ij}

the change of coordinates has a sort of regularity. In the next chapter we find out that all these geometric objects are tensors and they describe the language of General Relativity.

In the next chapter we prove both the tensorial aspect of the geodesic equations and the fact that a geodesic with respect to a metric is transformed by a change of coordinates into a geodesic of the transformed metric.

4.14 What if the Ambient Space is Not an Euclidean One?

Differential Geometry of surfaces was developed in the case when the ambient space is an Euclidean one. The Euclidean metric

$$ds^2 = dX^2 + dY^2 + dZ^2$$

transfers its geometry to surfaces contained in it. We know how it works, the parameterization of the surface

$$S : \begin{cases} X = X(x^1, x^2) \\ Y = Y(x^1, x^2) \\ Z = Z(x^1, x^2) \end{cases}$$

leads to

$$\begin{cases} dX = \dfrac{\partial X}{\partial x^1}dx^1 + \dfrac{\partial X}{\partial x^2}dx^2 \\ dY = \dfrac{\partial Y}{\partial x^1}dx^1 + \dfrac{\partial Y}{\partial x^2}dx^2 \\ dZ = \dfrac{\partial Z}{\partial x^1}dx^1 + \dfrac{\partial Z}{\partial x^2}dx^2, \end{cases}$$

therefore the metric of the surface results from the following equalities

$$ds^2 = \left[\left(\frac{\partial X}{\partial x^1}\right)^2 + \left(\frac{\partial Y}{\partial x^1}\right)^2 + \left(\frac{\partial Z}{\partial x^1}\right)^2\right](dx^1)^2 + \left[\frac{\partial X}{\partial x^1}\frac{\partial X}{\partial x^2} + \frac{\partial Y}{\partial x^1}\frac{\partial Y}{\partial x^2} + \frac{\partial Z}{\partial x^1}\frac{\partial Z}{\partial x^2}\right]dx^1dx^2 +$$

$$+ \left[\frac{\partial X}{\partial x^2}\frac{\partial X}{\partial x^1} + \frac{\partial Y}{\partial x^2}\frac{\partial Y}{\partial x^1} + \frac{\partial Z}{\partial x^2}\frac{\partial Z}{\partial x^1}\right]dx^2dx^1 + \left[\left(\frac{\partial X}{\partial x^2}\right)^2 + \left(\frac{\partial Y}{\partial x^2}\right)^2 + \left(\frac{\partial Z}{\partial x^2}\right)^2\right](dx^2)^2 =$$

$$= g_{ij}dx^i dx^j.$$

After obtaining the metric, we have all we need for the intrinsic geometry of a surface because we can compute:

- The Christoffel symbols of first kind:

$$\Gamma_{ij,k} := \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right);$$

- The Christoffel symbols of second kind:

$$\Gamma^i_{jk} := g^{is}\Gamma_{jk,s} = \frac{1}{2}\,g^{is}\left(\frac{\partial g_{js}}{\partial x^k} + \frac{\partial g_{ks}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^s}\right);$$

- The Riemann symbols of second kind:

$$R^h_{ijk} := \frac{\partial \Gamma^h_{ik}}{\partial x^j} - \frac{\partial \Gamma^h_{ij}}{\partial x^k} + \Gamma^h_{mj}\Gamma^m_{ik} - \Gamma^h_{mk}\Gamma^m_{ij}$$

- The Riemann symbol of first kind:

$$R_{ijkl} := g_{is}R^s_{jkl};$$

- The Ricci symbols

$$R_{ij} = R^s_{isj},$$

  which are obtained from Riemann symbols of second kind R^s_{imj} by contracting the indexes $s = m$.

- The geodesics, i.e. curves on the surface such that the coordinates satisfy the equations:

$$\ddot{x}^k(t) + \Gamma^k_{ij}(x(t)) \cdot \dot{x}^i(t) \cdot \dot{x}^j(t) = 0, \quad k \in \{1, 2\}.$$

  Let us observe that the previous equations can be written in the form

$$\frac{d^2x^k}{dt^2} + \Gamma^k_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \quad k \in \{1, 2\}.$$

- We can also compute the Gaussian curvature using the formula

$$K(x) = \frac{R_{1212}}{\det g_{ij}}.$$

- Furthermore, we can compute angles between geodesics or surfaces of triangles determined by geodesics. We can study the geometry of such surfaces as done in the numerous examples previously presented.

- How can we proceed if the surface is contained in a space endowed with a metric as

$$ds^2 = dX^2 - dY^2 - dZ^2 \text{ ?}$$

  This space transfers its geometry to surfaces contained in it exactly in the same way as the Euclidean space. Only the form of the metric coefficients is different. Let see why. Consider the parameterization of a surface in the new space in the same form as in the Euclidean space,

$$S : \begin{cases} X = X(x^1, x^2) \\ Y = Y(x^1, x^2) \\ Z = Z(x^1, x^2). \end{cases}$$

  Then we have the same formulas for

$$\begin{cases} dX = \dfrac{\partial X}{\partial x^1} dx^1 + \dfrac{\partial X}{\partial x^2} dx^2 \\ dY = \dfrac{\partial Y}{\partial x^1} dx^1 + \dfrac{\partial Y}{\partial x^2} dx^2 \\ dZ = \dfrac{\partial Z}{\partial x^1} dx^1 + \dfrac{\partial Z}{\partial x^2} dx^2. \end{cases}$$

  The metric of the surface is obtained using now

$$ds^2 = dX^2 - dY^2 - dZ^2,$$

  therefore

$$ds^2 = \left[\left(\frac{\partial X}{\partial x^1}\right)^2 - \left(\frac{\partial Y}{\partial x^1}\right)^2 - \left(\frac{\partial Z}{\partial x^1}\right)^2\right](dx^1)^2 + \left[\frac{\partial X}{\partial x^1}\frac{\partial X}{\partial x^2} - \frac{\partial Y}{\partial x^1}\frac{\partial Y}{\partial x^2} - \frac{\partial Z}{\partial x^1}\frac{\partial Z}{\partial x^2}\right]dx^1dx^2 +$$

$$+\left[\frac{\partial X}{\partial x^2}\frac{\partial X}{\partial x^1} - \frac{\partial Y}{\partial x^2}\frac{\partial Y}{\partial x^1} - \frac{\partial Z}{\partial x^2}\frac{\partial Z}{\partial x^1}\right]dx^2dx^1 + \left[\left(\frac{\partial X}{\partial x^2}\right)^2 - \left(\frac{\partial Y}{\partial x^2}\right)^2 - \left(\frac{\partial Z}{\partial x^2}\right)^2\right](dx^2)^2 =$$

$$= \bar{g}_{ij}dx^i dx^j.$$

Having now the metric coefficients $\bar{g}_{ij}$, all the formulas involving the Christoffel symbols, Riemann symbols, Ricci symbols, Gaussian curvature, geodesics are the same as above. The intrinsic geometry of the surface is preserved.

- The same happens if the metric of the surrounding space is

$$ds^2 = dX^2 + dY^2 - dZ^2.$$

The metric coefficients are computed taking into account the following equalities

$$ds^2 = \left[\left(\frac{\partial X}{\partial x^1}\right)^2 + \left(\frac{\partial Y}{\partial x^1}\right)^2 - \left(\frac{\partial Z}{\partial x^1}\right)^2\right](dx^1)^2 + \left[\frac{\partial X}{\partial x^1}\frac{\partial X}{\partial x^2} + \frac{\partial Y}{\partial x^1}\frac{\partial Y}{\partial x^2} - \frac{\partial Z}{\partial x^1}\frac{\partial Z}{\partial x^2}\right]dx^1dx^2 +$$

$$+\left[\frac{\partial X}{\partial x^2}\frac{\partial X}{\partial x^1} + \frac{\partial Y}{\partial x^2}\frac{\partial Y}{\partial x^1} - \frac{\partial Z}{\partial x^2}\frac{\partial Z}{\partial x^1}\right]dx^2dx^1 + \left[\left(\frac{\partial X}{\partial x^2}\right)^2 + \left(\frac{\partial Y}{\partial x^2}\right)^2 - \left(\frac{\partial Z}{\partial x^2}\right)^2\right](dx^2)^2 =$$

$$= \bar{\bar{g}}_{ij}dx^i dx^j.$$

And again all information about the intrinsic geometry is related to the same symbols now computed using the metric coefficients $\bar{\bar{g}}_{ij}$.

Let us present now examples induced by the previous two metrics.

Example 4.14.1 The time-like sphere in a Minkowski $(+--)$ **space** In the case when the $3D$ space is endowed with the metric

$$ds^2 = dX^2 - dY^2 - dZ^2,$$

we look at the geometry of the surface

$$X^2 - Y^2 - Z^2 = -R^2.$$

According to the signs, the above metric is a Minkowski $(+--)$ metric and it is induced by the quadratic form $\langle u, u\rangle$ attached to the Minkowski product

$$\langle u, v\rangle = u_x v_x - u_y v_y - u_z v_z.$$

We have three types of vectors according to the above Minkowsky product.

- Timelike vectors, when $\langle u, u \rangle > 0$. An example is the vector $u = (2, 1, 1)$, or the vector $u = (2x^2 + 1, x^2, x^2),\ x \in \mathbb{R}$.
- Space-like vectors, when $\langle u, u \rangle < 0$. An example is the vector $u = (1, 1, 1)$, or the vector $u = (x^2, x^2, x^2 + 2),\ x \in \mathbb{R}$.
- Null vectors, also known as lightlike vectors, if $\langle u, u \rangle = 0$. An example is the vector $u = (2, -1, \sqrt{3})$, or the vector $u = (2x^2, \sqrt{2}x^2, \sqrt{2}x^2),\ x \in \mathbb{R}$. All the lightlike vectors can be thought as part of the cone having the equation

$$X^2 - Y^2 - Z^2 = 0.$$

The time-like vectors have the origin (the application point) and the end in the interior of the previous cone, while the space-like vectors have the end outside the previous cone.
In a Minkowski $3D$ space there are two types of spheres.

- Space-like spheres, if the normal are time-like vectors. The sphere having the equation

$$X^2 - Y^2 - Z^2 = R^2$$

is such a sphere. To imagine it from the Euclidean point of view, this Minkowski sphere is two sheets hyperboloid.

- Timelike spheres, where normals are space-like vectors. The sphere having the equation

$$X^2 - Y^2 - Z^2 = -R^2$$

is such a sphere. In our Euclidean intuition, it corresponds to a one-sheet Euclidean hyperboloid. Accordingly the geometry, this is a Minkowski sphere. Therefore, in this example, we are interested to understand the intrinsic geometry features of this time-like sphere.
We can parameterize the above time-like sphere as

$$\begin{cases} X = X(x^1, x^2) = R \sinh x^1 \\ Y = Y(x^1, x^2) = R \cosh x^1 \cos x^2 \\ Z = Z(x^1, x^2) = R \cosh x^1 \sin x^2. \end{cases}$$

It results

$$\begin{cases} dX = R \cosh x^1 dx^1 \\ dY = R \sinh x^1 \cos x^2 dx^1 - R \cosh x^1 \sin x^2 dx^2 \\ dZ = R \sinh x^1 \sin x^2 dx^1 + R \cosh x^1 \cos x^2 dx^2. \end{cases}$$

The metric of the Minkowski time-like sphere is obtained using now

$$ds^2 = dX^2 - dY^2 - dZ^2,$$

therefore it is

$$ds^2 = R^2\,(dx^1)^2 - R^2\cosh^2 x^1\,(dx^2)^2.$$

The non-zero Christoffel symbols are

$$\Gamma^2_{12} = \Gamma^2_{21} = \tanh x^1,\ \Gamma^1_{22} = \cosh x^1 \sinh x^1$$

and

$$R^1_{212} = \frac{\partial \Gamma^1_{22}}{\partial x^1} - \frac{\partial \Gamma^1_{21}}{\partial x^2} + \Gamma^1_{s1}\Gamma^s_{22} - \Gamma^1_{s2}\Gamma^s_{21} = \cosh^2 x^1.$$

It results $R_{1212} = g_{11}R^1_{212} = R^2\cosh^2 x^1$, that is the Gaussian curvature of the Minkowski time-like sphere is $K = -\dfrac{1}{R^2}$.
The equations of geodesics are

$$\begin{cases} \ddot{x}^1\,(t) + \cosh x^1 \sinh x^1\,(x\,(t)) \cdot \left(\dot{x}^2(t)\right)^2 = 0 \\ \ddot{x}^2\,(t) + 2\tanh\,(x\,(t)) \cdot \dot{x}^1\,(t) \cdot \dot{x}^2\,(t) = 0 \end{cases}$$

and solutions are for $x^1(t) = t,\ x^2(t) = \alpha$, where α is a constant. Therefore the geodesics are "great Minkowski circles" meridian type, i.e. Euclidean hyperbolas

$$c(t) = (R\sinh t,\ R\cosh t\cos\alpha,\ R\cosh t\sin\alpha).$$

Arbitrary planes, passing through the origin which intersect the time-like Minkowsky sphere, produce curves which are geodesics too, the proof is related to a change of coordinates. They are the equivalent of the great circles of an Euclidean sphere.

Example 4.14.2 The time-like sphere in a Minkowski $(++-)$ space In the case where the $3D$ space is endowed with the metric

$$ds^2 = dX^2 + dY^2 - dZ^2,$$

we look at the geometry of the surface

$$X^2 + Y^2 - Z^2 = -R^2.$$

The above metric is a Minkowski $(++-)$ metric and it is induced by the quadratic form $\langle u, u\rangle$ attached to the Minkowski product

$$\langle u, v \rangle = u_x v_x + u_y v_y - u_z v_z.$$

As in the above example, three kinds of vectors exist according to the new Minkowsky product.

- Timelike vectors, where $\langle u, u \rangle > 0$. An example is the vector $u = (2, 1, 1)$, or the vector $u = (x^2 + 1, x^2 + 1, x^2),\ x \in \mathbb{R}$.

- Space-like vectors, where $\langle u, u \rangle < 0$. An example is the vector $u = (1, 1, 3)$, or the vector $u = (x^2, x^2, 2x^2 + 2),\ x \in \mathbb{R}$.

- Null vectors, also known as lightlike vectors, where $\langle u, u \rangle = 0$. An example is the vector $u = (1, -1, \sqrt{2})$, or the vector $u = (x^2, x^2, \sqrt{2}x^2),\ x \in \mathbb{R}$. All the lightlike vectors can be thought as part of the cone having the equation

$$X^2 + Y^2 - Z^2 = 0.$$

The time-like vectors have the origin (application point) and the end in the interior of the previous cone, while the space-like vectors have the end outside the previous cone.
These properties are exactly those of the other Minkowski space presented above. In this Minkowski $3D$ space, there are also two types of spheres.

- Space-like spheres, where the normals are time-like vectors. The sphere having the equation

$$X^2 + Y^2 - Z^2 = R^2$$

is such a sphere. To imagine it, from the Euclidean point of view, this Minkowski sphere is a one-sheet hyperboloid.

- Timelike spheres, where the normals are space-like vectors. The sphere having the equation

$$X^2 + Y^2 - Z^2 = -R^2$$

is such a sphere. This time-like Minkowski sphere corresponds to two sheets hyperboloid. Therefore, in this example, we are interested to understand the geometry of this time-like sphere.
We can parametrize the above time-like sphere as

$$\begin{cases} X = X(x^1, x^2) = R \sinh x^1 \cos x^2 \\ Y = Y(x^1, x^2) = R \sinh x^1 \sin x^2 \\ Z = Z(x^1, x^2) = R \cosh x^1. \end{cases}$$

Then we have

$$\begin{cases} dX = R\cosh x^1 \cos x^2 dx^1 - R\sinh x^1 \sin x^2 dx^2 \\ dY = R\cosh x^1 \sin x^2 dx^1 + R\sinh x^1 \cos x^2 dx^2 \\ dZ = R\sinh x^1 dx^1. \end{cases}$$

The metric of the Minkowski time-like sphere is obtained using now

$$ds^2 = dX^2 + dY^2 - dZ^2,$$

therefore it is

$$ds^2 = R^2\,(dx^1)^2 + R^2 \sinh^2 x^1\,(dx^2)^2.$$

The non-zero Christoffel symbols are

$$\Gamma^2_{12} = \Gamma^2_{21} = \coth x^1,\ \Gamma^1_{22} = -\cosh x^1 \sinh x^1$$

and

$$R^1_{212} = \frac{\partial \Gamma^1_{22}}{\partial x^1} - \frac{\partial \Gamma^1_{21}}{\partial x^2} + \Gamma^1_{s1}\Gamma^s_{22} - \Gamma^1_{s2}\Gamma^s_{21} = -\sinh^2 x^1.$$

It results $R_{1212} = g_{11} R^1_{212} = -R^2 \sinh^2 x^1$, that is the Gaussian curvature of the Minkowski time-like sphere is $K = -\frac{1}{R^2}$.
In this case, the equations of geodesics are

$$\begin{cases} \ddot{x}^1\,(t) - \cosh x^1 \sinh x^1\,(x\,(t)) \cdot \left(\dot{x}^2(t)\right)^2 = 0 \\ \ddot{x}^2\,(t) + 2\coth\,(x\,(t)) \cdot \dot{x}^1\,(t) \cdot \dot{x}^2\,(t) = 0 \end{cases}$$

with solutions for $x^1(t) = t,\ x^2(t) = \alpha$, where α is a constant. Therefore the geodesics are again "great Minkowski circles" meridian type but corresponding to this sphere, i.e. Euclidean hyperbolas

$$c(t) = (R\sinh t \cos\alpha,\ R\sinh t \sin\alpha,\ R\cosh t)$$

contained on the surface of the two sheets Euclidean hyperboloid which represent the time-like Minkowski sphere.
Again, arbitrary planes passing through the origin, which intersect the time-like Minkowsky sphere, produce curves which are geodesics too, the proof being related to a change of coordinates.
These curves are the equivalent of the great circles of an Euclidean sphere, too.

4.15 Transferring Metrics. Is Our Geometric Intuition Intrinsically Related to the Reality?

Let us consider the $3D$ Euclidean ambient space endowed with the metric

$$ds^2 = dX^2 + dY^2 + dZ^2.$$

The metric of the plane $Z = 0$ is

$$ds^2 = dX^2 + dY^2$$

with null Gaussian curvature, $K = 0$.
To relate this result with the content of this section, let us take into account the translated unit sphere having the center at $(0, 0, 1)$, i.e. its equation being

$$X^2 + Y^2 + (Z - 1)^2 = 1.$$

The parameterization

$$\begin{cases} X = \sin u \cos v \\ Y = \sin u \sin v \\ Z = 1 + \cos u \end{cases}$$

leads to the metric

$$ds^2 = du^2 + \sin^2 u \, dv^2$$

with constant Gaussian curvature, $K = 1$. In the case of a sphere of radius R the metric is

$$ds^2 = R^2 du^2 + R^2 \sin^2 u \, dv^2,$$

with constant Gaussian curvature $K = \dfrac{1}{R^2}$.

Therefore our intuition associates the null curvature to planes and non-null constant positive curvatures to spheres.

We have the right to think at the "form of geometric objects" having the Gaussian curvatures as above. This intuition emerges because we see the plane and the sphere as parts of the Euclidean $3D$ space.

More formally is thinking that the forms of objects has to be understood with respect to their intrinsic geometry, given by a metric, which is not necessarily induced by the Euclidean $3D$ space.

What can we prove if a sphere is endowed with a null Gaussian curvature metric?

Transferring the null Gaussian curvature to a sphere We start with the same images, the plane $Z = 0$ and the unit sphere $X^2 + Y^2 + (Z - 1)^2 = 1$ parameterized in the same way

$$\begin{cases} X = \sin u \cos v \\ Y = \sin u \sin v \\ Z = 1 + \cos u. \end{cases}$$

Let us consider a line starting from the North Pole of the sphere, $N(0, 0, 2)$, which passes through a point of the sphere, $(\sin u \cos v, \sin u \sin v, \cos u)$, different from the North Pole.

Its equation is

$$\frac{X}{\sin u \cos v} = \frac{Y}{\sin u \sin v} = \frac{Z - 2}{\cos u - 1}.$$

This line from the North Pole, called *stereographic projection*, intersects the plane $Z = 0$ at

$$\left(\frac{2 \sin u \cos v}{1 - \cos u}, \frac{2 \sin u \sin v}{1 - \cos u}, 0\right).$$

Therefore the parameterization

$$\begin{cases} X = \dfrac{2 \sin u \cos v}{1 - \cos u} \\ Y = \dfrac{2 \sin u \sin v}{1 - \cos u} \end{cases}$$

with respect to the metric of the plane

$$ds^2 = dX^2 + dY^2$$

induces on the "set of points of the unit sphere" the metric

$$ds^2 = \frac{1}{\sin^4 \frac{u}{2}} \left(du^2 + \sin^2 u \, dv^2\right).$$

If the initial sphere has radius R, the final metric is

$$ds^2 = \frac{1}{\sin^4 \frac{u}{2}} \left(R^2 du^2 + R^2 \sin^2 u \, dv^2\right).$$

Both metrics have null constant Gaussian curvature because we are transferring a null constant Gaussian curvature on the points of a sphere.
Our imagine about spheres as balls of constant positive Gaussian curvature has to be generalized.

The geometry of a sphere depends only on the metric on it. In this case, the sphere has a metric transferred from an Euclidean plane.

This sphere, without the North Pole N, becomes an image of an Euclidean plane with its Euclidean geometry.

The geometry of the sphere has geodesics (i.e. lines) transfered from the lines of the plane $Z = 0$ (which are geodesics). How they look like?
The line d, included in $Z = 0$ together the North Pole, determines a plane. This plane intersects the sphere after a circle $\mathcal{C}$ passing through the missing point N, the North Pole. So, the point N is missing from the “line” $\mathcal{C}$.
We can observe that the tangent to $\mathcal{C}$ at the missing point N is parallel to d. $\mathcal{C}$ is a line of the sphere having exactly the same properties as an Euclidean line.
Therefore two Euclidean parallel lines d_1 and d_2 of $Z = 0$ determines two Euclidean parallel lines, $\mathcal{C}'$ and $\mathcal{C}''$ on the sphere. These two lines are tangent at the missing point N.
All figures drawn into the plane $Z = 0$ have an image on the sphere, therefore the Euclidean figures we know in the plane have now “another aspect” on the sphere, which is now another form of the Euclidean plane.

The old Euclidean intuition can be replaced. And at least for very intelligent fishes in a spherical aquarium, there is a chance to visualize the figures for solving Euclidean geometry problems.

In the same way, we can generate a positive constant Gaussian curvature of a plane.

Transferring the constant positive Gaussian curvature to a plane We start from

$$\begin{cases} X = \dfrac{2\sin u \cos v}{1 - \cos u} \\ Y = \dfrac{2\sin u \sin v}{1 - \cos u}. \end{cases}$$

It is easy to see that

$$\begin{cases} \tan \dfrac{u}{2} = \dfrac{2}{\sqrt{X^2 + Y^2}} \\ \tan v = \dfrac{Y}{X}, \end{cases}$$

that is $\sin^2 u = \dfrac{16(X^2 + Y^2)}{(X^2 + Y^2 + 4)^2}$ and

$$\begin{cases} du = \dfrac{-4}{\sqrt{X^2 + Y^2}(X^2 + Y^2 + 4)}(XdX + YdY) \\ dv = \dfrac{1}{X^2 + Y^2}(-YdX + XdY). \end{cases}$$

Using

$$ds^2 = du^2 + \sin^2 u \, dv^2$$

we obtain the metric of the plane in the form

$$ds^2 = \frac{16}{(X^2 + Y^2 + 4)^2}(dX^2 + dY^2).$$

Its Gaussian curvature is $K = 1$.

From the intrinsic geometric point of view the plane is no longer a plane. It is another image of a unit sphere.

The stereographic projection transfers the geodesics of the initial unit sphere into geodesics of the new "sphere". Among them, the reader can recognize lines passing through the South Pole (0, 0, 0) and a circle centred at the South Pole corresponding to the equator of the initial unit sphere.

Any two distinct geodesics will intersect in exact two distinct points. Lines passing through the South Pole have the second intersection point "at infinity", where the image of the North Pole N can be imagined. Furthermore, we have another possible description of the elliptic geometry.

The intrinsic geometric is then related only to coordinates and metric and it is not related to our corresponding intuition of geometric properties of structures of the 3D Euclidean space.

Next example is particularly important.

Transferring the constant positive metric of a unit sphere to a plane where all geodesics are straight lines Let us consider the unit sphere $X^2 + Y^2 + Z^2 = 1$ centred at the origin. As we know, its metric is

$$ds^2 = d\phi^2 + \sin^2 \phi d\theta^2.$$

We transfer this metric to the tangent plane at the North Pole $N(0, 0, 1)$, denoted here by T_N, in the following way. Let $(u, v, 1)$ be an arbitrary point of the tangent plane T_N. The straight line

$$\frac{X}{u} = \frac{Y}{v} = \frac{Z}{1}$$

which passes through the origin $O(0, 0, 0)$ intersects the unit sphere at

$$M\left(\frac{u}{\sqrt{1 + u^2 + v^2}}, \frac{v}{\sqrt{1 + u^2 + v^2}}, \frac{1}{\sqrt{1 + u^2 + v^2}}\right)$$

because $X = uZ,\ Y = vZ$ and $u^2Z^2 + v^2Z^2 + Z^2 = 1$.

Choosing

$$\begin{cases} u = \tan\phi\cos\theta \\ v = \tan\phi\sin\theta \end{cases}$$

and replacing in the coordinates of the point M, we obtain the usual parameterization of the sphere

$$(\sin\phi\cos\theta, \sin\phi\sin\theta, \cos\phi)$$

whose metric was written above, i.e.

$$ds^2 = d\phi^2 + \sin^2 \phi d\theta^2.$$

Having this in mind, we intend to find the metric of the tangent plane T_N in terms of u and v.

We have $1 + u^2 + v^2 = \dfrac{1}{\cos^2\phi}$ and

$$\begin{cases} du = \dfrac{1}{\cos^2\phi}\cos\theta d\phi - \tan\phi\sin\theta d\theta \\ dv = \dfrac{1}{\cos^2\phi}\sin\theta d\phi + \tan\phi\cos\theta d\theta, \end{cases}$$

therefore

$$\frac{1}{1+u^2+v^2}(du^2+dv^2) = \frac{\sin^2\phi + \cos^2\phi}{\cos^2\phi}d\phi^2 + \sin^2\phi d\theta^2 = \tan^2\phi d\phi^2 + ds^2.$$

Since $\phi = \arctan\sqrt{u^2+v^2}$, it results

$$d\phi = \frac{udu + vdv}{\sqrt{u^2+v^2}(1+u^2+v^2)},$$

which leads us to the $K = 1$ metric in T_N

$$ds^2 = \frac{1}{1+u^2+v^2}(du^2+dv^2) - \frac{(udu+vdv)^2}{(1+u^2+v^2)^2}.$$

The geodesics in T_N are straight lines obtained from the intersection of the planes of the sphere geodesics with T_N.

Of course, we can force to define parallel lines in T_N, and the axiom of parallelism is fulfilled.

Let us observe that each straight line is in fact doubled by the antipodal points corresponding to M. The order of points can be distorted.

This is a very good example of geometry satisfying the Euclidean parallelism axiom without to be an Euclidean geometry. At the same time, this is an example of elliptic geometry of the plane where the intersection points of geodesics are seen "at infinity".

Taking into account the results of this section, i.e. Differential Geometry is only related to sets of coordinates and metrics and it is not related to our current intuition, we are going to introduce a chapter dedicated to the basic Differential Geometry concepts and their applications.

Chapter 5
Basic Differential Geometry Concepts and Their Applications

> *The theory of gravitation will not find its way into my colleagues' head for longtime yet, no doubt. Only one, Levi-Civita in Padua, has probably grasped the main point completely, because he is familiar with the mathematics used.*
>
> *Albert Einstein*

Einstein's point of view written above may seem malicious. But almost all the mathematics we present in this chapter is related to the work by Tullio Levi-Civita. And this work is standing on Gregorio Ricci-Curbastro contribution on tensor calculus. Let us recall: the previous chapter allowed us to step into the intrinsic geometry of surfaces. In fact we can step into another picture of geometry: the one where only the system of coordinates and the metric are necessary. Therefore the surface is no longer necessary, the form of the "geometric universe" we study is no longer important. The symbols, i.e. the multi-index quantities we studied before, will allow us to advance in the geometric structure induced by the metric. We do not know the nature of these symbols yet, but we know that, except for the Christoffel ones, they are invariant with respect to the changes of coordinates. The same seems to happen with the equations of geodesics. The covariant derivative was used with respect to vectors. We can think of extending it to the multi-index quantities and obtain new interesting geometric laws using it. It is important to see that all formulas we obtained, using the additional dimension of the space, can be obtained in the new context when we have only the system of coordinates and the metric. In this chapter, we will realize that it does not matter that we work with two variables or more in the coordinate system. From the very beginning, we shall work with coordinate systems of n components. All these possible features were somehow suggested by the formulas of coordinate changes for the coefficients of the metric, Christoffel symbols, Riemann symbols, Ricci symbols, etc.

To step forward, let us say that this is the language we need for General Relativity.

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*, UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_5

The physical objects described through multi-index quantities have to preserve their nature when changes of coordinates are applied. The substance of the General Relativity is related to the invariance of changes of coordinates for geometric objects which describe physical objects and this is related to the deep meaning of the Equivalence Principle formulated by Einstein generalizing the Galilei one.

5.1 Tensors in Differential Geometry. Definition and Examples

Differential Geometry deals with a set M endowed with a coordinate system $(x^1, x^2, .., x^n),\ x^i \in \mathbb{R}$. In Physics, the system of coordinates has four components and starts from x^0 instead x^1 to make a distinction between the first coordinate which represents time and the other three coordinates representing spatial coordinates, often denoted by Greek letters, α, β and γ. We may keep this notation with Greek letters even if the number of coordinates is greater than 4 or we can use the standard mathematical notation $(x^0, x^1, .., x^{n-1}),\ x^i \in \mathbb{R}$.

The dimension of our set M is n without insisting here on other possible structures of M. The reader has to imagine M as an open set of $\mathbb{R}^n$ whose geometry depends on a metric defined on it.

Let us observe that there is no extra dimensions to study M. In geometric examples below, we choose to use the coordinate system starting with x^1, or we use directly the letters x, y, etc. because the physical meaning will be discussed in the other chapters below.

The fundamental object we start to study is the *tensor*.

In simple words, a tensor is a multi-index quantity which, under a change of coordinates, is transforming linearly with respect to the indexes. The components of a tensor depend smoothly on the points of the space, in our case M. So, tensors are functions having derivatives of all order everywhere in M.

Let us suppose we have a quantity $T^{i_1 i_2 \dots i_k}_{l_1 l_2 \dots l_p}(x)$ in a given system of coordinates $(x^1, x^2, ..., x^n)$ and let us consider a change of coordinates

$$x^i = x^i(\overline{x}) = x^i(\overline{x}^1, \overline{x}^2, ..., \overline{x}^n),\ i \in \{1, 2, ..., n\}.$$

In a simpler form, it is

$$x^i = x^i(\overline{x}^j),\ i, j \in \{1, 2, ..., n\}.$$

The inverse transformation of coordinates is $\overline{x}^i = \overline{x}^i(x^j),\ i, j \in \{1, 2, ..., n\}$. It is worth noticing that the Einstein notation $a_i b^i = \sum_{i=1}^n a_i b^i$ can be used in this form or in its multi-index form. With these notations in mind, we can denote by $\overline{T}^{j_1 j_2 \dots j_k}_{q_1 q_2 \dots q_p}(\overline{x})$, the quantity $T^{i_1 i_2 \dots i_k}_{l_1 l_2 \dots l_p}(x)$ written with respect to the new coordinates.

Definition 5.1.1 $T^{i_1 i_2 \ldots i_k}_{l_1 l_2 \ldots l_p}$ is a tensor contravariant of rank k and covariant of rank p, or simply a (k, p) tensor, if, under the previous change of coordinates $x^i = x^i(\overline{x}^j)$, the formula of $\overline{T}^{j_1 j_2 \ldots j_k}_{q_1 q_2 \ldots q_p}$ is

$$\overline{T}^{j_1 j_2 \ldots j_k}_{q_1 q_2 \ldots q_p}(\overline{x}) = T^{i_1 i_2 \ldots i_k}_{l_1 l_2 \ldots l_p}(x) \frac{\partial x^{l_1}}{\partial \overline{x}^{q_1}} \frac{\partial x^{l_2}}{\partial \overline{x}^{q_2}} \cdots \frac{\partial x^{l_p}}{\partial \overline{x}^{q_p}} \frac{\partial \overline{x}^{j_1}}{\partial x^{i_1}} \frac{\partial \overline{x}^{j_2}}{\partial x^{i_2}} \cdots \frac{\partial \overline{x}^{j_k}}{\partial x^{i_k}}.$$

Example 5.1.2 $a_k(x)$ is a covariant tensor of rank 1, or a covariant vector, if, under a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{a}_i(\overline{x})$ is defined as

$$\overline{a}_i(\overline{x}) = a_k(x) \frac{\partial x^k}{\partial \overline{x}^i};$$

Example 5.1.3 $a_{kl}(x)$ is a covariant tensor of rank 2, if after a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{a}_{ij}(\overline{x})$ is defined as

$$\overline{a}_{ij}(\overline{x}) = a_{kl}(x) \frac{\partial x^k}{\partial \overline{x}^i} \frac{\partial x^l}{\partial \overline{x}^j};$$

We can observe that the metric coefficients are the components of a covariant tensor of rank 2.

Example 5.1.4 $a_{pqr}(x)$ is a covariant tensor of rank 3, if after a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{a}_{ijk}(\overline{x})$ is defined as

$$\overline{a}_{ijk}(\bar{x}) = a_{pqr}(x) \frac{\partial x^p}{\partial \overline{x}^i} \frac{\partial x^q}{\partial \overline{x}^j} \frac{\partial x^r}{\partial \overline{x}^k};$$

Example 5.1.5 $T_{i_1 i_2 \ldots i_p}(x)$ is a covariant tensor of rank p, if after a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{T}_{j_1 j_2 \ldots j_p}(\overline{x})$ is defined as

$$\overline{T}_{j_1 j_2 \ldots j_p}(\overline{x}) = T_{i_1 i_2 \ldots i_p}(x) \frac{\partial x^{i_1}}{\partial \overline{x}^{j_1}} \frac{\partial x^{i_2}}{\partial \overline{x}^{j_2}} \cdots \frac{\partial x^{i_p}}{\partial \overline{x}^{j_p}}.$$

Let us observe that, in the above definition of (k, p) tensor, the contravariant indexes change using the inverse transformation of coordinates. The matrix of change of coordinates is called Jacobian matrix. The transformation is regular, if the determinant of the Jacobian matrix is always finite and different from zero. Otherwise, the transformation is a singular one. In general, if an object transforms under any change of coordinates with a non-singular Jacobian determinant, the object is a tensor. In this case the transformation is called linear. The rank of the tensor determines the number of Jacobian matrixes concurring in the transformation. For example, a rank 2 tensor transforms, under coordinate changes, by the multiplication of two Jacobian matrices, one for each index. The same happens with contravariant tensors.

Example 5.1.6 $a^k(x)$ is a contravariant tensor of rank 1, or simply a contravariant vector, if at a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{a}^i(\overline{x})$ is defined as

$$\overline{a}^i(\overline{x}) = a^k(x)\frac{\partial \overline{x}^i}{\partial x^k}.$$

It is shown below that the tangent vector to a curve is a contravariant vector.

Observe that we have used the inverse change of coordinates formula.

Example 5.1.7 $T^{i_1 i_2 \ldots i_k}(x)$ is a contravariant tensor of rank k, if under a change of coordinates $x^i = x^i(\overline{x}^j)$, $\overline{T}^{j_1 j_2 \ldots j_k}(\overline{x})$ is defined as

$$\overline{T}^{j_1 j_2 \ldots j_k}(\overline{x}) = T^{i_1 i_2 \ldots i_k}(x)\frac{\partial \overline{x}^{j_1}}{\partial x^{i_1}}\frac{\partial \overline{x}^{j_2}}{\partial x^{i_2}}\cdots\frac{\partial \overline{x}^{j_k}}{\partial \overline{x}^{i_k}}.$$

The geometry on M is assigned by the following objects:
- The metric $ds^2 = g_{ij}(x^1, x^2, ..., x^n)dx^i dx^j$ where the matrix

$$\mathbf{G} = (g_{ij}),\ g_{ij} = g_{ij}(x^1, x^2, ..., x^n),\ i,\ j \in \{1, 2, ..., n\},$$

is a rank-2 tensor called *metric tensor*. The name is related to the transforming formula proved in the previous chapter,

$$\bar{g}_{kl} = g_{ij}\frac{\partial x^i}{\partial \overline{x}^k}\frac{\partial x^j}{\partial \overline{x}^l}.$$

Before continuing, we will show that, apart from the direct calculation, there is another way to find the coefficients of the metric after a change of coordinates. This way will be used in some cases in the chapter devoted to General Relativity.

Proposition 5.1.8 *A change of coordinates* $x^r = x^r(\overline{x}^h),\ r,\ h \in \{0, 1, ..., n\}$ *transforms the metric under the rule* $\bar{G}_{\bar{x}} = (dM_{\bar{\mathbf{x}}})^t \cdot G_{\mathbf{x}} \cdot (dM_{\bar{\mathbf{x}}})$.

Proof Suppose that $M : \bar{U} \to U$ is the previous change of coordinates which transforms $(\bar{x}^0, ..., \bar{x}^n)$ into the coordinates $(x^0, ..., x^n)$. The first metric is described by the matrix $\bar{G} = (\bar{g}_{ij}(\bar{\mathbf{x}}))$ and the second metric is described by the matrix $G = (g_{rs}(\mathbf{x}))$.

We first suggest why the formula should be as it is in the statement before.

Consider a quadratic form $\sum_{i,j=1}^n \alpha_{ij} y^i y^j$ written in its matrix form

$$\mathbf{y}^t \cdot \boldsymbol{\alpha} \cdot \mathbf{y},$$

where $\mathbf{y}$ is a column vector such that its transposed is $\mathbf{y}^t = (y^1, ..., y^n)$ and the matrix α is $\boldsymbol{\alpha} = (\alpha_{ij}),\ i \in \{1, ..., n\},\ j \in \{1, ..., n\}$.

The change of coordinates $\mathbf{y} = \mathbf{A} \cdot \mathbf{x}$ leads to

$$\mathbf{y}^t \cdot \boldsymbol{\alpha} \cdot \mathbf{y} = (\mathbf{A} \cdot \mathbf{x})^t \cdot \boldsymbol{\alpha} \cdot (\mathbf{A} \cdot \mathbf{x}),$$

that is $(\mathbf{x}^t \cdot \mathbf{A}^t) \cdot \boldsymbol{\alpha} \cdot (\mathbf{A} \cdot \mathbf{x})$, i.e.

$$\mathbf{x}^t \cdot (\mathbf{A}^t \cdot \boldsymbol{\alpha} \cdot \mathbf{A}) \cdot \mathbf{x}.$$

So, the transformed quadratic form has its matrix $\mathbf{B}$ given by the formula $\mathbf{B} = \mathbf{A}^t \cdot \boldsymbol{\alpha} \cdot \mathbf{A}$.

In our case, using Einstein's rule, the second metric is

$$g_{rs}(\mathbf{x}) dx^r dx^s = \sum_{r,s=0}^{n} g_{rs}(\mathbf{x}) dx^r dx^s.$$

Since the change of coordinates in terms of differentials is $\mathbf{dx} = \left(\frac{\partial x^i}{\partial \bar{x}^j}\right) \mathbf{d\bar{x}}$, i.e. $dM_{\bar{\mathbf{x}}} = \left(\frac{\partial x^i}{\partial \bar{x}^j}\right)$, the previous formula $\mathbf{B} = \mathbf{A}^t \cdot \boldsymbol{\alpha} \cdot \mathbf{A}$ for $\mathbf{B} = \bar{G}_{\bar{x}},\ \alpha = G_{\mathbf{x}},\ \mathbf{A} = dM_{\bar{\mathbf{x}}}$ leads to

$$\bar{G}_{\bar{x}} = (dM_{\bar{\mathbf{x}}})^t \cdot G_{\mathbf{x}} \cdot (dM_{\bar{\mathbf{x}}}).$$

□

Obviously, we have a similar formula for the inverse of our initial coordinates transformation.

Corollary 5.1.9 *A change of coordinates* $\overline{x}^r = \overline{x}^r(x^h),\ \ r \in \{0, 1, ..., n\},\ \ h \in \{0, 1, ..., n\}$ *transforms the metric according to the rule* $G_x = (dM_{\bar{\mathbf{x}}(\mathbf{x})})^t \cdot \bar{G}_{\bar{\mathbf{x}}(\mathbf{x})} \cdot (dM_{\bar{\mathbf{x}}(\mathbf{x})})$.

Going back to the coefficients of a metric, we notice the following properties:

1. $g_{ij}(x)$ is a smooth function of x on M;
2. at each point x, the metric tensor is symmetric, i.e. $g_{ij}(x) = g_{ji}(x)$;
3. at each point x, it exists the inverse $\mathbf{G}^{-1} = (g^{ij})$; using the Einstein notation, the inverse is described by the relations: $g^{ij} g_{jk} = \delta^i_k\ ,\ g_{js} g^{sk} = \delta^k_j$.

Imagine the attached bilinear form with coefficients g_{ij}, denoted $S(u, v) = g_{ij} u^i v^j$.

This one can have the property $S(u, u) > 0,\ \forall u \neq 0,\ u = (u^1, ..., u^n)$.

In this case, the metric is called a *Riemannian metric*.

Otherwise, if there are vectors such that $S(u, u) > 0$ and other vectors such that $S(v, v) < 0$, it is called a *non-Riemannian metric*; some textbooks use the equivalent terminology: *semi-Riemannian* or *pseudo-Riemannian*.

The *signature of a metric* is defined as the signature of the corresponding quadratic form. For the Minkowski metric, the signature can be $(+ - - -)$ or $(- + + +)$. Both signatures can be used. It is important to mention that if we decide to use a signature, the sign which decides the time-like vectors has to be given by the first sign in the signature, i.e. if the signature is $(+ - --)$, the time-like vectors satisfy $S(u, u) > 0$. If the signature is $(- + ++)$ the time-like vectors satisfy $S(u, u) < 0$.

Some examples: the Euclidean metric $ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2$ is a Riemannian metric. Its signature is $(+ + +)$ while the Minkowski metric $ds^2 = (dx^1)^2 - (dx^2)^2 - (dx^3)^2$ is a non-Riemannian metric with $(+ - -)$ signature. In a previous example, we saw a Minkowski space endowed with the metric $ds^2 = (dx^1)^2 + (dx^2)^2 - (dx^3)^2$ whose signature is $(+ + -)$. In four dimensions such Minkowski metrics, i.e. with signatures $(+ - --)$ and $(+ + --)$, are used to describe de Sitter and Anti-de Sitter space-times, which are examples of universes without matter. In three dimensions, we already saw the time-like Minkowski spheres of these spaces. They can be used as models for Non-Euclidean geometries as we will see later in this book.

• At each point x, it exists a *tangent space* of M, denoted by $T_x M$, whose coordinates are $(\dot{x}^1, \dot{x}^2, ..., \dot{x}^n)$. Consider a curve $x(t)$ in M, and the vector

$$\dot{x}(t) = \frac{dx}{dt} = (\dot{x}^1(t), \dot{x}^2(t), ..., \dot{x}^n(t)).$$

This vector belongs to the tangent space and, under a change of coordinates $x^i = x^i(\bar{x}^j)$, we have

$$\dot{x}^i(t) = \frac{dx^i}{dt} = \frac{dx^i}{d\overline{x}^j}\frac{d\overline{x}^j}{dt} = \dot{\bar{x}}^j(t)\frac{dx^i}{d\overline{x}^j},$$

that is

$$\dot{\bar{x}}^j(t) = \dot{x}^i(t)\frac{d\overline{x}^j}{dx^i}.$$

Therefore, a tangent vector to a curve is a contravariant vector.

Considering vectors, we prefer to write $V = (V^1, V^2, ..., V^n)$ in the simpler form $V = V^k$, or only V^k as we did before. That is, a vector can be seen through its components.

• The Christoffel symbols of first kind are:

$$\Gamma_{ij,k} := \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right).$$

A change of coordinates transforms this formula into

$$\frac{\partial^2 x^k}{\partial\overline{x}^i\partial\overline{x}^j} = -\Gamma^k_{rs}\frac{\partial x^r}{\partial\overline{x}^i}\frac{\partial x^s}{\partial\overline{x}^j} + \overline{\Gamma}^r_{ij}\frac{\partial x^k}{\partial\overline{x}^r}.$$

The above formula shows that the Christoffel symbols of first kind are not tensors.

• The Christoffel symbols of second kind:

$$\Gamma^i_{jk} := g^{is}\Gamma_{jk,s} = \frac{1}{2}\,g^{is}\left(\frac{\partial g_{js}}{\partial x^k} + \frac{\partial g_{ks}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^s}\right).$$

are also not tensors as can be easily shown.
• The mixed Riemannian curvature tensor is:

$$R^h_{ijk} := \frac{\partial \Gamma^h_{ik}}{\partial x^j} - \frac{\partial \Gamma^h_{ij}}{\partial x^k} + \Gamma^h_{mj}\Gamma^m_{ik} - \Gamma^h_{mk}\Gamma^m_{ij}$$

The name is provided by the formula presented in the previous chapter when we discussed how the Riemann symbol of second kind is changing under the action of a transformation of coordinates,

$$\overline{R}^a_{bgd}\frac{\partial x^i}{\partial \overline{x}^a} = R^i_{jkl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d},$$

which can be written as

$$\overline{R}^a_{bgd} = R^i_{jkl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d}\frac{\partial \overline{x}^a}{\partial x^i}.$$

Therefore the mixed Riemman curvature tensor is a $(1, 3)$ type tensor.

The same for the next two symbols.

• The covariant Riemannian curvature tensor is: $R_{ijkl} := g_{is}R^s_{jkl}$. Its transformation formula is

$$\overline{R}_{ebgd} = R_{rjkl}\frac{\partial x^r}{\partial \overline{x}^e}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d},$$

providing a $(0, 4)$ type tensor.
• The Ricci tensor: $R_{ij} = R^s_{isj}$ which is obtained from the curvature tensor R^s_{imj} by contracting the indexes $s = m$. Its transformation formula is

$$\overline{R}_{bg} = R_{jl}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^l}{\partial \overline{x}^g},$$

providing a $(0, 2)$ type tensor.
• The geodesics, i.e. curves $c(t) = (x^1(t), x^2(t), ..., x^n(t))$ which satisfy the equations

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq}\frac{dx^p}{dt}\frac{dx^q}{dt} = 0,\ r \in \{1, 2, ..., n\}.$$

In the previous chapter, we proved that a change of coordinates $x^r = x^r(\overline{x}^h),\ r,\ h \in \{1, 2, ..., n\}$ transforms the metric coefficients under the rule

$$\overline{g}_{kl} = g_{ij}\frac{\partial x^i}{\partial \bar{x}^k}\frac{\partial x^j}{\partial \bar{x}^l}.$$

This is a $(0, 2)$ type tensor.

Therefore we may summarize: the Christoffel symbols are not tensors, while $g_{ij},\ R^i_{jkl},\ R_{ijkl},\ R_{ij}$ are tensors and their form is preserved under any change of coordinates. Below, we will prove that geodesics equations are tensors, too.

5.2 Properties of Riemann and Ricci Tensors in the New Geometric Context

In the case of surfaces, the Riemann symbols were obtained by considering the partial derivatives of Gauss formulas. In this way Gauss equations

$$R_{ijkl} = h_{ik} \cdot h_{jl} - h_{il} \cdot h_{jk}$$

are highlighted.

Gauss equations were used to prove the symmetric properties

$$\begin{aligned} R_{ijkl} &= -R_{ijlk}; \\ R_{ijkl} &= -R_{jikl}; \\ R_{ijkl} &= R_{klij}; \\ R_{ijkl} &+ R_{iklj} + R_{iljk} = 0. \end{aligned}$$

Can we prove such formulas in this abstract framework which does not contain an extra dimension? The next theorem presents a formula for the covariant Riemann tensor.

Theorem 5.2.1

$$R_{ijkl} = \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial x^j \partial x^k} + \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} - \frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} - \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l}\right) + g_{mp}(\Gamma^m_{il}\Gamma^p_{jk} - \Gamma^m_{ik}\Gamma^p_{jl}).$$

Proof The first step of the proof is related to the chain of equalities

$$R_{ijkl} = g_{is}R^s_{jkl} = g_{is}\left(\frac{\partial \Gamma^s_{jl}}{\partial x^k} - \frac{\partial \Gamma^s_{jk}}{\partial x^l} + \Gamma^s_{mk}\Gamma^m_{jl} - \Gamma^s_{ml}\Gamma^m_{jk}\right) =$$

$$= g_{is}\left(\frac{\partial \Gamma^s_{jl}}{\partial x^k} - \frac{\partial \Gamma^s_{jk}}{\partial x^l}\right) + g_{is}\left(\Gamma^s_{mk}\Gamma^m_{jl} - \Gamma^s_{ml}\Gamma^m_{jk}\right) =$$

$$= g_{is}\left(\frac{\partial g^{sa}}{\partial x^k}\Gamma_{jl,a} - \frac{\partial g^{sa}}{\partial x^l}\Gamma_{jk,a}\right) + \left(\frac{\partial \Gamma_{jl,i}}{\partial x^k} - \frac{\partial \Gamma_{jk,i}}{\partial x^l}\right) + g_{is}(\Gamma^s_{mk}\Gamma^m_{jl} - \Gamma^s_{ml}\Gamma^m_{jk}) =$$

$$= g_{is}\left(\frac{\partial g^{sa}}{\partial x^k}\Gamma_{jl,a} - \frac{\partial g^{sa}}{\partial x^l}\Gamma_{jk,a}\right) + \frac{1}{2}\left(\frac{\partial^2 g_{li}}{\partial x^k \partial x^j} + \frac{\partial^2 g_{jk}}{\partial x^l \partial x^i} - \frac{\partial^2 g_{jl}}{\partial x^k \partial x^i} - \frac{\partial^2 g_{ki}}{\partial x^l \partial x^j}\right) +$$

$$+ g_{is}(\Gamma^s_{mk}\Gamma^m_{jl} - \Gamma^s_{ml}\Gamma^m_{jk})$$

In the last equality, we replace

$$\frac{\partial g^{sa}}{\partial x^k} := -\Gamma^s_{kr} g^{ra} - \Gamma^a_{kr} g^{sr}$$

and

$$\frac{\partial g^{sa}}{\partial x^l} := -\Gamma^s_{lr} g^{ra} - \Gamma^a_{lr} g^{sr}.$$

It results

$$g_{is}\left(\frac{\partial g^{sa}}{\partial x^k}\Gamma_{jl,a} - \frac{\partial g^{sa}}{\partial x^l}\Gamma_{jk,a}\right) = g_{is}[\Gamma_{jl,a}(-\Gamma^s_{kr}g^{ra} - \Gamma^a_{kr}g^{sr}) + \Gamma_{jk,a}(\Gamma^s_{lr}g^{ra} + \Gamma^a_{lr}g^{sr})].$$

Continuing,

$$g_{is}\Gamma_{jl,a}(-\Gamma^s_{kr}g^{ra} - \Gamma^a_{kr}g^{sr}) = -g_{is}\Gamma^r_{jl}\Gamma^s_{kr} - \Gamma^a_{ki}\Gamma_{jl,a} =$$

$$= -\Gamma_{kr,i}\Gamma^r_{jl} - \Gamma^a_{ki}\Gamma_{jl,a} = -\Gamma^m_{ki}\Gamma_{jl,m} - \Gamma^m_{jl}\Gamma_{km,i}.$$

Analogously

$$g_{is}\Gamma_{jk,a}(\Gamma^s_{lr}g^{ra} + \Gamma^a_{lr}g^{sr}) = \Gamma^m_{li}\Gamma_{jk,m} + \Gamma^m_{jk}\Gamma_{lm,i},$$

i.e.

$$g_{is}\left(\frac{\partial g^{sa}}{\partial x^k}\Gamma_{jl,a} - \frac{\partial g^{sa}}{\partial x^l}\Gamma_{jk,a}\right) = -\Gamma^m_{ki}\Gamma_{jl,m} - \Gamma^m_{jl}\Gamma_{km,i} + \Gamma^m_{li}\Gamma_{jk,m} + \Gamma^m_{jk}\Gamma_{lm,i}.$$

The next step is to compute

$$g_{is}(\Gamma^s_{mk}\Gamma^m_{jl} - \Gamma^s_{ml}\Gamma^m_{jk}) - \Gamma^m_{ki}\Gamma_{jl,m} - \Gamma^m_{jl}\Gamma_{km,i} + \Gamma^m_{li}\Gamma_{jk,m} + \Gamma^m_{jk}\Gamma_{lm,i}$$

which means

$$\cancel{g_{is}g^{sa}\Gamma_{mk,a}\Gamma^m_{jl}} - \cancel{g_{is}g^{sa}\Gamma_{ml,a}\Gamma^m_{jk}} - \Gamma^m_{ki}\Gamma_{jl,m} - \cancel{\Gamma^m_{jl}\Gamma_{km,i}} + \Gamma^m_{li}\Gamma_{jk,m} + \cancel{\Gamma^m_{jk}\Gamma_{lm,i}},$$

that is

$$g_{mp}(\Gamma^m_{li}\Gamma^p_{jk} - \Gamma^m_{ki}\Gamma^p_{jl}).$$

The formula we obtain, after arranging the indexes, is

$$R_{ijkl} = \frac{1}{2}\left(\frac{\partial^2 g_{il}}{\partial x^j \partial x^k} + \frac{\partial^2 g_{jk}}{\partial x^i \partial x^l} - \frac{\partial^2 g_{jl}}{\partial x^i \partial x^k} - \frac{\partial^2 g_{ik}}{\partial x^j \partial x^l}\right) + g_{mp}\left(\Gamma^m_{il}\Gamma^p_{jk} - \Gamma^m_{ik}\Gamma^p_{jl}\right).$$

□

It is now an exercise for the readers to quickly derive the following formulas

$$R_{ijkl} = -R_{ijlk};$$
$$R_{ijkl} = -R_{jikl};$$
$$R_{ijkl} = R_{klij};$$
$$R_{ijkl} + R_{iklj} + R_{iljk} = 0.$$

in their intrinsic form.

The last identity, $R_{ijkl} + R_{iklj} + R_{iljk} = 0$ is known as Bianchi's first formula and will be used to obtain the symmetry of the Ricci tensor. Let us recall that, in the case of surfaces when there are only two variables x^1 and x^2, the symmetry of Ricci symbols was obtained using Einstein theorem which states $R_{ij} = K \cdot g_{ij}$.

Theorem 5.2.2 *The Ricci tensor is symmetric, that is $R_{ij} = R_{ji}$.*

Proof We multiply the Bianchi first identity $R_{ijkl} + R_{iklj} + R_{iljk} = 0$ by g^{jl}. It results

$$g^{lj} R_{jilk} + g^{jl} R_{iklj} - g^{lj} R_{jkli} = 0,$$

that is

$$R^l_{ilk} + g^{jl} R_{iklj} - R^l_{kli} = 0,$$

or simply

$$R_{ik} + g^{jl} R_{iklj} - R_{ki} = 0.$$

If we show that $g^{jl} R_{iklj} = 0$, we complete the proof. This is a consequence of

$$g^{jl} R_{iklj} = g^{lj} R_{ikjl} = -g^{lj} R_{ikjl}.$$

□

5.3 Covariant Derivative for Vectors. Geodesics and Their Properties

In the abstract context described in the first section of this chapter, let

$$x(t) = (x^1(t), x^2(t), ..., x^n(t))$$

be a curve in the space of coordinates $(x^1, x^2, ..., x^n)$. We can define the contravariant vector $V^k(t) := V^k(x(t))$. We call this $V^k(t)$ a contravariant vector along the curve $x(t)$.

Having in mind the covariant derivative in the case of surfaces, we consider the vector with the components

$$\frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt}.$$

Let us show that this is a contravariant vector (or a $(1, 0)$ tensor) attached to this curve, i.e. the following formula holds

$$\frac{dV^k}{dt} + \Gamma^k_{ij} V^j = \left(\frac{d\overline{V}^r}{dt} + \overline{\Gamma}^r_{pq} \overline{V}^q \right) \frac{\partial x^k}{\partial \overline{x}^r}.$$

We start from the contravariant vector V having the components $V^r(x)$. Being a contravariant vector, under a change of coordinates $\overline{x}^i = \overline{x}^i(x^j)$, $i, j \in \{1, 2, , ..., n\}$, its components change according to the rule

$$\overline{V}^r(\bar{x}) = V^j(x) \frac{\partial \overline{x}^r}{\partial x^j}.$$

The partial derivative with respect to x^i leads to

$$\frac{\partial \overline{V}^r}{\partial \overline{x}^p} \frac{\partial \overline{x}^p}{\partial x^i} = \frac{\partial V^j}{\partial x^i} \frac{\partial \overline{x}^r}{\partial x^j} + V^j \frac{\partial^2 \overline{x}^r}{\partial x^i \partial x^j}.$$

Multiplying by $\dfrac{\partial x^k}{\partial \overline{x}^r}$, it results

$$\frac{\partial \overline{V}^r}{\partial \overline{x}^p} \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial x^k}{\partial \overline{x}^r} = \frac{\partial V^j}{\partial x^i} \frac{\partial \overline{x}^r}{\partial x^j} \frac{\partial x^k}{\partial \overline{x}^r} + V^j \frac{\partial^2 \overline{x}^r}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^r} = \frac{\partial V^k}{\partial x^i} + V^j \frac{\partial^2 \overline{x}^r}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^r},$$

which can be written in the form

$$\frac{\partial V^k}{\partial x^i} = \frac{\partial \overline{V}^r}{\partial \overline{x}^p} \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial x^k}{\partial \overline{x}^r} - V^j \frac{\partial^2 \overline{x}^r}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^r}.$$

From the Christoffel symbols of second kind transformation formula

$$\Gamma^k_{ij} = \overline{\Gamma}^r_{pq} \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial \overline{x}^q}{\partial x^j} \frac{\partial x^k}{\partial \overline{x}^r} + \frac{\partial^2 \overline{x}^p}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^p},$$

we deduce

$$V^j \Gamma^k_{ij} = V^j \overline{\Gamma}^r_{pq} \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial \overline{x}^q}{\partial x^j} \frac{\partial x^k}{\partial \overline{x}^r} + V^j \frac{\partial^2 \overline{x}^p}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^p} = \overline{V}^q \overline{\Gamma}^r_{pq} \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial x^k}{\partial \overline{x}^r} + V^j \frac{\partial^2 \overline{x}^p}{\partial x^i \partial x^j} \frac{\partial x^k}{\partial \overline{x}^p}.$$

If we add the two relations obtained above, it results

$$\frac{\partial V^k}{\partial x^i} + V^j \Gamma^k_{ij} = \left(\frac{\partial \overline{V}^r}{\partial \overline{x}^p} + \overline{V}^q \overline{\Gamma}^r_{pq} \right) \frac{\partial \overline{x}^p}{\partial x^i} \frac{\partial x^k}{\partial \overline{x}^r}.$$

Considering the coordinates x and $\bar{x}$ as functions of t, we can write the last equality in the form

$$\frac{\partial V^k}{\partial x^i}\frac{dx^i}{dt} + V^j\Gamma^k_{ij}\frac{dx^i}{dt} = \left(\frac{\partial \overline{V}^r}{\partial \overline{x}^p} + \overline{V}^q\overline{\Gamma}^r_{pq}\right)\frac{\partial \overline{x}^p}{\partial x^i}\frac{\partial x^k}{\partial \overline{x}^r}\frac{dx^i}{dt},$$

or equivalently

$$\frac{dV^k}{dt} + \Gamma^k_{ij}V^j = \left(\frac{d\overline{V}^r}{dt} + \overline{\Gamma}^r_{pq}\overline{V}^q\right)\frac{\partial x^k}{\partial \overline{x}^r}.$$

□

Now, let us understand the implications of the previous computations.

The formula

$$\frac{\partial V^k}{\partial x^i} + V^j\Gamma^k_{ij} = \left(\frac{\partial \overline{V}^r}{\partial \overline{x}^p} + \overline{V}^q\overline{\Gamma}^r_{pq}\right)\frac{\partial \overline{x}^p}{\partial x^i}\frac{\partial x^k}{\partial \overline{x}^r}$$

shows that the expression $\dfrac{\partial V^k}{\partial x^i} + V^j\Gamma^k_{ij}$ is a $(1, 1)$ tensor type, while the formula

$$\frac{dV^k}{dt} + \Gamma^k_{ij}V^j = \left(\frac{d\overline{V}^r}{dt} + \overline{\Gamma}^r_{pq}\overline{V}^q\right)\frac{\partial x^k}{\partial \overline{x}^r}$$

highlights a $(1, 0)$ tensor type. We can consider the following two definitions suggested by the two tensor types seen above.

Definition 5.3.1 For the contravariant vector $V^k(x)$, the covariant derivative is the $(1, 1)$ tensor with the components

$$\frac{\partial V^k}{\partial x^i} + V^j\Gamma^k_{ij}.$$

We denote the covariant derivative of the contravariant vector $V(x) = V^k(x)$ as

$$V^k_{;i} := \frac{\partial V^k}{\partial x^i} + V^j\Gamma^k_{ij}$$

Other possible notations for $V^k_{;i}$ are $\dfrac{\nabla V^k}{\partial x^i}$ or $\dfrac{\nabla V}{\partial x^i}$.

Definition 5.3.2 For the contravariant vector $V^k(t)$, the covariant derivative is the contravariant vector

$$\frac{dV^k}{dt} + \Gamma^k_{ij}V^j\frac{dx^i}{dt}.$$

We denote this in the form

$$V^k_; := \frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt}$$

Other possible notations for $V^k_;$ are $\frac{\nabla V^k}{dt}$ or $\frac{\nabla V}{dt}$.

Let us specifically write the connection:

$$\frac{\nabla V^k}{dt} = V^k_; = V^k_{;i} \frac{dx^i}{dt} = \frac{\nabla V^k}{\partial x^i} \frac{dx^i}{dt}$$

which comes from

$$\frac{\nabla V^k}{dt} = \frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt} = \left(\frac{\partial V^k}{\partial x^i} + V^j \Gamma^k_{ij} \right) \frac{dx^i}{dt} = \frac{\nabla V^k}{\partial x^i} \frac{dx^i}{dt}.$$

We can follow the approach used in the case of surfaces.

Definition 5.3.3 The contravariant vector $V^k(t)$ is parallel transported along the curve $x(t)$ if

$$\frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt} = 0, \ k \in \{1, 2, ..., n\}.$$

Definition 5.3.4 The curve $x = x(t) = (x^1(t), x^2(t), ..., x^n(t))$ is a geodesic if its contravariant tangent vector $\dot{x}(t)$ is parallel transported along the curve.

Proposition 5.3.5 *A curve* $x(t) = (x^1(t), x^2(t), ..., x^n(t))$ *whose components satisfy the equations*

$$\frac{d^2x^k}{dt^2} + \Gamma^k_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0, \ k \in \{1, 2, ..., n\}$$

is a geodesic of M.

Proof We replace V^k by $\frac{dx^k}{dt}$ in the formula

$$\frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt} = 0,$$

and we obtain

$$\frac{d^2x^k}{dt^2} + \Gamma^k_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} = 0.$$

□

The following statement shows that a change of coordinates transforms a geodesic into a geodesic.

Theorem 5.3.6 *The change of coordinates* $x^r = x^r(\overline{x}^h)$, $r \in \{1, 2, ..., n\}, h \in \{1, 2, ..., n\}$ *transforms the equations*

$$\frac{d^2\overline{x}^h}{dt^2} + \overline{\Gamma}^h_{ij}\frac{d\overline{x}^i}{dt}\frac{d\overline{x}^j}{dt} = 0$$

for the curve $\overline{c}(t) = (\overline{x}^1(t), \overline{x}^2(t), ..., \overline{x}^n(t))$ *into the equations*

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq}\frac{dx^p}{dt}\frac{dx^q}{dt} = 0$$

for the curve $c(t) = (x^1(t), x^2(t), ..., x^n(t))$.

Proof We use the result of the previous proposition for the contravariant vector $\dot{c}$:

$$\frac{d^2\overline{x}^h}{dt^2} + \overline{\Gamma}^h_{ij}\frac{d\overline{x}^i}{dt}\frac{d\overline{x}^j}{dt} = \left(\frac{d^2x^r}{dt^2} + \Gamma^r_{pq}\frac{dx^p}{dt}\frac{dx^q}{dt}\right)\frac{\partial\overline{x}^h}{\partial x^r}.$$

If

$$\frac{d^2\overline{x}^h}{dt^2} + \overline{\Gamma}^h_{ij}\frac{d\overline{x}^i}{dt}\frac{d\overline{x}^j}{dt} = 0$$

then

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq}\frac{dx^p}{dt}\frac{dx^q}{dt} = 0,$$

which ends the proof. □

Another important fact about geodesics is

Theorem 5.3.7 *If* $c(t)$ *is a geodesic, then* $||\dot{c}(t)||$ *is a constant.*

Proof Recall first that the length of a vector in a given metric $ds^2 = gijdx^idx^j$ obviously depends on the type of the vector. Therefore, in our case, the formula is

$$||\dot{c}(t)||^2 := \mp\, g_{ij}(x^0(t), x^1(t), ..., x^n(t))\frac{dx^i(t)}{dt}\frac{dx^j(t)}{dt}.$$

We continue using the + sign, in the other case the computations are the same. Having in mind that

$$\frac{d^2x^i}{dt^2} = -\Gamma^i_{lm}\frac{dx^l}{dt}\frac{dx^m}{dt}$$

we start to compute $\frac{d}{dt}\left(||\dot{c}(t)||^2\right)$. We obtain

$$\frac{d}{dt}\left(||\dot{c}(t)||^2\right) = \frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt} + g_{ij}\frac{d^2x^i}{dt^2}\frac{dx^j}{dt} + g_{ij}\frac{dx^i}{dt}\frac{d^2x^j}{dt^2} =$$

$$= \frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt} + 2g_{ij}\frac{d^2x^i}{dt^2}\frac{dx^j}{dt} =$$

$$= \frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt} - 2g_{ij}\Gamma^i_{lm}\frac{dx^l}{dt}\frac{dx^m}{dt}\frac{dx^j}{dt} = \frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt} - 2\Gamma_{lm,j}\frac{dx^l}{dt}\frac{dx^m}{dt}\frac{dx^j}{dt} =$$

$$= \frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt} - \left(\frac{\partial g_{lj}}{\partial x^m} + \frac{\partial g_{mj}}{\partial x^l} - \frac{\partial g_{lm}}{\partial x^l}\right)\frac{dx^l}{dt}\frac{dx^m}{dt}\frac{dx^j}{dt} = 0.$$

This expression contains four terms. After we relabel the summation indexes in the last three terms, we have in fact the term $\frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{dt}\frac{dx^i}{dt}\frac{dx^j}{dt}$ written four times, two with the sign plus and two with minus. It results $\frac{d}{dt}\left(||\dot{c}(t)||^2\right) = 0$, i.e. $||\dot{c}(t)||^2 = b$, where b is a constant. □

The concept of covariant derivative of tensors allows us to obtain the same result in the next section. The fact that, along a geodesic, the length of the tangent vector at the geodesic is a constant one, allows us to replace the parameter t with $s = \frac{t}{\sqrt{|b|}}$. The geodesic $c = c(s)$ has the property $||\dot{c}(s)|| = 1$.

Definition 5.3.8 A curve which fulfils such a property, i.e. $||\dot{c}(s)|| = 1$, is called a canonically parameterized curve.

These last two theorems will be used later to understand how geodesics of the disk are transformed by inversion in geodesic of the Poincaré half-plane. And more, how geodesics of the disk are transformed by inversion in geodesics outside the disk. All these facts will allow us to better understand the connections among Non-Euclidean Geometry basic models.

5.4 Covariant Derivative of Tensors and Applications

The previous definitions of covariant derivative for contravariant vectors allow us to think on how a covariant derivative for (k, p) type tensors looks like.

Definition 5.4.1 The covariant derivative of a (k, p) tensor $T^{i_1i_2..i_k}_{l_1l_2..l_p}(x)$ is the $(k, p+1)$ tensor defined by

$$T^{i_1i_2..i_k}_{l_1l_2..l_p;j}(x) =$$

$$= \frac{T^{i_1i_2..i_k}_{l_1l_2..l_p}}{\partial x^j} + T^{mi_2..i_k}_{l_1l_2..l_p}\Gamma^{i_1}_{mj} + \cdots + T^{i_1i_2..i_{k-1}m}_{l_1l_2...l_p}\Gamma^{i_k}_{mj} - T^{i_1i_2...i_k}_{ml_2..l_p}\Gamma^m_{l_1j} - \cdots - T^{i_1i_2..i_k}_{l_1l_2..l_{p-1}m}\Gamma^m_{l_pj}.$$

If we look at some particular cases we better understand the way in which the previous formula can be derived.

- The covariant derivative of a covariant vector is the (0, 2) covariant tensor

$$V_{i;j} = \frac{\partial V_i}{\partial x^j} - V_k \Gamma^k_{ij}.$$

- The covariant derivative of a covariant (0, 2) tensor is the (0, 3) covariant tensor

$$a_{ij;k} = \frac{\partial a_{ij}}{\partial x^k} - a_{sj}\Gamma^s_{ik} - a_{si}\Gamma^s_{jk}.$$

- The covariant derivative of a contravariant (2, 0) tensor is the (2, 1) tensor

$$b^{ij}_{;k} = \frac{\partial b^{ij}}{\partial x^k} + b^{lj}\Gamma^i_{lk} + b^{il}\Gamma^j_{kl}.$$

- The covariant derivative of a (1, 1) tensor is the (1, 2) tensor $a^j_{i;k} = \dfrac{\partial a^j_i}{\partial x^k} + a^m_i \Gamma^j_{mk} - a^j_m \Gamma^m_{ik}$.
- For two tensors T^I_L and S^J_P, where I, L, J, P are multi-index quantities, it makes sense to define the tensor $T^I_L S^J_P$ and the product rule of covariant derivative, that is

$$(T^I_L S^J_P)_{;m} = (T^I_L)_{;m} S^J_P + T^I_L (S^J_P)_{;m}.$$

For the metric tensor g_{ij} we can prove.

Theorem 5.4.2 $g_{ij;k} = 0$

Proof We have

$$\begin{aligned} g_{ij;k} = \frac{\partial g_{ij}}{\partial x^k} - g_{sj}\Gamma^s_{ik} - g_{si}\Gamma^s_{jk} &= \frac{\partial g_{ij}}{\partial x^k} - g_{sj} g^{sl}\Gamma_{ik,l} - g_{si} g^{sl}\Gamma_{jk,l} \\ &= \frac{\partial g_{ij}}{\partial x^k} - \Gamma_{ik,j} - \Gamma_{jk,i} = 0 \end{aligned}$$

□

A very important consequence appears.

Consider two contravariant vectors V^k and W^j and their "dot product" via the metric tensor, $\langle V^k, W^j \rangle = g_{kj} V^k W^j$.

Suppose that V^k and W^j are parallel transported along a curve $x(t)$, that is

$$\frac{dV^k}{dt} + \Gamma^k_{ij} V^j \frac{dx^i}{dt} = 0,$$

and

$$\frac{dW^j}{dt} + \Gamma^j_{il} W^l \frac{dx^i}{dt} = 0,$$

or simply, $\dfrac{\nabla V^k}{dt} = V^k_{;}(t) = 0$ and $\dfrac{\nabla W^j}{dt} = W^j_{;}(t) = 0$. The covariant derivative of the metric tensor g_{ij} vanishes,

$$g_{ij;k} = \frac{\partial g_{ij}}{\partial x^k} - g_{sj}\Gamma^s_{ik} - g_{si}\Gamma^s_{jk} = 0$$

and this can be written as

$$\frac{\nabla g_{ij}}{dt} = g_{ij;}(t) = \left(\frac{\partial g_{ij}}{\partial x^k} - g_{sj}\Gamma^s_{ik} - g_{si}\Gamma^s_{jk}\right)\frac{dx^k}{dt} = 0.$$

The derivative with respect to t of the "dot product" is, in fact, the covariant derivative of $g_{kj}V^kW^j$. Applying the product rule for the covariant derivative, we obtain

$$\left(g_{kj}V^kW^j\right)_{;i} = g_{kj;i}V^kW^j + g_{kj}V^k_{;i}W^j + g_{kj}V^kW^j_{;i}.$$

Therefore

$$\left(g_{kj}V^kW^j\right)_{;i}\frac{dx^i}{dt} = g_{kj;i}\frac{dx^i}{dt}V^kW^j + g_{kj}V^k_{;i}\frac{dx^i}{dt}W^j + g_{kj}V^kW^j_{;i}\frac{dx^i}{dt},$$

which can be written in the form

$$\frac{\nabla\left(g_{kj}V^kW^j\right)}{dt} = \frac{\nabla g_{ij}}{dt}V^kW^j + g_{kj}\frac{\nabla V^k}{dt}W^j + g_{kj}V^k\frac{\nabla W^j}{dt}.$$

All covariant derivatives in the right member of the equality are 0. Therefore the derivative of the "dot product" is 0. It results:

Theorem 5.4.3 *(i) The length of a vector is conserved if the vector is parallel transported along a curve.*
(ii) The length of the tangent vector to a geodesic is a constant.

In each geometric space where the meaning of

$$\frac{\langle V^k, W^j\rangle}{|V^k|\,|W^j|} = \frac{g_{kj}V^kW^j}{\sqrt{|g_{kj}V^kV^j|}\sqrt{|g_{kj}W^kW^j|}}$$

is related to an angle via a trigonometric function $f(\alpha)$ the following theorem holds:

Theorem 5.4.4 *(i) The angle between two parallel transported vectors along a curve is conserved.*
(ii) The angle between a parallel transported vector along a geodesic and the tangent vector to a geodesic is the same at each point of the geodesic.

For the inverse g^{ij} of the metric tensor g_{ij} we have

Exercise 5.4.5 $g^{ij}_{;k} = 0$.

Hint: Consider the covariant derivative of the expression $g_{is}g^{sj} = \delta^j_i$. It results

$$\delta^j_{i;k} = \frac{\partial \delta^j_i}{\partial x^k} + \delta^m_i \Gamma^j_{mk} - \delta^j_m \Gamma^m_{ik} = 0$$

and

$$0 = (g_{is}g^{sj})_{;k} = g_{is;k}g^{sj} + g_{is}g^{sj}_{;k} = g_{is}g^{sj}_{;k},$$

i.e.

$$g^{mi}g_{is}g^{sj}_{;k} = g^{mj}_{;k} = 0.$$

5.5 A Step Towards General Relativity: The Bianchi Second Formula

Next Lemma is necessary if we intend to offer a shorter proof to Bianchi's second formula.

Lemma 5.5.1 *The following equality holds:*

$$\frac{\partial R^s_{ijk}}{\partial x^l} + \frac{\partial R^s_{ikl}}{\partial x^j} + \frac{\partial R^s_{ilj}}{\partial x^k} + R^a_{ijk}\Gamma^s_{al} + R^a_{ikl}\Gamma^s_{aj} + R^a_{ilj}\Gamma^s_{ak}$$
$$= R^s_{mjk}\Gamma^m_{il} + R^s_{mkl}\Gamma^m_{ij} + R^s_{mlj}\Gamma^m_{ik}.$$

Proof The proof is simple. We have to compute

$$\frac{\partial R^s_{ijk}}{\partial x^l} + \frac{\partial R^s_{ikl}}{\partial x^j} + \frac{\partial R^s_{ilj}}{\partial x^k}$$

and to add

$$R^a_{ijk}\Gamma^s_{al} + R^a_{ikl}\Gamma^s_{aj} + R^a_{ilj}\Gamma^s_{ak}.$$

Arranging in an adequate form, we obtain the desired result stated in the Lemma.

We start from $\dfrac{\partial R^s_{ijk}}{\partial x^l} + \dfrac{\partial R^s_{ikl}}{\partial x^j} + \dfrac{\partial R^s_{ilj}}{\partial x^k}$ which means

$$\frac{\partial}{\partial x^l}\left(\frac{\partial \Gamma^s_{ik}}{\partial x^j} - \frac{\partial \Gamma^s_{ij}}{\partial x^k} + \Gamma^s_{aj}\Gamma^a_{ik} - \Gamma^s_{ak}\Gamma^a_{ij}\right) + \frac{\partial}{\partial x^j}\left(\frac{\partial \Gamma^s_{il}}{\partial x^k} - \frac{\partial \Gamma^s_{ik}}{\partial x^l} + \Gamma^s_{ak}\Gamma^a_{il} - \Gamma^s_{al}\Gamma^a_{ik}\right) + \\ + \frac{\partial}{\partial x^k}\left(\frac{\partial \Gamma^s_{ij}}{\partial x^l} - \frac{\partial \Gamma^s_{il}}{\partial x^j} + \Gamma^s_{al}\Gamma^a_{ij} - \Gamma^s_{aj}\Gamma^a_{il}\right),$$

that is

$$\frac{\partial}{\partial x^l}\left(\Gamma^s_{aj}\Gamma^a_{ik} - \Gamma^s_{ak}\Gamma^a_{ij}\right) + \frac{\partial}{\partial x^j}\left(\Gamma^s_{ak}\Gamma^a_{il} - \Gamma^s_{al}\Gamma^a_{ik}\right) + \frac{\partial}{\partial x^k}\left(\Gamma^s_{al}\Gamma^a_{ij} - \Gamma^s_{aj}\Gamma^a_{il}\right).$$

If we continue in computing and add the missing part

$$R^a_{ijk}\Gamma^s_{al} + R^a_{ikl}\Gamma^s_{aj} + R^a_{ilj}\Gamma^s_{ak}$$

we obtain

$$\frac{\partial \Gamma^s_{aj}}{\partial x^l}\Gamma^a_{ik} + \frac{\partial \Gamma^a_{ik}}{\partial x^l}\Gamma^s_{aj} - \frac{\partial \Gamma^s_{ak}}{\partial x^l}\Gamma^a_{ij} - \frac{\partial \Gamma^a_{ij}}{\partial x^l}\Gamma^s_{ak} + \frac{\partial \Gamma^s_{ak}}{\partial x^j}\Gamma^a_{ik} + \frac{\partial \Gamma^a_{il}}{\partial x^j}\Gamma^s_{ak} - \frac{\partial \Gamma^s_{al}}{\partial x^j}\Gamma^a_{ij} - \frac{\partial \Gamma^a_{ik}}{\partial x^j}\Gamma^s_{al} + \\ + \frac{\partial \Gamma^s_{al}}{\partial x^k}\Gamma^a_{ij} + \frac{\partial \Gamma^a_{ij}}{\partial x^k}\Gamma^s_{al} - \frac{\partial \Gamma^s_{aj}}{\partial x^k}\Gamma^a_{il} - \frac{\partial \Gamma^a_{il}}{\partial x^k}\Gamma^s_{aj} + R^a_{ijk}\Gamma^s_{al} + R^a_{ikl}\Gamma^s_{aj} + R^a_{ilj}\Gamma^s_{ak}$$

which can be successively arranged as

$$\Gamma^s_{al}\left(R^a_{ijk} - \frac{\partial \Gamma^a_{ik}}{\partial x^j} + \frac{\partial \Gamma^a_{ij}}{\partial x^k}\right) + \Gamma^s_{aj}\left(R^a_{ikl} - \frac{\partial \Gamma^a_{il}}{\partial x^k} + \frac{\partial \Gamma^a_{ik}}{\partial x^l}\right) + \Gamma^s_{ak}\left(R^a_{ilj} - \frac{\partial \Gamma^a_{ij}}{\partial x^l} + \frac{\partial \Gamma^a_{il}}{\partial x^j}\right) + \\ + \Gamma^a_{ik}\left(\frac{\partial \Gamma^s_{aj}}{\partial x^l} - \frac{\partial \Gamma^s_{al}}{\partial x^j}\right) + \Gamma^a_{ij}\left(\frac{\partial \Gamma^s_{al}}{\partial x^k} - \frac{\partial \Gamma^s_{ak}}{\partial x^l}\right) + \Gamma^a_{il}\left(\frac{\partial \Gamma^s_{ak}}{\partial x^j} - \frac{\partial \Gamma^s_{aj}}{\partial x^k}\right) = \\ = \Gamma^s_{al}\left(\Gamma^a_{mj}\Gamma^m_{ik} - \Gamma^a_{mk}\Gamma^m_{ij}\right) + \Gamma^s_{aj}\left(\Gamma^a_{mk}\Gamma^m_{il} - \Gamma^a_{ml}\Gamma^m_{ik}\right) + \Gamma^s_{ak}\left(\Gamma^a_{ml}\Gamma^m_{ij} - \Gamma^a_{mj}\Gamma^m_{il}\right) + \\ + \Gamma^m_{ik}\left(\frac{\partial \Gamma^s_{mj}}{\partial x^l} - \frac{\partial \Gamma^s_{ml}}{\partial x^j}\right) + \Gamma^m_{ij}\left(\frac{\partial \Gamma^s_{ml}}{\partial x^k} - \frac{\partial \Gamma^s_{mk}}{\partial x^l}\right) + \Gamma^m_{il}\left(\frac{\partial \Gamma^s_{mk}}{\partial x^j} - \frac{\partial \Gamma^s_{mj}}{\partial x^k}\right) = \\ \Gamma^m_{ik}\left(\frac{\partial \Gamma^s_{mj}}{\partial x^l} - \frac{\partial \Gamma^s_{ml}}{\partial x^j} + \Gamma^s_{al}\Gamma^a_{mj} - \Gamma^s_{aj}\Gamma^a_{ml}\right) + \Gamma^m_{ij}\left(\frac{\partial \Gamma^s_{ml}}{\partial x^k} - \frac{\partial \Gamma^s_{mk}}{\partial x^l} + \Gamma^s_{ak}\Gamma^a_{ml} - \Gamma^s_{al}\Gamma^a_{mk}\right) \\ + \Gamma^m_{il}\left(\frac{\partial \Gamma^s_{mk}}{\partial x^j} - \frac{\partial \Gamma^s_{mj}}{\partial x^k} + \Gamma^s_{aj}\Gamma^a_{mk} - \Gamma^s_{ak}\Gamma^a_{mj}\right) = R^s_{mlj}\Gamma^m_{ik} + R^s_{mkl}\Gamma^m_{ij} + R^s_{mjk}\Gamma^m_{il}.$$

□

Theorem 5.5.2 (Bianchi's second formula)

$$R^s_{ijk;l} + R^s_{ikl;j} + R^s_{ilj;k} = 0.$$

Proof Consider the covariant derivative of the $(1, 3)$ type Riemann tensor R^s_{ijk}. According to the definition formula we have:

$$R^s_{ijk;l} = \frac{\partial R^s_{ijk}}{\partial x^l} + R^a_{ijk}\Gamma^s_{al} - R^s_{mjk}\Gamma^m_{il} - R^s_{imk}\Gamma^m_{jl} - R^s_{ijm}\Gamma^m_{kl}.$$

In the same way, we obtain

$$R^s_{ikl;j} = \frac{\partial R^s_{ikl}}{\partial x^j} + R^a_{ikl}\Gamma^s_{aj} - R^s_{mkl}\Gamma^m_{ij} - R^s_{iml}\Gamma^m_{jk} - R^s_{ikm}\Gamma^m_{lj}$$

and

$$R^s_{ilj;k} = \frac{\partial R^s_{ilj}}{\partial x^k} + R^a_{ilj}\Gamma^s_{ak} - R^s_{mlj}\Gamma^m_{ik} - R^s_{imj}\Gamma^m_{lk} - R^s_{ilm}\Gamma^m_{kj}.$$

We add the three equalities and use the obvious equality $R^s_{ijk} = -R^s_{ikj}$.

It results

$$\begin{aligned} &R^s_{ijk;l} + R^s_{ikl;j} + R^s_{ilj;k} = \\ &= \frac{\partial R^s_{ijk}}{\partial x^l} + \frac{\partial R^s_{ikl}}{\partial x^j} + \frac{\partial R^s_{ilj}}{\partial x^k} + R^a_{ijk}\Gamma^s_{al} + R^a_{ikl}\Gamma^s_{aj} + R^a_{ilj}\Gamma^s_{ak} \\ &- R^s_{mjk}\Gamma^m_{il} - R^s_{mkl}\Gamma^m_{ij} - R^s_{mlj}\Gamma^m_{ik}. \end{aligned}$$

The previous Lemma asserts that the right member is 0 and this ends the proof. □

Chapter 6
Differential Geometry at Work: Two Ways of Thinking the Gravity. The Einstein Field Equations from a Geometric Point of View

Lectures which really teach will never be popular;
lectures which are popular will never really teach.

Michael Faraday

The gravitational field in General Relativity is described by the Einstein field equations. They look like

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} = k T_{ij}$$

and they act in a four-dimensional coordinate space (x^0, x^1, x^2, x^3).

We first observe that tensors involved in the left-hand side of the equality come from our previous Differential Geometry considerations. In the case of surfaces, i.e. $n = 2$, the left-hand side gives the so-called Einstein theorem,

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} = 0.$$

In the right-hand side, k is a constant and T_{ij} is the so-called stress–energy tensor which satisfies the condition

$$T_{ij;a} = 0\,,$$

which is a conservation law.

But if we can somehow "prove" that $T_{ij;a} = 0$, we have to prove that

$$\left(R_{ij} - \frac{1}{2} R \cdot g_{ij} \right)_{;a} = 0.$$

The null covariant derivative for both tensors is followed by their proportionality through the constant k.

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*, UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_6

However, since these tensors have to be well chosen, an analogy can be derived from the Newtonian Mechanics. The next chapter will be devoted to the Newtonian Mechanics and all the details of the short story we are going to tell here can be seen there. The same about the General Relativity aspects presented here: they will be discussed later in this book. Now we want to highlight how we can move from the Foundations of Geometry, through Differential Geometry, to General Relativity. This fact is almost amazing.

6.1 From Newtonian Gravity to the Geometry of Space-Time

General relativity is a theory describing gravity. Gravity was first described as a force in the framework of Newtonian Mechanics, so our short story passes through the definition offered by Isaac Newton. If the masses M_1 and m are concentrated at the points (x_1, y_1, z_1), (x, y, z), respectively, the *gravitational force* induced by the body of mass M_1 which acts on the body of mass m, $M_1 > m$ has the intensity

$$F = G\frac{mM_1}{r_1^2},$$

where

$$r_1 := \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$$

is the distance between the two bodies and $G = 6.67 \cdot 10^{-11}$ is the *gravitational constant* measured in the well-known SI units for length, mass, and time, i.e. $(m)^3 \cdot (kg)^{-1}(s)^{-2}$.

The gravitational force vector has the form

$$\vec{F} = -\frac{GmM_1}{r_1^2}\frac{\vec{r}_1}{r_1} = -\frac{GmM_1}{r_1^2} \cdot \left(\frac{x - x_1}{r_1}, \frac{y - y_1}{r_1}, \frac{z - z_1}{r_1}\right).$$

If we denote the unit vector in the brackets by $-\vec{u}$, the previous formula becomes

$$\vec{F} = m\frac{GM_1}{r_1^2}\vec{u}\ .$$

Later, in this book, the reader will find a chapter devoted to basic Newtonian Mechanics principles and consequences. According to the Second Principle of Newtonian Mechanics, the above formula can be written in the form

$$\vec{F} = m\vec{A}_1,$$

where

$$\vec{A}_1 = \frac{GM_1}{r_1^2}\,\vec{u}$$

is the *gravitational acceleration* or the *gravitational field* induced by the body of mass M_1.

The last formula can be seen in coordinates as

$$\vec{A}_1\,(x, y, z) = -\frac{GM_1}{r_1^2} \cdot \left(\frac{x - x_1}{r_1}, \frac{y - y_1}{r_1}, \frac{z - z_1}{r_1}\right).$$

Let us define the *gravitational potential* of the field $\vec{A}_1$ to be the function

$$\Phi_1(x, y, z) = -\frac{GM_1}{r_1}.$$

This definition makes sense at all points of the Euclidean three-dimensional space except (x_1, y_1, z_1) where the gravitational source is located.

The following mathematical facts can be proven in the case of a Universe in which only a mass exists:

$$1.\ \nabla\Phi_1(x, y, z) = \frac{GM_1}{r_1^2} \cdot \left(\frac{x - x_1}{r_1}, \frac{y - y_1}{r_1}, \frac{z - z_1}{r_1}\right) = -\vec{A}_1\,(x, y, z),$$

where $\nabla := \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$ is the *gradient operator* of simply, the gradient.

$$2.\ \nabla^2\Phi_1(x, y, z) = \frac{\partial^2\Phi_1}{\partial x^2} + \frac{\partial^2\Phi_1}{\partial y^2} + \frac{\partial^2\Phi_1}{\partial z^2} = 0,$$

where

$$\nabla^2 := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

is the *Laplace operator* or simply, the Laplacian.

If we have more masses, denoted M_k, located at the points (x_k, y_k, z_k), $k \in \{1, 2, ..., N\}$, we can define the gravitational field of these masses as

$$\vec{A}\,(x, y, z) := \sum_{k=1}^{N} \vec{A}_k\,(x, y, z) = -\sum_{k=1}^{N} \frac{GM_k}{r_k^2} \cdot \left(\frac{x - x_k}{r_k}, \frac{y - y_k}{r_k}, \frac{z - z_k}{r_k}\right)$$

and the gravitational potential of these masses as

$$\Phi(x, y, z) := \sum_{k=1}^{N} \Phi_k(x, y, z) = -\sum_{k=1}^{N} \frac{GM_k}{r_k}.$$

Of course, $r_k := \sqrt{(x - x_k)^2 + (y - y_k)^2 + (z - z_k)^2}$ and we kept (x, y, z) as the coordinate of the point in which it is located the mass m, now attracted by all the masses considered. Therefore, if the Universe contains only the masses $M_k,\ k \in \{1, 2, ..., N\}$, it is governed by the previous acceleration field $\vec{A}$ and the considered gravitational potential Φ. This Universe obeys the rules:

$$1'.\ \nabla\Phi = -\vec{A},$$

$$2'.\ \nabla^2\Phi = 0.$$

If the masses are in a medium described by its density function $\rho = \rho(x, y, z)$, $(x, y, z) \neq (x_k, y_k, z_k),\ k \in \{1, 2, ..., N\}$, the last formula becomes

$$3'.\ \nabla^2\Phi = 4\pi G\rho,$$

and these are not simple. A complete proof can be seen in [34, 46].

Let us consider again a Universe "with only one mass M_1. This time we are not interested to add more masses to the system, therefore this mass attached to the point (x_1, y_1, z_1) will be denoted by M. All the formulas written in the case of the index 1 are now without index, i.e.

$$\vec{F} = m\vec{A},$$

where

$$\vec{A} = \frac{GM}{r^2}\vec{u};$$

$$r = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2},$$

and for the gravitational potential is

$$\Phi(x, y, z) = -\frac{GM}{r}.$$

Since

$$\vec{A} = -\nabla\Phi,$$

the trajectories of free-falling particles, everyone of mass 1, are described by the equations

$$\frac{d^2x^k}{dt^2}(t) = -\frac{\partial \Phi}{\partial x^k}(\bar{x}(t)) \ , \ k \in \{1, 2, 3\}.$$

If the starting point of the system is the curve $c(q) = (x(q), y(q), z(q))$, the equations before are in fact

$$\frac{d^2x^k}{dt^2}(t, q) = -\frac{\partial \Phi}{\partial x^k}(\bar{x}(t, q)) \ , \ k \in \{1, 2, 3\},$$

where

$$\bar{x}(t, q) = (x^1(t, q), x^2(t, q), x^3(t, q))$$

is a point of the trajectory of the system of particles. We see the difference between the two parameters. The parameter t is an evolution parameter, we can name it "time". The second one, q, indicates only the initial point of the curve $c(q)$ from which the trajectory starts. If we consider a nearby initial point, say $c(q + \Delta q)$, the previous equations become

$$\frac{d^2x^k}{dt^2}(t, q + \Delta q) = -\frac{\partial \Phi}{\partial x^k}(\bar{x}(t, q + \Delta q)), \quad k \in \{1, 2, 3\}.$$

Now,

$$\lim_{\Delta q \to 0} \frac{\dfrac{d^2x^k}{dt^2}(t, q + \Delta q) - \dfrac{d^2x^k}{dt^2}(t, q)}{\Delta q} = -\lim_{\Delta q \to 0} \frac{\dfrac{\partial \Phi}{\partial x^k}(\bar{x}(t, q + \Delta q)) - \dfrac{\partial \Phi}{\partial x^k}(\bar{x}(t, q))}{\Delta q},$$

that is,

$$\frac{d^2}{dt^2}\frac{\partial x^k}{\partial q}(t, q) = -\frac{\partial^2 \Phi}{\partial q \partial x_k}(\bar{x}(t, q)) = -\sum_{i=1}^{3} \frac{\partial^2 \Phi}{\partial x_i \partial x_k}\frac{\partial x^i}{\partial q}(\bar{x}(t, q))$$

for each $k \in \{1, 2, 3\}$ and the last equality is obtained from the chain rule.

We highlighted the tidal vector

$$\frac{\partial \bar{x}}{\partial q} = \left(\frac{\partial \bar{x}^1}{\partial q}, \frac{\partial \bar{x}^2}{\partial q}, \frac{\partial \bar{x}^3}{\partial q}\right)$$

which satisfies the so-called tidal acceleration equations

$$\frac{d^2}{dt^2}\frac{\partial \bar{x}}{\partial q} = -d^2\Phi_{\bar{x}}\frac{\partial \bar{x}}{\partial q},$$

where the Hessian matrix $d^2\Phi_{\bar{x}} = \left(\frac{\partial^2\Phi(\bar{x})}{\partial x_i \partial x_k}\right)_{i,k}$ encapsulates in its trace the vacuum field equation

$$\nabla^2\Phi = 0.$$

The tidal vector measures the variation of nearby trajectories and the above result is a very geometric one, because it is related to the trajectories of a free-falling system of particles in a gravitational field generated by a mass.

The reader has to notice: all above results are obtained in the Euclidean space, therefore the Christoffel symbols Γ^i_{jk} are all null. Why are we telling this fact?

6.2 The Einstein Field Equations and the Energy–Momentum Tensor

Let us think how the equations of the tidal acceleration looks like in a four-dimensional space (x^0, x^1, x^2, x^3) endowed with the metric $ds^2 = g_{ij}dx^i dx^j$.

In such a space, the Christoffel symbols Γ^i_{jk} are not necessarily null. In the same way as in the Euclidean case, we consider that each coordinate depends on two parameters denoted (τ, q), therefore each point of the space-time has the form $(x^0(\tau, q), x^1(\tau, q), x^2(\tau, q), x^3(\tau, q))$.

Differential geometry highlighted the curves determined by the null covariant derivative condition

$$\frac{\nabla}{d\tau}\frac{dx^k}{d\tau} = \frac{d^2x^k}{d\tau^2} + \Gamma^k_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau} = 0,$$

i.e. the geodesics of the space-time. Therefore the equations described by

$$\frac{d^2x^k}{dt^2}(t, q) = -\frac{\partial\Phi}{\partial x^k}(\bar{x}(t, q)) \ , \ k \in \{1, 2, 3\}$$

in the Euclidean space are now replaced by the geodesic equations

$$\frac{d^2x^k}{d\tau^2} + \Gamma^k_{ij}\frac{dx^i}{d\tau}\frac{dx^j}{d\tau} = 0, \ k \in \{0, 1, 2, 3\}.$$

We can consider that each geodesic starts from the point of a curve

$$a(0, q) = (x^0(0, q), x^1(0, q), x^2(0, q), x^3(0, q))$$

in the direction

$$V(0, q) = \left(\frac{dx^0}{d\tau}(0, q), \frac{dx^1}{d\tau}(0, q), \frac{dx^2}{d\tau}(0, q), \frac{dx^3}{d\tau}(0, q)\right).$$

As in the previous case, the vector

$$\frac{\partial x}{\partial q} := \left(\frac{\partial x^0}{\partial q}, \frac{\partial x^1}{\partial q}, \frac{\partial x^2}{\partial q}, \frac{\partial x^3}{\partial q}\right)$$

is called a geometric tidal vector and measures the *rate of separation of the geodesics*.

We start from the covariant derivative $\dfrac{\nabla}{d\tau}$ of the tidal vector $\dfrac{\partial x^h}{\partial q}$:

$$\frac{\nabla}{d\tau}\frac{\partial x^h}{\partial q} = \frac{d}{d\tau}\left(\frac{\partial x^h}{\partial q}\right) + \Gamma^h_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}.$$

We apply again the covariant derivative formula,

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = \frac{\nabla}{d\tau}\left(\frac{\nabla}{d\tau}\frac{\partial x^h}{\partial q}\right) = \frac{d}{d\tau}\left(\frac{\nabla}{d\tau}\frac{\partial x^h}{\partial q}\right) + \Gamma^h_{mk}\left(\frac{\nabla}{d\tau}\frac{\partial x^m}{\partial q}\right)\frac{dx^k}{d\tau} =$$

$$= \frac{d}{d\tau}\left[\frac{d}{d\tau}\left(\frac{\partial x^h}{\partial q}\right) + \Gamma^h_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\right] + \Gamma^h_{mk}\left[\frac{d}{d\tau}\left(\frac{\partial x^m}{\partial q}\right) + \Gamma^m_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\right]\frac{dx^k}{d\tau} =$$

$$= \frac{\partial}{\partial q}\frac{d^2x^h}{d\tau^2} + \frac{\partial\Gamma^h_{ij}}{\partial x^k}\frac{dx^k}{d\tau}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q} +$$

$$+ \Gamma^h_{ij}\frac{d^2x^i}{d\tau^2}\frac{\partial x^j}{\partial q} + \Gamma^h_{ij}\frac{dx^i}{d\tau}\frac{\partial^2 x^j}{\partial\tau\partial q} + \Gamma^h_{mk}\frac{\partial^2 x^m}{\partial\tau\partial q}\frac{dx^k}{d\tau} + \Gamma^h_{mk}\Gamma^m_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau}.$$

Now, we replace $\dfrac{d^2x^k}{d\tau^2}$ by $-\Gamma^k_{ij}\dfrac{dx^i}{d\tau}\dfrac{dx^j}{d\tau}$ and, in some terms, we replace the dummy indexes in a convenient way. It results

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = \frac{\partial}{\partial q}\left(-\Gamma^h_{ik}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}\right) + \frac{\partial\Gamma^h_{ij}}{\partial x^k}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau} + \Gamma^h_{ij}\left(-\Gamma^m_{ik}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}\right)\frac{\partial x^j}{\partial q} +$$

$$+ \Gamma^h_{ij}\frac{dx^i}{d\tau}\frac{\partial^2 x^j}{\partial\tau\partial q} + \Gamma^h_{mk}\frac{\partial^2 x^m}{\partial\tau\partial q}\frac{dx^k}{d\tau} + \Gamma^h_{mk}\Gamma^m_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau} =$$

$$= -\frac{\partial\Gamma^h_{ik}}{\partial x^j}\frac{\partial x^j}{\partial q}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau} \cancel{- 2\Gamma^h_{ik}\frac{\partial^2 x^i}{\partial\tau\partial q}\frac{dx^k}{d\tau}} + \frac{\partial\Gamma^h_{ij}}{\partial x^k}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau} - \Gamma^h_{ij}\Gamma^m_{ik}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}\frac{\partial x^j}{\partial q} +$$

$$\cancel{+ 2\Gamma^h_{ij}\frac{dx^i}{d\tau}\frac{\partial^2 x^j}{\partial\tau\partial q}} + \Gamma^h_{mk}\Gamma^m_{ij}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau},$$

that is

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = \left(\frac{\partial \Gamma^h_{ij}}{\partial x^k} - \frac{\partial \Gamma^h_{ik}}{\partial x^j} + \Gamma^h_{mk}\Gamma^m_{ij} - \Gamma^h_{mj}\Gamma^m_{ik}\right)\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau}.$$

Therefore, we have obtained the tidal acceleration equations

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = R^h_{ikj}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau} = -R^h_{ijk}\frac{dx^i}{d\tau}\frac{\partial x^j}{\partial q}\frac{dx^k}{d\tau}.$$

The last line can be written in the more suggestive form

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -K^h_j\frac{\partial x^j}{\partial q},$$

where

$$K^h_j = R^h_{ijk}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}.$$

If we have tides, mathematically they can be treated in the Newtonian way. In fact, the field equation

$$\nabla^2 \Phi = 0$$

is hidden in the trace of the Hessian matrix $d^2\Phi$ involved in the tidal equations

$$\frac{d^2}{dt^2}\frac{\partial \bar{x}}{\partial q} = -d^2\Phi_{\bar{x}}\frac{\partial \bar{x}}{\partial q}.$$

Geometric tides can be treated in the new geometric way expressed above. The null trace of the Hessian matrix is now

$$K^h_h = R^h_{ihk}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau} = 0,$$

that is, the field equations are expressed via the Ricci tensor in the form

$$R_{ik} = 0.$$

The major step made by Einstein was replacing the three-dimensional flat Euclidean space with a four-dimensional space-time endowed with a metric

$$ds^2 = g_{ij}dx^i dx^j.$$

He was the first one who realized that the laws of Nature have to be expressed by equations involving tensors, which are covariant at all changes of coordinates.

For Einstein, the metric coefficients g_{ij} play the role of gravitational potential Φ. The Christoffel symbols Γ^i_{jk} play the role of gravitational field $\vec{A}$. In fact the changes he did can be summarized in the following table containing the two ways of thinking at the gravity

$$
\begin{array}{ccc}
Newton & & Einstein \\
\Phi & \longleftrightarrow & g_{ij} \\
\vec{A} & \longleftrightarrow & \Gamma^i_{jk} \\
\nabla^2\Phi = 0 & \longleftrightarrow & R_{ij} = 0 \\
\nabla^2\Phi = 4\pi G\rho & \longleftrightarrow & R_{ij} - \frac{1}{2}R\, g_{ij} = k\, T_{ij}.
\end{array}
$$

The change of paradigm is related to the equations of tidal acceleration, the geometric tidal acceleration, and the conclusion drawn above.

In order to achieve the last equality, we need to construct the correct framework. We follow the line sketched in [34].

Let us start from a symmetric contravariant tensor (T^{ij}), expressed as a 4×4 matrix and consider a four-dimensional space of coordinates (x^0, x^1, x^2, x^3) endowed with a metric

$$ds^2 = g_{ij}dx^i dx^j,\ i, j \in \{0, 1, 2, 3, \}$$

such that the parallel transport depends on the Christoffel symbols Γ^i_{jk} not all zero. The tensor (T^{ij}) looks like

$$(T^{ij}) = \begin{pmatrix} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{10} & T^{11} & T^{12} & T^{13} \\ T^{20} & T^{21} & T^{22} & T^{23} \\ T^{30} & T^{31} & T^{32} & T^{33} \end{pmatrix}$$

Because of symmetry, the first line $T^{1i} = (T^{10}, T^{11}, T^{12}, T^{13})$ coincides with the first column $T^{i1} = (T^{01}, T^{11}, T^{21}, T^{31})$, and can be seen as the representation of a contravariant 4-vector denoted by T^1, the same for the other rows and corresponding columns.

Before discussing about the physical aspects involved in the components of the energy–momentum tensor, let us assert that we can analyse the tensor like we analyse a vector F used to represent the flow of an incompressible fluid.

We can suppose the existence of a flow associated to the above tensor. For each line of the given tensor we have the corresponding force T^k described above. We can also think about a 4-parallelepiped centred at a given point (x^0, x^1, x^2, x^3) with sides parallel to the axes of coordinates and having the small dimensions Δx_0, Δx_1, Δx_2, Δx_3.

Small enough to let us suppose the vector T^k having the same components at each point of considered face.

The total outflow determined by each T^k through the parallel faces, corresponding to x^i direction, is analysed in the following way.

If $i = 1$, we have the differences

$$T^{k1}\left(x_0, x_1 + \frac{\Delta x_1}{2}, x_2, x_3\right)\Delta x_0 \Delta x_2 \Delta x_3 - T^{k1}\left(x_0, x_1 - \frac{\Delta x_1}{2}, x_2, x_3\right)\Delta x_0 \Delta x_2 \Delta x_3,$$

where $k \in \{0, 1, 2, 3\}$.

We have to consider the differences with respect to the parallel transport of the given vectors $\left(0, T^{k1}\left(x_0, x_1 - \frac{\Delta x_1}{2}, x_2, x_3\right), 0, 0\right)$, $k \in \{0, 1, 2, 3\}$ along the vector $(0, \Delta x_1, 0, 0)$, therefore we have

$$\left[T^{k1}\left(x_0, x_1 + \frac{\Delta x_1}{2}, x_2, x_3\right) - T^{k1}\left(x_0, x_1 - \frac{\Delta x_1}{2}, x_2, x_3\right) + \Gamma^1_{k1} T^{k1}\Delta x_1\right]\Delta x_0 \Delta x_2 \Delta x_3$$

with the approximation

$$\left[\frac{\partial T^{k1}}{\partial x^1}\Delta x_1 + \Gamma^1_{k1} T^{k1}\Delta x_1\right]\Delta x_0 \Delta x_2 \Delta x_3,\ k \in \{0, 1, 2, 3\},$$

that is,

$$T^{k1}_{;1}\Delta x_0 \Delta x_1 \Delta x_2 \Delta x_3, \quad k \in \{0, 1, 2, 3\}.$$

It results in a total outflow for the 4-parallelepiped equal to

$$T^{kl}_{;l}\Delta x_0 \Delta x_1 \Delta x_2 \Delta x_3.$$

What is happening in the interior of this small 4-parallelepiped? The quantity of matter and energy which enters in the interior of the given 4-parallelepiped leaves completely the interior. Therefore, at each moment of time, the quantity of matter inside can be considered a constant. The energy–momentum tensor is conserved, and since it is conserved as $\Delta x^k \to 0$, the limiting net flow approaches to 0, that is, $T^{kl}_{;l} = 0$.

Let us underline the fact that this is a sort of axiom. We suppose that the matter does not disappear and does not appear from nowhere inside the parallelepiped, and we are talking here about the matter and energy seen as an incompressible fluid described by the tensor. This "axiom" is in fact the key of the fact that these equations never match in a quantum mechanics context.

We know how to lower indexes using the metric tensor: $T^l_i = g_{ik}T^{kl}$. Since the covariant derivative of the metric tensor is 0, it is

$$T^l_{i;l} = g_{ik;l}T^{kl} + g_{ik}T^{kl}_{;l} = 0\,.$$

Lowering again the indexes, it is $T_{ji} = g_{jl}T^l_i$. In the same way

$$T_{ji;l} = g_{jl;l}T^l_i + g_{jl}T^l_{i;l} = 0.$$

We have proved an important property of the energy–momentum tensor, i.e.

$$T_{ij;l} = 0\,,$$

which is a conservation law.

Therefore, to obtain the previous equality, we have to prove that the so-called Einstein tensor $E_{ij} := R_{ij} - \frac{1}{2}R \cdot g_{ij}$ has the property

$$E_{ij;a} = 0.$$

This is the aim of the next theorem. Before proving it, the following comment is necessary. $E_{ij;a} = 0$ holds also if dimensions are different from four.

Theorem 6.2.1 (Covariant derivative of Einstein tensor) *If g_{ij} is the* $(0, 2)$ *metric tensor, g^{ij} its inverse contravariant* $(2, 0)$ *tensor, R_{ij} is the Ricci tensor, $R := R^s_s$ is the Ricci curvature scalar derived from the* $(1, 1)$ *mixed tensor $R^i_j = g^{is}R_{sj}$, it is possible to define the Einstein tensor*

$$E_{ij} := R_{ij} - \frac{1}{2}R \cdot g_{ij}\ .$$

Then

$$E_{ij;a} = 0.$$

Proof First of all, let us observe that R can be written as $g^{ij}R_{ij}$. Indeed,

$$g^{ij}R_{ij} = g^{ij}R_{ji} = R^i_i = R.$$

We have to prove that the covariant derivative of the Einstein tensor is null, i.e.

$$\left(R_{ij} - \frac{1}{2}R \cdot g_{ij}\right)_{;a} = 0.$$

We start from the Bianchi identity

$$R^s_{ijk;l} + R^s_{ikl;j} + R^s_{ilj;k} = 0\,,$$

and contract the indexes $j = s$. It results

$$R^s_{isk;l} + R^s_{ikl;s} + R^s_{ils;k} = 0$$

and we use $R_{ik;l} = R^s_{isk;l}$, $R^s_{ils;k} = -R^s_{isl;k} = -R_{il;k}$. We obtain

$$R_{ik;l} + R^s_{ikl;s} - R_{il;k} = 0.$$

Using the fact that the covariant derivative of g^{ij} is null, we can write

$$(g^{ia}R_{ik})_{;l} + (g^{ia}R^s_{ikl})_{;s} - (g^{ia}R_{il})_{;k} = 0,$$

i.e.

$$R^a_{k;l} + (g^{ia}R^s_{ikl})_{;s} - R^a_{l;k} = 0.$$

We contract $a = l$ and we have

$$R^a_{k;a} + (g^{ia}R^s_{ika})_{;s} - R^a_{a;k} = 0.$$

Now, $g^{ia}R^s_{ika} = g^{ia}g^{sb}R_{bika} = g^{ia}g^{sb}R_{ibak} = g^{sb}g^{ia}R_{ibak} = g^{sb}R^a_{bak} = R^s_k$ and we replace in the previous equality.

It results $R^a_{k;a} + R^s_{k;s} - R_{;k} = 0$, that is $2R^a_{k;a} - R_{;k} = 0$, which can be written in the form

$$\left(R^a_k - \frac{1}{2}\delta^a_k R\right)_{;a} = 0.$$

Because the covariant derivative of the metric tensor is null, then we have

$$\left(g_{ma}R^a_k - \frac{1}{2}\delta^a_k g_{ma} R\right)_{;a} = 0,$$

that is,

$$\left(R_{mk} - \frac{1}{2}R g_{mk}\right)_{;a} = 0.$$

□

A comment is necessary now. Let us consider the mixed matter–energy tensor $T^i_j = g_{jk}T^{ki}$. The formula $T := T^i_i$ highlights a scalar derived from the matter–energy tensor in the same way as the Ricci scalar is derived from the Ricci tensor R_{ij}. T is known as the Laue scalar and allows us to write the Einstein field equations in the equivalent form:

$$R_{ij} = k \cdot \left(T_{ij} - \frac{1}{2}T \cdot g_{ij}\right).$$

This form is useful when we are considering a Universe described by the condition $T_{ij} = 0$. For such Universes $T = 0$, too. Therefore their equations are

$$R_{ij} = 0.$$

If we are looking at the Einstein field equations

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} = kT_{ij},$$

in the left-hand side, we see the geometry of the space expressed in terms of tensors; in the right-hand side, there is a tensor describing mass and energy. The equality shows that the mass and energy together create the geometric structure of the space-time. Therefore, the geodesics of the space-time, i.e. the curves followed by "material points" under the action of the geometry itself, depends on the geometry created by the matter–energy tensor T_{ij}.

John Archibald Wheeler gave the most impressive description of the facts explained above:

> Space-time tells matter how to move; matter tells space-time how to curve.

6.3 Including the Cosmological Constant

In order to achieve a static universe, Einstein proposed a modified left member of his equations of the form

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} + \Lambda \cdot g_{ij},$$

where the constant Λ is the so-called *cosmological constant* . Its role was thought by Einstein to be a counterbalance at the attractive action of gravity. Later on, when Hubble discovered the expansion of the Universe, Einstein cancelled the term Λg_{ij} considering it a blunder. Recently, as we will discuss below, this term has been resumed in order to describe the accelerated expansion of the Universe.

Since $g_{ij;l} = 0$, it results

$$\left(R_{ij} - \frac{1}{2} R \cdot g_{ij} + \Lambda \cdot g_{ij} \right)_{;l} = 0.$$

Therefore the new Einstein field equations with cosmological constant become

$$R_{ij} - \frac{1}{2} R \cdot g_{ij} + \Lambda \cdot g_{ij} = k \cdot T_{ij}.$$

The Einstein field equations with cosmological constant are

$$R_{ij} - \Lambda \cdot g_{ij} = k \left(T_{ij} - \frac{1}{2} T \cdot g_{ij} \right),$$

where T is the Laue scalar. Again, in the case when $T_{ij} = 0$, since $T = 0$, the Einstein equations of such spaces reduce to

$$R_{ij} - \Lambda \cdot g_{ij} = 0.$$

The cosmological constant has a geometric nature in these cases.

Chapter 7
Differential Geometry at Work: Euclidean, Non-Euclidean, and Elliptic Geometric Models from Geometry and Physics

Measure what can be measured, and make measurable what cannot be measured.

Galileo Galilei

We are ready to present the two big pictures of non-Euclidean geometry models. The first one is the consequence of Euclidean $3D$ structure. The second one is revealed by Physics.

7.1 Euclidean, Non-Euclidean, and Elliptic Geometric Models from Geometry

We have already seen models of Euclidean geometry. In the case of plane $z = 0$, the metric

$$ds^2 = dx^2 + dy^2 + dz^2$$

induces directly the metric

$$ds^2 = dx^2 + dy^2$$

with null Gaussian curvature and straight lines as geodesics. The parallel transport is exactly the one which makes the lines to have the same slope in a system of coordinate (x, y). Therefore, for a line d and a point A not belonging to d, there is only a parallel line through A to d. This one is the geodesic through A having the slope m.

Having in mind this model of Euclidean geometry, which may be called the classical one, we can transfer it on a sphere using the stereographic projection. We did it in Sect. 1.15 when we obtained the metric with null Gaussian curvature

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_7

$$ds^2 = \frac{1}{\sin^4 \frac{u}{2}} \left(R^2 du^2 + R^2 \sin^2 u \, dv^2\right).$$

We have explained there how the geodesics look like and how the parallel lines look like. We can transfer the Euclidean metric in the same way on a different surface. In all these models, through a point $A \notin d$, there is only a parallel line to the given straight line d. We have to consider that straight line means geodesic. In the same way, all geodesics of the basic Euclidean model are transferred into geodesics in models derived after appropriate changes of coordinates.

Similarly, we studied how the metric

$$ds^2 = dx^2 + dy^2 + dz^2$$

is transferred on a sphere. The metric we obtained,

$$ds^2 = R^2 du^2 + R^2 \sin^2 u \, dv^2,$$

has constant positive Gaussian curvature, $K = \frac{1}{R^2}$. Geodesics of this metric are great circles; therefore, two geodesics intersect in two antipodal points. Denoting the two geodesics by c_1 and c_2 and

$$\{A, B\} := c_1 \cap c_2,$$

A, B and the centre of the sphere are collinear points. Therefore, in this elliptic geometry of the sphere, there are no parallel lines.

So, until now we saw only geometries in which there exists only one parallel line or no parallel lines. Not entering in technical details, let us call *non-secant lines* two lines which do not intersect.

Are there geometries in which through $A \notin d$ we can construct more than one non-secant line to d? Again, do not forget the meaning of line, i.e. a geodesic. The answer is yes, but we need some steps to describe it.

Step 1. Example 1.3.3 was related to the existence of a surface obtained by the rotation of a tractrix around an axis. The 3D metric

$$ds^2 = dx^2 + dy^2 + dz^2$$

acts on the parameterization of the pseudosphere producing the metric

$$ds^2 = \cot^2 u \, du^2 + \sin^2 u \, dv^2,$$

where $u \in (0, \frac{\pi}{2})$, $v \in (0, 2\pi)$.

Step 2. In Example 4.10.4, we determined its Gaussian constant curvature, $K = -1$.

Step 3. In Example 4.13.2, we have transferred the metric of the pseudosphere to the "half-plane"

$$H^2 := \{(x, y)|x, y \in \mathbb{R}, y > 0\}$$

using the transformation of coordinates

$$\begin{cases} u = u(x, y) = \arcsin \dfrac{1}{y} \\ v = v(x, y) = x,\ x \in \mathbb{R},\ y > 0. \end{cases}$$

The metric of half-plane becomes

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

and its Gaussian curvature is $K = -1$.

The half-plane H^2 endowed with the previous metric is called Poincaré half-plane and it is involved in the description of a non-Euclidean geometry model, in which, through $A \notin d$ pass more than two non-secant lines to given line d. We show this below.

Finally, it remains something obvious: using the inverse transform of coordinates, the pseudosphere will be the surface involved in the description of non-Euclidean geometry.

So, let us study the geodesics of the metric

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

We denote $x := x^1$, $y := x^2$. The Christoffel symbols are

$$\Gamma_{11,1} = \Gamma_{22,1} = \Gamma_{12,2} = \Gamma_{21,2} = 0,\ \Gamma_{12,1} = \Gamma_{21,1} = \Gamma_{22,2} = -\frac{1}{y^3},\ \Gamma_{11,2} = \frac{1}{y^3},$$

$$\Gamma^1_{11} = \Gamma^1_{22} = \Gamma^2_{12} = \Gamma^2_{21} = 0,\ \Gamma^1_{12} = \Gamma^1_{21} = \Gamma^2_{22} = -\frac{1}{y},\ \Gamma^2_{11} = \frac{1}{y}.$$

The equations of the geodesics are

$$\begin{cases} \ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0 \\ \ddot{y} + \frac{1}{y}\dot{x}^2 - \frac{1}{y}\dot{y}^2 = 0. \end{cases}$$

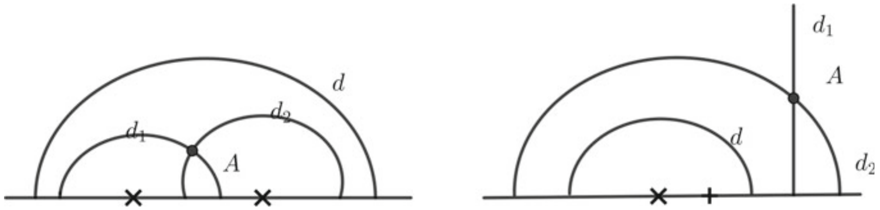


Fig. 7.1 Parallel lines through A to the line d in Poincaré half-plane H^2

Let us observe that

$$x = x(s) = a + R\tanh s; \quad y = y(s) = \frac{R}{\cosh s},$$

where a is a constant. It is a solution of the previous equations.

This first solution satisfies the equation

$$(x - a)^2 + y^2 = R^2.$$

This is, from the Euclidean point of view, the equation of a semicircle in the Poincaré half-plane. So, we draw the semicircle, but this is a geodesic, a straight line in the Poincaré half-plane. Another simple computation shows that $x = x(s) = a; \; y = y(s) = e^s$ satisfy the equations of geodesics.

From the Euclidean point of view, this is the parameterization of straight line perpendicular to $y = 0$. In fact, this geodesic is a straight line in Poincaré half-plane, too (Fig. 7.1). Now, it is only an exercise for the reader to write the equations of the geodesics of the pseudosphere.

Consider two semicircles, d_1 and d_2 which intersect at a point denoted A. Let us consider a bigger semicircle, d, such that d_1 and d_2 are drawn in its interior. Therefore, we are considering two geodesic lines through A non-secant to d. The Poincaré half-plane becomes a model of non-Euclidean geometry. Transferring this figure on pseudosphere, we give rise there to a model of non-Euclidean geometry. It is a simple exercise for the reader imagining a figure with a line d and two non-secant lines through A, one having the equation $x = a$.

Some remarks are now necessary. In the two-dimensional Euclidean plane E^2, let us consider the unit circle with the centre at $M(0, -1)$. Denote by $A(0, -2)$ a point which belongs to the circle, $B(X, Y)$ in the interior of the circle and $B'(x, y)$ in the superior half-plane such that $A, \; B, \; B'$ are collinear and $AB \cdot AB' = 4$. Using the previous two conditions, we can compute the coordinates x and y with respect to the coordinates (X, Y):

$$x(X, Y) = \frac{4X}{X^2 + (Y + 2)^2},$$

$$y(X, Y) = \frac{-2(X^2 + (Y + 1)^2 - 1)}{X^2 + (Y + 2)^2}.$$

Then

$$dx^2 + dy^2 = \frac{16}{(X^2 + (Y + 2)^2)^2}(dX^2 + dY^2)$$

and

$$\frac{1}{y^2} = \frac{(X^2 + (Y + 2)^2)^2}{4(X^2 + (Y + 1)^2 - 1)^2},$$

therefore the metric of the half-plane

$$ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$$

is transferred inside the unit circle centred at M as

$$ds^2 = \frac{4}{(1 - X^2 - (Y + 1)^2)^2}(dX^2 + dY^2).$$

If we translate the disc such as $M(0, -1)$ becomes $O(0, 0)$, the metric of the disc is

$$ds^2 = \frac{4}{(1 - X^2 - Y^2)^2}(dX^2 + dY^2).$$

The unit disc endowed with this metric is called the Poincaré disc. The Gaussian curvature of the metric is $K = -1$ and the geodesics of the disc are obtained transferring the geodesics of the half-plane inside the disc through the transformation of coordinates described above. Since the transformation is a geometric inversion (see [34]), we obtain two types of geodesics. Diameters of the Poincaré disc and arcs of circles inside the disc which are orthogonal to the circumference of the unit disc. Transforming a configuration like in Figure 2.1, we can see a non-Euclidean model of the geometry inside the Poincaré disc.

7.2 Euclidean, Non-Euclidean, and Elliptic Geometric Models from Physics

Let us prove that the metric

$$ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2$$

is related to Physics.

This is the Minkowski $(- + ++)$ metric of Special Relativity in geometric coordinates, that is when the speed of light is calibrated at $c = 1$. It comes from the quadratic form $\langle U, U \rangle$ attached to the Minkowski product

$$\langle U, V\rangle = -U_T V_T + U_X V_X + U_Y V_Y + U_Z V_Z.$$

In order to explain quickly why this metric is related to Special Relativity, let us consider only two coordinates T and X and a Lorentz transformation described as

$$T(t, x) = \frac{t + x\,v}{\sqrt{1 - v^2}};\;\; X(t, x) = \frac{t\,v + x}{\sqrt{1 - v^2}}.$$

As we know, this Lorentz transformation describes how the coordinates (t, x) of a frame which moves at constant speed v are seen in a frame at rest in coordinates (T, X). The Lorentz transformation implies the invariance of the quantity $-T^2 + X^2$ because

$$-T^2 + X^2 = -\left(\frac{t + x\,v}{\sqrt{1 - v^2}}\right)^2 + \left(\frac{t\,v + x}{\sqrt{1 - v^2}}\right)^2 = -t^2 + x^2.$$

Since this quantity depends on the Minkowski product

$$\langle U, U\rangle = -(U_T)^2 + (U_X)^2$$

it results that Lorentz transformations preserve the square of the Minkowski norm of vectors. The metric highlights this geometric aspect. In fact, if $v = \tanh\alpha$, the Lorentz transformation is a hyperbolic rotation of coordinates. All these facts will be completely explained and understood later in the chapter dedicated to Special Relativity.

Returning to geometric models, let us fix the time. At $T = T_0$, the space–time metric becomes

$$ds^2 = dX^2 + dY^2 + dZ^2$$

which is an Euclidean one.

This metric allows to obtain both the elliptic and the Euclidean geometries as we have in the first part of this section.

When we will discuss non-Euclidean models, a slice through a Minkowski sphere is involved. The sphere is related to the condition

$$-T^2 + X^2 + Y^2 + Z^2 = -1$$

and the slice will be described by the condition $Z = 0$.

Let us first understand why the unit space-like sphere

$$-T^2 + X^2 + Y^2 + Z^2 = -1$$

is important.

The parameterization we choose is

$$\begin{cases} T = \cosh t \\ X = \sinh t \cos x_1 \cos x_2 \\ Y = \sinh t \cos x_1 \sin x_2 \\ Z = \sinh t \sin x_1. \end{cases}$$

The space–time metric

$$ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2$$

endows the space-like sphere with the metric

$$ds^2 = dt^2 + \sinh^2 t \; dx_1^2 + \sinh^2 t \cos^2 x_1 dx_2^2.$$

After some computations, we can show that

$$R_{ij} + 2g_{ij} = 0 \text{ and } R = -6,$$

i.e.

$$R_{ij} - \frac{1}{2} R \; g_{ij} + \Lambda \; g_{ij} = 0$$

if the cosmological constant is $\Lambda = -1$.

Therefore, the Einstein field equations are satisfied for $\Lambda = -1$ and $T_{ij} = 0$.

We face a remarkable physics structure, a space-like unit sphere which is in fact a $3D$ Universe without matter called Anti-de Sitter space–time.

Let us now consider the $3D$ slice when $Z = 0$. This slice has the Minkowski $(- + +)$ metric

$$ds^2 = -dT^2 + dX^2 + dY^2.$$

The slice contains the surface

$$-T^2 + X^2 + Y^2 = -1,$$

which is still a Minkowski space-like sphere. From the Euclidean point of view, this space-like sphere is a two-sheeted hyperboloid $\mathcal{H}$ located in the interior of the light cone

$$-T^2 + X^2 + Y^2 = 0.$$

The parameterization of the space-like sphere is obtained for $x_1 = 0$ in the parameterization of the Anti-de Sitter space–time, i.e.

$$\mathcal{H}: \begin{cases} T = \cosh t \\ X = \sinh t \cos x_2 \\ Y = \sinh t \sin x_2. \end{cases}$$

The space–time metric

$$ds^2 = -dT^2 + dX^2 + dY^2$$

endows the unit space-like sphere $\mathcal{H}$ with the metric

$$ds^2 = dt^2 + \sinh^2 t \; dx_2^2.$$

This metric can be obtained directly from the Anti-de Sitter space–time metric for $x_1 = 0$ (Fig. 7.2).

The Christoffel symbols are

$$\Gamma^0_{11} = -\sinh t \cosh t, \;\; \Gamma^1_{01} = \Gamma^1_{10} = \coth t,$$

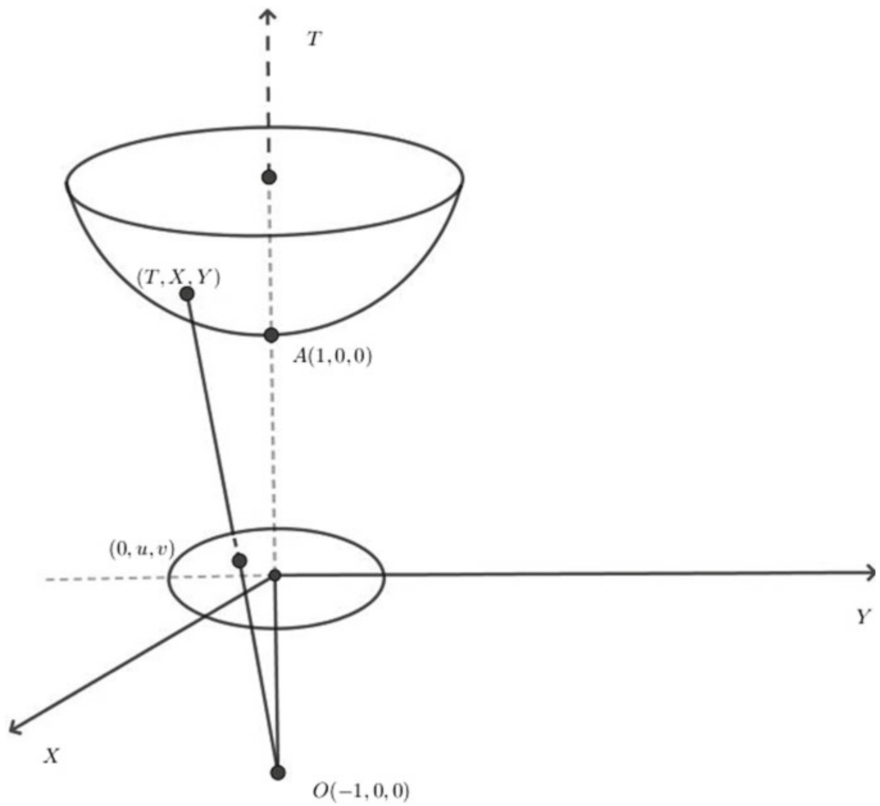


Fig. 7.2 Poincaré metric of the disc induced by a Minkowski sphere

therefore $R^0_{101} = -\sinh^2 t$; $R_{0101} = g_{00} R^0_{101} = -\sinh^2 t$. The Gaussian curvature is $K = -1$.

Next, we show that this Minkowski unit sphere is a model of non-Euclidean geometry.

Consider a point B with coordinates (T, X, Y), $T > 0$ which belongs to $\mathcal{H}$. We can consider the variables in the $3D$ slice to be (τ, u, v). Let $S(-1, 0, 0)$ be the "South Pole" of $\mathcal{H}$ and denote by A the intersection of SB with the plane $\tau = 0$.

The equation of the line SB is

$$\frac{\tau - T}{\tau + 1} = \frac{u - X}{u} = \frac{v - Y}{v},$$

therefore the intersection with $\tau = 0$ leads to

$$\begin{cases} u = \dfrac{X}{1+T} \\ \\ v = \dfrac{Y}{1+T}. \end{cases}$$

Using the parameterization of $\mathcal{H}$, the point A has the (u, v) coordinates

$$\begin{cases} u = \dfrac{\sinh t \cos x_2}{1 + \cosh t} \\ \\ v = \dfrac{\sinh t \sin x_2}{1 + \cosh t}. \end{cases}$$

It is easy to observe that

$$u^2 + v^2 < 1,$$

that is the point A belongs to a disc D in the plane $\tau = 0$. Moreover, these coordinates can be written as

$$\begin{cases} u(t, x_2) = \tanh \dfrac{t}{2} \cos x_2 \\ \\ v(t, x_2) = \tanh \dfrac{t}{2} \sin x_2. \end{cases}$$

The inverse transform is

$$\begin{cases} t(u, v) = 2 \tanh^{-1}(u^2 + v^2) \\ \\ x_2(u, v) = \arctan \dfrac{v}{u}. \end{cases}$$

After we compute the differentials dt and dx_2, it is relatively easy to replace into

$$ds^2 = dt^2 + \sinh^2 t \, dx_2^2.$$

The metric in D becomes

$$ds^2 = \frac{4}{(1 - u^2 - v^2)^2}(du^2 + dv^2)$$

which is the metric of the Poincaré disc. Since this Poicaré disc highlights a non-Euclidean geometry, the same non-Euclidean geometry exists on the unit Minkowski sphere $\mathcal{H}$. The geodesics are the images through S of the geodesics of D. In fact, the transformation we choose is a stereographic projection corresponding to the unit Minkowski sphere.

In the same way, let us obtain the Cayley–Klein metric of the non-Euclidean geometry. Consider the unit Minkowsky space-like sphere $-T^2 + X^2 + Y^2 = -1$ centred at the origin. As we know, its metric is (Fig. 7.3)

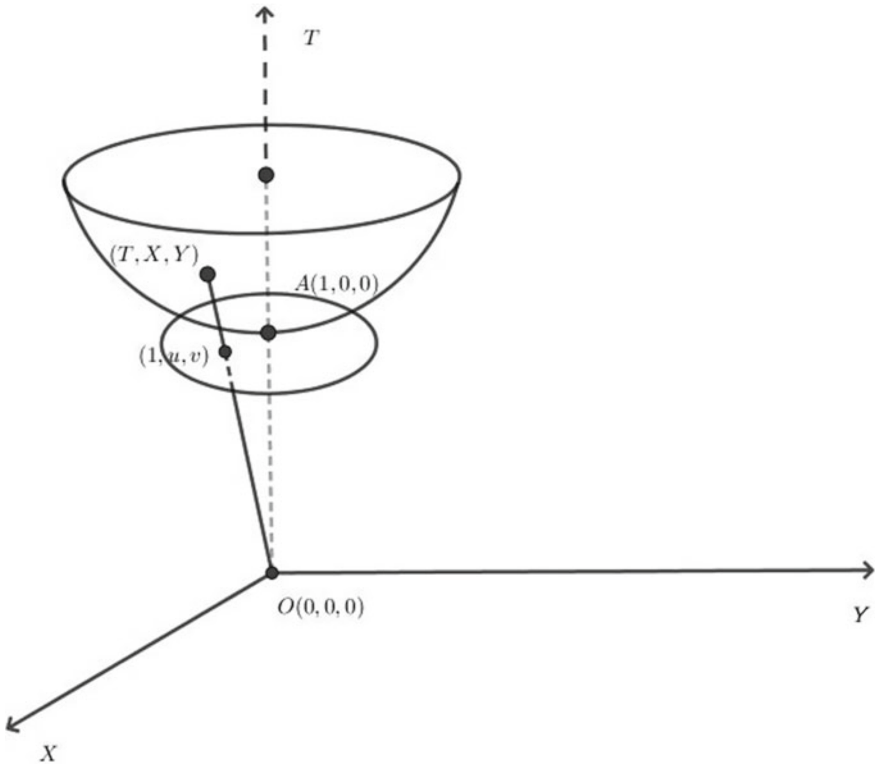


Fig. 7.3 The metric of Cayley–Klein non-Euclidian geometry

$$ds^2 = dt^2 + \sinh^2 t dx_2^2.$$

We transfer this metric to the tangent plane at the North Pole $N(1, 0, 0)$, denoted here by T_N, in the following way. Let $B(T, X, Y)$ be a point of $\mathcal{H}$. The straight line determined by this point B and the origin $O(0, 0, 0)$ intersects T_N at $(1, u, v)$. From

$$\frac{T}{1} = \frac{X}{u} = \frac{Y}{v}$$

we find the coordinates of B in the form

$$M\left(\frac{1}{\sqrt{1-u^2-v^2}}, \frac{u}{\sqrt{1-u^2-v^2}}, \frac{v}{\sqrt{1-u^2-v^2}}\right)$$

because $X = uT,\ Y = vT$ and $-T^1 + u^2T^2 + v^2T^2 = -1$.

Choosing

$$\begin{cases} u = \tanh t \cos x_2 \\ v = \tanh t \sin x_2 \end{cases}$$

and replacing in the coordinates of the point M, we obtain the usual parameterization of the Minkowski space-like sphere

$$(\cosh t, \sinh t \cos x_2, \sinh \sin x_2)$$

whose metric was written above, i.e.

$$ds^2 = dt^2 + \sinh^2 t dx_2^2.$$

We intend to find the metric of the tangent plane T_N in terms of u and v.

We have $1 - u^2 - v^2 = \dfrac{1}{\cosh^2 t}$ and $u^2 + v^2 = \tanh^2 t < 1$; therefore, the metric will be only in the interior of a unit disc D of T_N. From

$$\begin{cases} du = \dfrac{1}{\cosh^2 t} \cos x_2 dt - \tanh t \sin x_2 dx_2 \\[2ex] dv = \dfrac{1}{\cosh^2 t} \sin x_2 dt + \tanh t \cos x_2 dx_2, \end{cases}$$

it results

$$\frac{1}{1-u^2-v^2}(du^2 + dv^2) = \frac{\cosh^2 t - \sinh^2 t}{\cosh^2 t} dt^2 + \sinh^2 t dx_2^2 = -\tanh^2 t dt^2 + ds^2.$$

Since

$$udu + vdv = \frac{\sinh t}{\cosh^3 t} dt,$$

we obtain

$$\frac{(udu + vdv)^2}{(1 - u^2 - v^2)^2} = \tanh^2 t dt^2,$$

which leads us to the $K = -1$ metric in D

$$ds^2 = \frac{1}{1 - u^2 - v^2}(du^2 + dv^2) + \frac{(udu + vdv)^2}{(1 - u^2 - v^2)^2}.$$

The geodesics in D are straight lines obtained from the intersection of the planes of the geodesics of the Minkowski space-like sphere with D. The Cayley–Klein model of non-Euclidean geometry appears in D.

7.3 The Physical Interpretation

At the internet address https://archive.org/details/lascienceetlhypo00poin, it can be found the Poincaré famous book *Science et Hypothèse* [161]. Pages 83–87 offer a beautiful physical example of the Universe related to the non-Euclidean geometry in the disc model. The example is given in the interior of a sphere. Let us consider the interior of a sphere $S(O, 1)$ where O is the centre and 1 is the radius. This interior is the Universe for some intelligent inhabitants.

According to Poincaré, who conceived this particular Universe, both the Euclidean geometry and a temperature law are acting in the interior of the sphere.

The temperature is maximum at the centre and decreases to 0 on the surface of the sphere in which this Universe is included. The law of temperature variation is: if M is a point such that $OM = r$ then, the temperature at M is proportional to $1 - r^2$. Poincaré allows the temperature to contract or to dilate the length of the creatures according to their position after a rule he describes as the length of a ruler is proportional to its absolute temperature. So, a ruler having a side in O and the other side in M, such that the Euclidean length is $|OM| = r$, has a length proportional to $1 - r^2$.

The last Poincaré axiom is about how light travels in this Universe: the index of refraction of this Universe is inversely proportional to $1 - r^2$. We can suppose it as $\frac{4}{1 - r^2}$.

Having all these facts in mind, let us understand how the inhabitants will perceive their Universe. First of all, it is enough to understand the geometry of a disc containing the centre of the sphere. For Poincaré, this disc is Euclidean and it has the form of an open radius 1-disc. For the inhabitants, their length is smaller and smaller when they try to reach the border of this slice of Universe. They become shorter and shorter, their legs become shorter, and their steps become shorter. These things happen because the temperature acts by contracting the dimensions when they step to the border. The finite Euclidean Universe for Poincaré seems to be infinite for the small creatures. The rule established by Poicaré for distance will be understood by the inhabitants as

$$d^n(O, M) = \frac{1}{2} \cdot \ln \frac{1-x}{1-0} : \frac{-1-x}{-1-0},$$

that is, when x approaches 1, d^n approaches infinity. Of course, here OM is the x-axis. One inhabitant, mister "H", will observe that it is possible to describe this distance for two arbitrary points A, B in the form

$$d^n(A, B) := \ln \frac{\max_{P \in K} g_{AB}(P)}{\min_{P \in K} g_{AB}(P)}, \quad g_{AB} = \frac{|PA|}{|PB|}$$

where K is the boundary, that is the circle of radius 1, and $|PA|$ is the Euclidean distance between P in K and A in the interior of the disc.

The intelligent inhabitants will understand that light is moving on the "straight lines" of the Geometry of their Universe. Since the law of light propagation depends on the index of refraction, they will deduce the metric of their Universe as

$$ds^2 = \frac{4}{(1-(x^2+y^2))^2}(dx^2 + dy^2).$$

The straight lines (the geodesics), induced by the trajectories of ray lights, are diameters or arcs of circles bi-orthogonal to the border as we explained above.

There are two "parallel lines" to a given "line" through a given point. The sum of angles of a "triangle" is less than two right angles. Now they conclude they live in a non-Euclidean Universe.

Finally, the inhabitants have two ideas about their slice of the Universe, ideas which can be extended to the entire interior of the sphere:

(i) the Universe is infinite
(ii) the Universe is governed by the laws of non-Euclidean geometry and is curved. In each slice, the Gaussian curvature is a negative constant, $K(x, y) = -1$.

But this is not true, their Universe is a finite interior of a R-sphere and the underlying geometry is Euclidean, not hyperbolic!

Poincaré established that the inhabitants of his physical model are perfectly right to use hyperbolic geometry as the foundation of their Physics because it is convenient, but there is a non-sense to speak about the philosophical abstract truth or about an approximation of any truth, because intelligent inhabitants point of view is in collision with the way and laws their Universe was established.

Poincaré opinion is that the reality is not described by the most "realistic" Geometry "*la géométrie la plus vrai*", but by the most comfortable for the description of physical laws *(la géométrie la plus commode)*. Therefore, Poincaré believed that the Geometry of physical space is a conventional one.

7.4 Another Way to Obtain the Poincaré Disc Model Metric

The next theorem allows us to provide the Poincaré disc metric starting from the distance naturally related, in Chapter 3, to the Poincaré disc model. The theorem allows to obtain the metric of all non-Euclidean geometry models which come from the metrization procedure described in its statement. He was formulated by Dan Barbilian who is known for the generalizations of Poincaré work in the so-called Barbilian spaces (see [35, 38].)

Theorem 7.4.1 *(Barbilian's Theorem) Let K and J be two subsets of the Euclidean plane $\mathbb{R}^2$, and $K = \partial J$. Consider the function $f(M, A) = |MA|$, where, by $|MA|$, we denote the Euclidean distance. Consider*

$$g_{AB}(M) = \frac{f(M, A)}{f(M, B)} = \frac{|MA|}{|MB|},$$

and consider the semi-distance induced on J by the metrization procedure,

$$d(A, B) := \ln \frac{\max_{P \in K} g_{AB}(P)}{\min_{P \in K} g_{AB}(P)}.$$

Suppose furthermore that, for $M \in K$, the extrema $\max g_{AB}(M)$ *and* $\min g_{AB}(M)$ *for any A and B in J are reached each for a single point in K. Then:*

(i) For any $A \in J$ and any line d passing through A, there exist exactly two circles tangent to K and also to d at A.

(ii) The metric induced by the previous distance has the form

$$ds^2 = \frac{1}{4}\left(\frac{1}{R_1} + \frac{1}{r_1}\right)^2 (dx_1^2 + dx_2^2),$$

where R and r are the radii of the circles described in (i).

Proof Consider $A\,(x_1, x_2)$ and $B\,(y_1, y_2)$ in J and $M\,\left(x^1, x^2\right)$ in $J \cup K$.

The circle determined by the relation $\dfrac{|MA|}{|MB|} = \sqrt{\lambda}$ has the equation

$$\left(x^1 - x_1\right)^2 + \left(x^2 - x_2\right)^2 - \lambda(\left(x^1 - y_1\right)^2 + \left(x^2 - y_2\right)^2) = 0.$$

Its radius $\mathcal{R}$ is

$$\mathcal{R}^2 = \frac{\lambda |AB|^2}{(1-\lambda)^2} .$$

The maximum M_1 and the minimum m_1 values for the expression $\dfrac{|MA|^2}{|MB|^2}$ lead to the equalities

$$R_1^2 = \frac{M_1}{(1-M_1)^2}|AB|^2, \ r_1^2 = \frac{m_1}{(1-m_1)^2}|AB|^2.$$

The first equality becomes

$$\left(\frac{1+M_1}{1-M_1}\right)^2 = \frac{|AB|^2 + 4R_1^2}{|AB|^2},$$

and taking into account that $M_1 \geq 1$, it results

$$M_1 = 1 + \frac{2|AB|}{-|AB| + \sqrt{|AB|^2 + 4R_1^2}}.$$

In the same way, using $m_1 \leq 1$, we have

$$m_1 = 1 - \frac{2|AB|}{|AB| + \sqrt{|AB|^2 + 4r_1^2}} .$$

If A and B are close enough, i.e. $B = A + dA$, the Euclidean distance $|AB|^2$ becomes the arc element

$$d\sigma^2 = dx_1^2 + dx_2^2.$$

The distance between the points A and $A + dA$ leads to the new arc element $d\,(A, A + dA)$ denoted by ds.

So,

$$ds = d\,(A, A + dA) = \frac{1}{2}\frac{M_1 - m_1}{m_1}.$$

We have the approximations

$$\frac{2d\sigma}{-d\sigma + \sqrt{d\sigma^2 + 4R_1^2}} = \frac{d\sigma}{R_1} ,$$

and

$$\frac{2d\sigma}{d\sigma + \sqrt{d\sigma^2 + 4r_1^2}} = \frac{d\sigma}{r_1}.$$

The final computation leads to

$$ds = \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{r_1}\right) d\sigma,$$

i.e. the metric corresponding to the previous distance is

$$ds^2 = \frac{1}{4}\left(\frac{1}{R_1} + \frac{1}{r_1}\right)^2 \left(dx_1^2 + dx_2^2\right).$$

□

Theorem 7.4.2 *Consider the circle* Γ *centred at the origin and of radius* R. *Consider in the interior of the circle the Poincaré distance. Then, the associated metric, given by Barbilian's Theorem*

$$ds^2 = \frac{1}{4}\left(\frac{1}{R_1} + \frac{1}{r_1}\right)^2 \left(dx_1^2 + dx_2^2\right)$$

has the form

$$ds^2 = \frac{4R^2}{[R^2 - (x^2 + y^2)]^2} \cdot (dx^2 + dy^2).$$

Furthermore, the metric obtained by this procedure has the Gaussian curvature -1.

Proof In the case when Γ is a circle and J is its interior, we deal with a distance,

$$d(A, B) = \ln \frac{\max_{P\in\Gamma} g_{AB}(P)}{\min_{P\in\Gamma} g_{AB}(P)} = \ln \frac{\max_{P\in\Gamma} \frac{|PA|}{|PB|}}{\min_{P\in\Gamma} \frac{|PA|}{|PB|}},$$

called Poincaré distance of the disc. We would like to compute the metric of the disc induced by this distance and the previous theorem. Let A of coordinates (x_0, y_0), in the interior of Γ. Denote by $O_1(x_1, y_1)$ and $O_2(x_2, y_2)$ the centres of the two circles, each one tangent to the circle Γ and also tangent between them at A. Denote by m the slope of the tangent line Δ at A to both previous circles. Line O_1O_2 has the equation

$$y - y_0 = -\frac{1}{m}(x - x_0).$$

Therefore, the points O_1 and O_2 have the coordinates $\left(x_i, y_0 - \frac{1}{m}(x_i - x_0)\right)$, for $i = 1, 2$. Furthermore,

$$R_i^2 = |O_iA|^2 = \frac{m^2 + 1}{m^2}(x_i - x_0)^2, \quad i = 1, 2.$$

Without losing generality, we assume that $x_1 - x_0 \le 0$ and $x_2 - x_0 \ge 0$, with the equality case reached for $\Delta \| Ox$. It is worth remarking that $x_1 - x_0 < 0$, if $m > 0$. Thus

$$|O_1 A| = \frac{\sqrt{m^2+1}}{m}(x_0 - x_1),$$

and

$$|O_2 A| = \frac{\sqrt{m^2+1}}{m}(x_2 - x_0).$$

Therefore, the circles have the centres in $(x_1, y_0 - \frac{1}{m}(x_1 - x_0))$ and in $(x_2, y_0 - \frac{1}{m}(x_2 - x_0))$, and the radii $R_1 = \dfrac{\sqrt{m^2+1}}{m}(x_0 - x_1)$ and $R_2 = \dfrac{\sqrt{m^2+1}}{m}(x_2 - x_0)$.

To obtain the coordinates of the point T_1', we recall that it lies at the intersection between the circle $x^2 + y^2 = R^2$ and the line

$$y = \frac{1}{x_1}\left[y_0 - \frac{1}{m}(x_1 - x_0)\right]x,$$

which passes through the collinear points O, O_1 and T_1'. Solving the system, we get the coordinates of T_1' as follows:

$$\left(\frac{Rx_1}{\sqrt{x_1^2 + \left(y_0 - \frac{1}{m}(x_1 - x_0)\right)^2}}, \frac{R(y_0 - \frac{1}{m}(x_1 - x_0))}{\sqrt{x_1^2 + \left(y_0 - \frac{1}{m}(x_1 - x_0)\right)^2}}\right).$$

By direct computation, we get

$$|O_1 T_1'| = R - \sqrt{x_1^2 + \left(y_0 - \frac{1}{m}(x_1 - x_0)\right)^2}.$$

Since the segments $O_1 T_1'$ and $O_1 A$ are radii of the circle of centre O_1 and radius R_1, we set up the equalities

$$\frac{\sqrt{m^2+1}}{m}(x_0 - x_1) = R_1 = R - \sqrt{x_1^2 + (y_0 - \frac{1}{m}(x_1 - x_0))^2}.$$

It follows that

$$x_0 - x_1 = \frac{R_1 m}{\sqrt{m^2+1}},$$

therefore

$$(R - R_1)^2 = x_1^2 + \left(y_0 + \frac{R_1}{\sqrt{m^2+1}}\right)^2.$$

Since

$$x_1 = x_0 - \frac{R_1 m}{\sqrt{m^2+1}},$$

we have

$$(m^2+1)(R-R_1)^2 - (y_0\sqrt{m^2+1}+R_1)^2 = (x_0\sqrt{m^2+1}-R_1 m)^2$$

i.e.

$$R_1 = \frac{\sqrt{m^2+1}}{2} \cdot \frac{R^2 - x_0^2 - y_0^2}{R\sqrt{m^2+1} - x_0 m + y_0}.$$

In a similar way, we obtain

$$R_2 = \frac{\sqrt{m^2+1}}{2} \cdot \frac{R^2 - x_0^2 - y_0^2}{R\sqrt{m^2+1} + x_0 m - y_0}.$$

It results in the relation

$$\frac{1}{4}\left(\frac{1}{R_1} + \frac{1}{R_2}\right)^2 = \frac{4R^2}{(R^2 - x_0^2 - y_0^2)^2},$$

i.e. the *Poincaré metric of the disc* is

$$ds^2 = \frac{4R^2}{[R^2 - (x^2+y^2)]^2} \cdot (dx^2 + dy^2).$$

By a straightforward computation, we can easily see that the Gaussian curvature of this metric is $K(x, y) = -1$. □

The reader understands now all the possible connections which can be made when we intend to see the two big pictures of non-Euclidean geometries.

Chapter 8
Gravity in Newtonian Mechanics

Per Aspera ad Astra.

Newtonian mechanics is a branch of Physics which studies the way in which the bodies are changing in time their position in space. The space in which the objects are at rest (or they change their position) is the Euclidean three-dimensional space E^3. All objects, regardless of size, can be identified as points with a given mass in the previous space. So, the Euclidean frame of coordinates $Oxyz$ becomes the absolute place where all is happening. Newtonian Mechanics accepts an universal time in which all changes in position take place. Forces are seen as vectors. For a given point M in space, the vector $\vec{X}=\overrightarrow{OM}$ is called a position vector. If the point evolves in time, we write this as

$$\vec{X}\,(t) = (x(t), y(t), z(t)).$$

The velocity vector is

$$\dot{\vec{X}} = (\dot{x}(t), \dot{y}(t), \dot{z}(t))$$

and the acceleration vector is

$$\ddot{\vec{X}} = (\ddot{x}(t), \ddot{y}(t), \ddot{z}(t)).$$

Of course, we make the assumption that the coordinate functions are indefinitely differentiable on their domain of definition which differs from a model to another. The foundations of Newtonian Mechanics are based on three fundamental principles, the so-called Newton's laws of motion. They were introduced by Isaac Newton in "Philosophiae Naturalis Principia Mathematica", book published in 1687.

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_8

The Principle of Inertia, or the first law, asserts: “A physical body preserves its state of rest or will continue moving at its current velocity conserving its direction, until a force causes a change in its state of moving or rest. The physical body will change the velocity and the direction according to this force”. A particular case is related to the rectilinear uniform motion, when the body is moving on a straight line at constant speed. The frames where this principle is available are called inertial frames. These frames are at rest or they move rectilinear at constant speed. This fundamental principle was first enunciated by Galilei. We can say that this principle tells us where, according to Newton, the two other fundamental principles make sense: in inertial frames. In the same time, it tells us that it is impossible to make a distinction between the state “at rest” and the state “rectilinear motion at constant speed”. Imagine you are in the bowl of a ship and you have no possibility to observe outside. You slept and you waked up. You cannot distinguish between the two states without an observation, a possible comparison. You will play table tennis alike in both states, the object falls down in the same way in both states, etc. The two states are equivalent for you in the given conditions. Newton introduces a concept, the quantity of motion of a body as the product between the mass m and its velocity $\vec{v}$. This quantity of motion is known today as momentum and it is denoted by $\vec{p}$, therefore $\vec{p} := m\vec{v}$. The second law asserts: “The force who acts on a body is the variation in time of the quantity of motion”. Its differential form is $\vec{F} = \frac{d\vec{p}}{dt}$. If m does not depend on time, then

$$\vec{F} = \frac{d\vec{p}}{dt} = m\frac{d\vec{v}}{dt} = m\vec{a},$$

that is the force which acts on a body is proportional to the body acceleration through its mass. Newton’s third law: “When a body acts on a second body by the force $\vec{F}$, the second body simultaneously reacts on the first body by the force $-\vec{F}$.”

This chapter is devoted to gravity. We try to outline the basic facts about gravity, we prove the vacuum field equation and the general gravitational field equation. The artefact we use to express these laws is the gravitational potential. Later, in the chapter devoted to general relativity, the same gravitational potential is involved, in general, in metric components, and, specifically, in the coefficients of the Schwarzschild metric. The step towards general relativity is made when the tidal acceleration equations are written in a geometric form corresponding to a space endowed with a metric. However, our journey to relativity has to wait because we need some other tools until the moment we derive Einstein’s field equations via the Einstein–Hilbert action. We study Lagrangians and metrics induced by Lagrangians, where Euler–Lagrange equations become the geodesic equations of these metrics. Finally, we will connect these results to non-Euclidean geometry models. Kepler’s laws are derived. Later, in the same general relativity chapter, we understand how the conic curve, found as the trajectory of a planet, is still the geodesic trajectory approximation of the same planet in a given metric. An excellent discussion on Newtonian mechanics, in gravitational perspective, can be found in the book [66].

8.1 Gravity. The Vacuum Field Equation

Let us start to study about the gravity. Later on, in the book, gravity will be studied following Einstein's ideas. Now, we concentrate on gravity as a force trying to understand it from the classical mechanics point of view.

In the Euclidean 3-dimensional space E^3, let us consider two bodies of masses M and m, $M > m$, located at the points $X_1(x_1, y_1, z_1)$ and $X(x, y, z)$. The position vectors $\overrightarrow{OX_1}$ and $\overrightarrow{OX}$, where $O(0, 0, 0)$ is the origin, are simply denoted by $\vec{X_1}= (x_1, y_1, z_1)$ and $\vec{X}= (x, y, z)$. Let us define

$$\vec{r} := \vec{X} - \vec{X_1} = (x - x_1, y - y_1, z - z_1)\,.$$

The length of $\vec{r}$ is

$$r := \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2}$$

and the unit vector pointing the point X_1 from the point X is

$$\vec{u} = -\frac{\vec{X} - \vec{X_1}}{r} = -\frac{\vec{r}}{r} = -\left(\frac{x - x_1}{r}, \frac{y - y_1}{r}, \frac{z - z_1}{r}\right).$$

Newton stated that the *gravitational force* induced by the body of mass M which acts on the body of mass m has the intensity $F = G\dfrac{mM}{r^2}$, where $G = 6.67 \cdot 10^{-11} \dfrac{(m)^3}{(kg) \cdot (s)^2}$ is the *gravitational constant*. It can be described by the gravitational force vector

$$\vec{F} = \frac{GmM}{r^2}\,\vec{u} = -\frac{GmM}{r^2}\frac{\vec{r}}{r} = -\frac{GmM}{r^2}\left(\frac{x - x_1}{r}, \frac{y - y_1}{r}, \frac{z - z_1}{r}\right).$$

Before continuing, let us write the previous formula in the form

$$\vec{F} = m\frac{GM}{r^2}\,\vec{u}\,,$$

where $\vec{u}$ is a unitary vector. The mass m of the body gravitationally attracted seems to be like a "gravitational charge", if we compare $F = G\dfrac{mM}{r^2}$ with the similar formula which describes the intensity of an electric force, $F = k\dfrac{q_1q_2}{r^2}$. Therefore we can think at m to be a *gravitational mass* denoted by m_g.

In the special case, when we consider a body gravitational attracted by the Earth, M is the mass of the Earth, r is the radius of the Earth and G the gravitational constant, it results

$$F = m_g \cdot A,$$

where $A = \dfrac{GM}{r^2}$ is a constant acceleration denoted by g, where $g = 9.81\dfrac{(m)}{(s)^2}$.

In Newton's second law of motion, the mass m seems to be a constant which makes possible to compare the intensity of the force and the magnitude of acceleration, $F = ma$. This is an *inertial mass*, denoted by m_i, because the first Newton's law establishes the frames where the all three laws are true: the inertial frames. Therefore $F = m_i a$. In the case when F is the gravitational force exerted by the Earth on the body of mass m_i, $F = m_i g$. It results

$$\frac{m_g}{m_i} = \frac{gr^2}{GM} = k.$$

The constant k is not equal to 1 by definition, but, if we measure the weight, the space and the time with some other scaled units, the ratio $\dfrac{m_g}{m_i}$ results 1.

So we can accept that the gravitational mass is the same as the inertial mass, and we can denote by m the value $m_g = m_i$. This is the equivalence principle as formulated by Galileo[1]. We will see that it assumes a fundamental role in the formulation of general relativity.

Let us return to the formula

$$\vec{F} = m\frac{GM}{r^2}\vec{u}$$

seen as $\vec{F} = m\vec{A}$. We can define the *gravitational acceleration* as the vector

$$\vec{A} = \frac{GM}{r^2}\vec{u}\ .$$

This gravitational acceleration is also called a *gravitational field* induced by the body of mass M. This definition suggests how the gravity acts. In coordinates we have

$$\vec{A} = -\frac{GM}{r^2}\left(\frac{x - x_1}{r}, \frac{y - y_1}{r}, \frac{z - z_1}{r}\right).$$

We define the *gravitational potential* of the field $\vec{A}$ to be the function

[1] It is worth noticing that this is a peculiarity of gravitational force. For example, for the Coulomb force involving electric charges q, it is $m_i \neq q$. This means that equivalence principle is proper of gravity.

$$\Phi(x, y, z) = -\frac{GM}{r}.$$

This definition makes sense at all points of the Euclidean three-dimensional space except (x_1, y_1, z_1) where the gravitational source is located. It is easy to observe that

$$\frac{\partial \Phi}{\partial x} = \frac{GM}{r^2}\frac{\partial r}{\partial x} = \frac{GM}{r^2}\left(\frac{x - x_1}{r}\right).$$

If we define the *gradient* of the gravitational potential Φ by $\nabla \Phi := \left(\frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y}, \frac{\partial \Phi}{\partial z}\right)$, using the previous computation, we can prove

$$\nabla \Phi = \frac{GM}{r^2}\left(\frac{x - x_1}{r}, \frac{y - y_1}{r}, \frac{z - z_1}{r}\right) = -\vec{A}.$$

The *Laplace operator*, or simply, the Laplacian, denoted by ∇^2, is defined as

$$\nabla^2 := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

The Laplacian of the gravitational potential is

$$\nabla^2 \Phi = \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial y^2} + \frac{\partial^2 \Phi}{\partial z^2}$$

and can be computed. As we know

$$\frac{\partial \Phi}{\partial x} = \frac{GM}{r^2}\frac{x - x_0}{r},$$

therefore

$$\frac{\partial^2 \Phi}{\partial x^2} = GM \cdot \frac{r^3 - 3r^2\frac{\partial r}{\partial x}}{r^6} = GM\left(\frac{1}{r^3} - 3\frac{(x - x_1)^2}{r^5}\right),$$

i.e.

$$\nabla^2 \Phi = GM\left(\frac{3}{r^3} - 3\frac{r^2}{r^5}\right) = 0.$$

Therefore we showed that for all the points $(x, y, z) \neq (x_1, y_1, z_1)$ the gravitational potential

$$\Phi(x, y, z) = -\frac{GM}{r}$$

satisfies $\nabla^2 \Phi = 0$. Having in mind that the gravitational source is located at (x_1, y_1, z_1), we have proved that in the remaining "empty space", i.e. in vacuum, the Newtonian equation of the gravitational field, expressed with respect to its gravitational potential, is

$$\nabla^2 \Phi = 0.$$

The previous formula is known as *Newton's vacuum field equation*.

What is happening at a source point? We remember our previous construction with the gravitational potential

$$\Phi(x, y, z) = -\frac{1}{r}$$

where

$$r := \sqrt{x^2 + y^2 + z^2}$$

and

$$\nabla^2 \Phi(x, y, z) = 0$$

for all $(x, y, z) \neq (0, 0, 0)$.

Let us introduce the gravitational potential

$$\Phi_b(x, y, z) = -\frac{1}{\bar{r}_b}$$

where

$$\bar{r}_b := \sqrt{(x - b)^2 + y^2 + z^2},$$

that is the source is now $(b, 0, 0)$. The corresponding gravitational field is

$$\vec{A_b}\,(x, y, z) = -\nabla \Phi_b(x, y, z) = -\frac{1}{\bar{r}_b^2}\left(\frac{x - b}{\bar{r}_b}, \frac{y}{\bar{r}_b}, \frac{z}{\bar{r}_b}\right).$$

After easy computations

$$\frac{\partial \vec{A_b}}{\partial x}(0, 0, 0) = \left(\frac{2}{b^3}, 0, 0\right)$$

$$\frac{\partial \vec{A_b}}{\partial y}(0, 0, 0) = \left(0, -\frac{1}{b^3}, 0\right)$$

$$\frac{\partial \vec{A_b}}{\partial z}(0, 0, 0) = \left(0, 0, -\frac{1}{b^3}\right).$$

Now, we observe that the Hessian of the gravitational potential $d^2\Phi_b$ is the matrix with components $\frac{\partial \vec{A_b}}{\partial x_j}$, where $x_i \in \{x, y, z\}$, satisfying the relation

$$d^2\Phi_b(0,0,0) = \begin{pmatrix} \frac{\partial \vec{A_b}}{\partial x}(0,0,0) \\ \frac{\partial \vec{A_b}}{\partial y}(0,0,0) \\ \frac{\partial \vec{A_b}}{\partial z}(0,0,0) \end{pmatrix} = \begin{pmatrix} \frac{2}{b^3} & 0 & 0 \\ 0 & -\frac{1}{b^3} & 0 \\ 0 & 0 & -\frac{1}{b^3} \end{pmatrix}.$$

On the other hand, it can be seen as the matrix with the components $\frac{\partial^2\Phi_b}{\partial x_i \partial x_j}$, that is,

$$d^2\Phi_b(0,0,0) = \begin{pmatrix} \frac{\partial^2\Phi_b}{\partial x^2} & \frac{\partial^2\Phi_b}{\partial x\partial y} & \frac{\partial^2\Phi_b}{\partial x\partial z} \\ \frac{\partial^2\Phi_b}{\partial y\partial x} & \frac{\partial^2\Phi_b}{\partial y^2} & \frac{\partial^2\Phi_b}{\partial y\partial z} \\ \frac{\partial^2\Phi_b}{\partial z\partial x} & \frac{\partial^2\Phi_b}{\partial z\partial y} & \frac{\partial^2\Phi_b}{\partial z^2} \end{pmatrix}.$$

In fact the first line of the previous matrix is $\frac{\partial \vec{A_b}}{\partial x} = \left(\frac{\partial^2\Phi_b}{\partial x^2}, \frac{\partial^2\Phi_b}{\partial x\partial y}, \frac{\partial^2\Phi_b}{\partial x\partial z}\right)$, etc. Combining the previous results, the trace of Hessian matrix is the Laplacian of the gravitational potential, i.e.

$$Tr\left(d^2\Phi_b\right)(0,0,0) = \nabla^2\Phi_b(0,0,0) = \frac{2}{b^3} - \frac{1}{b^3} - \frac{1}{b^3} = 0$$

for all points $(x, y, z) \neq (b, 0, 0)$. When $b \to 0$, the gravitational potential $\nabla\Phi_b$ approaches the gravitational potential $\nabla\Phi$, therefore $\nabla\Phi_b^2$ $(0,0,0) = 0 \to \nabla\Phi^2(0,0,0)$. It means $\nabla\Phi^2(0,0,0) = 0$. We may conclude that the vacuum equation becomes

$$\nabla\Phi^2 = 0$$

everywhere, not only for all points without the source.

Let us now suppose that there are many gravitational sources, and we label the gravitational potentials. For each point (x_j, y_j, z_j), one can define

$$r_j := \sqrt{(x - x_j)^2 + (y - y_j)^2 + (z - z_j)^2}$$

and the gravitational potentials

$$\Phi_j(x, y, z) = -\frac{GM_j}{r_j}.$$

The total gravitational potential determined by the N sources is

$$\Phi(x, y, z) = \sum_{1}^{N} \Phi_j(x, y, z) = -\sum_{1}^{N} \frac{GM_j}{r_j}.$$

Theorem 8.1.1 *For* $(x, y, z) \neq (x_j, y_j, z_j)$, $j \in \{1, 2, ..., N\}$, *the total gravitational potential satisfies the gravitational field equation in vacuum*

$$\nabla^2 \Phi = 0.$$

Proof The linearity of Φ allows to work as previously, for all $j \in \{1, 2, ..., N\}$ having

$$\frac{\partial \Phi_j}{\partial x} = \frac{GM_j}{r_j^2} \frac{x - x_j}{r_j}.$$

Therefore

$$\frac{\partial^2 \Phi_j}{\partial x^2} = GM_j \cdot \frac{r_j^3 - 3r_j^2 \dfrac{\partial r_j}{\partial x}}{r_j^6} = GM_j \left(\frac{1}{r_j^3} - 3\frac{(x - x_j)^2}{r_j^5} \right),$$

i.e.

$$\nabla^2 \Phi = G \sum_{1}^{N} M_j \left(\frac{3}{r_j^3} - 3\frac{r_j^2}{r_j^5} \right) = 0.$$

The equation $\nabla^2 \Phi = 0$ is also known as the *Laplace equation for gravity.* □

In a similar way it can be proved.

Corollary 8.1.2 *For multiple sources, the equation* $\nabla \Phi^2 = 0$ *holds everywhere.*

8.2 Divergence of a Vector Field in a Euclidean 3D-Space

Let us consider an *incompressible fluid flow* described by the vector $\vec{F} := \rho \vec{V}$, where $\rho := \rho(x, y, z)$ is the *density of the incompressible fluid* at (x, y, z) and $\vec{V} = \vec{V}(x, y, z)$ is the speed vector at each point of a given region D of the Euclidean space.

If we are looking at the fact that F is measured in $\dfrac{(kg)}{(m)^2 \cdot (s)}$, we see in fact how much matter flows through a unit surface area in a unit time.

Consider a small parallelepiped centred at $(x, y, z) \in D$ and with sides of lengths Δx, Δy, Δz parallel to the axis of coordinates. The vector flow $\vec{F}$ has three components, $\vec{F}= (F_x, F_y, F_z)$. We can suppose the parallelepiped small enough to have the flow $\vec{F}$ constant over each face, that is at each point of a face, $\vec{F}$ has the same three given components. We are interested in expressing the *net outflow* through this parallelepiped, i.e. the algebraic sum of all outward flow vectors through the six faces.

The flow through the face of area $\Delta y \Delta z$ at the point $\left(x - \frac{\Delta x}{2}, y, z\right)$ is

$$F_x\left(x - \frac{\Delta x}{2}, y, z\right) \Delta y \Delta z.$$

Suppose this is an inflow. In the same way, the flow through the face of area $\Delta y \Delta z$ at the point $\left(x + \frac{\Delta x}{2}, y, z\right)$ is

$$F_x\left(x + \frac{\Delta x}{2}, y, z\right) \Delta y \Delta z$$

and this one is an outflow. Therefore the total outflow through these two parallel faces is

$$F_x\left(x + \frac{\Delta x}{2}, y, z\right) \Delta y \Delta z - F_x\left(x - \frac{\Delta x}{2}, y, z\right) \Delta y \Delta z \approx \frac{\partial F_x}{\partial x}(x, y, z)\Delta x \Delta y \Delta z,$$

where the last approximation was made taking into consideration the small dimensions of the parallelepiped.

Considering the contribution of the other two pairs of parallel faces, the total outflow through the parallelepiped faces becomes

$$\left(\frac{\partial F_x}{\partial x}(x, y, z) + \frac{\partial F_y}{\partial y}(x, y, z) + \frac{\partial F_z}{\partial z}(x, y, z)\right) \Delta x \Delta y \Delta z.$$

The *divergence* of $\vec{F}$ is defined by

$$\operatorname{div} \vec{F}:= \frac{\partial F_x}{\partial x}(x, y, z) + \frac{\partial F_y}{\partial y}(x, y, z) + \frac{\partial F_z}{\partial z}(x, y, z)$$

and a physical interpretation of it as total outflow over the parallelepiped is that presented above.

We can conclude: On the entire region D, the total outflow over D is

$$\mathcal{F}(D) := \int_D \text{div} \vec{F} \, d^3x = \text{div} \vec{F} \, (\vec{u_\eta}) \cdot \, volD,$$

where d^3x is the volume element $dxdydz$ and the last equality is a consequence of a mean value theorem for the given triple integral. A consequence of the last formula is

$$\lim_{D \to \vec{u}} \frac{\mathcal{F}(D)}{volD} = \text{div} \vec{F} \, (\vec{u}).$$

8.3 Covariant Divergence

We have discussed about a flow of an incompressible fluid in an Euclidean space. How this discussion changes if we are talking about an incompressible fluid in a region where the parallelism is not the Euclidean one? The problem appears when we consider the difference

$$F_x \left(x + \frac{\Delta x}{2}, y, z \right) \Delta y \Delta z - F_x \left(x - \frac{\Delta x}{2}, y, z \right) \Delta y \Delta z$$

because it means that we have moved by parallel transport the vector $\left(-F_x \left(x - \frac{\Delta x}{2} \right), 0, 0 \right)$ to the other face at the point $\left(x + \frac{\Delta x}{2}, y, z \right)$.

Therefore we parallel transport the contravariant vector $\left(-F_x \left(x - \frac{\Delta x}{2} \right), 0, 0 \right)$ along the infinitesimal vector $A^1 = (\Delta x, 0, 0)$.

Since, in general, $\Gamma^k_{ij} \neq 0$, the parallel transport along $A^1 = (\Delta x, 0, 0)$ for a contravariant vector $V = (V^1, 0, 0)$ leads to a vector whose first component is

$$V^1 \left(x - \frac{\Delta x}{2}, y, z \right) + \Delta V^1,$$

where

$$\Delta V^1 = -\Gamma^1_{ij} V^j \Delta x^i = -\Gamma^1_{1j} V^j \Delta x = -\Gamma^1_{11} V^1 \Delta x.$$

The difference

$$\left[V^1 \left(x + \frac{\Delta x}{2}, y, z \right) - V^1 \left(x - \frac{\Delta x}{2}, y, z \right) + \Gamma^1_{11} V^1 \Delta x \right] \Delta y \Delta z$$

is

$$\left(\frac{\partial V^1}{\partial x} + \Gamma^1_{11} V^1 \right) \Delta x \Delta y \Delta z,$$

i.e. the covariant derivative with respect to the first variable denoted by

$$V^1_{;1}\Delta x\Delta y\Delta z.$$

We have three pairs of opposite faces corresponding to the three directions, therefore the net outflow is

$$(V^1_{;1}+V^2_{;2}+V^3_{;3})\Delta x\Delta y\Delta z$$

for a parallelepiped in a region where the Euclidean parallel transport is replaced by the general parallel transport.

The quantity $V^s_s := V^1_{;1}+V^2_{;2}+V^3_{;3}$ is the covariant divergence of a contravariant vector (V^1, V^2, V^3).

In our case, we obtain

$$\left(F_x\left(x+\frac{\Delta x}{2}, y, z\right) - F_x\left(x-\frac{\Delta x}{2}, y, z\right) + F_x\Gamma^1_{11}\Delta x\right)\Delta y\Delta z \approx$$

$$\approx\left(\frac{\partial F_x}{\partial x}+F_x\Gamma^1_{11}\right)\Delta x\Delta y\Delta z = F_{x;1}\Delta x\Delta y\Delta z.$$

For the entire parallelepiped we have the total net outflow

$$(F_{x;1}+F_{y;2}+F_{z;3})\Delta x\Delta y\Delta z$$

expressed with respect to the covariant derivative.

Definition 8.3.1 The quantity $(F_{x;1}+F_{y;2}+F_{z;3})$ expressed with respect to the covariant derivatives of components is called a *covariant divergence* of the field F.

8.4 The General Newtonian Gravitational Field Equations

If a gravitational source of mass M is placed at (x_1, y_1, z_1) and no other gravitational source exists, we have deduced the vacuum fields equation

$$\nabla^2\Phi(x, y, z) = 0.$$

If there are many gravity sources $(x_j, y_j, z_j),\ j\in\{1, 2, ..., N\}$, we have defined

$$r_j := \sqrt{(x-x_j)^2+(y-y_j)^2+(z-z_j)^2}$$

and the corresponding gravitational potentials

$$\Phi_j(x, y, z) = -\frac{GM_j}{r_j}.$$

The total gravitational potential, determined by the N sources, was

$$\Phi(x, y, z) = \sum_1^N \Phi_j(x, y, z) = -\sum_1^N \frac{GM_j}{r_j}.$$

We have proved that the vacuum field equation, in this case, is

$$\nabla^2\Phi(x, y, z) = 0\,,$$

and it makes sense for all (x, y, z) of the space.

Now suppose that in a bounded region D of the Euclidean space E^3 there is a continuous distribution of matter and point sources. This continuous distribution of matter is defined by a density function $\rho = \rho(x, y, z)$ measured in $\frac{(kg)}{(m)^3}$. Outside D we have $\rho \equiv 0$.

How it looks like the gravitational field equation in this case? Let us prove the following.

Theorem 8.4.1 (General Gravitational Field Equation) *If D is a region of the space where it exists a continuous distribution of matter defined by the density function ρ, then*

$$\nabla^2\Phi(x, y, z) = 4\pi G\rho(x, y, z)$$

everywhere in D.

Proof Outside D, where $\rho = 0$, the theorem reduces to the vacuum field equation. It remains to prove the statement for all the points of D. We cover D with parallelepipeds. To do this, we consider points on Ox-axis and parallel planes to yOz through these points. In the same way, we take into account parallel planes to xOz through points on Oy and parallel planes to xOy through points on Oz. We obtain parallelepipeds with the faces parallel to the planes determined by the axes of coordinates. Some of parallelepipeds are completely inside D, some are completely outside D and some of them contain parts inside and outside.

Now we can index the points and we can denote the centres of parallelepipeds which cover D as being (x_i, y_j, z_k) and the corresponding dimensions of sides as $\Delta x_i, \Delta y_j, \Delta z_k$.

We can suppose the mass of such parallelepiped is $\rho(x_i, y_j, z_k)\Delta x_i \Delta y_j \Delta z_k$.

The corresponding gravitational potential at a point $(x, y, z) \in E^3$ is

$$\Phi(x, y, z) \approx \sum \Phi_j(x, y, z),$$

that is,

$$\Phi(x,y,z) = -\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{P} \frac{G\rho(x_i,y_j,z_k)}{\sqrt{(x-x_i)^2+(y-y_j)^2+(z-z_k)^2}}\Delta x_i \Delta y_j \Delta z_k.$$

We can improve the approximation of the gravitational potential formula considering more points on the axes and, at limit, we obtain

$$\Phi(x,y,z) = -G\int_D \rho(u,v,w)\frac{1}{\sqrt{(x-u)^2+(y-v)^2+(z-w)^2}}d^3u,$$

where d^3u is the volume element $dudvdw$. If $(x,y,z)\notin D$, the integral has sense. We are able to show that the integral has sense even for points $(x,y,z)\in D$. Consider a change of coordinates in E^3 defined by

$$u = x + r\sin x^2\cos x^1$$

$$v = y + r\sin x^2 \sin x^1$$

$$w = z + r\cos x^2.$$

We observe

$$r(x,y,z) = \sqrt{(x-u)^2+(y-v)^2+(z-w)^2}.$$

Then, according to our knowledge in calculus, the volume element for spherical coordinates is changing after the formula

$$dudvdw = r^2\sin x^2 dr dx^2 dx^1$$

and the integral becomes

$$\Phi(r,x^1,x^2) = -G\int_{D^*}\rho\frac{1}{r}r^2\sin x^2 dr dx^2 dx^1 = -G\int_{D^*}\rho r\sin x^2 dr dx^2 dx^1,$$

where D^* is the transformed of D with respect to the previous change of coordinates. The last integral is not singular, therefore the definition of the gravitational potential makes sense in D, too.

If we apply the Laplace operator

$$\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$$

to

$$\Phi(x,y,z) = -G\int_D \rho(u,v,w)\frac{1}{r(x,y,z)}d^3u,$$

we obtain

$$\nabla^2 \Phi(x, y, z) = \int_D \nabla^2 \left(-\frac{G\rho(u, v, w)}{r(x, y, z)} \right) d^3u.$$

The gradient operator $\nabla := \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right)$ leads to the Laplace operator via a formal dot product:

$$\nabla^2 := \nabla \cdot \nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

It implies

$$\nabla^2 \Phi = \nabla \cdot \nabla \Phi = -\nabla \cdot \vec{A} = -\left(\frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} \right) = -div\, \vec{A},$$

that is,

$$\nabla^2 \Phi(x, y, z) = -div\, \vec{A}\, (x, y, z).$$

Now, if $(x, y, z) \notin D$, we have proved $\nabla^2 \left(-\frac{G\rho}{r} \right) = 0$, therefore $\nabla^2 \Phi(x, y, z) = 0$. In the same time we have proved that $div\, \vec{A}\, (x, y, z) = 0$ when $(x, y, z) \notin D$.

If $(x, y, z) \in D$, let us make some considerations.

We define the gravitational field $\vec{A} = (A_x, A_y, A_z)$ attached to the potential Φ, $\vec{A} := -\nabla \Phi$. It remains to evaluate $-div\, \vec{A}\, (x, y, z)$ when $(x, y, z) \in D$. To do this, we consider a sphere $S(r)$ centred at (x, y, z) with a small radius r such that the mass-density ρ can be considered constant in all its interior, interior here denoted by $B(r)$. Therefore we suppose $\rho(u, v, w) = \rho(x, y, z)$ for all $(u, v, w) \in B(r)$. Let us decompose D in $B(r) \cup (D - B(r))$. We have

$$\vec{A} = \vec{A}_{B(r)} + \vec{A}_{D-B(r)}$$

and, since $(x, y, z) \notin D - B(r)$, using the previous case result, it follows

$$div\, \vec{A}_{D-B(r)}\, (x, y, z) = 0,$$

i.e.

$$div\, \vec{A}\, (x, y, z) = div\, \vec{A}_{B(r)}\, (x, y, z) + div\, \vec{A}_{D-B(r)}\, (x, y, z) = div\, \vec{A}_{B(r)}\, (x, y, z).$$

Now, the problem reduces to the evaluation of $div\, \vec{A}_{B(r)}\, (x, y, z)$.

Let us observe that the gravitational field $\vec{A}_{B(r)}$ at every $(\bar{x}, \bar{y}, \bar{z}) \in B(r)$ is

$$\overrightarrow{A}_{B(r)}(\bar{x}, \bar{y}, \bar{z}) = -\frac{G \cdot M_{B(r)}}{\bar{r}^2}\overrightarrow{n} = -\frac{G \cdot \rho \cdot vol B(r)}{\bar{r}^2}\overrightarrow{n},$$

where $\bar{r}$ is the length of the vector who points from (x, y, z) to $(\bar{x}, \bar{y}, \bar{z})$ and $\overrightarrow{n}$ is its unit vector. On the entire surface of $S(r)$, the gravitational field becomes the constant magnitude vector field

$$\overrightarrow{A}_{B(r)}(x, y, z) = -\frac{G \cdot \rho \cdot vol B(r)}{r^2}\overrightarrow{n}.$$

The total outflow over $B(r)$ is

$$\mathcal{F}(B(r)) = -\frac{G \cdot \rho \cdot vol B(r)}{r^2} \cdot 4\pi r^2 = -4\pi G \cdot \rho \cdot vol B(r).$$

Therefore

$$\lim_{r \to 0} \frac{\mathcal{F}(B(r))}{vol B(r)} = \operatorname{div} \overrightarrow{A}_{B(r)}(x, y, z) = -4\pi G \cdot \rho,$$

that is

$$\nabla^2 \Phi = 4\pi G \cdot \rho.$$

Since the ρ chosen is $\rho = \rho(x, y, z)$ and all computations are done at the point (x, y, z), the proof is complete. The previous equation is also known as the *Poisson equation for gravity.* □

8.5 Tidal Acceleration Equations

We met before the gravitational potential

$$\Phi_b(x, y, z) = -\frac{1}{\bar{r}_b}$$

determined by a source at $(b, 0, 0)$, $b > 0$. The denominator is

$$\bar{r}_b := \sqrt{(x - b)^2 + y^2 + z^2}$$

and the corresponding gravitational field is

$$\overrightarrow{A}_b(x, y, z) = -\nabla \Phi_b(x, y, z) = -\frac{1}{\bar{r}_b^2}\left(\frac{x - b}{\bar{r}_b}, \frac{y}{\bar{r}_b}, \frac{z}{\bar{r}_b}\right).$$

We have observed $\overrightarrow{A}_b(0, 0, 0) = \left(\frac{1}{b^2}, 0, 0\right)$.

Definition 8.5.1 The tidal acceleration $\vec{T}\,(x, y, z)$, generated by the gravitational field $\vec{A_b}\,(x, y, z)$ at $(0, 0, 0)$, is defined by the formula

$$\vec{T}\,(x, y, z) := \vec{A_b}\,(x, y, z) - \vec{A_b}\,(0, 0, 0).$$

We may use a Taylor approximation to compute the tidal acceleration at some points of the axes as follows:

$$\vec{T}\,(a, 0, 0) := \vec{A_b}\,(a, 0, 0) - \vec{A_b}\,(0, 0, 0) \approx a\frac{\partial \vec{A_b}}{\partial x}(0, 0, 0) = \left(\frac{2a}{b^3}, 0, 0\right).$$

In the same way

$$\vec{T}\,(0, a, 0) := \vec{A_b}\,(0, a, 0) - \vec{A_b}\,(0, 0, 0) \approx a\frac{\partial \vec{A_b}}{\partial y}(0, 0, 0) = \left(0, -\frac{a}{b^3}, 0\right)$$

and

$$\vec{T}\,(0, 0, a) := \vec{A_b}\,(0, 0, a) - \vec{A_b}\,(0, 0, 0) \approx a\frac{\partial \vec{A_b}}{\partial z}(0, 0, 0) = \left(0, 0, -\frac{a}{b^3}\right).$$

The effect of translation due to a tidal acceleration is called a *tidal effect*.

We can better see the tidal effect, if we consider slices in Oxy- and Oxz-planes.

We focus on Oxy-plane and let us consider the unit vector $(\cos u, \sin u)$.

If we compute $\vec{T}\,(a\cos u, a\sin u)$, we describe the tidal effect at all points of the circle centred at O having a as radius. Therefore

$$\vec{T}\,(a\cos u, a\sin u) := \vec{A_b}\,(a\cos u, a\sin u) - \vec{A_b}\,(0, 0) \approx$$

$$\approx a\cos u\frac{\partial \vec{A_b}}{\partial x}(0, 0) + a\sin u\frac{\partial \vec{A_b}}{\partial y}(0, 0),$$

the approximation being given by the directional derivative of $\vec{A_b}$ in the direction $(\cos u, \sin u)$. It results

$$\vec{T}\,(a\cos u, a\sin u) := \left(\frac{2a\cos u}{b^3}, -\frac{a\sin u}{b^3}\right).$$

This is the image of the tidal effect around $(0, 0, 0)$ in Oxy-plane. There is a similar image in Oxz-plane. In fact, if you rotate the Oxy-plane around Ox-axis, you have the big picture of the tidal effect at all the points of a sphere surface.

Now, you can imagine the Moon at $(b, 0, 0)$ and the Earth as a sphere centred at $(0, 0, 0)$ having radius a and the oceans tides appear when you rotate the sphere. This is the animated picture of the tidal effect.

The tidal effect appears and it can be studied as previously.

If we wish to highlight the equations of the tidal effect, we need to consider free-falling particles in the gravitational field created by the source which start from the points of a given curve $c(q) = (x(q), y(q), z(q))$ as we presented in the previous chapter. We do not repeat the results obtained there. As we will not repeat the equivalent formula of the tidal effect written with respect to the geometry of a metric as seen in the same previous chapter.

8.6 The Kepler Laws

We intend to obtain the three Kepler laws regarding the motion of planets around the Sun. It is necessary to understand how Newtonian mechanics together with Euclidean geometry describe these laws and, for this reason, let us prepare the geometric framework we need.

An *ellipse* of foci $F_1(f, 0)$ and $F_2(-f, 0),\ f > 0$ is the locus of points P in the Euclidean plane such that $|PF_1| + |PF_2| = 2a$, where a is a positive constant, $a > f$. The equation of the ellipse can be found after we transform the condition $|PF_1| + |PF_2| = 2a$ into the equation

$$\sqrt{(x-f)^2 + y^2} + \sqrt{(x+f)^2 + y^2} = 2a.$$

The result is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

where $b^2 = a^2 - f^2$.

The line F_1F_2 is called the *major axis* and the points where the ellipse cuts the major axis have the coordinates $(a, 0)$ and $(-a, 0)$.

The middle of the interval F_1F_2 is called the centre of the ellipse. In this case, the *centre of ellipse* is the origin $O(0, 0)$.

The minor axis is perpendicular to the major axis at $O(0, 0)$. The minor axis intersects the ellipse at the points $(0, b)$ and $(0, -b)$.

The *eccentricity of the ellipse* is, by definition, $e := \dfrac{f}{a} = \dfrac{\sqrt{a^2 - b^2}}{a} = \sqrt{1 - \dfrac{b^2}{a^2}}$.

The area enclosed by the previous ellipse can be computed using the function $y(x) = b\sqrt{1 - \dfrac{x^2}{a^2}}$ which describes the arc of the ellipse $\{(x, y),\ x \in (-a, a), y > 0\}$. If we use the change of variable $x = a \sin t$ the enclosed area is

$$\mathbb{A} = 2\int_{-a}^{a} y(x)dx = 2\frac{b}{a}\int_{-a}^{a}\sqrt{a^2 - x^2}dx = \pi ab.$$

If the ellipse has its centre at (x_0, y_0) and the axes parallel to the axes of the system, i.e. the foci are $(x_0 + f, y_0)$ and $(x_0 - f, y_0)$, the equation is

$$\frac{(x - x_0)^2}{a^2} + \frac{(y - y_0)^2}{b^2} = 1.$$

In fact, the previous ellipse is parallel shifted with respect to the axis such that the old centre $O(0, 0)$ becomes $O_1(x_0, y_0)$.

Consider an ellipse of eccentricity $0 < e < 1$ with a focus at $O(0, 0)$. Its major axis intersects the ellipse at the points $V\left(\frac{k}{1+e}, 0\right)$, $k > 0$, and $V'\left(-\frac{k}{1-e}, 0\right)$. The length of the major semi-axis is $a = \frac{k}{1 - e^2}$, the centre of the ellipse is $\left(-\frac{ke}{1 - e^2}, 0\right)$, and the length of the minor semi-axis is $b = \frac{k}{\sqrt{1 - e^2}}$. The equation of this ellipse is

$$\frac{\left(x + \frac{ke}{1 - e^2}\right)^2}{\frac{k^2}{(1 - e^2)^2}} + \frac{y^2}{\frac{k^2}{1 - e^2}} = 1.$$

Problem 8.6.1 Find the locus of points $M(x, y)$ such that

$$r = r(\theta) = \frac{k}{1 + e\cos\theta},$$

where $r = |OM| = \sqrt{x^2 + y^2}$ and θ is the counterclockwise angle $\angle VOM,\ V \in Ox$.

Hint. The geometric meaning of $r + er\cos\theta = k,\ \ k > 0$, leads to the equation $\sqrt{x^2 + y^2} + ex = k$, i.e. $\sqrt{x^2 + y^2} = k - ex$. If $e = 1$, we obtain a parabola. If $e \neq 1$, after squaring, the previous equation can be written in the form

$$\frac{\left(x + \frac{ke}{1 - e^2}\right)^2}{\frac{k^2}{(1 - e^2)^2}} + \frac{y^2}{\frac{k^2}{1 - e^2}} = 1.$$

Let us observe that, for $0 < e < 1$, we have an ellipse equation. For $e > 1$, the equation is

$$\frac{(x-x_0)^2}{a^2} - \frac{(y-y_0)^2}{b^2} = 1,$$

i.e. we deal with a hyperbola. □

We are ready to study the motion of planets under the action of gravitational force.

Consider the position of the Sun as $O(0, 0, 0)$. The motion of the Earth around the Sun depends on time, i.e. the position of the Earth is given by the vector $\vec{X}(t) = (x(t), y(t), z(t))$. Denote the length of this vector by

$$r(t) = \sqrt{x^2(t) + y^2(t) + z^2(t)}.$$

The Earth is attracted by the Sun via the gravitational force

$$\vec{F}(t) = -\frac{GmM}{r^3(t)}\vec{X}(t),$$

where M is the mass of the Sun, m is the mass of the Earth, and G is the gravitational constant. The equation of motion of the Earth around the Sun, established by the Newton's second law, is

$$m\ddot{\vec{X}}(t) = -\frac{GmM}{r^3(t)}\vec{X}(t),$$

which can be written as

$$\ddot{\vec{X}}(t) = -\frac{GM}{r^3(t)}\vec{X}(t),$$

due to the validity of Galileo's equivalence principle. Let us denote $\mu = GM$ and $\vec{V} = \dot{\vec{X}}$.

Theorem 8.6.2 *The motion of the Earth is planar, that is, the entire trajectory is included in a plane which contains the Sun.*

Proof If we consider the derivative of the cross product between $\vec{X}(t)$ and $\vec{V}(t)$, successively we have

$$\frac{d}{dt}\left(\vec{X} \times \vec{V}\right) = \dot{\vec{X}} \times \vec{V} + \vec{X} \times \dot{\vec{V}} = \vec{V} \times \vec{V} + \vec{X} \times \ddot{\vec{X}} = \vec{0},$$

that is, $\vec{X}(t) \times \vec{V}(t) = \vec{J}$, where $\vec{J}$ does not depend on t. Therefore, the vector $\vec{J}$ of length J is a constant vector, more precisely, it is the normal vector to the plane in which the motion of the Earth around the Sun happens. □

Let us consider $z = 0$ the equation of the plane of motion, that is, the position of the Earth is given by the vector $\vec{X}(t) = (x(t), y(t), 0)$. In the plane of motion, we consider polar coordinates $x = r\cos\theta$, $y = r\sin\theta$, with $r = r(t) = \sqrt{x^2(t) + y^2(t)}$; $\theta = \theta(t)$. We can prove:

Theorem 8.6.3 *If* $\overrightarrow{X}(t) = (r(t)\cos\theta(t), r(t)\sin\theta(t), 0)$ *and* $\overrightarrow{V}(t) = \dot{\overrightarrow{X}}(t)$ *it results*
(i) $\overrightarrow{J} = (0, 0, r^2\dot{\theta})$
(ii) $r^2\dot{\theta} = J$.

Proof We cancel t to write easier the next computations. Then

$$\overrightarrow{V} = \dot{\overrightarrow{X}} = (\dot{r}\cos\theta - r\dot{\theta}\sin\theta,\ \dot{r}\sin\theta + r\dot{\theta}\cos\theta, 0)$$

and

$$\overrightarrow{J} = \overrightarrow{X} \times \overrightarrow{V} = \left(0, 0, r^2\dot{\theta}\right).$$

Since $\overrightarrow{J}$ is a constant vector, the last component does not depend on time, therefore it is a positive constant equal to its length J. So, both assertions are proved. □

Theorem 8.6.4 *The equation of motion for* $\overrightarrow{X}(t)$ *is transformed into the equation*

$$r(t)\ddot{r}(t) = \frac{J^2}{r^2(t)} - \frac{\mu}{r(t)}.$$

Proof We started from the equation of motion

$$\ddot{\overrightarrow{X}}(t) = -\frac{\mu}{r^3(t)}\overrightarrow{X}(t)$$

and, using it, we obtained that the motion is planar. In the plane of motion, the polar coordinates allow us to describe the normal vector $\overrightarrow{J}$ and to obtain the relation $r^2\dot{\theta} = J$.

The derivative with respect to t of the relation $r^2 = \left\langle \overrightarrow{X}, \overrightarrow{X} \right\rangle$ leads to $r\dot{r} = \left\langle \overrightarrow{X}, \overrightarrow{V} \right\rangle$. Then, we have

$$(\dot{r})^2 + r\ddot{r} = \left\langle \overrightarrow{V}, \overrightarrow{V} \right\rangle + \left\langle \overrightarrow{X}, \dot{\overrightarrow{V}} \right\rangle,$$

i.e.

$$(\dot{r})^2 + r\ddot{r} = \left\langle \overrightarrow{V}, \overrightarrow{V} \right\rangle + \left\langle \overrightarrow{X}, -\frac{\mu}{r^3}\overrightarrow{X} \right\rangle,$$

that is,

$$(\dot{r})^2 + r\ddot{r} = |\overrightarrow{V}|^2 - \frac{\mu}{r}.$$

To compute $|\overrightarrow{V}|^2$, we start from the identity

$$\left\langle \overrightarrow{X}, \overrightarrow{V} \right\rangle^2 + |\overrightarrow{X} \times \overrightarrow{V}|^2 = |\overrightarrow{X}|^2 |\overrightarrow{V}|^2.$$

If we replace, in the previous identity, $\left\langle \vec{X}, \vec{V} \right\rangle$ with $r\dot{r}$, $| \vec{X} \times \vec{V} |^2$ with J^2, i.e. $(r^2\dot{\theta})^2$, and $| \vec{X} |^2$ with r^2, it results

$$(r\dot{r})^2 + (r^2\dot{\theta})^2 = r^2 | \vec{V} |^2,$$

thus

$$(\dot{r})^2 + r^2(\dot{\theta})^2 = | \vec{V} |^2.$$

Using $\dot{\theta} = \dfrac{J}{r^2}$, we obtain

$$| \vec{V} |^2 = (\dot{r})^2 + \frac{J^2}{r^2}.$$

It results

$$\cancel{(\dot{r})^2} + r\ddot{r} = \cancel{(\dot{r})^2} + \frac{J^2}{r^2} - \frac{\mu}{r},$$

which complete the proof. □

Theorem 8.6.5 *If $r = \dfrac{1}{u}$ and $u = u(\theta)$, the equation*

$$r\ddot{r} = \frac{J^2}{r^2} - \frac{\mu}{r}$$

becomes

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2}.$$

Proof We first show that $\dot{r} = -J\dfrac{du}{d\theta}$. To obtain this, let us observe that, successively, we have

$$\dot{r} = -\frac{\dot{u}}{u^2} = -\frac{1}{u^2}\frac{du}{dt} = -\frac{1}{u^2}\frac{du}{d\theta}\frac{d\theta}{dt} = -r^2\frac{du}{d\theta}\dot{\theta} = -J\frac{du}{d\theta}.$$

Then, $\ddot{r} = -J\dfrac{d^2u}{d\theta^2}\dot{\theta}$, i.e.

$$\ddot{r} = -J^2\frac{1}{r^2}\frac{d^2u}{d\theta^2}.$$

Taking into account $r = \dfrac{1}{u}$ and replacing into

$$r\ddot{r} = \frac{J^2}{r^2} - \frac{\mu}{r},$$

we obtain the desired equation

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2}.$$

□

The general solution is $u(\theta) = A\cos(\theta - \theta_0) + \frac{\mu}{J^2}$, where A is an arbitrary constant and θ_0 is the initial value, called phase, which leads to the starting point of the trajectory.

If we are interested only in the shape of the solution, we may consider

$$u(\theta) = A\cos\theta + \frac{\mu}{J^2}.$$

The solution in r is

$$r(\theta) = \frac{\frac{J^2}{\mu}}{1 + \frac{J^2 A}{\mu}\cos\theta}.$$

If A is in such a way that

$$0 < e := \frac{J^2 A}{\mu} < 1$$

the trajectory is an ellipse. Therefore we have proved.

Theorem 8.6.6 (Kepler's first law) *In the case of the pair {Sun, Earth}, the gravity makes Earth to move around the Sun after an elliptical orbit having the Sun as one of the foci.*

This is the Kepler first law. It generally describes how a planet moves around a star.

Let us see again the big picture of the motion of the Earth around Sun. We have the Sun at the origin of the coordinate system and the Earth position given by the vector $\vec{X}(t) = (x(t), y(t), z(t))$. The gravitational force acting between the two bodies leads to the equation of motion

$$\ddot{\vec{X}}(t) = -\frac{GM}{r^3(t)}\vec{X}(t).$$

The motion is planar. Using polar coordinates, we can transform the equation into the new equation

$$r\ddot{r} = \frac{J^2}{r^2} - \frac{\mu}{r},$$

and finally into the equation

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2}$$

which can be solved. The solution, in polar coordinates, is the ellipse

$$r(\theta) = \frac{\dfrac{J^2}{\mu}}{1 + \dfrac{J^2 A}{\mu}\cos\theta}.$$

Defining $k := \dfrac{J^2}{\mu}$ and $e := \dfrac{J^2 A}{\mu}$, in Cartesian coordinates, the equation is

$$\frac{\left(x + \dfrac{ke}{1-e^2}\right)^2}{\dfrac{k^2}{(1-e^2)^2}} + \frac{y^2}{\dfrac{k^2}{1-e^2}} = 1,$$

where the semi-axes are $a := \dfrac{k}{(1-e^2)}$ and $b := \dfrac{k}{\sqrt{(1-e^2)}}$. The *perihelion* of the trajectory, that is, the closest position to the Sun, is at the point $V\left(\dfrac{k}{1+e}, 0\right)$. The *aphelion*, that is, the furthest position from the Sun, is located at the point $V'\left(-\dfrac{k}{1-e}, 0\right)$.

If we look at comets, the trajectories can be elliptic, hyperbolic, and parabolic. The case $e = 1$ is a possible case, but it is difficult for an astronomer to say that a comet has a parabolic orbit. It is more probable to have a hyperbolic orbit with $e > 1$ but very close to 1. We prefer to remain at the case {planet, Sun} where the trajectories are always ellipses. Now we are able to prove the Kepler second law.

Theorem 8.6.7 (Kepler's second law) *Areas swept out by $\vec{OX}$ in equal time intervals are equal.*

Proof Consider two close positions of $\vec{OX}$, that is $\vec{OX'}$ and $\vec{OX''}$. The angle between these two positions is $d\theta$. The infinitesimal area swept by $\vec{OX}$ is $dA = \frac{1}{2}r^2 d\theta$. It results $\dfrac{dA}{dt} = \dfrac{1}{2}r^2\dot{\theta}$, i.e.

$$\frac{dA}{dt} = \frac{1}{2}J,$$

which ends the proof. □

Let us continue with the Kepler third law. The time necessary to have a complete revolution around the Sun is called the *orbital period of a planet.* It is denoted by T.

Theorem 8.6.8 (Kepler's third law) *The ratio between the square of the orbital period and the cube of the major semi-axis is a constant, that is,* $\dfrac{T^2}{a^3} = \dfrac{4\pi^2}{GM}$.

Proof Let us observe that $\overrightarrow{OX}$ sweeps the area of the ellipse during a revolution. Thus

$$\pi ab = T\frac{1}{2}r^2\dot{\theta}.$$

It results $\dfrac{T}{a} = 2\pi\dfrac{b}{J}$, that is, $\dfrac{T^2}{a^2} = 4\pi^2\dfrac{b^2}{J^2}$.

According to previous formulas for semi-axes we have $b^2 = \dfrac{k^2}{1-e^2} = k\dfrac{k}{1-e^2} = ka$, therefore

$$\frac{T^2}{a^2} = 4\pi^2\frac{ka}{J^2}.$$

Taking into account that $k := \dfrac{J^2}{\mu}$ we finally obtain

$$\frac{T^2}{a^3} = 4\pi^2\frac{1}{\mu}.$$

□

The third law is called the *Harmony law* because, if we consider two different planets moving around the Sun, the same constant is the ratio between $\dfrac{T_1^2}{a_1^3}$ and $\dfrac{T_2^2}{a_2^3}$.

8.7 Circular Motion, Centripetal Force, Deflection of Light Effect, and Dark Matter Problem

Before continuing, let us discuss a little bit about circular motion and observe the differences with respect to the elliptical motion presented above. Circular motion means a movement of an object along the circumference of a circle. A boy rotating a tide up ball with a chord, a car moving at constant speed on a circular track, or even a satellite on its orbit around the Earth can be mathematically modelled as circular motions. So, the trajectory is a circle of radius R, the object in circular motion can be imagined as a point (with a mass, say m) moving at constant speed v. The speed vector $\overrightarrow{v}$ is tangent at each point of the circle. To maintain the point on this trajectory, the force vector (that is the acceleration vector, too) has to be imagined as an arrow oriented from the point to the centre. Of course, the magnitude of the force has to be the same for all the possible positions, because there are no differences between these vectors except the possible directions. This force is called a *centripetal force*. The corresponding acceleration is called *centripetal acceleration*.

Let us consider two tangent vectors corresponding to two close points on the circumference separated by a $d\theta$ angle. Denote by dx the length of the arc determined by the two points and observe that between the two tangent vectors there is the same angle $d\theta$. We have $dx = Rd\theta$ and $v = \frac{dx}{dt}$. If dv is the vector which connects their ends, we may approximate $d\theta = \frac{dv}{v}$. It results

$$dx = Rd\theta = R\frac{dv}{v} = vdt,$$

that is,

$$a := \frac{dv}{dt} = \frac{v^2}{R}.$$

This is the formula of the centripetal acceleration which allows to write the formula of the centripetal force:

$$F_c := m\frac{v^2}{R}.$$

How it can be imagined the rotation of the Earth around the Sun using this force? The mathematical answer is

$$F_c = \frac{mv^2}{R} = \frac{GMm}{R^2} = F,$$

i.e.

$$v^2 = \frac{GM}{R},$$

thanks to the equivalence principle by which m can be simplified in both sides of the equation. This is important because if the radius R is increasing the orbital speed has to decrease.

Now, since the period of revolution around the circular trajectory is $T = \frac{2\pi R}{v}$, we obtain

$$T^2 = 4\pi^2 R^2 \frac{1}{v^2} = \frac{4\pi^2 R^3}{GM},$$

that is, T^2 is proportional to R^3, or

$$\frac{T^2}{R^3} = \frac{4\pi^2}{GM}.$$

It is a sort of approximation of the third Kepler law.

Let us imagine our Sun or a spherical galaxy and a plane section through their centre. Therefore we can imagine both sections as discs. If R is the radius of such a disc it remains only to imagine a ray of light passing through the vicinity of the disc. It makes sense to consider the time Δt when the ray of light is affected by gravity to be

$$\Delta t = \frac{2R}{c},$$

where c is the speed of light. The gravitational force F which acts to bend the ray of light has the gravitational acceleration described by the formula

$$a = \frac{GM}{R^2},$$

where M is the mass of, let us say, the Sun. Now, denote by Δv the variation of the speed of light when it is affected by Sun's gravity. We have

$$\Delta v = a\Delta t = \frac{GM}{R^2} \cdot \frac{2R}{c} = \frac{2GM}{Rc}.$$

A right triangle of speeds can be imagined. The deflected ray is the hypotenuse, the other two sides are "c" and "Δv" and the acute angle of deflection is α, such that

$$\tan\alpha = \frac{\Delta v}{c} = \frac{2GM}{Rc^2}.$$

Taking into consideration that α is small, $\tan\alpha \approx \alpha$, therefore

$$\alpha = \frac{2GM}{Rc^2}.$$

This value is half the value we measure in practice. Therefore the real value of deflection is

$$\alpha = \frac{4GM}{Rc^2}$$

and will be computed using a metric in the chapter dedicated to general relativity.

The centripetal force is often used in approximations of trajectories in astronomy. An interesting application of the centripetal force is the possible existence of *dark matter* or, according to Fritz Zwicky, the *missing matter* [200]. The formula

$$v^2 = \frac{GM}{R},$$

which asserts that if R is increasing, the speed v decreases (if M remains constant), is crucial.

In a galaxy, there are billions of stars. We may think that these stars are in an imaginary sphere having as a centre, the centre of the galaxy. Some stars are closer to the centre of the galaxy, some of them are far. Some other stars are out of the edge of the galaxy, or, more precisely, they are in the area where, if we increase the radius of the galaxy, we add few stars. For stars in the zone with a lot of stars, if we increase the radius we have more stars, i.e. more mass. Here, the fact that the observed speed of stars rotating around the centre is the same it is not a problem. The speed v can be kept constant, if the mass M increases when R increases. But for distant stars, when we increase the radius, we do not add more mass inside. However the measured speed v is the same and it is more or less constant also very far from the galactic centre (more than 10 kiloparsec). According to this situation, we have to suppose the existence of a sort of (sub-luminous) matter that cannot be detected by the standard electromagnetic emission. However, the amount of such a matter increases with the increase of the distance from the centre. The problem is known as the dark matter problem and can be solved in two alternative ways: Either one suppose the existence of exotic matter interacting only gravitationally, or one assumes deviation from the Kepler laws at large distances. More details can be found in [56, 141, 142].

Later in the book, we will study the trajectory of planets in a given metric. Specifically, we will study the trajectory of planets both in the Schwarzschild metric

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)dt^2 - \frac{1}{1 - \frac{2GM}{c^2 r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2$$

and in the Einstein metric

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)dt^2 - \left(1 + \frac{2GM}{c^2 r}\right)dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2.$$

The planet equation of motion in both metric is

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2}u^2,$$

where c is the constant speed of light in vacuum and $\mu = GM$ as previously defined.

Why we study the equation of motion in a metric and how close is the solution of this new equation to the above classical solution? These topics will be discussed in the General Relativity chapter of this book.

8.8 The Mechanical Lagrangian

In a system of coordinates (t, x), let $(t, x(t))$ be the trajectory of a particle of mass m moving under the influence of a force derived from a time-independent *potential* V. Since V depends only on the position, we denote this by $V := V(x)$.

Newton's equation of motion is

$$m\ddot{x}(t) = F(x),$$

where the force acting on the particle is $F(x) = -\dfrac{dV}{dx}$.

Given some initial conditions, the trajectory $(t, x(t))$ is comprised between the initial point $(t_1, x(t_1))$ and the final point $(t_2, x(t_2))$.

Let us underline that this trajectory is the expression of the force acting on the particle under some initial conditions. Therefore, there is an unique trajectory determined by the force and the initial conditions.

Now let us consider all the paths connecting $(t_1, x(t_1))$ and $(t_2, x(t_2))$. They can be thought as $y(t) + \eta(t)$, with $y(t_1) = x(t_1),\ y(t_2) = x(t_2),\ \eta(t_1) = \eta(t_2) = 0$.

Having all these paths, what new theory do we need to imagine in order to discover the original path described by Newton's equation of motion?

To answer this question, we need some technical details (see also [102]).

Let us insist on this first part when we have described what we want to do. We have used V such that $F = -\dfrac{dV}{dx}$. We defined V as an independent potential and we suggested its connection with the force F, $dV = -Fdx$.

Is this definition connected to the facts seen in our previous sections when we have studied the gravitational force and the gravitational potential? The answer is yes, but we need to point out a major difference between this V and the gravitational potential Φ.

Consider a body of mass M at the origin O of a line whose current coordinate is denoted by x. Suppose that at point $N(x)$, a body of mass m exists. The gravitational force in this case has the intensity $F = \dfrac{GMm}{x^2}$. The work done by the body of mass M to move the body of mass m from x to $x - dx$ is $-Fdx$. There is an energy transferred to do this work. Its variation ΔE is $-Fdx$. By definition, the *gravitational potential energy* $P_E(r)$ is related to the work done to move the body of mass m from the infinity to the point having coordinate r, that is,

$$P_E(r) = \int_{\infty}^{r} Fdx = \int_{\infty}^{r} \frac{GMm}{x^2}dx = -\frac{GMm}{r}.$$

The potential energy can be denoted by P_E. If one looks at the formula obtained and takes into account the formula of the gravitational potential $-\dfrac{GM}{r}$, we can understand both the explanations above and the relation

$$P_E = m\Phi.$$

Therefore, another definition for the gravitational potential appears: the work (energy transferred) per unit mass necessary to move a body from infinity to the point having the coordinate r. Indeed,

$$\Phi(r) = \frac{1}{m}\int_{\infty}^{r} F dx = \frac{1}{m}\int_{\infty}^{r} \frac{GMm}{x^2} dx = -\frac{GM}{r}.$$

In the case when we consider the constant gravitational field determined by the constant acceleration g between the origin O and a point H at the coordinate h, the potential energy is expressed by the formula $P_E = mgh$. The explanation is related to the difference of formal integrals

$$P_E := P_E(h) - P_E(0) = \int_{\infty}^{h} gmdx - \int_{\infty}^{0} gmdx = \int_{0}^{h} gmdx = gmh$$

which describes the amount of energy necessary to move the body at h to 0.

In the same way, we can define the kinetic energy. Let us start from $F = ma = m\frac{dv}{dt}$ written in its discrete form, $F = m\frac{\Delta v}{\Delta t}$. If we multiply by Δr, we obtain

$$F\Delta r = m\frac{\Delta v}{\Delta t}\Delta r = m\frac{\Delta r}{\Delta t}\Delta v = mv\Delta v,$$

which can be written in the differential way as

$$Fdr = mvdv.$$

Now, the amount of energy necessary to bring a body initially at rest to the speed v is

$$T(v) = \int_{0}^{v} F dx = \int_{o}^{v} mxdx = m\frac{v^2}{2}.$$

Since v can be seen as $\dot{x}(t)$, we may consider the *kinetic energy of the mechanical system* defined by the formula $T = T(\dot{x}) := \frac{1}{2}m(\dot{x}(t))^2$. Another possible notation is K_E. Here, with *mechanical system* we intend a system of elements that interact on mechanical principles. A material point and a force which acts on it is a possible example. Two material points which interact through the gravitational force offer another example. In this perspective, the next exercise has important consequences in Newtonian mechanics.

Exercise 8.8.1 Consider a mechanical system whose kinetic energy is $T(\dot{x}) := \frac{1}{2}m(\dot{x}(t))^2$ and its potential energy is V (such that the force which acts is $F(x) = -\frac{dV}{dx}$). Show that the total energy of the system, $T + V$, is a constant.

Hint. If we derive with respect to t the total energy, we obtain

$$\frac{d}{dt}(T+V) = \left(m\dot{x}(t)\ddot{x}(t) + \frac{dV}{dx}\frac{dx}{dt}\right) = (m\ddot{x}(t) - F)\,\dot{x}(t) = 0,$$

that is, $T + V$ is a constant.

We define the *mechanical Lagrangian* of the system by

$$L = L(x, \dot{x}) := T - V = \frac{1}{2}m(\dot{x}(t))^2 - V(x).$$

In this section, where there is no possibility of confusion, we simply use the definition "Lagrangian" instead of mechanical Lagrangian. Later in the book, we will see that there exist general Lagrangians which come from Geometry, therefore we have to well understand the nature of the Lagrangian we are considering.

Let us observe that, even if x and $\dot{x}$ depends on t, this Lagrangian is only implicitly a function of time.

In this formalism, it makes sense to consider a functional called *action*,

$$S[y] = \int_{t_1}^{t_2}\left[\frac{1}{2}m(\dot{y}(t))^2 - V(y)\right]dt$$

which exists for any path $y(t)$, not only for the "physical right on" which is $x(t)$.

Now consider the action corresponding to $y(t) + \eta(t)$,

$$S[y+\eta] = \int_{t_1}^{t_2}\left[\frac{1}{2}m(\dot{y}(t) + \dot{\eta}(t))^2 - V(y(t) + \eta(t))\right]dt\,.$$

We have, after expanding V in Taylor series with respect to $y(t)$,

$$S[y+\eta] = S[y] + \int_{t_1}^{t_2}\left[m\dot{y}(t)\dot{\eta}(t) - \frac{dV}{dy}(y(t))\eta(t)\right]dt + \mathbb{O}(\eta^2),$$

where $\mathbb{O}(\eta^2)$ are terms of order $\eta^2 := \eta^2(t)$ or higher. We can write

$$S[y+\eta] = S[y] + \delta S + \mathbb{O}(\eta^2),$$

where

$$\delta S = \int_{t_1}^{t_2}\left[m\dot{y}(t)\dot{\eta}(t) - \frac{dV}{dy}(y(t))\eta(t)\right]dt$$

is called the *first-order variation of the action* S. Since $\eta(t_1) = \eta(t_2) = 0$, we obtain

$$\delta S = \int_{t_1}^{t_2} \left[m\dot{y}(t)\dot{\eta}(t) - \frac{dV}{dy}(y(t)\eta(t)) \right] dt =$$

$$= \int_{t_1}^{t_2} \left[m\frac{d(\dot{y}(t)\eta(t))}{dt} - m\ddot{y}(t)\eta(t) - \frac{dV}{dy}(y(t)\eta(t)) \right] dt =$$

$$= m\dot{y}(t_2)\eta(t_2) - m\dot{y}(t_1)\eta(t_1) - \int_{t_1}^{t_2} \left[m\ddot{y}(t) + \frac{dV}{dy}(y(t)) \right] \eta(t)dt =$$

$$= -\int_{t_1}^{t_2} \left[m\ddot{y}(t) + \frac{dV}{dy}(y(t)) \right] \eta(t)dt.$$

Therefore, $\delta S \equiv 0$ means

$$\int_{t_1}^{t_2} \left[m\ddot{y}(t) + \frac{dV}{dy}(y(t)) \right] \eta(t)dt = 0$$

for every η, and it happens if and only if $m\ddot{y}(t) + \dfrac{dV}{dy}(y(t)) = 0$, i.e. for $y(t) = x(t)$. We have proved:

Theorem 8.8.2 *The first-order variation of the action S vanishes, i.e.*

$$\delta S = \int_{t_1}^{t_2} \left[m\dot{y}(t)\dot{\eta}(t) - \frac{dV}{dy}(y(t))\eta(t) \right] dt = 0,$$

if and only if $y(t)$ satisfies Newton's equation of motion

$$m\ddot{x}(t) - F(x) = 0.$$

So, the answer is: *The "physical right path" happens when the first-order variation δS vanishes*. Therefore the right path is described by the condition $\delta S \equiv 0$. This is known as *Hamilton's stationary action principle*.

8.9 Geometry Induced by a Lagrangian

Now, let us consider another problem.

Can we find an equation, satisfied by a general function $L(x, \dot{x})$, not only by the mechanical Lagrangian $L = T - V$ as before, such that the functionx $= x(t)$, which connects the given points $(t_1, x(t_1))$; $(t_2, x(t_2))$ where the functional

$$S[x] = \int_{t_1}^{t_2} L(x(t), \dot{x}(t))dt\,.$$

is extremized?

Let us explain first what is the mathematical meaning of the words "extremizes the functional S". Consider all the perturbation of $x(t)$, say

$$y_\lambda(t) = x(t) + \lambda\eta(t),\ \ \lambda \in \mathbb{R}$$

which preserves the endpoints $(t_1, x(t_1))$; $(t_2, x(t_2))$, that is, $\eta(t_1) = \eta(t_2) = 0$ and construct the action

$$S_\lambda[y_\lambda] = \int_{t_1}^{t_2} L_\lambda(y_\lambda(t), \dot{y}_\lambda(t))dt = \int_{t_1}^{t_2} L_\lambda(x(t) + \lambda\eta(t), \dot{x}(t) + \lambda\dot{\eta}(t))dt.$$

Extremizing the functional $S[x]$ means or $S_\lambda[y_\lambda] \geq S[x]$ for any $\lambda \in \mathbb{R}$ or $S_\lambda[y_\lambda] \leq S[x]$ for any $\lambda \in \mathbb{R}$, where the equality works if and only if $\lambda = 0$.

Therefore, *extremizing the functional* $S[x]$ *implies the condition* $\left.\dfrac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0.$

Since

$$\frac{dL_\lambda}{d\lambda} = \frac{\partial L_\lambda}{\partial y_\lambda}\frac{\partial y_\lambda}{\partial \lambda} + \frac{\partial L_\lambda}{\partial \dot{y}_\lambda}\frac{\partial \dot{y}_\lambda}{\partial \lambda} = \frac{\partial L_\lambda}{\partial y_\lambda}\eta(t) + \frac{\partial L_\lambda}{\partial \dot{y}_\lambda}\dot{\eta}(t)\,,$$

it results

$$\left.\frac{dL_\lambda}{d\lambda}\right|_{\lambda=0} = \frac{\partial L}{\partial x}\eta(t) + \frac{\partial L}{\partial \dot{x}}\dot{\eta}(t),$$

therefore the condition $\left.\dfrac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0$ is written as

$$\left.\frac{dS_\lambda}{d\lambda}\right|_{\lambda=0} = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x}\eta(t) + \frac{\partial L}{\partial \dot{x}}\dot{\eta}(t)\right] dt \equiv 0.$$

Definition 8.9.1 The curve $x = x(t)$ which extremizes the functional

$$S[x] = \int_{t_1}^{t_2} L(x(t), \dot{x}(t))dt$$

is called a stationary point of the functional $S[x]$.

Theorem 8.9.2 (Euler-Lagrange equation) *The curve* $x = x(t)$ *which connects the given points* $(t_1, x(t_1))$, $(t_2, x(t_2))$ *satisfies the Euler–Lagrange equation*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0$$

if and only if it is a stationary point of the functional

$$S[x] = \int_{t_1}^{t_2} L(x(t), \dot{x}(t))dt.$$

Proof Using the integration by parts

$$\int_{t_1}^{t_2} \frac{\partial L}{\partial x}\eta(t)dt + \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{x}}\dot{\eta}(t)dt =$$

$$= \int_{t_1}^{t_2} \frac{\partial L}{\partial x}\eta(t)dt + \frac{\partial L}{\partial \dot{x}}\eta(t_2) - \frac{\partial L}{\partial \dot{x}}\eta(t_1) - \int_{t_1}^{t_2} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right)\eta(t)dt =$$

$$= \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right)\right]\eta(t)dt.$$

The condition $\left.\frac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0$ means

$$\int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right)\right]\eta(t)dt = 0\,,$$

for every function η. We obtain

$$\frac{\partial L}{\partial x} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) = 0.$$

□

Another proof can be considered for the Euler–Lagrange equation. As previously, let us consider the action

$$S[y] = \int_{t_1}^{t_2} L(y(t), \dot{y}(t))dt\,.$$

Now consider the action corresponding to $y(t) + \eta(t)$,

$$S[y+\eta] = \int_{t_1}^{t_2} L(y(t)+\eta(t), \dot{y}(t)+\dot{\eta}(t))dt,$$

where $\eta(t_1) = \eta(t_2) = 0$. After expanding L in Taylor series with respect to the variables y and $\dot{y}$ we obtain

$$L(y(t)+\eta(t), \dot{y}(t)+\dot{\eta}(t)) = L(y(t), \dot{y}(t)) + \frac{\partial L}{\partial y}\eta + \frac{\partial L}{\partial \dot{y}}\dot{\eta} + \mathbb{O}(\eta^2) + \mathbb{O}(\dot{\eta}^2).$$

The first-order variation of the action S is

$$\delta S = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} \eta(t) + \frac{\partial L}{\partial \dot{y}} \dot{\eta}(t) \right] dt.$$

Using the integration by parts and the conditions $\eta(t_1) = \eta(t_2) = 0$, it results successively

$$\delta S = \int_{t_1}^{t_2} \frac{\partial L}{\partial y} \eta(t) dt + \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{y}} \dot{\eta}(t) dt =$$

$$= \int_{t_1}^{t_2} \frac{\partial L}{\partial y} \eta(t) dt + \frac{\partial L}{\partial \dot{y}} \eta(t_2) - \frac{\partial L}{\partial \dot{y}} \eta(t_1) - \int_{t_1}^{t_2} \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) \eta(t) dt =$$

$$= \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) \right] \eta(t) dt.$$

The first-order variation of action vanishes if the last integral vanishes, i.e. $\delta S \equiv 0$ iff

$$\int_{t_1}^{t_2} \left[\frac{\partial L}{\partial y} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) \right] \eta(t) dt = 0$$

unction η. This means

$$\frac{\partial L}{\partial y} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) = 0.$$

□

Both proofs reported before hold even if the Lagrangian is $L(t, y(t), \dot{y}(t))$ instead of $L(y(t), \dot{y}(t))$. In the particular case, when the Lagrangian does not depend explicitly on t, the Euler–Lagrange equation reduces to the Beltrami identity. The following theorem holds.

Theorem 8.9.3 (Beltrami's identity) *If the Lagrangian does not depend explicitly on t, then a constant C exists such that*

$$L(y, \dot{y}) - \dot{y} \frac{\partial L(y, \dot{y})}{\partial \dot{y}} = C.$$

Proof The total derivative of $L(t, y(t), \dot{y}(t))$ is

$$\frac{dL}{dt} = \frac{\partial L}{\partial t} + \frac{\partial L}{\partial y} \dot{y} + \frac{\partial L}{\partial \dot{y}} \ddot{y},$$

i.e.

$$\frac{\partial L}{\partial y} \dot{y} = \frac{dL}{dt} - \frac{\partial L}{\partial t} - \frac{\partial L}{\partial \dot{y}} \ddot{y}.$$

If $\dfrac{\partial L}{\partial t} = 0$, the previous equality becomes

$$\frac{\partial L}{\partial y}\dot{y} = \frac{dL}{dt} - \frac{\partial L}{\partial \dot{y}}\ddot{y}.$$

Multiplying the Euler–Lagrange equation by $\dot{y}$, we obtain

$$\dot{y}\frac{\partial L}{\partial y} = \dot{y}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right),$$

therefore, after combining the last two equalities we have

$$\frac{dL}{dt} - \dot{y}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) - \frac{\partial L}{\partial \dot{y}}\ddot{y} = 0,$$

that is

$$\frac{d}{dt}\left(L - \dot{y}\frac{\partial L}{\partial \dot{y}}\right) = 0,$$

which is equivalent to the statement. □

Let us consider now an important problem solved first using the equilibrium of the forces involved, afterwards using the Euler–Lagrange equation. We are talking about the problem of hanging rope.

Problem 8.9.4 The catenary problem: Suppose that a rope is hanged with its ends at the same height above the floor and its mass on the unit length is ρ. Find the function which describes the shape of the rope.

Solution I: Consider a frame of coordinates such that the two given points are $A(-a, b)$, $B(a, b)$ and the shape is described by the points of the curve $(x, y(x))$. The statement conditions allow us to consider a minimum point at $O(0, 0)$, a symmetry with respect to Oy-axis and Ox-axis as a tangent to the curve at O. Consider a point $M(x, y(x))$ on the arc OB and the tangent at M. Let us denote by s the length of the arc OM, that is,

$$s(x) = \int_0^x \sqrt{1 + (\dot{y}(t))^2}dt.$$

From Leibniz integral rule,

$$s(x) = \int_0^x \sqrt{1 + (\dot{y}(t))^2}dt = F(x) - F(0)$$

where $\dfrac{dF}{dt} = \sqrt{1 + (\dot{y}(t))^2}$. Therefore

$$\frac{ds}{dx} = \sqrt{1 + (\dot{y}(x))^2}.$$

There are three forces at equilibrium which act on the given arc. The tension $(-T_0, 0)$ at O, the weight of the arc $(0, -\rho g s)$, where g is the acceleration due to gravity, and the tension of magnitude T at M, $(T \cos\theta, T \sin\theta)$, this one acting along the tangent to $(x, y(x))$ at M. Therefore

$$(-T_0, 0) + (0, -\rho g s) + (T \cos\theta, T \sin\theta) = (0, 0).$$

The equilibrium conditions are

$$\begin{cases} T \cos\theta = T_0 \\ T \sin\theta = \rho g s. \end{cases}$$

It results

$$\dot{y}(x) = \frac{dy}{dx} = \tan\theta = \frac{\rho g}{T_0} s,$$

i.e.

$$\ddot{y}(x) = \frac{\rho g}{T_0} \frac{ds}{dx} = \frac{\rho g}{T_0} \sqrt{1 + (\dot{y}(x))^2}.$$

If we denote $b := \dfrac{\rho g}{T_0}$ and $u = \dot{y}(x)$, it remains to solve the equation

$$\frac{\dot{u}(x)}{\sqrt{1 + (u(x))^2}} = b$$

which leads to

$$\int \frac{du}{\sqrt{1 + u^2}} = b \int dx.$$

Since $u(0) = \dot{y}(0) = 0$, the equality

$$u + \sqrt{1 + u^2} = e^{bx+l}$$

implies $l = 0$ and

$$u(x) = \sinh 2bx,$$

i.e.

$$y(x) = \frac{T_0}{2\rho g} \cosh\left(\frac{2\rho g}{T_0} x\right) - \frac{T_0}{2\rho g}.$$

□

Solution II: The rope has a given length

$$\mathbf{l}_{a,b} = \int_{-a}^{a} \sqrt{1 + (\dot{y}(x))^2} dx \, ,$$

and we can think at a Lagrangian induced by the potential energy of the rope combined with the constraint of finite length for the rope,

$$\mathcal{L} = \rho g y(x) \sqrt{1 + (\dot{y}(x))^2} + \alpha \left(\sqrt{1 + (\dot{y}(x))^2} - \mathbf{l}_{a,b} \right),$$

where α is a constant. Without the length constraint, the potential energy is smaller and smaller while the rope is longer and longer. Finally, we can try to derive the curve starting from the Lagrangian

$$\mathcal{L} = (\rho g y + \alpha) \sqrt{1 + \dot{y}^2} + \beta,$$

where β is a constant. Since $\dfrac{\partial \mathcal{L}}{\partial t} = 0$, we can use Beltrami's identity. Therefore there exists a constant C such that

$$\mathcal{L} - \dot{y} \frac{\partial \mathcal{L}}{\partial \dot{y}} = C$$

which means

$$(\rho g y + \alpha) \sqrt{1 + \dot{y}^2} - \dot{y} (\rho g y + \alpha) \frac{\dot{y}}{\sqrt{1 + \dot{y}^2}} = C,$$

i.e.

$$(\rho g y + \alpha) \frac{1}{\sqrt{1 + \dot{y}^2}} = C.$$

It remains to solve

$$\dot{y}^2 = \frac{(\rho g y + \alpha)^2}{C^2} - 1.$$

The substitution $Cu = \rho g y + \alpha$ leads to

$$\frac{C}{\rho g} \dot{u} = \sqrt{u^2 - 1},$$

i.e.

$$\frac{du}{\sqrt{u^2 - 1}} = \frac{\rho g}{C} dx$$

with the solution $u(x) = \cosh \left(\dfrac{\rho g}{C} x + \gamma \right)$, where γ is a constant. Therefore

$$y(x) = \frac{C}{\rho g} \cosh\left(\frac{\rho g}{C} x + \gamma\right) - \alpha.$$

The constants are determined from the symmetry condition with respect to Oy-axis, that is, $\gamma = 0$, $y(0) = 0$ that is $\alpha = \dfrac{C}{\rho g}$ and C from

$$\mathbf{l}_{a,b} = \int_{-a}^{a} \sqrt{1 + (\dot{y}(x))^2} dx.$$

□

Let us return at the first proof we offered for the Euler–Lagrange equation. That proof can be used to obtain the general Euler–Lagrange equations.

For the Lagrangian $L = L(x^0, x^1, ..., x^n, \dot{x}^0, \dot{x}^1, ..., \dot{x}^n)$, we obtain

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0, \quad k = 0, 1, ..., n.$$

It is easy to see that we have to act as before on each pair of variables $x^k, \dot{x}^k, \ k = 0, 1, ..., n$. We are looking for a system of equations satisfied by the previous Lagrangian, such that a curve $x = x(t) = (x^0(t), x^1(t), .., x^n(t))$, which connects the given points $(t_1, x^0(t_1), x^1(t_1), ..., x^n(t_1))$, $(t_2, x^0(t_2), x^1(t_2), ..., x^n(t_2))$, extremizes the functional

$$S[x] = \int_{t_1}^{t_2} L(x^0(t), \dot{x}^0(t), x^1(t), \dot{x}^1(t), ..., x^n(t), \dot{x}^n(t)) dt.$$

As previously, a perturbation of $x(t)$ which preserves the endpoints is

$$y_\lambda(t) = (y_\lambda^0(t), y_\lambda^1(t), ..., y_\lambda^n(t)) = (x^0(t) + \lambda\eta_0(t), x^1(t) + \lambda\eta_1(t), ..., x^n(t) + \lambda\eta_n(t)),$$

$\lambda \in \mathbb{R}$ with $\eta_k(t_1) = \eta_k(t_2) = 0, \ k = 0, 1, ..., n$. Consider

$$S_\lambda[y_\lambda] = \int_{t_1}^{t_2} L_\lambda(y_\lambda^0(t), \dot{y}_\lambda^0(t), ..., y_\lambda^n(t), \dot{y}_\lambda^n(t)) dt =$$

$$= \int_{t_1}^{t_2} L_\lambda(x^0(t) + \lambda\eta_0(t), \dot{x}^0(t) + \lambda\dot{\eta}_0(t), ..., x^n(t) + \lambda\eta_n(t), \dot{x}^n(t) + \lambda\dot{\eta}_n(t)) dt.$$

Extremizing the functional S implies the condition $\left.\dfrac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0$. Or,

$$\frac{dL_\lambda}{d\lambda} = \sum_{k=0}^{n} \left[\frac{\partial L_\lambda}{\partial y_\lambda^k}\frac{\partial y_\lambda^k}{\partial \lambda} + \frac{\partial L_\lambda}{\partial \dot{y}_\lambda^k}\frac{\partial \dot{y}_\lambda^k}{\partial \lambda}\right] = \sum_{k=0}^{n}\left[\frac{\partial L_\lambda}{\partial y_\lambda^k}\eta_k(t) + \frac{\partial L_\lambda}{\partial \dot{y}_\lambda^k}\dot{\eta}_k(t)\right],$$

therefore

$$\left.\frac{dL_\lambda}{d\lambda}\right|_{\lambda=0} = \sum_{k=0}^{n}\left[\frac{\partial L}{\partial x^k}\eta_k(t) + \frac{\partial L}{\partial \dot{x}^k}\dot{\eta}_k(t)\right].$$

The condition $\left.\frac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0$ becomes

$$\left.\frac{dS_\lambda}{d\lambda}\right|_{\lambda=0} = \sum_{k=0}^{n}\int_{t_1}^{t_2}\left[\frac{\partial L}{\partial x^k}\eta_k(t) + \frac{\partial L}{\partial \dot{x}^k}\dot{\eta}_k(t)\right]dt \equiv 0.$$

Definition 8.9.5 The curve $x = x(t) = (x^0(t), x^1(t), .., x^n(t))$ which extremizes the functional

$$S[x] = \int_{t_1}^{t_2} L(x^0(t), \dot{x}^0(t), x^1(t), \dot{x}^1(t), ..., x^n(t), \dot{x}^n(t))dt$$

is called a stationary point of the functional.

Theorem 8.9.6 (Euler–Lagrange equations) *The curve* $x = x(t) = (x^0(t), x^1(t), .., x^n(t))$ *which connects the given points* $(t_1, x^0(t_1), x^1(t_1), ..., x^n(t_1))$, $(t_2, x^0(t_2), x^1(t_2), ..., x^n(t_2))$ *satisfies the Euler–Lagrange equations*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) - \frac{\partial L}{\partial x^k} = 0,\ k = 0, 1, ..., n$$

if and only if $x = x(t) = (x^0(t), x^1(t), .., x^n(t))$ *is a stationary point of the functional*

$$S[x] = \int_{t_1}^{t_2} L(x^0(t), \dot{x}^0(t), x^1(t), \dot{x}^1(t), ..., x^n(t), \dot{x}^n(t))dt.$$

Proof Using the integration by parts, it is

$$\sum_{k=0}^{n}\int_{t_1}^{t_2}\frac{\partial L}{\partial x^k}\eta_k(t)dt + \sum_{k=0}^{n}\left[\frac{\partial L}{\partial \dot{x}^k}\eta_k(t_2) - \frac{\partial L}{\partial \dot{x}^k}\eta_k(t_1)\right] - \sum_{k=0}^{n}\int_{t_1}^{t_2}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right)\eta_k(t)dt =$$

$$= \sum_{k=0}^{n}\int_{t_1}^{t_2}\left[\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right)\right]\eta_k(t)dt.$$

Therefore the condition $\left.\frac{dS_\lambda}{d\lambda}\right|_{\lambda=0} \equiv 0$ reduces to

$$\sum_{k=0}^{n}\int_{t_1}^{t_2}\left[\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right)\right]\eta_k(t)dt = 0$$

for every function η_k. We obtain

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0,\ k = 0, 1, ..., n.$$

□

These equations are called the *Euler–Lagrange equations*.

They represent an equivalent way to express Newton's equations of motion in several variables for the Lagrangian $L = T - V$. However, they are more general than the Newton equations because accelerations are not required in an explicit form. See [14] for a general discussion.

Example 8.9.7 Consider a curve in the Euclidean plane, $c(t) = (t, x(t))$, $t \in [a, b] \subset \mathbb{R}$. We know, from standard calculus textbooks, that its length between the points $c(a)$ and $c(b)$ is given by the formula

$$l_a^b = \int_a^b ||\dot{c}(t)||dt = \int_a^b \sqrt{1 + \dot{x}^2(t)}dt.$$

For the Lagrangian $L(x, \dot{x}) = \sqrt{1 + \dot{x}^2}$, extremizing the functional

$$S[x] = \int_a^b \sqrt{1 + \dot{x}^2}dt\ ,$$

means to find out a curve connecting the points $A(a, x(a)),\ B(b, x(b))$ such that it has minimum length. Any other curve has a longer length. Such a curve is a line and its minimum length is the length of the segment $[AB]$.

Let us see what happens if we use the Euler–Lagrange equation. We have $\dfrac{\partial L}{\partial x} = 0$ and $\dfrac{\partial L}{\partial \dot{x}} = \dfrac{\dot{x}}{\sqrt{1 + \dot{x}^2}}$. Therefore the Euler–Lagrange equation is $\dfrac{d}{dt}\left(\dfrac{\dot{x}}{\sqrt{1 + \dot{x}^2}}\right) = 0$.

It results $\dfrac{\dot{x}}{\sqrt{1 + \dot{x}^2}} = k = \text{constant}$, i.e. $\dot{x} = \dfrac{k}{\sqrt{1 - k^2}} := m$, and finally $x = mt + n$, that is, a line equation in the Euclidean plane. The reader has to try to understand why $\sqrt{1 - k^2}$ exists.

Let us observe that the Euclidean metric is obtained from the previous Lagrangian, that is,

$$ds^2 = L^2dt^2 = \left(\sqrt{(\dot{t})^2 + \dot{x}^2}\right)^2 dt^2 = dt^2 + dx^2.$$

We may conclude that this is another proof for the fact that Euclidean lines are the geodesics of the Euclidean metric.

Example 8.9.8 Using the rule $ds^2 = L^2dt^2$, the Poincaré metric of the half-plane written as

$$ds^2 = \frac{1}{(x^2)^2}\left[(dx^1)^2 + (dx^2)^2\right]$$

allows us to highlight a Lagrangian. This is

$$L(x^1, x^2, \dot{x}^1, \dot{x}^2) := \sqrt{\frac{1}{(x^2)^2}\left[(\dot{x}^1)^2 + (\dot{x}^2)^2\right]}.$$

Let us write some modified equations in which L^2 is involved, in the form

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^i}\right) - \frac{\partial L^2}{\partial x^i} = 0, \quad i \in \{1, 2\}.$$

Denote $x := x^1$, $y := x^2$. The first one becomes

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}}\right) - \frac{\partial L^2}{\partial x} = 0,$$

that is,

$$\ddot{x} - \frac{2}{y}\dot{x}\dot{y} = 0.$$

The second one becomes

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{y}}\right) - \frac{\partial L^2}{\partial y} = 0,$$

that is,

$$\ddot{y} + \frac{1}{y}\dot{x}^2 - \frac{1}{y}\dot{y}^2 = 0.$$

Therefore, we observe that we have obtained the equations of the geodesics of the Poincaré half-plane. The solutions are

$$x = x(t) = c + R\tanh t, \quad y = y(t) = \frac{R}{\cosh t}$$

and

$$x(t) = a, \ \ y(t) = e^t,$$

therefore the curves $c_1(t) = \left(c + R\tanh t, \ \ \frac{R}{\cosh t}\right)$ and $c_2(t) = (a, \ \ e^t)$ are the stationary points of the functional

$$S[c] = \int_{t_1}^{t_2} \frac{1}{y^2}\left[\dot{x}^2 + \dot{y}^2\right] dt.$$

If we look back at the first example and we work with L^2 instead L, we obtain the same segment line as a geodesic.

These facts involving the extremization of a functional and the examples rise some fundamental questions.

- *Is there Geometry involved?*
- *Are the Euler–Lagrange equations, the geodesic equations for a given metric in which the Lagrangian is involved?*
- *Why L^2 appeared?*

The next theorem answers at all these questions.

Theorem 8.9.9 *Consider the Lagrangian $L = \sqrt{g_{ij}\dot{x}^i\dot{x}^j}$ where $g_{ij} = g_{ji}$ and g_{ij} depends only on the variables $(x^0, x^1, ..., x^n)$. Then the Euler–Lagrange equations*

$$\frac{\partial L}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0,\ \ k = 0, 1, ..., n,$$

are the geodesic equations of the metric $ds^2 = L^2 dt^2$.

Proof First, we prove that Euler–Lagrange equations have an equivalent form written with respect to L^2,

$$\frac{\partial L^2}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = -2\frac{dL}{dt}\frac{\partial L}{\partial \dot{x}^k}.$$

Let us start from the Euler–Lagrange equations and multiply by $2L$. We have

$$2L\frac{\partial L}{\partial x^k} - 2L\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0,$$

that is,

$$\frac{\partial L^2}{\partial x^k} - 2L\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = 0.$$

Next, we compute $\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right)$. We obtain

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = \frac{d}{dt}\left(2L\frac{\partial L}{\partial \dot{x}^k}\right) = 2\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k} + 2L\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right),$$

therefore

$$2L\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}^k}\right) = \frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) - 2\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k}.$$

So, the transformed equations

$$\frac{\partial L^2}{\partial x^k} - \frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = -2\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k}$$

are obtained.

Second, using $L^2 = g_{ij}\dot{x}^i\dot{x}^j$, where $g_{ij} = g_{ij}(x^0, x^1, ..., x^n)$, we have

$$\left(\frac{\partial L^2}{\partial x^k}\right) = \frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j.$$

Third: we prove the relation

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = 2g_{ks}\ddot{x}^s + 2\frac{\partial g_{ks}}{\partial x^m}\dot{x}^m\dot{x}^s.$$

This is not difficult. Successively

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = \frac{d}{dt}\left(\frac{\partial}{\partial \dot{x}^k}\left[g_{ij}\dot{x}^i\dot{x}^j\right]\right) = \frac{d}{dt}\left(g_{ij}\frac{\partial \dot{x}^i}{\partial \dot{x}^k}\dot{x}^j + g_{ij}\dot{x}^i\frac{\partial \dot{x}^j}{\partial \dot{x}^k}\right) =$$

$$= \frac{d}{dt}\left(g_{kj}\dot{x}^j + g_{ik}\dot{x}^i\right) = \frac{d}{dt}\left(2g_{ks}\dot{x}^s\right),$$

then

$$\frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = 2g_{ks}\ddot{x}^s + 2\frac{\partial g_{ks}}{\partial x^m}\frac{dx^m}{dt}\dot{x}^s = 2g_{ks}\ddot{x}^s + 2\frac{\partial g_{ks}}{\partial x^m}\dot{x}^m\dot{x}^s.$$

The forth relation to be proved is

$$\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k} = \frac{\ddot{S}}{\dot{S}}\, g_{ks}\dot{x}^s,$$

where

$$S = \int_{t_0}^{t} L d\tau = \int_{t_0}^{t}\sqrt{g_{ij}\dot{x}^i\dot{x}^j}d\tau,\ \dot{S} = L,\ \ddot{S} = \frac{dL}{dt}.$$

Step by step, we have

$$\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k} = \frac{dL}{dt}\cdot\frac{\partial}{\partial \dot{x}^k}\left[\sqrt{g_{ij}\dot{x}^i\dot{x}^j}\right] = \frac{dL}{dt}\cdot\left[\frac{1}{2\sqrt{g_{ij}\dot{x}^i\dot{x}^j}}\frac{\partial}{\partial \dot{x}^k}\left[g_{ij}\dot{x}^i\dot{x}^j\right]\right] =$$

$$= \frac{dL}{dt}\cdot\left[\frac{1}{2L}(2g_{ks}\dot{x}^s)\right] = \frac{\ddot{S}}{\dot{S}}\, g_{ks}\dot{x}^s.$$

Now, replacing in the modified Euler–Lagrange equations

$$-\frac{\partial L^2}{\partial x^k} + \frac{d}{dt}\left(\frac{\partial L^2}{\partial \dot{x}^k}\right) = 2\frac{dL}{dt}\cdot\frac{\partial L}{\partial \dot{x}^k}$$

we obtain

$$-\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j + 2g_{ks}\ddot{x}^s + 2\frac{\partial g_{ks}}{\partial x^m}\dot{x}^m\dot{x}^s = 2\frac{\ddot{S}}{\dot{S}}\, g_{ks}\dot{x}^s.$$

Manipulating the dummy indexes, the previous relation can be written in the form

$$2g_{ks}\ddot{x}^s + \left(\frac{\partial g_{ks}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^s}\right)\dot{x}^m\dot{x}^s - \frac{\partial g_{ms}}{\partial x^k}\dot{x}^m\dot{x}^s = 2\frac{\ddot{S}}{\dot{S}}\, g_{ks}\dot{x}^s.$$

The Christoffel symbols appear if we put together the last two terms of the left member,

$$g_{ks}\ddot{x}^s + \frac{1}{2}\left(\frac{\partial g_{ks}}{\partial x^m} + \frac{\partial g_{km}}{\partial x^s} - \frac{\partial g_{ms}}{\partial x^k}\right)\dot{x}^m\dot{x}^s = \frac{\ddot{S}}{\dot{S}}\, g_{ks}\dot{x}^s,$$

therefore, after multiplying by g^{ik}, we have

$$\ddot{x}^i + \Gamma^i_{ms}\dot{x}^m\dot{x}^s = \frac{\ddot{S}}{\dot{S}}\dot{x}^i,\ i \in \{0, 1, ..., n\}.$$

Still we have not the desired geodesic equations, but we are close. It remains to consider the parameter t in such a way to have a curve which is canonically parameterized.

So, we choose t such that $L = \dfrac{dS}{dt} = \dot{S} = 1$. It results $\dfrac{dL}{dt} = \ddot{S} = 0$, i.e.

$$\ddot{x}^i + \Gamma^i_{ms}\dot{x}^m\dot{x}^s = 0,\ i \in \{0, 1, ..., n\}.$$

□

We can see a new feature of Lagrangians: they are important because they induce metrics whose geodesics are described by the Euler–Lagrange equations.

Finally, we can see a possible switch between the traditional mechanical point of view for several models in Physics to the geometric point of view. Somehow the forces, the energies, some other functions involved in describing “the reality” can be replaced by geometric objects from differential geometry. The trajectories created by forces are now geodesics of spaces with metrics induced by Lagrangians. As we will see below, this point of view is fundamental in general relativity.

Chapter 9
Special Relativity

Numerus omnium aptantur.
Pythagoras

In seventeenth-century, Newton considered light as a collection of particles, now called photons according to Quantum Mechanics, traveling through space. Reflection and refraction of light were explained in a satisfactory way interpreting light rays as trajectory of photons.

James Clark Maxwell results on Electrodynamics, in the middle of the nineteenth-century, offered another view: the light is an electromagnetic wave.
Maxwell's equations of Electromagnetism are not simple at all, and, putting them in accordance with Newton's theory, points out the necessity of considering a medium in which the electromagnetic waves travel through space. This hypothetical medium was called "ether".

Ernst Mach did not agree with the idea of ether and observed the necessity of the revision of all fundamental concepts of Physics. Michelson-Morley experiment, who initially was designed to reveal such an ether, had a result completely different with respect the expectations and hard to interpret in view of Classical Mechanics. Albert Einstein explained the result of the experiment in a theory, the Special Relativity, where he revised, in a fundamental way, the ideas of space and time. After this achievement, no place remained for ether. For a comprehensive exposition of Special Relativity, see [100].

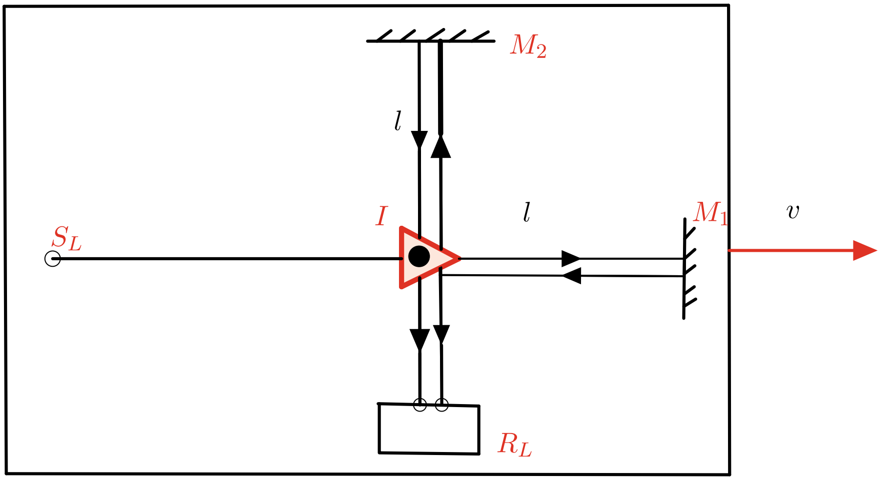
W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_9

9.1 Principles of Special Relativity

Let us first discuss about Michelson-Morley experiment.

Suppose we have a platform of a railway train wagon, an open one, on an existing straight railway line. During the Michelson-Morley experiment, the platform is considered at rest or it moves at constant speed v.

On this platform, let us imagine two perpendicular lines which intersect at I, one, say d_1, coincident to the sense of motion, the other one, say d_2, perpendicular to the sense of motion. On d_1, called the longitudinal direction, in this order, there exist: a source of light denoted by S_L, an interferometer placed in I and a mirror denoted by M_1, such that the distance between I and M_1 is l.



The interferometer is a device able to split a light-ray in the two perpendicular directions d_1 and d_2, but also to receive two light-rays from perpendicular directions and to send them separately to another given direction.

On the line d_2, which corresponds to the transversal direction, there is another mirror denoted by M_2, such that the distance between I and M_2 is the same l and a receiver-device R_L such that the interferometer I is between M_2 and R_L.

The receiver-device is able to capture the light rays coming from the interferometer and to decide which one reached first the device (Figs. 9.1 and 9.2).

The experiment is like this: when the platform is at rest or it is moving at constant speed v in the $S_L I$ longitudinal direction, a light-ray is sent by the source S_L to the interferometer I. The interferometer splits the light-ray in two light-rays. The first one is sent to the mirror M_1, it is reflected by the mirror and it is returned to the interferometer which sends it to R_L. The second one is directed to M_2, it is reflected and sent to the interferometer which sends it to R_L. Which one reaches first R_L?

This is something as: we are interested in identifying the influence of the speed v on the splitted light-rays. There is, or there is not, a difference between what is

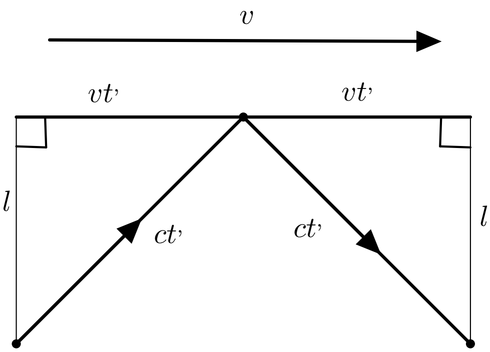


Fig. 9.1 Fig. 2.1.1$_b$

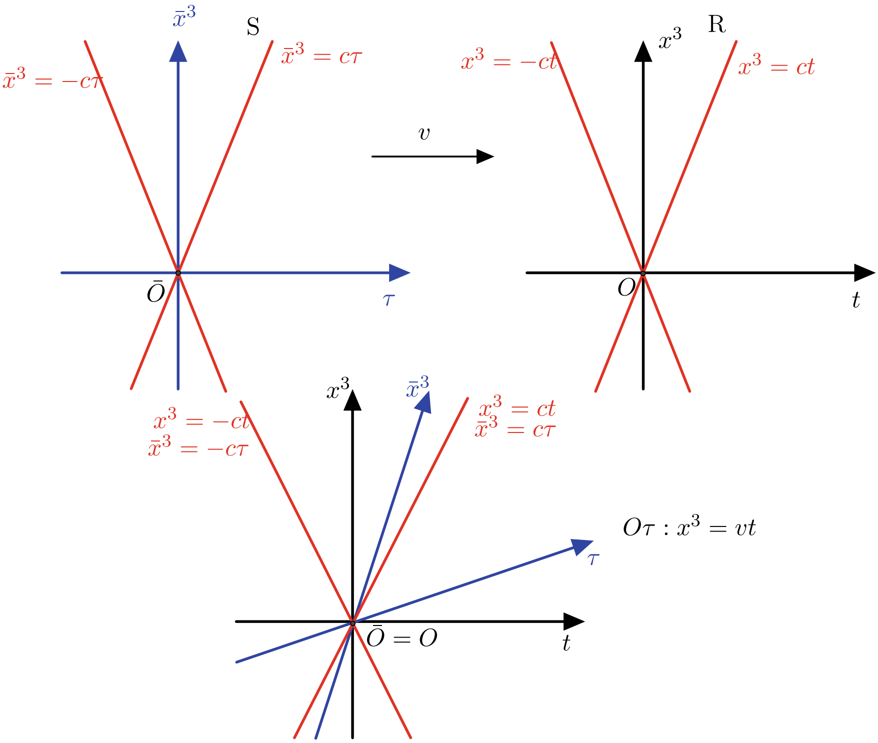


Fig. 9.2 Inertial frames and Lorentz transformation

happening when the platform is at rest comparing with the case when the platform is moving at constant speed v?

Let us observe something obvious: if the platform is at rest, both light-rays reach at same time R_L.

Now, let us try to use Classical Mechanics to describe what is happening when the platform is moving at constant speed v. First at all, let us observe that it is enough to establish only the time necessary to cover the routes IM_1I and IM_2I and to compare them.

Denote by c the speed of light. The time to cover the longitudinal route IM_1I is

$$t_1 = \frac{l}{c-v} + \frac{l}{c+v} = \frac{2lc}{c^2 - v^2},$$

because $c - v$ and $c + v$ are in Newtonian mechanics the speeds for the directions $IM_1,\ M_1I$ respectively. To be sure that the reader understands why the speeds are like this, let us focus on the first direction case. Moving at constant speed v in the sense IM_1, the photon is slowed down by the air, that is by the medium in which it is traveling, with the speed $-v$. Therefore, according to mechanics rules, the speed of the photon traveling in IM_1 direction is $c - v$.

For the transversal direction, denote by t' the time necessary for the light-ray to reach the mirror M_2. During this time, the platform, therefore the mirror, travels in the longitudinal direction a $t'v$ space. The Pythagoras theorem in the rectangle triangle formed is $(t'c)^2 = l^2 + (t'v)^2$, that is

$$t' = \frac{l}{\sqrt{c^2 - v^2}}.$$

It is obvious that the time necessary to the transversal ray to reach again the interferometer I is $t_2 := 2t'$, so we have

$$t_2 = \frac{2l}{\sqrt{c^2 - v^2}}.$$

Therefore

$$\frac{t_2}{t_1} = \sqrt{1 - \frac{v^2}{c^2}},$$

which implies

$$t_2 < t_1,$$

i.e. the transversal light-ray reaches earlier R_L compared to the longitudinal light-ray.

The mathematical model made with respect to the rules of Classical Mechanics has a prediction, let us repeat it: the first light-ray arriving in R_L is the transversal one.

If we make the experiment the result is: the transversal and the longitudinal light-rays reach R_L at the same time. If we repeat it, the same results holds. There is not a difference between what happens when the platform is at rest, comparing with the case when the platform is moving at constant speed v.

As we explained in the introduction, the error is in the model: it is related to the fact we thought that v could affect the speed of light. It seems that $c - v$ and $c + v$ are not correctly thought, therefore we have not to consider Classical Mechanics when we try to understand this experiment. Another rule has to be applied when we "add" velocities.

This experiment can be also seen making a parallel between the platform moving in Earth atmosphere at constant speed v and the Earth moving through the ether at constant speed v. After we establish a new theory to explain the experimental result, the main consequence is the fact that there is no ether[1].

The consequences of Einstein's postulates give the chance to understand how the light propagates in the context of a new physical theory, the Special Relativity, which changes the rules of Classical Mechanics when we are dealing with bodies moving at very large speeds.

Part of these results were also obtained by Henry Poincaré in his effort to explain the Michelson-Morley experiment.

Essentially, Einstein formulated the Special Relativity starting from two main postulates:

1. The laws of Physics are the same in all inertial reference frames.

2. The speed of light in vacuum, denoted by $c \approx 2,99 \cdot 10^8\ m/s$, **is the same for all the observers and it is the maximal speed reached by a moving object.**

Einstein used the word *observer* with the meaning of *reference frame* from which a set of objects or events are measured. Since the measurement are generally made with respect to the center O of the frame, this special point is often called the "O observer" or we may refer to a frame with "the observer placed at O". We know that the laws of mechanics are the same in all inertial frames. The first postulate asks for the same form of electromagnetic laws in any inertial reference frames, as the mechanics laws have. And in general, all laws of Physics must have the same form in all reference frames (this result will be fully achieved in General Relativity).

The second postulates plays a key role in Special Relativity being involved in the way in which we derive the Lorentz transformations.

The framework of Newton's laws of mechanics is the three-dimensional Euclidean space. Each object is described by a point or by a collection of points of it. Time is given by a universal clock and allows us to see the evolution of objects.

In Special Relativity, we have to work in a four-dimensional space, but not in an usual one. Three of the dimensions are the standard dimensions used in mechanics.

---

[1] In modern physics, it has been realized that "ether" is the "physical vacuum" that is a maximally symmetric configuration of space-time where no physical field is present. This means that matter-energy density is extremely low. In this "vacuum", electromagnetic waves propagate at the speed of light.

We can denote them with the letters as x^1, x^2, x^3. The fourth dimension is related to time.

Definition 9.1.1 A frame of coordinates (t, x^1, x^2, x^3) is called a space-time.

The Geometry of a space-time is in fact what we are trying to develop, and this is made according to some given physical postulates we have to accept.

Definition 9.1.2 Each point of such a space-time is called an event.

Definition 9.1.3 A curve of the space-time is called a world line and represents a successions of events.

Example 9.1.4 Suppose we work in a two dimensional slice of the previous frame, with the coordinates (t, x^3). Consider a world line starting from the the origin $O(0, 0)$. Suppose the next point is $A(1, x_0^3)$. Then the object remain t_0 seconds at rest with respect our perspective. This means that the world line has to be continued with the segment AB, where B has the coordinates $B(1 + t_0, x_0^3)$. Next, suppose the object advances in the direction $-v_1$. The line followed has the equation $x^3 - x_0^3 = -v_1(t - (1 + t_0))$, etc.

Example 9.1.5 From the origin $O(0, 0)$ an object is moving t_1 seconds in the direction $-v$. It reaches the point $M(t_1, -vt_1)$. Negative speed means only the direction of evolution in time.

Example 9.1.6 A photon is released from the origin O. There are two possible directions, c and $-c$. If it is released in the direction c, its trajectory will be the line $x^3 = ct$. Or, it can be released in the direction $-c$. Its trajectory in this case is $x^3 = -ct$. In this case, after $t_0 > 0$ seconds, the photon reaches the point $L(t_0, -ct_0)$.

In order to advance into the theory, we have to consider two local frames of coordinates, one moving at constant speed v, denoted by S, and another one considered at rest, denoted by R. The letters are chosen from the words "speed" and "rest". Two observers are placed at the origins of each system denoted by $\bar{O}$, respectively O. The first local frame S is considered described by the coordinates $(\tau = \bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3)$, while the frame R is described by the coordinates $(t = x^0, x^1, x^2, x^3)$.

Now, the reference frames of the two observers have to adapt to the second postulate of the Special Relativity. To be easier in our reasonings, let us suppose the bidimensional case when the frame S consists of the coordinates $(\tau = \bar{x}^0, \bar{x}^3)$ and it is moving, at constant speed v, in the same plane as the one determined by R, here denoted as $(t = x^0, x^3)$.

First at all, how can we express the fact that S is moving at constant speed v with respect to R? The simple mathematical answer is: the axis $\bar{O}\tau$ in R has the equation $x^3 = vt$.

Even if the light can be seen as an electromagnetic wave and we check the conservation of the form of Maxwell's equations by the Lorentz transformations, in order to develop Special Relativity, we can consider the light-rays as trajectories of photons.

What can we say about the world line of a photon in these inertial reference frames? With respect to the observers in each frame, two world lines are highlighted: a photon moving at constant speed c with a trajectory $x^3 = ct$ in R and $\bar{x}^3 = c\tau$ in S, while, for a photon moving at speed $-c$, we have the lines $x^3 = -ct$ in R and $\bar{x}^3 = -c\tau$ in S.

The two world lines of photons at O form the *light cone* of the frame R. A similar definition holds in S.

Therefore, if we use a same diagram for both frames, the second postulate has the following mathematical expression:

1. The lines $x^3 = ct$ in R and $\bar{x}^3 = c\tau$ in S have the same image;

2. The lines $x^3 = -ct$ in R and $\bar{x}^3 = -c\tau$ in S have the same image.

In other words, the two light cones are coincident.

Since we deal with inertial frames, as a rule, objects moving at constant speed in S move at constant speed in R, and vice versa. So, a straight line representing a world line of an object moving at constant speed in S, it is seen as a straight line representing the world line of the same object moving at (another) constant speed in R and vice versa. Transforming lines into lines, the change of coordinates between the two frames is described by a linear map; we denote it by L_v and we call it a *Lorentz transformation* corresponding to the speed v.

Theorem 9.1.7 *In the context described before, the matrix of the Lorentz transformation corresponding to the speed v has the form*

$$L_v = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}.$$

Proof A linear map $L_v : S \to R$ has the form

$$L_v = \begin{pmatrix} a & b \\ d & e \end{pmatrix}.$$

Since $\bar{O}\tau$ axis in R has the equation $x^3 = vt$ we have

$$\begin{pmatrix} a & b \\ d & e \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} t \\ vt \end{pmatrix},$$

that is $d = va$. In mathematical language, the second postulate is:
The eigenvectors of L_v are $\begin{pmatrix} 1 \\ c \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -c \end{pmatrix}$, that is

$$L_v \cdot \begin{pmatrix} 1 \\ c \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ c \end{pmatrix}$$

and

$$L_v \cdot \begin{pmatrix} 1 \\ -c \end{pmatrix} = \lambda_2 \begin{pmatrix} 1 \\ -c \end{pmatrix}.$$

To preserve the sense of movement of photons, it is necessary to impose two inequalities for the eigenvalues $\lambda_1 > 0,\ \lambda_2 > 0$.
Replacing L_v, it results the equations

$$\begin{cases} a\ c + b\ c^2 = a\ v + e\ c \\ \\ -a\ c + b\ c^2 = a\ v - e\ c \end{cases}$$

that is

$$L_v = a \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}.$$

To determine a, we need to observe who is the inverse of the considered Lorentz transformation.
L_v^{-1} has to act from R to S, such that $L_v\ L_v^{-1} = L_v^{-1}\ L_v = I_2$. It is standard to think at $L_v^{-1} := L_{-v}$, that is to see S at rest and R moving at constant speed $-v$. This leads to

$$I_2 = a^2 \begin{pmatrix} 1 - v^2/c^2 & 0 \\ 0 & 1 - v^2/c^2 \end{pmatrix},$$

i.e. $a^2 = \dfrac{1}{1 - v^2/c^2}$.
To determine the right sign of a, we use the Cayley Theorem. It is a simple matrix exercise: For a 2×2 real matrix B, it is

$$B^2 - 2\ Tr B \cdot B\ + \det B \cdot I_2 = O_2.$$

In our case, $Tr L_v = 2a = \lambda_1 + \lambda_2 > 0$.
The Lorentz transformation, in final form, is

$$L_v = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}.$$

□

We can write how the transformation looks like in four dimensions:

$$\begin{cases} t = \dfrac{\tau + \bar{x}^3\ v/c^2}{\sqrt{1 - v^2/c^2}} \\ x^1 = \bar{x}^1 \\ x^2 = \bar{x}^2 \\ x^3 = \dfrac{\tau\ v + \bar{x}^3}{\sqrt{1 - v^2/c^2}}. \end{cases}$$

Exercise 9.1.8 Express in four dimensions the corresponding inverse of the Lorentz transformation L_v.

Solution. According to the proof, the inverse transformation is $L_{-v} : R \to S$. In four dimensions, we have

$$\begin{cases} \tau = \dfrac{t - x^3\, v/c^2}{\sqrt{1 - v^2/c^2}} \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = \dfrac{-t\, v + x^3}{\sqrt{1 - v^2/c^2}}. \end{cases}$$

Let us observe that, for a small velocitiy v with respect to c, the ratios v/c^2 and v^2/c^2 are small enough. We can consider the influence of these terms almost zero, that is the Lorentz transformations become the usual way, in Classical Mechanics to pass from the inertial reference frame S to the inertial reference frame$R$, that is

$$\begin{cases} t = \tau \\ x^1 = \bar{x}^1 \\ x^2 = \bar{x}^2 \\ x^3 = \tau\, v + \bar{x}^3. \end{cases}$$

These formulas are called *Galilean transformations* for Classical Mechanics.

Consider three inertial reference frames, S', S and R, such that S' is moving at constant speed w with respect to S and S is moving at constant speed v with respect to R.

The two corresponding Lorentz transformations are $L_w = \dfrac{1}{\sqrt{1 - w^2/c^2}} \begin{pmatrix} 1 & w/c^2 \\ w & 1 \end{pmatrix}$ and $L_v = \dfrac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix}$.

The natural question is: which is the speed of S' with respect to R?
The answer is: We have to describe the linear map between S' and R via S, that is $L_v \cdot L_w$.

Theorem 9.1.9 *$L_v \cdot L_w = L_{v \oplus w}$, where $v \oplus w = \dfrac{v + w}{1 + vw/c^2}$.*

Proof After multiplying, we have

$$L_v \cdot L_w = \frac{1}{\sqrt{1 - v^2/c^2}} \frac{1}{\sqrt{1 - w^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & w/c^2 \\ w & 1 \end{pmatrix} =$$

$$= \frac{1 + vw/c^2}{\sqrt{(1 - v^2/c^2)(1 - w^2/c^2)}} \begin{pmatrix} 1 & \dfrac{v + w}{1 + vw/c^2} \cdot \dfrac{1}{c^2} \\ \dfrac{v + w}{1 + vw/c^2} & 1 \end{pmatrix} =$$

$$= \frac{1}{\sqrt{1 - \left(\frac{v+w}{1+vw/c^2}\right)^2 \cdot \frac{1}{c^2}}} \begin{pmatrix} 1 & \frac{v+w}{1+vw/c^2} \cdot \frac{1}{c^2} \\ \frac{v+w}{1+vw/c^2} & 1 \end{pmatrix} = L_{v \oplus w},$$

where

$$v \oplus w = \frac{v+w}{1+vw/c^2}.$$

□

Definition 9.1.10 The last formula is called the *relativistic velocities addition.*

The relativistic velocities-addition formula, in the case of small velocities, reduces to the standard sum of velocities of Classical Mechanics.

Exercise 9.1.11 Show that the set $K = (-c, c)$ endowed with the operation

$$v \oplus w = \frac{v+w}{1+vw/c^2}$$

is an abelian group.

Exercise 9.1.12 Show that the set of Lorentz transformations

$$L := \{L_v \in M_{2\times 2}(\mathbb{R}) |\ v \in (-c, c)\}$$

endowed with the usual product of matrices is an Abelian group.

9.2 Lorentz Transformations in Geometric Coordinates and Consequences

In Physics, systems of coordinates are thought with axes whose coordinates are related to the physical units as second, meter, etc. The systems of coordinates corresponding to the physical units can be called systems of *physical coordinates*. In the previous sections, we worked in physical coordinates. The units of measure in Physics were thought before understanding how deeply is the Geometry involved in the description of the physical phenomena. If we choose an appropriate "length" (e.g. the meter) and an appropriate "time duration" (e.g. the second), the speed of light can be $c = 1$. We call these new units *geometric units*. All formulas become simpler and the geometric images are more intuitive.

Definition 9.2.1 The coordinates corresponding to geometric units are called geometric coordinates.

If we adapt the second postulate conditions, seen on the same diagram, we have:
1. The lines $x^3 = t$ in R and $\bar{x}^3 = \tau$ in S have the same image
2. The lines $x^3 = -t$ in R and $\bar{x}^3 = -\tau$ in S have the same image,
in geometric coordinates, it is easier to understand how it looks like the frame S seen in R: since $O = \bar{O}$, the axis $O\bar{x}^3$ and $O\tau$ are symmetric with respect the line $x^3 = t$.

Before obtaining the Lorentz transformations in geometric coordinates, let us consider the concept of simultaneity.

9.2.1 The Relativity of Simultaneity

Two events, E_1 and E_2, are called *simultaneous* in S, if they happen at the same moment of time τ_0 in S, that is they are $E_1(\tau_0, \tau_0)$ and $E_2(\tau_0, -\tau_0)$. The same, two events, U_1 and U_2, are called simultaneous in R if they happen at the same moment of time t_0 in R, i.e. they are $U_1(t_0, t_0)$ and $U_2(t_0, -t_0)$.

On the same diagram, it is easy to see that U_1 and U_2 are simultaneous in R, but $U_1(t_0, t_0)$ and $V_2\left(t_0\dfrac{1-v}{1+v}, -t_0\dfrac{1-v}{1+v}\right)$ are simultaneous in S.

Let us explain the result from the mathematical point of view.
It is not very difficult to show that, in geometric coordinates, if $O\tau$ has the equation $x^3 = vt$, then $O\bar{x}^3$ has the equation $x^3 = \dfrac{1}{v}t$. Therefore the line $y - t_0 = \dfrac{1}{v}(t - t_0)$ intersects $x^3 = -t$, if $t = t_0\dfrac{1-v}{1+v}$.

For the observer in R, the events $U_1(t_0, t_0)$ and $U_2(t_0, -t_0)$ happen simultaneously. The observer in S cannot agree: for him $U_1(t_0, t_0)$ and $V_2\left(t_0\dfrac{1-v}{1+v}, -t_0\dfrac{1-v}{1+v}\right)$ happen simultaneously. Therefore it exists the Relativity of the *simultaneity*.

9.2.2 The Lorentz Transformations in Geometric Coordinates

In geometric coordinates, we choose the Lorentz transformation as the linear map $L_v : S \to R$,

$$L_v = \begin{pmatrix} a & b \\ d & e \end{pmatrix}.$$

Since $\bar{O}\tau$ axis in R has the equation $x^3 = vt$, we have

$$\begin{pmatrix} a & b \\ d & e \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} t \\ vt \end{pmatrix},$$

that is $d = va$. In mathematical language, the second postulate is:
The eigenvectors of L_v are $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, that is

$$L_v \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

and

$$L_v \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

To preserve the direction of movement of photons, it is necessary to impose $\lambda_1 > 0,\ \lambda_2 > 0$.

Replacing L_v, the following equations result

$$\begin{cases} a + b = a\ v + e \\ \\ -a + b = a\ v - e \end{cases}$$

that is

$$L_v = a \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}.$$

In the same way, as in the physical coordinates case, the inverse of the Lorentz transformation L_v in geometric coordinates is $L_v^{-1} := L_{-v}$. It results

$$I_2 = a^2 \begin{pmatrix} 1 - v^2 & 0 \\ 0 & 1 - v^2 \end{pmatrix},$$

that is $a^2 = \dfrac{1}{1 - v^2}$.
To determine the right sign of a, we use the same Cayley theorem: For a 2×2 real matrix B, it is

$$B^2 - 2\ TrB \cdot B\ + \det B \cdot I_2 = O_2.$$

In our case $\ TrL_v = 2a = \lambda_1 + \lambda_2 > 0$.
For those who do not understand this result, we invite to look at the characteristic equation

$$\det(B - \lambda I_2) = 0.$$

The final form of Lorentz transformation (corresponding to the velocity v), in geometric coordinates, is

$$L_v = \frac{1}{\sqrt{1 - v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}.$$

We can write how the transformation looks like in geometric coordinates in four dimensions:

$$\begin{cases} t = \dfrac{\tau + \bar{x}^3\, v}{\sqrt{1 - v^2}} \\ x^1 = \bar{x}^1 \\ x^2 = \bar{x}^2 \\ x^3 = \dfrac{\tau\, v + \bar{x}^3}{\sqrt{1 - v^2}}. \end{cases}$$

In the same case as in physical coordinates, let us consider three inertial reference frames, S', S and R, such that S' is moving at constant speed w with respect to S and S is moving at constant speed v with respect to R. Here $v,\ w$ are in $(-1, 1)$. The two corresponding Lorentz transformations are $L_w = \dfrac{1}{\sqrt{1 - w^2}} \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix}$ and $L_v = \dfrac{1}{\sqrt{1 - v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix}$.

Exercise 9.2.2 What is the speed of S' with respect to R?

Hint. We must find the linear map between S' and R, that is $L_v \cdot L_w$. A similar computation as the one made in physical coordinates leads to

$$L_v \cdot L_w = \frac{1}{\sqrt{1 - v^2}} \frac{1}{\sqrt{1 - w^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix} =$$

$$= \frac{1 + vw}{\sqrt{(1 - v^2)(1 - w^2)}} \begin{pmatrix} 1 & \dfrac{v + w}{1 + vw} \\ \dfrac{v + w}{1 + vw} & 1 \end{pmatrix} =$$

$$= \frac{1}{\sqrt{1 - \left(\dfrac{v + w}{1 + vw}\right)^2}} \begin{pmatrix} 1 & \dfrac{v + w}{1 + vw} \\ \dfrac{v + w}{1 + vw} & 1 \end{pmatrix} = L_{v \oplus w},$$

where

$$v \oplus w = \frac{v + w}{1 + vw}.$$

The last formula can be called the *addition of relativistic velocities in geometric coordinates*.

Exercise 9.2.3 Show that the set $K = (-1, 1)$ endowed with the operation

$$v \oplus w = \frac{v + w}{1 + vw}$$

is an Abelian group.

Exercise 9.2.4 Show that the set of Lorentz transformations $\{L_v \in M_{2\times 2}(\mathbb{R}) | v \in (-1, 1)\}$ endowed with the standard product of matrices is an Abelian group.

9.2.3 The Minkowski Geometry of Inertial Frames in Geometric Coordinates and Consequences: Time Dilation and Length Contraction

Let us observe that the addition of velocities was deduced using Einstein's postulates and more, it is related to the Minkowski Geometry attached to S and R frames. Why? Because if we choose

$$v = \tanh\alpha \ ; \ w = \tanh\beta,$$

we obtain the known geometric formula

$$\tanh(\alpha + \beta) = \frac{\tanh\alpha + \tanh\beta}{1 + \tanh\alpha \tanh\beta}$$

for the addition of velocities in geometric coordinates.

The Lorentz transformation corresponding to the constant speed v is now

$$L_{\tanh\alpha} = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}.$$

It is well known that the matrices $L_{\tanh\alpha}$ are hyperbolic rotations in the two-dimensional Minkowski space denoted by $\mathbb{M}^2$, where the Minkowski product of the vectors $x = (t_1, x_1^3)$ and $y = (t_2, x_2^3)$ is defined by

$$\langle x, y\rangle_M := t_1 t_2 - x_1^3 x_2^3.$$

It is also known that each matrix $L_{\tanh\alpha}$ preserves the Minkowski product.

The last property suggests another way to think at the Lorentz transformations in the case of geometric coordinates: they preserve the quantity $t^2 - (x^3)^2$.

Exercise 9.2.5 Show that Lorentz transformation implies the equality

$$\tau^2 - (\bar{x}^3)^2 = t^2 - (x^3)^2.$$

Hint.

$$t^2 - (x^3)^2 = \left(\frac{\tau + \bar{x}^3\, v}{\sqrt{1 - v^2}}\right)^2 - \left(\frac{\tau\, v + \bar{x}^3}{\sqrt{1 - v^2}}\right)^2 = \tau^2 - (\bar{x}^3)^2.$$

It results

Corollary 9.2.6 *The Lorentz transformations preserves the square of the Minkowski norm of vectors.*

Theorem 9.2.7 (Time dilation) *A clock slows down when it is moving at constant speed.*

Proof Denote by $\Delta\tau$ the unit interval of a clock moving at constant speed v. It means to consider the unit of τ axis in S to be $\Delta\tau$. Denote by Δt the corresponding element of $\Delta\tau$ after a Lorentz transformation L_v in geometric coordinates. We have

$$L_v \cdot \begin{pmatrix} \Delta\tau \\ 0 \end{pmatrix} = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta\tau \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta t \\ * \end{pmatrix},$$

where $*$ meaning is related to the fact we are not intereasted in. Therefore

$$\Delta t = \frac{\Delta\tau}{\sqrt{1-v^2}},$$

that is

$$\Delta\tau < \Delta t.$$

□

Example 9.2.8 Let us consider two twins separated. The first one is sent in space with a cosmic vehicle having the constant speed $v = 4/5$. The other one remains on Earth. When they separated they are 20 years old. After 15 years in space, according with his time, the brother from space returned. He is now, according to his time, 35 years old. How old is the brother remained on Earth, according to his perspective? The factor $\sqrt{1-v^2}$ is $3/5$. From the formula $\Delta t = \dfrac{\Delta\tau}{\sqrt{1-v^2}}$, after we replace, we obtain $3\Delta t = 5\Delta\tau$. Now, for the observer in S fifteen years have passed, that is $\Delta\tau = 15$. It results $\Delta t = 25$. Therefore his brother is 45 years old.

Theorem 9.2.9 *(Length contraction) The lengths are contracting when the frame is moving at constant speed.*

Proof Denote by $\Delta\bar{l}$ the unit length of S. Let Δl be the corresponding element of $\Delta\bar{l}$ after a Lorentz transformation L_v. In order to compare the two lengths, we compute

$$L_v \cdot \begin{pmatrix} 0 \\ \Delta\bar{l} \end{pmatrix} = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \Delta\bar{l} \end{pmatrix} = \begin{pmatrix} * \\ \Delta l \end{pmatrix},$$

where $*$ meaning is related to the fact we are not interested in. It results

$$\Delta l = \frac{\Delta\bar{l}}{\sqrt{1-v^2}},$$

that is

$$\Delta\bar{l} < \Delta l.$$

□

Example 9.2.10 A cosmic vehicle is 125 m long at rest. Suppose it is sent in space and it is moving at constant speed $v = 3/5$. How long is this moving cosmic vehicle for an observer at rest? We apply $\Delta l = \dfrac{\Delta\bar{l}}{\sqrt{1-v^2}}$ formula for $v = 3/5$ and $\Delta l = 125$. It results $\Delta\bar{l} = 100\,m$.

9.2.4 Relativistic Mass, Rest Mass and Energy

Newton's second law involves the concept of inertial mass. As we have seen at that time, the mass was considered as a constant. We have discussed about the inertial mass and the gravitational mass and how the mass is part of the so called quantity of motion, also known as momentum. In Classical Mechanics momentum means inertial mass in motion and redefined in a relativistic way, will lead to important consequences.

Let us think at an object at rest, having a rest mass denoted by $m_0 \neq 0$. Is the mass of the object "moving at constant speed" the same as its rest mass? The answer is related to how the relativistic momentum is changing with respect to the Lorentz transformations.

Let us denote by $\mathbb{P} = \begin{pmatrix} m \\ mv \end{pmatrix}$ the *relativistic momentum* of a classical body moving at constant speed v. The second component of the relativistic momentum is the classical momentum.

The relativistic momentum of a classical body at rest in S has to be $\mathbb{P}_0 = \begin{pmatrix} m_0 \\ 0 \end{pmatrix}$. According to the theory we are developing, the formula of the relativistic momentum at constant speed v is obtained from the relativistic momentum at rest, changed with respect to the Lorentz transformation L_v. This was the key point where Einstein applied, in a brilliant way, the idea that all physical formulas have to be invariant under Lorentz transformations. The consequences can be seen in the following two theorems.

Theorem 9.2.11 *If $m_0 \neq 0$ is the rest mass of a body moving at constant speed v, then*

$$m = m(v) = \frac{m_0}{\sqrt{1-v^2}}.$$

Proof Using the Lorentz transformation L_v we have $\mathbb{P} = L_v \cdot \mathbb{P}_0$. It results

$$\begin{pmatrix} m \\ mv \end{pmatrix} = \frac{1}{\sqrt{1-v^2}} \begin{pmatrix} 1 & v \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix},$$

which leads to the so called *relativistic mass* formula

$$m = m(v) = \frac{m_0}{\sqrt{1 - v^2}}.$$

□

We may observe that the mass of an object is increasing when the object travel at constant speed v. Another consequence is related to the fact that an object having its rest mass $m_0 \neq 0$ can not reach the speed of light.

Definition 9.2.12 $m(v)$ is called relativistic mass corresponding to the constant speed v of an object having the rest mass m_0.

The previous obtained formula has sense when $m_0 \neq 0$. The physicists know that there is no rest mass for the photon. Therefore this formula does not work for photon or for any other physical particle with no rest mass.

The following theorem explains why it is a good choice to consider the relativistic momentum if we intend to show how the mass is changing when it is moving at constant speed. Even if the proof is done using the geometric coordinates, the reader can change it to adapt the result to physical coordinates.

Theorem 9.2.13 *The relativistic mass formula is preserved by the Lorentz transformations.*

Proof If we consider the inertial frame S, moving at constant speed v with respect to R, we have

$$L_v \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \dfrac{m_0}{\sqrt{1 - v^2}} \\ \dfrac{m_0 v}{\sqrt{1 - v^2}} \end{pmatrix}.$$

In the same way, for the inertial frame S, moving at constant speed V with respect to R_1, we have

$$L_V \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix} = \begin{pmatrix} \dfrac{m_0}{\sqrt{1 - V^2}} \\ \dfrac{m_0 V}{\sqrt{1 - V^2}} \end{pmatrix}.$$

If the frame R is moving at constant speed w with respect to R_1, we have to compute $L_w \cdot L_v \begin{pmatrix} m_0 \\ 0 \end{pmatrix}$ and we wish the result to be coincident with $L_V \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix}$. We have

$$L_w \cdot L_v \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix} = L_w \cdot \begin{pmatrix} \dfrac{m_0}{\sqrt{1 - v^2}} \\ \dfrac{m_0 v}{\sqrt{1 - v^2}} \end{pmatrix} = \frac{1}{\sqrt{1 - w^2}} \frac{m_0}{\sqrt{1 - v^2}} \begin{pmatrix} 1 & w \\ w & 1 \end{pmatrix} \begin{pmatrix} 1 \\ v \end{pmatrix} =$$

$$= \frac{1}{\sqrt{1-w^2}} \frac{m_0}{\sqrt{1-v^2}} \begin{pmatrix} 1+wv \\ w+v \end{pmatrix} = \frac{1+wv}{\sqrt{1-w^2}} \frac{m_0}{\sqrt{1-v^2}} \begin{pmatrix} 1 \\ \dfrac{w+v}{1+wv} \end{pmatrix} =$$

$$= \frac{m_0}{\sqrt{1-\left(\dfrac{w+v}{1+wv}\right)^2}} \begin{pmatrix} 1 \\ \dfrac{w+v}{1+wv} \end{pmatrix} = L_{w\oplus v} \begin{pmatrix} m_0 \\ 0 \end{pmatrix} = L_V \begin{pmatrix} m_0 \\ 0 \end{pmatrix},$$

that is $V = w \oplus v = \dfrac{w+v}{1+wv}$. □

We are close to prove a very important consequence of the previous relativistic mass formula.

Theorem 9.2.14 *In geometric coordinates, mass means energy.*

Proof Denote by f', f'' the first and the second derivative of a real function f. It is easy to prove that

$$f(x) = f(0) + \frac{x}{1!} f'(0) + \frac{x^2}{2!} f''(0) + B[x^3],$$

where $B[x^3]$ contains only terms in x with powers greater than 3.
If we neglect the B terms, when we consider the real function

$$f(v) = \frac{1}{\sqrt{1-v^2}}$$

and the formula of the relativistic mass, we can write

$$m(v) = \frac{m_0}{\sqrt{1-v^2}} = m_0 + \frac{1}{2} m_0 v^2.$$

Looking at both members we can observe how, in geometric coordinates, the relativistic mass is related to the rest mass and the kinetic energy, that is the statement: "mass is energy" is confirmed. □

9.3 Consequences of Lorentz Physical Transformations: Time Dilation, Length Contraction, Relativistic Mass and Rest Energy

In the previous section, we used Lorentz transformations in geometric coordinates which can be called *Lorentz geometric transformations*. When we obtained, for the first time, the Lorentz transformations, we worked in physical coordinates. therefore the Lorentz transformations found there can be called *Lorentz physical transformations*. How can we adapt the previous results in the case of physical coordinates?

9.3.1 The Minkowski Geometry of Inertial Frames in Physical Coordinates and Consequences: Time Dilation and Length Contraction

If we choose

$$v = c\tanh\alpha\ ;\ \ w = c\tanh\beta,$$

we obtain the known geometric formula

$$c \cdot \tanh(\alpha+\beta) = c \cdot \frac{\tanh\alpha + \tanh\beta}{1+\tanh\alpha\tanh\beta}$$

for the velocities addition, the Lorentz transformation corresponding to the constant speed v being

$$L_{c\tanh\alpha} = \begin{pmatrix} \cosh\alpha & \frac{1}{c}\sinh\alpha \\ c\sinh\alpha & \cosh\alpha \end{pmatrix}.$$

In the two-dimensional Minkowski space, denoted by $\mathbb{M}^2$, where the Minkowski product of the vectors $x = (t_1, x_1^3)$ and $y = (t_2, x_2^3)$ is defined by

$$\langle x, y\rangle_M := c^2 t_1 t_2 - x_1^3 x_2^3\,,$$

each matrix $L_{c\tanh\alpha}$ preserves the Minkowski product.
Indeed, for $j \in \{1,2\}$ we have

$$L_{c\tanh\alpha}\cdot\begin{pmatrix}\tau_j \\ \bar{x}_j^3\end{pmatrix} = \begin{pmatrix} \cosh\alpha & \frac{1}{c}\sinh\alpha \\ c\sinh\alpha & \cosh\alpha \end{pmatrix}\cdot\begin{pmatrix}\tau_j \\ \bar{x}_j^3\end{pmatrix} = \begin{pmatrix}\tau_j\cosh\alpha + \frac{1}{c}\bar{x}_j^3\sinh\alpha \\ c\tau_j\sinh\alpha + \bar{x}_j^3\cosh\alpha\end{pmatrix},$$

and

$$c^2\left(\tau_1\cosh\alpha + \frac{1}{c}\bar{x}_1^3\sinh\alpha\right)\left(\tau_2\cosh\alpha + \frac{1}{c}\bar{x}_2^3\sinh\alpha\right) -$$

$$-\left(c\tau_1\sinh\alpha + \bar{x}_1^3\cosh\alpha\right)\left(c\tau_1\sinh\alpha + \bar{x}_1^3\cosh\alpha\right) = c^2\tau_1\tau_2 - \bar{x}_1^3\bar{x}_2^3.$$

The last property suggests another way to think at the Lorentz transformations in the case of physical coordinates: they preserve the quantity $c^2t^2 - (x^3)^2$.

Exercise 9.3.1 Show that

$$c^2\tau^2 - (\bar{x}^3)^2 = c^2t^2 - (x^3)^2.$$

Hint.

$$c^2t^2 - (x^3)^2 = c^2\left(\frac{\tau + \bar{x}^3\ v/c^2}{\sqrt{1 - v^2/c^2}}\right)^2 - \left(\frac{\tau\ v + \bar{x}^3}{\sqrt{1 - v^2/c^2}}\right)^2 = c^2\tau^2 - (\bar{x}^3)^2.$$

Now it becomes clear how the physical coordinates can be transformed into "geometric physical coordinates": The Ox^0 axis has the units done with respect to ct in R. In S, the corresponding axis becomes $c\tau$.

In this way, the unit of measure for the first axis is a length, the same as the unit for the spatial axes.

Theorem 9.3.2 *Lorentz physical transformations preserves the square of the Minkowski norm of vectors.*

However, in the case in which we are not interested in highlighting the Minkowski Geometry, we prefer to work in our initial R and S systems of coordinates.

Consider an infinitesimal time-like interval between the points (t, x) and $(t + dt, x + dx)$ and its arclength expressed in the form suggested by the previous invariant, that is

$$ds^2 = c^2(dt)^2 - (dx)^2.$$

We denoted by x the x^3 coordinate to make the notations easier. The same interval can be seen in a frame such that, at each time τ, the moving point which describes the interval is at rest. Denote by (τ, x_τ) the world line whose coordinates express the moving point at rest. Taking into account the conservation law seen before, we have

$$ds^2 = c^2(dt)^2 - (dx)^2 = c^2(d\tau)^2 - (dx_\tau)^2 = c^2(d\tau)^2.$$

Therefore

$$ds = cd\tau,$$

that is we can define

$$\Delta\tau = \int_l d\tau = \int_l \frac{ds}{c},$$

where l is the notation for the chosen time-like infinitesimal interval. We observe

$$\Delta\tau = \int_l \frac{\sqrt{c^2dt^2 - dx^2}}{c} = \int_l \sqrt{1 - \frac{1}{c^2}\frac{dx^2}{dt^2}}dt = \int_l \sqrt{1 - \frac{v^2(t)}{c^2}}dt,$$

where $v(t)$ is the usual speed.

Definition 9.3.3 $\Delta\tau$ is called a proper time interval.

Therefore, we can say that proper time measured along the time-like world line above is the time measured by a clock following point by point the considered world line. Let us give now an important property of the proper time $\Delta\tau$ in Special Relativity.

Theorem 9.3.4 (Time dilation in physical coordinates) *A clock slows down when it is moving at constant speed.*

Proof Denote by $\Delta\tau$ the unit interval of a clock moving at constant speed v. This clock measures the proper time defined above. It is like you consider the unit of τ axis in S to be $\Delta\tau$. We are interested in knowing the connection between the proper time and the time coordinate t of the frame at rest, R. Denote by Δt the corresponding element of $\Delta\tau$ after a Lorentz transformation L_v in geometric coordinates. We have

$$L_v \cdot \begin{pmatrix} \Delta\tau \\ 0 \end{pmatrix} = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} \Delta\tau \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta t \\ * \end{pmatrix},$$

where $*$ meaning is related to the fact we are not intereasted in. Therefore

$$\Delta t = \frac{\Delta\tau}{\sqrt{1 - v^2/c^2}},$$

that is

$$\Delta\tau < \Delta t.$$

□

Theorem 9.3.5 (Length contraction in physical coordinates) *The length are contracting when the frame is moving at constant speed.*

Proof Denote by $\Delta\bar{l}$ the unit length of S. Let Δl be the corresponding element of $\Delta\bar{l}$ after a Lorentz transformation L_v. In order to compare the two lengths, we compute

$$L_v \cdot \begin{pmatrix} 0 \\ \Delta\bar{l} \end{pmatrix} = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \Delta\bar{l} \end{pmatrix} = \begin{pmatrix} * \\ \Delta l \end{pmatrix},$$

where $*$ meaning is related to the fact we are not interested in. It results

$$\Delta l = \frac{\Delta\bar{l}}{\sqrt{1 - v^2/c^2}},$$

that is

$$\Delta\bar{l} < \Delta l.$$

□

9.3.2 Relativistic Mass, Rest Mass and Rest Energy in Physical Coordinates

Let us see how it looks like the relativistic mass in the case of physical coordinates. We start from an object at rest, having a rest mass denoted by $m_0 \neq 0$ with its relativistic momentum as in the case of geometrical coordinates in S, $\mathbb{P}_0 = \begin{pmatrix} m_0 \\ 0 \end{pmatrix}$.

Let us denote by $\mathbb{P} = \begin{pmatrix} m \\ mv \end{pmatrix}$ the relativistic momentum of a classical body moving at constant speed v.

Theorem 9.3.6 *If $m_0 \neq 0$ is the rest mass of a body moving at constant speed v, then*

$$m = m(v) = \frac{m_0}{\sqrt{1 - v^2/c^2}}.$$

Proof Using the Lorentz transformation L_v, we have $\mathbb{P} = L_v \cdot \mathbb{P}_0$, i.e.

$$\begin{pmatrix} m \\ mv \end{pmatrix} = \frac{1}{\sqrt{1 - v^2/c^2}} \begin{pmatrix} 1 & v/c^2 \\ v & 1 \end{pmatrix} \cdot \begin{pmatrix} m_0 \\ 0 \end{pmatrix},$$

which leads to the so called relativistic mass, now in physical coordinates,

$$m = m(v) = \frac{m_0}{\sqrt{1 - v^2/c^2}}.$$

□

As in the case of geometrical coordinates, the previous formula holds when $m_0 \neq 0$.

We are talking about the rest energy, of course, in the same case $m_0 \neq 0$. The discussion is almost the same as when we proved that, in geometric coordinates, mass means energy.

If we consider the real function

$$f(v) = \frac{1}{\sqrt{1 - v^2/c^2}}$$

and the formula of the relativistic mass, we can neglect the B terms because $1/c^4$ modify a given quantity in an irrelevant mode. We may write

$$\frac{m_0}{\sqrt{1 - v^2/c^2}} = m_0 + \frac{1}{2} m_0 v^2/c^2.$$

Let us define the *kinetic relativistic energy* by

$$E(v) := \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}.$$

The previous formula becomes

$$E(v) = m_0 c^2 + \frac{1}{2} m_0 v^2.$$

We may call *rest energy* the formula $E := m_0 c^2$; it makes sense when $m_0 \neq 0$.

A comment. It is useful, at this point, after the discussion about the relativistic mass, saying some words about the light energy which is not 0, even if the rest mass of photons is 0. To understand why, we have to accept the alternative way to consider the light as explained by Maxwell equations, that is light is an electromagnetic wave. We have also to accept the dual behavior of light and to define the photon as the particle attached to the wave[2]. The equation of photon energy is $E = hf = hc/\lambda$, where h is the Planck constant, f is the photon frequency, λ is the photon wavelength and, of course, c is the speed of light in vacuum. Therefore, in the case of a photon, we have a relativistic equivalent of mass given by the formula E/c^2.

9.4 The Maxwell Equations

The Maxwell equations are the "core" of Special Relativity. Essentially, this theory has been developed in view of explaining their invariance under Lorentz transformations. In order to discuss Maxwell's equations, which describes the electromagnetic field, we need some preliminary algebraic result.

Theorem 9.4.1 *If*

$$A = (A_1, A_2, A_3),\ B = (B_1, B_2, B_3),\ C = (C_1, C_2, C_3),$$

$$B \times C := \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix},$$

$$A \cdot B := A_1 B_1 + A_2 B_2 + A_3 B_3,\ A \cdot C := A_1 C_1 + A_2 C_2 + A_3 C_3,$$

[2] The dual nature of light, and of any particle, is better framed in the context of Quantum Mechanics in relation to the concept of wave-particle. For a discussion, see [79].

then

$$A \times (B \times C) = (A \cdot C)B - (A \cdot B)C.$$

Proof We have

$$(A \cdot C)B - (A \cdot B)C =$$

$$= (A_1C_1 + A_2C_2 + A_3C_3)(B_1, B_2, B_3) - (A_1B_1 + A_2B_2 + A_3B_3)(C_1, C_2, C_3) =$$

$$= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ A_1 & A_2 & A_3 \\ B_2C_3 - B_3C_2 & -B_1C_3 + B_3C_1 & B_1C_2 - B_2C_1 \end{vmatrix} = A \times (B \times C).$$

□

Now, consider both the gradient operator and the Laplace operator in spatial coordinates denoted by (x^1, x^2, x^3), that is

$$\nabla := \left(\frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3} \right),$$

$$\nabla^2 := \frac{\partial^2}{(\partial x^1)^2} + \frac{\partial^2}{(\partial x^3)^2} + \frac{\partial^2}{(\partial x^3)^2}.$$

The last formula can be also seen written in the formal way

$$\nabla^2 := \nabla \cdot \nabla$$

We formally define

$$\nabla \cdot A := \frac{\partial A_1}{\partial x^1} + \frac{\partial A_2}{\partial x^2} + \frac{\partial A_3}{\partial x^3}$$

and

$$\nabla \times A := \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ A_1 & A_2 & A_3 \end{vmatrix} = \left(\frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3}, \frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1}, \frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} \right).$$

Using these operators, a consequence of the above theorem is

Corollary 9.4.2

$$\nabla \times (\nabla \times A) = (\nabla \cdot A)\nabla - (\nabla \cdot \nabla)A.$$

Another comment is in order. We know the meaning of $\nabla^2\phi$, where ϕ is a scalar function. The meaning of $\nabla^2 A$ is related to the fact that ∇^2 acts on each component of A, i.e.

$$\nabla^2 A := (\nabla^2 A_1, \nabla^2 A_2, \nabla^2 A_3).$$

Therefore we can write

$$\nabla \times (\nabla \times A) = (\nabla \cdot A)\nabla - \nabla^2 A.$$

If $\nabla \cdot A = 0$, the previous formula becomes

Corollary 9.4.3

$$\nabla \times (\nabla \times A) = -\nabla^2 A.$$

We will use this result later.
Denote by

$$E = E(t, x^1, x^2, x^3) := (E_1(t, x^1, x^2, x^3), E_2(t, x^1, x^2, x^3), E_3(t, x^1, x^2, x^3))$$

the electric force vector and by

$$H = H(t, x^1, x^2, x^3) := (H_1(t, x^1, x^2, x^3), H_2(t, x^1, x^2, x^3), H_3(t, x^1, x^2, x^3))$$

the magnetic force vector;

In geometric units, the *Maxwell equations*, in the frame R considered as an empty space, are

$$\begin{cases} \nabla \cdot E = 0 \\ \nabla \times E = -\dfrac{\partial H}{\partial t} \\ \nabla \cdot H = 0 \\ \nabla \times H = \dfrac{\partial E}{\partial t} \end{cases}$$

The first equation reveals the existence of an electric field in the absence of electric charge. If we are not in vacuum, the first equation is $\nabla \cdot E = \rho$, where ρ is the electric charge, therefore the first equation describes how an electric charge acts as source for the electric force, here seen as an electric field.

The second equation $\nabla \times E = -\dfrac{\partial H}{\partial t}$ shows how a time varying magnetic field gives rise to an electric field.

The third equation $\nabla \cdot H = 0$ shows that there are no magnetic charges.

The forth equation $\nabla \times H = \dfrac{\partial E}{\partial t}$ shows how the time variation of electric field creates the magnetic field.

Let us consider the derivative with respect t of the second equation.

$$-\frac{\partial^2 H}{\partial t^2} = \frac{\partial}{\partial t}(\nabla \times E) = \frac{\partial}{\partial t}\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ E_1 & E_2 & E_3 \end{vmatrix} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ \frac{\partial E_1}{\partial t} & \frac{\partial E_2}{\partial t} & \frac{\partial E_3}{\partial t} \end{vmatrix} = \nabla \times \frac{\partial E}{\partial t}.$$

Using the last Maxwell equation and the above results, we find

$$-\frac{\partial^2 H}{\partial t^2} = \nabla \times \frac{\partial E}{\partial t} = \nabla \times (\nabla \times H) = -\nabla^2 H,$$

that is

$$\frac{\partial^2 H}{\partial t^2} = \nabla^2 H.$$

If we denote by

$$\Box := \frac{\partial^2}{\partial t^2} - \nabla^2$$

the d'Alembert operator, the previous equation is

$$\Box H = 0.$$

This is the wave equation corresponding to the magnetic field. Therefore, for each component $H_i,\ i \in \{1, 2, 3\}$ we have

$$\frac{\partial^2 H_i}{\partial t^2} = \nabla^2 H_i = \frac{\partial^2 H_i}{(\partial x^1)^2} + \frac{\partial^2 H_i}{(\partial x^2)^2} + \frac{\partial^2 H_i}{(\partial x^3)^2}.$$

Now, let us consider the derivative with respect t of the last equation.

$$\frac{\partial^2 E}{\partial t^2} = \frac{\partial}{\partial t}(\nabla \times H) = \frac{\partial}{\partial t}\begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ H_1 & H_2 & H_3 \end{vmatrix} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ \frac{\partial H_1}{\partial t} & \frac{\partial H_2}{\partial t} & \frac{\partial H_3}{\partial t} \end{vmatrix} = \nabla \times \frac{\partial H}{\partial t}.$$

Using the second Maxwell's equation and the consequence, we find that

$$\frac{\partial^2 E}{\partial t^2} = \nabla \times \frac{\partial H}{\partial t} = -\nabla \times (\nabla \times E) = \nabla^2 E,$$

i.e.

$$\Box E = 0.$$

This one is the wave equation corresponding to the electric field. We have now a picture of the electromagnetic field described by the Maxwell equations: The two waves equations of electric and magnetic field are interconnected by the four Maxwell equations. We understand that one field can not exist without the other. Each one generates the other.

Are these wave equations invariant under Lorentz transformations? The answer is yes, but we need to perform more steps in order to achieve these results.

In the same way as before, for each component $E_i,\ i \in \{1, 2, 3\}$, we have

$$\frac{\partial^2 E_i}{\partial t^2} = \nabla^2 E_i = \frac{\partial^2 E_i}{(\partial x^1)^2} + \frac{\partial^2 E_i}{(\partial x^2)^2} + \frac{\partial^2 E_i}{(\partial x^3)^2}.$$

To simplify, let us suppose that the electric field E depends only on the variables t and x^3, as in the case of a plane wave. The previous equations become

$$\frac{\partial^2 E_i}{\partial t^2} - \frac{\partial^2 E_i}{(\partial x^3)^2} = 0.$$

To continue, let us choose a component only, say $i = 1$. Since for the other two components, the following computations are the same, we prefer instead to use E_1, to denote this chosen component by the letter $\mathbb{E}$. The previous equation becomes

$$\frac{\partial^2 \mathbb{E}}{\partial t^2} - \frac{\partial^2 \mathbb{E}}{(\partial x^3)^2} = 0.$$

How this simple equation looks like in S, frame considered with coordinates $\tau, \bar{x}_3$, if S is supposed to move at constant speed v along the x_3 axis in R? We have to use the Lorentz inverse transformation L_{-v}, that is

$$\begin{cases} \tau = \dfrac{t - x^3\ v}{\sqrt{1 - v^2}} \\ \bar{x}^3 = \dfrac{-t\ v + x^3}{\sqrt{1 - v^2}}. \end{cases}$$

Denote by $\bar{\mathbb{E}}(\tau, \bar{x}^3) = \bar{\mathbb{E}}\left(\dfrac{t - x^3\ v}{\sqrt{1 - v^2}}, \dfrac{-t\ v + x^3}{\sqrt{1 - v^2}}\right) := \mathbb{E}(t, x^3)$ the corresponding component of the electric field in S, which, obviously have to be the same as in R. We would like to prove that

$$\frac{\partial^2 \mathbb{E}}{\partial t^2} - \frac{\partial^2 \mathbb{E}}{(\partial x^3)^2} = \frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2} - \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}.$$

We have

$$\frac{\partial \mathbb{E}}{\partial t}=\frac{\partial \bar{\mathbb{E}}}{\partial \tau}\frac{\partial \tau}{\partial t}+\frac{\partial \bar{\mathbb{E}}}{\partial \bar{x}^3}\frac{\partial \bar{x}^3}{\partial t}=\frac{\partial \bar{\mathbb{E}}}{\partial \tau}\frac{1}{\sqrt{1-v^2}}+\frac{\partial \bar{\mathbb{E}}}{\partial \bar{x}^3}\frac{-v}{\sqrt{1-v^2}}$$

and

$$\frac{\partial^2 \mathbb{E}}{\partial t^2}=\frac{1}{\sqrt{1-v^2}}\left(\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2}\frac{\partial \tau}{\partial t}+\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau \partial \bar{x}^3}\frac{\partial \bar{x}^3}{\partial t}\right)-\frac{v}{\sqrt{1-v^2}}\left(\frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau}\frac{\partial \tau}{\partial t}+\frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}\frac{\partial \bar{x}^3}{\partial t}\right),$$

that is

$$\frac{\partial^2 \mathbb{E}}{\partial t^2}=\frac{1}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2}-\frac{2v}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau}+\frac{v^2}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}.$$

In the same way

$$\frac{\partial^2 \mathbb{E}}{(\partial x^3)^2}=\frac{v^2}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2}-\frac{2v}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau}+\frac{1}{1-v^2}\frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2},$$

therefore the desired relation is obtained by subtracting the two expressions. Now, from

$$\frac{\partial^2 \mathbb{E}}{\partial t^2}-\frac{\partial^2 \mathbb{E}}{(\partial x^3)^2}=0,$$

in R, we obtain

$$\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2}-\frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}=0,$$

in S, that is the corresponding equation is the same as it has to be. Therefore, in a moving inertial frame, the Maxwell equations are the same as in a frame at rest. We have proved

Theorem 9.4.4 *Lorentz transformations preserve Maxwell's equations.*

If the reader try to prove if the equality

$$\frac{\partial^2 \mathbb{E}}{\partial t^2}-\frac{\partial^2 \mathbb{E}}{(\partial x^3)^2}=\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2}-\frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}$$

holds for the inverse of Galilean transformations $\bar{\mathbb{E}}(\tau,\bar{x}^3)=\bar{\mathbb{E}}\left(t,-vt+x^3\right):=\mathbb{E}(t,x^3)$, the answer is no, that is the Galilean transformations fail for the Maxwell equations. This can be easily shown. If the reader computes

$$\frac{\partial \mathbb{E}}{\partial t}=\frac{\partial \bar{\mathbb{E}}}{\partial \tau}\frac{\partial \tau}{\partial t}+\frac{\partial \bar{\mathbb{E}}}{\partial \bar{x}^3}\frac{\partial \bar{x}^3}{\partial t}=\frac{\partial \bar{\mathbb{E}}}{\partial \tau}-v\frac{\partial \bar{\mathbb{E}}}{\partial \bar{x}^3}$$

and

$$\frac{\partial^2 \mathbb{E}}{\partial t^2} = \left(\frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2} \frac{\partial \tau}{\partial t} + \frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau \partial \bar{x}^3} \frac{\partial \bar{x}^3}{\partial t} \right) - v \left(\frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau} \frac{\partial \tau}{\partial t} + \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2} \frac{\partial \bar{x}^3}{\partial t} \right) =$$

$$= \frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2} - 2v \frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau} + v^2 \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}.$$

Then,

$$\frac{\partial \mathbb{E}}{\partial x^3} = \frac{\partial \bar{\mathbb{E}}}{\partial \bar{x}^3}$$

and

$$\frac{\partial^2 \mathbb{E}}{(\partial x^3)^2} = \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2},$$

that is

$$\frac{\partial^2 \mathbb{E}}{\partial t^2} - \frac{\partial^2 \mathbb{E}}{(\partial x^3)^2} = \frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2} - \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2} - 2v \frac{\partial^2 \bar{\mathbb{E}}}{\partial \bar{x}^3 \partial \tau} + v^2 \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2} \neq \frac{\partial^2 \bar{\mathbb{E}}}{\partial \tau^2} - \frac{\partial^2 \bar{\mathbb{E}}}{(\partial \bar{x}^3)^2}.$$

Theorem 9.4.5 *Galilei's transformations do not preserve Maxwell's equations.*

The final conclusion is: Classical Mechanics through Galilei's transformations does not preserve Maxwell's equations while the Special Relativity, through Lorentz transformations, does it.

9.5 The Doppler Effect in Special Relativity

We have proved that the speed of light does not depend on the speed of the source of light. Let us now focus on the frequency of light signals. We prove that the frequency of light signals depends on the speed of the source, that is, we show that light frequency is increasing when the source is approaching to the observer O at rest in R, then, when the source is moving away, the light frequency is decreasing. This is the so called *Doppler's effect* or *relativistic Doppler's effect*.

Definition 9.5.1 Doppler's effect is a change in frequency of light-wave when a source is moving at constant speed with respect to the frequency perceived by an observer at rest.

Therefore we have two different formulas, one to estimate the frequency of the source which is approaching, and another one for the frequency in the case when the source is moving away. Let us translate this in a mathematical way.

Consider, as usual, two local frames of geometric coordinates, one moving at constant speed v, denoted by S, and another one considered at rest, denoted by R.

The first local frame S is described by the coordinates $(\tau = \bar{x}^0, \bar{x}^3)$, while the frame R is described by the coordinates $(t = x^0, x^3)$.

Consider a source of light in S which, for each $\Delta\tau$ seconds, releases a light signal. If the *frequency* is denoted by ν, the connection between the two physical quantities is

$$\nu = \frac{1}{\Delta\tau}.$$

This formula is related to the behavior of a light wave.
The quantity $\Delta\tau$ is the *period* of a light wave in the frame S. The light wave, with frequency ν, is imagined as emitted light signals of duration $\Delta\tau$ seconds. They produces light cones with the vertexes on the τ-axis.

So, the source is moving in S along the τ-axis and, in R, this τ axis becomes the line $x^3 = vt$. The observer, at the origin O of R, perceives the source first as approaching, then as moving away.

To simplify, let us consider the moment when the two origins are coincident and, on the τ-axis, we draw $\Delta\tau$ intervals to the left and to the right. The light cones, considered in S, determine two kinds of equal intervals Δt on the t-axis in R. Until the origin, we denote them Δt_{app}, after we denote them by Δt_{ma}, each one determining its corresponding frequency in R. The subscript *app* and *ma* are obviously from the words "approaching" and "moving away".

Therefore, two kinds of frequencies appear in R, that is

$$f_{app} := \frac{1}{\Delta t_{app}}$$

and

$$f_{ma} := \frac{1}{\Delta t_{ma}}.$$

Theorem 9.5.2 *According to the above conditions, we have:*
(i). If the source is approaching,

$$f_{app} = \nu\sqrt{\frac{1+v}{1-v}} > \nu.$$

(ii). If the source is moving away,

$$f_{ma} = \nu\sqrt{\frac{1-v}{1+v}} < \nu.$$

Proof Denote by $(0, 0)$ and (b, vb), the coordinates at the ends of the first interval $\Delta\tau$ on the τ-axis, as seen in R. The light-ray emitted at the point (b, vb) reaches the t-axis at $(b + vb, 0)$. Of course, in this case, we used the photon corresponding to speed -1. Therefore

$$\Delta t_{ma} = b(1 + v).$$

Now, consider the points $(0, 0)$ and $(-b, -vb)$ as the coordinates of a $\Delta\tau$ interval when the source is approaching. The light-ray emitted at the point $(-b, -vb)$ reaches the t-axis at $(-b + vb, 0)$ because we used the photon corresponding to speed 1. In this case

$$\Delta t_{app} = b(1 - v).$$

If we consider the Minkowski arc length corresponding to a $\Delta\tau$ interval, we have

$$\Delta\tau^2 = b^2 - b^2 v^2,$$

that is

$$b = \frac{\Delta\tau}{\sqrt{1 - v^2}}.$$

Now using this last formula and the two formulas for $f_{app} = \frac{1}{\Delta t_{app}}$ and $f_{ma} = \frac{1}{\Delta t_{ma}}$, the statement is proved. □

Let us observe that we can write the two above formulas in the form

$$f = \nu\sqrt{\frac{1 - v}{1 + v}},$$

if we perceive the approaching wave as moving away with speed $-v$.

9.6 Gravity in Special Relativity: The Case of the Constant Gravitational Field

The fact that Special Relativity had to be improved towards General Relativity is essentially due to two main reasons: From one side, Einstein, according to the Mach criticisms [21], realized that the laws of Physics must be written in the same way for *any* (inertial or non-inertial) observer (Invariance Principle). Secondly, considering the gravitational phenomena, he realized that one needs to introduce accelerating frames. According to these observations, Special Relativity is inadequate to enclose gravity.

In order to discuss gravity in the framework of Special Relativity and show their basic incompatibility, let us begin considering a very simple result.

In a Minkowski space, for every t, let $v(t)$ be a vector of constant norm.

It results $\langle v(t), v(t)\rangle_M = k$. If we consider the derivative with respect to t, we obtain

$$\langle \dot{v}(t), v(t)\rangle_M = 0.$$

We have proved the following

Proposition 9.6.1 *(i)The derivative of a constant norm vector is a vector orthogonal on the given vector, that is $\dot{v}(t) \perp_M v(t)$,*
(ii) The vectors $\dot{v}(t)$ and $v(t)$ are Minkowski type different, that is, if $v(t)$ is space-like vector, the derivative $\dot{v}(t)$ is time-like vector, and vice versa.

A second very important observation is this one:
In a local frame S of coordinates $(\tau = \bar{x}^0, \bar{x}^3)$, let us consider an event $E(\tau, \bar{x}^3)$, $\bar{x}^3 > 0$. There are only two events on the τ-axis, say $E_1(\tau_1, 0)$ and $E_2(\tau_2, 0)$ with $\tau_1 < \tau_2$, such that the event E is connected to the events E_1 and E_2 by light-rays. Indeed, considering that the slopes of the lines E_1E and E_2E have to be 1 and -1 respectively, the connections among the coordinates are

$$\tau = \frac{\tau_1 + \tau_2}{2}; \quad \bar{x}^3 = \frac{\tau_2 - \tau_1}{2},$$

or equivalently

$$\tau_1 = \tau - \bar{x}^3; \quad \tau_2 = \bar{x}^3 + \tau.$$

Therefore we have proved.

Proposition 9.6.2 *Suppose the event $E(\tau, \bar{x}^3)$, $\bar{x}^3 > 0$ is connecting the events E_1 and E_2 by light-rays. If the coordinates are $E_1(\tau_1, 0)$, $E_2(\tau_2, 0)$, $\tau_1 < \tau_2$, then, between the above coordinates there are the relations*

$$\tau_1 = \tau - \bar{x}^3; \quad \tau_2 = \bar{x}^3 + \tau.$$

The physical image is the following: a light-ray from E_1 reaches E and is reflected to E_2. The coordinates are like in the previous proposition.

Let us now suppose that τ-axis is seen in the frame R as a curve. To move forward, let us suppose that τ-axis is parameterized by

$$\tau - axis : \begin{cases} t(\tau) = \dfrac{1}{\alpha} \sinh \alpha\tau \\ x^3(\tau) = \dfrac{1}{\alpha} \cosh \alpha\tau. \end{cases}$$

Consider the event E, now in coordinates of R, that is $E(t, x^3)$.

The events E_1 and E_2 belong now to the curve which represents the τ-axis in R, that is $E_1(t_1, x_1^3)$ with $t_1 = \dfrac{1}{\alpha} \sinh \alpha\tau_1$; $x_1^3 = \dfrac{1}{\alpha} \cosh \alpha\tau_1$

and

$E_2(t_2, x_2^3)$ with $t_2 = \dfrac{1}{\alpha} \sinh \alpha\tau_2$; $x_2^3 = \dfrac{1}{\alpha} \cosh \alpha\tau_2$,

in such a way that a light-ray from E_1 reaches E and is reflected to E_2.
Since the slopes E_1E and E_2E have to be 1 and -1 respectively, we have

$$\frac{x^3 - x_1^3}{t - t_1} = 1; \ \frac{x^3 - x_2^3}{t - t_2} = -1.$$

It results the system of equations

$$\begin{cases} -t + x^3 = -t_1 + x_1^3 \\ t + x^3 = t_2 + x_2^3 \end{cases}$$

with the solution

$$\begin{cases} t = \dfrac{t_1 + t_2 + x_2^3 - x_1^3}{2} \\ x^3 = \dfrac{-t_1 + t_2 + x_2^3 + x_1^3}{2}. \end{cases}$$

The first formula becomes

$$t = \frac{\sinh \alpha\tau_1 + \sinh \alpha\tau_2 + \cosh \alpha\tau_2 - \cosh \alpha\tau_1}{2\alpha} = \frac{e^{\alpha\tau_2} - e^{-\alpha\tau_1}}{2\alpha},$$

that is

$$t = \frac{e^{\alpha(\tau+\bar{x}^3)} - e^{-\alpha(\tau-\bar{x}^3)}}{2\alpha} = \frac{e^{\alpha\bar{x}^3}}{\alpha} \sinh \alpha\tau.$$

In the same way

$$x^3 = \frac{e^{\alpha\bar{x}^3}}{\alpha} \cosh \alpha\tau,$$

that is we found out a coordinate transformation $G : S \to R$,

$$G : \begin{cases} t(\tau, \bar{x}^3) = \dfrac{e^{\alpha\bar{x}^3}}{\alpha} \sinh \alpha\tau \\ x^3(\tau, \bar{x}^3) = \dfrac{e^{\alpha\bar{x}^3}}{\alpha} \cosh \alpha\tau. \end{cases}$$

This is the proof of the following

Theorem 9.6.3 *Consider a system R of coordinates (t, x^3) in which the τ-axis is the curve parameterized by*

$$\begin{cases} t(\tau) = \dfrac{1}{\alpha} \sinh \alpha\tau \\ x^3(\tau) = \dfrac{1}{\alpha} \cosh \alpha\tau. \end{cases}$$

Suppose it exists three events E_1, E, E_2 such that a light-ray from E_1 reaches E and is reflected to E_2. Then, between the coordinates of $E_1(t_1, x_1^3)$, $E_2(t_2, x_2^3)$ with

$$t_1 = \frac{1}{\alpha}\sinh\alpha\tau_1 \;,\; x_1^3 = \frac{1}{\alpha}\cosh\alpha\tau_1 \;;\; t_2 = \frac{1}{\alpha}\sinh\alpha\tau_2 \;,\; x_2^3 = \frac{1}{\alpha}\cosh\alpha\tau_2 \;;$$

and the coordinates of the event $E(t, x^3)$*, there are the relations*

$$\begin{cases} t(\tau, \bar{x}^3) = \dfrac{e^{\alpha\bar{x}^3}}{\alpha}\sinh\alpha\tau \\ x^3(\tau, \bar{x}^3) = \dfrac{e^{\alpha\bar{x}^3}}{\alpha}\cosh\alpha\tau, \end{cases}$$

where

$$\tau_1 = \tau - \bar{x}^3; \quad \tau_2 = \bar{x}^3 + \tau.$$

Now, we consider a local frame S with coordinates $(\tau = \bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3)$ in which a constant gravitational field exists. This constant gravitational field can be imagined as a vector $-\vec{\alpha}$ acting along the $\bar{x}^3$ axis in its negative direction, therefore as a vector with the spatial coordinates $(0, 0, -\alpha)$.

Let us consider another frame of coordinates R, whose coordinates are $(t = x^0, x^1, x^2, x^3)$. This frame is in free fall in the previous constant gravitational field.

We may assume that, at $\tau = t = 0$, the two frames can be seen together with axes corresponding in notation of indexes. Let us suppose that the second frame R is moving along the $\bar{x}^3$ axis in its negative direction. So, we can think of a transformation which describes the constant gravitational field in S, involving only the pairs of axis $(\tau, \bar{x}^3)$ of S and (t, x^3) of R.

To obtain it, we change the perspective: We consider R at rest and the frame S accelerating along the x^3 axis with the constant acceleration $(0, \alpha)$.

When we determine the Lorentz transformation, our first concern is describing the τ axis in R when S is moving at constant speed v. The question is: If S is $\vec{\alpha}$-accelerating with respect to R, what becomes the τ axis of S in R?

Let us think of a line as a trajectory of a moving point. The speed is constant along the line, that is the vector speed of a given line has constant norm; in the same time, the acceleration vector is null. If we consider the current point $(\tau, 0)$ on τ-axis, the speed vector is $\vec{V} = (1, 0)$ and the acceleration vector is $\vec{A} = (0, 0)$. Therefore the τ-axis at rest is characterized by

$$|| \vec{V} ||_M = 1; \; || \vec{A} ||_M = 0.$$

Looking at the accelerated frame S, the R observer sees a modified τ-axis, denoted now

$$c(\tau) := (t(\tau), x^3(\tau))$$

and characterized by

$$||\overrightarrow{\dot{c}}\ (\tau)||_M = 1;\ ||\overrightarrow{\ddot{c}}\ (\tau)||_M = \alpha.$$

Now, we observe that, according to the first proposition, the speed vector $\overrightarrow{\dot{c}}$ is a time-like one, while the acceleration vector $\overrightarrow{\ddot{c}}$ is a space-like one. The two conditions become the system of differential equations

$$\begin{cases} (\dot{t}(\tau))^2 - (\dot{x}^3(\tau))^2 = 1 \\ -(\ddot{t}(\tau))^2 + (\ddot{x}^3(\tau))^2 = \alpha^2 \end{cases}$$

with the general solution

$$\begin{cases} t(\tau) = \dfrac{1}{\alpha} \sinh \alpha(\tau + \tau_0) + t_0 \\ x^3(\tau) = \mp \dfrac{1}{\alpha} \cosh \alpha(\tau + \tau_0) + x_0^3. \end{cases}$$

From the Euclidean point of view, we deal with the hyperbola

$$(t - t_0)^2 - (x^3 - x_0^3)^2 = \frac{1}{\alpha^2}$$

having the center at (t_0, x_0^3) and the parallel asymptotes along the light cone. This is a good exercise for the reader.

Of course, in R, where the Minkowski Geometry is acting, this curve is a Minkowski space-like circle. From symmetry reason, we may choose the center of this hyperbola at $(0, 0)$, $\tau_0 = 0$ and the sign $+$. That is, we have an image of the τ-axis of S in R,

$$\begin{cases} t(\tau) = \dfrac{1}{\alpha} \sinh \alpha\tau \\ x^3(\tau) = \dfrac{1}{\alpha} \cosh \alpha\tau. \end{cases}$$

We have proved the following

Theorem 9.6.4 *If a coordinates frame S is $\vec{\alpha}$-accelerated with respect to a frame at rest R, then the image of the τ-axis of S in R is a curve $c(\tau) := (t(\tau), x^3(\tau))$ characterized by the equations*

$$\begin{cases} t(\tau) = \dfrac{1}{\alpha} \sinh \alpha\tau \\ x^3(\tau) = \dfrac{1}{\alpha} \cosh \alpha\tau. \end{cases}$$

Now, we have the complete image: It exists a local change of coordinates between S and R described by the transformation G, $G : S \to R$

$$G : \begin{cases} t(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \sinh \alpha\tau \\ x^3(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \cosh \alpha\tau. \end{cases}$$

The transformation G, which was defined by using the idea of accelerating frame, allows us to understand how the constant gravitational field $-\vec{\alpha}$ in the frame of coordinates S can be seen via the system of coordinates R. Consequently, in the future, we will be able to compute the metric of S.

Exercise 9.6.5 Show that the inverse transformation $G^{-1} : R \to S$ is

$$G^{-1} : \begin{cases} \tau(t, x^3) = \dfrac{1}{\alpha} \tanh^{-1}\left(\dfrac{t}{x^3}\right) \\ \bar{x}^3(t, x^3) = \dfrac{1}{2\alpha} \ln\left[\alpha^2\left(\left(x^3\right)^2 - t^2\right)\right]. \end{cases}$$

9.6.1 The Doppler Effect in Constant Gravitational Field and Consequences

We know, up to this point, that frames at rest and frames moving at constant speed are inertial frames. The laws of mechanics and the new laws of Special Relativity have the same form and hold in such frames. There are no evidences that frames in which acts a constant gravitational field are non-inertial frames. Are they really inertial frames? The answer is related to the Doppler effect in a constant gravitational field.

We are interested in finding out how the frequency of light in S is affected by the constant gravitational field $-\vec{\alpha}$ which acts in S.

To obtain a formula which connects the frequency of the light and $-\vec{\alpha}$, we need to change the perspective as we have done before. We use two frames of coordinates S and R. Instead of looking at the frame of coordinates R in free fall in the previous constant gravitational field, we look at R at rest and at the frame S accelerated along the x^3 axis with the constant acceleration $\vec{\alpha}$.

Our study is done again in the two corresponding slices of S and R, taking into account the coordinates $(\tau = \bar{x}^0, \bar{x}^3)$, respectively $(t = x^0, x^3)$.

Let us pose the problem.

From the origin $O(0, 0)$ of S is emitted a light signal with frequency ν. Consider $C(0, h)$, a point on $\bar{x}^3$ axis at height h. The level h is reached by the light-ray at the point $H(h, h)$. In order to obtain the frequency at the level h, we need to consider the frame R. Denote by f_h the frequency of the light-ray in R corresponding to the level h in S. We have

Theorem 9.6.6

$$f_h = \nu e^{-\alpha h}.$$

Proof Let us remember the transformation G,

$$G : \begin{cases} t(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \sinh \alpha\tau \\ x^3(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \cosh \alpha\tau. \end{cases}$$

The points O and C, from S, are seen through G in R with the coordinates $(t_0, x_0^3) = \left(0, \dfrac{1}{\alpha}\right)$ and $(t_C, x_C^3) = \left(0, \dfrac{e^{\alpha h}}{\alpha}\right)$ respectively.

The equivalent of the point H in R has the coordinates $\left(\dfrac{e^{\alpha h}}{\alpha} \sinh \alpha h, \dfrac{e^{\alpha h}}{\alpha} \cosh \alpha h\right)$. Since, through $\left(0, \dfrac{1}{\alpha}\right)$, the new τ-axis passes in R, that is the curve $c(\tau) = \left(\dfrac{1}{\alpha} \sinh \alpha\tau, \dfrac{1}{\alpha} \cosh \alpha\tau\right)$, equivalent to the line $\bar{x}^3 = h$, is the curve $c_h(\tau) = \left(\dfrac{e^{\alpha h}}{\alpha} \sinh \alpha\tau, \dfrac{e^{\alpha h}}{\alpha} \cosh \alpha\tau\right)$.

The speed vector at h has the components $\left(e^{\alpha h} \cosh \alpha h, e^{\alpha h} \sinh \alpha h\right)$, that is

$$v_h = \tanh \alpha h.$$

We replace this formula in the general formula found before for the relativistic Doppler's effect and it results

$$f_h = \nu \sqrt{\frac{1 - v_h}{1 + v_h}},$$

that is

$$f_h = \nu \sqrt{\frac{1 - \tanh h}{1 + \tanh h}} = \nu e^{-\alpha h}.$$

□

Let us observe that if we denote by

$$\Delta\tau = \frac{1}{\nu}$$

the corresponding period in S, and by

$$\Delta t = \frac{1}{f_h}$$

the cooresponding period in R, we obtain a formula connecting the two periods, that is

$$\Delta\tau = e^{-\alpha h}\Delta t.$$

If $h > 0$, that is, if the point C belongs to the upper half-plane of S, comparing the periods in S with the one in R, we have

$$\Delta\tau < \Delta t.$$

If $h < 0$, that is if the point C is in the complementary half-plane of S, we have

$$\Delta\tau > \Delta t.$$

If h is very close to 0, we may consider the approximation

$$e^{-\alpha h} = 1 - \alpha h.$$

From a physical point of view, $-\alpha h$ corresponds to a potential energy for an object whose mass is 1. Therefore we can write the formula

$$\Delta t = \frac{1}{1-\alpha h}\Delta\tau$$

written with respect to the potential energy.

Now, let us take into account two clocks, one in O and one in C. Suppose the first one ticking at each $\Delta\tau$ seconds. The second clock at C is ticking in Δt seconds.

The results $\Delta\tau < \Delta t$ if $h > 0$ and $\Delta\tau > \Delta t$, if $h < 0$ hold.

This situation shows that S cannot be an inertial reference frame. In an inertial reference frame, the position cannot affect the way in which time is running. In the entire frame S, we should have a same result.

Therefore we have.

Corollary 9.6.7 *The frames in which a constant gravitational field is acting are not inertial frames.*

A further remark is the following. Let us suppose we are on the surface of a planet. Consider $0 < h_1 < h_2$. It results $-\alpha h_1 > -\alpha h_2$, that is $1 - \alpha h_1 > 1 - \alpha h_2$. Suppose that $h_1,\ h_2$ are so small than the quantities $1 - \alpha h_1$ and $1 - \alpha h_2$ are positive. We obtain

$$\Delta t_1 = \frac{1}{1-\alpha h_1}\Delta\tau < \frac{1}{1-\alpha h_2}\Delta\tau = \Delta t_2.$$

Therefore, while h is decreasing, the clock, from C is approaching O and it is ticking slower and slower, *that is the gravity slows down the clocks.* This effect is taken into consideration in the case of GPS systems where we need to have same times at ground level and at the GPS satellite level[3].

9.6.2 Bending of Light-Rays in a Constant Gravitational Field

Theorem 9.6.8 *The light-rays are bending in a constant gravitational field* $-\vec{\alpha}$.

Proof The main idea of the proof that the light is bending in a constant gravitational field is related to the fact that the projection of a line to a plane is a line or a point. For the proof, the trajectory of a photon included in a given plane, in our case $x^3 = \frac{1}{\alpha}$, is transferred into the frame of coordinates $(\tau = \bar{x}^0, \bar{x}^2, \bar{x}^3)$ and then it is projected to the plane $(\bar{x}^2, \bar{x}^3)$. The result is neither a line nor a point. Therefore the light-ray is bent by the constant gravitational field.

Let us focus on G^{-1}, now defined for a three dimensional slice in R. The result is

$$G^{-1} : \begin{cases} \tau(t, x^2, x^3) = \frac{1}{\alpha} \tanh^{-1}\left(\frac{t}{x^3}\right) \\ \bar{x}^2(t, x^2, x^3) = x^2 \\ \bar{x}^3(t, x^2, x^3) = \frac{1}{2\alpha} \ln\left[\alpha^2\left(\left(x^3\right)^2 - t^2\right)\right] \end{cases}$$

and we look at the image of the plane $x^3 = \frac{1}{\alpha}$. In the next formulas, we suppress the (t, x^2) coordinates, therefore

$$G^{-1}\left(x^3 = \frac{1}{\alpha}\right) : \begin{cases} \tau = \frac{1}{\alpha} \tanh^{-1}(\alpha t) \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = \frac{1}{2\alpha} \ln\left(1 - \alpha^2 t^2\right). \end{cases}$$

We observe: $G^{-1}\left(x^3 = \frac{1}{\alpha}\right)$ is a cylinder containing the $\bar{x}^2$ axis.

If we consider the trajectory of a photon in the $x^3 = \frac{1}{\alpha}$ plane, this has to be the line

[3] The acronym GPS stays for Global Positioning System. It is a satellite-based radio-navigation system that provides geolocation and time information to a receiver anywhere on or near the Earth where there is an unobstructed line of sight to the fleet of GPS satellites. Obstacles, such as mountains, block or weaken the GPS signals.

$c(s) = \left(s, s, \dfrac{1}{\alpha}\right)$. The system $G^{-1}(c(s))$ is described by the equations

$$G^{-1}(c(s)) : \begin{cases} \tau = \dfrac{1}{\alpha} \tanh^{-1}(\alpha s) \\ \bar{x}^2 = s \\ \bar{x}^3 = \dfrac{1}{2\alpha} \ln\left(1 - \alpha^2 s^2\right). \end{cases}$$

If we project the previous trajectory of a photon, that is trajectory seen in S to the plane $(\bar{x}^2, \bar{x}^3)$, the result, denoted by $\bar{c}(s)$, is parameterized as

$$\bar{c}(s) = \left(s, \ \frac{1}{2\alpha} \ln\left(1 - \alpha^2 s^2\right)\right).$$

It is obvious that $\bar{c}(s)$ is neither a point nor a line. □

9.6.3 *The Basic Incompatibility Between Gravity and Special Relativity*

We can conclude this chapter pointing out the basic incompatibility between gravity and Special Relativity.

Let us suppose we are in a local frame S, where a constant gravitational field $-\vec{\alpha}$ is acting and let us consider a photon emitted at the origin O from a source moving along the τ-axis. Taking into account the frequency ν and the formula $\Delta\tau = \dfrac{1}{\nu}$, the next photon is emitted by the source at the point $A(\Delta\tau, 0)$. The frame S is not an inertial one and we have proved that the trajectories of photons are bending, that is they are not straight lines but curves. This means that there is a specific curve starting at the emitting point of the photon, in our case O, which reaches the line $\bar{x}^3 = h$ at a point denoted by M. The second photon, emitted in A has an identical trajectory to the one emitted at O. This second trajectory reaches the line $\bar{x}^3 = h$ in a point denoted by N.

The quadrilateral $OANM$ has the property $\Delta\tau = OA = MN$. The length MN is the period Δt corresponding to the frequency f_h in R.
We have

$$\Delta\tau = \Delta t,$$

instead of

$$\Delta\tau = e^{-\alpha h} \Delta t.$$

This contradiction shows that the gravity cannot be integrated into the framework of Special Relativity. Another theory has to be developed in order to fix this shortcoming. This is General Relativity.

Chapter 10
General Relativity and Relativistic Cosmology

Quod erat demostrandum.

An imaginary discussion between Newton and Einstein could be the following.

....*..

Isaac Newton: *Dear Prof. Einstein, my Universe is very simple. I can describe it using vectors and calculus. Between any two objects, a gravitational force is acting and, according to the masses of objects and the distance between them, the gravitational force law is* $F = G\dfrac{mM}{r^2}$. *The gravitational field, in this case, is* $A = \dfrac{GM}{r^2}$. *However, there exists an artifact, the gravitational potential* $\Phi = \dfrac{GM}{r}$. *After me, the brilliant experimental physicist, Henry Cavendish, measured the gravitational constant* $G = 6,67 \times 10^{-11}\mathrm{Nm}^2/\mathrm{kg}^2$, *considered "universal". The potential is related to the gravitational field through the formula* $\nabla\Phi = -\vec{A}$, *the vacuum field equation is* $\nabla^2\Phi = 0$, *as established by Pierre Simon Laplace, and the general gravitational field equation is* $\nabla^2\Phi = 4\pi G\rho$ *as pointed out by Siméon Denis Poisson, once the density of matter is known. The objects are moving in this gravitational field according to* $\vec{F} = m\vec{A}$ *and the trajectories are conics because my gravitational universal law gives a mathematical proof for the Kepler laws. What do you think?*

Albert Einstein: *Very simple indeed, Sir Isaac! Conversely, my Universe is geometric and has four dimensions, it is called space-time! I need more mathematics to describe it. Differential Geometry is essential, but, my dear Sir, this was invented after you passed away! My Universe is expressed by a metric* $ds^2 = g_{ij}dx^i dx^j$, *where the coefficients* g_{ij} *play the role of your gravitational potential* Φ. *The Christoffel symbols* Γ^i_{jk} *are related to your gravitational field* A. *This means that "my gravitational field" has more variables and structures than yours. The vacuum field equations are*

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_10

$$R_{ij} = 0$$

and my general field equations are

$$R_{ij} - \frac{1}{2} R\, g_{ij} = \frac{8\pi G}{c^4} T_{ij}.$$

Starting from them, I can recover the Laplace and Poisson equations in the weak field limit so, my dear Sir Isaac,....I am coherent with your picture! The metric I mentioned before is the one that satisfies the field equations. Objects are always moving on geodesics of the metric, therefore their equations are

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq} \frac{dx^p}{dt} \frac{dx^q}{dt} = 0$$

These geodesic equations are my way of saying $\vec{F} = m\ \vec{A}$ *that I recover, indeed, in the weak field limit. To conclude, one of my collaborators, John Archibald Wheeler, said that the better description of my theory can be reduced to the sentence* **"Space-time tells matter how to move; matter tells space-time how to curve"** [141].

....*..

Let us insist on the last sentence. How the space is curved appears from the Einstein field equations

$$R_{ij} - \frac{1}{2} R\, g_{ij} = \frac{8\pi G}{c^4} T_{ij}.$$

In the left-hand side, we have the "Geometry": Metric g_{ij} *and its derivatives are involved; in the right-hand side, we have a tensor depending on matter, the so called energy-momentum tensor. Once we have a metric* g_{ij}*, according to the Equivalence Principle, we have also the geodesics of the metric as we will discuss below. Which is the meaning of the geodesics described by the equations*

$$\frac{d^2x^r}{dt^2} + \Gamma^r_{pq} \frac{dx^p}{dt} \frac{dx^q}{dt} = 0\,?$$

The simplest answers is: They are trajectories of objects moving accordingly to the Geometry of space-time.

We start this chapter with some general considerations on what a good theory of gravity should do and enunciating the basic principles on which General Relativity lies. After, we take into account the differences between the Classical Newtonian Mechanics and the Einstein picture of gravity based on Geometry. We discuss how it works looking at the differences between the constant gravitational field, as conceived in Classical Mechanics, and the General Relativity counterpart. Finally, we provide Einstein's field equations from the Einstein-Hilbert variational principle and briefly discuss possible generalizations like the so called $f(R)$ *gravity.*

The Schwarzschild solution of the Einstein vacuum field equations is presented. The orbits of planets and the bending of light rays are computed in the framework of Schwarzschild metric. Even if it does not verify the field equations, the Einstein metric is presented because Einstein used it to compute the orbits of planets and the bending of light rays. The full computation for the perihelion drift is presented. The same, in both metrics, is presented the bending of light rays passing near our Sun.

Fermi's viewpoint on Einstein's vacuum field equations is presented with implications related to the study of the week gravitational field; the classical counterparts of the relativistic equations are obtained in this way. We analyze Einstein static universe and the basic considerations on the cosmological constant, as a part related to the standard approach to the General Relativity. A "cosmological metric" is discussed when we study the Friedmann-Lemaître-Robertson-Walker metrics of a Universe in expansion. The way we obtain it is related to the way we considered the energy-momentum tensor. An interesting introductory section devoted to black holes mathematics is also presented. To have a more complete view on Relativity, we offer a short introduction on cosmic strings, wormholes and gravitational waves.

Particular hypothetical universes without global time coordinate, as Gödel's one and without masses, as de Sitter one and not only these, are presented to enlarge the possibilities of solutions of Einstein's field equations.

This is the most important chapter of the book. The main references for the topics we are developing can be found in [40, 44, 46, 56, 58, 75, 77, 137, 141, 142, 154, 157, 163, 169, 171, 176].

10.1 What is a Good Theory of Gravity?

Before entering the details of General Relativity, some considerations are in order. We need them to discuss the change of perspective introduced by the Einstein theory.

As it is well known, General Relativity is based on the fundamental assumption that space and time are entangled into a single space-time structure assigned on a pseudo-Riemann manifold. Being a dynamical structure, it has to reproduce, in the absence of gravitational field, the Minkowski space-time.

General Relativity has to match some minimal requirements to be considered a self-consistent physical theory. First of all, it has to reproduce the Newtonian dynamics in the weak-energy limit, hence it must be able to explain the astronomical dynamics related to the orbits of planets and the self-gravitating structures. Moreover, it passed some observational tests in the Solar System that constitute its experimental foundation [196].

However, General Relativity should be able to explain the Galactic dynamics, taking into account the observed baryonic constituents (e.g. luminous components as stars, sub-luminous components as planets, dust and gas), radiation and Newtonian potential which is, by assumption, extrapolated to Galactic scales. Besides, it should address the problem of large-scale structure as the clustering of galaxies. On cosmological scales, it should address the dynamics of the Universe, which means to reproduce the cosmological parameters as the expansion rate, the density parameter, and so on, in a self-consistent way. Observations and experiments, essentially, probe the standard baryonic matter, the radiation, and an attractive overall interaction, acting at all scales and depending on distance: this interaction is gravity.

In particular, Einstein's General Relativity is based on four main assumptions. They are

> The "*Relativity Principle*" - there is no preferred inertial frames, i.e. all frames are good frames for Physics.

> The "*Equivalence Principle*" - inertial effects are locally indistinguishable from gravitational effects (which means the equivalence between the inertial and the gravitational masses). In other words, any gravitational field can be locally cancelled.

> The "*General Covariance Principle*" - field equations must be "covariant" in form, i.e. they must be invariant in form under the action of space-time diffeomorphisms.

> The "*Causality Principle*" - each point of space-time has to admit a universally valid notion of past, present and future.

On these bases, Einstein postulated that, in a four-dimensional space-time manifold, the gravitational field is described in terms of the metric tensor field $ds^2 = g_{ij}dx^i dx^j$, with the same signature of Minkowski metric. The metric coefficients have the physical meaning of gravitational potentials. Moreover, he postulated that space-time is curved by the distribution of the energy–matter sources.

The above principles require that the space-time structure has to be determined by either one or both of the two following fields: a Lorentzian metric g and a linear connection Γ, assumed by Einstein to be torsionless. The metric g fixes the causal structure of space-time (the light cones) as well as its metric relations (clocks and rods); the connection Γ fixes the free fall, i.e. the locally inertial observers. They have, of course, to satisfy a number of compatibility relations which amount to require that photons follow null geodesics of Γ, so that Γ and g can be independent, a priori, but constrained, a posteriori, by some physical restrictions. These, however, do not impose that Γ has necessarily to be the Levi-Civita connection of g.

It should be mentioned, however, that there are many shortcomings in General Relativity, both from a theoretical point of view (non-renormalizability, the presence of singularities, and so on), and from an observational point of view. The latter indeed clearly shows that General Relativity is no longer capable of addressing Galactic, extra-galactic, and cosmic dynamics, unless the source side of field equations contains some exotic form of matter–energy. These new elusive ingredients, as mentioned above, are usually addressed as "*dark matter*" and *dark energy* and constitute up to the 95% of the total cosmological amount of matter–energy [52].

On the other hand, instead of changing the source side of the Einstein field equations, one can ask for a "geometrical view" to fit the missing matter–energy of the observed Universe. In such a case, the dark side could be addressed by extending General Relativity including more geometric invariants into the standard Einstein–Hilbert Action. Such effective Lagrangians can be easily justified at fundamental level by any quantization scheme on curved space-times. However, at present stage of the research, this is nothing else but a matter of taste, since no final probe discriminating between dark matter and extended gravity has been found up to now. Finally, the bulk of observations that should be considered is so high that an effective Lagrangian or a single particle will be difficult to account for the whole phenomenology at all astrophysical and cosmic scales.

10.1.1 Metric or Connections?

As we will see below, in the General Relativity formulation, Einstein assumed that the metric g of the space-time is the fundamental object to describe gravity. The connection Γ is constituted by coefficients with no dynamics. Only g has dynamics. This means that the single object g determines, at the same time, the causal structure (light cones), the measurements (rods and clocks) and the free fall of test particles (geodesics). Space-time is therefore a couple $\{M, g\}$ constituted by a pseudo-Riemannian manifold and a metric. Even if it was clear to Einstein that gravity induces freely falling observers and that the Equivalence Principle selects an object that cannot be a tensor (the connection Γ)—since it can be switched off and set to zero at least in a point—he was obliged to choose it (the Levi-Civita connection) as being determined by the metric structure itself.

In the Palatini formalism, a (symmetric) connection Γ and a metric g are given and varied independently. Space-time is a triple $\{M, g, \Gamma\}$ where the metric determines rods and clocks (i.e. it sets the fundamental measurements of space-time) while Γ determines the free fall. In the Palatini formalism, Γ are differential equations. The fact that Γ is the Levi-Civita connection of g is no longer an assumption but becomes an outcome of the field equations.

The connection is the gravitational field and, as such, it is the fundamental field in the Lagrangian. The metric g enters the Lagrangian with an "ancillary" role. It reflects the fundamental need to define lengths and distances, as well as areas and volumes. It defines rods and clocks that we use to make experiments. It defines also the causal structure of space-time. However, it has no dynamical role. There is no whatsoever reason to assume g to be the potential for Γ, nor that it has to be a true field just because it appears in the action. We will not develop any more the Palatini formalism in this book. For a detailed discussion see [56].

10.1.2 The Role of Equivalence Principle

The Equivalence Principle is strictly related to the above considerations and could play a very relevant role in order to discriminate among theories. In particular, it could specify the role of g and Γ selecting between the metric and Palatini formulation of gravity. In particular, precise measurements of Equivalence Principle could say us if

Γ is only Levi-Civita or a more general connection disentangled, in principle, from g. Before, we discussed the Equivalence Principle starting from the early Galileo consideration stating that $m_i \equiv m_g$. Besides this result, in General Relativity, Equivalence Principle states that accelerations can be set to zero in given reference frame. According to this result, the free fall along geodesics, given by the connection, is ruled by the metric, as we will discuss below.

Before entering into details, let us discuss some topics related to the Equivalence Principle. Summarizing, the relevance of this principle comes from the following points:

- Competing theories of gravity can be discriminated according to the validity of Equivalence Principle.
- Equivalence Principle holds at classical level but it could be violated at quantum level.
- Equivalence Principle allows to investigate independently geodesic and causal structure of space-time.

From a theoretical point of view, Equivalence Principle lies at the physical foundation of metric theories of gravity. The first formulation of Equivalence Principle comes out from the theory of gravitation formulates by Galileo and Newton, i.e. the Weak Equivalence Principle (the above Galilean Equivalence Principle) which asserts the inertial mass m_i and the gravitational mass m_g of any physical object are equivalent. The Weak Equivalence Principle statement implies that it is impossible to distinguish, locally, between the effects of a gravitational field from those experienced in uniformly accelerated frames using the simple observation of the free-falling particles behaviour.

A generalization of Weak Equivalence Principle claims that Special Relativity is locally valid. Einstein realized, after the formulation of Special Relativity, that the mass can be reduced to a manifestation of energy and momentum as discussed in previous chapter. As a consequence, it is impossible to distinguish between a uniform acceleration and an external gravitational field, not only for free-falling particles, but whatever is the experiment. According to this observation, Einstein Equivalence Principle states:

- The Weak Equivalence Principle is valid.
- The outcome of any local non-gravitational test experiment is independent of the velocity of free-falling apparatus.
- The outcome of any local non-gravitational test experiment is independent of where and when it is performed in the Universe.

One defines as "local non-gravitational experiment" an experiment performed in a small size of a free-falling laboratory. Immediately, it is possible to realize that the gravitational interaction depends on the curvature of space-time, i.e. the postulates of any metric theory of gravity have to be satisfied. Hence the following statements hold:

- Space-time is endowed with a metric g_{ij}.
- The world lines of test bodies are geodesics of the metric.
- In local freely falling frames, called local Lorentz frames, the non-gravitational laws of physics are those of Special Relativity.

One of the predictions of this principle is the gravitational redshift, experimentally verified by Pound and Rebka in 1960 [196]. Notice that gravitational interactions are excluded from the Weak Equivalence Principle and the Einstein Equivalence Principle.

In order to classify alternative theories of gravity, the gravitational Weak Equivalence Principle and the Strong Equivalence Principle have to be introduced. On the other hand, the Strong Equivalence Principle extends the Einstein Equivalence Principle by including all the laws of physics in its terms. That is:

- Weak Equivalence Principle is valid for self-gravitating bodies as well as for test bodies (gravitational weak equivalence principle).
- The outcome of any local test experiment is independent of the velocity of the free-falling apparatus.
- The outcome of any local test experiment is independent of where and when in the Universe it is performed.

Alternatively, the Einstein Equivalence Principle is recovered from the Strong Equivalence Principle as soon as the gravitational forces are neglected. Many authors claim that the only theory coherent with Strong Equivalence Principle is General Relativity.

A very important issue is the consistency of Equivalence Principle with respect to the Quantum Mechanics. General Relativity is not the only theory of gravitation and several alternative theories of gravity have been investigated from the 1960s of last century. Considering the space-time to be special relativistic at a background level, gravitation can be treated as a Lorentz-invariant field on the background. Assuming the possibility of General Relativity extensions, two different classes of experiments can be conceived:

- Tests for the foundations of gravitational theories considering the various formulations of Equivalence Principle.
- Tests of metric theories where space-time is a priori endowed with a metric tensor and where the Einstein Equivalence Principle is assumed always valid.

The subtle difference between the two classes of experiments lies on the fact that Equivalence Principle can be postulated a priori or, in a certain sense, "recovered" from the self-consistency of the theory. What is today clear is that, for several fundamental reasons, extra fields are necessary to describe gravity with respect to the other interactions. Such fields can be scalar fields or higher order corrections of curvature invariants. For these reasons, two sets of field equations can be considered: The first set couples the gravitational field to the non-gravitational contents of the Universe, i.e. the matter distribution, the electromagnetic fields, etc. The second set of equations gives the evolution of non-gravitational fields. Within the framework of metric theories, these laws depend only on the metric and this is a consequence of the Einstein Equivalence Principle. In the case where Einstein field equations are modified and matter field are minimally coupled with gravity, we are dealing with the so-called *Jordan frame*. In the case where Einstein field equations are preserved and matter field are non-minimally coupled, we are dealing with the so-called *Einstein frame*. Both frames are conformally related but the very final issue is to understand if passing from one frame to the other (and vice versa) is physically significant. See [56] for details. Clearly, Equivalence Principle plays a fundamental role in this discussion. In particular, the question is if it is always valid or it can be violated at quantum level. See [12, 182, 184].

After these preliminary considerations, let us start with the geometric construction of General Relativity. However, we recommend the reader to consider again these introductory sections after he/she finishes to read the book because some current problems in General Relativity are reported.

10.2 Gravity Seen Through Geometry in General Relativity

Even if we repeat some ideas, let us go back to our previous discussion on the gravitational potential in Newtonian Mechanics. We start from the tidal acceleration equations

$$\frac{d^2}{dt^2}\frac{\partial \bar{x}}{\partial q} = -d^2\Phi_{\bar{x}}\frac{\partial \bar{x}}{\partial q},$$

where the Hessian matrix of the gravitational potential Φ

$$d^2\Phi_{\bar{x}} = \left(\frac{\partial^2\Phi(\bar{x})}{\partial x_i \partial x_k}\right)_{i,k}$$

is encapsulated in its trace by the Laplace equation $\nabla^2\Phi = 0$ in vacuum. In a space endowed with a metric $ds^2 = g_{ij}dx^i dx^j$, it is possible to find the equivalent

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -K^h_j\frac{\partial x^j}{\partial q},$$

where

$$K^h_j = R^h_{ijk}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}$$

plays the role of the Hessian of the gravitational potential.

It seems to be natural to think of the trace of the matrix K^h_j to obtain an equivalent of classical vacuum field equations $\nabla^2\Phi = 0$. Since

$$K_h^h = R_{ijh}^h \frac{dx^i}{d\tau} \frac{dx^j}{d\tau},$$

the Ricci tensor has to be involved in General Relativity field equations. It is why Einstein, and then Hilbert, considered a way to express the field equations through the Ricci tensor.

So, let us repeat their main idea. The gravitational field is not constant. There are small variations of the gravitational field induced by some other bodies or by changing the distance r between bodies. If we are on the surface of the Earth, our legs will experience a higher intensity of the gravitational field of the Earth than our head. To understand this, it is enough to look at the formula $A = \frac{GM}{r^2}$, M being the mass of the Earth, G being the gravitational constant and r being the radius R of the Earth at the legs level and $r = R + h$ at the level of our head, h being our height. For the same reason, a person at the first floor of a building experiences a greater intensity of the gravitational field comparing with another person which is at the 33th floor of the same building. The Moon makes ocean tides and we see how these are related to the tidal effects.

If we have tides, mathematically they can be treated under the Newtonian standard, the field equation $\nabla^2 \Phi = 0$ being hidden in the trace of the Hessian matrix $d^2\Phi$ involved in the tidal equations

$$\frac{d^2}{dt^2} \frac{\partial \bar{x}}{\partial q} = -d^2 \Phi_{\bar{x}} \frac{\partial \bar{x}}{\partial q}.$$

Tides can be dealt with a geometric approach considering the Ricci tensor of a given metric from the equations

$$\frac{\nabla^2}{d\tau^2} \frac{\partial x^h}{\partial q} = -K_j^h \frac{\partial x^j}{\partial q},$$

where

$$K_j^h = R_{ijk}^h \frac{dx^i}{d\tau} \frac{dx^k}{d\tau}.$$

Einstein had the power to break the standard Newtonian approach, describing gravity with the language of Differential Geometry.

According to Einstein, the components g_{ij} of a metric $ds^2 = g_{ij} dx^i dx^j$ play the role of gravitational potential Φ, which is just one of the potentials in g_{ij}. The Christoffel symbols Γ_{jk}^i play the role of the gravitational field $\vec{A}$. Let us consider again a table of analogies containing the two ways of conceiving at the gravity

$$\begin{array}{ccc} \textit{Newton} & & \textit{Einstein} \\ \Phi & \longleftrightarrow & g_{ij} \\ \vec{A} & \longleftrightarrow & \Gamma^i_j k \\ \nabla^2 \Phi = 0 & \longleftrightarrow & ? \\ \nabla^2 \Phi = 4\pi G \rho & \longleftrightarrow & ? \end{array}$$

The first question mark seems to be replaced by $R_{ij} = 0$, but we still have to work to obtain it. At this moment we know that the second question mark has to be replaced by the Einstein field equations. Because the meaning of the two ways of conceiving the gravity will be clarified latter after we introduce Fermi coordinates.

Einstein was the first who realized that the laws of Nature have to be expressed by equations which hold for any system of coordinates, that is, they must be covariant with respect to any change of coordinates.

Taking into account also the discussion of previous sections, *Einstein's Principle of General Covariance* states:

The laws of Nature have to be expressed as equalities of different tensors.
The changes of coordinates become part of the core of General Relativity. Why are they so important? They allow us to describe the laws of Nature from the point of view of different observers or/and they allow us to describe a new state of a given system.

Let us consider a region of space where the gravitational field can be neglected. Consider a spacecraft there. Suppose that there are no other forces acting there. Therefore all objects are moving on straight lines with constant velocity. The spacecraft does the same. Locally, the space-time system of coordinates (x^0, x^1, x^2, x^3) can be thought to describe an inertial frame, that is, the local metric tensor is the Minkowski one

$$g_{ij}(x^0, x^1, x^2, x^3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Since $\Gamma^i_{jk} = 0$, the geodesics equations are

$$\ddot{x}^j(t) = 0, \;\; j \in \{0, 1, 2, 3\},$$

i.e. all objects there experience a free fall. So, the law of motion is described by the previous equations which express in fact the equality $\vec{F} = m \cdot \vec{A}$ for $\vec{F} = \vec{0}$.

Let us now suppose the engines of the spacecraft start and the space craft is accelerated. This is described by a map M which switches from the coordinates

$(\bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3)$ to (x^0, x^1, x^2, x^3), i.e. we have to describe the old coordinates with respect to the new ones.

We know, from Differential Geometry, how the new metric looks like.

The new components $\bar{g}_{ij}$ are found after the rule $dM^t \cdot (g_{ij}) \cdot dM$. In this new metric, we can compute the new $\bar{\Gamma}^i_{jk}$ and the geodesic equations are

$$\ddot{\bar{x}}^i(t) + \bar{\Gamma}^i_{jk}\dot{\bar{x}}^j(t)\dot{\bar{x}}^j(t) = 0,\ \ j \in \{0, 1, 2, 3\}.$$

Since under a change of coordinates, geodesics are transformed into geodesics, and the meaning is kept, the old law of motion becomes the new law of motion, therefore the equations

$$\ddot{\bar{x}}^i(t) = -\bar{\Gamma}^i_{jk}\dot{\bar{x}}^j(t)\dot{\bar{x}}^j(t),\ \ j \in \{0, 1, 2, 3\}$$

describes $\vec{F} = m \cdot \vec{A}$ for $\vec{F} \neq \vec{0}$.

Let us try to understand the constant gravitational field under this more general approach.

10.2.1 The Einstein Landscape for the Constant Gravitational Field

We consider a local frame of coordinates S, $(\tau = \bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3)$ in which acts a constant gravitational field and another frame of coordinates R, whose coordinates are $(t = x^0, x^1, x^2, x^3)$, frame which is in free fall with respect to the previous constant gravitational field. The metric in the second frame is the Minkowski one,

$$g_{ij}(x^0, x^1, x^2, x^3) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Let us assume that, for $\tau = t = 0$, the two frames can be seen as an unique frame with axes corresponding in index notation. We assume also that the second frame is moving along the $\bar{x}^3$-axis in its negative direction. We saw already the transformation which involves the constant gravitational field $-\alpha$ in S. It is, in fact, a change of coordinates between S and R. If we consider only the pairs of axis $(\tau, \bar{x}^3)$ and (t, x^3), this is

$$G : \begin{cases} t(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \sinh \alpha\tau \\ x^3(\tau, \bar{x}^3) = \dfrac{e^{\alpha \bar{x}^3}}{\alpha} \cosh \alpha\tau. \end{cases}$$

with

$$G^{-1}: \begin{cases} \tau(t, x^3) = \dfrac{1}{\alpha} \tanh^{-1}\left(\dfrac{t}{x^3}\right) \\ \bar{x}^3(t, x^3) = \dfrac{1}{2\alpha} \ln\left[\alpha^2\left(\left(x^3\right)^2 - t^2\right)\right]. \end{cases}$$

In the considered slice, in R, the metric is

$$\begin{pmatrix} g_{00} & g_{03} \\ g_{30} & g_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The metric in S is determined by $dG^t \cdot (g_{ij}) \cdot dG$, where

$$dG = dG^t = \begin{pmatrix} e^{\alpha\bar{x}^3} \cosh\alpha\tau & e^{\alpha\bar{x}^3} \sinh\alpha\tau \\ e^{\alpha\bar{x}^3} \sinh\alpha\tau & e^{\alpha\bar{x}^3} \cosh\alpha\tau \end{pmatrix} = e^{\alpha\bar{x}^3} \begin{pmatrix} \cosh\alpha\tau & \sinh\alpha\tau \\ \sinh\alpha\tau & \cosh\alpha\tau \end{pmatrix}.$$

It results in

$$\begin{pmatrix} \bar{g}_{00} & \bar{g}_{03} \\ \bar{g}_{30} & \bar{g}_{33} \end{pmatrix} = \begin{pmatrix} e^{2\alpha\bar{x}^3} & 0 \\ 0 & -e^{2\alpha\bar{x}^3} \end{pmatrix}.$$

The metric which describes the constant gravitational field in the corresponding slice of S is

$$ds^2 = e^{2\alpha\bar{x}^3}\left[d\bar{x}^0 - d\bar{x}^3\right].$$

In S, locally, the metric tensor is

$$\bar{g}_{ij}(\bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3) = \begin{pmatrix} e^{2\alpha\bar{x}^3} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -e^{2\alpha\bar{x}^3} \end{pmatrix}.$$

The Christoffel first kind symbols are

$$\Gamma_{30,0} = \Gamma_{03,0} = \alpha e^{2\alpha\bar{x}^3},\ \Gamma_{00,3} = \Gamma_{33,3} = -\alpha e^{2\alpha\bar{x}^3},\ \Gamma_{30,3} = \Gamma_{03,3} = \Gamma_{00,0} = \Gamma_{33,0} = 0.$$

The Christoffel second kind symbols are

$$\Gamma^0_{30} = \Gamma^0_{03} = \Gamma^3_{00} = \Gamma^3_{33} = \alpha,\ \Gamma^3_{30} = \Gamma^3_{03} = \Gamma^0_{00} = \Gamma^0_{33} = 0.$$

The geodesic equations, in the considered slice, with respect to the geodesic parameter λ are

$$\begin{cases} \dfrac{d^2\bar{x}^0}{d\lambda^2} = -2\alpha \dfrac{d\bar{x}^0}{d\lambda}\dfrac{d\bar{x}^3}{d\lambda} \\ \dfrac{d^2\bar{x}^3}{d\lambda^2} = -\alpha\left(\dfrac{d\bar{x}^0}{d\lambda}\right)^2 - \alpha\left(\dfrac{d\bar{x}^3}{d\lambda}\right)^2. \end{cases}$$

In S, we have

$$\begin{cases} \dfrac{d^2\bar{x}^0}{d\lambda^2} = -2\alpha \dfrac{d\bar{x}^0}{d\lambda}\dfrac{d\bar{x}^3}{d\lambda} \\ \dfrac{d^2\bar{x}^1}{d\lambda^2} = 0 \\ \dfrac{d^2\bar{x}^2}{d\lambda^2} = 0 \\ \dfrac{d^2\bar{x}^3}{d\lambda^2} = -\alpha\left(\dfrac{d\bar{x}^0}{d\lambda}\right)^2 - \alpha\left(\dfrac{d\bar{x}^3}{d\lambda}\right)^2 \end{cases}$$

with the general solutions

$$\begin{cases} \bar{x}^0 = k_3 + \dfrac{1}{2\alpha}\ln(k_1+\lambda) - \dfrac{1}{2\alpha}\ln(k_2-\lambda) \\ \bar{x}^1 = k_4\lambda + k_5 \\ \bar{x}^2 = k_6\lambda + k_7 \\ \bar{x}^3 = \dfrac{1}{\alpha}\ln\alpha + \dfrac{1}{2\alpha}\ln(k_1+\lambda) + \dfrac{1}{2\alpha}\ln(k_2-\lambda). \end{cases}$$

A very good exercise for the reader is to prove that the above formulas verify the equations of the geodesics.

Let us analyse the trajectories of photons. They come from the equations $x^3 = t + b$ or $x^3 = -t + b$. The constant b is arbitrary and the speed of light is assumed 1. We consider only the case $x^3 = t + b$, the other case can be analysed in a similar way.

Let us introduce the previous formula in G^{-1}. It results in

$$G^{-1}: \begin{cases} \tau(t) = \dfrac{1}{\alpha}\tanh^{-1}\left(\dfrac{t}{t+b}\right) = \dfrac{1}{2\alpha}\ln\dfrac{1+\dfrac{t}{t+b}}{1-\dfrac{t}{t+b}} = \dfrac{1}{2\alpha}\ln(2t+b) - \dfrac{1}{2\alpha}\ln b \\ \bar{x}^3(t) = \dfrac{1}{2\alpha}\ln\left[\alpha^2\left((t+b)^2 - t^2\right)\right] = \dfrac{1}{2\alpha}\ln(2t+b) + \dfrac{1}{\alpha}\ln\alpha, \end{cases}$$

that is,

$$\bar{x}^3(t) = \tau(t) + \beta,$$

where β is a constant. The trajectories of photons are lines having the slope $+1$ (or -1). Of course these lines are geodesics because they come from the geodesics of R.

In the case $x^3 = k$, let us express $\bar{x}^3$ as a function of τ. From

$$\tau(t) = \frac{1}{\alpha}\tanh^{-1}\left(\frac{t}{k}\right)$$

it results in

$$t = k\tanh(\alpha\tau),$$

that is,

$$\bar{x}^3(\tau) = \frac{1}{2\alpha}\ln\left[\alpha^2k^2(1-\tanh^2(\alpha\tau))\right].$$

Therefore

$$\bar{x}^3(\tau) = \frac{1}{\alpha}\ln(\alpha k) - \frac{1}{\alpha}\ln(\cosh(\alpha\tau)).$$

Since

$$\bar{x}^3(0) = \frac{1}{\alpha}\ln(\alpha k);\ \frac{d\bar{x}^3}{d\tau}(0) = 0;\ \frac{d^2\bar{x}^3}{d\tau^2}(0) = -\alpha;$$

the second-order approximation of $\bar{x}^3$ is the parabola

$$\bar{x}^3(\tau) = \frac{1}{\alpha}\ln(\alpha k) - \frac{\alpha}{2}\tau^2,$$

which, in the case $k = \frac{1}{\alpha}$ and $\tau = \frac{\tau_1}{v}$, becomes

$$\bar{x}^3(\tau_1) = -\frac{\alpha}{2v^2}\tau_1^2.$$

This is the parabola seen in the case of constant gravitational field in Classical Mechanics, that is, the trajectory function of time.

Since the second kind Christoffel symbols are constant, it is easy to compute R^0_{303}. We find $R^0_{303} = \Gamma^0_{h0}\Gamma^h_{33} - \Gamma^0_{h3}\Gamma^h_{30} = \alpha^2 - \alpha^2 = 0$.

In fact, all sectional curvatures are 0, but, in general, the geodesics are not straight lines as we saw, they only come from lines of R.

In simple words, we can say that the constant gravitational field bends geodesics of space.

How the constant gravitational field affects the proper time can be found out by looking at the metrics involved in this description. For the frame R, free falling in the constant gravitational field $-\alpha$ of S, the metric is the Minkowski one, i.e.

$$ds^2 = dt^2 - dx^2.$$

The clock ticks in Δt seconds. The constant gravitational field induces in S, as we saw, the metric

$$ds^2 = e^{2\alpha\bar{x}}(d\tau^2 - d\bar{x}^2).$$

Here, the clock ticks in $\Delta\tau$ seconds. Between the observers of R and S, if $\bar{x} \to 0,\ \bar{x} > 0$, there is the connection

$$\Delta t = e^{\alpha\bar{x}}\Delta\tau \geq (1 + \alpha\bar{x})\Delta\tau \geq \Delta\tau,$$

that is, the clock of R ticks slower and slower as $\bar{x} \to 0,\ \bar{x} > 0$.

The clock of a person A at the ground level of a building ticks less than the clock of a person B at the 33th floor. Therefore the ground level person A ages slower than the person B. Or, everyone legs are younger than the brain. Of course even at the level of lifetime of a person, the effects are imperceptible.

Therefore, according to Newton, the constant gravitational field landscape exists in $n = 3$ dimensions. The gravitational field is $\vec{A}$ and the gravitational potential Φ is related to it by the formula

$$\vec{A} = -\nabla\Phi = (0, 0, -\alpha).$$

The constant gravitational field satisfies the vacuum field equation

$$\nabla^2\Phi = 0.$$

The equations of motion are

$$\frac{d^2x}{dt^2} = 0;\ \frac{d^2y}{dt^2} = 0;\ \frac{d^2z}{dt^2} = -\alpha.$$

The solution, in appropriate initial conditions if we consider a plane (t, z), is

$$z(t) = -\frac{\alpha}{2v^2}t^2.$$

Einstein's constant gravitational field landscape exists in four dimensions.

The gravitational potential appears in the coefficients of the metric tensor

$$\bar{g}_{ij}(\bar{x}^0, \bar{x}^1, \bar{x}^2, \bar{x}^3) = \begin{pmatrix} e^{2\alpha\bar{x}^3} & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -e^{2\alpha\bar{x}^3} \end{pmatrix}.$$

The gravitational field is described by the Christoffel second kind symbols:

$$\Gamma^0_{30} = \Gamma^0_{03} = \Gamma^3_{00} = \Gamma^3_{33} = \alpha,\ \Gamma^3_{30} = \Gamma^3_{03} = \Gamma^0_{00} = \Gamma^0_{33} = 0$$

and satisfies

$$R_{ij} = 0.$$

The equations of motions are the geodesic equations:

$$\begin{cases} \dfrac{d^2\bar{x}^0}{d\lambda^2} = -2\alpha \dfrac{d\bar{x}^0}{d\lambda}\dfrac{d\bar{x}^3}{d\lambda} \\ \dfrac{d^2\bar{x}^1}{d\lambda^2} = 0 \\ \dfrac{d^2\bar{x}^2}{d\lambda^2} = 0 \\ \dfrac{d^2\bar{x}^3}{d\lambda^2} = -\alpha\left(\dfrac{d\bar{x}^0}{d\lambda}\right)^2 - \alpha\left(\dfrac{d\bar{x}^3}{d\lambda}\right)^2 \end{cases}$$

with the general solutions

$$\begin{cases} \bar{x}^0 = k_3 + \dfrac{1}{2\alpha}\ln(k_1+\lambda) - \dfrac{1}{2\alpha}\ln(k_2-\lambda) \\ \bar{x}^1 = k_4\lambda + k_5 \\ \bar{x}^2 = k_6\lambda + k_7 \\ \bar{x}^3 = \dfrac{1}{\alpha}\ln\alpha + \dfrac{1}{2\alpha}\ln(k_1+\lambda) + \dfrac{1}{2\alpha}\ln(k_2-\lambda). \end{cases}$$

Locally, the particular solution presented before,

$$\bar{x}^3(\tau) = \frac{1}{\alpha}\ln(\alpha k) - \frac{\alpha}{2}\tau^2,$$

can be approximated by the classical solution

$$\bar{x}^3(\tau_1) = \frac{\alpha}{2v^2}\tau_1^2.$$

This intuitive description of Einstein's pictures can be fully formalized considering the Hilbert approach by which the gravitational field equations come out from a variational principle.

10.3 The Einstein–Hilbert Action and The Einstein Field Equations

Under a change of coordinates $x^r = x^r(\overline{x}^h)$, $r \in \{0, 1, ..., n\}$, $h \in \{0, 1, ..., n\}$, the second kind Christoffel symbols change according to the rule

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\Gamma^k_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\Gamma}^r_{ij} \frac{\partial x^k}{\partial \overline{x}^r}.$$

Suppose we vary the metric. It means that the coefficients g_{ij} are changed in some new coefficients $\bar{g}_{ij} := g_{ij} + \delta g_{ij}$. This second metric produces first-type and second-type Christoffel symbols. Let us denote them by $\gamma_{ij,k}$ and γ^i_{jk}. The same change of coordinates gives for these new Christoffel symbols a similar formula

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\gamma^k_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\gamma}^r_{ij} \frac{\partial x^k}{\partial \overline{x}^r}.$$

The difference of the previous formulas leads to

Proposition 10.3.1 *The variation difference* $\delta\Gamma^i_{jk} := \Gamma^i_{jk} - \gamma^i_{jk}$ *satisfies*

$$\delta\Gamma^k_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} = \delta\overline{\Gamma}^r_{ij} \frac{\partial x^k}{\partial \overline{x}^r},$$

i.e. $\delta\Gamma^i_{jk}$ *is a* $(1, 2)$ *mixed tensor.*

Let g_{ij} be the matrix of the metric $ds^2 = g_{ij}dx^i dx^j$ and let g be the determinant of g_{ij}. Suppose this determinant is negative as in the case of the Minkowski metric.

Theorem 10.3.2 *The formula which expresses the variation of* $\sqrt{-g}$ *is*

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\, g_{ij}\delta g^{ij}.$$

Proof The inverse of the matrix g_{ij} is g^{ij} such that $g^{is}g_{sj} = \delta^i_j$.

Consider a given element g_{ij} of the matrix and denote by M^{ij} the determinant of the matrix obtained from the initial one after we cancel both the line i and the column j.

The corresponding inverse element is $g^{ij} = \dfrac{(-1)^{i+j}M^{ij}}{g}$ and, in this respect, using the column j, the determinant can be thought as $g = \sum_i (-1)^{i+j} g_{ij} M^{ij}$.

Then, for the variation of $\sqrt{-g}$, we have:

$$\delta\sqrt{-g} = \frac{\partial}{\partial g_{ij}}\left(\sqrt{-g}\right)\delta g_{ij} =$$

$$= -\frac{1}{2\sqrt{-g}}\frac{\partial g}{\partial g_{ij}}\delta g_{ij} = -\frac{1}{2\sqrt{-g}}(-1)^{i+j}M^{ij}\delta g_{ij} = -\frac{1}{2\sqrt{-g}}\cdot g\cdot g^{ij}\delta g_{ij}.$$

It results in

$$\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}g^{ij}\delta g_{ij}.$$

From $g^{ks}g_{sl} = \delta^k_l$, it is $\delta g^{ks}g_{sl} + g^{ks}\delta g_{sl} = 0$, that is, $g^{ks}\delta g_{sl} = -\delta g^{ks}g_{sl}$.
Multiplying by g_{mk}, we obtain

$$g_{mk}g^{ks}\delta g_{sl} = -g_{mk}g_{sl}\delta g^{ks}$$

and, after considering $s = m = i,\ l = j$, it is

$$\delta g_{ij} = -g_{ik}g_{ij}\delta g^{ik}.$$

Replacing in the formula of the variation of $\sqrt{-g}$ we obtain

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g^{ij}g_{ik}g_{ji}\delta g^{ik},$$

that is,

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{ik}\delta g^{ik}.$$

□

Theorem 10.3.3 (Palatini's Formula) $\delta R_{ij} = \delta\Gamma^s_{ij;s} - \delta\Gamma^s_{is;j}$.

Proof We start from

$$R_{ij} = R^s_{isj} = \frac{\partial\Gamma^s_{ij}}{\partial x^s} - \frac{\partial\Gamma^s_{is}}{\partial x^j} + \Gamma^s_{su}\Gamma^u_{ij} - \Gamma^s_{ju}\Gamma^u_{is}.$$

Then the variation of the Ricci tensor is

$$\delta R_{ij} = \frac{\partial(\delta\Gamma^s_{ij})}{\partial x^s} - \frac{\partial\left(\delta\Gamma^s_{is}\right)}{\partial x^j} + \delta\Gamma^s_{su}\Gamma^u_{ij} + \Gamma^s_{su}\delta\Gamma^u_{ij} - \delta\Gamma^s_{ju}\Gamma^u_{is} - \Gamma^s_{ju}\delta\Gamma^u_{is}.$$

The variation $\delta\Gamma^s_{ij}$ is a $(1, 2)$ tensor type. Its covariant derivative is

$$\delta\Gamma^s_{ij;s} = \frac{\partial(\delta\Gamma^s_{ij})}{\partial x^s} + \Gamma^s_{su}\delta\Gamma^u_{ij} - \delta\Gamma^s_{ju}\Gamma^u_{is} - \delta\Gamma^s_{iu}\Gamma^u_{js}.$$

In the same way the covariant derivative of $\delta\Gamma^s_{is}$ is

$$\delta\Gamma^s_{is;j} = \frac{\partial(\delta\Gamma^s_{is})}{\partial x^j} + \delta\Gamma^u_{is}\Gamma^s_{ju} - \delta\Gamma^s_{su}\Gamma^u_{ij} - \delta\Gamma^s_{iu}\Gamma^u_{js}.$$

Subtracting the second relation from the first we obtain the Palatini formula. □

Theorem 10.3.4 *If V is a compact region of the Universe whose volume element is dV, such that on its boundary ∂V, the variations $\delta\Gamma^i_{jk}$ vanish, then*

$$\int_V g^{ij}\ \delta R_{ij}\ dV = 0.$$

Proof (*Palatini's Formula Consequence*) Since the volume element dV is expressed with respect to the given metric by

$$dV = \sqrt{-g}\ dx_0 dx_1 dx_2 dx_3,$$

our integral becomes

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x,$$

where we denoted $dx_0dx_1dx_2dx_3$ by d^4x.

It exists a corresponding 3D-surface element $d\sigma$ on ∂V, $d\sigma = \sqrt{-g'}\ d^3x$.

At each point of ∂V, it exists a normal outward vector n of components n_s, i.e. $n = n_s$.

All these results help us to express the divergence formula which, in the classical form, looks like

$$\int_V \div B\ dV = \int_{\partial V} B \cdot n\ d\sigma,$$

here, in its covariant form, being

$$\int_V B^s_{;s}\sqrt{-g}\ d^4x = \int_{\partial V} B^s n_s \sqrt{-g'}\ d^3x.$$

Now, Palatini's formula leads to

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x = \int_V g^{ij}\ \left(\delta\Gamma^s_{ij;s} - \delta\Gamma^s_{is;j}\right)\sqrt{-g}\ d^4x\ .$$

Taking into account that $g^{ij}_{;s} = 0$ and changing the dummy indexes, we can write

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x = \int_V \left[\left(g^{ij}\,\delta\Gamma^s_{ij} \right)_{;s} - \left(g^{ij}\,\delta\Gamma^s_{is} \right)_{;j} \right] \sqrt{-g}\ d^4x =$$

$$= \int_V \left[\left(g^{ij}\,\delta\Gamma^s_{ij} \right)_{;s} - \left(g^{is}\,\delta\Gamma^j_{ij} \right)_{;s} \right] \sqrt{-g}\ d^4x = \int_V \left[g^{ij}\,\delta\Gamma^s_{ij} - g^{is}\,\delta\Gamma^j_{ij} \right]_{;s} \sqrt{-g}\ d^4x.$$

Let us denote the contravariant vector $g^{ij}\delta\Gamma^s_{ij} - g^{is}\delta\Gamma^j_{ij}$ by B^s. Our initial integral

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x,$$

according to the covariant above form becomes

$$\int_{\partial V} B^s n_s\ \sqrt{-g'}\ d^3x.$$

Since $B^s = g^{ij}\delta\Gamma^s_{ij} - g^{is}\delta\Gamma^j_{ij}$ vanishes on ∂V, the last integral is 0, that is

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x = 0.$$

□

These considerations lead us to the following.

Theorem 10.3.5 (Einstein's Field Equations in vacuum) *If V is a compact region of the Universe without matter and energy inside it, such that, on its boundary ∂V, the variations $\delta\Gamma^i_{jk}$ vanish, then*

$$R_{ij} - \frac{1}{2} R\ g_{ij} = 0.$$

Proof (*Hilbert*) We have proved both

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\ g_{ij}\delta g^{ij}$$

and

$$\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\ d^4x = 0.$$

To derive Einstein's field equations, we have to choose an appropriate Lagrangian.

Hilbert's idea was to consider the Lagrangian expressed through the Ricci curvature scalar R, that is, the *Einstein–Hilbert action* is

$$S_{EH} = \int_V R\sqrt{-g}\ d^4x.$$

Let us compute the first variation of S_{EH}. It is

$$\delta S_{EH} = \delta \int_V R\sqrt{-g}\, d^4x = \delta \int_V g^{ij} R_{ij}\sqrt{-g}\, d^4x = \int_V \delta\left(g^{ij} R_{ij}\sqrt{-g}\right) d^4x =$$

$$= \int_V \left(\delta g^{ij}\right) R_{ij}\sqrt{-g}\, d^4x + \int_V g^{ij}\left(\delta R_{ij}\right)\sqrt{-g}\, d^4x + \int_V R\,\delta\left(\sqrt{-g}\right) d^4x =$$

$$= \int_V \delta g^{ij} R_{ij}\sqrt{-g}\, d^4x + \int_V g^{ij}\delta R_{ij}\sqrt{-g}\, d^4x - \frac{1}{2}\int_V R\sqrt{-g}\, g_{ij}\delta g^{ij} d^4x$$

After rearranging the right side, we have

$$\delta S_{EH} = \int_V \left[R_{ij} - \frac{1}{2}R\, g_{ij}\right]\sqrt{-g}\,\delta g^{ij}\, d^4x + \int_V g^{ij}\,\delta R_{ij}\,\sqrt{-g}\, d^4x.$$

Since

$$\int_V g^{ij}\,\delta R_{ij}\,\sqrt{-g}\, d^4x = 0,$$

the condition $\delta S_{EH} = 0$ for g^{ij} arbitrary leads to the *Einstein field equations in vacuum*. Therefore in a region of space as the one described by the previous statement, without matter and energy, the Einstein field equations are

$$R_{ij} - \frac{1}{2}R\, g_{ij} = 0.$$

□

Theorem 10.3.6 (Einstein's field equations in presence of matter) *If V is a region of the Universe containing matter and energy, such that, on its boundary ∂V, the variations $\delta\Gamma^i_{jk}$ vanish, then there exists a* $(2, 0)$ *covariant tensor T_{ij} such that*

$$R_{ij} - \frac{1}{2}R\, g_{ij} = K\, T_{ij},$$

where K is a coupling constant.

Proof (*Hilbert*) In the previous theorem, we have used an action which describes the geometry of space without matter and energy. If we want to describe the geometry of a space with matter and energy inside it, the Einstein–Hilbert action has to contain a further term, denoted by S_M, depending on the matter–energy distribution in space-time.

So, the *general Einstein–Hilbert action* S_{GEH} has the form $kS_{EH} + S_M$, where k is a constant. This can be written in the form

$$S_{GEH} = \int_V (kR + S)\sqrt{-g}\, d^4x,$$

that is,

$$\delta S_{GEH} = k \int_V \left[R_{ij} - \frac{1}{2} R\, g_{ij} \right] \sqrt{-g}\, \delta g^{ij} d^4x + \int_V \left[\frac{\delta S}{\delta g^{ij}} \right] \sqrt{-g}\, \delta g^{ij} d^4x,$$

because we have already computed the variation

$$\delta S_{EH} = \int_V \left[R_{ij} - \frac{1}{2} R\, g_{ij} \right] \sqrt{-g}\, \delta g^{ij} d^4x.$$

If the first variation of S_{GEH} vanishes, it results in

$$R_{ij} - \frac{1}{2} R\, g_{ij} = K\, T_{ij},$$

where $T_{ij} := -\dfrac{\delta S}{\delta g^{ij}}$ and $K := \dfrac{1}{k}$.

The *general Einstein field equations* are obtained. □

10.4 An Introduction to $f(R)$ Gravity

It is very interesting to observe that the previous theorems can be generalized if instead of R we use any smooth function $f(R)$ in the Einstein–Hilbert action. In this way, we obtain the field equations of the so-called $f(R)$ *gravity*. They are the straightforward generalization of Einstein field equations and have recently acquired a lot of interest in view of solving several problems in cosmology and astrophysics (for a comprehensive discussion, see, for example, [56]). For example, the model of $f(R) = R + \dfrac{R^2}{6M^2}$, where M has the dimension of mass, gives rise to the so-called Starobinski inflation [175] which gives rise to the accelerated expansion of the early Universe capable of addressing several issues of Cosmological Standard Model based on General Relativity and Standard Model of Particles [127]. This kind of theories can be useful also to address issues related to the late Universe, like recent accelerated expansion, often dubbed dark energy epoch [53, 57, 149] or astrophysical issues like dark matter [62].

A detailed discussion of these problems is out of the scope of this book but it is worth pointing out that they are very active research areas. We refer the interested reader to the cited bibliography.

In view of the present discussion, it is interesting to develop how the Einstein field equations can be generalized in the $f(R)$ gravity framework. In particular, it is interesting to point out that metric and Palatini's formalisms give different field equations that, however, can be related to each other, see [64].

Taking into account the previous results related to the Einstein field equations, let us derive here the $f(R)$ gravity field equations. We shall use the following facts proven in the previous section, that is:

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\, g_{ij}\delta g^{ij},$$

$$\int_V g^{ij}\, \delta R_{ij}\, \sqrt{-g}\, d^4x = 0.$$

The second formula is a consequence of Palatini's identity

$$\delta R_{ij} = \delta\Gamma^s_{ij;s} - \delta\Gamma^s_{is;j}.$$

To begin, let us consider the basic objects introduced in the Differential Geometry chapter: $g_{ij},\ g^{ij},\ \Gamma_{ij,k},\ \Gamma^k_{ij},\ R^i_{jkl},\ R_{ij},\ R_{ijkl},\ R^i_j,\ R$. They are smooth functions, that is functions having derivatives of all order everywhere in the domain, here V. Therefore, if we assume $f(R)$ as a smooth function of $R(V)$, then $f(R)$ is a smooth function for $x \in V$. If V is compact and connected region in the Euclidean n-dimensional space, then $R(V)$ is a compact interval in $\mathbb{R}$. In other words, the Ricci scalar $R(V)$ is the image of V through R. In particular $f,\ f'$, ... are at least continuous functions on a real compact interval, here $R(V)$. The values of $f(R),\ f'(R)$, ... are in real compact intervals, too. The prime indicates derivative with respect to the Ricci scalar R.

Theorem 10.4.1 *If V is a compact and connected region of the universe without matter and energy inside it, such that on its boundary ∂V the variations $\delta\Gamma^i{}_{jk}$ vanish and f is a smooth real valued arbitrary function on $R(V)$, then*

$$f'(R)R_{ij} - \frac{1}{2}f(R)g_{ij} = 0.$$

Proof The line of the proof is similar to Theorem 9.3.5. The appropriate Lagrangian is

$$S_f = \int_V f(R)\sqrt{-g}\, d^4x.$$

Let us compute the first variation of S_f.

$$\delta S_f = \delta \int_V f(R)\sqrt{-g}\, d^4x = \int_V \delta[f(R)\sqrt{-g}]\, d^4x =$$

$$= \int_V f'(R)\delta R\sqrt{-g}\, d^4x + \int_V f(R)\delta\left(\sqrt{-g}\right) d^4x =$$

$$= \int_V f'(R)\left(\delta g^{ij}\right) R_{ij}\sqrt{-g}\, d^4x + \int_V f'(R)g^{ij}\left(\delta R_{ij}\right)\sqrt{-g}\, d^4x + \int_V f(R)\,\delta\left(\sqrt{-g}\right) d^4x =$$

$$= \int_V f'(R)\delta g^{ij} R_{ij}\sqrt{-g}\, d^4x + \int_V f'(R)g^{ij}\ \delta R_{ij}\sqrt{-g}\, d^4x - \frac{1}{2}\int_V f(R)\sqrt{-g}\, g_{ij}\delta g^{ij} d^4x$$

After rearranging the right-hand side, we have

$$\delta S_f = \int_V \left[f'(R)R_{ij} - \frac{1}{2}f(R)\, g_{ij}\right]\sqrt{-g}\,\delta g^{ij}\ d^4x + \int_V f'(R)\, g^{ij}\ \delta R_{ij}\ \sqrt{-g}\, d^4x.$$

Now, the mean value theorem implies the existence of a point $x \in V$ such that

$$\int_V f'(R)g^{ij}\ \delta R_{ij}\ \sqrt{-g}\, d^4x = f'(R(x))\int_V g^{ij}\ \delta R_{ij}\ \sqrt{-g}\, d^4x.$$

The last integral is 0, therefore the condition $\delta S_f = 0$ for g^{ij} arbitrary leads to $f(R)$ *field equations in vacuum*. Therefore, in the condition of the above statement, in a region of space without matter and energy, the $f(R)$ field equations in vacuum are

$$f'(R)R_{ij} - \frac{1}{2}f(R)\, g_{ij} = 0.$$

Let us observe that, for $f(R) = R$, we obtain the Einstein field equations in vacuum. □

Theorem 10.4.2 *If V is a compact and connected region of universe containing matter and energy, such that on its boundary ∂V the variations $\delta\Gamma^i_{jk}$ vanish and f is a smooth real valued arbitrary function on $R(V)$, then there exists a* (2, 0) *covariant tensor T_{ij} such that*

$$f'(R)R_{ij} - \frac{1}{2}f(R)g_{ij} = KT_{ij},$$

where K is a constant.

Proof If we choose the action

$$S_{Gf} = \int_V (kf(R) + S)\sqrt{-g}\, d^4x,$$

that is,

$$\delta S_{Gf} = k \int_V \left[f'(R)R_{ij} - \frac{1}{2} f(R)\, g_{ij} \right] \sqrt{-g}\, \delta g^{ij} d^4x + \int_V \left[\frac{\delta S}{\delta g^{ij}} \right] \sqrt{-g}\, \delta g^{ij} d^4x,$$

in the same way as in Theorem 9.3.6, we obtain the $f(R)$ *generalized field equations*

$$f'(R)R_{ij} - \frac{1}{2} f(R)\, g_{ij} = KT_{ij}.$$

□

Exercise 10.4.3 If the reader is interested in the differences between the Palatini and metric formalisms in $f(R)$ gravity, we propose the following exercise whose notation can be found in [56]. Starting from the action of $f(R)$, show that

$$f'(R)R_{ij} - \frac{1}{2} f(R) g_{ij} - f'(R)_{;ij} + g_{ij} \Box f'(R) = KT_{ij}\,,$$

with the trace

$$3\Box f'(R) + f'(R)R - 2f(R) = KT\,,$$

are the field equations obtained by varying with respect to g_{ij} without using the above Palatini identity. Demonstrate that they are equivalent to

$$f'(R)R_{ij} - \frac{1}{2} f(R)\, g_{ij} = KT_{ij}\,,$$

unless a divergence free current. Here $\Box$ is the d'Alembert operator defined as $\Box := \nabla_i \nabla^i$ with ∇_i the covariant derivative.

Hint. Use results in [64].

We will consider again $f(R)$ gravity in view of the discussion of the de Sitter space-time which is a solution of this theory.

10.5 The Schwarzschild Solution of Vacuum Field Equations

We intend to solve the Einstein field equations in vacuum, i.e. $R_{ij} = 0$ obtained previously assuming the spherical symmetry of space-time. The Schwarzschild solution is an exact solution for the vacuum field equations. Another way to find Schwarzschild solution is presented in [89, 142].

Theorem 10.5.1 *Consider the vacuum field equations $R_{ij} = 0$. Then*

$$ds^2 = c^2 \left(1 + \frac{B}{r}\right) dt^2 - \frac{1}{1 + \frac{B}{r}} dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

is the Schwarzschild solution for an arbitrary constant B.

Proof Karl Schwarzschild had the intuition to look for a spherically symmetric solution which describes the relativistic field outside of a non-rotating, massive body. This was the first exact solution of the Einstein field equations. Instead of the ordinary Cartesian coordinates $(x^0 = ct, x^1, x^2, x^3)$, Schwarzschild used spherical coordinates for the spatial part. The new coordinate system $(x^0 = ct, r, \varphi, \theta)$ is related to the old one by the formulas

$$x^1 = r \sin \varphi \cos \theta, \quad x^2 = r \sin \varphi \sin \theta, \quad x^3 = r \cos \varphi,$$

so then, for the spatial part, it is

$$(dx^1)^2 + (dx^2)^2 + (dx^3)^2 = dr^2 + r^2 d\varphi^2 + r^2 \sin^2 \varphi d\theta^2.$$

Far from the source, the solution has to approximate the Minkowski metric

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2.$$

In fact, the solution has to approximate

$$ds^2 = c^2 dt^2 - dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2,$$

which is the *Minkowski metric in spherically spatial coordinates.*[1]

Therefore, it is natural to think the Schwarzschild metric in the form

$$ds^2 = c^2 \cdot e^T dt^2 - (e^Q - 1) dr^2 - dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

where $T := T(r)$, $Q := Q(r)$ are two real functions that we need to determine from the vacuum field equations $R_{ij} = 0$. As we previously discussed, both $e^T \to 1$ and $e^Q \to 1$ have to go as $r \to \infty$. For the metric

$$ds^2 = c^2 \cdot e^T dt^2 - e^Q dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

[1] This request is an important property that any physical solution has to possess. In fact, very far from the source, a gravitational field has to go to zero. This means that Minkowski space-time has to be recovered. This property is called "asymptotic flatness" and characterizes any physical gravitational field. It is worth noticing that this feature is fundamental for black hole solutions having physical meaning.

the coefficients are

$$g_{00} = e^T, \quad g_{11} = -e^Q, \quad g_{22} = -r^2, \quad g_{33} = -r^2 \sin^2 \varphi.$$

The inverse matrix coefficients are

$$g^{00} = e^{-T}, \qquad g^{11} = -e^{-Q}, \quad g^{22} = -\frac{1}{r^2}, \quad g^{33} = -\frac{1}{r^2 \sin^2 \varphi}.$$

Let us observe that

$$\frac{\partial g_{ij}}{\partial x^0} = 0, \quad i, j = 0, .., 3; \quad \Gamma^i_{jk} = 0, \quad i \neq j \neq k.$$

The non-zero Christoffel symbols are

$$\Gamma^0_{01} = \Gamma^0_{10} = \frac{T'}{2}, \quad \Gamma^1_{00} = -\frac{T'}{2} e^{T-Q}, \quad \Gamma^1_{11} = \frac{Q'}{2},$$
$$\Gamma^1_{22} = -re^{-Q}, \quad \Gamma^1_{33} = -re^{-Q} \sin^2 \varphi,$$

$$\Gamma^2_{21} = \Gamma^2_{12} = \frac{1}{r}, \quad \Gamma^2_{33} = -\sin\varphi\cos\varphi, \quad \Gamma^3_{31} = \Gamma^3_{13} = \frac{1}{r}, \quad \Gamma^3_{23} = \Gamma^3_{32} = \cot\varphi.$$

The only non-zero components of the Ricci tensor are

$$R_{00} = e^{T-Q}\left(\frac{T''}{2} + \frac{(T')^2}{4} - \frac{T'Q'}{4} + \frac{T'}{r}\right), \quad R_{11} = -\left(\frac{T''}{2} + \frac{(T')^2}{4} - \frac{T'Q'}{4} - \frac{Q'}{r}\right)$$

$$R_{22} = 1 - e^Q + re^{-Q}\left(\frac{Q'}{2} - \frac{T'}{2}\right), \quad R_{33} = \sin^2\varphi \, R_{22}.$$

The conditions $R_{00} = 0$ and $R_{11} = 0$ determine both T and Q.

Indeed, $e^{Q-T} R_{00} + R_{11} = 0$ implies $T' + Q' = 0$, that is $T + Q$ = constant = k. Thus $e^T = e^{-Q} e^k$, i.e. the metric is

$$ds^2 = c^2 \cdot e^{-Q} e^k dt^2 - e^Q dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2.$$

If we let $t = e^{k/2} u$, then $dt^2 = e^k du^2$, and the metric becomes

$$ds^2 = c^2 \cdot e^{-Q} du^2 - e^Q dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2.$$

So, we may choose $T + Q = 0$, i.e. $Q = -T$. Replacing in the second equation, we have

$$rT'' + r(T')^2 + 2T' = \left(re^T\right)'' = 0.$$

It results in $\left(re^T\right)' = A$, that is, $re^T = Ar + B$, i.e. $e^T = A + \frac{B}{r}$. We impose that $e^T \to 1$ as $r \to \infty$; it results in $A = 1$. Therefore $e^T = 1 + \frac{B}{r}$ and $e^Q = e^{-T} = \frac{1}{1+\frac{B}{r}}$.

Let us observe that for T and Q so far determined, $R_{22} = R_{33} = 0$.

The *Schwarzschild metric* is exactly the solution in the theorem statement, that is,

$$ds^2 = c^2\left(1+\frac{B}{r}\right)dt^2 - \frac{1}{1+\frac{B}{r}}dr^2 - r^2 d\varphi^2 - r^2\sin^2\varphi d\theta^2 .$$

□

It is important to note that the Schwarzschild solution is independent of time. According to this property, the solution is not only stationary but also static. This is the statement of the Birkhoff theorem. See [113] for a detailed proof.

10.5.1 Orbit of a Planet in the Schwarzschild Metric

The above result can be immediately applied to celestial mechanics. Let us recall the classical orbit from first Kepler's law. The differential equation which describes the gravitational attraction between a planet and a star is

$$\ddot{\vec{X}} = -\frac{GM}{r^3}\cdot\vec{X},$$

where $\vec{X}$ is the position vector, $G = 6.67\cdot 10^{-11}(\mathrm{m})^3/(\mathrm{kg})\cdot(\mathrm{s})^2$ is the gravitational constant, M is the mass of the star, and $r = \| \vec{X} \|$. Let J be the magnitude of the angular moment of the planet. If we consider polar coordinates and $r = r(\theta) = \frac{1}{u(\theta)}$, then the previous equation becomes

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2}, \quad \mu = GM.$$

The classical solution is

$$u(\theta) = \frac{\mu}{J^2} + A\cos(\theta-\theta_0),$$

where A is an arbitrary constant which can be obtained from the initial condition and θ_0 is a phase shift. Since the phase shift alters the position of the planet at time $t = 0$

and we are interested only in the orbit itself, we may consider $\theta_0 = 0$. Denoting by e the eccentricity $e := \frac{AJ^2}{\mu}$, the orbit described by the solution

$$u(\theta) = \frac{\mu}{J^2}(1 + e\cos\theta) \quad \text{is the conic} \quad r(\theta) = \frac{\frac{J^2}{\mu}}{1 + e\cos\theta}.$$

The next result provides the differential equation which predicts the orbit of a planet in its movement around the Sun in the new context of Schwarzschild metric.

Theorem 10.5.2 *The orbit of a planet in the Schwarzschild metric is described by the equation*

$$\frac{d^2u}{d\theta^2} + u = -\frac{c^2B}{2J^2} - \frac{3B}{2}u^2.$$

Proof In the same way as before, we denote $x^0 := ct$. The worldcurve of the planet is the geodesic $\zeta(\tau) := (t(\tau), r(\tau), \varphi(\tau), \theta(\tau))$ of the Schwarzschild metric. We are looking for a solution in the (x, y) plane, that is $\varphi = \frac{\pi}{2}$. The reduced metric is

$$ds^2 = \left(1 + \frac{B}{r}\right)(dx^0)^2 - \frac{1}{1 + \frac{B}{r}}dr^2 - r^2 d\theta.$$

Since $\Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r}$ and $\Gamma^3_{ij} = 0$, the equation corresponding to the variable θ is:

$$\ddot{\theta}\,(\tau) + \frac{2}{r(\tau)} \cdot \dot{r}\,(\tau) \cdot \dot{\theta}\,(\tau) = \frac{1}{r^2}\left(r^2(\tau)\,\dot{\theta}\,(\tau)\right)' = 0.$$

We denote $r^2\,\dot{\theta} = J$ and this J describes the magnitude of the angular momentum of the planet exactly as in the classical case. We cancel τ in the next computations.

Let us continue with the geodesic equation corresponding to the variable x^0. Since only $\Gamma^0_{01} = \Gamma^0_{10} = -\frac{1}{\left(1 + \frac{B}{r}\right)} \cdot \frac{B}{2r^2}$, the equation in x^0 is

$$\ddot{x}^0 - \frac{B}{r^2} \cdot \frac{1}{\left(1 + \frac{B}{r}\right)} \cdot \dot{x}^0 \cdot \dot{r} = 0.$$

By replacing x^0 with ct, it results in

$$\ddot{t} - \frac{B}{r^2} \cdot \frac{1}{\left(1 + \frac{B}{r}\right)} \cdot \dot{t} \cdot \dot{r} = 0 \quad \text{i.e.} \quad \left(\left(1 + \frac{B}{r}\right) \cdot \dot{t}\right)' = 0,$$

that is,

$$\dot{t} = \frac{E}{1 + \frac{B}{r}},$$

where E is a constant.

In the case of the equation corresponding to the variable r, we use directly the metric condition taking into account that $ds^2 = c^2 d\tau^2$. After cancelling $d\tau^2$, it results in

$$c^2 = c^2 \left(1 + \frac{B}{r}\right) \dot{t}^2 - \frac{1}{1 + \frac{B}{r}} \dot{r}^2 - r^2 \dot{\theta}^2 .$$

Let us replace $\dot{t}$ and $\dot{\theta}$ in the previous equation, we have

$$c^2 \left(1 - E^2\right) + c^2 \cdot B \cdot \frac{1}{r} = - \dot{r}^2 - \frac{J^2}{r^2} - \frac{B \cdot J^2}{r^3}.$$

Consider $r = r(\theta)$. It results in

$$\dot{r} = \frac{dr}{d\theta} \cdot \dot{\theta} = \frac{dr}{d\theta} \cdot \frac{J}{r^2}.$$

If $r := \frac{1}{u}$, then $\frac{dr}{d\theta} = -\frac{1}{u^2} \frac{du}{d\theta}$, i.e.

$$\dot{r} = -J \cdot \frac{du}{d\theta}.$$

Since $\dot{r}^2 = J^2 \cdot \left(\frac{du}{d\theta}\right)^2$, the previous equation becomes

$$c^2(1 - E^2) + c^2 Bu = -J^2 \left(\frac{du}{d\theta}\right)^2 - J^2 u^2 - BJ^2 u^3.$$

If we differentiate with respect to θ, then we divide by $\frac{du}{d\theta}$, we obtain the equation

$$\frac{d^2 u}{d\theta^2} + u = -\frac{c^2 B}{2J^2} - \frac{3B}{2} u^2.$$

□

10.5.2 *Relativistic Solution of the Mercury Perihelion Drift Problem*

Now we need to clarify who is B in the Schwarzschild metric. We have requested that, as $r \to \infty$, the Schwarzschild metric approaches the ordinary Minkowski metric. Let us continue by taking into account the following two equations.

1. The classical orbit is described by

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2}, \quad \mu = GM$$

2. The relativistic orbit is described, in the Schwarzschild metric, by

$$\frac{d^2u}{d\theta^2} + u = -\frac{c^2 B}{2J^2} - \frac{3B}{2}u^2 .$$

From $g_{00} = c^2\left(1 + \frac{B}{r}\right)$, we compute Γ^1_{00} which is the only non-zero Γ^i_{00}. So, the r-component of the geodesic equation is

$$\frac{d^2r}{d\tau^2} = \Gamma^1_{00}\frac{d\tau}{d\tau}\frac{d\tau}{d\tau},$$

that is, after canceling the third order term,

$$\frac{d^2r}{d\tau^2} = \frac{c^2B}{2r^2}.$$

As r approaches the infinity, $d\tau$ becomes dt and the previous equation is the original Newton equation $\frac{d^2r}{dt^2} = -\frac{GM}{r^2}$ if and only if $B = -\frac{2GM}{c^2}$.

It results in

$$1 + \frac{B}{r} = 1 - \frac{2GM}{c^2} \cdot \frac{1}{r}.$$

In this way, the gravitational Newtonian potential $\phi(x, y, z) = -\frac{GM}{r}$ is involved in the coefficients of the metric. The coefficient $\frac{1}{c^2}$ highlights the weak gravitational field which we will discuss later. See also [46].

The quantity $r_M := \frac{2GM}{c^2}$ has the dimension of a length and it is called *gravitational radius*, or *the Schwarzschild radius*, corresponding to the mass M. It is an intrinsic characteristic of any body with mass.

In General Relativity, we can define a *proper time interval* $\Delta\tau$ between two events along a time-like path l following the definition given in Special Relativity. Using constant space coordinates, the proper time satisfies the same equality

$$ds^2 = c^2(d\tau)^2$$

as in Special Relativity. Therefore using the same constant coordinates $x^1,\ x^2,\ x^3$, it results in

$$\Delta\tau = \int_l ds = \int_l \frac{1}{c}\sqrt{g_{ij}dx^i dx^j} = \int_l \frac{1}{c}\sqrt{g_{00}}dx^0.$$

A discussion about how gravity influences the proper time is in [154].

Next result allows to make distinction between the proper time and the time coordinate in the case of Schwarzschild metric.

Theorem 10.5.3 *The gravitational field described by the Schwarzschild metric*

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)dt^2 - \frac{1}{1 - \frac{2GM}{c^2 r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2$$

causes the slowdown of clocks.

Proof Let us consider the Schwarzschild metric in a frame at rest R and apply the previous results in the following way. The source of the gravitational field is at the origin O of R. Consider two motionless observers, one close to the source O, denoted by O_1, and the other one far from the source, O_2. Each observer has a clock. For both observers the variation of the space coordinates is 0. We have $dr = d\varphi = d\theta = 0$ for the first observer, therefore, according to him

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)^2 dt^2.$$

For the second motionless observer at rest, far from source, the influence of the gravitational field is almost not observable. There, for $r \to \infty$, we have $ds^2 = c^2 d\tau^2$. Therefore the proper time is affected by the gravity according to the rule

$$c^2 d\tau^2 = ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)dt^2.$$

Considering the clocks, it results in

$$\not{c}^2 \Delta\tau^2 = \not{c}^2\left(1 - \frac{2GM}{c^2 r}\right)\Delta t^2,$$

that is, the time interval $\Delta\tau$ of O_2's clock appears to be less than Δt on O_1's clock. If you are close to a source, your clock will slow down and will continue to slow down if you come closer and closer to the source. □

Let c be the speed of light in vacuum. If we write formally an expression Q as a Taylor series in powers of $\frac{1}{c}$

$$Q = a_0 + a_1 \cdot \frac{1}{c} + a_2 \cdot \frac{1}{c^2} + a_3 \cdot \frac{1}{c^3} + \cdots + a_k \cdot \frac{1}{c^k} + \cdots,$$

we say that the order of Q is $O\left(\frac{1}{c^m}\right)$ if $a_0 = a_1 = \cdots = a_{m-1} = 0$ and $a_m \neq 0$. How is this formal definition working in a given physical context? Let us write each relativistic expression (components of the gravitational field, metric tensor, equations) as a Taylor series in powers of $\frac{1}{c}$. The computations with these series can be truncated at the term that is appropriate for the physical context we are considering.

Theorem 10.5.4 *In the relativistic field described by the Schwarzschild metric*

$$ds^2 = c^2\left(1 - \frac{2GM}{c^2 r}\right)dt^2 - \frac{1}{1 - \frac{2GM}{c^2 r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

the planet equation of motion

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2}u^2$$

has the solution

$$u(\theta) = \frac{\mu}{J^2}\left(1 + e\cos(\theta - F\theta)\right) + O\left(\frac{1}{c^2}\right),$$

where $F := \frac{3\mu^2}{c^2 J^2}$, *being* $\mu = GM$.

Proof We start from the classical equation of an orbit of a planet,

$$\frac{d^2a}{d\theta^2} + a = \frac{\mu}{J^2}$$

with the classical solution $a(\theta) = \frac{\mu}{J^2}(1 + e\cos\theta)$.

The new equation of the orbit

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2} \cdot u^2$$

differs from the classical one by $\dfrac{3\mu}{c^2}u^2$. This "correction" of the classical orbit is due to the gravity related to the Schwarzschild metric. It is natural to search for the solution as

$$u(\theta) := a(\theta) + \frac{w(\theta)}{c^2}.$$

If we replace it in the new orbit equation

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2} \cdot u^2,$$

we obtain

$$\frac{d^2a}{d\theta^2} + a + \frac{1}{c^2}\left(\frac{d^2w}{d\theta^2} + w\right) = \frac{\mu}{J^2} + \frac{3\mu}{c^2}\left(\frac{\mu}{J^2}(1 + e\cos\theta) + \frac{w}{c^2}\right)^2,$$

or equivalently

$$\frac{1}{c^2}\left(\frac{d^2w}{d\theta^2} + w\right) = \frac{3\mu^3}{c^2J^4}(1 + e\cos\theta)^2 + O\left(\frac{1}{c^4}\right).$$

The term $O\left(\dfrac{1}{c^4}\right)$ has a small influence on w. It remains to solve

$$\frac{d^2w}{d\theta^2} + w = \frac{3\mu^3}{J^4}\left(1 + \frac{e^2}{2} + 2e\cos\theta + \frac{e^2}{2}\cos 2\theta\right).$$

The solutions of the following three equations

$$\frac{d^2w_1}{d\theta^2} + w_1 = \frac{3\mu^3}{J^4}\left(1 + \frac{e^2}{2}\right); \quad \frac{d^2w_2}{d\theta^2} + w_2 = \frac{6e\mu^3}{J^4}\cos\theta;$$
$$\frac{d^2w_3}{d\theta^2} + w_3 = \frac{3\mu^3e^2}{2J^4}\cos 2\theta$$

are

$$w_1 = \frac{3\mu^3}{J^4}\left(1 + \frac{e^2}{2}\right), \quad w_2 = \frac{3e\mu^3}{J^4}\theta\cos\theta, \quad w_3 = -\frac{3\mu^3e^2}{2J^4}\cos 2\theta.$$

Therefore, the solution of the new orbit equation is

$$u(\theta) = \frac{\mu}{J^2}(1 + e\cos\theta) + \frac{3\mu^3}{J^4c^2}\left(1 + \frac{e^2}{2} + e\theta\sin\theta - \frac{e^2}{2}\cos 2\theta\right).$$

Einstein's idea was to use only the non-periodic term in the classical solution. Then

$$u(\theta) = \frac{\mu}{J^2}\left[1 + e\left(\cos\theta + \frac{3\mu^2}{c^2 J^2}\theta\sin\theta\right)\right] + O\left(\frac{1}{c^2}\right),$$

which can be written as

$$u(\theta) = \frac{\mu}{J^2}\left[1 + e\cos(\theta - F\theta)\right] + O\left(\frac{1}{c^2}\right),$$

where $F := \frac{3\mu^2}{c^2 J^2}$. Neglecting the term $O\left(\frac{1}{c^2}\right)$ which adds only a small contribution, the trajectory is still the old conic. □

The correction to the classical trajectory, described in the Schwarzschild metric, reaches the perihelion for $\cos(\theta - F\theta) = 1$, therefore, it is $\theta = \theta_n = \frac{2n\pi}{1-F}$ for an integer n. It results in $\theta \approx 2n\pi\left(1 + F + O\left(F^2\right)\right)$; that is $2\pi F$ is the perihelion drift for each revolution.

If N is the number of orbits for a given period of time T, then the perihelion drift P_d is

$$P_d = \frac{6\pi G^2 M^2}{c^2 J^2}\cdot N.$$

For Mercury, if we replace the constants, we obtain 43 arcseconds per century which was observed by astronomers, without explanation, in the context of Classical Mechanics. This was considered one of the first confirmations of General Relativity. See Sect. 9.8 and [154] for a detailed discussion also in the historical context.

10.5.3 Speed of Light in a Given Metric

Consider a Minkowski space-time and suppose the worldcurve $\mathbb{X}(t) = (ct, x^1(t), x^2(t), x^3(t))$ of a spatial object parameterized by the time t.

Then, its relativistic speed is

$$\left\|\frac{d\mathbb{X}}{dt}\right\|^2 = c^2 - \left(\frac{dx^1}{dt}\right)^2 - \left(\frac{dx^2}{dt}\right)^2 - \left(\frac{dx^3}{dt}\right)^2 = c^2 - v^2,$$

where

$$v = \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2}$$

is the ordinary velocity of the object.

If the object is a photon,

$$\left\|\frac{d\mathbb{X}}{dt}\right\|^2 = c^2 - c^2 = 0$$

as we expected.

If we consider the same worldcurve $\mathbb{X}(t) = (ct, x^1(t), x^2(t), x^3(t))$ in the metric

$$ds^2 = g_{00}(dx^0)^2 + g_{\alpha\beta}dx^\alpha dx^\beta,$$

where α, β are spatial indexes; according to the above formalism, we have

$$ds^2 = g_{00}(dx^0)^2 + g_{11}(dx^1)^2 + g_{22}(dx^2)^2 + g_{33}(dx^3)^2 + \sum_{\alpha\neq\beta=1}^{3} g_{\alpha\beta}dx^\alpha dx^\beta.$$

Then

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (g_{00} - 1)(dx^0)^2 + (g_{11} + 1)(dx^1)^2 +$$

$$+ (g_{22} + 1)(dx^2)^2 + (g_{33} + 1)(dx^3)^2 + \sum_{\alpha\neq\beta=1}^{3} g_{\alpha\beta}dx^\alpha dx^\beta,$$

i.e.

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 + (g_{00} - 1)(dx^0)^2 + \sum_{\alpha,\beta=1}^{3} \bar{g}_{\alpha\beta}dx^\alpha dx^\beta$$

where $\bar{g}_{\alpha\beta} = g_{\alpha\beta}$, if $\alpha \neq \beta$ and $\bar{g}_{\alpha\beta} = 1 + g_{\alpha\beta}$, if $\alpha = \beta$.

This means that, for a photon, we obtain

$$0 = \left(\frac{d\mathbb{X}}{dt}\right)^2 = c^2 - \gamma^2 + (g_{00} - 1)\left(\frac{dx^0}{dt}\right)^2 + \sum_{\alpha,\beta=1}^{3} \bar{g}_{\alpha\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt}$$

where

$$\gamma = \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2}.$$

It results in

$$\gamma = \sqrt{c^2 + (g_{00} - 1)\left(\frac{dx^0}{dt}\right)^2 + \sum_{\alpha,\beta=1}^{3} \bar{g}_{\alpha\beta}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt}}.$$

γ is called *the speed of light in a gravitational field* derived by the above metric.
We say that γ does not *violate the speed of light limit* if $\gamma \leq c$.

10.5.4 Bending of Light in the Schwarzschild Metric

Let us consider now the light travelling in the space-time described by the Schwarzschild metric . First of all, we need to compute the speed γ of light in the gravitational field induced by the metric above.

Theorem 10.5.5 *Consider the Schwarzschild metric*

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \frac{1}{1-\frac{2\mu}{c^2 r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2.$$

(i) In Cartesian coordinates this metric has the form

$$ds^2 = \sum_{i=0}^{3}\left(dx^i\right)^2 - \frac{2\mu}{c^2 r}\left((dx^0)^2 + \frac{1}{1-\frac{2\mu}{c^2 r}}\sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}dx^\alpha dx^\beta\right).$$

(ii) A deflected photon in the (x^1, x^2) plane, which comes from the undeflected photon
$X(t) = (ct, h, ct, 0)$, has the speed

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu}{cr}\cdot\frac{(x^2)^2}{r^2}\cdot\frac{1}{1-\frac{2\mu}{c^2 r}}.$$

(iii) The deflected photon does not violate the speed of light limit c and γ can be written in the equivalent form

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu(x^2)^2}{cr^3} + O\left(\frac{1}{c^3}\right).$$

Proof (i) Let $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ and $r^2 = x^2 + y^2 + z^2$.

It results in $rdr = xdx + ydy + zdz$ which gives

$$dr = \sum_{\alpha=1}^{3}\frac{x^\alpha}{r}dx^\alpha, \quad dr^2 = \sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}dx^\alpha dx^\beta.$$

Taking into account that

$$dx^2 + dy^2 + dz^2 = dr^2 + r^2 d\varphi^2 + r^2 \sin^2 \varphi d\theta^2 = \sum_{\alpha=1}^{3} (dx^\alpha)^2 \ ,$$

it results in

$$ds^2 = c^2dt^2 - \frac{2\mu}{c^2r}c^2dt^2 - \frac{1 - \frac{2\mu}{c^2r} + \frac{2\mu}{c^2r}}{1 - \frac{2\mu}{c^2r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

$$= c^2dt^2 - dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2 - \frac{2\mu}{c^2r}\left(c^2dt^2 + \frac{1}{1 - \frac{2\mu}{c^2r}}dr^2\right)$$

$$= (dx^0)^2 - \sum_{\alpha=0}^{3} (dx^\alpha)^2 - \frac{2\mu}{c^2r}\left((dx^0)^2 + \frac{1}{1 - \frac{2\mu}{c^2r}} \sum_{\alpha,\beta=1}^{3} \frac{x^\alpha x^\beta}{r^2} dx^\alpha dx^\beta\right).$$

(ii) According to the technique previously described, suppose that $X(t) = \big(ct, x^1(t), x^2(t), x^3(t)\big)$ is the worldcurve of an object parameterized by the time t. In the Minkowski metric, it is

$$\left(\frac{ds}{dt}\right)^2 = \| \dot{X}\ (t)\|^2 = c^2 - \left((\dot{x}^1\ (t))^2 + (\dot{x}^2\ (t))^2 + (\dot{x}^3\ (t))^2\right) = c^2 - v^2,$$

where v is the usual spatial speed of the object.

If the object is a photon, then $\left(\frac{ds}{dt}\right)^2 = c^2 - c^2 = 0$ and so

$$0 = \left(\frac{ds}{dt}\right)^2 = c^2 - \gamma^2 - \frac{2\mu}{c^2r}\left(c^2 + \frac{1}{1 - \frac{2\mu}{c^2r}} \sum_{\alpha,\beta=1}^{3} \frac{x^\alpha x^\beta}{r^2} \frac{dx^\alpha}{dt} \frac{dx^\beta}{dt}\right) \ ,$$

where

$$\gamma(t) = \sqrt{(\dot{x}^1\ (t))^2 + (\dot{x}^2\ (t))^2 + (\dot{x}^3\ (t))^2}$$

is the speed of the photon in the gravitational field described by the above metric. In fact,

$$\gamma = \sqrt{c^2 - \frac{2\mu}{c^2r}\left(c^2 + \frac{1}{1 - \frac{2\mu}{c^2r}} \sum_{\alpha,\beta=1}^{3} \frac{x^\alpha x^\beta}{r^2} \frac{dx^\alpha}{dt} \frac{dx^\beta}{dt}\right)}.$$

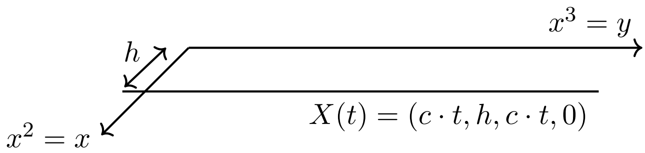


Fig. 10.1 Trajectory of an undeflected photon

We determine γ along the worldcurve X of an undeflected photon in the (x^1, x^2) plane at the fixed distance h from the x^2-axis (Fig. 10.1).
The undeflected worldcurve of the photon is $X(t) := (ct, h, ct, 0)$. The deflection will add only lower order terms, therefore the deflected photon has, in the same plane, a worldcurve in which components have extra terms of order $O\left(\frac{1}{c}\right)$. The deflected photon is parameterized by

$$X_d(t) := \left(ct, h + O\left(\frac{1}{c}\right), ct + O(1), 0\right).$$

Since $\dot{X_d}\ (t) := \left(c, O\left(\frac{1}{c}\right), c + O(1), 0\right)$, we have

$$\frac{dx^1}{dt} = O\left(\frac{1}{c}\right), \quad \frac{dx^2}{dt} = c + O(1), \quad \frac{dx^3}{dt} = 0.$$

It results in the approximation

$$\gamma^2 = c^2 - \frac{2\mu}{c^2 r}\left(c^2 + \frac{1}{1 - \frac{2\mu}{c^2 r}} \frac{(x^2)^2}{r^2} \cdot c^2\right),$$

equivalent to

$$\gamma^2 = c^2\left(1 - \frac{2\mu}{c^2 r} - \frac{2\mu(x^2)^2}{c^2 r^3} \frac{1}{1 - \frac{2\mu}{c^2 r}}\right),$$

i.e.

$$\gamma = c \cdot \sqrt{1 - \frac{2\mu}{c^2 r} - \frac{2\mu(x^2)^2}{c^2 r^3} \frac{1}{1 - \frac{2\mu}{c^2 r}}}.$$

Taking into account that $\sqrt{1 + 2A} \approx 1 + A$, the result is

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu}{cr} \cdot \frac{(x^2)^2}{r^2} \cdot \frac{1}{1 - \frac{2\mu}{c^2 r}}.$$

(iii) Since $\frac{1}{1-\frac{2\mu}{c^2 r}} \approx 1 + \frac{2\mu}{c^2 r} + O\left(\frac{1}{c^3}\right)$ it results in both the formula

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu (x^2)^2}{cr^3} + O\left(\frac{1}{c^3}\right)$$

and the fact that the deflected photon does not violate the light limit speed. □

Theorem 10.5.6 *The total deflection of the trajectory $X_d(t)$ of a deflected photon in the gravitational field described by the Schwarzschild metric*

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \frac{1}{1-\frac{2\mu}{c^2 r}}dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

is $T_D = \frac{4GM}{c^2 h}$.

Proof Let us recall that the trajectory of the deflected photon is

$$X_d(t) = \left(ct, h + O\left(\frac{1}{c}\right), ct + O(1), 0\right)$$

and it comes from the undeflected photon trajectory $X(t) = (ct, h, ct, 0)$, in which the deflection added small contribution terms.

The previous theorem proves that the speed of a deflected photon in (x, y)-plane is

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu (x^2)^2}{cr^3} + O\left(\frac{1}{c^3}\right).$$

Let us imagine a line l and two points on it having coordinates x and $x + dx$, respectively. Two parallel lines constructed through the given points make the same $\Delta\theta$ angle with the perpendicular to the l-direction. These lines can be imagined as trajectories of photons, the first one travelling with the speed $\gamma(x)$, the second one travelling with the speed $\gamma(x + \Delta x)$.

After Δt seconds, the last two parallel lines change the trajectories into other two parallel lines, etc. The first photon travelled $\gamma(x)\Delta t$, the second one travelled $\gamma(x + \Delta x)\Delta t$.

Let us suppose now that $\gamma(x)\Delta t > \gamma(x + \Delta x)\Delta t$. It is easy to see that there is a rectangle triangle which leads to the relation

$$\Delta\theta \approx \sin \Delta\theta = \frac{\gamma(x)\Delta t - \gamma(x + \Delta x)\Delta t}{dx},$$

that is,

$$\frac{\Delta\theta}{dt} \approx \frac{\gamma(x) - \gamma(x + \Delta x)}{dx}.$$

As $\Delta t \to 0$ and $\Delta x \to 0$, the last relation becomes

$$\frac{\Delta\theta}{dt} = -\frac{\partial\gamma}{\partial x}.$$

If $s = ct$,

$$\frac{\Delta\theta}{ds} = -\frac{1}{c}\frac{\partial\gamma}{\partial x}.$$

At the same time, $\frac{\Delta\theta}{ds}$ is the geometric curvature determined for the photons' trajectories when the parameter is s. If we denote x by x^1, the perpendicular direction coordinate by x^2, the *total deflection* is related to the integral of the geometric curvature $\frac{\Delta\theta}{ds}$, that is,

$$T_D := -\frac{1}{c}\int_{-\infty}^{\infty}\frac{\partial\gamma}{\partial x^1}dx^2.$$

To perform the computation, we start from cancelling the $O\left(\frac{1}{c^4}\right)$ term. We have

$$\left.\frac{\partial\gamma}{\partial x^1}\right|_{X_d} = \left.\frac{GMx^1}{cr^3}\right|_{X_d} + \left.\frac{3GM(x^2)^2x^1}{cr^5}\right|_{X_d} = \frac{GMh}{c(h^2+(x^2)^2)^{\frac{3}{2}}} + \frac{3GMh(x^2)^2}{c(h^2+(x^2)^2)^{\frac{5}{2}}}$$

and elementary computations lead to

$$\frac{1}{c}\int_{-\infty}^{\infty}\frac{\partial\gamma}{\partial x^1}dx^2 = \frac{GMh}{c^2}\left(\int_{-\infty}^{\infty}\frac{1}{(h^2+(x^2)^2)^{\frac{3}{2}}}dx^2 + \int_{-\infty}^{\infty}\frac{3(x^2)^2}{(h^2+(x^2)^2)^{\frac{5}{2}}}dx^2\right) =$$

$$= \left(\frac{2}{h^2}+\frac{2}{h^2}\right)\frac{GMh}{c^2}.$$

The total deflection is then $T_D = \frac{4GM}{c^2h}$. □

At the surface of the Sun, we have

$$h = \text{radius of the Sun} = 7\times 10^8(\text{m});\ G = 6,67\times 10^{-11}(\text{m}^3)/(\text{kg})\cdot(\text{s}^2),$$
$$M = \text{mass of the Sun} = 2\times 10^{30}(\text{kg});\ c = 3\times 10^8(\text{m})/(\text{s}^2).$$

It results for $T_D s \approx 1,75''$. This was another sensational confirmation of General Relativity due to Dyson and Eddington in 1919. See [154] for details.

10.6 The Einstein Metric: Einstein's Computations Related to the Perihelion Drift and the Bending of Light Rays

Even if Einstein was the one who discovered the vacuum field equations, he did not solve them. In order to make computations possible, he choose a spherically symmetric metric, independent of time, metric who approximates the Minkowski metric as $r \to \infty$. He took care to involve the gravitational potential $\Phi = -\dfrac{GM}{r}$ in the first two coefficients. Therefore, the chosen metric was

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \left(1 + \frac{2\mu}{c^2 r}\right)dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2,$$

being always $\mu = GM$. Obviously, this metric does not satisfy the field equations $R_{ij} = 0$.

Einstein's computations on perihelion drift and bending of light were performed with this metric.

Theorem 10.6.1 (Einstein's First Theorem) *In the relativistic field described by Einstein's metric*

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \left(1 + \frac{2\mu}{c^2 r}\right)dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2$$

the planet equation of motion

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2}u^2$$

has the solution

$$u(\theta) = \frac{\mu}{J^2}\left(1 + e\cos(\theta - F\theta)\right) + O\left(\frac{1}{c^2}\right),$$

where $F := \dfrac{3\mu^2}{c^2 J^2}$.

Proof In the same way as before, we denote $x^0 := ct$. The worldcurve of the planet is the geodesic $\zeta(\tau) := (t(\tau), r(\theta), \varphi(\tau), \theta(\tau))$ of Einstein's metric. We are looking for a solution in the (x, y)-plane, that is, $\varphi = \dfrac{\pi}{2}$. The reduced metric is

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \left(1 + \frac{2\mu}{c^2 r}\right)dr^2 - r^2 d\theta^2 .$$

We cancel out τ in the next computations. Since $\Gamma^3_{13} = \Gamma^3_{31} = \dfrac{1}{r}$ and $\Gamma^3_{ij} = 0$ in the other cases, the equation corresponding to the variable θ is

$$\ddot{\theta} + \frac{2}{r} \cdot \dot{r} \cdot \dot{\theta} = \frac{1}{r^2} \left(r^2 \, \dot{\theta} \right)' = 0.$$

It results in $r^2 \dot{\theta}$ =constant. We denote $J := r^2 \, \dot{\theta}$.

The constant J describes the magnitude of the angular momentum of the planet exactly as in the classical case.

Let us continue with the geodesic equation corresponding to the variable x^0.

Since only $\Gamma^0_{01} = \Gamma^0_{10} = \dfrac{1}{1 - \dfrac{2\mu}{c^2 r}} \cdot \dfrac{\mu}{c^2 r^2}$, the equation in x^0 is

$$\ddot{x}^0 + \frac{2\mu}{c^2 r^2} \cdot \frac{1}{1 - \frac{2\mu}{c^2 r}} \cdot \dot{x}^0 \cdot \dot{r} = 0.$$

Replacing x^0 by ct, it results in

$$\ddot{t} + \frac{2\mu}{c^2 r^2} \cdot \frac{1}{1 - \frac{2\mu}{c^2 r}} \cdot \dot{t} \cdot \dot{r} = 0,$$

that is,

$$\dot{t} = \frac{E}{1 - \frac{2\mu}{c^2 r}},$$

where E is a constant.

In the case of the equation corresponding to the variable r, we use directly the metric condition. Taking into account that $ds^2 = c^2 d\tau^2$, after we cancel $d\tau^2$, it results in

$$c^2 = c^2 \left(1 - \frac{2\mu}{c^2 r} \right) \dot{t}^2 - \left(1 + \frac{2\mu}{c^2 r} \right) \dot{r}^2 - r^2 \, \dot{\theta}^2 .$$

Let us replace $\dot{t} = \dfrac{E}{1 - \frac{2\mu}{c^2 r}}$, and $\dot{\theta} = \dfrac{J}{r^2}$ in the previous equation. We have

$$c^2 \left(1 - E^2 \right) - \frac{2\mu}{r} = \left(\frac{4\mu^2}{c^4 r^2} - 1 \right) \dot{r}^2 - r^2 \dot{\theta}^2 \left(1 - \frac{2\mu}{c^2 r} \right).$$

Consider $r = r(\theta)$. It results in

$$\dot{r} = \frac{dr}{d\theta} \cdot \dot{\theta} = \frac{dr}{d\theta} \cdot \frac{J}{r^2}.$$

If $r := \frac{1}{u}$, then $\frac{dr}{d\theta} = -\frac{1}{u^2}\frac{du}{d\theta}$, i.e. $\dot{r} = -J \cdot \frac{du}{d\theta}$.

Since $\dot{r}^2 = J^2 \cdot \left(\frac{du}{d\theta}\right)^2$, the previous equation becomes

$$c^2(1 - E^2) - 2\mu \cdot u = -J^2 \left(\frac{du}{d\theta}\right)^2 - J^2u^2 \left(1 - \frac{2\mu}{c^2r}u\right) + O\left(\frac{1}{c^4}\right).$$

We can neglect the $O\left(\frac{1}{c^4}\right)$ terms which add only a small contribution to the trajectory. If we differentiate with respect to θ and then we divide by $\frac{du}{d\theta}$, we obtain exactly the equation derived in the Schwarzschild metric case, that is,

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{J^2} + \frac{3\mu}{c^2}u^2.$$

Of course, Einstein found the same solution and the same perihelion drift as in the case of Schwarzschild metric. □

Now, let us compute the Einstein metric

$$ds^2 = c^2 \left(1 - \frac{2\mu}{c^2r}\right) dt^2 - \left(1 + \frac{2\mu}{c^2r}\right) dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2$$

in Cartesian coordinates. As in the case of Schwarzschild metric, let $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$ and $r^2 = x^2 + y^2 + z^2$. It results $rdr = xdx + ydy + zdz$ which gives

$$dr = \sum_{\alpha=1}^{3} \frac{x^\alpha}{r} dx^\alpha, \quad dr^2 = \sum_{\alpha,\beta=1}^{3} \frac{x^\alpha x^\beta}{r^2} dx^\alpha dx^\beta.$$

Taking into account that

$$dx^2 + dy^2 + dz^2 = dr^2 + r^2 d\varphi^2 + r^2 \sin^2 \varphi d\theta^2 = \sum_{\alpha=1}^{3} (dx^\alpha)^2 \ ,$$

it results in

$$ds^2 = c^2dt^2 - dr^2 - r^2d\varphi^2 - r^2 \sin^2 \varphi d\theta^2 - \frac{2\mu}{c^2r}\left(c^2dt^2 + dr^2\right).$$

Therefore, *Einstein's metric in Cartesian coordinates* is

$$ds^2 = (dx^0)^2 - \sum_{\alpha=0}^{3}(dx^\alpha)^2 - \frac{2\mu}{c^2 r}\left((dx^0)^2 + \sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}dx^\alpha dx^\beta\right).$$

We determine the speed of light in the gravitational field described by Einstein's metric. We use the same technique as in the case of Schwarzschild metric.

If $X(t) = \left(ct, x^1(t), x^2(t), x^3(t)\right)$ is the worldcurve of an object parameterized by the time t, then, in the Minkowski metric, it is

$$\left(\frac{ds}{dt}\right)^2 = \|\dot{X}(t)\|^2 = c^2 - \left((\dot{x}^1(t))^2 + (\dot{x}^2(t))^2 + (\dot{x}^3(t))^2\right) = c^2 - v^2,$$

where v is the usual spatial speed of the object.

If the object is a photon, then $\left(\frac{ds}{dt}\right)^2 = c^2 - c^2 = 0$ and so

$$0 = \left(\frac{ds}{dt}\right)^2 = c^2 - \gamma^2 - \frac{2\mu}{c^2 r}\left(c^2 + \sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt}\right),$$

where $\gamma(t) = \sqrt{(\dot{x}^1(t))^2 + (\dot{x}^2(t))^2 + (\dot{x}^3(t))^2}$ is the speed of the photon in the gravitational field described by the metric above. In fact

$$\gamma = c \cdot \sqrt{1 - \frac{2\mu}{c^2 r}\left(1 + \frac{1}{c^2}\sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}\frac{dx^\alpha}{dt}\frac{dx^\beta}{dt}\right)}.$$

It remains, as an exercise, to determine the speed of a deflected photon using the same technique as in the case of Schwarzschild metric. We highlight the quick answer.

We determine γ along the worldcurve X of an undeflected photon in the (x^1, x^2) plane at the fixed distance h from x^2-axis.

So, the undeflected worldcurve of the photon is $X(t) := (ct, h, ct, 0)$. The deflection will add only lower order terms, therefore the deflected photon has in the same plane extra terms of order $O\left(\frac{1}{c}\right)$. Then, the deflected photon is parameterized by

$$X_d(t) := \left(ct, h + O\left(\frac{1}{c}\right), ct + O(1), 0\right).$$

Since $\dot{X}_d\ (t) := \left(c, O\left(\frac{1}{c}\right), c + O(1), 0\right)$ we have

$$\frac{dx^1}{dt} = O\left(\frac{1}{c}\right), \quad \frac{dx^2}{dt} = c + O(1), \quad \frac{dx^3}{dt} = 0.$$

It results in the approximation

$$\gamma^2 = c^2 - \frac{2\mu}{c^2 r}\left(c^2 + \frac{(x^2)^2}{r^2} \cdot c^2\right),$$

equivalent to

$$\gamma = c \cdot \sqrt{1 - \frac{2\mu}{c^2 r} - \frac{2\mu(x^2)^2}{c^2 r^3}}.$$

Taking into account

$$\sqrt{1 + 2A} \approx 1 + A,$$

the final result is

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu}{cr} \cdot \frac{(x^2)^2}{r^2}.$$

The total deflection is computed as in the Schwarzschild case. Therefore, we succeeded to prove

Theorem 10.6.2 (Einstein's Second Theorem) *Consider the Einstein metric*

$$ds^2 = c^2\left(1 - \frac{2\mu}{c^2 r}\right)dt^2 - \left(1 + \frac{2\mu}{c^2 r}\right)dr^2 - r^2 d\varphi^2 - r^2 \sin^2\varphi d\theta^2.$$

(i) In Cartesian coordinates, the above metric has the form

$$ds^2 = \sum_{i=0}^{3}\left(dx^i\right)^2 - \frac{2\mu}{c^2 r}\left((dx^0)^2 + \sum_{\alpha,\beta=1}^{3}\frac{x^\alpha x^\beta}{r^2}dx^\alpha dx^\beta\right).$$

(ii) A deflected photon in the (x^1, x^2)-plane, which comes from the undeflected photon
$X(t) = (ct, h, ct, 0)$, has the speed

$$\gamma = c - \frac{\mu}{cr} - \frac{\mu}{cr} \cdot \frac{(x^2)^2}{r^2}$$

and does not violate the speed of light limit.

(iii) The total deflection of the trajectory $X_d(t)$ of a deflected photon in the gravitational field, described by the Einstein metric, is $\frac{4GM}{c^2 h}$.

All the computations made in Einstein's metric lead to the same results in Schwarzschild's metric. According to these considerations, we can say that the Schwarzschild metric reduces to the Einstein metric in the weak field limit.

10.7 Black Holes: A Mathematical Introduction

Black hole physics is undoubtedly one of the most fascinating topic of General Relativity. Here we give only a short mathematical introduction and refer the interested reader to the book by Frolov and Novikov *Black Hole Physics* [103] for a detailed discussion.

10.7.1 Escape Velocity and Black Holes

Suppose we stay on the surface of the Earth imagined as a sphere and we vertically throw a ball. Depending on the speed of throwing, the ball can be higher and higher throwed, but after it reaches a maximum altitude it falls down attracted by the Earth. Which is the speed necessary such that the ball never return?

So, the gravitational force acts between the two involved bodies, the Earth and the ball. If the ball is at distance r from the centre of the Earth, we have

$$F = \frac{GMm}{r^2}; \; K_E = \frac{mv^2}{2}; \; P_E = -\frac{GMm}{r}.$$

Consider, on the surface of the Earth, the escape velocity v_e. The ball goes higher and higher loosing in time its speed. At infinity, its kinetic energy is 0, the same for the potential energy $-\frac{GMm}{r}$. This means that, at infinity, the total energy is 0. At each point of this rectilinear trajectory, the total energy, which is a constant, has to be 0, that is the *escape velocity* can be computed from the condition

$$\frac{mv^2}{2} - \frac{GMm}{r} = 0.$$

It results in the escape velocity formula

$$v_e^2 = \frac{2GM}{R_E},$$

where R_E is the radius of the Earth. If someone replaces the values, the escape velocity from the Earth gravitational field is almost 11 (km)/(s).

By definition, a black hole is a "cosmic body" having the escape velocity $> c$, where c is the speed of light in vacuum. According to the fact that there are no speeds greater than the speed of light in vacuum, let us compute how small should be the Earth such that $v_e = c$. We obtain

$$r = \frac{2GM}{c^2} \approx \frac{2 \cdot 6.67 \cdot 10^{-11} 6 \cdot 10^{24}}{9 \cdot 10^{16}} \approx 8.8\,(\text{mm}).$$

Therefore, if all the mass of the Earth is concentrated in a sphere with 8.8 (mm) radius, the Earth should be a black hole and not even photons can leave its surface.

Let us see the difference between the Earth, as we know, and the Earth as a black hole. We have to compute in both cases the gravitational force exerted to a $1(kg)$ body.

For the "usual Earth":

$$F = \frac{6.67 \cdot 10^{-11} \cdot 6 \cdot 10^{24} \cdot 1}{(24 \cdot 10^6)^2} \approx 9.8(\text{kg} \cdot \text{ m})/(\text{s}^2) \approx \text{g}$$

where g is the constant gravitational acceleration as we expected.
For the "black hole Earth", we have

$$F = \frac{6.67 \cdot 10^{-11} \cdot 6 \cdot 10^{24} \cdot 1}{(8 \cdot 10^{-3})^2} \approx \frac{1}{2} \cdot 10^{19}(\text{kg} \cdot \text{ m})/(\text{s}^2),$$

that is, a tremendous huge force exerted by the black hole to all the bodies on its surface.

Suppose now the Earth transformed instantaneously into a black hole. Is the Moon trajectory affected? This is only a mathematical discussion, of course. Let us look at the formula

$$F = \frac{GM\cancel{m}}{r^{\cancel{2}}} = \frac{\cancel{m}v^2}{\cancel{r}}.$$

It is the same in both cases, because r is measured from the centre of the Earth. We deduce that the Moon continues to orbit the black hole Earth such it does now.

Some other considerations about black holes can be seen when we study them using metrics.

10.7.2 The Rindler Metric and Pseudo-Singularities

Let us define the *Rindler metric* as

$$ds^2 = \frac{(\bar{x}^1)^2}{b^2}(\bar{x}^0)^2 - (\bar{x}^1)^2.$$

What is happening with this metric when $\bar{x}^1 \to 0$? If we are looking only at the first term, the first coefficient approaches 0. We can think that the metric fails to exist. A singularity seems to be highlighted. However, we will show that a suitable change of coordinates transforms the Rindler metric into the ordinary Minkowski metric. We may conclude that $\bar{x}^1 = 0$ is not a physical singularity, but a *pseudo-singularity* (or a *geometric singularity*), that is one which can be removed by a convenient change of coordinates.

Consider the change of coordinates

$$C : \begin{cases} \bar{x}^0(x^0, x^1) = b \tanh^{-1} \dfrac{x^0}{x^1} \\ \bar{x}^1(x^0, x^1) = \sqrt{(x^1)^2 - (x^0)^2}, \end{cases}$$

where $\tanh^{-1}(y) = \dfrac{1}{2} \ln \dfrac{1+y}{1-y}$. If we compute

$$dC = \begin{pmatrix} \dfrac{\partial \bar{x}^0}{\partial x^0} & \dfrac{\partial \bar{x}^0}{\partial x^1} \\ \dfrac{\partial \bar{x}^1}{\partial x^0} & \dfrac{\partial \bar{x}^1}{\partial x^1} \end{pmatrix}$$

the four components are

$$\frac{\partial \bar{x}^0}{\partial x^0} = \frac{bx^1}{(x^1)^2 - (x^0)^2}; \quad \frac{\partial \bar{x}^0}{\partial x^1} = \frac{-bx^0}{(x^1)^2 - (x^0)^2};$$

$$\frac{\partial \bar{x}^1}{\partial x^0} = \frac{-x^0}{\sqrt{(x^1)^2 - (x^0)^2}}; \quad \frac{\partial \bar{x}^1}{\partial x^1} = \frac{x^1}{\sqrt{(x^1)^2 - (x^0)^2}}.$$

Exercise 10.7.1 Compute $dC^t \cdot R\, dC$.

Solution. We have to compute

$$\begin{pmatrix} \frac{bx^1}{(x^1)^2-(x^0)^2} & \frac{-x^0}{\sqrt{(x^1)^2-(x^0)^2}} \\ \frac{-bx^0}{(x^1)^2-(x^0)^2} & \frac{x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix} \begin{pmatrix} \frac{(\bar{x}^1)^2}{b^2} & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{bx^1}{(x^1)^2-(x^0)^2} & \frac{-bx^0}{(x^1)^2-(x^0)^2} \\ \frac{-x^0}{\sqrt{(x^1)^2-(x^0)^2}} & \frac{x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix},$$

that is,

$$\begin{pmatrix} \frac{bx^1}{(x^1)^2-(x^0)^2} & \frac{-x^0}{\sqrt{(x^1)^2-(x^0)^2}} \\ \frac{-bx^0}{(x^1)^2-(x^0)^2} & \frac{x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix} \begin{pmatrix} \frac{(x^1)^2-(x^0)^2}{b^2} & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{bx^1}{(x^1)^2-(x^0)^2} & \frac{-bx^0}{(x^1)^2-(x^0)^2} \\ \frac{-x^0}{\sqrt{(x^1)^2-(x^0)^2}} & \frac{x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{bx^1}{(x^1)^2-(x^0)^2} & \frac{-x^0}{\sqrt{(x^1)^2-(x^0)^2}} \\ \frac{-bx^0}{(x^1)^2-(x^0)^2} & \frac{x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix} \begin{pmatrix} \frac{x^1}{b} & -\frac{x^0}{b} \\ \frac{x^0}{\sqrt{(x^1)^2-(x^0)^2}} & \frac{-x^1}{\sqrt{(x^1)^2-(x^0)^2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad \square$$

The metric in coordinates (x^0, x^1) becomes

$$ds^2 = (dx^0)^2 - (dx^1)^2.$$

Therefore we have proved

Theorem 10.7.2 *The change of coordinates*

$$C : \begin{cases} \bar{x}^0(x^0, x^1) = b \tanh^{-1} \frac{x^0}{x^1} \\ \bar{x}^1(x^0, x^1) = \sqrt{(x^1)^2 - (x^0)^2} \end{cases}$$

transforms the Rindler metric

$$ds^2 = \frac{(\bar{x}^1)^2}{b^2}(\bar{x}^0)^2 - (\bar{x}^1)^2$$

into the Minkowski metric

$$ds^2 = (dx^0)^2 - (dx^1)^2.$$

As we discussed, $\bar{x}^1 = 0$ is not a physical singularity, but a pseudo-singularity.

The lines $x^1 = x^0$ and $x^1 = -x^0$ are called the *horizon of the geometric singularity* $\bar{x}^1 = 0$. Removing of geometric singularities is part of the mathematical theory of black holes we present.

10.7.3 Black Holes in the Schwarzschild Metric

When we studied the vacuum field equations $R_{ij} = 0$, we started from the Schwarzschild intuition to look for a spherically symmetric solution which describes the relativistic field outside of a non-rotating, massive body.

In the coordinate system $(x^0 = ct, r, \varphi, \theta)$ Schwarzschild chose the form of the solution as

$$ds^2 = c^2 \cdot e^T dt^2 - e^Q dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2,$$

with $T := T(r)$, $Q := Q(r)$ two real functions we need to determine. (In [154], a more general approach is presented considering $T := T(r, t)$ and $Q := Q(r, t)$).

The non-zero Christoffel symbols are

$$\Gamma^0_{01} = \Gamma^0_{10} = \frac{T'}{2},\ \Gamma^1_{00} = -\frac{T'}{2}e^{T-Q},\ \Gamma^1_{11} = \frac{Q'}{2},\quad \Gamma^1_{22} = -re^{-Q},\quad \Gamma^1_{33} = -re^{-Q}\sin^2\varphi,$$

$$\Gamma^2_{21} = \Gamma^2_{12} = \frac{1}{r},\quad \Gamma^2_{33} = -\sin\varphi\cos\varphi,\quad \Gamma^3_{31} = \Gamma^3_{13} = \frac{1}{r},\quad \Gamma^3_{23} = \Gamma^3_{32} = \cot\varphi,$$

where we denote by $'$ the derivative with respect to r. The computations lead to

$$T = -Q = \ln\left(1 + \frac{B}{r}\right).$$

The obtained Schwarzschild metric is

$$ds^2 = c^2 \cdot \left(1 + \frac{B}{r}\right)dt^2 - \frac{1}{\left(1 + \frac{B}{r}\right)}dr^2 - r^2 d\varphi^2 - r^2\sin^2\varphi d\theta^2.$$

Then we find $B = \frac{-2GM}{c^2}$, that is, the gravitational Newtonian potential $\phi(x, y, z) = -\frac{GM}{r}$ is involved in the coefficients of Schwarzschild metric. We have remembered here these results for two reasons. The first one is the following exercise we need to understand the behaviour of the Riemann curvature tensor at the surface of a black hole.

Exercise 10.7.3 Compute R^0_{101}.

Solution. Replacing in R^i_{jkl} formula (in the case $i = k = 0,\ j = l = 1$) the above corresponding Christoffel symbols, we find

$$R^0_{101} = -\frac{1}{2}T'' + \frac{1}{4}T'Q' - \frac{1}{4}(Q')^2,$$

that is,

$$R^0_{101} = \frac{-B}{r^3}\left(\frac{1}{1 + \frac{B}{r}}\right).\ \square$$

The second reason is related to the quantity $r_s := \frac{2GM}{c^2}$, called the Schwarzschild radius, which gives Schwarzschild metric in the form

$$ds^2 = c^2 \cdot \left(1 - \frac{r_s}{r}\right) dt^2 - \frac{1}{\left(1 - \frac{r_s}{r}\right)} dr^2 - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2.$$

If we remember the nature of the Schwarzschild metric, the first term is positive, the other three are negative. If we look at the first term in the case of the Sun, if we replace the gravitational constant G, $M = M_{sun}$, the speed of light in vacuum c and $r = r_{sun}$ we have $1 - \frac{r_s}{r} > 0$. If r approaches $r_s = \frac{2GM}{c^2}$ the first two terms of the metric have the properties $g_{00} \to 0$ and $g_{11} \to -\infty$. So, the Schwarzschild metric becomes singular. If we compute r_s in the case of our Sun, we find $r_s \approx 3km$. So, the anomaly appears when the entire mass of our Sun is concentrated in a sphere with a radius as r_s.

Such a sphere is a *black hole*. The interior of the sphere is called an *interior of the black hole* and it is characterized by the condition $r < r_s$. The surface of the sphere is called the *event horizon* and it is characterized by the condition $r = r_s$. The *exterior of the black hole* is characterized by the condition $r > r_s$.

A clock at $r = r_s$ has its proper time $d\tau = \sqrt{1 - \frac{r_s}{r}} dt \to 0$. Which means that the clock is slowed down at maximum; a clock outside the black hole works faster.

What is going on in the interior of the black hole?

We observe $g_{00} = c^2 \left(1 - \frac{r_s}{r}\right) < 0$ and $g_{11} = -\frac{1}{1 - \frac{1}{r_s}} > 0$. Therefore the signs are opposite with respect to the standard ones. It results that t becomes a spatial coordinate and r becomes a temporal coordinate inverting their roles!

However we can prove that the singularity $r = r_s$ is in fact a pseudo-singularity.

First, let us see what is happening to Riemann curvature tensor at $r = r_s$.

If we denote $t := x^0;\ r := x^1;\ \varphi := x^2;\ \theta := x^3$ the old coordinates, we may construct the new coordinates:

$\bar{x}^0 = (x^0 - t_0)\sqrt{1 - \frac{r_s}{r}}; \quad \bar{x}^1 = \frac{x^1 - r_0}{\sqrt{1 - \frac{r_s}{r}}}; \quad \bar{x}^2 = r_0 \left(x^2 - \frac{\pi}{2}\right); \quad \bar{x}^3 = r_0 \left(x^3 - \theta_0\right).$

It results

$$\frac{\partial x^0}{\partial \bar{x}^0} = \frac{1}{\sqrt{1 - \frac{r_s}{r}}}; \quad \frac{\partial x^1}{\partial \bar{x}^1} = \sqrt{1 - \frac{r_s}{r}}; \quad \frac{\partial x^2}{\partial \bar{x}^2} = \frac{1}{r_0}; \quad \frac{\partial x^3}{\partial \bar{x}^3} = \frac{1}{r_0};$$

all the other possible partial derivatives are null.

The new Riemann curvature tensor $\bar{R}^0_{101}$ is

$$\bar{R}^0_{101} = \frac{\partial \bar{x}^0}{\partial x^0} \frac{\partial x^1}{\partial \bar{x}^1} \frac{\partial x^0}{\partial \bar{x}^0} \frac{\partial x^1}{\partial \bar{x}^1} R^0_{101} = \left(1 - \frac{r_s}{r}\right) \frac{r_s}{r_0^3} \frac{1}{1 - \frac{r_s}{r}} = \frac{r_s}{r_0^3}.$$

Therefore, at the surface of the black hole, that is when $r = r_s$, the Riemann curvature tensor is well defined. The surface of a black hole is not a physical singularity.

At the surface of a black hole, that is when $r = r_s$, the Kruskal–Szekeres metric gives more information than Schwarzschild metric.

We act only on the first two coordinates of the Schwarzschild metric, the other two remain unchanged. The Kruskal–Szekeres coordinates look different inside the black hole comparing to the case of the exterior of the black hole.

In the interior of the black hole the Kruskal–Szekeres coordinates are

$$KS(r < r_s) : \begin{cases} V(t,r) = \sqrt{1 - \dfrac{r}{r_s}} e^{r/2r_s} \cosh \dfrac{t}{2r_s} \\ U(t,r) = \sqrt{1 - \dfrac{r}{r_s}} e^{r/2r_s} \sinh \dfrac{t}{2r_s} \end{cases}$$

$$U^2 - V^2 = \left(\frac{r}{r_s} - 1\right) e^{r/r_s} < 0; \quad \frac{U}{V} = \tanh \frac{t}{2r_s}, \text{ i.e. } t = 2r_s \tanh^{-1} \frac{V}{U}.$$

When $r < r_s$, that is in the interior of the black hole, we have

$\dfrac{\partial V}{\partial t} = A \cdot \sinh \dfrac{t}{2r_s}$ and $\dfrac{\partial U}{\partial t} = A \cdot \cosh \dfrac{t}{2r_s}$ where $A = \sqrt{1 - \dfrac{r}{r_s}} e^{r/r_s} \dfrac{t}{2r_s}$.

$\dfrac{\partial V}{\partial r} = B \cdot \cosh \dfrac{t}{2r_s}$ and $\dfrac{\partial U}{\partial r} = B \cdot \sinh \dfrac{t}{2r_s}$ where $B = \dfrac{-1\dfrac{1}{r_s}}{2\sqrt{1 - \dfrac{r}{r_s}}} e^{r/2r_s} + \dfrac{1}{2r_s}\sqrt{1 - \dfrac{r}{r_s}} e^{r/2r_s}$.

$$\begin{pmatrix} A \sinh \frac{t}{2r_s} & A \cosh \frac{t}{2r_s} \\ B \cosh \frac{t}{2r_s} & B \sinh \frac{t}{2r_s} \end{pmatrix} \begin{pmatrix} \frac{4r_s^3}{r} e^{-r/r_s} & 0 \\ 0 & -\frac{4r_s^3}{r} e^{-r/r_s} \end{pmatrix} \begin{pmatrix} A \sinh \frac{t}{2r_s} & B \cosh \frac{t}{2r_s} \\ A \cosh \frac{t}{2r_s} & B \sinh \frac{t}{2r_s} \end{pmatrix}$$

$$= \frac{4r_s^3}{r} e^{-r/r_s} \cdot \begin{pmatrix} -A^2 & 0 \\ 0 & B^2 \end{pmatrix} = \begin{pmatrix} 1 - \dfrac{r_s}{r} & 0 \\ 0 & -\dfrac{1}{1 - \dfrac{r_s}{r}} \end{pmatrix},$$

the last matrix equality because

$$-\frac{4r_s^3}{r} e^{-r/r_s} \cdot A^2 = 1 - \frac{r_s}{r}$$

and

$$B^2 \cdot \frac{4r_s^3}{r} e^{-r/r_s} = -\frac{1}{1 - \frac{r_s}{r}}.$$

We have proved the following

Theorem 10.7.4 *When the metric is the Schwarzschild one, in the interior of a black hole described by the condition $r < r_s$, the Kruskal–Szekeres coordinates*

$$KS(r < r_s) : \begin{cases} V(t,r) = \sqrt{1 - \frac{r}{r_s}} e^{r/2r_s} \cosh \frac{t}{2r_s} \\ U(t,r) = \sqrt{1 - \frac{r}{r_s}} e^{r/2r_s} \sinh \frac{t}{2r_s} \end{cases}$$

transform the Schwarzschild metric into the Kruskal–Szekeres metric

$$ds^2 = \frac{4r_s^3}{r} e^{-r/r_s} \left(dV^2 - dU^2\right) - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2.$$

In the case of the exterior of the black hole, the Kruskal–Szekeres coordinates are

$$KS(r > r_s) : \begin{cases} V(t,r) = \sqrt{\frac{r}{r_s} - 1} e^{r/2r_s} \cosh \frac{t}{2r_s} \\ U(t,r) = \sqrt{\frac{r}{r_s} - 1} e^{r/2r_s} \sinh \frac{t}{2r_s} \end{cases}$$

$U^2 - V^2 = \left(\frac{r}{r_s} - 1\right) e^{r/r_s} > 0$; $\frac{U}{V} = \tanh \frac{t}{2r_s}$, i.e. $t = 2r_s \tanh^{-1} \frac{V}{U}$. Similar computations lead to

Theorem 10.7.5 *When the metric is the Schwarzschild one, in the exterior of a black hole described by the condition $r > r_s$, the Kruskal–Szekeres coordinates*

$$KS(r > r_s) : \begin{cases} V(t,r) = \sqrt{\frac{r}{r_s} - 1} e^{r/2r_s} \cosh \frac{t}{2r_s} \\ U(t,r) = \sqrt{\frac{r}{r_s} - 1} e^{r/2r_s} \sinh \frac{t}{2r_s} \end{cases}$$

transform the Schwarzschild metric into the same Kruskal–Szekeres metric

$$ds^2 = \frac{4r_s^3}{r} e^{-r/r_s} \left(dV^2 - dU^2\right) - r^2 d\varphi^2 - r^2 \sin^2 \varphi d\theta^2.$$

In both cases at $r = r_s$ the singularity has been removed. The only physical singularity corresponds to $r = 0$.

Corollary 10.7.6 *The event horizon of a black hole is a geometric singularity only.*

Let us discuss now the entropy of a black hole. From the Special Relativity chapter we know the following formulas and their meaning:

$$E = hf = h\frac{c}{\lambda};\ E = mc^2.$$

From this section, we know that the dimension of a black hole is related to its Schwarzschild radius r_s. If a photon is captured by a black hole of mass M and energy E, the black hole changes its mass and energy. This fact can be represented by the formula

$$\Delta M = \frac{\Delta E}{c^2} = \frac{h}{\lambda c} = \frac{h}{r_s c},$$

because λ becomes r_s.

But $\lambda = r_s = \dfrac{2MG}{c^2}$ suggests a variation of the radius described by the formula

$$\Delta r = \frac{2\Delta MG}{c^2} = \frac{2G}{c^2}\Delta M = \frac{2G}{c^2}\frac{h}{r_s c},$$

that is

$$r_s \Delta r = \frac{2Gh}{c^3}.$$

It is worth noticing that the right member is a constant. Furthermore, it is easy to see that the area of a black hole is the area of a sphere of radius r_s,

$$A = 4\pi r_s^2.$$

The derivative with respect to r leads to

$$\frac{dA}{dr} = 8\pi r_s,$$

i.e.

$$dA = 8\pi r_s dr = \frac{16\pi Gh}{c^3}.$$

Therefore, if a photon is captured by a black hole, the black hole area is increasing by the quantity

$$dA = \frac{16\pi Gh}{c^3}.$$

If dS is a unitary entropy, the variation of the area becomes

$$dA = \frac{16\pi Gh}{c^3} dS,$$

therefore

$$S = \frac{c^3}{16\pi G h} A.$$

This formula is known as *Bekenstein–Hawking formula* for the *black hole entropy*. See [103] for a detailed discussion.

With these considerations in mind, it is straightforward to define the *black hole temperature*. The formula which connects the variation of energy, the temperature, and the variation in entropy is

$$dE = TdS.$$

If we consider only "one unit" of variation for entropy, i.e. a single photon which changes the black hole energy, we have that

$$dE = T = \frac{hc}{\lambda} = \frac{hc}{r_s} = \frac{hc^3}{2MG}.$$

That is, a single photon changes the temperature of the black hole such that T is proportional to $\frac{1}{M}$. We deduce that smaller black holes are warmer than the massive ones.

10.7.4 The Light Cone in the Schwarzschild Metric

In Minkowski metric, the trajectories of light rays are determined by the condition $ds^2 = 0$. Since the Minkowski metric in geometric coordinates is

$$ds^2 = (dx^0)^2 - (dx^1)^2,$$

the previous condition becomes

$$x^0 = x^1, \quad x^0 = -x^1,$$

therefore the light cone is highlighted.

In (t, r) coordinates, let us consider the 3-plane $\varphi = 0$ and the corresponding Schwarzschild metric in geometric coordinates

$$ds^2 = \left(1 - \frac{r_s}{r}\right) dt^2 - \frac{1}{\left(1 - \frac{r_s}{r}\right)} dr^2.$$

The condition $ds^2 = 0$ leads to

$$\left(1 - \frac{r_s}{r}\right)^2 dt^2 = dr^2,$$

i.e.

$$\frac{dr}{dt} = \pm\left(1 - \frac{r_s}{r}\right).$$

It is obvious that this is different with respect to the Minkowski condition

$$\frac{dr}{dt} = \pm 1.$$

We can write the equivalent formula

$$\pm \frac{d\left(\frac{r}{r_s}\right)}{\left(1 - \frac{r_s}{r}\right)} = d\left(\frac{t}{r_s}\right).$$

Let us denote $X := \left(\frac{r}{r_s}\right)$ and $Y := \left(\frac{t}{r_s}\right)$. It remains to solve

$$\pm \int \frac{XdX}{X+1} = \int dY.$$

The solutions are $Y = X + \ln|x-1|$ and $Y = -X - \ln|x-1|$.

The graphs of the functions $f(X) = X + \ln(1-X)$ and $g(X) = -X + \ln(1-X)$, both defined on $(0, 1)$, highlight the light cone in the interior of the black hole.

Suppose $x_0 \in (0, 1)$ and the tangent lines at $(x_0, f(x_0))$ and $(x_0, f(x_0))$. The parallel lines to the tangents in $(x_0, 0)$ show how the light cone looks like in the interior of the black hole.

The same, the graphs of the functions $h(X) = X + \ln(X-1)$ and $l(X) = -X + \ln(X-1)$, both defined when $x \in (1, \infty)$ highlight the light cone outside the black hole.

In summary, we gave only some main features of black holes but the physics and the mathematics of these gravitational systems is extremely reach and deserve to be explored in details. Furthermore, after the direct detection of the black hole shadow by the Event Horizon Telescope collaboration, a new era started in this fascinating sector of Physics. These gravitational objects, considered only exotic theoretical objects until recently, have become an amazing arena for observational astrophysics and cosmology.

10.8 Cosmological Solutions of the Einstein Field Equations: The Friedmann–Lemaître–Robertson–Walker Models of Universe

If we intend to find a metric for the general Einstein field equations describing the Universe, we have to consider the fact that the observed Universe appears homogeneous and isotropic beyond a given scale according to the Cosmological Principle, therefore we have to consider, at the beginning, a spherical symmetry for the cosmic space-time.[2]

The spatial part has to be as

$$dr^2 + q^2(r)\left(d\theta^2 + \sin^2\theta d\phi^2\right),$$

where $q(r)$ will be determined, so, we can try with the metric

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + q^2(r)d\theta^2 + q^2(r)\sin^2\theta d\phi^2\right]$$

which introduces a new function $a(t)$ necessary to preserve the spherical symmetry of the spatial part of the metric which can, eventually, expand under a homothetic transformation. In this way, $a(t)$ becomes an expansion factor of the Universe. We will discuss this fact a little bit later. Observe that we are working in geometric coordinates, that is $c = 1$.

Let us search for $a(t)$ and $q(r)$ such that the previous metric satisfies the Einstein field equations. To address the answer, there are three possible forms for $q(r)$ depending on a constant of integration, while $a(t)$ is determined from Einstein's field equations. We prove

Theorem 10.8.1 *The following three metrics*

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

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

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2\right]$$

satisfy the Einstein field equations

$$R_{ij} - \frac{1}{2}Rg_{ij} = KT_{ij}$$

---

[2] From observational surveys, the Universe can be considered homogeneous and isotropic beyond scales of the order 100–120 Megaparsecs. See [159] for details. This means that, over these scales, no large-scale structure, like clusters or super-clusters of galaxies are detected. According to these data, matter density can be considered homogeneously distributed in all directions.

in the case when the contravariant energy–momentum tensor is describing a perfect fluid with components

$$T^{ij} = (\rho_0 + p_0)u^i u^j - p_0 g^{ij},$$

where g^{ij} are the inverse components of the metric tensor matrix which satisfies Einstein's field equations, ρ_0 is the density of the fluid, p_0 is the pressure of the fluid, and u^i are the components $(u^t, v_x u^t, v_y u^t, v_z u^t)$ of the fluid 4-velocity. For the moment, ρ_0 and p_0 are assumed constant.

Proof We start by calculating the Ricci symbols.

$$g_{00} = 1,\ g_{11} = -a^2(t),\ g_{22} = -a^2(t)q^2(r),\ g_{33} = -a^2(t)q^2(r)\sin^2\theta$$

$$g^{00} = 1,\ g^{11} = -\frac{1}{a^2(t)},\ g^{22} = -\frac{1}{a^2(t)q^2(r)},\ g^{33} = -\frac{1}{a^2(t)q^2(r)\sin^2\theta}.$$

We observe

$$\Gamma^i_{jk} = g^{is}\Gamma_{jk,s} = g^{ii}\Gamma_{jk,i};\ \Gamma^i_{jk} = 0,\ i \neq j \neq k.$$

Therefore

$$\Gamma^0_{11} = a\cdot\dot{a},\ \Gamma^0_{22} = a\cdot\dot{a}\cdot q^2,\ \Gamma^0_{33} = a\cdot\dot{a}\cdot q^2\cdot\sin^2\theta, \text{otherwise } \Gamma^0_{ij} = 0,$$
$$\Gamma^1_{01} = \Gamma^1_{10} = \frac{\dot{a}}{a},\ \Gamma^1_{22} = -q\cdot q',\ \Gamma^1_{33} = -q\cdot q'\cdot\sin^2\theta, \text{otherwise } \Gamma^1_{ij} = 0,$$
$$\Gamma^2_{02} = \Gamma^2_{20} = \frac{\dot{a}}{a},\ \Gamma^2_{12} = \Gamma^2_{21} = \frac{q'}{q},\ \Gamma^2_{33} = -\sin\theta\cos\theta, \text{otherwise } \Gamma^2_{ij} = 0,$$
$$\Gamma^3_{03} = \Gamma^3_{30} = \frac{\dot{a}}{a},\ \Gamma^3_{13} = \Gamma^3_{31} = \frac{q'}{q},\ \Gamma^3_{32} = \Gamma^3_{23} = -\cot\theta, \text{otherwise } \Gamma^3_{ij} = 0.$$

If we compute

$$R_{00} = R^s_{0s0} = -\frac{\partial\Gamma^1_{01}}{\partial t} - \frac{\partial\Gamma^2_{02}}{\partial t} - \frac{\partial\Gamma^3_{03}}{\partial t} - \Gamma^1_{01}\Gamma^1_{10} - \Gamma^2_{02}\Gamma^2_{20} - \Gamma^3_{03}\Gamma^3_{30},$$

it results

$$R_{tt} = R_{00} = -3\frac{\ddot{a}}{a}.$$

We obtain

$$R_{rr} = R_{11} = 2\dot{a}^2 + \ddot{a}\cdot a - 2\frac{q''}{q},$$

$$R_{\theta\theta} = R_{22} = 2q^2\cdot\dot{a}^2 + q^2\cdot a\cdot\ddot{a} - q\cdot q'' + 1 - (q')^2,$$

$$R_{\phi\phi} = R_{33} = R_{22}\sin^2\theta.$$

Using $R^i_j = g^{is} R_{sj}$, we rise an index, therefore

$$R^t_t = -3\frac{\ddot{a}}{a},$$

$$R^r_r = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} + 2\frac{q''}{a^2 \cdot q},$$

$$R^\theta_\theta = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} - \frac{1}{a^2 \cdot q^2}\left(1 - q \cdot q'' - (q')^2\right),$$

$$R^\phi_\phi = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} - \frac{1}{a^2 \cdot q^2}\left(1 - q \cdot q'' - (q')^2\right) = R^\theta_\theta.$$

The key of finding the metric is related to the way the physicists describe the energy–momentum tensor. They look at the galaxies in the Universe such that they are imagined as the molecules of an ideal gas which move arbitrarily. In this case, the gas is described as in the statement of the theorem, by the contravariant energy–momentum tensor

$$T^{ij} = (\rho_0 + p_0)u^i u^j - p_0 g^{ij},$$

where

- g^{ij} are the inverse components of the metric tensor matrix which satisfies Einstein's field equations $R_{ij} - \frac{1}{2}R g_{ij} = K T_{ij}$,
- ρ_0 is the density of the gas,
- p_0 is the constant pressure of the gas, and
- u^i are the components $(u^t, v_x u^t, v_y u^t, v_z u^t)$ of the gas 4-velocity.

It is convenient to use the (1, 1) tensor T^i_j by lowering the second index, so

$$T^i_j = (\rho_0 + p_0)u^i g_{jk} u^k - p_0 \delta^i_j.$$

Our chosen metric has $g_{tt} = 1$.

This ideal fluid is, by definition, at rest in these comoving coordinates, therefore the conditions $u^r = u^\theta = u^\phi = 0$ for every t and $u^i g_{ij} u^j = 1$ lead to $1 = g_{tt}(u^t)^2$, that is, $u^t = 1$.

It follows that $T^t_t = \rho_0$ and $T^r_r = T^\theta_\theta = T^\phi_\phi = -p_0$, that is, $T = T^i_i = \rho_0 - 3p_0$.

If we arrange Einstein's field equation in the form

$$R^i_j = K \cdot \left(T^i_j - \frac{1}{2}\delta^i_j T\right)$$

we obtain $R^r_r = R^\theta_\theta = R^\phi_\phi = -\frac{K}{2}(\rho_0 - p_0)$.

The condition $R_r^r = R_\theta^\theta$ highlights the equality

$$2\frac{q''}{a^2 \cdot q} = -\frac{1}{a^2 \cdot q^2}\left(1 - q \cdot q'' - (q')^2\right)$$

and it remains to solve the differential equation

$$(q')^2 - q \cdot q'' = 1.$$

I. Determining $q(r)$.

From the beginning, we observe that $q(r) = r$ is a possible solution.

We continue: for $p := q' = \frac{dq}{dr}$ we obtain $q'' = \frac{dp}{dr} = \frac{dp}{dq}\frac{dq}{dr} = \frac{dp}{dq}p$.

The differential equation transforms to $p^2 - qp\frac{dp}{dq} = 1$, that is,

$$2\frac{dq}{q} = \frac{2pdp}{p^2 - 1}.$$

The solution written as $2\ln|q| = \ln q^2 = \ln|p^2 - 1| - \ln|k|$ leads first to $q^2 = \frac{p^2 - 1}{k}$, then, after replacing in $(q')^2 - q \cdot q'' = 1$, to $q'' = kq$. It results in

$$\left(\frac{dq}{dr}\right)^2 = (q')^2 = 1 + q \cdot q'' = 1 + kq^2.$$

Since in the metric appears q^2, we are not interested in the solutions with minus. Without loosing the generality, we can suppose that $q(0) = 0$ and $q'(0) = 1$. Therefore, having these initial conditions, we have to solve

$$\frac{dq}{\sqrt{1 + kq^2}} = dr.$$

Case $k > 0$. We choose $k = \frac{1}{R^2}$. We have

$$r = \int \frac{1}{\sqrt{1 + \left(\frac{q}{R}\right)^2}}dq = R\sinh^{-1}\frac{q}{R},$$

that is, $q = R\sinh\frac{r}{R}$. In this case the metric is

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + R^2\sinh^2\frac{r}{R}d\theta^2 + R^2\sinh^2\frac{r}{R}\sin^2\theta d\phi^2\right].$$

Case $k < 0$. We choose $k = -\frac{1}{R^2}$. We have

$$r = \int \frac{1}{\sqrt{1-\left(\frac{q}{R}\right)^2}}dq = R\arcsin\frac{q}{R},$$

that is $q = R\sin\frac{r}{R}$. In this case the metric is

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

For $q(r) = r$ the metric is

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + r^2 d\theta^2 + r^2\sin^2\theta d\phi^2\right].$$

□

Let us now determine $a(t)$. To proceed, we consider again the above field equations:

$$R_t^t = -3\frac{\ddot{a}}{a} = K\cdot\left(T_t^t - \frac{1}{2}T\right) = \frac{K}{2}\cdot(\rho_0 + 3p_0)$$

$$R_r^r = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} + 2\frac{q''}{a^2\cdot q} = K\cdot\left(T_\theta^\theta - \frac{1}{2}T\right) = -\frac{K}{2}(\rho_0 - p_0).$$

Since $q'' = kq$, in the case when $k = \pm\frac{1}{R^2}$, the following two equations have to be considered:

$$\frac{\ddot{a}}{a} = -\frac{K}{6}\cdot(\rho_0 + 3p_0),$$

$$2\frac{\dot{a}^2}{a^2} + \frac{\ddot{a}}{a} - 2\frac{k}{a^2} = \frac{K}{2}(\rho_0 - p_0).$$

It results in the equation

$$\frac{\dot{a}^2}{a^2} - \frac{k}{a^2} = \frac{K}{3}\rho_0,$$

that is,

$$\dot{a}^2 - \frac{K}{3}\rho_0\cdot a^2 = k,$$

which can be solved. The equation can be written as

$$\dot{a}^2 - B \cdot a^2 = k,$$

where $B = \frac{K}{3}\rho_0 > 0$. Since in metric appears a^2, as in the case of $q(r)$, we are not interested in the solutions with minus. Furthermore, we are not interested in using constants which can be eliminated by a convenient change of coordinates.

In the case $k = \frac{1}{R^2} > 0$, if we arrange the equation in the form

$$\frac{1}{k}\dot{a}^2 - \frac{B}{k} \cdot a^2 = 1,$$

the solution is $a(t) = \frac{1}{R\sqrt{B}} \sinh(t\sqrt{B})$. Replacing B, it results in

$$a(t) = \frac{1}{R \cdot \sqrt{\frac{K}{3}\rho_0}} \sinh\left(t\sqrt{\frac{K}{3}\rho_0}\right).$$

In the case $k = -\frac{1}{R^2} < 0$, if we arrange the equation in the form

$$-R^2\dot{a}^2 + R^2 B \cdot a^2 = 1,$$

the solution is $a(t) = \frac{1}{R\sqrt{B}} \cosh(t\sqrt{B})$. After replacing B,

$$a(t) = \frac{1}{R \cdot \sqrt{\frac{K}{3}\rho_0}} \cosh\left(t\sqrt{\frac{K}{3}\rho_0}\right).$$

If $q(r) = r$ it results in $q''(r) = 0$. Let us consider the first two equations:

$$\frac{\ddot{a}}{a} = -\frac{K}{6} \cdot (\rho_0 + 3p_0),$$

$$2\frac{\dot{a}^2}{a^2} + \frac{\ddot{a}}{a} = \frac{K}{2}(\rho_0 - p_0).$$

This kind of differential equations in $a(t)$ is called the $(FLRW)$ equations. We obtain

$$\frac{\dot{a}^2}{a^2} = \frac{K}{3}\rho_0.$$

Taking into account our notation $B = \dfrac{K}{3}\rho_0$, two solutions are possible: $a_1(t) = e^{t\sqrt{B}}$ and $a_2(t) = e^{-t\sqrt{B}}$.

We may observe that as $t \to +\infty,\ a_2(t) \to 0$ which does not correspond to the known expansion of the Universe, related to the observational evidences. The other solution can be accepted. As we will see in a following subsection, it is related to the Hubble constant.

Let us stress again that these metrics have been obtained in the case when T_{ij} has the above special form. We may conceive other possible T_{ij} having the property $T_r^r = T_\theta^\theta = T_\phi^\phi$ and some other metrics can appear. □

Denote $d\Omega^2 := d\theta^2 + \sin^2\theta d\phi^2$. In the process of finding the metric, we have used

$$\frac{dq^2}{1+kq^2} = dr^2.$$

Replace this position in the metric

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + q^2 d\theta^2 + q^2\sin^2\theta d\phi^2\right],$$

we can write all possible metrics in the form

$$ds^2 = dt^2 - a^2(t)\left[\frac{dq^2}{1+kq^2} + q^2 d\Omega^2\right].$$

This metric is known as the *Friedmann–Lemaître–Robertson–Walker metric (or FLRW metric) of the Universe.*

Problema 10.8.2 Consider the case of the cosmological fluid such that the contravariant energy–momentum tensor is

$$T^{ij} = (\rho_0 + p_0)u^i u^j - p_0 g^{ij} + \frac{\Lambda}{K}g^{ij},$$

where Λ is the cosmological constant. Under the conditions of the previous theorem, let us find the coefficients if the metric for Universe is

$$ds^2 = c^2dt^2 - a^2(t)\left[dr^2 + q^2(r)d\theta^2 + q^2(r)\sin^2\theta d\phi^2\right].$$

Hint: We obtain $T_t^t = \rho_0 + \dfrac{\Lambda}{K}$, $T_r^r = T_\theta^\theta = T_\phi^\phi = -p_0 + \dfrac{\Lambda}{K}$ and $T = \rho_0 - 3p_0 + \dfrac{4\Lambda}{K}$, but we have to complete the computations using

$$R_j^i - \Lambda\delta_j^i = K\cdot\left(T_j^i - \frac{1}{2}\delta_j^i T\right).$$

Problema 10.8.3 Consider the metric

$$ds^2 = \alpha(x+y+z)dt^2 - \frac{1}{2}\left(dx^2 + dy^2 + dz^2\right),$$

where α is a constant. Compute $R_{ij} - \frac{1}{2} R\, g_{ij}$.

Hint. Denote $x^0 := t,\ x^1 := x,\ x^2 := y,\ x^3 := z$. It is easy to obtain $\Gamma^0_{10} = \Gamma^0_{01} = \Gamma^0_{20} = \Gamma^0_{02} = \Gamma^0_{30} = \Gamma^0_{03} = \dfrac{1}{2(x+y+z)}$ and $\Gamma^1_{00} = \Gamma^2_{00} = \Gamma^3_{00} = \alpha$. Then

$$R_{ii} = -\frac{3\alpha}{2(x+y+z)},\ R_{00} = \frac{\alpha}{4(x+y+z)^2},\ i = 1, 2, 3,$$

that is,

$$R = R^i_i = -3\frac{\alpha}{(x+y+z)^2}.$$

It results in

$$R_{ij} - \frac{1}{2} R\, g_{ij} = -\frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & (x+y+z)^{-2} & 0 & 0 \\ 0 & 0 & (x+y+z)^{-2} & 0 \\ 0 & 0 & 0 & (x+y+z)^{-2} \end{pmatrix}.$$

Can you derive some conclusions about T_{ij} tensor?

The equations of geodesics are

$$\begin{cases} \dfrac{d^2t}{d\tau^2} + \dfrac{3}{x+y+z}\dfrac{dt}{d\tau}\left(\dfrac{dx}{d\tau} + \dfrac{dy}{d\tau} + \dfrac{dz}{d\tau}\right) = 0 \\ \dfrac{d^2x}{d\tau^2} + \alpha\left(\dfrac{dt}{d\tau}\right)^2 = 0 \\ \dfrac{d^2y}{d\tau^2} + \alpha\left(\dfrac{dt}{d\tau}\right)^2 = 0 \\ \dfrac{d^2z}{d\tau^2} + \alpha\left(\dfrac{dt}{d\tau}\right)^2 = 0 \end{cases}$$

An important question for the reader is

Exercise 10.8.4 Can these equations be the geodesic equations of the classical constant gravitational field $(-\alpha, -\alpha, -\alpha)$?

Hint. Start by analysing the necessary condition $\frac{dt}{d\tau} = 1$ and the norm (with respect to the metric) of the tangent vector to the geodesic. The answer is no.

10.8.1 More About FLRW Universes

In the previous subsection, we saw that the problem to find a metric for the general Einstein field equations describing the Universe started from the fact that the observed Universe appears homogeneous and isotropic beyond a given scale, therefore we have to consider, at the beginning, a spherical symmetry for the cosmic space-time.

The spatial part first considered there was

$$dr^2 + q^2(r)\left(d\theta^2 + \sin^2\theta d\phi^2\right),$$

where $q(r)$ will be determined, so, the first trial metric was

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + q^2(r)d\theta^2 + q^2(r)\sin^2\theta d\phi^2\right].$$

The function $a(t)$ introduced is necessary to preserve the proposed spherical symmetry of the spatial part of the metric which can, eventually, expand under a homothetic transformation. This way $a^2(t)$ becomes an expansion factor of the Universe. We have to mention that Friedmann and Lemaître used only the expansion factor $a^2(t)$ in the front of the spatial part of the metric without to consider the $q(r)$ factor described above.

We are working in geometric coordinates, that is, $c = 1$.

FLRW universes with or without cosmological constant are obtained only if the contravariant energy-momentum tensor is describing a perfect fluid with components

$$T^{ij} = (\rho_0 + p_0)u^i u^j - p_0 g^{ij},$$

where g^{ij} are the inverse components of the metric tensor matrix which satisfies Einstein's field equations, ρ_0 is the constant density of the fluid, p_0 is the constant pressure of the fluid, and u^i are the components $(u^t, v_x u^t, v_y u^t, v_z u^t)$ of the fluid 4-velocity. To finish the classical description, $a(t)$ is determined from the equation

$$\dot{a}^2 - \frac{8\pi G}{3}\rho_0 \cdot a^2 = k_1$$

where k_1 can have the values $1/R^2$; $-1/R^2$; 0.

What if we are interested to explore a little beat more? Let us propose the spatial part to be

$$dr^2 + q^2(r)\left(d\theta^2 + A^2(\theta)d\phi^2\right),$$

where A is a function we have to determine together with $q(r)$ and $a(t)$ such that the metric

$$ds^2 = dt^2 - a^2(t)\left[dr^2 + q^2(r)d\theta^2 + q^2(r)A^2(\theta)d\phi^2\right]$$

satisfies the Einstein field equations in the same conditions as in the classical case presented above (see [33]). Of course, we expect to obtain at least the solution $A(\theta) = \sin\theta$, together with the classical known $q(r)$ and $a(t)$.

Let us prove

Theorem 10.8.5 *A FLRW universe appears if and only if the metric*

$$d\Omega_A^2 = d\theta^2 + A^2(\theta)d\phi^2$$

describing the 2D generating spatial part has constant Gaussian curvature.

Proof We start by calculating the Ricci symbols of the metric proposed in the introduction above. From

$$g_{00} = 1,\ g_{11} = -a^2(t),\ g_{22} = -a^2(t)q^2(r),\ g_{33} = -a^2(t)q^2(r)A^2(\theta)$$

$$g^{00} = 1,\ g^{11} = -\frac{1}{a^2(t)},\ g^{22} = -\frac{1}{a^2(t)q^2(r)},\ g^{33} = -\frac{1}{a^2(t)q^2(r)A^2(\theta)}$$

we observe

$$\Gamma^i_{jk} = g^{is}\Gamma_{jk,s} = g^{ii}\Gamma_{jk,i};\ \Gamma^i_{jk} = 0,\ i \neq j \neq k.$$

After denoting by $\dot{a}$ the derivative with respect to t and by q' the derivative with respect to$r$ we successively obtain

$$\Gamma^0_{11} = a\cdot\dot{a},\ \Gamma^0_{22} = a\cdot\dot{a}\cdot q^2,\ \Gamma^0_{33} = a\cdot\dot{a}\cdot q^2\cdot A^2(\theta),$$
$$\Gamma^1_{01} = \Gamma^1_{10} = \frac{\dot{a}}{a},\ \Gamma^1_{22} = -q\cdot q',\ \Gamma^1_{33} = -q\cdot q'\cdot A^2(\theta),$$
$$\Gamma^2_{02} = \Gamma^2_{20} = \frac{\dot{a}}{a},\ \Gamma^2_{12} = \Gamma^2_{21} = \frac{q'}{q},\ \Gamma^2_{33} = -A'(\theta)A(\theta),$$
$$\Gamma^3_{03} = \Gamma^3_{30} = \frac{\dot{a}}{a},\ \Gamma^3_{13} = \Gamma^3_{31} = \frac{q'}{q},\ \Gamma^3_{32} = \Gamma^3_{23} = \frac{A'(\theta)}{A(\theta)}, \text{otherwise } \Gamma^k_{ij} = 0.$$

We have

$$R_{tt} = -3\frac{\ddot{a}}{a}.$$

We obtain

$$R_{rr} = 2\dot{a}^2 + \ddot{a}\cdot a - 2\frac{q''}{q},$$

$$R_{\theta\theta} = 2q^2 \cdot \dot{a}^2 + q^2 \cdot a \cdot \ddot{a} - q \cdot q'' - (q')^2 + \frac{A''(\theta)}{A(\theta)},$$

$$R_{\phi\phi} = A^2(\theta) \cdot R_{\theta\theta}.$$

Using $R^i_j = g^{is} R_{sj}$, we rise an index:

$$R^t_t = -3\frac{\ddot{a}}{a},$$

$$R^r_r = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} + 2\frac{q''}{a^2 \cdot q},$$

$$R^\theta_\theta = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} + \frac{1}{a^2 \cdot q^2}\left(\frac{A''(\theta)}{A(\theta)} + q \cdot q'' + (q')^2\right),$$

$$R^\phi_\phi = R^\theta_\theta.$$

The key of finding the metric is related to the lowering of indexes of the contravariant energy–momentum tensor mentioned above. It follows that $T^t_t = \rho_0$ and $T^r_r = T^\theta_\theta = T^\phi_\phi = -p_0$, that is $T = T^i_i = \rho_0 - 3p_0$.

If we arrange the Einstein's field equation in the form

$$R^i_j = 8\pi G \cdot \left(T^i_j - \frac{1}{2}\delta^i_j T\right)$$

we obtain

$$R^t_t = -3\frac{\ddot{a}}{a} = 8\pi G \cdot \left(T^t_t - \frac{1}{2}T\right) = 4\pi G \cdot (\rho_0 + 3p_0)$$

and $R^r_r = R^\theta_\theta = R^\phi_\phi = -4\pi G\,(\rho_0 - p_0)$.

Now, the condition $R^r_r = R^\theta_\theta$ highlights the equality

$$2\frac{q''}{a^2 \cdot q} = \frac{1}{a^2 \cdot q^2}\left(\frac{A''(\theta)}{A(\theta)} + q \cdot q'' + (q')^2\right)$$

which makes sense if and only if the right member doesn't depend on θ. And this means that the ratio $\dfrac{A''(\theta)}{A(\theta)}$ is a constant denoted by K.

Considering the following possible situations $K < 0;\ K = 0;\ K > 0$, we determine the function $q(r)$ from the equation

$$q \cdot q'' - (q')^2 = K$$

and $A(\theta)$ from the differential equation

$$A''(\theta) - KA(\theta) = 0.$$

$a(t)$ is determined from the equation

$$\frac{\dot{a}^2}{a^2} - \frac{q''}{a^2 \cdot q} = \frac{8\pi G}{3}\rho_0.$$

This one is obtained after we replace

$$R_t^t = -3\frac{\ddot{a}}{a} = 4\pi G \cdot (\rho_0 + 3p_0)$$

in

$$R_r^r = -2\frac{\dot{a}^2}{a^2} - \frac{\ddot{a}}{a} + 2\frac{q''}{a^2 \cdot q} = -4\pi G\,(\rho_0 - p_0)\,.$$

I. The case $K = -\dfrac{1}{R^2} < 0$.

This case allows the particular solutions $A_1(\theta) = \sin\dfrac{\theta}{R}$ and $A_2(\theta) = \cos\dfrac{\theta}{R}$. The general solution is a linear combination of these two solutions.

Then, the equation $q \cdot q'' - (q')^2 = -\dfrac{1}{R^2}$ and the corresponding Friedmann equation

$$\frac{\dot{a}^2}{a^2} - \frac{q''}{a^2 \cdot q} = \frac{8\pi G}{3}\rho_0$$

describe the classical FLRW case.

Let us underline that this case is highlighted from a 2D-metric of constant Gaussian positive curvature,

$$K_G = -K = \frac{1}{R^2}.$$

The new situations appear from the analysis of the other cases.

II. The case $K = 0$. We have $A(\theta) = h\theta$, or $A(\theta) = h$.
It follows

$$(q')^2 - q \cdot q'' = 0$$

with the solution $q(r) = e^{kr}$, where k is a constant.
So, the ratio $\dfrac{q''}{q} = k^2$ and $a(t)$ is determined by the equation

$$\dot{a}^2 - \frac{8\pi G}{3}\rho_0 a^2 = k^2.$$

We know to obtain the general solution of an equation as

$$\dot{a}^2 - A^2 a^2 = B^2,$$

that is,

$$a(t) = \frac{C_1}{2A} e^{At} - \frac{B^2}{2C_1 A} e^{-At}.$$

To maintain the way in which the FLRW solution is done, we choose $C_1 = B$, therefore we consider the solution in the form

$$a(t) = \frac{k}{\sqrt{\frac{8\pi G}{3}\rho_0}} \sinh\left(\sqrt{\frac{8\pi G}{3}\rho_0} \cdot t\right).$$

So, in the case $K = 0$ the two metrics are described by the previous $a(t)$, $q(r) = e^{kr}$ and $A_1(\theta) = h\theta$ and $A_2(\theta) = h$, that is,

$$ds^2 = dt^2 - \frac{3k^2}{8\pi G\rho_0} \sinh^2\left(\sqrt{\frac{8\pi G}{3}\rho_0} \cdot t\right)\left[dr^2 + e^{2kr}(d\theta^2 + A_i^2(\theta)d\phi^2)\right].$$

We can observe that the initial generating 2D metric $d\Omega_A^2 = d\theta^2 + A_i^2(\theta)d\phi^2$ used to create the FLRW metric above has in both cases null Gaussian curvature, $K_G = -K = 0$.

III. The case $\dfrac{A''(\theta)}{A(\theta)} = K = \dfrac{1}{R^2} > 0$.

In this case $A_1(\theta) = e^{\theta/R}$; $A_1(\theta) = e^{-\theta/R}$; $A_3(\theta) = \sinh\dfrac{\theta}{R}$; $A_4(\theta) = \cosh\dfrac{\theta}{R}$ and linear possible combinations.
We determine $q(r)$ from the differential equation

$$q \cdot q'' - (q')^2 = K = \frac{1}{R^2}.$$

The solution is

$$q(r) = \cosh\frac{r}{R}.$$

$a(t)$ is determined by the equation

$$\dot{a}^2 - \frac{8\pi G}{3}\rho_0 a^2 = \frac{1}{R^2}.$$

The solution is obtained as above,

$$a(t) = \frac{1}{R\sqrt{\frac{8\pi G}{3}\rho_0}} \sinh \sqrt{\frac{8\pi G}{3}\rho_0} \cdot t.$$

This $a(t)$ together with $q(r) = \cosh \frac{r}{R}$ is part of all the metrics corresponding to this case, that is, we get

$$ds^2 = dt^2 - \frac{3k^2}{8\pi G\rho_0} \sinh^2 \left(\sqrt{\frac{8\pi G}{3}\rho_0} \cdot t\right) \left[dr^2 + \cosh^2 \frac{r}{R} \left(d\theta^2 + A_i^2(\theta) d\phi^2\right)\right].$$

Now, if you look at the 2D metrics $d\Omega_A^2 = d\theta^2 + A_i^2(\theta)d\phi^2$ used to create the FLRW metric in this case, all are constant Gaussian negative curvature, $K_G = -K = -\frac{1}{R^2}$. □

10.8.2 A Remarkable Universe without Matter from FLWR Conditions

It is possible to use the FLRW conditions for a universe suggested by the Poincaré half-plane, extended to four dimensions in a Minkowski space-time. The properties of a Minkowski–Poincaré half-plane in two dimensions are described by the metric

$$ds^2 = \frac{1}{x^2}(dt^2 - dx^2).$$

We obtain

$$\Gamma^0_{01} = \Gamma^0_{10} = \Gamma^1_{00} = \Gamma^1_{11} = -\frac{1}{x};\ R^0_{101} = -\frac{2}{x^2},\ R_{0101} = -\frac{2}{x^4},$$

that is the Gauss-Minkowski curvature is 2 and the geodesic equations

$$\ddot{t} - \frac{2}{x}\dot{t}\dot{x} = 0,\ \ddot{x} - \frac{1}{x}(\dot{t})^2 - \frac{1}{x}(\dot{x})^2 = 0$$

are satisfied by $x(s) = t(s) = e^{s\sqrt{2}}$ and $x(s) = -t(s) = e^{s\sqrt{2}}$. Of course, the light-cones induced by the condition $ds^2 = 0$ are formed by geodesics of this metric.

According to [33], we can consider a system of coordinates where the metric

$$ds^2 = \frac{1}{z^2}\left[dt^2 - a^2(t)\left(dx^2 + dy^2 + dz^2\right)\right]$$

can be written by the FLRW conditions. Let us denote by $\dot{a}$ and $\ddot{a}$, the first and the second derivative of the function $a(t)$ with respect to t. Computing the non-zero second-type Christoffel symbols, we have

$$\Gamma^0_{03} = \Gamma^0_{30} = -\frac{a^2}{z};\ \Gamma^3_{00} = \Gamma^1_{13} = \Gamma^1_{31} = \Gamma^2_{23} = \Gamma^2_{32} = \Gamma^3_{33} = -\frac{1}{z};\ \Gamma^3_{11} = \Gamma^3_{22} = \frac{1}{z};$$

$$\Gamma^0_{11} = \Gamma^0_{22} = \Gamma^0_{33} = a\dot{a};\ \ \Gamma^1_{10} = \Gamma^1_{01} = \Gamma^2_{02} = \Gamma^2_{20} = \Gamma^3_{30} = \Gamma^3_{03} = \frac{\dot{a}}{a}.$$

The only non-zero Ricci tensor components are

$$R_{00} = \frac{4 - a^2}{z^2} - 3\frac{\ddot{a}}{a};\ \ R_{11} = R_{22} = -\frac{2 + a^2}{z^2} + a\ddot{a} + 2\dot{a}^2;$$
$$R_{33} = -\frac{2 + a^2}{z^2} + a\ddot{a} + 3\dot{a}^2$$

therefore

$$R^0_0 = 4 - a^2 - 3z^2\,\frac{\ddot{a}}{a};\ \ R^1_1 = R^2_2 = \frac{2 + a^2}{a^2} - z^2\,\frac{\ddot{a}}{a} - 2z^2\,\frac{\dot{a}^2}{a^2};$$
$$R^3_3 = \frac{2 + a^2}{a^2} - z^2\,\frac{\ddot{a}}{a} - 3z^2\,\frac{\dot{a}^2}{a^2}.$$

The FLWR conditions $R^1_1 = R^2_2 = R^3_3$ lead to the simple Friedmann equation $\dot{a} = 0$. Rescaling the variables x, y, z, we can consider the solution $a(t) = 1$.

The FLWR conditions give us the particular metric

$$ds^2 = \frac{1}{z^2}\left[dt^2 - \left(dx^2 + dy^2 + dz^2\right)\right].$$

Looking at the above formulas, we get the metric of an Einstein manifold, which satisfies the formulas

$$R_{ij} = 3\, g_{ij}.$$

Since $R = 12$, the Einstein equation can be written in the form

$$R_{ij} - \frac{1}{2}R\, g_{ij} + \Lambda\, g_{ij} = 0$$

for the cosmological constant $\Lambda = 3$. So, we have obtained a universe without matter because $T_{ij} = 0$. If we look at the condition $ds^2 = 0$ which gives the structure of the lightcone, we have a structure similar to the one appearing in the usual Minkowski space-time.

10.8.3 The Cosmological Expansion

This subsection is dedicated to the expansion of the Universe. We saw that Einstein Static Universe imposed the existence of a new term in the fields equations, because in a Universe in which the matter is constrained to interact only by gravity, all the matter sources will be concentrated in the same region, in contrast with the desired Einstein static structure.

The new term was proposed to establish a repulsive effect to counterweight the attractive effect of gravity. However, Einstein discarded the cosmological term when Hubble discovered evidences for cosmological expansion. In any case, Theorem 10.8.1 suggests that we can obtain an expanding universe even if the cosmological constant is not considered. Is Hubble's law related to the cosmological metrics obtained above? The answer is yes!

Let us describe the Hubble law for recession of galaxies.

First, we have to mention that Hubble used Doppler's effect to establish his result related to the redshift of distant galaxies. The light in the Universe is produced by stars. The hydrogen of stars, in thermonuclear fusion, produces primarily helium and energy that radiates in space, some of it in form of light. Hubble considered the four lines of the hydrogen light spectrum. For distant galaxies, the same four lines of hydrogen spectrum are seen shifted to the right in comparison to normal pattern of light decomposition detected in laboratory. Hubble realized that this is a Doppler effect and the observed redshift means that the distant galaxies are moving away from us. He stated that the redshifts in spectra of distant galaxies are proportional to the distance of galaxies from us. The mathematical form is

$$V = H \cdot D,$$

where D is the proper distance from us to the galaxy, $V := \dot{D}$ is the proper speed of the galaxy, and H is a constant called the Hubble constant. The farther away the galaxy is, the faster it moves away from us. The entire space, the entire texture of the Universe is moving away from us carrying the galaxies in it.

Alternatively, let us suppose we have a ruler of coordinates marked $0,\ 1,\ 2,\ 3,\ 4, \ldots$. The distance between two consecutive coordinates is denoted by a. The distance measured with this ruler is denoted by D. $D = a \cdot \Delta x$, where Δx is the difference between the coordinates of the chosen points we wish to measure.

Now, suppose we have a rubber band marked in the same way as our ruler; we pin the origin and start to stretch. The coordinate points remain drawn on the rubber band but the distance between them increases. Therefore a depends on time, it is $a(t)$. The distance D becomes $D(t) = \Delta x \cdot a(t)$. We have

$$V := \dot{D} = \Delta x \cdot \dot{a}.$$

This relation can be written as

$$V \cdot a = a \cdot \Delta x \cdot \dot{a} = D \cdot \dot{a},$$

that is,

$$V = \frac{\dot{a}(t)}{a(t)} D.$$

We define $H := \dfrac{\dot{a}(t)}{a(t)}$ and obtain Hubble's law

$$V = H \cdot D.$$

What is new in this approach is the fact that it is suggested the stretch of the texture of the universe. Such a stretch was seen in Sect. 9.9 when we discussed about a possible metric for the Cosmos. The metric proposed was

$$ds^2 = c^2 dt^2 - a^2(t) \left[dr^2 + q^2(r) d\theta^2 + q^2(r) \sin^2 \theta d\phi^2 \right],$$

where $q(t)$ and $a(t)$ were determined from the general Einstein field equations under some conditions imposed by the energy–momentum tensor T_{ij}.
The differential equation for $a(t)$ will be found now under some physical conditions and important consequences will come out.

Consider two galaxies in the Universe such that the distance between them is D. Let us consider one of them and the sphere of radius D centred at the chosen galaxy. Denote by M the total mass of galaxies inside the sphere and by m the mass of the second galaxy. This galaxy moves away from the galaxy at the centre of the sphere with speed $V = H \cdot D$. The gravitational force which acts on the galaxy of mass m is

$$F = \frac{GMm}{D^2}.$$

The potential energy for that galaxy is

$$P_E = -\frac{GMm}{D}$$

and the kinetic energy is

$$K_E = \frac{mV^2}{2}.$$

The total energy acting on the second galaxy is a constant,

$$P_E + K_E = const = k_1.$$

Thanks to the Equivalence Principle, we can divide by m and then it results

$$-\frac{2GM}{D} + V^2 = k.$$

But $D(t) = \Delta x \cdot a(t)$ and $V(t) = \Delta x \cdot \dot{a}(t)$, that is

$$(\Delta x \cdot \dot{a}(t))^2 - \frac{2GM}{\Delta x \cdot a(t)} = k.$$

Some remarks are in order now. $M = Vol \times density$. If the volume increases when the Universe is expanding but the number of galaxies does not change, the density decreases. Since

$$M = \frac{4}{3}\pi \cdot D^3 \cdot \rho(t) = \frac{4}{3}\pi \cdot (\Delta x \cdot a(t))^3 \cdot \rho(t),$$

we have

$$(\Delta x)^2 \cdot (\dot{a}(t))^2 - \frac{8\pi G}{3} \cdot (\Delta x)^2 \cdot (a(t))^2 \cdot \rho(t) = k.$$

We arrange in a dimensional way the previous formula replacing k by $-K \cdot \Delta x$. Finally, we obtain the differential equation

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 - \frac{8\pi G}{3} \cdot \rho(t) = -\frac{K}{a^2(t)}.$$

This is a sort of Friedmann–Lemaître–Robertson–Walker equation as obtained in Sect. 10.8.1. The term $\frac{8\pi G}{3} \cdot \rho(t)$ is always positive. If K is negative, the equation written in the form

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \rho(t) - \frac{K}{a^2(t)}$$

can be solved. Such an equation describes a spatially open Universe. If, for some t, K is such that the right member becomes negative at a point, this Universe will increase until that point; then it can remain unchanged or even it can contract. This kind of Universe is called a spatially closed universe. If $K = 0$, the universe will be called a spatially flat Universe. This Universe expands too. In such a Universe there is a perfect balance between the kinetic and the potential energy.

The observational evidences show that our Universe is a flat one. So, it remains to solve the equation

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \rho(t).$$

In a flat, matter-dominated Universe (mdu), in a cube of side $a(t)$, having inside galaxies whose total mass is M, the density is expressed by the formula $\rho_{mdu}(t) = \frac{M}{a^3(t)}$. The corresponding $(FLRW)$ equation is

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \frac{M}{a^3(t)}.$$

The solution, expressing the expansion of a matter-dominated Universe, is then

$$a(t) = B \cdot t^{2/3},$$

where B is a positive constant.

After the Big Bang and inflation [159], there was a period when the universe was radiation dominated (rdu). To describe its expansion, we consider the same cube of side $a(t)$, now full of photons. Since the energy is expressed by the formula $E = h\nu = h\frac{c}{\lambda}$ and, when $a(t)$ is increasing, the wavelength λ is increasing too, we can suppose that $E = \frac{C}{a(t)}$ is describing the energy formula. Here C is a constant. The density of such a Universe is given by

$$\rho_{rdu} = \frac{E}{a^3(t)} = \frac{C}{a^4(t)}.$$

The corresponding $(FLRW)$ equation is

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \frac{C}{a^4(t)}.$$

The solution, which expresses the expansion of a radiation-dominated universe, is

$$a(t) = A \cdot t^{1/2},$$

where A is a positive constant.

Now, let us observe something crucial. We know two important physic formulas, Planck's one $E = h\nu = h\frac{c}{\lambda}$ and Boltzmann's one $E = k_B T$. Using the same reasoning as before we deduce the direct proportionality between the temperature T and $\frac{1}{a(t)}$. As said, after Big Bang and inflation, our Universe was radiation dominated, and the temperature at which the atoms can form is less than 3×10^3 K degrees. Now, the today cosmic background radiation has approximatively 3 K degrees. Therefore, if we suppose that in the period in our Universe started to be matter dominated, the temperature decreases of the ratio $\sim \frac{3 \times 10^3}{3}$, it is easy to see the ratio $\frac{a_{today}}{a_{ionized}}$, where $a_{ionized}$ is the epoch in which ionized atoms appear. It is

$$\frac{T_{ionized}}{T_{today}} = \frac{a_{today}}{a_{ionized}} \simeq 10^3 \simeq \frac{t_{today}^{2/3}}{t_{ionized}^{2/3}}.$$

Since t_{today} is about 10^{10} years, i.e. the age of the observed Universe, $t_{ionized}$ becomes about 3×10^5 years after Big Bang. It means that the Universe was radiation dominated for almost 3×10^5 years. More precisely, it takes about 3×10^5 years for the Universe, expanding and cooling after Big Bang, to allow electrons and protons to couple and form neutral atoms. At this point, even the photons are free to move and get to us, providing us with the first "photograph" of the Universe that can be obtained, that is, the Cosmic Microwave Background Radiation.[3] Clearly, this is only a rough calculation to derive the order of magnitudes. For a detailed discussion on primordial Universe phenomenology, see [158, 159].

We are now ready to understand some basic facts about dark energy and the pressure exerted to expand our Universe. Specifically, dark energy is the hypothetical fluid fueling the observed accelerated expansion revealed at the end of twentieth century [13]. Let us begin by analysing the pressure exerted on the faces of a cube imagined in our Universe. Obviously, there is no pressure in a matter-dominated Universe because the galaxies inside the cube do not exert any pressure on the faces of the cube.

In a radiation-dominated universe, it is possible to study the pressure in the following way. Let us consider a photon which can move between "the extremities" of a segment line of length L. The small amount of time necessary to move between the extremities can be denoted as dt and we have the formula $dt = \frac{2L}{c}$. The force which produces the pressure on the extremities is

$$F = \frac{dp}{dt} = \frac{2p}{\frac{2L}{c}} = \frac{pc}{L} = \frac{E}{L}.$$

If we denote by L the length of the side of a cube in a radiation-dominated Universe and by dA the infinitesimal area of a square drawn on a face (the sides parallel to the sides of the face), now corresponding to the perpendicular direction on the given face, we have

$$P = \frac{F}{dA} = \frac{E}{L dA}.$$

[3] Actually the recombination of hydrogen happened at a redshift $z = 1089$ corresponding to a period of 3.79×10^5 years after Big Bang. Here the redshift correspond to the above $a_{today}/a_{ionized}$. See [198].

Therefore the pressure P exerted can be though as the ratio between the energy and a volume corresponding to dA and the above-mentioned perpendicular direction, that is, an energy density ρ. In fact we have

$$P = w\rho,$$

where $w = 0$ in the case of matter-dominated Universe and $w = \frac{1}{3}$ in an radiation-dominated Universe; 3 appears because we have three perpendicular direction on faces.

These numbers represent two possibilities for the equation of state of a standard perfect fluid where $0 \leq w \leq 1$ is the so-called Zel'dovich interval [199]. Being $w = \left(\frac{c_s}{c}\right)^2$, with c_s the sound speed, the fluids in the Zel'dovich interval agree with the causality condition implying that the speed of light has to be $c > c_s$. In other words, standard matter cannot be constituted by tachyons, that is, particles faster than light.

Suppose now that the pressure expands the cosmic cube of a dV volume. Taking into account the work done by the force F, $F \cdot d = P \cdot A \cdot d = P \cdot dV$, and the variation of the energy E, we have

$$dE = -P \cdot dV.$$

At the same time,

$$E = \rho \cdot V.$$

It results in

$$dE = d\rho \cdot V + \rho \cdot dV = -P \cdot dV,$$

i.e.

$$V \cdot d\rho = -(P + w)dV = -\rho(w + 1)dV.$$

We have obtained the differential equation

$$\frac{d\rho}{\rho} = -(w + 1)\frac{dV}{V}$$

with the solution

$$\rho = NV^{-(w+1)} = Na^{-3(w+1)},$$

where N is a constant.

For $w = 0$, we obtain the formula corresponding to a matter-dominated Universe, while for $w = \frac{1}{3}$ we obtain the formula of a radiation-dominated Universe.

Let us insert this last formula in the $(FLRW)$ equation, we get

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \frac{N}{(a(t))^{3(w+1)}}.$$

Clearly, in the above discussion, the functions P and ρ are functions of time and the above definition of the energy–momentum tensor can be generalized to describe a perfect fluid of the form

$$T^{ij} = (\rho + P)u^i u^j - pg^{ij}.$$

What happens if $w = -1$? The $(FLRW)$ equation becomes

$$\left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3} \cdot \rho_0,$$

where ρ_0 is a constant. Now we are in the case of a Universe expanding according to the law

$$\frac{\dot{a}(t)}{a(t)} = H_0 = \sqrt{\frac{8\pi G \rho_0}{3}}.$$

The solution of the expansion is exponential,[4] that is,

$$a(t) = a_0 e^{H_0 t},$$

where a_0 is a constant related to the initial value of the scale factor $a(t)$. The value $w = -1$ is clearly out of the above Zel'dovich interval, i.e. it is not a standard perfect fluid, and corresponds to "something" which determines the exponential accelerated expansion. Such an expansion is in agreement with the existence of a possible cosmological constant Λ (that is, ρ_0). This "something" manifests itself as a pressure implying an energy density. As said, this energy is neither produced by the ordinary matter nor by the radiation.[5] This is a simple example of *dark energy* that gives rise to accelerated expansion. The mechanism can work both in early Universe, giving rise to inflation, and in late Universe, giving the observed accelerated expansion of the Hubble flow. Clearly the scales of energy are completely different and between inflation and recent accelerated epoch there are radiation- and matter-dominated eras. It is worth noticing that, according to data, the dark energy constitutes $\sim$70% of the total amount of matter–energy content of the Universe [13]. Understanding nature and dynamics of dark energy is one of the main challenges of modern cosmology.

---

[4] H_0 is assumed constant because ρ_0 is constant.

[5] It is important to note that any form of standard matter, in the interval $0 \leq w \leq 1$, gives rise to decelerated expansion.

10.9 Measuring the Cosmos

Let us consider some physical constants and astronomic units necessary to evaluate the age, the radius, the volume, the mass, and the density of our observable Universe.

We know already both the gravitational constant $G = 6.67 \cdot 10^{-11} \frac{(\text{m})^3}{(\text{kg}) \cdot (\text{s})^2}$ and the speed of light in vacuum $c = 3 \cdot 10^5 \frac{(\text{m})}{(\text{s})}$. Let us write the value of the Hubble constant

$$H = 70 \frac{\text{Km}}{\text{s}} \cdot \frac{1}{1 \text{ parsec} \cdot 10^6}$$

and 1 parsec $= 3,086 \cdot 10^{16}$ (m). If we compute the value of Hubble constant we obtain

$$H = 70 \cdot 10^3 \frac{(\text{m})}{(\text{s})} \cdot \frac{1}{3,086 \cdot 10^{16} \cdot 10^6 (\text{m})} = 2,3 \cdot 10^{-18} \frac{1}{(\text{s})}.$$

The number of seconds of a year is

$$N_s = 365 \cdot 24 \cdot 60 \cdot 60 \text{ (s)} = 3,15 \cdot 10^7 (\text{s}).$$

Since the age of the Universe is $A_U = \frac{1}{H}$, it results that the age of the Universe measured in years is

$$A_U = \frac{1}{2,3} \cdot 10^{18} \frac{1}{3,15 \cdot 10^7} \; years = 1,38 \cdot 10^{10} \; years.$$

In words means 13, 8 billions years.

Now, let us give a "sketch" on how the radius of the observable Universe can be computed. When we are talking about the radius of the observable Universe, we are taking into consideration the facts explained in the previous sections. In principle, a primordial photon can travel to reach our eye from 13, 8 billions years. According to the Hubble law, the texture of the Universe is in continuous expansion. Therefore, to establish the distance crossed by the photon it is not enough to multiply the time, i.e. the age of this Universe, by the speed of light in vacuum. Therefore a formula as $R = C \cdot A_U = c \cdot dt$ will not describe the exact value of the radius R_O of our observable Universe. We need to take into account the "expansion factor". The expansion factor is

$$a(t) := \frac{dx(t)}{dx(t_0)}.$$

Here $t_0 := 13,8By$ and $dx(t_0)$ is the variation of the real distance crossed by photon in t_0 years. We obtain

$$dx(t_0) = \frac{dx(t)}{a(t)} = \frac{cdt}{a(t)},$$

therefore, for a chosen $a(t)$, we have

$$R_O = \int_0^{t_0} dx(t_0) = \int_0^{t_0} \frac{c}{a(t)} dt \approx 46\text{Bly}.$$

Therefore, this leads to the unrealistic radius of the observable Universe of 46 billion light years.

Now we can imagine the $3D$ spatial part of the observable Universe as a sphere with the radius R_O. The volume of this observable Universe is

$$Vol_U = 4/3\pi\, R_O^3 \approx 5 \cdot 10^{80} (\text{m})^3.$$

The real value has 3,566 instead of 5, but this value was obtained both because our simplified value of speed of light in vacuum and our approximations.

The mass of the observable Universe can be computed thinking at the escape velocity formula seen in the black hole section. In this case, we have to think about a galaxy escaping from our observable Universe. In the formula seen there,

$$\frac{1}{2}\cancel{m}v^2 = \frac{GM\cancel{m}}{R},$$

the mass m of the escaping galaxy is cancelled. M becomes the mass of all the observable Universe, G is the gravitational constant, R must be the radius of the observable Universe, R_O, and the escape velocity must be the speed of light, c. It results in

$$M = \frac{c^2 R_O}{2G}.$$

Replacing the values, we obtain the mass of the observable Universe, $M \approx 10^{53} (\text{Kg})$.

The density of the observable Universe is computed using the formula

$$\rho_U = \frac{M}{\text{Vol}_U}.$$

The value of this density is $\approx 2 \cdot 10^{-28} \frac{(\text{Kg})}{(\text{m})^3}$ if we compute in our approximation. This means about 10 protons per each cubic metre of space.

The above calculations give a rough image of some basic observables of our Universe. Clearly, they have to be confronted with data and some results can be completely different. This fact points out that more details on cosmological dynamics are needed.

10.10 The Fermi Coordinates

After the above summary on cosmological expansion, let us define a system of coordinates very useful to describe the geodesic motion. From a mathematical point of view, Fermi's coordinates are *local coordinates adapted to a geodesic*, that is, at a given point P on a geodesic $c(\tau)$, there exists a local system of coordinates around P such that:

- the geodesic locally becomes $(x^0, 0, 0, .., 0)$;
- the metric tensor along geodesic is the Minkowski metric (or the Euclidean metric; it depends on the context);
- all the Christoffel symbols vanish along geodesic.

A nice treatment of this subject[6] and its applications can be seen in [46, 137].

In our context, we intend to describe the topic in a simplified way.

Consider a coordinate frame at rest denoted by $R : (y^0, y^1, y^2, y^3)$ together with a given metric $ds^2 = \bar{g}_{ij} dy^i dy^j$.

We intend to describe the free fall of an observer F in the gravitational field induced by $\bar{g}_{ij}$.

1. In the coordinate frame at rest, R, the freely falling observer F is moving on a geodesic of the metric $ds^2 = \bar{g}_{ij} dy^i dy^j$, say $c(\tau)$.
   The geodesic equations of $c(\tau)$ are $\dfrac{d^2 y^i}{d\tau^2} + \bar{\Gamma}^i_{jk} \dfrac{dy^j}{d\tau} \dfrac{dy^k}{d\tau} = 0$.
   This geodesic is the world line of F in R.
2. From F point of view, there is no field. Consider F in a spacecraft, somewhere in an almost empty region of the space. That is, to describe the free falling, means to create a coordinate frame $F : (x^0, x^1, x^2, x^3)$ such that, along the world line of F in R in these coordinates, we have $\Gamma^i_{jk} = 0$. For F, the geodesic equations become $\dfrac{d^2 x^i}{d\tau^2} = 0$, that is, F should move on a straight line.
   We make the assumption: Let x^0-axis be the world line of F in R.
3. Now, more clearly, we have to construct a map $M : F \to R$ which transfers x^0-axis into the geodesic $c(\tau)$, in such a way that the x^0-axis becomes a geodesic in F endowed with the metric $g_{ij} = dM_x^t \cdot \bar{g}_{ij} \cdot dM_x$.

Therefore, M maps the x^0-axis into the image of the geodesic $c(\tau)$.

If τ is the geodesic parameter for the curve $c(\tau)$, we can consider the same parameter for the x^0-axis, i.e. $x^0 = \tau$ is the current coordinate of this axis.
At each point τ, we have

$$c(\tau) = (y^0(\tau), y^1(\tau), y^2(\tau), y^3(\tau)) \in R,$$

[6] It is interesting saying that the paper reporting these results was the first one written by Enrico Fermi when he was student at Scuola Normale Superiore di Pisa [101].

therefore the map $M : F \to R$ gives rise to

$$(x^0, 0, 0, 0) \to (y^0(x^0), y^1(x^0), y^2(x^0), y^3(x^0)),$$

where $(y^0(x^0), y^1(x^0), y^2(x^0), y^3(x^0))$ are the coordinates of the points of the geodesic in R. Therefore our transformation $M : F \to R$ can be thought

$$(\tau, 0, 0, 0) \to c(\tau),$$

with some considerations on the functions y^k we need to describe.

Let us keep in our mind that we are interested in transferring the property "*c is a geodesic in R*" to the x^0-axis in F. So, we have

Lemma 10.10.1 *Along the geodesic c in R, it can be highlighted an orthonormal frame with respect to the metric $\bar{g}_{ij}$ whose time-like vector is the tangent vector* $\dfrac{dc}{d\tau}$.

Proof We know that at each point $c(\tau)$, i.e. along the geodesic c, the tangent vector $\dfrac{dc}{d\tau} = \left(\dfrac{dy^0}{d\tau}, \dfrac{dy^1}{d\tau}, \dfrac{dy^2}{d\tau}, \dfrac{dy^3}{d\tau}\right)$ is a time-like unit vector. We denote it by $e_0(\tau)$. We know that $e_0(\tau)$ is parallel transported along the geodesic c in R preserving all its properties.

Consider the point corresponding to $\tau = 0$, that is, the point $c(0)$ on the geodesic. We choose the spatial vectors $e_1(0),\ e_2(0), e_3(0)$ such that the frame $\{e_0(0), e_1(0), e_2(0), e_3(0)\}$ is orthonormal with respect to the metric $\bar{g}_{ij}$ and we parallel transport it along the geodesic c.

At each point $c(\tau)$, the vectors $\{e_0(\tau), e_1(\tau), e_2(\tau), e_3(\tau)\}$ form an orthonormal frame with respect to the metric $\bar{g}_{ij}$. □

Lemma 10.10.2 *Every point (x^0, x^1, x^2, x^3) in F can be uniquely described in the form $(\tau, l\ \vec{v})$, where $l\ \vec{v}$ is an appropriate Euclidean description of its spatial part.*

Proof Consider a point (x^0, x^1, x^2, x^3) in F which does not belong to x^0-axis. This point is (τ, x^1, x^2, x^3) and at least one spatial component is non-zero.

Denote

$$l := \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}$$

and construct the vector

$$\vec{v} := \left(\frac{x^1}{l}, \frac{x^2}{l}, \frac{x^3}{l}\right) := (v^1, v^2, v^3).$$

It appears the possibility to describe the point (x^0, x^1, x^2, x^3) by (τ, lv^1, lv^2, lv^3) or, simply, by $(\tau, l\ \vec{v})$. □

10.10.1 Determining the Fermi Coordinates

Consider in $T_{c(\tau)}R$ the vector

$$\vec{V}(\tau) := v^1 e_1(\tau) + v^2 e_2(\tau) + v^3 e_3(\tau).$$

Observe

$$ds^2\left(\vec{V}(\tau), \vec{V}(\tau)\right) = \bar{g}_{\alpha\beta} v^\alpha v^\beta = -(v^1)^2 - (v^2)^2 - (v^3)^2.$$

We may impose $ds^2\left(\vec{V}(\tau), \vec{V}(\tau)\right) = -1$, that is $\vec{V}$ is a spatial unit vector. Let us observe that this spatial part has the same property $(v^1)^2 + (v^2)^2 + (v^3)^2 = 1$ as the vector $\vec{v}$ from the above lemma.

According to the equations of geodesics, we may conclude that it exists a unique geodesic of R, denoted by $y_{\vec{V}}(s)$ passing through the point $c(\tau)$ at $s = 0$, such that its tangent vector at origin is $\vec{V}$, that is, $\dfrac{dy_{\vec{V}}}{ds}(0) = \vec{V}$.

According to the above notations, the local map $M : F \to R$, describing the Fermi coordinates, is

$$M(x^0, x^1, x^2, x^3) = M(\tau, s\ \vec{v}) := y_{\vec{V}}(s).$$

Observe that the tangent vector along the spatial geodesic $y_{\vec{V}}(s)$ is a unit vector.

The immediate consequence is: For a given point $(\tau,\ s_0\ \vec{v})$, the spatial distance to $(\tau, 0, 0, 0)$ is s_0. The length of the spatial geodesic between its initial point $c(\tau) = y_{\vec{V}}(0)$ and $y_{\vec{V}}(s_0)$ is also s_0, because the length formula is

$$\int_0^{s_0} \left\| \frac{dy_{\vec{V}}}{ds}(s) \right\| ds = \int_0^{s_0} ds = s_0.$$

The coordinates induced in F by M are called *Fermi's coordinates*.

It remains to prove that in F, in Fermi's coordinates, with respect to the induced metric g_{ij}, the x^0 axis $(\tau, 0, 0, 0)$ is a geodesic and $\Gamma^i_{jk}(\tau, 0, 0, 0) = 0$.

Let us discuss the consequences on the map M.

Theorem 10.10.3 *The map M in invertible in the neighbourhood of each point $P(\tau, 0, 0, 0)$ of the x^0-axis.*

Proof According to the inverse function theorem, it is enough to prove that the matrix dM_P transforms a basis of the tangent space $T_P F$ into linear independent vectors of $T_{M(P)}R$.

We know

$$M(P) = M(\tau, 0, 0, 0) = c(\tau).$$

Consider the standard basis of $T_P F$, denoted by ε_i, $i \in \{0, 1, 2, 3\}$, ε_i having 1 on the ith row, 0 elsewhere. Therefore, by the way we defined M, $dM_P(\varepsilon_i) = e_i(\tau)$, $i \in \{0, 1, 2, 3\}$, i.e. M is locally invertible. □

The meaning of the word "neighborhood" in this context is "tube around the geodesic".

Now, it makes sense the metric $g_{ij} = dM_x^t \cdot \bar{g}_{ij} \cdot dM_x$ as a metric of F.

Theorem 10.10.4 *The x^0-axis is a geodesic of F with respect to the metric g_{ij}.*

Proof The previous theorem allows us to observe that the tangent vector ε_0 is parallel transported along the x^0-axis, therefore x^0 axis is a geodesic of F. □

Exercise 10.10.5 All orthogonal lines to x^0 axis are geodesics of F with respect to the g_{ij} metric.

Hint. M maps these orthogonal lines into geodesics $y_{\vec{V}}$.

Proposition 10.10.6 *At each point $(\tau, 0, 0, 0)$ of the x^0-axis, it is*

$$g_{ij}(P) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Proof We have

$$ds_F^2(\varepsilon_i, \varepsilon_j) = ds_R^2(dM_P(\varepsilon_i), dM_P(\varepsilon_j)) = ds_R^2(e_i(\tau), e_j(\tau)) = \delta_{ij},$$

where δ_{ij} is the Kronecker symbol. □

According to the local definition of M, we can prove the main

Theorem 10.10.7 *In Fermi coordinates, at every point P belonging of the x^0 axis, the gravitational field is null, that is $\Gamma^i_{jk}(P) = 0$.*

Proof From the above exercise, we know that the line $\gamma(s) := (\tau, sv^1, sv^2, sv^3)$ is a geodesic. The geodesic equations are

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{jk}(\tau, s\ \vec{v})\frac{dx^j}{ds}\frac{dx^k}{ds} = 0,\ i, j, k \in \{0, 1, 2, 3\}.$$

Since $\dfrac{dx^\alpha}{ds} = v^\alpha$, $\alpha \in \{1, 2, 3\}$, it results $\dfrac{d^2x^\alpha}{ds^2} = 0$. Then, being $\dfrac{dx^0}{ds} = 0$ (because x^0 is parameterized by τ), it is $\dfrac{d^2x^i}{ds^2} = 0$. From the geodesic equations, it remains only

$$\Gamma^i_{\alpha\beta}(\tau, s\,\vec{v})v^\alpha v^\beta = 0,\ i \in \{0, 1, 2, 3\},\ \alpha, \beta \in \{1, 2, 3\}.$$

Now, for $s = 0$, we have

$$\Gamma^i_{\alpha\beta}(\tau, 0, 0, 0)v^\alpha v^\beta = 0,\ i \in \{0, 1, 2, 3\},\ \alpha, \beta \in \{1, 2, 3\}$$

for any given vector $\vec{v}$, therefore

$$\Gamma^i_{\alpha\beta}(\tau, 0, 0, 0) = 0,\ i \in \{0, 1, 2, 3\},\ \alpha, \beta \in \{1, 2, 3\}.$$

It remains to prove

$$\Gamma^i_{j0}(\tau, 0, 0, 0) = 0,\ i, j \in \{0, 1, 2, 3\}.$$

We know that the vectors ε_i are parallel transported along x^0-axis. Let us write the parallel transport equations for these vectors with all components constant, $\varepsilon_k = \delta^i_k$. It is

$$\frac{d\delta^i_k}{d\tau} + \Gamma^i_{jl}(\tau, 0, 0, 0)\delta^j_k \frac{dx^l}{d\tau} = 0.$$

The only non-null terms are obtained when $j = k$ and $l = 0$, which ends the proof. □

Three consequences can immediately be proved:

1. From $\Gamma_{ij,k} = g_{kr}\Gamma^r_{ij}$ we have

$$\Gamma_{ij,k}(\tau, 0, 0, 0) = 0.$$

2. From $\dfrac{\partial g_{ij}}{\partial x^k} = \Gamma_{ik,j} + \Gamma_{jk,i}$ it results

$$\frac{\partial g_{ij}}{\partial x^k}(\tau, 0, 0, 0) = 0.$$

3. From

$$\lim_{h\to 0} \frac{\dfrac{\partial g_{ij}}{\partial x^k}(\tau + h, 0, 0, 0) - \dfrac{\partial g_{ij}}{\partial x^k}(\tau, 0, 0, 0)}{h} = 0$$

we obtain

$$\frac{\partial^2 g_{ij}}{\partial x^0 \partial x^k}(\tau, 0, 0, 0) = 0.$$

Starting from these considerations, the Fermi coordinates offer another view, more physical than geometrical, about the field equations in vacuum.

10.10.2 The Fermi Viewpoint on the Einstein Field Equations in Vacuum

Consider the tidal acceleration equations, written in Fermi's coordinates, with respect to a freely falling observer whose world line has the equation $a^h(\tau) = (\tau, 0, 0, 0, 0)$. Suppose this world line is part of a family of geodesic $x^h(\tau, q)$ such that $x^h(\tau, 0) = a^h(\tau)$.

Therefore
$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -K^h_j \frac{\partial x^j}{\partial q},$$

where
$$K^h_j = R^h_{ijk}\frac{dx^i}{d\tau}\frac{dx^k}{d\tau}.$$

For the components of our curve $a^h(\tau) = (\tau, 0, 0, 0)$, there is only the term R^h_{0j0} for K^h_j.

So, the relativistic equations of tidal acceleration vector along the curve $a^h(\tau) = (\tau, 0, 0, 0)$ are
$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -R^h_{0j0}\frac{\partial x^j}{\partial q}, \ h, j \in \{0, 1, 2, 3\}.$$

Theorem 10.10.8 *In Fermi's coordinates, the tidal acceleration equations along the curve $a^h(\tau) = (\tau, 0, 0, 0)$ have the form*
$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -\frac{\partial \Gamma^h_{00}}{\partial x^j}\frac{\partial x^j}{\partial q}.$$

Proof Denote by A a point belonging to the curve a^h. Since $\Gamma^i_{jk}(A) = 0$, it results
$$R^h_{0j0}(A) = \frac{\partial \Gamma^h_{00}}{\partial x^j}(A) - \frac{\partial \Gamma^h_{0j}}{\partial x^0}(A).$$

Now, $\Gamma^h_{0j}(\tau, 0, 0, 0) = 0$ for every τ, that is $\dfrac{\partial \Gamma^h_{0j}}{\partial x^0}(A) = 0$, therefore
$$K^h_j(A) = R^h_{0j0}(A) = \frac{\partial \Gamma^h_{00}}{\partial x^j}(A).$$

The tidal acceleration equations for all points $(\tau, 0, 0, 0)$ become
$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -\frac{\partial \Gamma^h_{00}}{\partial x^j}\frac{\partial x^j}{\partial q}.$$

□

It remains to compute $\dfrac{\partial \Gamma^h_{00}}{\partial x^j}(A)$.

Theorem 10.10.9 *It is*

$$\frac{\partial \Gamma^h_{00}}{\partial x^j}(A) = \pm \frac{\partial^2 g_{00}}{\partial x^h \partial x^j}$$

Proof $\Gamma^h_{00} = g^{hs}\Gamma_{00,s} = \dfrac{g^{hs}}{2}\left(2\dfrac{\partial g_{0s}}{\partial x^0} - \dfrac{\partial g_{00}}{\partial x^s}\right)$

$$\frac{\partial \Gamma^h_{00}}{\partial x^j} = \frac{1}{2}\frac{\partial g^{hs}}{\partial x^j}\left(2\frac{\partial g_{0s}}{\partial x^0} - \frac{\partial g_{00}}{\partial x^s}\right) + \frac{g^{hs}}{2}\left(2\frac{\partial^2 g_{0s}}{\partial x^j \partial x^0} - \frac{\partial^2 g_{00}}{\partial x^j \partial x^s}\right).$$

We know $2\dfrac{\partial g_{0s}}{\partial x^0}(A) - \dfrac{\partial g_{00}}{\partial x^s}(A) = 0$ and $\dfrac{\partial^2 g_{0s}}{\partial x^j \partial x^0}(A) = 0$, therefore

$$\frac{\partial \Gamma^h_{00}}{\partial x^j}(A) = -\frac{g^{hs}}{2}\frac{\partial^2 g_{00}}{\partial x^j \partial x^s}(A).$$

Since $g^{00}(A) = 1,\ g^{\alpha\alpha}(A) = -1,\ \alpha \in \{1, 2, 3\},\ g^{hs} = 0$ when $h \neq s$, we obtain

$$\frac{\partial \Gamma^h_{00}}{\partial x^j}(A) = \pm \frac{1}{2}\frac{\partial^2 g_{00}}{\partial x^h \partial x^j}.$$

□

Let us construct now the matrix K^h_j. It results in $\dfrac{\partial \Gamma^\alpha_{00}}{\partial x^\beta}(A) = \dfrac{1}{2}\dfrac{\partial^2 g_{00}}{\partial x^\alpha \partial x^\beta}(A),\ \alpha, \beta \in \{1, 2, 3\}$, i.e.

$$K^\alpha_\beta(A) = K^\beta_\alpha(A) = \frac{1}{2}\frac{\partial^2 g_{00}}{\partial x^\alpha \partial x^\beta}(A),\ \alpha, \beta \in \{1, 2, 3\}.$$

Using $\dfrac{\partial^2 g_{00}}{\partial x^k \partial x^0}(A) = 0,\ k \in \{0, 1, 2, 3\}$, it is

$$K^0_j = \frac{\partial \Gamma^0_{00}}{\partial x^j}(A) = \frac{1}{2}\frac{\partial^2 g_{00}}{\partial x^j \partial x^0}(A) = 0,\ j \in \{0, 1, 2, 3\}$$

and

$$K^h_0 = \frac{\partial \Gamma^h_{00}}{\partial x^0}(A) = \frac{1}{2}\frac{\partial^2 g_{00}}{\partial x^0 \partial x^h}(A) = 0,\ h \in \{0, 1, 2, 3\}.$$

Therefore, in Fermi's coordinates, the tidal acceleration equations

$$\frac{\nabla^2}{d\tau^2}\frac{\partial x^h}{\partial q} = -K^h_j \frac{\partial x^j}{\partial q}$$

along the world line $a^h(\tau) = (\tau, 0, 0, 0)$ highlight the symmetric matrix

$$K_j^h(A) = \frac{1}{2}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \dfrac{\partial^2 g_{00}}{(\partial x^1)^2} & \dfrac{\partial^2 g_{00}}{\partial x^1 \partial x^2} & \dfrac{\partial^2 g_{00}}{\partial x^1 \partial x^3} \\ 0 & \dfrac{\partial^2 g_{00}}{\partial x^2 \partial x^1} & \dfrac{\partial^2 g_{00}}{(\partial x^2)^2} & \dfrac{\partial^2 g_{00}}{\partial x^2 \partial x^3} \\ 0 & \dfrac{\partial^2 g_{00}}{\partial x^3 \partial x^1} & \dfrac{\partial^2 g_{00}}{\partial x^3 \partial x^2} & \dfrac{\partial^2 g_{00}}{(\partial x^3)^2} \end{pmatrix}$$

having, as components, second-order partial derivatives.

The information about the gravitational field depends on the gravitational potential.

If in these Fermi's coordinates, we identify the classical gravitational potential Φ as $\frac{1}{2}g_{00}$, the Hessian matrix of the gravitational potential Φ,

$$d^2\Phi_{\bar{x}} = \left(\frac{\partial^2 \Phi(\bar{x})}{\partial x^i \partial x^k}\right)_{i,k}$$

can be identified with the "spatial part" $\frac{1}{2}\left(\frac{\partial^2 g_{00}}{\partial x^\alpha \partial x^\beta}\right)_{\alpha,\beta\in\{1,2,3\}}$ of the matrix K_j^h.

The information encapsulates in the trace of the Hessian of the gravitational field, that is, the vacuum field equation $\nabla^2\Phi = 0$ appears when we consider the trace of entire matrix K_j^h in the form $TrK_j^h = K_h^h = 0$. This means that $K_h^h = R_{ihk}^h = 0$, i.e. $R_{ik} = 0$.

Now, we apply the Principle of General Covariance.

The equations

$$R_{ij} = 0$$

represent, in any system of coordinates, the *relativistic field equations in vacuum*.

10.10.3 The Gravitational Coupling in the Einstein Field Equations: $K = \frac{8\pi G}{c^4}$

Let us consider the energy–momentum tensor as a perfect fluid. We can choose such a tensor as a 4×4 symmetric matrix (T^{ij})

$$(T^{ij}) = \begin{pmatrix} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{10} & T^{11} & T^{12} & T^{13} \\ T^{20} & T^{21} & T^{22} & T^{23} \\ T^{30} & T^{31} & T^{32} & T^{33} \end{pmatrix}$$

whose most important property is its null divergence expressed in terms of covariant derivative,

$$T^{kl}_{;l} = 0.$$

This property means that "at each moment, the quantity of matter and energy in the interior of a given infinitesimal parallelepiped is constant".

A first example of energy–momentum tensor was related to the Friedmann–Lemaître–Robertson–Walker metric of the Universe. In fact, the key point in the computations of the metric in geometric coordinates was related to the chosen form of energy–momentum tensor. Physicists proposed to look at galaxies as molecules of an ideal gas. In this case, the contravariant energy–momentum tensor was

$$T^{ij} = (\rho + p)u^i u^j - pg^{ij},$$

where g^{ij} are the inverse components of the metric tensor matrix which satisfies Einstein's field equations $R_{ij} - \frac{1}{2}Rg_{ij} = KT_{ij}$,

- ρ is the density;
- p is the pressure;
- u^i are the components $(u^t, v_x u^t, v_y u^t, v_z u^t)$ of the gas 4-velocity. The previous null divergence property is obviously recovered.

Energy and matter can be seen in different ways according to physics models.

The next description, known as the *energy–momentum tensor of a swarm of particles*, is useful to determine the constant K in Einstein's field equations

$$R_{ij} - \frac{1}{2}Rg_{ij} = KT_{ij}.$$

How can we describe a *swarm of particles*?

They have to be identical, they have to be uniformly distributed in space, and they have to be non-interacting. Each particle has the rest mass m_0 and we suppose that, in an unit of volume of a given space-time, if the swarm is at rest, there are exactly n_0 particles (see [46]).

The mass can be incorporated in a 4-*momentum vector*

$$\mathbb{P} := (m, m \overrightarrow{v}),$$

where m is the mass of each non-interacting particle which moves at speed v. If the swarm is at rest,

$$\mathbb{P}_0 = (m_0, \overrightarrow{0}).$$

The proper 4-*velocity of a particle moving at speed v* is

$$\mathbb{V} := \left(\frac{1}{\sqrt{1 - v^2}}, \frac{\vec{v}}{\sqrt{1 - v^2}} \right),$$

that is, $\mathbb{P} = m_0 \mathbb{V}$.

Another 4-vector can be related to the number of particles, denoted by n, in the unit of volume of the previous space-time which move at speed v,

$$\mathbb{N} := (n, n\ \vec{v}).$$

At rest, we choose

$$\mathbb{N}_0 = (n_0, \vec{0}).$$

It results in $\mathbb{N} = n_0 \mathbb{V}$.

If we define the *density of mass for the swarm* by the product between the mass of a particle and the number of particles in a unit volume of the space-time, we have

$$\rho_v := mn$$

if the swarm is moving at speed v and

$$\rho_0 := m_0 n_0$$

if the swarm is at rest. Even if

$$mn = \frac{m_0}{\sqrt{1 - v^2}} \frac{n_0}{\sqrt{1 - v^2}} = \frac{m_0 n_0}{1 - v^2},$$

that is, the product of first components is not a covariant quantity, the $(1, 0)$ contravariant vectors $\mathbb{P} = (p^0, p^1, p^2, p^3)$ and $\mathbb{N} = (n^0, n^1, n^2, n^3)$ produce a $(2, 0)$ contravariant tensor,

$$T^{ij} := p^i n^j = \begin{pmatrix} p^0 n^0 & p^0 n^1 & p^0 n^2 & p^0 n^3 \\ p^1 n^0 & p^1 n^1 & p^1 n^2 & p^1 n^3 \\ p^2 n^0 & p^2 n^1 & p^2 n^2 & p^2 n^3 \\ p^3 n^0 & p^3 n^1 & p^3 n^2 & p^3 n^3 \end{pmatrix},$$

such that the mass-density is incorporated in the T^{00}-component.

Now let us cancel the geometric coordinates which helped us to find a possible energy–momentum tensor and consider the dimensional coordinates $(x^0, x^1, x^2, x^3) = (ct, x, y, z)$. It results in

$$\mathbb{V} = (c, \dot{x}, \dot{y}, \dot{z}) = (c, v^1, v^2, v^3) = (c, \vec{v}),$$

$$\mathbb{P} = \left(\frac{E}{c}, p^1, p^2, p^3\right),$$

where $E = mc^2$ is the relativistic energy of a particle of the swarm and

$$\mathbb{N} = (nc, nv^1, nv^2, nv^3).$$

The *energy–momentum tensor* becomes

$$T^{ij} := \begin{pmatrix} En & Env^1/c & Env^2/c & Env^3/c \\ cp^1n & p^1nv^1 & p^1nv^2 & p^1nv^3 \\ cp^2n & p^2nv^1 & p^2nv^2 & p^2nv^3 \\ cp^3n & p^3nv^1 & p^3nv^2 & p^3nv^3 \end{pmatrix}.$$

We have

$$T^{00} = En = mnc^2 = \rho c^2,$$

therefore we can call T^{00} the *density of the relativistic energy of the swarm.* One may describe all the components of the energy–momentum tensor according to the physic units. However only T^{00} is used to determine K.

Suppose we are working in Fermi's coordinates with a swarm of non-interacting particles which move together such that the world line of a particle is the x^0-axis. Therefore

$$\mathbb{V} = (c, 0, 0, 0)$$

and the energy–momentum T^{ij} has only one term,

$$T^{00} = T_0^0 = T_{00} = T = \rho c^2.$$

Along the x^0-axis, the Einstein equations, written in the form

$$R_{ij} = K\left(T_{ij} - \frac{1}{2}T g_{ij}\right),$$

become the only equation

$$R_{00} = K\left(T_{00} - \frac{1}{2}T g_{00}\right) = K\left(\rho c^2 - \frac{1}{2}\rho c^2\right),$$

that is

$$R_{00} = K\frac{1}{2}\rho c^2.$$

Since dimensionally we have

$$R_{00} = K^s_s = \frac{1}{2}\sum_{\alpha=1}^{3}\frac{\partial^2 g_{00}}{\partial(x^\alpha)^2} = \frac{1}{c^2}\nabla^2\Phi = \frac{1}{c^2}4\pi G\rho,$$

it results

$$\frac{1}{c^2}4\pi G\not\rho = K\frac{1}{2}\not\rho c^2,$$

that is

$$K = \frac{8\pi G}{c^4}.$$

Therefore *Einstein's field equations* are

$$R_{ij} - \frac{1}{2}Rg_{ij} = \frac{8\pi G}{c^4}T_{ij},$$

where the gravitational coupling is written in physical constants.

10.11 Weak Gravitational Field and the Classical Counterparts of the Relativistic Equations

We are interested in seeing under which conditions it is possible to recover the Classical Mechanics basic formulas involving gravity from the relativistic formulas seen in the present chapter.

Let us discuss this point in a mathematical language: In this section we show that, in the case of a "week gravitational field", for "particles with slow motion", the classical field equations emerge from their *relativistic counterparts*, that is,

$$\frac{d^2x^i}{d\tau^2} = -\Gamma^i_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau} \longrightarrow \frac{d^2x^\alpha}{dt^2} = -\frac{\partial\Phi}{\partial x^\alpha},$$

$$R_{ij} - \frac{1}{2}g_{ij}R = \frac{8\pi G}{c^4}T_{ij} \longrightarrow \nabla^2\Phi = 4\pi\rho,$$

$$R_{ij} = 0 \longrightarrow \nabla^2\Phi = 0.$$

A complete treatment of these results can be found in [46]. Of course, the basic facts were presented by Einstein himself in [133].

Consider the Minkowski metric which describes a frame with no gravity

$$J := J_{ij} = J^{ij} := \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Adding small variations of order $\frac{1}{c^k}$, $k \geq 2$ we introduce gravitational effects. Therefore, the following definition is necessary to introduce our working frame:

Definition 10.11.1 A weak gravitational field is described by a metric

$$g_{ij} = J_{ij} + \frac{1}{c^2} g_{ij}^{(2)} + \frac{1}{c^3} g_{ij}^{(3)} + O\left(\frac{1}{c^4}\right)$$

with the supplementary properties

$$g_{ij}^{(m)} = O(1), \ \frac{\partial g_{ij}^{(m)}}{\partial t} = O(1), \ \frac{\partial g_{ij}^{(m)}}{\partial x^{\alpha}} = O(1), \ m \in \{2, 3\}, \ \alpha \in \{1, 2, 3\}.$$

Here, $g_{ij}^{(2)}$ are coefficients of the metric g_{ij} related to the factor $\frac{1}{c^2}$, etc. In other words, $\frac{1}{c^2}$ is our expansion parameter related to the strength of the field.

Let us first observe that, for a weak gravitational field, it is

$$g_{ij} = J_{ij} + O\left(\frac{1}{c^2}\right)$$

and

$$g^{ij} = J^{ij} + O\left(\frac{1}{c^2}\right).$$

Therefore, for a weak gravitational field, we have the following consequences of the previous definition:

$$g_{0j} = O\left(\frac{1}{c^2}\right), \ g_{\alpha\alpha} = O(1), \ \frac{\partial g_{ij}}{\partial x^{\alpha}} = O\left(\frac{1}{c^2}\right) \text{ and } det(g_{ij}) = -1 + O\left(\frac{1}{c^2}\right).$$

Then, it is easy to see that

$$\frac{\partial g_{ij}^{(m)}}{\partial x^0} = \frac{\partial g_{ij}^{(m)}}{\partial t} \frac{\partial t}{\partial x^0} = O(1) \frac{1}{c} = O\left(\frac{1}{c}\right).$$

In the same way,

$$\frac{\partial g_{ij}}{\partial x^0} = \frac{\partial g_{ij}}{\partial t}\frac{\partial t}{\partial x^0} = O\left(\frac{1}{c^2}\right)\frac{1}{c} = O\left(\frac{1}{c^3}\right).$$

Theorem 10.11.2 *The Christoffel symbols of a weak gravitational field have the properties*

$$\Gamma_{k0,k} = O\left(\frac{1}{c^3}\right),\ \Gamma_{ij,k} = O\left(\frac{1}{c^2}\right),\ \Gamma^0_{00} = \Gamma^h_{h0} = O\left(\frac{1}{c^3}\right),$$

$$\Gamma^\alpha_{00} = \frac{1}{2c^2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c^3}\right),\ \Gamma^i_{\alpha\beta} = O\left(\frac{1}{c^2}\right).$$

Proof We present two computations and we leave to the reader the details.
The first one:

$$\Gamma^0_{00} = g^{0i}\Gamma_{00,i} = g^{00}\Gamma_{00,0} + g^{0\alpha}\Gamma_{00,\alpha} =$$

$$\left(1 + O\left(\frac{1}{c^2}\right)\right)O\left(\frac{1}{c^2}\right) + O\left(\frac{1}{c^2}\right)\left(\frac{1}{c^3}\right) = O\left(\frac{1}{c^3}\right).$$

The second one:

$$\Gamma^\alpha_{00} = g^{\alpha i}\Gamma_{00,i} = g^{\alpha\alpha}\Gamma_{00,\alpha} + g^{\alpha 0}\Gamma_{00,0} + g^{\alpha\beta}\Gamma_{00,\beta} =$$

$$= \left(-1 + O\left(\frac{1}{c^2}\right)\right)\Gamma_{00,\alpha} + O\left(\frac{1}{c^2}\right)\left(\Gamma_{00,0} + \Gamma_{00,\beta}\right).$$

Replacing $\Gamma_{00,\alpha}$, it results in

$$\Gamma^\alpha_{00} = \left(-1 + O\left(\frac{1}{c^2}\right)\right)\Gamma_{00,\alpha} + O\left(\frac{1}{c^3}\right) =$$

$$= \left(-1 + O\left(\frac{1}{c^2}\right)\right)\left(\frac{\partial g_{0\alpha}}{\partial x^0} - \frac{1}{2}\frac{\partial g_{00}}{\partial x^\alpha}\right) + O\left(\frac{1}{c^3}\right),$$

that is,

$$\Gamma^\alpha_{00} = \frac{1}{2}\frac{\partial g_{00}}{\partial x^\alpha} + O\left(\frac{1}{c^3}\right) = \frac{1}{2c^2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c^3}\right).$$

□

Denote by $X(t) := (x^1(t), x^2(t), x^3(t))$ the trajectory of a classical particle; its classical speed is $\dot{X} := (\dot{x}^1(t), \dot{x}^2(t), \dot{x}^3(t))$.

Definition 10.11.3 The particle is "slow" if $\dot{x}^\alpha(t) = O(1),\ \alpha \in \{1, 2, 3\}$.

The corresponding worldcurve is $\overrightarrow{X} = (ct, X(t))$ and its relativistic speed is $\overrightarrow{V} = (ct, \dot{X}(t))$. Observe that

$$L(t) = \int_{t_0}^{t} || \overrightarrow{V}(s)||_g ds = \int_{t_0}^{t} \sqrt{g_{ij}\dot{x}^i(s)\dot{x}^j(s)} ds$$

has length dimension.

Parameterizing $\overrightarrow{X}$ by proper time means to consider $\tau(t) := \frac{1}{c}L(t)$. Let us observe that $\tau(t)$ has time dimension.

Theorem 10.11.4 *In a Minkowski metric, if a particle is moving "slow" uniformly along a curve parameterized by proper time, then*

$$\frac{d\tau}{dt} = 1 + O\left(\frac{1}{c^2}\right).$$

Proof From

$$d\tau = \frac{|| \overrightarrow{V} ||_M}{c} dt = \sqrt{1 - \frac{1}{c^2}\sum_{\alpha=1}^{3}(\dot{x}^\alpha(t))^2 dt} = \sqrt{1 - \frac{1}{c^2}O(1)},$$

it results

$$\frac{d\tau}{dt} = \sqrt{1 - O\left(\frac{1}{c^2}\right)} dt.$$

Since $\sqrt{1-A} \approx 1 + \frac{A}{2}$, we have

$$\frac{d\tau}{dt} = 1 + O\left(\frac{1}{c^2}\right).$$

□

Now, parameterizing with respect to τ in a metric g_{ij}, we have

$$d\tau = \frac{|| \overrightarrow{V} ||_g}{c} dt = \sqrt{\frac{g_{ij}\dot{x}^i\dot{x}^j}{c^2}} dt = \sqrt{g_{ij}\frac{\dot{x}^i}{c}\frac{\dot{x}^j}{c}} dt.$$

Theorem 10.11.5 *If a particle is moving "slow" in a weak gravitational field along a curve parameterized by proper time, then*

$$\frac{d\tau}{dt} = 1 + O\left(\frac{1}{c^2}\right).$$

Proof We have $\dot{x}^0 = \dfrac{d}{dt}(ct) = c$. The particle is “slow” and, by definition, this means $\dot{x}^\alpha = O(1)$. Therefore

$$g_{ij}\frac{\dot{x}^i}{c}\frac{\dot{x}^j}{c} = g_{00} + 2g_{0\alpha}\frac{\dot{x}^\alpha}{c} + g_{\alpha\beta}\frac{\dot{x}^\alpha}{c}\frac{\dot{x}^\beta}{c} = 1 + O\left(\frac{1}{c^2}\right).$$

We have used $g_{00} = 1,\ g_{\alpha\alpha} = O(1),\ g_{\alpha\beta} = O\left(\frac{1}{c^2}\right),\ \alpha \neq \beta,\ g_{0\alpha} = O\left(\frac{1}{c^2}\right)$.

Finally,

$$\frac{d\tau}{dt} = \sqrt{g_{ij}\frac{\dot{x}^i}{c}\frac{\dot{x}^j}{c}}dt = \sqrt{1 + O\left(\frac{1}{c^2}\right)}dt,$$

i.e.

$$\frac{d\tau}{dt} = 1 + O\left(\frac{1}{c^2}\right).$$

Observe we can also obtain

$$\frac{dt}{d\tau} = 1 + O\left(\frac{1}{c^2}\right).$$

□

Theorem 10.11.6 *If a particle is moving “slow” along a curve parameterized by proper time, then it is* $\dfrac{d^2x^0}{d\tau^2} = O\left(\dfrac{1}{c}\right)$ *and* $\dfrac{d^2x^\alpha}{d\tau^2} = \ddot{x}^\alpha + O\left(\dfrac{1}{c^2}\right)$.

Proof From $\dfrac{dx^0}{d\tau} = \dfrac{dx^0}{dt}\dfrac{dt}{d\tau} = c\left(1 + O\left(\dfrac{1}{c^2}\right)\right) = c + O\left(\dfrac{1}{c}\right)$, we obtain

$$\frac{d^2x^0}{d\tau^2} = \frac{d}{dt}\left(c + O\left(\frac{1}{c}\right)\right) = O\left(\frac{1}{c}\right),$$

and, from

$$\frac{dx^\alpha}{d\tau} = \frac{dx^\alpha}{dt}\frac{dt}{d\tau} = \dot{x}^\alpha\left(1 + O\left(\frac{1}{c^2}\right)\right) = \dot{x}^\alpha + O\left(\frac{1}{c^2}\right),$$

it results in

$$\frac{d^2x^\alpha}{d\tau^2} = \frac{d}{dt}\left(\dot{x}^\alpha + O\left(\frac{1}{c^2}\right)\right) = \ddot{x}^\alpha + O\left(\frac{1}{c^2}\right).$$

□

Theorem 10.11.7 *In a weak gravitational field, the four geodesic equations for "slow" particles reduce to the three classical equations of motion, that is,*

$$\frac{d^2x^i}{d\tau^2} = -\Gamma^i_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau},\ i, j, k \in \{0, 1, 2, 3\} \ \longrightarrow \ \frac{d^2x^\alpha}{dt^2} = -\frac{\partial \Phi}{\partial x^\alpha},\ \alpha \in \{1, 2, 3\}.$$

Proof We already proved that, if a particle is moving "slow," then $\frac{d^2x^0}{d\tau^2} = O\left(\frac{1}{c}\right)$, so the left-hand side part is a $O\left(\frac{1}{c}\right)$ quantity.

We consider

$$\frac{d^2x^0}{d\tau^2} = -\Gamma^0_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau} = -\Gamma^0_{00}\left(\frac{dx^0}{d\tau}\right)^2 - 2\Gamma^0_{0\alpha}\frac{dx^0}{d\tau}\frac{dx^\alpha}{d\tau} - \Gamma^0_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}$$

and we observe

$$\Gamma^0_{00}\left(\frac{dx^0}{d\tau}\right)^2 = O\left(\frac{1}{c^3}\right)\left(c + O\left(\frac{1}{c}\right)\right)^2 = O\left(\frac{1}{c}\right),$$

$$\Gamma^0_{0\alpha}\frac{dx^0}{d\tau}\frac{dx^\alpha}{d\tau} = O\left(\frac{1}{c^2}\right)\left(c + O\left(\frac{1}{c}\right)\right)\left(\dot{x}^\alpha + O\left(\frac{1}{c^2}\right)\right) = O\left(\frac{1}{c}\right)$$

$$\Gamma^0_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = O\left(\frac{1}{c^2}\right)\left(\dot{x}^\alpha + O\left(\frac{1}{c^2}\right)\right)\left(\dot{x}^\beta + O\left(\frac{1}{c^2}\right)\right) = O\left(\frac{1}{c^2}\right).$$

The right-hand side part of the equation of geodesic is $O\left(\frac{1}{c}\right)$, therefore, for a "slow" particle, the first geodesic equation is an equality between "very small" quantities. As a consequence, we can neglect it.

We already proved that the left-hand side part of the geodesic equations is

$$\frac{d^2x^\alpha}{d\tau^2} = \ddot{x}^\alpha + O\left(\frac{1}{c^2}\right).$$

Now, for the right-hand side part, we proceed as above.

$$\frac{d^2x^\alpha}{d\tau^2} = -\Gamma^\alpha_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau} = -\Gamma^\alpha_{00}\left(\frac{dx^0}{d\tau}\right)^2 - 2\Gamma^\alpha_{0\beta}\frac{dx^0}{d\tau}\frac{dx^\beta}{d\tau} - \Gamma^\alpha_{\beta\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau}$$

and we observe

$$\Gamma^\alpha_{00}\left(\frac{dx^0}{d\tau}\right)^2 = \left(\frac{1}{2c^2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c^3}\right)\right)\left(c^2 + O\left(1\right)\right)^2 = \frac{1}{2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c}\right),$$

$$\Gamma^{\alpha}_{0\beta}\frac{dx^0}{d\tau}\frac{dx^\beta}{d\tau} = O\left(\frac{1}{c^3}\right)\left(c + O\left(\frac{1}{c}\right)\right)\left(\dot{x}^\alpha + O\left(\frac{1}{c^2}\right)\right) = O\left(\frac{1}{c^2}\right),$$

$$\Gamma^{\alpha}_{\beta\gamma}\frac{dx^\beta}{d\tau}\frac{dx^\gamma}{d\tau} = O\left(\frac{1}{c^2}\right)\left(\dot{x}^\beta + O\left(\frac{1}{c^2}\right)\right)\left(\dot{x}^\gamma + O\left(\frac{1}{c^2}\right)\right) = O\left(\frac{1}{c^2}\right),$$

that is the right-hand side of the geodesic equations is

$$-\frac{1}{2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c}\right), \alpha \in \{1, 2, 3\}.$$

Neglecting the "small" quantities, the geodesic equations

$$\frac{d^2x^i}{d\tau^2} = -\Gamma^i_{jk}\frac{dx^j}{d\tau}\frac{dx^k}{d\tau},\ \ i, j, k \in \{0, 1, 2, 3\}$$

reduce to

$$\frac{d^2x^\alpha}{dt^2} = -\frac{\partial \Phi}{\partial x^\alpha},\ \ \alpha \in \{1, 2, 3\},\ \ \Phi = \frac{1}{2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha}.$$

□

Theorem 10.11.8 *The relativistic equations of the weak gravitational field reduce to the classical Poisson field equation:*

$$R_{ij} - \frac{1}{2}g_{ij}R = \frac{8\pi G}{c^4}T_{ij} \quad \longrightarrow \quad \nabla^2\Phi = 4\pi G\rho.$$

Proof We consider the relativistic equation written with respect to the Laue scalar

$$R_{ij} = \frac{8\pi G}{c^4}\left(T_{ij} - \frac{1}{2}g_{ij}T\right).$$

Suppose the matter–energy tensor written in the previous form consisting of a swarm of identical non-interacting particles having density ρ. The only non-zero component is $T_{00} = \rho c^2$. The right member is

$$\begin{aligned}\frac{8\pi G}{c^4}\left(T_{ij} - \frac{1}{2}g_{ij}T\right) &= \frac{8\pi G}{c^4}\left[\rho c^2 - \frac{1}{2}\left(1 + O\left(\frac{1}{c^2}\right)\right)\rho c^2\right]\\ &= \frac{8\pi G}{c^4}\left[\frac{\rho c^2}{2} + O(1)\right] =\\ &= \frac{1}{c^2}\cdot 4\pi G\rho + O\left(\frac{1}{c^4}\right).\end{aligned}$$

Now, let us look at the left member.

$$R_{00} = R^s_{0s0} = R^0_{000} + R^\alpha_{0\alpha 0} = R^\alpha_{0\alpha 0} = \frac{\partial \Gamma^\alpha_{00}}{\partial x^\alpha} - \frac{\partial \Gamma^\alpha_{0\alpha}}{\partial x^0} + \Gamma^m_{00}\Gamma^\alpha_{m\alpha} - \Gamma^m_{0\alpha}\Gamma^\alpha_{m0}.$$

The two products of Christoffel symbols are at least $O\left(\frac{1}{c^4}\right)$.

Then, since $\Gamma^\alpha_{0\alpha} = O\left(\frac{1}{c^3}\right)$, it results $\frac{\partial \Gamma^\alpha_{0\alpha}}{\partial x^0} = \frac{\partial \Gamma^\alpha_{0\alpha}}{\partial t}\frac{\partial t}{\partial x^0} = O\left(\frac{1}{c^3}\right)\frac{1}{c} = O\left(\frac{1}{c^4}\right)$.

If we consider the derivative with respect to x^α of the equality $\Gamma^\alpha_{00} = \frac{1}{2c^2}\frac{\partial g^{(2)}_{00}}{\partial x^\alpha} + O\left(\frac{1}{c^3}\right)$, it is

$$\frac{\partial \Gamma^\alpha_{00}}{\partial x^\alpha} = \frac{1}{2c^2}\frac{\partial^2 g^{(2)}_{00}}{(\partial x^\alpha)^2} + O\left(\frac{1}{c^3}\right), \ \alpha \in \{1, 2, 3\},$$

that is,

$$R_{00} = \frac{1}{2c^2}\sum_{\alpha=1}^{3}\frac{\partial^2 g^{(2)}_{00}}{(\partial x^\alpha)^2} + O\left(\frac{1}{c^3}\right).$$

Since $\Phi = \frac{1}{2}g^{(2)}_{00}$, we finally obtain the right member as

$$R_{00} = \frac{1}{c^2}\nabla^2\Phi + O\left(\frac{1}{c^3}\right).$$

Neglecting the small quantities, the relativistic weak field equations reduce to the classical Poisson field equation

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

□

Corollary 10.11.9 *The relativistic equations of the weak gravitational field in vacuum reduce to the classical Laplace field equation in vacuum:*

$$R_{ij} = 0 \quad \longrightarrow \quad \nabla^2\Phi = 0.$$

10.12 The Einstein Static Universe and the Cosmological Constant

On 8 February 1917, the Prussian Academy of Science in Berlin published a paper by Albert Einstein where the first application of his theory, published on 25 November 1915, was presented [92]. The paper discussed a dynamical system describing a static space-time representing the Universe. It can be considered the birth of modern Cosmology. The model proved wrong after the discovery or recession of galaxies by Hubble, however it is important because several concepts presented in it were used in the further developments of this science. Let us give now a quick presentation of it.

Consider before some mathematical preliminaries. Let $U := (0, 2\pi) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$, $(\alpha, \beta, \theta) \in U$ and the map $f : U \to \mathbb{R}^4$,

$$f(\alpha, \beta, \theta) := \begin{cases} u_1 = r \cos\alpha \cos\beta \cos\theta \\ u_2 = r \sin\alpha \cos\beta \cos\theta \\ u_3 = r \sin\beta \cos\theta \\ u_4 = r \sin\theta \end{cases}$$

It is easy to see that $u_1^2 + u_2^2 + u_3^2 + u_4^2 = r^2$.

The image of f in $\mathbb{R}^4$, $f(U)$, is the 3-sphere centred at the origin having radius r.

In classical notation, it is $S^3(O, r)$.

The coefficients of the metric are computed with the Euclidean inner product of the partial derivatives of f. The only non-zero coefficients of the metric are

$$g_{\alpha\alpha} = r^2 \cos^2\beta \cos^2\theta, \ g_{\beta\beta} = r^2 \cos^2\theta, \ g_{\theta\theta} = r^2,$$

therefore the metric induced by the Euclidean 4-space in the tangent 3-planes of this surface is

$$ds^2 = g_{\alpha\alpha}(dx^\alpha)^2 + g_{\beta\beta}(dx^\beta)^2 + g_{\theta\theta}(dx^\theta)^2.$$

The volume of $S^3(O, r)$ is

$$Vol[S^3(O, r)] = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} \sqrt{det(g_{ij})} d\alpha d\beta d\theta =$$

$$= r^3 \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \int_{-\pi/2}^{\pi/2} \cos\beta \cos^2\theta d\alpha d\beta d\theta = 2\pi^2 r^3.$$

Einstein's static universe is $\mathbb{E} := \mathbb{R} \times S^3(O, r)$ and it does not evolve. If we accept the Einstein point of view, the Universe is static. Let us study the mathematical formalism to find its properties.

Let us choose the homogeneous coordinates $(x^0, x^1, x^2, x^3) := (ct, \alpha, \beta, \theta)$. Now, we use the Minkowski product of partial derivatives to determine the metric of $\mathbb{E}$. It results in a simple form for the metric of Einstein's static universe

$$ds^2 = (dx^0)^2 - r^2 \cos^2 x^1 \cos^2 x^3 (dx^1)^2 - r^2 \cos^2 x^3 (dx^2)^2 - r^2 (dx^3)^2.$$

The only non-zero Christoffel symbols are

$$\Gamma^1_{12} = \Gamma^1_{21} = -\tan x^2;\ \Gamma^1_{13} = \Gamma^1_{31} = \Gamma^2_{23} = \Gamma^2_{32} = -\tan x^3;$$

$$\Gamma^2_{11} = \sin x^2 \cos x^2;\ \Gamma^3_{11} = \cos^2 x^2 \sin x^3 \cos x^3;\ \Gamma^3_{22} = \sin x^3 \cos x^3.$$

Using

$$R_{jl} = R^h_{jhl} = \frac{\partial \Gamma^h_{jl}}{\partial x^h} - \frac{\partial \Gamma^h_{jh}}{\partial x^l} + \Gamma^s_{jl}\Gamma^h_{sh} - \Gamma^s_{jh}\Gamma^h_{sl}$$

it results in

$$R_{00} = \frac{\partial \Gamma^h_{00}}{\partial x^h} - \frac{\partial \Gamma^h_{0h}}{\partial x^0} + \Gamma^s_{00}\Gamma^h_{sh} - \Gamma^s_{0h}\Gamma^h_{s0} = 0.$$

In the same way, we compute $R_{ii} = -\frac{2}{r^2} g_{ii},\ i \in \{1, 2, 3\}$.

If we consider the Einstein field equations with the cosmological constant included, it is

$$R_{ij} - \frac{1}{2} R g_{ij} + \Lambda g_{ij} = \frac{8\pi G}{c^4} T_{ij,}$$

in the equivalent form

$$R_{ij} - \Lambda g_{ij} = \frac{8\pi G}{c^4} \left(T_{ij} - \frac{1}{2} T g_{ij} \right).$$

If $i = j = 0$, the Einstein field equations reduced to:

$$R_{00} - \Lambda g_{00} = \frac{8\pi G}{c^4} \left(T_{00} - \frac{1}{2} T g_{00} \right).$$

If $i = j \neq 0$,

$$R_{ii} - \Lambda g_{ii} = \frac{8\pi G}{c^4} \left(T_{ii} - \frac{1}{2} T g_{ii} \right).$$

Einstein assumed that the matter–energy appears only in the form of a swarm of non-interacting particles of uniform density ρ. Only $T^{00} = \rho c^2$, all the other components are 0. The first equation becomes

$$\Lambda = -\frac{4\pi G}{c^2}\rho.$$

The other three equations lead to a single equation,

$$-\frac{2}{r^2} - \Lambda \cdot (-1) = \frac{8\pi G}{c^4}\left(0 - \frac{1}{2}\rho c^2 \cdot (-1)\right),$$

i.e.

$$-\frac{1}{r^2} = \Lambda$$

Combining the two equations, it results in

$$r = \frac{c}{2\sqrt{\pi G \rho}}.$$

The total amount of matter, denoted as $\mathbb{M}_{\mathbb{E}}$, in Einstein's universe is finite. It is computed as the product between the spatial density ρ and the volume of $S^3(O, r)$, that is,

$$\mathbb{M}_{\mathbb{E}} = \rho \cdot 2\pi^2 r^3 = \rho \cdot 2\pi^2 \cdot r \cdot r^2 = \cancel{\rho} 2\pi^{\cancel{2}} \frac{1}{\sqrt{-\Lambda}} \frac{c^2}{\cancel{4\pi} G \cancel{\rho}} = \frac{\pi c^2}{2G\sqrt{-\Lambda}}.$$

Of course, the radius of this universe is the constant computed before, that is,

$$r = \frac{c}{2\sqrt{\pi G \rho}}.$$

Einstein did not consider any more this model after Hubble discovered the evidence of cosmological expansion, however the concept of cosmological constant, used here, was considered later in view of the issues of cosmological inflation and dark energy discussed above.

10.13 Cosmic Strings

Cosmic strings are one-dimensional hypothetical structures emerged as topological defects of space-time in some phase transition after the Big Bang. They should have acted like seeds for cosmological large-scale structure formation [158]. A nice presentation of this topic is in [142]. Here, we adapted it for a metric with signature $(+ - - -)$.

In a system of geometric coordinates (t, r, ϕ, z), the metric which describes a static string around and along the z-axis is

$$ds^2 = dt^2 - dr^2 - f^2(r)d\phi^2 - dz^2,$$

where $f(r)$ has to be determined. The only non-zero Christoffel symbols are

$$\Gamma^1_{22} = f(r)f'(r);\ \Gamma^2_{12} = \Gamma^2_{21} = \frac{f'(r)}{f(r)}.$$

We have

$$\frac{\partial\Gamma^2_{12}}{\partial r} = \frac{\partial\Gamma^2_{21}}{\partial r} = \frac{f''(r)f(r) - (f'(r))^2}{f^2(r)}; \quad \frac{\partial\Gamma^1_{22}}{\partial r} = (f'(r))^2 + f''(r)f(r).$$

Then

$$R^2_{121} = \cancel{\frac{\partial\Gamma^2_{11}}{\partial\phi}} - \frac{\partial\Gamma^1_{12}}{\partial r} + \cancel{\Gamma^2_{s2}\Gamma^s_{11}} - \Gamma^2_{s1}\Gamma^s_{12} = -\frac{f''(r)f(r) - \cancel{(f'(r))^2}}{f^2(r)} - \cancel{\frac{(f'(r))^2}{f^2(r)}}$$
$$= \frac{-f''(r)}{f(r)},$$

that is

$$R_{11} = R^s_{1s1} = \frac{-f''(r)}{f(r)}.$$

From

$$R^1_{212} = \frac{\partial\Gamma^1_{22}}{\partial r} - \cancel{\frac{\partial\Gamma^1_{21}}{\partial\phi}} + \cancel{\Gamma^1_{s1}\Gamma^s_{22}} - \Gamma^1_{s2}\Gamma^s_{21}$$
$$= \cancel{(f'(r))^2} + f''(r)f(r) - \cancel{f(r)f'(r)\frac{f'(r)}{f(r)}},$$

we deduce

$$R_{22} = R^s_{2s2} = f''(r)f(r).$$

Since

$$R = R^i_i = g^{11}R_{11} + g^{22}R_{22} = \cancel{\frac{f''(r)}{f(r)}} + \cancel{\left(\frac{-1}{f^2(r)}\right)f''(r)f(r)} = 0,$$

we obtain

$$\begin{cases} R_{11} + \frac{1}{2}\cancel{R}g_{11} = \dfrac{-f''(r)}{f(r)} \\ R_{22} + \frac{1}{2}\cancel{R}g_{22} = f''(r)f(r). \end{cases}$$

Einstein's field equations in geometric coordinates are

$$R_{ij} - \frac{1}{2}Rg_{ij} = 8\pi G T_{ij}.$$

Now using the previous result, that is, $R_1^1 = -R_2^2 = \dfrac{f''(r)}{f(r)}$, to choose the tensor T_j^i: $T_j^i = 0$ except $T_1^1 = -T_2^2 = \sigma(r)$, where σ is a given positive smooth function expressing the unit energy density of the string. It remains to find f which satisfies the equation

$$f''(r) = -8\pi G\sigma(r)f(r).$$

In order to avoid singularities, the metric has to reduce to the flat Minkowski metric at the origin. This means that $f(r)$ approaches r for small r. Therefore, two conditions for f have to be given: $f(0) = 0$ and $f'(0) = 1$.

Denote by r_s the value of r such that $\sigma(r) = 0$ if $r \geq r_s$. The physical area of a ring of radius r and width dr in the given metric is

$$\int_0^{2\pi}\int_0^{dr}\sqrt{-\det g_{ij}}\,d\phi dz = 2\pi f(r)dr.$$

It implies that the string energy per unit of length is

$$E = \int_0^{r_s}\sigma(r)2\pi f(r)dr,$$

therefore, integrating the equation $R_{11} = -8\pi G\sigma(r)$, we obtain

$$f'(r_s) - f'(0) = \int_0^{r_s} f''(r)dr = -4G\int_0^{r_s}\sigma(r)2\pi f(r)dr = -4GE.$$

If $r > r_s$, it is

$$f'(r) = 1 - 4GE.$$

Integrating again, it results in

$$f(r) = (1 - 4GE)r + K,$$

where K is a constant that should be 0 because $f(0) = 0$.

The metric obtained outside the string is

$$ds^2 = dt^2 - dr^2 - (1 - 4GE)^2 r^2 d\phi^2 - dz^2,$$

which is a flat metric. In this way, we have the simplest expression of a metric describing an infinite, straight, and independent of time string lying along the z-axis of our chosen coordinate system.

10.14 Planar Gravitational Waves

Also the issue of gravitational waves can be dealt under the standard of our geometric approach. Here, we shall give just a short summary of this important topic. For a detailed discussion on the history, the theoretical foundation, and the discovery, we refer the reader to specialized texts and papers [1, 135, 178].

In order to deal with gravitational waves, we have to obtain metrics in a geometric coordinate system as

$$ds^2 = (\alpha_{ij} + \epsilon h_{ij}) dx^i dx^j,$$

such that both Einstein's field equations and $\Box h_{ij} = 0$ are satisfied.
Here,

$$\Box := (\partial^0)^2 - (\partial^1)^2 - (\partial^2)^2 - (\partial^3)^2,$$

where

$$(\partial^k)^2 := \frac{\partial^2}{(\partial x^k)^2}.$$

If μ_{ij} are the coefficients of the classical Minkowski metric, the previous *d'Alembert operator* definition can be written in a simpler form as

$$\Box := \mu_{ij} \partial^i \partial^j.$$

In any case it is difficult to find such kind of metrics because we have to develop the whole theory of tensor perturbations in General Relativity [135]. Instead of trying to find out general gravitational wave solutions, let us focus on planar gravitational waves which are easier to obtain. See [171] for details. We follow this last reference to offer a first glance on this subject. Consider the metric

$$ds^2 = (1 + \cos(t - x)[2 + \cos(t - x)])dt^2 - (1 - \cos^2(t - x))dx^2 - dy^2 - dz^2 -$$

$$- 2\cos(t - x)(1 + \cos(t - x))dtdx.$$

The previous metric can be seen as a slightly perturbation of the Minkowski metric μ_{ij} because

$$g_{ij} = \mu_{ij} + h_{ij}.$$

A metric which such coefficients is called a *linearized metric*. If it satisfies both Einstein's vacuum field equations and the conditions $\Box\, h_{ij} = 0$, such a linearized metric describes *gravitational planar waves*.

Theorem 10.14.1 *The previous metric having the following non-zero coefficients*

$$g_{00} = 1 + \cos(t-x)(2+\cos(t-x));\ \ g_{11} = 1 - \cos^2(t-x);\ g_{22} = g_{33} = -1;$$

$$g_{01} = g_{10} = -\cos(t-x)(1+\cos(t-x));$$

describes gravitational planar waves.

Proof We have the non-zero perturbations

$$h_{00} = \cos(t-x)\ (2+\cos(t-x));\ \ h_{11} = -\cos^2(t-x);$$

$$h_{01} = h_{10} = -\cos(t-x)(1+\cos(t-x)).$$

It is easy to see that

$$\Box\ \cos(t-x) = (\partial^0)^2 \cos(t-x) - (\partial^1)^2 \cos(t-x) = 0$$

and

$$\Box\ \cos(t-x) = (\partial^0)^2 \cos^2(t-x) - (\partial^1)^2 \cos^2(t-x) = 0,$$

therefore the conditions

$$\Box\, h_{ij} = 0$$

are fulfilled.

Now it seems we have a lot of difficult computations to do in order to prove $R_{ij} = 0$.

It is easy to provide a coordinate transformation for which the Ricci tensor can be computed. Let us consider the Minkowski metric

$$ds^2 = d\bar{t}^2 - d\bar{x}^2 - d\bar{y}^2 - d\bar{z}^2$$

and the transformation

$$\bar{t} = t + \sin(t - x),\ \bar{x} = x,\ \bar{y} = y,\ \bar{z} = z.$$

Since

$$d\bar{t} = (1 + \cos(t - x))dt - \cos(t - x)dx,\ d\bar{x} = dx,\ d\bar{y} = dy,\ d\bar{z} = dz,$$

the Minkowski metric turns into our metric

$$ds^2 = (1 + \cos(t - x)[2 + \cos(t - x)])dt^2 - (1 - \cos^2(t - x))dx^2 - dy^2 - dz^2 -$$

$$- 2\cos(t - x)(1 + \cos(t - x))dtdx.$$

Therefore $\bar{R}_{ij} = 0$ transforms into the desired $R_{ij} = 0$. □

Gravitational waves were among the early predictions of Einstein's General Relativity. Their discovery a century later can be considered one of the greatest achievements of modern Science.

10.15 The Gödel Universe

Another interesting metric is the one describing the so-called Gödel Universe [108] published[7] in 1949.

First, let us show that the metric

$$ds^2 = (dx^0)^2 - (dx^1)^2 + \frac{e^{2x^1}}{2}(dx^2)^2 - (dx^3)^2 + 2e^{x^1}dx^0dx^2$$

written in geometric coordinates satisfies Einstein's field equations in the case when the cosmological constant is $\Lambda = \dfrac{1}{2}$ and the stress-energy tensor describes dust with constant density $\rho = \dfrac{1}{8\pi G}$. This is called Gödel's first metric . The coefficients involved in computations are

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & e^{x^1} & 0 \\ 0 & -1 & 0 & 0 \\ e^{x^1} & 0 & \dfrac{e^{2x^1}}{2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},\quad (g^{ij}) = \begin{pmatrix} -1 & 0 & 2e^{-x^1} & 0 \\ 0 & -1 & 0 & 0 \\ 2e^{-x^1} & 0 & -2e^{-2x^1} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Since only the derivative with respect to x^1 of the metric coefficients can be non-zero, the first-type Christoffel symbols are

[7] The story of this solution is very nice. Kurt Gödel gave it to Albert Einstein as a gift for his 70th birthday when they both lived in Princeton.

$$\Gamma_{12,2} = \Gamma_{21,2} = \frac{1}{2}e^{2x^1},\ \Gamma_{01,2} = \Gamma_{10,2} = \frac{1}{2}e^{x^1};$$

$$\Gamma_{02,1} = \Gamma_{20,1} = -\frac{1}{2}e^{x^1};\ \Gamma_{22,1} = -\frac{1}{2}e^{2x^1};$$

$$\Gamma_{12,0} = \Gamma_{21,0} = \frac{1}{2}e^{x^1}.$$

The non-zero second-type Christoffel symbols are

$$\Gamma^0_{10} = \Gamma^0_{01} = 1,\ \Gamma^0_{12} = \Gamma^0_{21} = \frac{1}{2}e^{x^1};$$

$$\Gamma^1_{02} = \Gamma^1_{20} = \frac{1}{2}e^{x^1};\ \Gamma^1_{22} = \frac{1}{2}e^{2x^1};$$

$$\Gamma^2_{01} = \Gamma^2_{10} = -e^{-x^1}.$$

Then

$$\begin{aligned} R_{00} = R^i_{0i0} &= \cancel{\frac{\partial\Gamma^i_{00}}{\partial x^i}} - \cancel{\frac{\partial\Gamma^i_{0i}}{\partial x^0}} + \cancel{\Gamma^i_{si}\Gamma^s_{00}} - \Gamma^i_{s0}\Gamma^s_{0i} = -\Gamma^i_{10}\Gamma^1_{0i} - \Gamma^i_{20}\Gamma^2_{0i} \\ &= -\Gamma^2_{10}\Gamma^1_{02} - \Gamma^1_{20}\Gamma^2_{01} = 1 \end{aligned}$$

$$\begin{aligned} R_{22} = R^i_{2i2} &= \frac{\partial\Gamma^i_{22}}{\partial x^i} - \cancel{\frac{\partial\Gamma^i_{2i}}{\partial x^2}} + \Gamma^i_{si}\Gamma^s_{22} - \Gamma^i_{s2}\Gamma^s_{2i} \\ &= \frac{\partial\Gamma^1_{22}}{\partial x^1} + \Gamma^i_{1i}\Gamma^1_{22} - \Gamma^1_{02}\Gamma^0_{21} - \Gamma^0_{12}\Gamma^1_{20} = e^{2x^1} \end{aligned}$$

$$R_{02} = R_{20} = R^i_{2i0} = \frac{\partial\Gamma^i_{20}}{\partial x^i} - \cancel{\frac{\partial\Gamma^i_{2i}}{\partial x^0}} + \Gamma^i_{si}\Gamma^s_{20} - \Gamma^i_{s0}\Gamma^s_{2i} = \frac{\partial\Gamma^1_{20}}{\partial x^1} + \Gamma^i_{1i}\Gamma^1_{20} = e^{x^1}$$

The others R_{ij} are null. Now,

$$R^0_0 = g^{0s}R_{s0} = g^{00}R_{00} + g^{02}R_{20} = 1,$$

$$R^2_2 = g^{2s}R_{s2} = g^{20}R_{02} + g^{22}R_{22} = 0,$$

i.e. the trace is

$$R = 1.$$

Consider the contravariant vector $u^i := (1, 0, 0, 0) = (u^0, u^1, u^2, u^3)$. The corresponding covariant vector is $u_i := g_{is}u^s = (1, 0, e^{x^1}, 0) = (u_0, u_1, u_2, u_3)$.

Let us observe $R_{00} = 1 = u_0 u_0$; $R_{22} = e^{2x^1} = u_2 u_2$; $R_{02} = R_{20} = e^{x^1} = u_0 u_2$, that is, in our case, for $\Lambda = \frac{1}{2}$ and $T_{ij} = \rho u_i u_j = \frac{1}{8\pi G} u_i u_j$ we have

$$R_{ij} - \cancel{\frac{1}{2} R g_{ij}} + \cancel{\Lambda g_{ij}} = u_i u_j = 8\pi G T_{ij}.$$

Therefore Gödel's metric is a solution for Einstein's field equations when the cosmological constant is $\Lambda = \frac{1}{2}$ and $T_{ij} = \frac{1}{8\pi G} u_i u_j$.

Consider Gödel's change of coordinates $(t, r, \phi, y) \to (x^0, x^1, x^2, x^3)$

$$\begin{cases} x^0 = 2t - \phi\sqrt{2} + 2\sqrt{2} \arctan\left(\tan\left(\frac{\phi}{2}\right) e^{-2r}\right), \phi \neq \pi;\ x^0 = 2t \ if \ \phi = \pi \\ x^1 = \ln\left[\cosh(2r) + \cos\phi \sinh(2r)\right] \\ x^2 = \dfrac{\sqrt{2} \sin\phi \sinh(2r)}{\cosh(2r) + \cos\phi \sinh(2r)} \\ x^3 = 2y. \end{cases}$$

It can be written in the form

$$\begin{cases} \tan\left(\frac{\phi}{2} + \frac{x^0 - 2t}{2\sqrt{2}}\right) = \tan\left(\frac{\phi}{2}\right) e^{-2r} \\ e^{x^1} = \cosh(2r) + \cos\phi \sinh(2r) \\ x^2 e^{x^1} = \sqrt{2} \sin\phi \sinh(2r) \\ x^3 = 2y. \end{cases}$$

Let us look at the coordinates x^1 and x^2 when $r \geq 0$; $0 \leq \phi \leq \pi$. It can be seen a 2π periodicity of x^1 and x^2 when r is fixed. These coordinates can be called cylindrical coordinates for the manifold M. Computing Gödel's metric in the new coordinates, we find another form of the previous solution of the Einstein field equations, that is,

$$ds^2 = 4\left[dt^2 - dr^2 - dy^2 + (\sinh^4 r - \sinh^2 r) d\phi^2 + 2\sqrt{2} \sinh^2 r d\phi dt\right],$$

called second Gödel's second metric.

We do not present the computations because they are heavy to be reported. We leave the calculations as an exercise for the reader. We can prove:

Theorem 10.15.1 *Denote by* $M := \mathbb{R}^4$ *the set having the coordinates* (x^0, x^1, x^2, x^3). *Then*

1. *For any two events A and B there is a transformation on M carrying A into B, that is, there are not privileged points. From the physical point of view, it means that M is homogeneous.*

2. *M has rotational symmetry, i.e. there exists a transformation of coordinates depending on one parameter only such that A is carried into A.*

Proof 1. Consider the transformation

$$\begin{cases} \bar{x}^0 = x^0 + a \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 \end{cases}$$

Let us check that this is an isometry of M. We observe $d\bar{x}^k = dx^k$ and $d\bar{s}^2 = ds^2$. Then, consider two points $A_0(x^0(\lambda_0), x^1(\lambda_0), x^2(\lambda_0), x^3(\lambda_0))$ and $A_1(x^0(\lambda_1), x^1(\lambda_1), x^2(\lambda_1), x^3(\lambda_1))$ joined by the curve $c(\lambda) = (x^0(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda))$, $\lambda \in [\lambda_0, \lambda_1]$ and their images $\bar{A}_0(\bar{x}^0(\lambda_0), \bar{x}^1(\lambda_0), \bar{x}^2(\lambda_0), \bar{x}^3(\lambda_0))$, $\bar{A}_1(\bar{x}^0(\lambda_1), \bar{x}^1(\lambda_1), \bar{x}^2(\lambda_1), \bar{x}^3(\lambda_1))$ joined by the curve $\bar{c}(\lambda) = (\bar{x}^0(\lambda), \bar{x}^1(\lambda), \bar{x}^2(\lambda), \bar{x}^3(\lambda))$, $\lambda \in [\lambda_0, \lambda_1]$. It results in

$$l_c(A_0, A_1) = \int_{\lambda_0}^{\lambda_1} ||\dot{c}(\lambda)|| d\lambda,$$

where the norm is expressed with respect to ds^2. The same,

$$l_{\bar{c}}(\bar{A}_0, \bar{A}_1) = \int_{\lambda_0}^{\lambda_1} ||\dot{\bar{c}}(\lambda)||_1 d\lambda,$$

where this second norm is expressed with respect $d\bar{s}^2$. Since $||\dot{c}(\lambda)|| = ||\dot{\bar{c}}(\lambda)||_1$ it results in $l_c(A_0, A_1) = l_{\bar{c}}(\bar{A}_0, \bar{A}_1)$, that is we deal with an isometry of M. Three other isometries can be highlighted:

$$\begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 + b \\ \bar{x}^2 = e^{-b} x^2 \\ \bar{x}^3 = x^3 \end{cases}, \quad \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 + c \\ \bar{x}^3 = x^3 \end{cases}, \quad \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 + d \end{cases}.$$

Combining all four previous transformations, any point of M can be mapped into any point of M without changing metric properties of M, i.e. any point can be seen as an origin. Therefore M is homogeneous.

2. The previous discussion allows us to consider any point A as the origin of M. Therefore in the new coordinates (t, r, ϕ, y), $r_A = 0$. Consider the group of transformations with respect to the parameter $K \in \mathbb{R}$,

$$(t, r, \phi, y) \to (t, r, \phi + K, y).$$

The point A is a fixed point of this group, and according to the previous observation, there exists a 2π periodicity experienced by any other point. Therefore M allows rotations with respect any given point of it. □

A space-time is time orientable, if the time-like and null vectors can be classified into two classes, the future-pointing and the past-pointing vectors as we did it in the case of the 2-Minkowski space (that is, with respect to a given vector).

Let us remember: in the case of the Minkowski metric $ds^2 = dt^2 - dx^2$, provided by the Minkowski product $< p, q >_M = t_p t_q - x_p x_q$, the vector $e_1 = (1, 0)$ is a time-like vector because $< e_1, e_1 >_M > 0$ and the time-like vector $v = (3, 2)$ becomes a future-pointing time-like vector because $< v, e_1 >_M > 0$. The vector $-v$ becomes a past-pointing time-like vector. The vector $w = (1, 1)$ is a null future-pointing vector, etc.

A curve ψ is called time-like if the tangent vectors $\dot{\psi}$ are time-like future-pointing vectors.

If we choose two events $E_0(\lambda_0)$ and $E_1(\lambda_1)$ connected by a time-like curve, we say that E_0 is in the past of E_1 (or, equivalently, E_1 is in the future of E_0) if $\lambda_0 < \lambda_1$.

In the case of Gödel's second metric, the vector $u^i = (1, 0, 0, 0)$ has the property $g_{ij}u^i u^j = 4 > 0$, that is, u^i is a time-like vector. If v^j is a time-like vector, that is, $g_{ij}v^i v^j > 0$, we say that v^j is future pointing if $g_{ij}u^i v^j > 0$. If $g_{ij}u^i v^j < 0$ the vector v^j is called past pointing. The same, if w^k is a null vector, we can define past pointing and future-pointing null vectors according to the sign of $g_{ij}u^i w^k$. Therefore M becomes time orientable.

Theorem 10.15.2 *The time orientable Gödel's universe allows*

1. *closed time-like curves;*

2. *time-like loops, i.e. any two events connected by a time-like curve can be connected by a closed time-like curve.*

Proof 1. Consider the curve $\alpha(s) := (0, R, bs, 0),\ b \in \mathbb{N}$.
Its velocity vector is $\dot{\alpha}(s) = v^j = (0, 0, b, 0)$. The norm of this vector depends on

$$g_{ij}v^i v^j = (\sinh^4 R - \sinh^2 R)b^2.$$

If we choose, from the beginning $R > \ln(1 + \sqrt{2})$, i.e. $\sinh R > 1$, this vector is a time-like one. The chosen curve is a time-like curve according to the second statement of the previous theorem and $\alpha(0) = \alpha(2\pi)$. We have obtained a closed time-like curve in M.

2. First, if we look only at the coordinates (r, ϕ, t), we observe that they determine completely the coordinates (x^1, x^2, x^3). More precisely, for particular given (r, ϕ, t), we have particular corresponding (x^1, x^2, x^3). It remains the coordinate x^0 which depends on t; therefore the $t-$lines of matter, in cylindrical coordinates, are x^0-lines of matter.

Consider the point B_{t_1} with the coordinates $(t_1, R, 0, 0)$. The curve

$$\gamma(s) = \left(t_2 + \frac{t_1 - t_2}{2\pi n}s, R, bs, 0\right)$$

is time-like because the vector

$$\dot{\gamma}(s) = v^i = \left(\frac{t_1 - t_2}{2\pi n}, 0, b, 0\right)$$

is time-like. Indeed, for a chosen n, big enough, and $R > \ln(1 + \sqrt{2})$, we have

$$g_{ij}v^i v^j = 4\left[\frac{(t_1 - t_2)^2}{4\pi^2 n^2} + (\sinh^4 R - \sinh^2 R)b^2 - \sinh^2 R \cdot \left(\frac{t_1 - t_2}{2\pi n}\right) b\right] > 0.$$

We observe that $\gamma(0) = B_{t_2}$ and $\gamma(2n\pi) = B_{t_1}$.
Using the same idea, we can derive time-like curves between B_{t_2} and B_{t_3} and between B_{t_3} and B_{t_1}. The concatenation of the three time-like curves is a time-like loop starting from B_{t_1}, passing through B_{t_2}, then to B_{t_3}, to finally reach B_{t_1}. We can conclude that t is not a proper time coordinate, because if it is so, moving forward in time we return in our past. Therefore no global time coordinate exists in Gödel's universe. □

More about this very nice and difficult subject can be found in [108, 113, 176]. An exhaustive discussion on closed time-like curves and their physical implications can be found in [103]. Here we developed this Universe model because it can be easily framed in our geometric picture.

10.16 Is it Possible a Space-Time without Matter and Time?

Let us start from the metric [32]

$$ds^2 = (dx^0)^2 + (dx^1)^2 + (dx^2)^2 - (dx^3)^2.$$

This is a Minkowski (3, 1) signature metric. Let us denote by M the set involving the coordinates (x^0, x^1, x^2, x^3). We are dealing with a metric derived from

$$R_{ij} = 0.$$

So, we have a space-time without matter.

Let us consider a simple change of coordinates $(t, r, \phi, y) \to (x^0, x^1, x^2, x^3)$:

$$\begin{cases} x^0 = t + y, \ t, y \in \mathbf{R}, \\ x^1 = r \sin\phi, \ r > 0, \ \phi \in \mathbf{R}, \\ x^2 = r\cos\phi, \\ x^3 = 2y + 2r. \end{cases}$$

The new metric obtained after considering

$$\begin{cases} dx^0 = dt + dy \\ dx^1 = \sin\phi \, dr + r\cos\phi \, d\phi \\ dx^2 = \cos\phi \, dr - r\sin\phi \, d\phi \\ dx^3 = 2dy + 2dr \end{cases}$$

is

$$d\bar{s}^2 = dt^2 - 3dr^2 + r^2 d\phi^2 - 3dy^2 + 2dtdy - 8drdy.$$

Denote by $\bar{M}$ the new space having the coordinates (t, r, ϕ, y). If we look at the coordinates x^1 and x^2 when $r > 0, \ \phi \in \mathbf{R}$, it can be seen a 2π periodicity of x^1 and x^2 for r is fixed. These are the cylindrical coordinates for the initial set $M(x^0, x^1, x^2, x^3)$. The points (t, r, ϕ, y) and $(t, r, \phi + 2k\pi, y), \ k \in \mathbf{Z}$ in $\bar{M}$ describe the same point (x^0, x^1, x^2, x^3) in M. We can prove, for the initial set of coordinates denoted by M, the following

Theorem 10.16.1 *(i) For any two events A and B in M there exists a transformation of M carrying A into B, that is there is no privileged point. From the physical point of view, it means that M is homogeneous.*

(ii) M has rotational symmetry, i.e. there exists a transformation of coordinates depending on one parameter only such that A is carried into A.

Proof (i) The transformations of coordinates below

$$\begin{cases} \bar{x}^0 = x^0 + a \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 + b \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 + c \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 + d \end{cases}.$$

are obviously isometries of M because they preserve the metric, i.e. $ds^2 = d\bar{s}^2$. Combining the four previous transformations, any point of M can be mapped into any point of M without changing metric properties of M, i.e. any point can be seen as an origin. There is no privileged point. Therefore M is homogeneous.

(ii) Consider the group of transformations with respect to the parameter $b \in \mathbb{R}$,

$$(t, r, \phi, y) \to (t, r, \phi + b, y)$$

and apply it to any point of M. The 2π periodicity shows that M allows rotations with respect to any given point of it. □

As we know, a space-time is time orientable if the time-like vectors can be classified into two classes, the future- and the past-pointing vectors. If the definition of time-like vectors in this case is something classic, only a well-chosen time-like vector e allows the classification in the two classes.

In the case of our second metric with coefficients denoted here by $\bar{g}_{ij}$, the vector $e = (e^0, e^1, e^2, e^3) = (1, 0, 1, 0)$ has the property $\bar{g}_{ij}(t, r, \phi, y)e^i e^j = 1 + r^2 > 0$, that is, e is a time-like vector. If $v = (v^0, v^1, v^2, v^3)$ is a time-like vector, i.e. $\bar{g}_{ij}(t, r, \phi, y)v^i v^j > 0$ we say that v is future pointing if $\bar{g}_{ij}(t, r, \phi, y)e^i v^j > 0$. If $\bar{g}_{ij}(t, r, \phi, y)e^i v^j < 0$, the vector v is called past pointing. Let us observe that if v is time-like and future pointing then $-v$ is still time-like but past pointing. It is easy to see that $-e$ is past pointing because $\bar{g}_{ij}(t, r, \phi, y)e^i(-e^j) = -1 - r^2 < 0$. Therefore $\bar{M}$ becomes time orientable.

Theorem 10.16.2 *The time orientable set $\bar{M}$ allows time-like loops.*

Proof Consider the curve $\alpha(s) := (0, R, 2ns, 0)$.
Its velocity vector is $\dot{\alpha}(s) = (v^0, v^1, v^2, v^3) = (0, 0, 2n, 0)$. We have at each point of this curve that

$$d\bar{s}^2(\dot{\alpha}(s), \dot{\alpha}(s)) = \bar{g}_{ij}(0, R, 2ns, 0)v^i v^j = 4n^2R^2 > 0,$$

i.e. this vector is a time-like one. More,

$$d\bar{s}^2(\dot{\alpha}(s), e) = \bar{g}_{ij}(0, R, 2ns, 0)v^i e^j = 2nR^2 > 0,$$

that is, the vector $\dot{\alpha}(s)$ is future pointing. At the same time, $\alpha(0)$ and $\alpha(2\pi)$ have the same image in M, therefore the image of the curve α in M is starting from a point of M to return at the same point. We have obtained a loop in M.

We have to observe that if we choose any point B of this loop, the first arch $AB \subset M$ is time oriented by the corresponding arch of α in $\bar{M}$, therefore the event B is in the future of the event A, while the arch $BA \subset M$, using the same reasoning, makes the event A to be in the future of B. We can say, in simple words, that going towards the future, we return to the past. □

Theorem 10.16.3 *(i) Any two points induced in M by the coordinates $(t_1, R, 0, 0)$ and $(t_2, R, 0, 0)$ of $\bar{M}$ (such that $R^2 > \frac{1}{2\pi}|t_2 - t_1|$) can be joined by a time-like curve such that the second point is in the future of the initial first point.*

(ii) M allows time-like closed curves.

Proof (i) First, if we look only at the coordinates (r, ϕ, y), we observe that they determine completely the coordinates (x^1, x^2, x^3). It remains the coordinate x^0 which depends on t, therefore the $t-$lines of matter in cylindrical coordinates are

in fact x^0- lines of matter. This observation allows to consider the x^0 coordinate as a time coordinate in the first metric. Specifically, the initial metric has signature $(+, +, +, -)$ with the time type coordinate followed by $(+, +, -)$. It is definitely different by the Lorentz-type metric having signature $(+, -, -, -)$ whose spatial part has three components with the same sign minus.
Therefore, let us consider the point B_{t_1}, $B_{t_2} \in \bar{M}$ with the coordinates $(t_1, R, 0, 0)$, $(t_1, R, 0, 0)$, respectively. We proceed as we did it in the previous theorem. The curve

$$\gamma(s) = \left(t_1 + \frac{t_2 - t_1}{2\pi}s, R, s, 0\right) \subset \bar{M}$$

is a time-like curve because the vector

$$\dot{\gamma}(s) = (w^0, w^1, w^2, w^3) = \left(\frac{t_2 - t_1}{2\pi}, 0, 1, 0\right)$$

has the property

$$d\bar{s}^2(\dot{\gamma}(s), \dot{\gamma}(s)) = \bar{g}_{ij} w^i w^j = \left(\frac{t_2 - t_1}{2\pi}\right)^2 + R^2 > 0.$$

Furthermore,

$$d\bar{s}^2(\dot{\gamma}(s), e) = \bar{g}_{ij} w^i e^j = \frac{t_2 - t_1}{2\pi} + R^2.$$

The last expression is positive for the condition of theorem, that is, $\dot{\gamma}(s)$ is pointing the future. We observe that $\gamma(0) = B_{t_1}$ and $\gamma(2\pi)$ and B_{t_2} have the same image in M. Therefore, for the images in M, i.e. E_1 and E_2, we can say that the event E_2 occurs after the event E_1.

(ii) Using the previous idea we can generate the same time-like curves between B_{t_2} and B_{t_3} and between B_{t_3} and B_{t_1}. The concatenation of the three time-like curves is the time-like "chain-curve" starting from B_{t_1}, passing through B_{t_2}, then to B_{t_3}, to finally reach B_{t_1} again. Taking into consideration how events occur in relation to the future time-like tangent vectors of the curves, we can conclude that neither t nor x^0 can be proper time coordinates, because if it is so, moving forward in time would mean to return in our past. Therefore no global time coordinate exists in this model. So we have derived a universe without both time and without matter. □

The Case of Planar Gravitational Waves in this Space-time

The issue of gravitational waves can be dealt in the context of the space-time without matter and time. Again, the power of coordinates change together with the convenient physical interpretation is the key for the desired result. Let us consider the initial metric [32]

$$ds^2 = (dx^0)^2 + (dx^1)^2 + (dx^2)^2 - (dx^3)^2 \,,$$

and the transformation of coordinates

$$x^0 = t + \sin(t - z),\ x^1 = x,\ x^2 = y,\ x^3 = z.$$

Since

$$dx^0 = (1 + \cos(t - z))dt - \cos(t - z)dz,\ dx^1 = dx,\ dx^2 = dy,\ dx^3 = dz,$$

the initial metric turns into the metric

$$ds^2 = (1 + \cos(t - z)[2 + \cos(t - z)])dt^2 + dx^2 + dy^2 - (1 - \cos^2(t - z))dz^2$$

$$- 2\cos(t - z)(1 + \cos(t - z))dtdz.$$

We observe that the metric can be arranged in the form

$$ds^2 = g_{ij}dx^i dx^j = (\alpha_{ij} + h_{ij})dx^i dx^j,$$

where the $\alpha_{00} = \alpha_{11} = \alpha_{22} = 1,\ \alpha_{33} = -1, \alpha_{ij} = 0$ if $i \neq j$.

The non-zero coefficients are

$$g_{00} = 1 + \cos(t - z)(2 + \cos(t - z));\ g_{11} = 1;\ g_{22} = 1;\ g_{33} = -1 + \cos^2(t - z);$$

$$g_{03} = g_{30} = -\cos(t - z)(1 + \cos(t - z));$$

therefore the non-zero "perturbations" which appear are

$$h_{00} = \cos(t - z)\ (2 + \cos(t - z));\ h_{33} = -\cos^2(t - z);$$

$$h_{03} = h_{30} = -\cos(t - z)(1 + \cos(t - z)).$$

According to the theory in [171], the metric

$$ds^2 = g_{ij}dx^i dx^j = (\alpha_{ij} + h_{ij})dx^i dx^j$$

describes planar gravitational waves if it satisfies both the Einstein field equations (in our case $R_{ij} = 0$) and the wave equation, derived from them,

$$\Box\, h_{ij} = 0$$

for each perturbation involved in the metric coefficients. We remember that the wave equation is written with respect to the d'Alembert operator here in the form induced by the metric coefficients,

$$\Box := (\partial^0)^2 + (\partial^1)^2 + (\partial^2)^2 - (\partial^3)^2,$$

where

$$(\partial^k)^2 := \frac{\partial^2}{(\partial x^k)^2}.$$

It is easy to see that

$$\Box \ \cos(t-z) = (\partial^0)^2 \cos(t-z) - (\partial^3)^2 \cos(t-z) = 0$$

and

$$\Box \ \cos^2(t-z) = (\partial^0)^2 \cos^2(t-z) - (\partial^3)^2 \cos^2(t-z) = 0,$$

therefore the conditions

$$\Box \, h_{ij} = 0$$

are fulfilled. Our space-time without matter and time has a texture describing planar gravitational waves. This section shows again the power of the coordinate change when we model physical aspects.

10.17 A Remarkable Universe without Time

A natural development of the previous considerations is the following. Let us consider the metric

$$ds^2 = f(x^3)(dx^0)^2 + (dx^1)^2 + (dx^2)^2 - (dx^3)^2$$

for an appropriate function f. The above computations show that the only non-zero Christoffel symbols are

$$\Gamma^0_{03} = \Gamma^0_{30} = \frac{1}{2}\frac{f'(x^3)}{f(x^3)}; \ \Gamma^3_{00} = \frac{1}{2} f'(x^3),$$

therefore the only non-zero Ricci tensor components are

$$R_{00} = \frac{1}{2} f"(x^3) - \frac{1}{4}\frac{\left(f'(x^3)\right)^2}{f(x^3)}; \ R_{33} = -\frac{1}{4}\frac{\left(f'(x^3)\right)^2}{f^2(x^3)}.$$

Using the Gödel idea to adopt exponential functions, if we choose $f(x^3) = e^{x^3}$, we obtain

$$R_{00} = \frac{1}{4} e^{x^3}; \ R_{33} = -\frac{1}{4},$$

that is

$$R = \frac{1}{2}.$$

The Einstein field equations are

$$R_{ij} - \frac{1}{2} R\, g_{ij} = 8\pi G\, T_{ij}$$

where

$$T_{ij} = \frac{1}{8\pi G} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{4} & 0 & 0 \\ 0 & 0 & -\frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Therefore the metric

$$ds^2 = e^{x^3}(dx^0)^2 + (dx^1)^2 + (dx^2)^2 - (dx^3)^2$$

satisfies Einstein's field equations on $M = \mathbf{R}^4$.

The change of coordinates $(t, r, \phi, y) \to (x^0, x^1, x^2, x^3)$:

$$\begin{cases} x^0 = t, \ t \in \mathbf{R}, \\ x^1 = r \sin\phi, \ r > 0, \ \phi \in \mathbf{R}, \\ x^2 = r\cos\phi, \\ x^3 = y, \ y \in \mathbf{R} \end{cases}$$

transforms the initial metric into the metric

$$d\bar{s}^2 = e^y dt^2 + dr^2 + r^2 d\phi^2 - dy^2$$

on the set $\bar{M}$ described by the new coordinates.
In this case, it is important to split the time-like vectors in two classes.

Denote the second metric coefficients by $\bar{g}_{ij}$. The vector $e = (e^0, e^1, e^2, e^3) = (1, 1, 1, 0)$ has the property $\bar{g}_{ij} e^i e^j = e^y + 1 + r^2 > 0$, that is e is a time-like vector. If $v = (v^0, v^1, v^2, v^3)$ is a time-like vector, i.e. $\bar{g}_{ij} v^i v^j > 0$, we say that v is future pointing for $\bar{g}_{ij} e^i v^j > 0$. If $\bar{g}_{ij} e^i v^j < 0$, the vector v is past pointing. Let us observe that if v is time-like and future pointing, then $-v$ is still time-like but past pointing. It is easy to see that $-e$ is past pointing because $\bar{g}_{ij} e^i(-e^j) = -e^y - 1 - r^2 < 0$. This way $\bar{M}$ becomes time orientable.

All results stated in the three theorems of the previous section remain valid. We have also to mention the form of isometries of M in this case:

$$\begin{cases} \bar{x}^0 = x^0 + a \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 + b \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 + c \\ \bar{x}^3 = x^3 \end{cases}, \begin{cases} \bar{x}^0 = e^{-d/2} x^0 \\ \bar{x}^1 = x^1 \\ \bar{x}^2 = x^2 \\ \bar{x}^3 = x^3 + d. \end{cases}.$$

Therefore we described a universe without a proper global time coordinate. If we look at the chain of curves, we understand that the "history" can change for someone who travels between two events in the way the observer decides. It is easy to see that chain curves do not satisfy the geodesic equations in M. Therefore a chain curve, or only a part of it, is not an imposed geometric trajectory of M, it is completed by the "desire" of a moving point to follow the chosen trajectory. It remains only to think if it is acceptable to say that at each point B_t this universe splits in an infinity of similar universes offering the superposition of all possible futures to a given inhabitant of it.

10.18 Another Exact Solution of Einstein Field Equations Induced by the Gödel One

In this section we present an solution of Einstein field equations generated by a particular form of matter–energy tensor. This solution is related to an attempt to find another Gödel-type metric [32].

Let us keep in mind the Gödel first metric description in a more general form

$$ds^2 = (dx^0)^2 - (dx^1)^2 + f(x^1)(dx^2)^2 - (dx^3)^2 + 2g(x^1)dx^0dx^2.$$

The coefficients, after cancelling x^1, involved in computations are

$$(g_{ij}) = \begin{pmatrix} 1 & 0 & g & 0 \\ 0 & -1 & 0 & 0 \\ g & 0 & f & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (g^{ij}) = \begin{pmatrix} \dfrac{f}{f-g^2} & 0 & \dfrac{-g}{f-g^2} & 0 \\ 0 & -1 & 0 & 0 \\ \dfrac{-g}{f-g^2} & 0 & \dfrac{1}{f-g^2} & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

We find the non-zero Ricci tensor coefficients:

$$R_{00} = \frac{1}{2}\frac{(g')^2}{g^2 - f}$$

$$R_{22} = \frac{1}{2}f'' - \frac{1}{4}\frac{gg'f'}{g^2 - f} + \frac{1}{2}\frac{(g')^2 f}{g^2 - f}$$

$$R_{02} = R_{20} = \frac{1}{2}g'' + \frac{1}{4}\frac{f'g'}{g^2 - f},$$

where f', f'' and g', g'' are the first and the second derivatives with respect to x^1 of the functions f, respectively, g.

Of course, if $f(x^1) = \frac{1}{2}e^{2x^1}$ and $g(x^1) = e^{x^1}$, we obtain exactly the results obtained in the case of Gödel first metric, i.e.

$$R_{00} = 1$$

$$R_{22} = e^{2x^1}$$

$$R_{02} = R_{20} = e^{x^1}$$

and then $R = 1$.

The issue is to find some other coefficients such that the Einstein field equations are satisfied, in particular using polynomials instead of exponential functions. If we choose $f(x^1) = \frac{1}{2}(x^1)^2$ and $g(x^1) = x^1$, we obtain

$$R_{00} = \frac{1}{(x^1)^2}$$

$$R_{22} = \frac{1}{2}$$

$$R_{02} = R_{20} = \frac{1}{2x^1},$$

therefore

$$R = R^i_i = R^0_0 + R^2_2 = g^{0s}R_{s0} + g^{2s}R_{s2} = g^{00}R_{00} + g^{02}R_{20} + g^{20}R_{02} + g^{22}R_{22} = 0.$$

The Einstein field equations

$$R_{ij} - \frac{1}{2}Rg_{ij} + \Lambda g_{ij} = 8\pi G T_{ij}$$

are satisfied for $\Lambda = 0$ and the "artificial matter" created by using the covariant vector

$$u := \left(\frac{1}{x^1}, 0, \frac{1}{2}, 0\right) = (u_0, u_1, u_2, u_3)$$

after the rule

$$T_{ij} = \rho\, a_{ij} u_i u_j, \ \ \rho = \frac{1}{8\pi G},$$

where the matrix (a_{ij}) can be chosen with the form

$$(a_{ij}) = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & a & 0 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & b \end{pmatrix} \text{ where } a \text{ and } b \text{ are real parameters.}$$

Even if this solution cannot highlight another Gödel-type universe, it is important because it shows us how we can create an artificial matter in a form such that the Einstein field equations are satisfied.

10.19 The Wormhole Solutions

In 1973, two interesting papers on traversable wormholes appeared. We are talking about results by Ellis [93] and Bronnikov [39], who independently studied possible wormhole solutions of the Einstein field equations. We present here the conclusions of these papers to give an idea of wormhole solutions.

Consider the metric

$$ds^2 = c^2dt^2 - dr^2 - (r^2 + a^2)\, d\theta^2 - (r^2 + a^2)\sin^2\theta\, d\phi^2,$$

where a is a real positive parameter.

Denoting $t := x^0,\ r := x^1,\ \theta := x^2$ and $\phi := x^3$, the form of the metric is

$$ds^2 = c^2(dx^0)^2 - (dx^1)^2 - ((x^1)^2 + a^2)(dx^2)^2 - ((x^1)^2 + a^2)\sin^2 x^2 (dx^3)^2$$

and we are ready for necessary computations. Before starting, let us observe that if $a = 0$ the metric is

$$ds^2 = c^2dt^2 - dr^2 - r^2\, d\theta^2 - r^2\sin^2\theta\, d\phi^2,$$

i.e. it is the Minkowski metric

$$ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$$

written in spherical coordinates,

$$\begin{cases} x = r\sin\theta\cos\phi \\ y = r\sin\theta\sin\phi \\ z = r\cos\theta. \end{cases}$$

We are interested only by the case $a \neq 0$. The only non-zero Christoffel symbols are

$$\Gamma^1_{22} = -x^1;\ \Gamma^1_{33} = -x^1 \sin^2 x^2;\ \Gamma^2_{12} = \Gamma^2_{21} = \frac{x^1}{(x^1)^2 + a^2};\ \Gamma^2_{33} = -\sin x^2 \cos x^2;$$

$$\Gamma^3_{23} = \Gamma^3_{32} = \cot x^2;\ \Gamma^3_{13} = \Gamma^3_{31} = \frac{x^1}{(x^1)^2 + a^2}.$$

It is easy to see that $R_{00} = 0$. Then, for $R_{11} = R^i_{1i1}$, it is

$$R_{11} = \frac{\partial \Gamma^i_{11}}{\partial x^i} - \frac{\partial \Gamma^1_{1i}}{\partial x^1} + \Gamma^i_{si}\Gamma^s_{11} - \Gamma^i_{s1}\Gamma^s_{1i} = -\frac{\partial \Gamma^2_{12}}{\partial x^1} - \frac{\partial \Gamma^3_{13}}{\partial x^1} - \Gamma^2_{21}\Gamma^2_{12} - \Gamma^3_{31}\Gamma^3_{13}$$
$$= \frac{-2a^2}{((x^1)^2 + a^2)^2}.$$

Similar computations lead to

$$R_{22} = R_{33} = 0.$$

Now, let us represent the results in the more suggestive way given by the first coordinates used to write the metric. So, for

$$ds^2 = c^2 dt^2 - dr^2 - (r^2 + a^2)\, d\theta^2 - (r^2 + a^2) \sin^2\theta\, d\phi^2$$

the only non-zero Christoffel symbols are

$$\Gamma^r_{\theta\theta} = -r;\ \Gamma^r_{\phi\phi} = -r \sin^2\theta;\ \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} = \frac{r}{r^2 + a^2};\ \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta;$$

$$\Gamma^\phi_{\theta\phi} = \Gamma^\phi_{\phi\theta} = \cot\theta;\ \Gamma^\phi_{r\phi} = \Gamma^\phi_{\phi r} = \frac{r}{r^2 + a^2}$$

and

$$R_{tt} = R_{\theta\theta} = R_{\phi\phi} = 0 \text{ and } R_{rr} = \frac{-2a^2}{(r^2 + a^2)^2}.$$

It results in

$$R = R_{rr} = \frac{-2a^2}{(r^2 + a^2)^2}.$$

Therefore, the Einstein field equations

$$R_{ij} - \frac{1}{2} R g_{ij} = \frac{8\pi G}{c^4} T_{ij}$$

lead to

$$T_{tt} = \frac{a^2 c^6}{8\pi G} \frac{1}{(r^2 + a^2)^2}.$$

In the same way we obtain

$$T_{rr} = \frac{-3a^2 c^4}{8\pi G} \frac{1}{(r^2 + a^2)^2},$$

$$T_{\theta\theta} = \frac{-a^2 c^4}{8\pi G} \frac{1}{r^2 + a^2},$$

$$T_{\phi\phi} = \frac{-a^2 c^4}{8\pi G} \frac{\sin^2\theta}{r^2 + a^2}.$$

Ellis imagined the Einstein field equations written in the form

$$R_{ij} - \frac{1}{2} R g_{ij} = -2 \left(\frac{\partial \Psi}{\partial x^i} \frac{\partial \Psi}{\partial x^j} - \frac{1}{2} \sum_{k=0}^{3} \left(\frac{\partial \Psi}{\partial x^k} \right)^2 g_{ij} \right) = \frac{8\pi G}{c^4} T_{ij},$$

with respect to a function $\Psi = \Psi(x^0, x^1, x^2, x^3)$. He observed that, in the case of his metric, the function has one variable only, i.e. $\Psi = \Psi(r)$, and has to satisfy the equation

$$\frac{\partial \Psi}{\partial r} = \frac{a}{r^2 + a^2}.$$

Such a function can be

$$\Psi(r) = \arctan \frac{r}{a}.$$

The idea to write the stress–energy tensor with respect to a scalar function $\Psi = \Psi(r)$ is the key point of this result.

Before continuing, let us present the revolution surface called catenoid. This surface was discovered by Euler and it is arising by rotating a catenary curve (see Problem 8.9.4) around an axis. We consider the parameterization

$$\begin{cases} x = l \cosh \dfrac{x^1}{l} \cos x^2 \\ y = l \cosh \dfrac{x^1}{l} \sin x^2 \\ z = x^1. \end{cases}$$

The metric induced by the Euclidean $3D$ metric is

$$ds^2 = \cosh^2 \frac{x^1}{l} (dx^1)^2 + l^2 \cosh^2 \frac{x^1}{l} (dx^2)^2.$$

The mirror metric, corresponding to $t = 0$; $\theta = \dfrac{\pi}{2}$; $a = l$ and

$$ds^2 = c^2 dt^2 - dr^2 - (r^2 + a^2)\, d\theta^2 - (r^2 + a^2)\sin^2\theta\, d\phi^2,$$

is

$$d\bar{s}^2 = dr^2 + (r^2 + l^2) d\phi^2.$$

A change of coordinates,

$$\begin{cases} r = l \sinh \dfrac{x^1}{l} \\ \phi = x^2 \end{cases}$$

switches the metric $d\bar{s}^2$ into the catenoid metric

$$ds^2 = \cosh^2 \frac{x^1}{l} (dx^1)^2 + l^2 \cosh^2 \frac{x^1}{l} (dx^2)^2.$$

Now, imagine the catenoid with his throat surrounded by the circle

$$x^2 + y^2 = l^2$$

obtained in our parameterization for $x^1 = 0$. The wormhole throat is stable because there is no evolution of it in the considered choice $\left(t = 0 \text{ and } \theta = \dfrac{\pi}{2}\right)$.

Let us now take into account the geodesics of the metric

$$d\bar{s}^2 = dr^2 + (r^2 + l^2) d\phi^2 \,.$$

The geodesic equations are

$$\begin{cases} \ddot{r} - r\dot{\phi}^2 = 0 \\ \ddot{\phi} + \dfrac{2r}{r^2 + l^2}\dot{r}\dot{\phi} = 0 \end{cases}$$

The second equation leads to

$$\dot{\phi} = \frac{k}{r^2 + l^2},$$

where k is a constant. Let us observe that the derivative of

$$(\dot{r})^2 + \frac{k^2}{r^2 + l^2} = 1,$$

taking into account the above equation,

$$\dot{\phi} = \frac{k}{r^2 + l^2},$$

is exactly the first equation

$$\ddot{r} - r\dot{\phi}^2 = 0.$$

There are two possible cases.

In the first case, the condition $\dot{r} \equiv 0$ leads to a constant k,

$$k^2 = r^2 + l^2 \geq l^2.$$

at the same time $\dot{\phi} = \frac{1}{k}$, i.e.

$$\phi(s) = \frac{1}{k}s + k_1 \text{ and } r(s) = \sqrt{k^2 - l^2}.$$

Looking at the first equation we see that only $k = l$ can be considered in order to solve it.

The second case is obtained if $\dot{r}$ is not identically null. We combine the two equations, written now in the form,

$$\begin{cases} \left(\frac{dr}{ds}\right)^2 + \frac{k^2}{r^2 + l^2} = 1 \\ \frac{d\phi}{ds} = \frac{k}{r^2 + l^2} \end{cases}$$

into the equation

$$\left(\frac{d\phi}{dr}\right)^2 = \frac{k^2}{(r^2 + l^2)(r^2 + l^2 - k^2)}.$$

If $k^2 > l^2$, the solution is related to the condition $r^2 \geq k^2 - l^2 > 0$, therefore the geodesic is included in one of the two parts of the throat.

If $k^2 = l^2$, we have again $r = 0$ separating the parts where a geodesic is included.

If $k^2 < l^2$, the geodesic can traverse the wormhole.

It is easy to draw numerically the geodesics corresponding to the described situations.

To conclude, we can say that this is just a simple prototype of wormhole solutions. Interested reader can consider the so-called Morris–Thorne wormhole [143] where traversability and stability conditions can be related to the energy conditions and the presence of exotic matter. A detailed discussion of the problem is reported in the book by Frolov and Novikov [103]. It is also possible to derive wormhole solutions where geometric contributions stabilize and make the structures traversable [59]. Also if wormholes have not been observed yet, this is a very important research area. See, for example [84], for a general discussion on the topic in Metric-Affine Theories of Gravity.

Chapter 11
A Geometric Realization of Relativity: The de Sitter Space-time

Rem tene, verba sequentur.

Cato

We want to conclude this book considering a gravity theory without masses which can be constructed in Minkowski spaces using a geometric Minkowski potential. From the point of view of this book, this can be considered a full geometric realization of the relativistic approach. The affine space-like spheres can be seen as the regions of the Minkowski space-like vectors characterized by a constant Minkowski gravitational potential. They highlight, for each dimension $n \geq 3$, a model of space-time, the de Sitter one, which satisfies Einstein's field equations in the absence of matter, and it is now intuitive why. This chapter is based on results that can be found in [28, 31, 82, 112, 173].

11.1 About the Minkowski Geometric Gravitational Force

Denote by $\mathbb{M}^n$ the Minkowski n-dimensional space, $n \geq 3$, endowed by the Minkowski product

$$\langle a, b \rangle_M := a_0 b_0 - \sum_{\alpha=1}^{n-1} a_\alpha b_\alpha$$

With respect to given $b = (b_0, b_1, ..., b_{n-1})$, we consider all vectors $x = (x_0, x_1, ..., x_{n-1})$ such that $x - b$ is a space-like vector, that is $\langle x - b, x - b \rangle_M < 0$. We denote by

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*, UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_11

$$r := \sqrt{-(x_0-b_0)^2+\sum_{\alpha=1}^{n-1}(x_\alpha-b_\alpha)^2}$$

the Minkowski "length" of the space-like vector $x-b$ and by

$$u=-\frac{1}{r}(x_0-b_0,x_1-b_1,...,x_{n-1}-b_{n-1})$$

the unit vector of $b-x$. We can define the *Minkowski geometric gravitational* force as

$$F_M^n := \frac{1}{n-1}\,\frac{1}{r^{n-1}}\,u.$$

If

$$A_M^n := \frac{n-2}{r^{n-1}}\,u$$

is by definition the *Minkowski geometric gravitational field*, we have the following *"Minkowski-Newton second principle"*:

$$F_M^n=\frac{1}{(n-1)(n-2)}A_M^n.$$

Let us define the *Minkowski gradient* and the *Minkowski Laplacian*:

$$\nabla_M := \left(-\frac{\partial}{\partial x_0},\frac{\partial}{\partial x_1},...,\frac{\partial}{\partial x_{n-1}}\right)$$

$$\nabla_M^2 := \langle\nabla_M,\nabla_M\rangle_M=\frac{\partial^2}{\partial x_0^2}-\frac{\partial^2}{\partial x_1^2}-...-\frac{\partial^2}{\partial x_{n-1}^2}.$$

For each dimension n, we can define the *Minkowski gravitational potential*

$$\Phi_M^n := -\frac{1}{r^{n-2}}.$$

The following computations

$$\frac{\partial\Phi_M^n}{\partial x_0}=(2-n)\frac{x_0-b_0}{r^n};\quad \frac{\partial\Phi_M^n}{\partial x_\alpha}=(n-2)\frac{x_\alpha-b_\alpha}{r^n},\ \alpha\in\{1,2,...,n-1\};$$

$$\frac{\partial^2\Phi_M^n}{\partial x_0^2}=(2-n)\frac{r^2+n(x_0-b_0)^2}{r^{n+2}};\quad \frac{\partial^2\Phi_M^n}{\partial x_\alpha^2}=(n-2)\frac{r^2-n(x_\alpha-b_\alpha)^2}{r^{n+2}},\ \alpha\in\{1,2,...,n-1\}.$$

lead us to the following two theorems.

Theorem 11.1.1 *The Minkowski gradient of the Minkowski gravitational potential is the opposite of the Minkowski gravitational field.*

Proof It is easy to check the equality $\nabla_M \Phi_M^n = -A_M^n$. □

Theorem 11.1.2 *The Minkowski Laplacian of the Minkowski gravitational potential is null.*

Proof The same, it is easy to check $\nabla_M^2 \Phi_M^n = 0$. □

The last relation is the equation of the Minkowski geometric gravitational field.

In the case when b is the origin of the Minkowski space, the Minkowski unitary space-like sphere can be thought as the set of points of $\mathbb{M}^n$ described by the constant gravitational Minkowski potential $\Phi_M^n = -1$. In this theory, at each dimension n, the Minkowski geometrical gravitational force and the Minkowski geometric gravitational field have the physical dimension $\frac{1}{(l)^{n-1}}$. The Minkowski gravitational potential has the physical dimension $\frac{1}{(l)^{n-2}}$ where (l) is a length.

We may conclude: for each dimension, in the Minkowski space-like vectors region, a natural geometric Minkowski gravity appears in the absence of matter. An equivalent of the Newton gravity theory can be constructed starting from the Minkowski geometric gravitational potential. The affine space-like spheres can be seen as the regions of the Minkowski space-like vectors characterized by a constant Minkowski gravitational potential. They highlight, at each dimension $n \geq 3$, a model of space-time, the de Sitter one, which satisfies Einstein's field equations in the absence of matter, and it is now intuitive why.

11.2 De Sitter Spacetime and Its Cosmological Constant

In the case $n = 3$, we choose to represent the 2-surface as

$$X_0^2 - X_1^2 - X_2^2 = -a^2,$$

in the form $f : \mathbb{R} \times (-\pi, \pi) \longrightarrow \mathbb{M}^3$,

$$f(t, x_1) = (a \sinh t, a \cosh t \cos x_1, a \cosh t \sin x_1).$$

Some computations lead to the metric

$$ds_2^2 = a^2 dt^2 - a^2 \cosh^2 t \, dx_1^2.$$

The non-zero Christoffel symbols are

$$\Gamma^1_{01} = \Gamma^1_{10} = \tanh t,\ \ \Gamma^0_{11} = \cosh t \sinh t$$

and

$$R^0_{101} = \frac{\partial \Gamma^0_{11}}{\partial t} - \frac{\partial \Gamma^0_{10}}{\partial x_1} + \Gamma^0_{s0}\Gamma^s_{11} - \Gamma^0_{s1}\Gamma^s_{10} = \cosh^2 t.$$

It results $R_{0101} = g_{00}R^0_{101} = a^2\cosh^2 t$, that is $K^M_f = -\dfrac{1}{a^2}$.

For this 2-de Sitter space-time, according to the Einstein theorem for surfaces $R_{ij} = K^M_f\, g_{ij}$, we have

$$R_{ij} + \frac{1}{a^2}\, g_{ij} = 0.$$

This last equation can be written also as $R_{ij} - \dfrac{1}{2} R\, g_{ij} = 0$, that is $\Lambda = 0$ and $T_{ij} = 0$.

In the case $n = 4$, the 3-de Sitter space-time is the Minkowski time-like sphere of $\mathbb{M}^4$ given by the equation

$$X_0^2 - X_1^2 - X_2^2 - X_3^2 = -a^2.$$

The standard parameterization is

$$f(t, x_1, x_2) = (a \sinh t \cos x_2, a \cosh t \cos x_1 \cos x_2, a \cosh t \sin x_1 \cos x_2, a \sin x_2).$$

The metric is

$$ds_3^2 = a^2\cos^2 x_2\, dt^2 - a^2\cosh^2 t \cos^2 x_2\, dx_1^2 - a^2 dx_2^2.$$

We observe

$$ds_3^2 = \cos^2 x_2 (\, a^2 dt^2 - a^2\cosh^2 t\ dx_1^2) - a^2 dx_2^2.$$

therefore

$$ds_3^2 = \cos^2 x_2\, ds_2^2 - a^2 dx_2^2.$$

The non-zero Christoffel symbols are

$$\Gamma^0_{02} = \Gamma^0_{20} = -\tan x_2,\ \ \Gamma^0_{11} = \cosh t \sinh t,$$

$$\Gamma^1_{01} = \Gamma^1_{10} = \tanh t,\ \ \Gamma^1_{12} = \Gamma^1_{21} = -\tan x_2,$$

$$\Gamma^2_{00} = -\sin x_2 \cos x_2,\ \ \Gamma^2_{11} = \cosh^2 t \cos x_2 \sin x_2.$$

Now, if we compute

$$R_{ii} = R^s_{isi} = \frac{\partial \Gamma^s_{ii}}{\partial x^s} - \frac{\partial \Gamma^s_{is}}{\partial x^i} + \Gamma^h_{ii}\Gamma^s_{hi} - \Gamma^h_{is}\Gamma^s_{hi},$$

we find

$$R_{00} = -2\cos^2 x_2;\ \ R_{11} = 2\cosh^2 t \cos^2 x_2;\ \ R_{22} = 2.$$

The other Rici symbols are null, $R_{ij} = 0, i \neq j$. Therefore

$$R_{ij} + \frac{2}{a^2}\, g_{ij} = 0.$$

If we compute $R := R^i_i$, taking into account $R^i_j = g^{is} R_{sj}$, it results $R = -\frac{6}{a^2}$. The left hand of Einstein's field equations become

$$R_{ij} - \frac{1}{2}\left(-\frac{6}{a^2}\right) g_{ij} + \Lambda g_{ij}.$$

If we choose $\Lambda = -\frac{1}{a^2}$, the left hand becomes

$$R_{ij} + \frac{2}{a^2}\, g_{ij},$$

that is the left hand becomes 0. The de Sitter space-time presented above satisfies the Einstein field equations

$$R_{ij} - \frac{1}{2} R\, g_{ij} + \Lambda\, g_{ij} = \frac{8\pi G}{c^4} T_{ik}$$

for $R = -\frac{6}{a^2}$, $\Lambda = -\frac{1}{a^2}$ and $T_{ij} = 0$. A space-time without matter appears as we expected.

In the case $n = 5$, the parameterization is

$$\begin{cases} X_0 = a \sinh t \cos x_2 \cos x_3 \\ X_1 = a \cosh t \cos x_1 \cos x_2 \cos x_3 \\ X_2 = a \cosh t \sin x_1 \cos x_2 \cos x_3 \\ X_3 = a \sin x_2 \cos x_3 \\ X_4 = a \sin x_3 \end{cases}$$

The metric related to this parameterization is

$$ds_4^2 = a^2 \cos^2 x_2 \cos^2 x_3 dt^2 - a^2 \cosh^2 t \cos^2 x_2 \cos^2 x_3 dx_1^2 - a^2 \cos^2 x_3 dx_2^2 - a^2 dx_3^2.$$

In the same way, as above, it is

$$ds_4^2 = \cos^2 x_3(a^2\cos^2 x_2 dt^2 - a^2\cosh^2 t\cos^2 x_2 dx_1^2 - a^2 dx_2^2) - a^2 dx_3^2,$$

therefore

$$ds_4^2 = \cos^2 x_3\, ds_3^2 - a^2 dx_3^2.$$

Again, if we compute

$$R_{ii} = R^s_{isi} = \frac{\partial \Gamma^s_{ii}}{\partial x^s} - \frac{\partial \Gamma^s_{is}}{\partial x^i} + \Gamma^h_{ii}\Gamma^s_{hi} - \Gamma^h_{is}\Gamma^s_{hi},$$

we find

$$R_{00} = -3\cos^2 x_2\cos^2 x_3,\ R_{11} = 3\cosh^2 t\cos^2 x_2\cos^2 x_3,\ R_{22} = 3\cos^2 x_3,\ R_{33} = 3,$$

that is

$$R_{ij} + \frac{3}{a^2}\, g_{ij} = 0,$$

which leads to

$$R = -\frac{12}{a^2},\ \Lambda = -\frac{3}{a^2},\ T_{ij} = 0,$$

for Einstein's field equations.

In the general case, the $(n-1)$-*de Sitter space-time* is the Minkowski $(n-1)$-sphere determined by the ends of all the space-like vectors with Minkowski length a.

This is a hypersurface of the Minkowski n-dimensional space $\mathbb{M}^n$ having the algebraic equation

$$X_0^2 - X_1^2 - ... - X_{n-1}^2 = -a^2.$$

The related parameterization is

$$\begin{cases} X_0 = a\sinh t\cos x_2\cos x_3 .... \cos x_{n-2} \\ X_1 = a\cosh t\cos x_1\cos x_2\cos x_3 .... \cos x_{n-2} \\ X_2 = a\cosh t\sin x_1\cos x_2\cos x_3 .... \cos x_{n-2} \\ X_3 = a\sin x_2\cos x_3\cos x_4 .... \cos x_{n-2} \\ X_4 = a\sin x_3\cos x_4 .... \cos x_{n-2} \\ \dots\dots\dots\dots\dots\dots \\ X_{n-2} = a\sin x_{n-3}\cos x_{n-2} \\ X_{n-1} = a\sin x_{n-2}. \end{cases}$$

This parameterization makes sense for $n \geq 5$.

For $n \geq 6$, we can denote $X_{0,n} := X_0;\ X_{1,n} := X_1;\ X_{n-1,n} := X_{n-1}$ and we can write

$$\begin{cases} X_{0,n} = X_{0,n-1} \cos x_{n-2} \\ X_{1,n} = X_{1,n-1} \cos x_{n-2} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X_{n-2,n} = X_{n-2,n-1} \cos x_{n-2} \\ X_{n-1,n} = a \sin x_{n-2} \end{cases}$$

with

$$X_{0,n-1}^2 - X_{1,n-1}^2 - \dots - X_{n-2,n-1}^2 = -a^2.$$

A direct consequence is

$$X_{0,n-1} dX_{0,n-1} - X_{1,n-1} dX_{1,n-1} - \dots - X_{n-2,n-1} dX_{n-2,n-1} = 0.$$

Using

$$\begin{cases} dX_{0,n} = dX_{0,n-1} \cos x_{n-2} - X_{0,n-1} \sin x_{n-2} dx_{n-2} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ dX_{n-2,n} = dX_{n-2,n-1} \cos x_{n-2} - X_{n-2,n-1} \sin x_{n-2} dx_{n-2} \\ dX_{n-1,n} = a \cos x_{n-2} dx_{n-2} \end{cases}$$

and denoting by

$$ds_k^2 = dX_{0,k+1}^2 - dX_{1,k+1}^2 - \dots - dX_{k,k+1}^2 ,$$

we obtain

$$ds_{n-1}^2 = a^2 \cos^2 x_{n-2}\, ds_{n-2}^2 - a^2 dx_{n-2}^2,\ n \geq 6,$$

a formula which is the generalization of the formulas obtained for the previous cases $n = 4$ and $n = 5$.

Therefore, in all cases, we proved that the metric is a diagonal one and we have a recursive method to obtain it. Finally it looks like:

$$ds_{n-1}^2 = a^2 \cos^2 x_{n-2}\, ds_{n-2}^2 - a^2 dx_{n-2}^2,\ n \geq 4$$

and

$$ds_2^2 = a^2 dt^2 - a^2 \cosh^2 t\, dx_1^2,\ n = 3.$$

Now, other considerations are in order. If

$$f(t, x_1, x_2, .., x_{n-2}) = \left(X_{0,n-1} \cos x_{n-2}, \dots, X_{n-2,n-1} \cos x_{n-2}, a \sin x_{n-2}\right)$$

the direct consequence of the above results is $\left\langle f, \dfrac{\partial f}{\partial t} \right\rangle_M = 0.$

Another computation leads to $\left\langle f, \dfrac{\partial f}{\partial x_k} \right\rangle_M = 0$, while $\left\langle \dfrac{\partial f}{\partial x_k}, \dfrac{\partial f}{\partial t} \right\rangle_M = 0$, $\left\langle \dfrac{\partial f}{\partial x_k}, \dfrac{\partial f}{\partial x_j} \right\rangle_M = 0$ are the consequences of the diagonal form of the metric and highlight the orthogonal frame of the tangent space at each point.

Finally, the Minkowski normal to the hypersurface is

$$N(t, x_1, ..., x_{n-2}) = \frac{1}{a} f(t, x_1, ..., x_{n-2}),$$

that is the Minkowski distance from the origin to the tangent hyperplane at a given point of the hypersurface is a and all the coefficients of the second fundamental form are computed with the formula established for the case $n = 2$,

$$h_{ij} = \left\langle \frac{\partial N}{\partial x_i}, \frac{\partial f}{\partial x_j} \right\rangle_M ,$$

therefore

$$h_{ij} = \frac{1}{a} g_{ij}.$$

Since $\langle N, N \rangle_M = -1 < 0$, we have

$$R_{ijkl} = -\left(h_{ik}h_{jl} - h_{il}h_{jk}\right).$$

It results

$$R_{ijkl} = -\frac{1}{a^2}\left(g_{ik}g_{jl} - g_{il}g_{jk}\right), \; i, j, k, l \in \{0, 1, ..., n-2\}.$$

Therefore each sectional curvature is

$$K = -\frac{1}{a^2}.$$

From

$$R_{ijij} = -\frac{1}{a^2}\left(g_{ii}g_{jj} - g_{ij}g_{ji}\right),$$

it results

$$g^{mi} R_{ijij} = -\frac{1}{a^2}\left(g^{mi} g_{ii}g_{jj} - g^{mi} g_{ij}g_{ji}\right),$$

that is

$$R^m_{jij} = -\frac{1}{a^2}\left(\delta^m_i g_{jj} - \delta^m_j g_{ji}\right).$$

For $m = i$, it remains

$$R^m_{jmj} = -\frac{1}{a^2} g_{jj},$$

for each $m \neq j$. Finally,

$$R_{jj} = \sum_{m=0,\ m\neq j}^{n-2} R^m_{jmj} = -\frac{n-2}{a^2}\ g_{jj}.$$

If we start from

$$R_{ijkl} = -\frac{1}{a^2} \left(g_{ik} g_{jl} - g_{il} g_{jk}\right),$$

the same reasoning leads to

$$R^m_{jml} = -\frac{1}{a^2} g_{jl},$$

i.e. for $j \neq l$, we have $R_{jl} = 0$. Therefore

$$R_{ij} + \frac{n-2}{a^2}\ g_{ij} = 0$$

for all i and j. From this formula, we obtain

$$R = -(n-1)(n-2)\frac{1}{a^2}.$$

Since

$$R_{ij} + \frac{1}{2}(n-1)(n-2)\frac{1}{a^2}\ g_{ij} - \frac{(n-2)(n-3)}{2}\frac{1}{a^2}\ g_{ij} = R_{ij} + \frac{n-2}{a^2}\ g_{ij} = 0\,,$$

it results that, if we choose

$$\Lambda = -\frac{(n-2)(n-3)}{2}\frac{1}{a^2},$$

the previous metric satisfies the Einstein field equations

$$R_{ij} - \frac{1}{2} R\ g_{ij} + \Lambda\ g_{ij} = \frac{8\pi G}{c^4} T_{ij}$$

in the absence of matter, that is with $T_{ij} = 0$.

11.3 Some Physical Considerations

Let us return to the 4th dimensional de Sitter space-time described by the parameterization

$$\begin{cases} X_0 = \sinh t \cos x_2 \cos x_3 \\ X_1 = \cosh t \cos x_1 \cos x_2 \cos x_3 \\ X_2 = \cosh t \sin x_1 \cos x_2 \cos x_3 \\ X_3 = \sin x_2 \cos x_3 \\ X_4 = \sin x_3 \end{cases}$$

with the metric

$$ds_4^2 = \cos^2 x_2 \cos^2 x_3 dt^2 - \cosh^2 t \cos^2 x_2 \cos^2 x_3 dx_1^2 - \cos^2 x_3 dx_2^2 - dx_3^2.$$

It is difficult to talk about photons travelling in this Universe but if we consider a slice in the previous de Sitter space, determined by $x_2 = x_3 = 0$, we obviously highlight the 2-de Sitter space-time

$$\begin{cases} X_0 = \sinh t \\ X_1 = \cosh t \cos x_1 \\ X_2 = \cosh t \sin x_1 \end{cases}$$

denoted here by

$$f(t, x_1) = (X_0, X_1, X_2) = (\sinh t, \cosh t \cos x_1, \cosh t \sin x_1),$$

$f : \mathbb{R} \times (-\pi, \pi) \longrightarrow \mathbb{M}^3$, with the metric

$$ds_2^2 = dt^2 - \cosh^2 t dx_1^2$$

and we can hope for a simpler approach of the problem.

It exists two coordinate curves at each given point (t^0, x_1^0). The first one is

$$c_0(t) = f(t, x_1^0)$$

where x_1^0 is a constant. Since

$$\dot{c}_0(t) = \frac{\partial f}{\partial t} = \left(\cosh t, \sinh t \cos x_1^0, \sinh t \sin x_1^0\right)$$

is a time-like vector, i.e. $\left\langle \frac{\partial f}{\partial t}, \frac{\partial f}{\partial t} \right\rangle = 1$, the curve c_0 is a world line for an observer, that is, we are talking about the evolution in time of an event. The relation

$$||\dot{c}_0(t)||_M = 1$$

shows that the parameter t is the proper time because

$$\tau(t) = \int_0^t ||\dot{c}_0(q)||_M dq = t.$$

The other possible curve

$$c_1(x_1) = f(t^0, x_1)$$

where t^0 is a constant, is a "circle" which cuts the "Euclidean hyperboloid", and, at the same time, a space-like curve because $\left\langle \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_1} \right\rangle = -\cosh^2 t = -\cosh^2 \tau < 0.$ Let us analyse the "circumnavigation problem", that is the possibility to go around the "hyperboloid" in a finite amount of time. The length of c_1 is

$$\int_{-\pi}^{\pi} ||\dot{c}_1(\tau)||_M dx_1 = \int_{-\pi}^{\pi} \cosh \tau dx_1 = 2\pi \cosh \tau.$$

The limit, as τ approaches to ∞, is infinite, therefore this space-time is unbounded in both given directions.

We are interested in understanding how photons travel in this de Sitter space-time. Firstly, the metric

$$ds_2^2 = dt^2 - \cosh^2 t \; dx_1^2$$

is described by the metric tensor

$$\begin{pmatrix} 1 & 0 \\ 0 & -\cosh^2 t \end{pmatrix}$$

whose light-cone vectors, in the (t, x) plane, are $L^+ = \begin{pmatrix} x \\ x \dfrac{1}{\cosh t} \end{pmatrix}$ and $L^- = \begin{pmatrix} x \\ -x \dfrac{1}{\cosh t} \end{pmatrix}$, $x \in \mathbb{R}$.

These vectors were deduced in the same way we deduced them in a Minkowski space, whose metric

$$ds^2 = dt^2 - dx^2$$

is described by the metric tensor

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The light-cone vectors are $E^+ = \begin{pmatrix} x \\ x \end{pmatrix}$ and $E^- = \begin{pmatrix} x \\ -x \end{pmatrix}$, $x \in \mathbb{R}$.

If we are looking at the vectors $L^+,\ L^-$ in their transposed form on $\mathbb{M}^3$, from

$$\left(x, \pm\frac{x}{\cosh t}\right) = x\cdot(1,0) \pm \frac{x}{\cosh t}\cdot(0,1),$$

the formulas

$$df_x\left(L^+\right) = x\cdot\frac{\partial f}{\partial t} + \frac{x}{\cosh t}\frac{\partial f}{\partial x_1};\ \ df_x\left(L^-\right) = x\cdot\frac{\partial f}{\partial t} - \frac{x}{\cosh t}\frac{\partial f}{\partial x_1},$$

results. This means that the velocity of photons is the ratio (with a sign) between the norms of the spatial vector $\dfrac{x}{\cosh t}\dfrac{\partial f}{\partial x_1}$ and the temporal vector $x\dfrac{\partial f}{\partial t}$, i.e.

$$\pm\frac{1}{\cosh t}\cdot\frac{||\dot{c}_0(t)||_M}{||\dot{c}_1(x_1)||_M} = \pm 1$$

as we expected.

Let us look again at the tx_1 plane and suppose we have the trajectory of a photon described by a function $x_1(t)$ which is $x_1(\tau)$. In fact, taking into account L^+, we have

$$x_1(q) = \int\frac{2}{e^q+e^{-q}}dq = \int\frac{2e^q}{e^{2q}+1} = 2\arctan(e^q) + C_1,$$

where C_1 is a constant. This function is increasing. The limit as q approaches $-\infty$ is C_1 and the limit as q approaches $+\infty$ is $\pi + C_1$, therefore a photon image curve in tx_1 plane is completely included in a strip with width π. The same happens for the photon described by

$$x_1(q) = -\int\frac{2}{e^q+e^{-q}}dq = -\int\frac{2e^q}{e^{2q}+1} = -2\arctan e^q + C_2,$$

where C_2 is a constant.

If we ask for photons having, at the origin of the tx_1 plane the vectors $L^+(0)$, $L^-(0)$ as tangent vectors respectively, we obtain the curves $x_1(\tau) = 2\arctan e^\tau - \dfrac{\pi}{2}$, and $x_1(\tau) = -2\arctan e^\tau + \dfrac{\pi}{2}$, respectively. The images of these two curves are the trajectories of photons in the de Sitter space-time. Therefore, if we choose the first curve, its image in the de Sitter space-time is

$$\begin{cases} X_0(\tau) = \sinh\tau \\ X_1(\tau) = \cosh\tau\cos\left(2\arctan e^\tau - \dfrac{\pi}{2}\right) \\ X_2(\tau) = \cosh t\sin\left(2\arctan e^\tau - \dfrac{\pi}{2}\right) \end{cases}.$$

Being

$$\cos\left(2\arctan e^{\tau}-\frac{\pi}{2}\right)=\frac{1}{\cosh\tau}$$

and

$$\sin\left(2\arctan e^{\tau}-\frac{\pi}{2}\right)=\tanh\tau,$$

therefore it is

$$\begin{cases} X_0(\tau)=\sinh\tau \\ X_1(\tau)=1 \\ X_2(\tau)=\sinh\tau. \end{cases}$$

We leave, as an exercise for the reader, to prove that the second curve is

$$\begin{cases} X_0(\tau)=\sinh\tau \\ X_1(\tau)=1 \\ X_2(\tau)=-\sinh\tau. \end{cases}$$

Finally, the trajectories of photons in the de Sitter space-time are lines with slopes 1 and -1 (as we expected) which belong, in this case, to the plane $X_1=1$.

Even if it is just the investigation of a light cone at a single point, the reader has to imagine that, at each point of the de Sitter space-time, the situation is the same: the Euclidean hyperboloid have, at each point, a pair of straight lines embedded into its surface.

11.4 A FLRW Metric for de Sitter Space-time Given by the Flat Slicing Coordinates Attached to the Affine Sphere

In [112] it is presented a very interesting parameterization of the affine sphere

$$X_0^2-X_1^2-\ldots-X_{n-1}^2=-a^2,$$

using the *flat slicing coordinates*:

$$f:\begin{cases} X_0=a\sinh\dfrac{t}{a}+\dfrac{r^2}{2a}\cdot e^{t/a} \\ \\ X_1=a\cosh\dfrac{t}{a}-\dfrac{r^2}{2a}\cdot e^{t/a} \\ X_2=y_1\,e^{t/a} \\ X_3=y_2\,e^{t/a} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ X_{n-1}=y_{n-2}\,e^{t/a} \end{cases}$$

with

$$y_1^2 + y_2^2 + ... + y_{n-2}^2 = r^2.$$

It results the relation

$$y_1 dy_1 + y_2 dy_2 + ... + y_{n-2} dy_{n-2} = r dr$$

which helps us to find the corresponding metric.

Now,

$$dX_0 = \left[\cosh\frac{t}{a} + \frac{r^2}{2a^2}e^{t/a}\right]dt + \frac{r}{a}e^{t/a}dr$$

$$dX_1 = \left[\sinh\frac{t}{a} - \frac{r^2}{2a^2}e^{t/a}\right]dt - \frac{r}{a}e^{t/a}dr$$

$$dX_k = e^{t/a}\left[dy_{k-1} + \frac{1}{a}y_{k-1}dt\right],\ k \in \{2, 3, ..., n-1\}.$$

If we compute the metric, firstly we obtain

$$dX_0^2 - dX_1^2 = \left[1 + \frac{r^2}{a^2}e^{2t/a}\right]dt^2 + \frac{2r}{a}e^{2t/a}dtdr.$$

Since

$$\sum_{k=2}^{n-1} dX_k^2 = e^{2t/a}\sum_{k=2}^{n-1}\left[dy_{k-1}^2 + \frac{2}{a}y_{k-1}dy_{k-1}dt + \frac{1}{a^2}y_{k-1}^2 dt^2\right] =$$

$$= e^{2t/a}\sum_{k=2}^{n-1} dy_{k-1}^2 + e^{2t/a}\frac{2r}{a}drdt + e^{2t/a}\frac{r^2}{a^2}dt^2,$$

finally we find

$$ds^2 = dt^2 - e^{2t/a}\left(dy_1^2 + dy_2^2 + ... + dy_{n-2}^2\right).$$

For this metric we have

$$\begin{cases} \Gamma^{\alpha}_{0\alpha} = \Gamma^{\alpha}_{\alpha 0} = \dfrac{1}{a} \\ \Gamma^{0}_{\alpha\alpha} = \dfrac{1}{a}\, e^{2t/a},\ \alpha \in \{1, 2, ..., n-2\}, \end{cases}$$

all the other Christoffel symbols are null. Then,

$$R_{00} = R^{0}_{000} + R^{\beta}_{0\beta 0} = \sum_{\beta=1}^{n-2} R^{\beta}_{0\beta 0} = \sum_{\beta=1}^{n-2} \left[\frac{\partial \Gamma^{\beta}_{00}}{\partial x^{\beta}} - \frac{\partial \Gamma^{\beta}_{0\beta}}{\partial x^{0}} + \Gamma^{m}_{00}\Gamma^{\beta}_{m\beta} - \Gamma^{m}_{0\beta}\Gamma^{\beta}_{m0} \right] =$$

$$= -\frac{n-2}{a^2} = -\frac{n-2}{a^2} g_{00};$$

$$R_{\alpha\alpha} = R^{s}_{\alpha s \alpha} = R^{\alpha}_{\alpha\alpha\alpha} + \sum_{s=0, s\neq\alpha}^{n-2} R^{s}_{\alpha s \alpha} = \sum_{s=0, s\neq\alpha}^{n-2} \left[\frac{\partial \Gamma^{s}_{\alpha\alpha}}{\partial x^{s}} - \frac{\partial \Gamma^{s}_{\alpha s}}{\partial x^{\alpha}} + \Gamma^{m}_{\alpha\alpha}\Gamma^{s}_{sm} - \Gamma^{m}_{\alpha s}\Gamma^{s}_{m\alpha} \right] =$$

$$= \frac{2}{a^2} e^{2t/a} + (\Gamma^{1}_{01} + \Gamma^{2}_{02} + ... + \Gamma^{n-2}_{0n-2})\Gamma^{0}_{\alpha\alpha} - \Gamma^{\alpha}_{0\alpha}\Gamma^{0}_{\alpha\alpha} - \Gamma^{0}_{\alpha\alpha}\Gamma^{\alpha}_{0\alpha} = -\frac{n-2}{a^2} e^{2t/a} = -\frac{n-2}{a^2} g_{\alpha\alpha}.$$

Therefore

$$R = -(n-1)(n-2)\frac{1}{a^2},$$

that is choosing

$$\Lambda = -\frac{(n-2)(n-3)}{2}\frac{1}{a^2},$$

the previous metric satisfies the Einstein field equations in absence of matter,

$$R_{ij} - \frac{1}{2} R\, g_{ij} + \Lambda g_{ij} = 0.$$

This metric can be written in the form

$$ds^2 = dt^2 - e^{2t/a} dy^2$$

where

$$dy^2 = dy_1^2 + dy_2^2 + ... + dy_{n-2}^2$$

is the flat metric in the y_k coordinates, which explains the name.

This is an example of a *FLRW metric for de Sitter space-time*.

Example 11.4.1 Consider de Sitter space-time having one of the previous metric g_{ij}. We know that

$$R_{ij} = -\frac{n-2}{a^2} g_{ij}$$

and

$$R = -(n-1)(n-2)\frac{1}{a^2}.$$

Therefore, in the case of $f(R) = R$ gravity, we obtain Einstein's field equations

$$R_{ij} - \frac{1}{2} R\, g_{ij} + \Lambda\, g_{ij} = 0$$

with the cosmological constant

$$\Lambda = -\frac{(n-2)(n-3)}{2} \frac{1}{a^2},$$

because we have

$$R_{ij} + \frac{1}{2}(n-1)(n-2)\frac{1}{a^2}\, g_{ij} - \frac{(n-2)(n-3)}{2} \frac{1}{a^2}\, g_{ij} = R_{ij} + \frac{n-2}{a^2}\, g_{ij} = 0.$$

If we consider the case of $f(R) = R^2$ gravity, we have

$$f'(R) R_{ij} - \frac{1}{2} f(R) g_{ij} + \Lambda_f\, g_{ij} = 0$$

if

$$\Lambda_f = \frac{(n-1)(n-2)^2(n-5)}{2a^4} = -R\, \frac{(n-2)(n-5)}{2a^2}.$$

Indeed,

$$2R\, R_{ij} - \frac{1}{2} R^2\, g_{ij} - R\, \frac{(n-2)(n-5)}{2a^2}\, g_{ij} = 2R \left(R_{ij} - \frac{1}{4} R\, g_{ij} - \frac{(n-2)(n-5)}{4a^2}\, g_{ij} \right) =$$

$$= 2R \left(R_{ij} + \frac{(n-1)(n-2)}{4a^2}\, g_{ij} - \frac{(n-2)(n-5)}{4a^2}\, g_{ij} \right) = 2R \left(R_{ij} + \frac{n-2}{a^2}\, g_{ij} \right) = 0.$$

Therefore we have the following statement.

The $f(R) = R^2$ gravity equations in absence of matter

$$f'(R) R_{ij} - \frac{1}{2} f(R) g_{ij} + \Lambda_f\, g_{ij} = 0$$

in the case of the cosmological constant

$$\Lambda_f = \frac{(n-1)(n-2)^2(n-5)}{2a^4}$$

are satisfied by any metric

$$ds^2 = g_{ij} dx^i dx^j,\ i, j \in \{0, 1, ..., n-2\}$$

having the property

$$R_{ij} = -\frac{n-2}{a^2} g_{ij}.$$

An example is de Sitter space-time metric

$$ds^2 = dt^2 - e^{2t/a}\left(dy_1^2 + dy_2^2 + ... + dy_{n-2}^2\right).$$

A particular situation happens when $n = 5$. The $f(R)$ equations in vacuum are satisfied in the their original form

$$f'(R)R_{ij} - \frac{1}{2} f(R) g_{ij} = 0,$$

that is no cosmological constant is needed.

Exercise 11.4.2 For the unit space-like affine sphere

$$X_0^2 - X_1^2 - ... - X_{n-1}^2 = -1$$

consider the parameterization

$$f : \begin{cases} X_0 = a \sinh t \\ X_1 = y_1 \cosh t \\ \dots\dots\dots\dots\dots \\ X_{n-1} = y_{n-1} \cosh t \end{cases}$$

with

$$y_1^2 + y_2^2 + ... + y_{n-1}^2 = 1,$$

that is

$$y_1 dy_1 + y_2 dy_2 + ... + y_{n-2} dy_{n-2} = 0.$$

(i) Show that the corresponding metric is

$$ds^2 = dt^2 - \cosh^2 t\left(dy_1^2 + dy_2^2 + ... + dy_{n-1}^2\right).$$

(ii) Try to understand why the metric can be written in the form

$$ds^2 = dt^2 - \cosh^2 t \; d\Omega_{n-2}^2,$$

where $d\Omega_{n-2}^2$ is the metric of the S^{n-2} sphere.

11.5 Deriving Cosmological Singularities in the Context of de Sitter Space-time

Let us consider the case $n = 3$. The Euclidean one-sheet hyperboloid, which is in fact a Minkowki sphere, has the algebraic equation

$$X_0^2 - X_1^2 - X_2^2 = -1.$$

Using the flat slicing coordinates

$$f : \begin{cases} X_0 = \sinh t + \dfrac{r^2}{2} \cdot e^t \\ X_1 = \cosh t - \dfrac{r^2}{2} \cdot e^t \\ X_2 = r \cdot e^t. \end{cases}$$

we find, as a particular case of our previous discussion, the metric

$$ds^2 = dt^2 - e^{2t} dr^2.$$

For this metric, we have

$$\begin{cases} \Gamma^1_{01} = \Gamma^1_{10} = 1 \\ \Gamma^0_{11} = e^{2t}, \end{cases}$$

all the other Christoffel symbols are null.

Let us consider the Minkowski sphere. At $t = 0$, consider the curve

$$m(r) = \left(\frac{r^2}{2}, 1 - \frac{r^2}{2}, r \right)$$

obtained by replacing $t = 0$ in the parameterization of f. This curve is the intersection between the plane

$$X_0 + X_1 = 1$$

and the Minkowski sphere

$$X_0^2 - X_1^2 - X_2^2 = -1.$$

We can conceive this curve as at the initial singularity at the origin of the de Sitter space-time. Let us follow the evolution in time of it and choose a point of the singularity,

$$m(r_0) = \left(\frac{r_0^2}{2},\ 1 - \frac{r_0^2}{2},\ r_0 \right).$$

The evolution in time of this point is the line

$$c(t) = \left(\sinh t + \frac{r_0^2}{2} \cdot e^t,\ \cosh t + \frac{r_0^2}{2} \cdot e^t,\ r_0 \cdot e^t\right).$$

Theorem 11.5.1 *1. $c(t)$ is the intersection between the plane $X_0 + X_1 - \frac{1}{r_0} X_2 = 0$ and the previous "hyperboloid" $X_0^2 - X_1^2 - X_2^2 = -1$.*
2. The Minkowski product $< \dot{c}(t), \dot{c}(t) >_M$ is 1, i.e. the tangent vector, is a time-like vector.
3. $c(t)$ is a time-like geodesic of de Sitter space-time.

Proof We leave for the reader the proof of the first two points which are simple exercises.
The equations of the geodesics are

$$\begin{cases} \dfrac{d^2 t}{d\tau^2} + \Gamma^0_{11} \dfrac{dr}{d\tau} \dfrac{dr}{d\tau} = 0 \\[2ex] \dfrac{d^2 r}{d\tau^2} + 2\Gamma^1_{01} \dfrac{dt}{d\tau} \dfrac{dr}{d\tau} = 0, \end{cases}$$

i.e.

$$\begin{cases} \dfrac{d^2 t}{d\tau^2} + e^{2t} \dfrac{dr}{d\tau} \dfrac{dr}{d\tau} = 0 \\[2ex] \dfrac{d^2 r}{d\tau^2} + 2 \dfrac{dt}{d\tau} \dfrac{dr}{d\tau} = 0. \end{cases}$$

The solution $t = \tau,\ r = r_0$ corresponds to the curve c. □

Therefore, the line is the evolution in time of the point and it is the first line we considered in the part of de Sitter space-time out of the singularity curve.
Now, from each point of

$$m(r) = \left(\frac{r^2}{2},\ 1 - \frac{r^2}{2},\ r\right)$$

consider the corresponding curve

$$c_r(t) = \left(\sinh t + \frac{r^2}{2} \cdot e^t,\ \cosh t + \frac{r^2}{2} \cdot e^t,\ r \cdot e^t\right).$$

All these time-like geodesics for $t = \tau > 0$ starting from the initial singularity are part of the texture of the de Sitter space-time. Observe that not all of the Minkowski sphere is the texture of the "evolution started from the singularity". Furthermore the parameterization makes sense also for $t = \tau < 0$. We have the image of an evolution of a singularity corresponding to the de Sitter space-time in a chosen time direction.

De Sitter space-time is only a possible geometric realization of Relativity. Another possible realization is analysed in the next chapter. We are talking about the Anti-de Sitter space-time. We prefer to warn the reader about the fact that some geometric properties of de Sitter and Anti-de Sitter spaces-times are different if we prefer different metric signatures and other types of Minkowski spheres chosen in the basic definitions. Other interesting geometric realizations of space-times can be found in [167].

Chapter 12
Another Geometric Realization of Relativity: The Anti-de Sitter Space–Time

It is wrong to think that the task of Physics is to find out how Nature is. Physics concerns what we say about Nature.
Niels Bohr

There are a lot of similarities between de Sitter and Anti-de Sitter space–times; therefore, we want to present here some geometric aspects which are not highlighted in the previous chapter. Before developing a geometric theory of gravity which can be extended in Minkowski $M^{(p,q)}$ *spaces and before reaching the subject of Anti-de Sitter space–times as examples of universes without matter, we want to introduce some facts related to the centro-affine geometry, a branch of geometry developed by Gheorghe Tzitzeica. These facts are essential for the development of the chapter (see* [30]*).*

The theory of Tzitzeica surfaces in Euclidean spaces was the starting point of the affine differential geometry in which the differential invariants depend on the metric, but they are also preserved by affine transformations (see [151, 186, 187]*). The geometric nature of de Sitter and Anti-de Sitter space–times is related to the Minkowski affine differential geometry and this is the main aspect we are going to present here.*

Shortly, in the case of Euclidean 3*-dimensional space, the definition of a Tzitzeica surface is this: we choose a point* $f(p)$ *of the surface locally written in the form* $f(x, y) = (x, y, u(x, y))$ *and we compute the Gaussian curvature* $K_f(p)$ *at that point. Then, we compute the distance, denoted by* $d_f(p)$*, from the origin* $O(0, 0, 0)$ *to the tangent plane of the surface at the point* $f(p)$*. The surface is called a Tzitzeica surface, if the ratio* $R^f(p) := \dfrac{K_f(p)}{d_f^4(p)}$ *is a constant at each* p*, that is* $R^f(p) = R^f$*. The constant* R^f *is called "affine radius" and becomes an intrinsic number attached to the surface.*

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*, UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_12

Both de Sitter and Anti-de Sitter space–times are Minkowski-affine Tzitzeica hypersurfaces and they have the affine radius related to their Ricci curvature scalar R. We have to point out that only the Gauss constant geometric curvature is not essential for a surface to be an affine sphere. The pseudosphere has constant Gaussian curvature but it is not a Tzitzeica surface; de Sitter and Anti-de Sitter space–times have constant Ricci curvature and both fulfil the Tzitzeica property, which in fact makes them remarkable.

Let us explain how it works in the case of surfaces. Suppose we have computed the affine radius

$$R^f(p) := \frac{K_f(p)}{d_f^4(p)}.$$

By definition, a centro-affine transformation of the surface f consists in the product between the surface f and 3×3 matrix $\mathbb{A}$, with $\det \mathbb{A} \neq 0$. Therefore the surface obtained is now

$$\bar{f}(x, y) = (a_{11}x + a_{21}y + a_{31}u(x, y), a_{12}x + a_{22}y + a_{32}u(x, y), a_{13}x + a_{23}y + a_{33}u(x, y)),$$

where the coefficients a_{ij} are the components of the matrix $\mathbb{A}$.
For this surface $\bar{f}$, when we compute $R^{\bar{f}}(q) = \frac{K_{\bar{f}}(q)}{d_{\bar{f}}^4(q)}$, we obtain a constant if and only if the initial $R^f(p) = \frac{K_f(p)}{d_f^4(p)}$ is a constant. The constant is the same if and only if $\det \mathbb{A} = 1$ and this can be called the total affine invariance condition.

The previous observation allows to extend the definition of Tzitzeica surfaces to hypersurfaces in Euclidean and Minkowski n-dimensional spaces, the affine radius being there the ratio $\frac{K_f(p)}{d_f^{(n+1)}(p)}$.

A general theory about curves and surfaces in Minkowski spaces $M^{(1,n)}$ is reported in [132], while the extension of the concept of Tzitzeica surfaces to Minkowski three-dimensional spaces is presented in [28]. Other basic results in Minkowski $M^{(1,n)}$ spaces are reported in [31]. In the following, we present a generalization for hypersurfaces in Minkowski $M^{(2,n)}$ spaces in a way that can be immediately extended to Minkowski $M^{(p,q)}$ spaces. All these results will be related to the Physics of Anti-de Sitter space–times (see [30]).

12.1 The Minkowski $M^{(2,4)}$ Geometric Gravitational Force

The theory we present below is reported in details in [30] and can be formulated in any Minkowski $M^{(p,q)}$ space. In the case of $M^{(1,n)}$ Minkowski spaces, it was first presented in [31]. Let us say that $M^{(1,n)}$ is the Minkowski n-dimensional space endowed with the Minkowski product

$$\langle a, b\rangle_M := a_0 b_0 - \sum_{\alpha=1}^{n-1} a_\alpha b_\alpha.$$

We discussed its properties in the previous chapter when we considered the de Sitter space–time.

Let us denote by $M^{(2,4)}$ the Minkowski 4-dimensional space endowed with the Minkowski product

$$\langle a, b\rangle_M := a_0 b_0 + a_1 b_1 - \sum_{\alpha=2}^{3} a_\alpha b_\alpha$$

Therefore, (2, 4) is related to the fact that the signature we choose is $(++--)$. With respect to a given $b = (b_0, b_1, b_2, b_3)$, we consider all vectors $x = (x_0, x_1, x_2, x_3)$ such that $x - b$ is a space-like vector, that is $\langle x - b, x - b\rangle_M < 0$.

We denote by

$$r := \sqrt{-(x_0 - b_0)^2 - (x_1 - b_1)^2 + (x_2 - b_2)^2 + (x_3 - b_3)^2}$$

the Minkowski "length" of the space-like vector $x - b$ and by

$$u = -\frac{1}{r}(x_0 - b_0, x_1 - b_1, x_2 - b_2, x_3 - b_3)$$

the unit vector of $b - x$.

We define the Minkowski $M^{(2,4)}$ geometric gravitational force by

$$F := \frac{1}{3}\,\frac{1}{r^3}\,u.$$

If

$$A := \frac{2}{r^3}\,u$$

is by definition the geometric gravitational Minkowski $M^{(2,4)}$ field, we have the adapted "Minkowski-Newton second principle":

$$F = \frac{1}{6}A.$$

Let us define the Minkowski $M^{(2,4)}$ gradient and the Minkowski $M^{(2,4)}$ Laplacian:

$$\nabla := \left(-\frac{\partial}{\partial x_0}, -\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right);$$

$$\nabla^2 := \langle \nabla, \nabla\rangle_M = \frac{\partial^2}{\partial x_0^2} + \frac{\partial^2}{\partial x_1^2} - \frac{\partial^2}{\partial x_2^2} - \frac{\partial^2}{\partial x_3^2}.$$

If we define the Minkowski $M^{(2,4)}$ gravitational potential

$$\Phi := -\frac{1}{r^2}$$

the following computations:

$$\frac{\partial \Phi}{\partial x_0} = -2\frac{x_0 - b_0}{r^4}; \quad \frac{\partial \Phi}{\partial x_1} = -2\frac{x_1 - b_1}{r^4}; \quad \frac{\partial \Phi}{\partial x_2} = 2\frac{x_2 - b_2}{r^4}; \quad \frac{\partial \Phi}{\partial x_3} = 2\frac{x_3 - b_3}{r^4};$$

$$\frac{\partial^2 \Phi}{\partial x_0^2} = -2\frac{r^2 + 4(x_0 - b_0)^2}{r^6}; \quad \frac{\partial^2 \Phi}{\partial x_1^2} = -2\frac{r^2 + 4(x_1 - b_1)^2}{r^6};$$

$$\frac{\partial^2 \Phi}{\partial x_2^2} = 2\frac{r^2 - 4(x_2 - b_2)^2}{r^6}; \quad \frac{\partial^2 \Phi}{\partial x_3^2} = 2\frac{r^2 - 4(x_3 - b_3)^2}{r^6};.$$

lead us to the following two theorems.

Theorem 12.1.1 *The Minkowski $M^{(2,4)}$ gradient of the Minkowski $M^{(2,4)}$ gravitational potential is the opposite of the Minkowski $M^{(2,4)}$ gravitational field.*

Proof It is easy to check the equality $\nabla \Phi = -A$. □

Theorem 12.1.2 *The Minkowski $M^{(2,4)}$ Laplacian of the Minkowski $M^{(2,4)}$ gravitational potential is null.*

Proof The same, it is easy to check $\nabla^2 \Phi = 0$. □

The last relation is here the equation of the Minkowski $M^{(2,4)}$ geometric gravitational field.

In the case when b is the origin of the Minkowski $M^{(2,4)}$ space, the unitary space-like sphere can be thought as the set of points of $M^{(2,4)}$ described by the constant gravitational Minkowski potential $\Phi_M^4 = -1$.

The Minkowski $M^{(2,4)}$ geometrical gravitational force and the Minkowski $M^{(2,4)}$ geometrical gravitational field have dimension $\frac{1}{(l)^3}$ where (l) is a length. Therefore, we can replace by $\frac{1}{(m)^3}$. In the same way, the Minkowski $M^{(2,4)}$ gravitational potential has as dimension $\frac{1}{(m)^2}$.

We may conclude: For each dimension, generalizing the previous theory, in the Minkowski $M^{(p,q)}$ space-like vectors region, a natural geometric Minkowski gravity appears in the absence of matter. Adapting the number of positive and negative signs, we can construct adapted Minkowski $M^{(p,q)}$ gradients and Laplacians. An equivalent of the Newton gravity theory can be constructed this way. The affine space-like spheres can be seen as the regions of the Minkowski $M^{(p,q)}$ space-like vectors characterized by a constant Minkowski gravitational potential. They highlight a model of space–time, the Anti-de Sitter one, which satisfies Einstein's field equations in the absence of matter and this happens because they are surfaces generated by a Minkowski gravitational force described without mass distributions.

12.2 The Minkowski–Tzitzeica Surfaces

The theory we present here is developed in [31]. However, according to the geometric structure of Minkowski $M^{(2,q)}$ some formulas are changed with respect to the cited reference. See also [30].

In a Minkowski $M^{(2,3)}$ space, we have:

- the Minkowski $M^{(2,3)}$ product of vectors

$$\langle a, b\rangle_M := a_0b_0 + a_1b_1 - a_2b_2 \,,$$

- the Minkowski $M^{(2,3)}$ crossproduct of vectors

$$a \times_M b := (a_1b_2 - a_2b_1, -a_0b_2 + a_2b_0, a_1b_0 - a_0b_1)\,,$$

which can be understood from the formal determinant components,

$$\begin{vmatrix} \vec{i} & \vec{j} & -\vec{k} \\ a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 \end{vmatrix} ,$$

- a surface locally represented by $f : U = \overset{\circ}{U} \subset \mathbb{R}^2 \longrightarrow M^{(2,3)}$ having the form

$$f(x, y) = (x, y, u(x, y))\,,$$

with the metric

$$ds^2 = \left(1 - \left(\frac{\partial u}{\partial x}\right)^2\right) dx^2 - 2\frac{\partial u}{\partial x}\frac{\partial u}{\partial y}dxdy + \left(1 - \left(\frac{\partial u}{\partial y}\right)^2\right) dy^2\,,$$

- the Gauss–Minkowski curvature $K_f^M(p)$ of f at the point $f(p)$, where $p = (x, y) \in U$,

$$K_f^M(p) = \frac{R_{1212}}{\det g_{ij}} = \frac{\dfrac{\partial^2 u}{\partial x^2}\dfrac{\partial^2 u}{\partial y^2} - \left(\dfrac{\partial^2 u}{\partial x \partial y}\right)^2}{\left[1 - \left(\dfrac{\partial u}{\partial x}\right)^2 - \left(\dfrac{\partial u}{\partial y}\right)^2\right]^2},$$

- the equation of the tangent plane α at the point $f(p)$ is

$$\alpha : (X - x)\left(-\frac{\partial u}{\partial x}\right) - (Y - y)\left(\frac{\partial u}{\partial y}\right) + (Z - u(x, y)) = 0\,,$$

- the Minkowski $M^{(2,3)}$ "distance", denoted $d_f^M(p)$, from the origin to the tangent plane of the surface f at the point $f(p)$ computed after the formula

$$d_f^M(p) = \frac{\left|x\dfrac{\partial u}{\partial x} + y\dfrac{\partial u}{\partial y} - u(x, y)\right|}{\sqrt{\left|1 - \left(\dfrac{\partial u}{\partial x}\right)^2 - \left(\dfrac{\partial u}{\partial y}\right)^2\right|}}.$$

An immediate consequence of the above formula is

$$\frac{K_f^M(p)}{(d_f^M)^4(p)} = \frac{\dfrac{\partial^2 u}{\partial x^2}\dfrac{\partial^2 u}{\partial y^2} - \left(\dfrac{\partial^2 u}{\partial x \partial y}\right)^2}{\left(x\dfrac{\partial u}{\partial x} + y\dfrac{\partial u}{\partial y} - u(x, y)\right)^4}.$$

Consider a matrix $\mathbb{A} \in \mathcal{M}_3(\mathbb{R})$, such that $\det \mathbb{A} \neq 0$.

By definition, a centro-affine transformation of f is a surface $\bar{f} : U = \overset{\circ}{U} \subset \mathbb{R}^2 \longrightarrow M^{(2,3)}$ given by the formula $\bar{f}(x, y) = f(x, y) \cdot \mathbb{A}$.

$$\bar{f}(x, y) = (x, y, u(x, y)) \cdot \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Therefore,

$$\bar{f}(x, y) = (a_{11}x + a_{21}y + a_{31}u(x, y), a_{12}x + a_{22}y + a_{32}u(x, y), a_{13}x + a_{23}y + a_{33}u(x, y))\,.$$

We denote by $\bar{K}_{\bar{f}}^M$, the Minkowski–Gauss curvature of $\bar{f}$ at the point $\bar{f}(p)$, where $p = (x, y)$, and by $d^M{}_{\bar{f}}$, the Minkowski distance from the origin to $\bar{\alpha}$, the tangent plane of the surface $\bar{f}$ at the point $\bar{f}(p)$. One may observe that a centro-affine transformation changes the "shape" of a surface, the curvature of it at the new corresponding point and changes the distance between the origin and the tangent plane at the new corresponding point of the surface. Something remains invariant and the conclusion, after the following theorem, highlights this invariant.

Theorem 12.2.1

$$\frac{\bar{K}^M_{\bar{f}}(p)}{(d^M{}_{\bar{f}})^4(p)} = \frac{1}{(det\mathbb{A})^2} \cdot \frac{\dfrac{\partial^2 u}{\partial x^2}\dfrac{\partial^2 u}{\partial y^2} - \left(\dfrac{\partial^2 u}{\partial x \partial y}\right)^2}{\left(x\dfrac{\partial u}{\partial x} + y\dfrac{\partial u}{\partial y} - u(x,y)\right)^4} = \frac{1}{(det\mathbb{A})^2} \cdot \frac{K^M_f(p)}{(d^M_f)^4(p)}.$$

We prefer, instead of a very long computational proof of this theorem, to offer the latter a very nice geometric argument. The previous theorem leads to a very important conclusion:

Knowing that f has the property $\dfrac{K^M_f(p)}{(d^M_f)^4(p)}$ is a constant, using the previous theorem, we obtain $\dfrac{\bar{K}^M_{\bar{f}}(p)}{(d^M{}_{\bar{f}})^4(p)}$ is a constant, too. □

An important class of surfaces is highlighted in Minkowski $M^{(2,3)}$ in the same way as it was presented in the case of Minkowski $M^{(1,3)}$ space ([28, 31]).

Definition 12.2.2 In a Minkowski $M^{(2,3)}$ space a surface f is called a Minkowski–Tzitzeica surface if $\dfrac{K^M_f(p)}{(d^M_f)^4(p)}$ is a constant.

For a Minkowski–Tzitzeica surface, the quantity $\dfrac{K^M_f(p)}{(d^M_f)^4(p)}$ is called an affine radius.

12.3 The Geometric Nature of the Affine Radius in a Minkowski $M^{(2,3)}$ Space

Let us look at the ratio $\dfrac{K^M_f(p)}{d^{M\,4}_{\ f}(p)}$. We can observe the numerator

$$L := \frac{\partial^2 u}{\partial x^2}\frac{\partial^2 u}{\partial y^2} - \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2$$

which can be written with respect to the vectors basis

$$\frac{\partial f}{\partial x} = \left(1, 0, \frac{\partial u}{\partial x}\right); \ \frac{\partial f}{\partial y} = \left(0, 1, \frac{\partial u}{\partial y}\right)$$

and four second-order vectors

$$\frac{\partial^2 f}{\partial x^2} = \left(0, 0, \frac{\partial^2 u}{\partial x^2}\right); \ \frac{\partial^2 f}{\partial y^2} = \left(0, 0, \frac{\partial^2 u}{\partial y^2}\right); \ \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \left(0, 0, \frac{\partial^2 u}{\partial x \partial y}\right)$$

in the form

$$L = \begin{vmatrix} 0 & 0 & \frac{\partial^2 u}{\partial x^2} \\ 1 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & \frac{\partial u}{\partial y} \end{vmatrix} \cdot \begin{vmatrix} 0 & 0 & \frac{\partial^2 u}{\partial y^2} \\ 1 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & \frac{\partial u}{\partial y} \end{vmatrix} - \begin{vmatrix} 0 & 0 & \frac{\partial^2 u}{\partial x \partial y} \\ 1 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & \frac{\partial u}{\partial y} \end{vmatrix} \cdot \begin{vmatrix} 0 & 0 & \frac{\partial^2 u}{\partial y \partial x} \\ 1 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & \frac{\partial u}{\partial y} \end{vmatrix}.$$

L is in fact a difference of two products of 3-volumes.

The denominator can be written in the form

$$\begin{vmatrix} x & y & u(x, y) \\ 1 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & \frac{\partial u}{\partial y} \end{vmatrix}^4.$$

Therefore is the 4th power of a 3-volume. The ratio $\frac{K_f^M(p)}{d^{M^4}_f(p)}$ has at the denominator a dimension given as the second power of a 3-volume. The cubic root

$$\sqrt[3]{\frac{K_f^M(p)}{d^{M^4}_f(p)}}$$

has as dimension $\frac{1}{(m)^2}$ and we use this information later.

And more, if we are looking at a centro-affine transformation, the vector basis we use are

$$\frac{\partial \bar{f}}{\partial x} = \left(a_{11} + a_{31}\frac{\partial u}{\partial x}, a_{12} + a_{32}\frac{\partial u}{\partial x}, a_{13} + a_{33}\frac{\partial u}{\partial x}\right)$$

$$\frac{\partial \bar{f}}{\partial y} = \left(a_{21} + a_{31}\frac{\partial u}{\partial y}, a_{22} + a_{32}\frac{\partial u}{\partial y}, a_{23} + a_{33}\frac{\partial u}{\partial y}\right)$$

and the four other vectors are

$$\frac{\partial^2 \bar{f}}{\partial x^2} = \frac{\partial^2 u}{\partial x^2} \cdot (a_{31}, a_{32}, a_{33}),\ \frac{\partial^2 \bar{f}}{\partial x \partial y} = \frac{\partial^2 \bar{f}}{\partial y \partial x} = \frac{\partial^2 u}{\partial x \partial y} \cdot (a_{31}, a_{32}, a_{33}),\ \frac{\partial^2 \bar{f}}{\partial y^2} = \frac{\partial^2 u}{\partial y^2} \cdot (a_{31}, a_{32}, a_{33}).$$

The first determinant of the numerator is

$$\begin{vmatrix} a_{31}\dfrac{\partial^2 u}{\partial x^2} & a_{32}\dfrac{\partial^2 u}{\partial x^2} & a_{33}\dfrac{\partial^2 u}{\partial x^2} \\ a_{11} + a_{31}\dfrac{\partial u}{\partial x} & a_{12} + a_{32}\dfrac{\partial u}{\partial x} & a_{13} + a_{33}\dfrac{\partial u}{\partial x} \\ a_{21} + a_{31}\dfrac{\partial u}{\partial y} & a_{22} + a_{32}\dfrac{\partial u}{\partial y} & a_{23} + a_{33}\dfrac{\partial u}{\partial y} \end{vmatrix} = det\mathbb{A} \cdot \frac{\partial^2 u}{\partial x^2},$$

therefore it becomes easy to see that the numerator has the value $det\mathbb{A}^2 \cdot \left(\frac{\partial^2 u}{\partial x^2}\frac{\partial^2 u}{\partial y^2} - \left(\frac{\partial^2 u}{\partial x \partial y}\right)^2\right)$.

The denominator can be described by the following 4th power of a determinant together a sign we can neglect,

$$\begin{vmatrix} a_{11}x + a_{21}y + a_{31}u(x,y) & a_{12}x + a_{22}y + a_{32}u(x,y) & a_{13}x + a_{23}y + a_{33}u(x,y) \\ a_{11} + a_{31}\dfrac{\partial u}{\partial x} & a_{12} + a_{32}\dfrac{\partial u}{\partial x} & a_{13} + a_{33}\dfrac{\partial u}{\partial x} \\ a_{21} + a_{31}\dfrac{\partial u}{\partial y} & a_{22} + a_{32}\dfrac{\partial u}{\partial y} & a_{23} + a_{33}\dfrac{\partial u}{\partial y} \end{vmatrix}^4 .$$

It is easy to compute the determinant above using the usual properties of the determinants.

Finally, the denominator has the value $\det \mathbb{A}^4 \cdot \left(x\frac{\partial u}{\partial x} + y\frac{\partial u}{\partial y} - u\,(x, y)\right)^4$.

Therefore, Tzitzeica surfaces are related, both in the Minkowski $M^{(2,3)}$ case and in the initially considered Minkowski $M^{(1,3)}$ cases (see [28, 31]), to the volume invariance.

One more comment we made also in [31]. Let us understand the meaning of an affine transformation. The vectors $(1, 0, 0)$; $(0, 1, 0)$, $(0, 0, 1)$ transform into the vectors

$$(a_{11}, a_{12}, a_{13});\ (a_{21}, a_{22}, a_{23});\ (a_{31}, a_{32}, a_{33})$$

described by the rows of the matrix $\mathbb{A}$. Therefore, the initial volume determined by the vectors $(1, 0, 0)$; $(0, 1, 0)$, $(0, 0, 1)$, that is 1, is transformed into the volume $\det \mathbb{A}$.

In the same way we considered unit vectors on the initial axes, we may consider the vectors $\frac{1}{\sqrt[3]{\det \mathbb{A}}}(a_{11}, a_{12}, a_{13})$; ..., which determine the unit volume after the centro-affine transformation. The affine radius is now fully preserved after a centro-affine transformation.

12.4 Geometrical Considerations Related to the Affine Radius in the Minkowski $M^{(2,4)}$ Space

In a Minkowski 4−dimensional space $M^{(2,4)}$, the 3-surface we consider is $f(x, y, z) = (x, y, z, u(x, y, z))$, with (x, y, z) belonging to an open three-dimensional domain.

The vector we consider are related to the tangent 3-space, that is

$$\frac{\partial f}{\partial x} = \left(1, 0, 0, \frac{\partial u}{\partial x}\right); \; \frac{\partial f}{\partial y} = \left(0, 1, 0, \frac{\partial u}{\partial y}\right); \frac{\partial f}{\partial z} = \left(0, 0, 1, \frac{\partial u}{\partial z}\right),$$

together with the other six second-order derivatives

$$\frac{\partial^2 f}{\partial x^2} = \left(0, 0, 0, \frac{\partial^2 u}{\partial x^2}\right); \; \frac{\partial^2 f}{\partial x \partial y} = \left(0, 0, 0, \frac{\partial^2 u}{\partial x \partial y}\right); \; \frac{\partial^2 f}{\partial x \partial z} = \left(0, 0, 0, \frac{\partial^2 u}{\partial x \partial z}\right);$$

$$\frac{\partial^2 f}{\partial y^2} = \left(0, 0, 0, \frac{\partial^2 u}{\partial y^2}\right); \; \frac{\partial^2 f}{\partial y \partial z} = \left(0, 0, 0, \frac{\partial^2 u}{\partial y \partial z}\right); \; \frac{\partial^2 f}{\partial z^2} = \left(0, 0, 0, \frac{\partial^2 u}{\partial z^2}\right);$$

If the surface is seen as a vector, there are involved seven 4-determinants we need to highlight the affine Minkowski radius. It is easy to compute $g_{ij} = \left\langle \frac{\partial f}{\partial x^i}(x), \frac{\partial f}{\partial x^j}(x) \right\rangle_M$. Therefore,

$$\det g_{ij} = -1 + \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - \left(\frac{\partial u}{\partial z}\right)^2 = \epsilon \left| \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - \left(\frac{\partial u}{\partial z}\right)^2 - 1 \right|,$$

where ϵ is the algebraic sign of $\det g_{ij}$.

Then, the 4-Minkowski $M^{(2,4)}$ normal N to the surface is related to the formal developing of the determinant

$$\begin{vmatrix} \vec{i} & \vec{j} & -\vec{k} & -\vec{l} \\ 1 & 0 & 0 & \frac{\partial u}{\partial x} \\ 0 & 1 & 0 & \frac{\partial u}{\partial y} \\ 0 & 0 & 1 & \frac{\partial u}{\partial z} \end{vmatrix},$$

that is

$$N := \frac{1}{\sqrt{\left| \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 - \left(\frac{\partial u}{\partial z}\right)^2 - 1 \right|}} \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, -\frac{\partial u}{\partial z}, 1\right).$$

If we denote $x_1 = x,\ x_2 = y,\ x_3 = z$, the coefficients of the second fundamental form h_{ij} in the case when the normal is a space-like vector are

$$h_{ij} := -\left\langle N, \frac{\partial^2 f}{\partial x^i \partial x^j} \right\rangle_M = \frac{\dfrac{\partial^2 u}{\partial x_i \partial x_j}}{\sqrt{\left|\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial u}{\partial y}\right)^2 - \left(\dfrac{\partial u}{\partial z}\right)^2 - 1\right|}}$$

and the defined 3-Minkowski curvature at $f(p)$, where $p = (x, y, z)$ is

$$K_f^M(p) := -\frac{det h_{ij}}{det g_{ij}} = -\frac{\sum_{\sigma\in\Sigma_3} \varepsilon(\sigma) \dfrac{\partial^2 u}{\partial x_1 \partial x_{\sigma(1)}} \dfrac{\partial^2 u}{\partial x_2 \partial x_{\sigma(2)}} \dfrac{\partial^2 u}{\partial x_3 \partial x_{\sigma(3)}}}{\epsilon \left|\left(\dfrac{\partial u}{\partial x_1}\right)^2 + \left(\dfrac{\partial u}{\partial x_2}\right)^2 - \left(\dfrac{\partial u}{\partial x_3}\right)^2 - 1\right|^{\frac{5}{2}}}.$$

If the normal is a timelike vector the sign $-$ in previous formulas becomes $+$. The distance from origin to the tangent plane at $f(p)$ is

$$d_f^M(p) = \frac{\left|\sum_{i=1}^3 x_i \dfrac{\partial u}{\partial x_i} - u(x_1, x_2, x_3)\right|}{\sqrt{\left|\left(\dfrac{\partial u}{\partial x}\right)^2 + \left(\dfrac{\partial u}{\partial y}\right)^2 - \left(\dfrac{\partial u}{\partial z}\right)^2 - 1\right|}};$$

therefore, the Minkowski affine radius is modulo a sign the ratio

$$\frac{K_f^M(p)}{(d^M(p))^5} = -\frac{\sum_{\sigma\in\Sigma_3} \varepsilon(\sigma) \dfrac{\partial^2 u}{\partial x_1 \partial x_{\sigma(1)}} \dfrac{\partial^2 u}{\partial x_2 \partial x_{\sigma(2)}} \dfrac{\partial^2 u}{\partial x_3 \partial x_{\sigma(3)}}}{\left|\sum_{i=1}^3 x_i \dfrac{\partial u}{\partial x_i} - u(x_1, x_2, x_3)\right|^5}.$$

It becomes clear that $\dfrac{\partial^2 u}{\partial x_i \partial x_{\sigma(i)}}$ is the value of the determinant

$$\begin{vmatrix} 0\ 0\ 0 & \dfrac{\partial^2 u}{\partial x_i \partial x_{\sigma(i)}} \\ 1\ 0\ 0 & \dfrac{\partial u}{\partial x} \\ 0\ 1\ 0 & \dfrac{\partial u}{\partial y} \\ 0\ 0\ 1 & \dfrac{\partial u}{\partial z} \end{vmatrix}$$

and $\left|\sum_{i=1}^{3} x_i \dfrac{\partial u}{\partial x_i} - u(x_1, x_2, x_3)\right|$ is the absolute value of the determinant

$$\begin{vmatrix} x\ y\ z & u(x, y, z) \\ 1\ 0\ 0 & \dfrac{\partial u}{\partial x} \\ 0\ 1\ 0 & \dfrac{\partial u}{\partial y} \\ 0\ 0\ 1 & \dfrac{\partial u}{\partial z} \end{vmatrix}.$$

At the numerator, we have six terms as products of 4-determinants; therefore, the dimension is the third power of a 4-volume. At the denominator, we have the 5th power of a 4-volume, therefore the affine radius has, as dimension at denominator; the second power of a 4-volume, therefore

$$\sqrt[4]{\left|\frac{K_f\,(p)}{d_f^5\,(p)}\right|}$$

is measured in $\dfrac{1}{(m)^2}$.

After a centro-affine transformation $\mathbb{A} \in \mathcal{M}_4(\mathbb{R})$ of the surface f, the connection between the two affine radii is

$$\frac{\bar{K}^M_{\bar{f}}}{(d^M{}_{\bar{f}})^5} = \frac{(det\mathbb{A})^3}{(det\mathbb{A})^5} \cdot \frac{K^M_f}{(d^M{}_f)^5};$$

therefore, the same relation holds

$$\frac{\bar{K}^M_{\bar{f}}}{(d^M{}_{\bar{f}})^5} = \frac{1}{(det\mathbb{A})^2} \cdot \frac{K^M_f}{(d^M{}_f)^5}.$$

Now we can understand how it can be extended to any dimension.

12.5 Anti-de Sitter Space–Times as Affine Hypersurfaces. Their Cosmological Constant and Its Connection with the Affine Radius

All de Sitter space–times are affine hypersurfaces whose cosmological constants are connected to their affine radii (see [31]). A similar conclusion can be seen in the case of Anti-de Sitter space–times. Let's see why.

In the case $n = 3$, the Anti-de Sitter space–time here denoted $AdS(2, 3)$, is the Minkowski time-like sphere of $M^{(2,3)}$ having the equation

$$X_0^2 + X_1^2 - X_2^2 = -a^2.$$

The two numbers involved in the above notation $AdS(2, 3)$ are 2, from the dimension of the space-like sphere and 3 from the total dimension of the ambient Minkowski space.

We met it when we studied non-Euclidean geometries, more precisely, the hyperboloid model.

For this Anti-de Sitter space–time, we have, according to the Einstein theorem for surfaces

$$R_{ij} = K_f^M \, g_{ij}$$

therefore

$$R_{ij} + \frac{1}{a^2} \, g_{ij} = 0.$$

This last equation can be written as

$$R_{ij} - \frac{1}{2} R \, g_{ij} = 0,$$

that is $\Lambda = 0$ and $T_{ij} = 0$. R is still depending on the centro-affine invariant

$$\sqrt[3]{\left| \frac{K_f(p)}{d_f^4(p)} \right|}.$$

In the case $n = 4$, the Anti-de Sitter space–time, called $AdS(3, 4)$, is the Minkowski space-like sphere of $M^{(2,4)}$ having the equation

$$X_0^2 + X_1^2 - X_2^2 - X_3^2 = -a^2.$$

The standard parameterization is

$$f(t, x_1, x_2) = (a \sinh t \cos x_1, a \sinh t \sin x_1, a \cosh t \cos x_2, a \cosh t \sin x_2).$$

The metric is

$$ds_3^2 = (a^2dt^2 + a^2\sinh^2 t\, dx_1^2) - a^2\cosh^2 t dx_2^2.$$

We observe

$$ds_3^2 = ds_2^2 - a^2\cosh^2 t\, dx_2^2.$$

The non-zero Christoffel symbols are

$$\Gamma^0_{11} = -\sinh t\cosh t,\ \ \Gamma^0_{22} = \sinh t\cosh t,$$

$$\Gamma^1_{01} = \Gamma^1_{10} = \coth t,$$

$$\Gamma^2_{02} = \Gamma^2_{20} = \tanh t.$$

Now, if we compute

$$R_{ii} = R^s_{isi} = \frac{\partial\Gamma^s_{ii}}{\partial x^s} - \frac{\partial\Gamma^s_{is}}{\partial x^i} + \Gamma^h_{ii}\Gamma^s_{hi} - \Gamma^h_{is}\Gamma^s_{hi},$$

we find

$$R_{00} = -2;\ \ R_{11} = -2\sinh^2 t;\ \ R_{22} = 2\cosh^2 t.$$

The other Ricci symbols are null, $R_{ij} = 0, i \neq j$. Therefore

$$R_{ij} + \frac{2}{a^2}\, g_{ij} = 0.$$

If we compute $R := R^i_i$, taking into account $R^i_j = g^{is}R_{sj}$, it results $R = -\dfrac{6}{a^2}$. The left-hand side of Einstein's field equations becomes

$$R_{ij} - \frac{1}{2}\left(-\frac{6}{a^2}\right) g_{ij} + \Lambda g_{ij}.$$

If we choose $\Lambda = -\dfrac{1}{a^2}$, the left hand becomes

$$R_{ij} + \frac{2}{a^2}\, g_{ij},$$

that is the left hand becomes 0. The de Sitter space–time presented above satisfies the Einstein field equations

$$R_{ij} - \frac{1}{2}R\, g_{ij} + \Lambda\, g_{ij} = 8\pi G\, T_{ik}$$

for $R = -\frac{6}{a^2}$, $\Lambda = -\frac{1}{a^2}$ and $T_{ij} = 0$.

A space–time without matter appears as we expected.

The normal is

$$N = (\sinh t \cos x_1, \sinh t \sin x_1, \cosh t \cos x_2, \cosh t \sin x_2) = \frac{1}{a} f,$$

therefore, it is

$$K_f^M(p) := -\frac{\det h_{ij}}{\det g_{ij}} := -\frac{1}{a^3}.$$

It is easy to see that $\left\langle f, \frac{\partial f}{\partial x^i} \right\rangle_M = 0,\ i \in \{0, 1, 2\}$, that is $d_f^M(p) = a$. The Minkowski sphere becomes a Tzitzeica–Minkowski affine sphere because

$$\frac{K_f^M(p)}{(d_f^M(p))^5} = -\frac{1}{a^8}.$$

According to both the affine invariant meaning and the fact that the quantity is measured in $\frac{1}{(m)^2}$, the geometric quantity

$$\sqrt[4]{\left| \frac{K_f^M(p)}{(d_f^M(p))^5} \right|}$$

is involved in the dimensional constants of the Anti-de Sitter $AdS(3, 4)$ space–time. So also in its cosmological constant.

In the case of Minkowski $M^{(2,5)}$ space, the parameterization of the $AdS(4, 5)$ space–time

$$X_0^2 + X_1^2 - X_2^2 - X_3^2 - X_4^2 = -a^2.$$

is

$$\begin{cases} X_0 = a \sinh t \cos x_1 \\ X_1 = a \sinh t \sin x_1 \\ X_2 = a \cosh t \cos x_2 \\ X_3 = a \cosh t \ \sin x_2 \cos x_3 \\ X_4 = a \cosh t \sin x_2 \sin x_3 \end{cases}$$

The metric related to this parameterization is

$$ds_4^2 = (a^2 dt^2 + a^2 \sinh^2 t dx_1^2 - a^2 \cosh^2 t dx_2^2) - a^2 \cosh^2 t \sin^2 x_2\, dx_3^2.$$

As in the previous case, we obtain

$$ds_4^2 = ds_3^2 - a^2 \cosh^2 t \sin^2 x_2 dx_3^2.$$

Again, if we compute

$$R_{ii} = R^s_{isi} = \frac{\partial \Gamma^s_{ii}}{\partial x^s} - \frac{\partial \Gamma^s_{is}}{\partial x^i} + \Gamma^h_{ii}\Gamma^s_{hi} - \Gamma^h_{is}\Gamma^s_{hi},$$

we find

$$R_{00} = -3,\ R_{11} = -3\sinh^2 t,\ R_{22} = 3\cosh^2 t,\ R_{33} = 3\cosh^2 t \sin^2 x_2,$$

that is

$$R_{ij} + \frac{3}{a^2}\, g_{ij} = 0,$$

which leads to

$$R = -\frac{12}{a^2},\ \Lambda = -\frac{3}{a^2},\ T_{ij} = 0$$

for the Einstein's field equations.

As in the previous $n = 4$ case, the Minkowski normal to the 4-hypersurface is

$$N(t, x_1, x_2, x_3) = \frac{1}{a} f(t, x_1, x_2, x_3),$$

a space-like vector.

It means that the Minkowski distance from the origin to the tangent 4-hyperplane, at a given point of the 4-hypersurface, is a and all the coefficients of the second fundamental form have the property

$$h_{ij} = \frac{1}{a} g_{ij}.$$

Therefore,

$$K_f^M = -\frac{\det h_{ij}}{\det g_{ij}} = -\frac{1}{a^4}$$

and

$$\frac{K_f^M(p)}{(d_f^M(p))^6} = -\frac{1}{a^{10}},$$

that is the absolute value of its 5th root establishes the invariant involved in the coefficients of the Einstein field equations.

Let us consider now $x_1 = 0$, which implies $X_1 = 0$. The previous description of the Anti-de Sitter $AdS(4, 5)$ space–time

$$X_0^2 + X_1^2 - X_2^2 - X_3^2 - X_4^2 = -a^2\,,$$

in the case of the Minkowski $M^{(2,5)}$ space, leads to the parameterization

$$\begin{cases} X_0 = a \sinh t \\ X_2 = a \cosh t \cos x_2 \\ X_3 = a \cosh t \ \sin x_2 \cos x_3 \\ X_4 = a \cosh t \sin x_2 \sin x_3 \end{cases}$$

for the surface

$$X_0^2 - X_2^2 - X_3^2 - X_4^2 = -a^2,$$

which is a de Sitter space–time we can call $dS(3, 4)$, whose ambient is a Minkowski $M^{(1,4)}$ space whose metric signature obviously is $(+ - --)$.

The metric attached to this parameterization is

$$ds^2 = a^2 dt^2 - a^2 \cosh^2 t dx_2^2 - a^2 \cosh^2 t \sin^2 x_2 \, dx_3^2.$$

We obtain

$$R_{ij} + \frac{2}{a^2} \, g_{ij} = 0,$$

that is the de Sitter space–time presented above satisfies the Einstein field equations

$$R_{ij} - \frac{1}{2} R \, g_{ij} + \Lambda \, g_{ij} = 8\pi G T_{ik}$$

for $R = -\dfrac{6}{a^2}$, $\Lambda = -\dfrac{1}{a^2}$ and $T_{ij} = 0$ exactly as $AdS(3, 4)$. Let us observe that if we consider $X_4 = 0$ (for $x_3 = 0$), we obtain the inclusion $AdS(3, 4) \subset AdS(4, 5)$.

This structure inclusion,

$$dS(3, 4) = AdS(4, 5) \cap \{X_2 = 0\} \subset AdS(4, 5) \subset M^{(2,5)}$$

has an important physical consequence.

Light can travel in $AdS(4, 5)$ because it can travel in any $dS(2, 3)$ (choosing X_0 and two among the three variables X_2, X_3, X_4), as we have seen in the previous chapter or in citeCallahan,Georges1.

In the general case, the Anti-de Sitter $AdS(n - 1, n)$ space–time is the Minkowski $(n - 1)$-sphere determined by the ends of all the space-like vectors with Minkowski length a.

This is a hypersurface of the Minkowski n-dimensional space $M^{(2,n)}$ having the algebraic equation

$$X_0^2 + X_1^2 - ... - X_{n-1}^2 = -a^2.$$

The parameterization we present is

$$\begin{cases} X_0 = a \sinh t \cos x_1 \\ X_1 = a \sinh t \sin x_1 \\ X_2 = a \cosh t \cos x_2 \\ X_3 = a \cosh t \sin x_2 \cos x_3 \\ X_4 = a \cosh t \sin x_2 \sin x_3 \cos x_4 \\ X_5 = a \cosh t \sin x_2 \sin x_3 \sin x_4 \cos x_5 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_{n-3} = a \cosh t \sin x_2 \sin x_3 \ldots \sin x_{n-4} \cos x_{n-3} \\ X_{n-2} = a \cosh t \sin x_2 \sin x_3 \ldots \sin x_{n-3} \cos x_{n-2} \\ X_{n-1} = a \cosh t \sin x_2 \sin x_3 \ldots \sin x_{n-3} \sin x_{n-2} \end{cases}$$

This parameterization makes sense for $n \geq 4$.
We obtain

$$ds_{n-1}^2 = ds_{n-2}^2 - a^2 \sin^2 x_2 \sin^2 x_3 \ldots \sin^2 x_{n-3}\, dx_{n-2}^2,$$

a formula which is the generalization of the formulas obtained for the previous cases $n = 4$ and $n = 5$.

Therefore, in all cases, we proved that the metric is a diagonal one.

Now, let us observe something else, a geometric property used in [31], which allows to substitute very complicated computations in order to achieve the main formula necessary to compute the cosmological constant value. We are talking about

$$R_{jl} = -\frac{n-2}{a^2} g_{jl}.$$

We have $\left\langle f, \frac{\partial f}{\partial t} \right\rangle_M = 0$, $\left\langle f, \frac{\partial f}{\partial x_k} \right\rangle_M = 0$, while $\left\langle \frac{\partial f}{\partial x_k}, \frac{\partial f}{\partial t} \right\rangle_M = 0$, $\left\langle \frac{\partial f}{\partial x_k}, \frac{\partial f}{\partial x_j} \right\rangle_M = 0$ are the consequences of the diagonal form of the metric and highlight the orthogonal frame of the tangent space at each point.

Therefore, the Minkowski normal to the hypersurface is

$$N(t, x_1, \ldots, x_{n-2}) = \frac{1}{a} f(t, x_1, \ldots, x_{n-2}),$$

that is the Minkowski distance from the origin to the tangent hyperplane at a given point of the hypersurface is a and all the coefficients of the second fundamental form are computed with the formula established for the case $n = 2$,

$$h_{ij} = \left\langle \frac{\partial N}{\partial x_i}, \frac{\partial f}{\partial x_j} \right\rangle_M,$$

therefore

$$h_{ij} = \frac{1}{a} g_{ij}.$$

Since $\langle N, N\rangle_M = -1 < 0$, we have

$$R_{ijkl} = -\left(h_{ik}h_{jl} - h_{il}h_{jk}\right).$$

It results

$$R_{ijkl} = -\frac{1}{a^2}\left(g_{ik}g_{jl} - g_{il}g_{jk}\right),\ i, j, k, l \in \{0, 1, ..., n-2\}.$$

Therefore, each sectional curvature is

$$K = -\frac{1}{a^2}$$

and

$$R^m_{jkl} = g^{mi}R_{ijkl} = -\frac{1}{a^2}\left(\delta^m_k g_{jl} - \delta^m_l g_{jk}\right).$$

Finally,

$$R_{jl} = \sum_{m=0}^{n-2} R^m_{jml} = -\frac{1}{a^2}\sum_{m=0}^{n-2}(g_{jl} - \delta^m_l g_{jm}) = -\frac{n-2}{a^2}g_{jl}.$$

Since

$$R_{ij} + \frac{1}{2}(n-1)(n-2)\frac{1}{a^2}\ g_{ij} - \frac{(n-2)(n-3)}{2}\frac{1}{a^2}\ g_{ij} = R_{ij} + \frac{n-2}{a^2}\ g_{ij} = 0,$$

it results that if we choose

$$\Lambda = -\frac{(n-2)(n-3)}{2}\frac{1}{a^2},$$

the previous metric satisfies the Einstein field equations

$$R_{ij} - \frac{1}{2}R\ g_{ij} + \Lambda\ g_{ij} = 8\pi G T_{ij}$$

in the absence of matter, that is with $T_{ij} = 0$. At the same time,

$$K^M_f := -\frac{\det h_{ij}}{\det g_{ij}} = -\frac{1}{a^{n-1}};\ d^M_f := a.$$

The generalization is immediate, for the $AdS(n-1, n)$, the constants are written with respect to the Minkowski-affine invariant

$$\sqrt[n]{\left|\frac{K_f^M(p)}{(d_f^M(p))^{n+1}}\right|} = \frac{1}{a^2}.$$

Therefore, this affine invariant is also involved in the cosmological constant of this space–time.

Let us observe that the structure inclusion,

$$dS(n-2, n-1) = AdS(n-1, n) \cap \{X_2 = 0\} \subset AdS(n-1, n) \subset M^{(2,n)}$$

allows light to travel in $AdS(n-1, n)$ because it can travel in any $dS(2, 3) \subset dS(n-2, n-1)$ (choosing X_0 and two among the components $X_2, X_3, ...X_{n-1}$. So, generally, we can claim that light travels in any Anti-de Sitter space–time because it travels in all de Sitter space–times included in the chosen Anti-de Sitter space–time.

Chapter 13
More Than Metric: Geometric Objects for Alternative Pictures of Gravity

Everything we call real is made of things that cannot be regarded as real.

Niels Bohr

In the previous discussions, we presented General Relativity as a theory essentially based on metric as the main object capable of being related to physical observables. Thanks to the Equivalence Principle, the Einstein formulation is a "metric theory of gravity" where causal structure (related to metric and light cones) and geodesic structure (related to geodesics and motions along them) coincide. This is not the only possibility and more general formulations of gravitational interaction can require other mathematical tools besides the metric [49, 55]. In other words, metric may have an "ancillary role" instead of being the main object of investigation as shortly discussed in Chap. 10. The need to extend or modify the Einstein theory comes mainly from several issues related to Physics (see, for example, [45, 61] for a discussion). We intend to present here, in a more abstract way, the geometric objects considered in the chapters devoted to Differential Geometry. We will see that the metric may not even be involved in the description of gravity. After presenting basic results about differentiable manifolds, tensors, exterior forms, differential forms, vector fields, and general affine connections, we will take into account a special affine connection, the Levi-Civita one. This connection is related to metric and then involves the Equivalence Principle. Then General Relativity is only a particular theory in the wide world of theories of gravity. Concepts as tetrad fields, the commutator of covariant derivatives, and the spin connection will give us the tools to develop metric-affine theories and then alternative pictures as Teleparallel Gravity.

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_13

13.1 Differentiable Manifolds

Let us remember the way we studied the unit sphere in the chapter devoted to surfaces. "The algebraic point of view is related to equations defining manifolds. $X^2 + Y^2 + Z^2 = 1$ is the algebraic definition of a sphere centred at the origin with radius 1. In Differential Geometry, we deal with smooth functions describing a surface.

The previous sphere can be seen as the smooth function $f : (0, \pi) \times (0, 2\pi) \to \mathbb{E}^3$,

$$f(x^1, x^2) = (\sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1),\ x^1 \in (0, \pi),\ x^2 \in (0, 2\pi)''.$$

Now, let us observe that the entire sphere is not the image of the previous function. The meridian arc between the North Pole $N(0, 0, 1)$ and the South Pole $S(0, 0, -1)$ corresponding to $x^2 = 0$ is missing. We have obtained important results about the Differential Geometry of the unit sphere, but all results were obtained for the unit sphere minus the NS meridian arc. Therefore all geodesics of the unit sphere have a point which is missing, at least one. Of course, we understood the power of changing coordinates. Rotations can be used to give information about another image of the sphere, therefore the missing points can be included in the image of geodesics. The same rotations can be used to obtain information about Gaussian curvature of a missing point. Somehow, in our mind, it is clear that we can obtain the missing point information using a suitable coordinate transformation. Can we write all these in a formal mathematical language? The answer is yes and it is related to the notion of differentiable manifold. In simple words, a differentiable manifold is something which is locally similar to a system of coordinates to allow us to apply the calculus and the formulas seen in the chapter on Differential Geometry. This can be obtained using a collection of charts called atlas.

In order to start in formalizing the problem, let us consider the Euclidean n-dimensional space E^n. As we know, there is the Euclidean inner product which generates the Euclidean norm $||\cdot||$ of vectors and the Euclidean distance d between two points. An open ball $B_r(x)$ with x as centre and r as radius is the set

$$B_r(x) := \{y \in E^n |\, d(x, y) < r\}.$$

Taking into account the Euclidean n-dimensional space E^n, we can say that the Euclidean topology on $\mathbb{R}^n$ is the topology generated by these balls. Therefore the open sets of the Euclidean topology on $\mathbb{R}^n$ consist of arbitrary unions of open balls $B_r(x),\ r > 0,\ x \in \mathbb{R}^n$.

We denote by $\mathcal{T}$ the topology on $\mathbb{R}^n$ described above. It will be used to define an atlas $\mathbb{R}^n$-type as we will understand from the following definition.

Definition 13.1.1 Let M be an arbitrary non-empty set. Consider a set of indexes denoted by A. A smooth atlas on M is a collection of charts $\mathcal{A} = \{(U_a, h_a)\ |\, a \in A\}$ such that:

(A_1) $U_a \subset M$, for all $a \in A$ and $\bigcup_{a \in A} U_a = M$;

(A_2) $h_a : U_a \longrightarrow \mathbb{R}^n$ is an injective function, i.e. it maps distinct elements into distinct elements;

(A_3) $h_a \left(U_a \bigcap U_b \right)$ is an open set of $\mathbb{R}^n$ for all $a, b \in A$ satisfying $U_a \bigcap U_b \neq \emptyset$;

(A_4) For all $a, b \in A$ such that $U_a \bigcap U_b \neq \emptyset$, the chart change defined by the map $h_b \circ h_a^{-1} : h_a \left(U_a \bigcap U_b \right) \longrightarrow \mathbb{R}^n$ is a smooth one.

Definition 13.1.2 The set M together with the smooth atlas $\mathcal{A}$ is called a differentiable manifold.

Let us understand the previous definitions looking at an example. Of course, we can choose the previous sphere considering a differentiable manifold structure. In this way, we understand the preliminaries made before the definitions.

Example 13.1.3 Consider $S^2 = \left\{ (x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1 \right\}$ and the stereographic projections on the equatorial plane corresponding to the points $N\,(0, 0, 1)$ and $S\,(0, 0, -1)$ having the form

$$h_N : S^2 \backslash \{N\} \longrightarrow \mathbb{R}^2, \; h_N\,(M) = \left(\frac{x_0}{1 - z_0}, \frac{y_0}{1 - z_0} \right) \text{ and}$$
$$h_S : S^2 \backslash \{S\} \longrightarrow \mathbb{R}^2, \; h_S\,(M) = \left(\frac{x_0}{1 + z_0}, \frac{y_0}{1 + z_0} \right).$$

The atlas $\mathcal{A} = \left\{ \left(S^2 \backslash \{N\}, h_N \right), \left(S^2 \backslash \{S\}, h_S \right) \right\}$ fulfills the conditions (A_1), (A_2) and (A_3) of the first definition. It remains to prove that a $h_S \circ h_N^{-1} : \mathbb{R}^2 \backslash \{(0, 0)\} \longrightarrow \mathbb{R}^2 \backslash \{(0, 0)\}$ is a smooth map.

Since $h_N^{-1} : \mathbb{R}^2 \backslash \{(0, 0)\} \longrightarrow S^2 \backslash \{N\}$ is

$$h_N^{-1}\,(X, Y) = \left(\frac{2X}{X^2 + Y^2 + 1}, \frac{2Y}{X^2 + Y^2 + 1}, \frac{X^2 + Y^2 - 1}{X^2 + Y^2 + 1} \right),$$

we have $h_S \circ h_N^{-1}\,(X, Y) = \left(\frac{X}{X^2 + Y^2}, \frac{Y}{X^2 + Y^2} \right)$ which obviously is a smooth map. In other words, it is a differentiable manifold structure to the unit sphere with all described implications.

Starting from this point all results can be done in the context of differentiable manifolds. To preserve our style, we continue to work with one chart of the atlas only. We preserve the name of this chart, i.e. system of coordinates, and the chart changes restricted only to our chart will be called as before, i.e. changes of coordinates.

13.2 Abstract Frame for Tensors, Exterior Forms, and Differential Forms

To succeed in giving an intuitive description, we start remembering some definitions used in the chapter devoted to surfaces.

In the $3D$ Euclidean space E^3, each surface is an image of a function $f : U \to E^3$, where U is an open set of $\mathbb{R}^2$.
• Denote by p a point of U.
• Therefore $p := (x^1, x^2)$.
• The vector tangents to the surface at the point $f(p)$ are $\frac{\partial f}{\partial x^1}$ and $\frac{\partial f}{\partial x^1}$ computed at p.
• The vector space generated by these vectors is called a tangent space at $f(p)$ to the surface f and it is denoted by $T_p f$.
• $T_p f$ contains all vectors v_p written in the form $v_p = v^1 \frac{\partial f}{\partial x^1}(p) + v^2 \frac{\partial f}{\partial x^1}(p)$.
• According to the *Theorema Egregium* by Gauss, the extra dimension necessary to obtain geometric information on the surface can be cancelled.

Following these considerations, let us see the above picture in a more abstract way, i.e. in the absence of the surface and increasing the number of dimensions.
• Let M be a set of real coordinates $(x^1, x^2, ..., x^n)$ in which we wish to develop geometric concepts.
• $p := (x^1, x^2, ..., x^n)$ is a given point of M.
• The tangent space at p will be denoted by $T_p M$ and consists in all vectors v_p such that they are described by

$$v_p = \sum_{k=1}^{n} v^k \left(\frac{\partial}{\partial x^k} \right)_p .$$

Of course, we can write a vector v_p using the Einstein summation rule,

$$v_p = v^k \left(\frac{\partial}{\partial x^k} \right)_p .$$

• At any p, the above vector can be identified by its components, i.e. $v_p = (v^1, v^2, ..., v^n)$.
• The tangent space $T_p M$ is a n-dimensional vector space.
• Denote by $TM := \bigsqcup_{p \in M} T_p M = \bigcup_{p \in M} \{p\} \times T_p M$ the tangent bundle of M.
• Denote by $\mathcal{F}(\mathcal{M})$ the set of all smooth functions on M to $\mathbb{R}$. Remember that these functions are smooth at each $p \in M$.
• For such a function $h : M \to \mathbb{R}$ at each $p \in N$, it makes sense the operation $v_p(h)$ defined by the rule

$$v_p(h) := \sum_{k=1}^{n} v^k \left(\frac{\partial h}{\partial x^k}\right)_p .$$

• The last operation allows, for each $p \in M$, to define the differential at p of the map $h \in \mathcal{F}(\mathcal{M})$.
$(dh)_p : T_pM \to \mathbb{R}$ is by definition

$$(dh)_p(v) := v_p(h).$$

• Denote by T_p^*M the dual of T_pM. It is called a cotangent space at p and contains all linear functions defined on T_pM to $\mathbb{R}$. Obviously, $(dh)_p \in T_p^*M$.
• Denote $(dx^k)_p(v) := v^k,\ k \in \{1, 2, ..., n\}$. Analogously, the set containing

$$\left(\frac{\partial}{\partial x^1}\right)_p, \left(\frac{\partial}{\partial x^2}\right)_p, ..., \left(\frac{\partial}{\partial x^n}\right)_p$$

is the basis of the vector space T_pM, while the set

$$(dx^1)_p, (dx^2)_p, ..., (dx^n)_p$$

is the basis of the vector space T_p^*M. This basis is the dual of that of T_pM.
• It is obvious that the differential of h at p is a linear function and it can be written in the form

$$(dh)_p = \sum_{k=1}^{n} \left(\frac{\partial h}{\partial x^k}\right)_p (dx^k)_p.$$

• Generally, a cotangent vector at p to M looks like

$$w_p = \sum_{i=1}^{n} w_i (dx^i)_p.$$

Starting from this point, we can work with general n-dimensional real vector spaces V and V^* instead T_pM and T_p^*M.

In fact, all these finite dimensional real vector spaces are isomorphic.

If $\{e_1, e_2, ..., e_n\}$ is the canonical basis of V and $\{p_1, p_2, ..., p_n\}$ is the canonical basis of V^*, the connection between them is $p_i(e_j) = \delta_{ij}$, exactly as in the case of $dx^i \left(\frac{\partial}{\partial x^j}\right)_p = \delta^i_j$.

More we talk about all these geometric objects and their properties without assuming that they can be also smooth functions. This property can be added at any moment when we need to describe aspects of these objects involving the smoothness. We will consider this aspect later in this section. The considerations below are made in the absence of smoothness.

Denote by

$$V^{0,m} = \{T \mid T : V \times V \times ... \times V \longrightarrow \mathbb{R},\ T \text{ is a m-linear map } \}$$

the set of m-covariant tensors. This set can be endowed with a vector space structure considering the operations

$$(T_1 \oplus T_2)(v_1, ..., v_m) := T_1 (v_1, ..., v_m) + T_2 (v_1, ..., v_m),$$

$$(\lambda T_1)(v_1, ..., v_m) := \lambda \cdot T_1 (v_1, ..., v_m),\ \lambda \in \mathbb{R},\ v_1, ..., v_m \in V.$$

A tensor product $\otimes : V^{0,m} \times V^{0,r} :\longrightarrow V^{0,m+r}$ can also be defined by the formula

$$(T \otimes S)(v_1, ..., v_m; u_1, ..., u_r) := T (v_1, ..., v_m) \cdot S (u_1, ..., u_r)$$

for $T \in V^{0,m},\ S \in V^{0,r}$ and $v_1, ..., v_m; u_1, ..., u_r \in V$.

Denote by S_m the set of permutations of m items.

Definition 13.2.1 A m-covariant tensor $\omega \in V^{0,m}$ is called an alternating $m-$ tensor if

$$\omega \left(v_{\pi(1)}, ..., v_{\pi(m)}\right) = \varepsilon(\pi) \cdot \omega (v_1, ..., v_m)$$

for all the permutations $\pi \in S_m$ and $v_1, ..., v_m \in V$.

By definition,

$$\Lambda^m \left(\mathbb{R}^n, \mathbb{R}\right) := \left\{\omega \in V^{0,m} \mid \omega \text{ is a m-alternating tensor}\right\}$$

is the set of all m alternating tensors.

The previous vector space operations on $V^{0,m}$ transform $\Lambda^m (\mathbb{R}^n, \mathbb{R})$ into a vector subspace of the vector space $\left(V^{0,m}, \oplus, \cdot_\lambda\right) / \mathbb{R}$.

Definition 13.2.2 The map $\mathcal{A} : V^{0,m} \longrightarrow \Lambda^m (\mathbb{R}^n, \mathbb{R})$,

$$\mathcal{A}(\omega)(v_1, ..., v_m) := \sum_{\pi \in S_m} \varepsilon(\pi) \cdot \omega \left(v_{\pi(1)}, ..., v_{\pi(m)}\right)$$

is called an alternating multilinear map.

It allows us to define the exterior product formula

$$\omega_1 \wedge \omega_2 := \mathcal{A}(\omega_1 \otimes \omega_2),$$

for all $\omega_1 \in \Lambda^m(\mathbb{R}^n, \mathbb{R})$ and $\omega_2 \in \Lambda^r(\mathbb{R}^n, \mathbb{R})$. Since

$$(\omega_1 \wedge \omega_2)(v_1, ..., v_{m+r}) = \sum_{\pi \in S_{m+r}} \varepsilon(\pi)\, \omega_1\left(v_{\pi(1)}, ..., v_{\pi(m)}\right) \cdot \omega_2\left(v_{\pi(m+1)}, ..., v_{\pi(m+r)}\right),$$

it results in $\omega_1 \wedge \omega_2 \in \Lambda^{m+r}(\mathbb{R}^n, \mathbb{R})$.

Therefore the exterior product is obtained by the alternatization (we may say also alternatization) of the tensor product.

Some textbooks define the wedge product in the form

$$\omega_1 \wedge \omega_2 := \frac{(m+r)!}{m!r!}\mathcal{A}(\omega_1 \otimes \omega_2),$$

for all $\omega_1 \in \Lambda^m(\mathbb{R}^n, \mathbb{R})$ and $\omega_2 \in \Lambda^r(\mathbb{R}^n, \mathbb{R})$.

This wedge product is often called "exterior product". If one compares to our definition, it is only a constant which makes the difference. We continue to use our definition.

Let us observe that V^* is in fact $V^{0,1}$.

The same, we can observe that if $\alpha \in V^*,\ \beta \in V^*$, then $\alpha \otimes \beta \in V^{0,2}$.

There is an immediate generalization: if $\alpha_1 \in V^*, ..., \alpha_m \in V^*$, then $\alpha_1 \otimes ... \otimes \alpha_m \in V^{0,m}$.

This way the set $\left\{p_{i_1} \otimes p_{i_2} \otimes ... \otimes p_{i_m}\right\}_{1 \leqslant i_1, i_2, ..., i_m \leqslant n}$ becomes a vector basis of $V^{0,m}$, i.e.

$$\omega = \sum_{1 \leqslant i_1, i_2, ..., i_m \leqslant n} \omega_{i_1 i_2 ... i_m} \cdot p_{i_1} \otimes p_{i_2} \otimes ... \otimes p_{i_m}$$

for some coefficients $\omega_{i_1 i_2 ... i_m} \in \mathbb{R}$.

The following natural question occurs: which is the vector basis of $\Lambda^m(\mathbb{R}^n, \mathbb{R})$?

To answer, it is necessary to better understand the properties of the exterior product. Using the definition, we observe that

$$(\omega_1 \wedge \omega_2)(v_1, v_2) = \omega_1(v_1) \cdot \omega_2(v_2) - \omega_1(v_2) \cdot \omega_2(v_1)$$

for $\omega_1, \omega_2 \in \Lambda^1(\mathbb{R}^n, \mathbb{R})$.

It results in

$$\omega \wedge \omega = 0$$

for all $\omega \in \Lambda^1(\mathbb{R}^n, \mathbb{R})$.

Proposition 13.2.3 *The following equalities hold:*

(i) $(\lambda_1\omega_1 + \lambda_2\omega_2) \wedge \omega_3 = \lambda_1 \cdot \omega_1 \wedge \omega_3 + \lambda_2 \cdot \omega_2 \wedge \omega_3$
for any $\omega_1, \omega_2 \in \Lambda^m(\mathbb{R}^n, \mathbb{R})$, $\omega_3 \in \Lambda^r(\mathbb{R}^n, \mathbb{R})$, $\lambda_1, \lambda_2 \in \mathbb{R}$

(ii) $(\omega_1 \wedge \omega_2) \wedge \omega_3 = \omega_1 \wedge (\omega_2 \wedge \omega_3)$
for any $\omega_1 \in \Lambda^m(\mathbb{R}^n, \mathbb{R})$, $\omega_2 \in \Lambda^r(\mathbb{R}^n, \mathbb{R})$, $\omega_3 \in \Lambda^q(\mathbb{R}^n, \mathbb{R})$

(iii) $\omega_1 \wedge \omega_2 = (-1)^{m \cdot r} \cdot \omega_2 \wedge \omega_1$
for any $\omega_1 \in \Lambda^m(\mathbb{R}^n, \mathbb{R})$, $\omega_2 \in \Lambda^r(\mathbb{R}^n, \mathbb{R})$

Proof (i) and (ii) are proven by straight computation.
For (iii) we define the permutation $\sigma : \{1, ..., m+r\} \longrightarrow \{1, ..., m+r\}$,

$$\sigma(i) = \begin{cases} i+m, \ \text{dacă} 1 \leqslant i \leqslant r \\ i-r, \ \text{dacă} r+1 \leqslant i \leqslant m+r \end{cases}$$

with the following property related to its signature, $\varepsilon(\sigma) = (-1)^{m \cdot r}$.

Let us show that

$$\omega_1 \otimes \omega_2 = \sigma(\omega_2 \otimes \omega_1).$$

Successively we have

$$\sigma(\omega_2 \otimes \omega_1)(v_1, ..., v_{m+r}) = \omega_2\left(v_{\sigma(1)}, ..., v_{\sigma(r)}\right) \cdot \omega_1\left(v_{\sigma(r+1)}, ..., v_{\sigma(r+m)}\right) =$$

$$= \omega_2(v_{m+1}, ..., v_{m+r}) \cdot \omega_1(v_1, ..., v_m) = (\omega_1 \otimes \omega_2)(v_1, ..., v_{m+r}).$$

Since

$$(\sigma\mathcal{A})(\omega_1 \otimes \omega_2) = \varepsilon(\sigma) \cdot \mathcal{A}(\omega_2 \otimes \omega_1),$$

we obtain the desired result

$$\omega_1 \wedge \omega_2 = (-1)^{m \cdot r} \cdot \omega_2 \wedge \omega_1.$$

Proposition 13.2.4 *For any* $\omega_1 \in V^{0,1}, ..., \omega_m \in V^{0,1}$ *we have*

$$(\omega_1 \wedge ... \wedge \omega_m)(v_1, ..., v_m) = \det\left(\omega_i\left(v_j\right)\right)_{1 \leqslant i,j \leqslant m}.$$

Proof Directly,

$$(\omega_1 \wedge ... \wedge \omega_m)(v_1, ..., v_m) = \mathcal{A}(\omega_1 \otimes ... \otimes \omega_m)(v_1, ..., v_m) =$$

$$= \sum_{\sigma \in S_m} \varepsilon(\sigma) \cdot \omega_1\left(v_{\sigma(1)}\right) \cdot ... \cdot \omega_m\left(v_{\sigma(m)}\right) = \det\left(\omega_i\left(v_j\right)\right)_{1 \leqslant i,j \leqslant m}.$$

□

Corollary 13.2.5 *If* $\omega_i = \omega_j$ *for* $i \neq j$, *then* $\omega_1 \wedge ... \wedge \omega_m = 0$.

This happens because if in a determinant two lines or two columns are identical, the determinant is null. We can observe something extra, i.e. $\Lambda^m(\mathbb{R}^n, \mathbb{R}) = \{0\}$ if $m > n$.

Proposition 13.2.6 *If* $\omega_1, ..., \omega_m \in V^{0,1}$ *and* $1 \leqslant m \leqslant n = \dim \mathbb{R}^n/\mathbb{R}$, *the following assertions are equivalent:*
(i) $\{\omega_1, ..., \omega_m\}$ *are linear independent vectors;*
(ii) $\omega_1 \wedge ... \wedge \omega_m \neq 0$.

Proof $(i) \Rightarrow (ii)$

Since the set $\{\omega_1, ..., \omega_m\}$ consists of linear independent vectors, it exists $\{v_1, ..., v_m\}$ such that $\omega_i\left(v_j\right) = \delta_{ij}$.

It results in $(\omega_1 \wedge ... \wedge \omega_m)(v_1, ..., v_m) = \det\left(\omega_i\left(v_j\right)\right)_{1\leqslant i,j\leqslant m} = \det \delta_{ij} = 1$, therefore

$$(\omega_1 \wedge ... \wedge \omega_m) \neq 0.$$

$(ii) \Rightarrow (i)$

Ad absurdum, if $\omega_1 = \sum_{i=2}^{m} \alpha_i \omega_i$, it results in

$$\omega_1 \wedge ... \wedge \omega_m = \sum_{i=2}^{m} \alpha_i \omega_i \wedge \omega_2 \wedge ... \wedge \omega_m = 0,$$

in contrast with (ii).

□

Theorem 13.2.7 (i) *The set*

$$\left\{p_{i_1} \wedge p_{i_2} \wedge ... \wedge p_{i_m} \mid 1 \leqslant i_1 < i_2 < ... < i_m \leqslant n\right\}$$

induced by the dual vector basis of V^* is a vector basis for $\Lambda^m(\mathbb{R}^n, \mathbb{R})$.

(ii) $\dim \Lambda^m(\mathbb{R}^n, \mathbb{R}) = C_m^n = \dfrac{n!}{m!(n-m)!}$ *(combination of n taken m).*

Proof Using the previous proposition, it is sufficient to prove the following equality:

$$\left(\sum_{1\leqslant i_1<i_2<...<i_m\leqslant n} \alpha_{i_1...i_m} \cdot p_{i_1} \wedge p_{i_2} \wedge ... \wedge p_{i_m}\right)\left(e_{i_1}, ..., e_{i_m}\right) =$$
$$= \alpha_{i_1...i_m} \cdot \det\left(p_{i_k}\left(e_{i_k}\right)\right) = \alpha_{i_1...i_m}.$$

□

Definition 13.2.8 Consider an open set $U \subset \mathbb{R}^n$. A smooth map

$$\omega : U \longrightarrow \Lambda^m\left(\mathbb{R}^n, \mathbb{R}\right)$$

is called a m-differential form on the set U.

The set of m-differential forms on $U \subset \mathbb{R}^n$ is denoted by $\Omega^m(U)$ and can be endowed with a real vector space structure whose dimension in C_m^n.

First, let us observe that $\Omega^0(U) = \mathcal{F}(\mathrm{U})$, i.e. the 0-differential forms on U are coincident with the smooth maps on U.

If $\omega \in \Omega^m(U)$ and $x \in U$ we have

$$\omega(x) = \sum_{1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n} \omega_{i_1 \ldots i_m}(x) \cdot p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_m},$$

where $p_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, defined by $p_i(x) = p_i(x_1, \ldots, x_n) := x_i$ are the canonical projections and the maps $\omega_{i_1 \ldots i_m} : U \longrightarrow \mathbb{R}$ are smooth on the entire domain of definition $U \subset \mathbb{R}^n$. Taking into consideration that the canonical projections are linear maps, it results $dp_i = p_i' = p_i$.

Some important particular cases are

(a) If $\omega \in \Omega^1(U)$, then $\omega = \sum_{i=1}^{n} \omega_i \; p_i$.

(b) If $\omega \in \Omega^2(U)$, then $\omega = \sum_{1 \leqslant i < j \leqslant n} \omega_{ij} \; p_i \wedge p_j$.

(c) If $\omega \in \Omega^n(U)$, then $\omega = f \cdot p_1 \wedge p_2 \wedge \ldots \wedge p_n$.

Definition 13.2.9 The map $d : \Omega^m(U) \longrightarrow \Omega^{m+1}(U)$ such that

(i) $d\omega := df = \sum_{k=1}^{n} \frac{\partial f}{\partial x^k} p_k$ if ω is the 0-differential form $f \in \Omega^0(U)$;

(ii) $d\omega := \sum_{1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n} d\omega_{i_1 \ldots i_m} \wedge p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_m}$ if ω is the m-differential form $\omega = \sum_{1 \leqslant i_1 < i_2 < \ldots < i_m \leqslant n} \omega_{i_1 \ldots i_m} \cdot p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_m}$ is called the exterior derivative of the differential form ω.

The definition implies that the exterior derivative is a linear map.

Let us see some examples:

(1) If $U = \overset{o}{U} \subset \mathbb{R}^2$ and $\omega = \omega_1 p_1 + \omega_2 p_2 \in \Omega^1(U)$, then

$$d\omega = \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2} \right) p_1 \wedge p_2.$$

Indeed, the previous definition leads to

$$d\omega = d\omega_1 \wedge p_1 + d\omega_2 \wedge p_2 = \left(\sum_{i=1}^{2} \frac{\partial \omega_1}{\partial x^i} p_i \right) \wedge p_1 + \left(\sum_{i=1}^{2} \frac{\partial \omega_2}{\partial x^i} p_i \right) \wedge p_2 =$$

$$= \frac{\partial \omega_1}{\partial x^2} \cdot p_2 \wedge p_1 + \frac{\partial \omega_2}{\partial x^1} \cdot p_1 \wedge p_2 = \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2} \right) p_1 \wedge p_2.$$

(2) The same way from $\omega = \omega_1 p_1 + \omega_2 p_2 + \omega_3 p_3 \in \Omega^1(U)$, where $U = \overset{o}{U} \subset \mathbb{R}^3$. We obtain:

$$d\omega = \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2} \right) p_1 \wedge p_2 + \left(\frac{\partial \omega_3}{\partial x^2} - \frac{\partial \omega_2}{\partial x^3} \right) p_2 \wedge p_3 + \left(\frac{\partial \omega_3}{\partial x^1} - \frac{\partial \omega_1}{\partial x^3} \right) p_1 \wedge p_3.$$

(3) If $U = \overset{o}{U} \subset \mathbb{R}^3$ and $\omega = \omega_1 \cdot p_2 \wedge p_3 + \omega_2 \cdot p_3 \wedge p_1 + \omega_3 \cdot p_1 \wedge p_2 \in \Omega^2(U)$, then

$$d\omega = \left(\sum_{i=1}^{3} \frac{\partial \omega_i}{\partial x^i} \right) p_1 \wedge p_2 \wedge p_3.$$

Again, applying the definition, we have

$$d\omega = d\omega_1 \wedge p_2 \wedge p_3 + d\omega_2 \wedge p_3 \wedge p_1 + d\omega_3 \wedge p_1 \wedge p_2 =$$

$$= \left(\sum_{i=1}^{3} \frac{\partial \omega_1}{\partial x^i} \cdot p_i \right) \wedge p_2 \wedge p_3 + \left(\sum_{i=1}^{3} \frac{\partial \omega_2}{\partial x^i} \cdot p_i \right) \wedge p_3 \wedge p_1 + \left(\sum_{i=1}^{3} \frac{\partial \omega_3}{\partial x^i} \cdot p_i \right) \wedge p_1 \wedge p_2 =$$

$$= \frac{\partial \omega_1}{\partial x^1} \cdot p_1 \wedge p_2 \wedge p_2 + \frac{\partial \omega_2}{\partial x^2} \cdot p_2 \wedge p_3 \wedge p_1 + \frac{\partial \omega_3}{\partial x^3} \cdot p_3 \wedge p_1 \wedge p_2 =$$

$$= \left(\frac{\partial \omega_1}{\partial x^1} + \frac{\partial \omega_2}{\partial x^2} + \frac{\partial \omega_3}{\partial x^3} \right) \cdot p_1 \wedge p_2 \wedge p_3.$$

Definition 13.2.10 For $\alpha \in \Omega^r(U)$ and $\beta \in \Omega^m(U)$, we define $\alpha \wedge \beta$ by

$$(\alpha \wedge \beta)(x) := \alpha(x) \wedge \beta(x), \quad \text{for any } x \in U.$$

Proposition 13.2.11 *The following formula holds:*

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + (-1)^r \cdot \alpha \wedge d\beta.$$

Proof Consider

$$\alpha = \sum_{1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant n} \alpha_{i_1 \ldots i_r} \cdot p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_r}$$

and

$$\beta = \sum_{1 \leqslant j_1 < j_2 < \ldots < j_m \leqslant n} \beta_{j_1 \ldots j_m} \cdot p_{j_1} \wedge p_{j_2} \wedge \ldots \wedge p_{j_m}.$$

To simplify the next long formulas, we cancel the summation indexes $1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant n$ and $1 \leqslant j_1 < j_2 < \ldots < j_m \leqslant n$. Since

$$\alpha \wedge \beta = \sum \alpha_{i_1 \ldots i_r} \cdot \beta_{j_1 \ldots j_m} \cdot p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_r} \wedge p_{j_1} \wedge p_{j_2} \wedge \ldots \wedge p_{j_m},$$

it results in

$$d(\alpha \wedge \beta) = \sum d\alpha_{i_1 \ldots i_r} \cdot \beta_{j_1 \ldots j_m} \wedge p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_r} \wedge p_{j_1} \wedge p_{j_2} \wedge \ldots \wedge p_{j_m} +$$

$$+ \sum \alpha_{i_1 \ldots i_r} \cdot d\beta_{j_1 \ldots j_m} \wedge p_{i_1} \wedge p_{i_2} \wedge \ldots \wedge p_{i_r} \wedge p_{j_1} \wedge p_{j_2} \wedge \ldots \wedge p_{j_m} =$$

$$= \sum d\alpha_{i_1 \dots i_r} \wedge p_{i_1} \wedge p_{i_2} \wedge \dots \wedge p_{i_r} \wedge \beta_{j_1 \dots j_m} p_{j_1} \wedge p_{j_2} \wedge \dots \wedge p_{j_m} +$$

$$+ \sum \alpha_{i_1 \dots i_r} p_{i_1} \wedge p_{i_2} \wedge \dots \wedge p_{i_r} \wedge (-1)^r \, d\beta_{j_1 \dots j_m} \wedge p_{j_1} \wedge p_{j_2} \wedge \dots \wedge p_{j_m} = d\alpha \wedge \beta + (-1)^r \, \alpha \wedge d\beta.$$

□

Proposition 13.2.12 *If* $f \in \Omega^0(U)$ *then* $d^2 f = 0$.

Proof From

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} p_i$$

it results in

$$d^2 f = d(df) = \sum_{i<j} \left(\frac{\partial^2 f}{\partial x^j \partial x^i} - \frac{\partial^2 f}{\partial x^i \partial x^j} \right) p_i \wedge p_j = 0.$$

□

Theorem 13.2.13 *If* $\omega \in \Omega^m(U)$ *then* $d^2\omega = 0$.

Proof Using the linearity of the exterior derivative map d we can consider ω in the form

$$\omega = f \cdot p_{i_1} \wedge \dots \wedge p_{i_m}.$$

First of all we have

$$d\omega = df \wedge p_{i_1} \wedge \dots \wedge p_{i_m} = df \wedge p_{i_1} \wedge \dots \wedge p_{i_m},$$

therefore

$$d^2\omega = d(df) \wedge p_{i_1} \wedge \dots \wedge p_{i_m} - df \wedge d\left(1 \cdot p_{i_1} \wedge \dots \wedge p_{i_m}\right) =$$

$$= d^2 f \wedge p_{i_1} \wedge \dots \wedge p_{i_m} - df \wedge 0 \wedge p_{i_1} \wedge \dots \wedge p_{i_m} = 0.$$

□

13.3 Vector Fields and the Structure Equations of $\mathbb{R}^n$

Let M be the set of coordinates $(x^1, x^2, \dots, x^n)$ and U a subset of M.

Definition 13.3.1 A smooth map $X : M \longrightarrow TM$, which assigns to any $p \in M$ a vector $X_p \in T_pM$ is called a vector field on M.

In coordinates

$$X_p = \sum_{i=1}^{n} X^i (p) \left(\frac{\partial}{\partial x^i} \right)_p , \ X^i \in \mathcal{F} (M, \mathbb{R}) .$$

The previous formula tells us that each vector field is a directional derivative of a function f at p.

We denote the set of vector fields on M by $\mathcal{X} (M)$.

For any $f \in \mathcal{F} (M)$ and $X \in \mathcal{X} (M)$, it makes sense the formula

$$(f X)_p := f (p) X_p$$

which defines the vector field $f X \in \mathcal{X} (M)$. If $X \in \mathcal{X} (M)$ and $f \in \mathcal{F} (M)$, another formula can be derived,

$$(Xf) (p) := X_p f,$$

which defines a function $Xf \in \mathcal{F} (M)$.

For any two vector fields $X, Y \in \mathcal{X} (M)$, it can be defined a vector field $[X, Y]$, called commutator(or the Lie bracket), by the formula

$$[X, Y]_p f := X_p (Yf) - Y_p (Xf) .$$

In coordinates, if $X = X^i \frac{\partial}{\partial x^i}$, $Y = Y^i \frac{\partial}{\partial x^i}$ and $f = x^i$, we have

$$[X, Y]_p \left(x^i\right) = X^j_p \left.\frac{\partial}{\partial x^j}\right|_p \left(Y^i\right) - Y^j_p \left.\frac{\partial}{\partial x^j}\right|_p \left(X^i\right),$$

that is, the i-vector component $[X, Y]^i_p$ is

$$[X, Y]^i_p = X^j_p \left.\frac{\partial Y^i}{\partial x^j}\right|_p - Y^j_p \left.\frac{\partial X^i}{\partial x^j}\right|_p .$$

Theorem 13.3.2 *For any $X, \ Y, Z \in \mathcal{X} (M)$ the following identity holds:*

$$[[X, Y] , Z] + [[Y, Z] , X] + [[Z, X] , Y] = 0.$$

Proof This identity is called the *Jacobi identity*.

To prove it, let first observe that for $f \in \mathcal{F} (M)$, we have

$$[[X, Y] , Z] (f) = [X, Y] (Z (f)) - Z ([X, Y] (f)) ,$$

i.e.

$$[[X, Y] , Z] (f) = X (Y (Z (f))) - Y (X (Z (f))) - Z (X (Y (f))) + Z (Y (X (f))) .$$

The permutations $X \to Y,\ Y \to Z,\ Z \to X$ lead to other two formulas. Summing them, we obtain the assertion of the theorem. □

We are ready to present the structure equations of $\mathbb{R}^n$..

For $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ in $\mathbb{R}^n$, the usual inner product is

$$\langle x, y\rangle = x_1 y_1 + x_2 y_2 + ... + x_n y_n.$$

In fact we can consider at each point $p \in \mathbb{R}^n$ the above vectors and we can write

$$\langle x, y\rangle_p = x_1 y_1 + x_2 y_2 + ... + x_n y_n.$$

This way we considered at each point $p \in \mathbb{R}^n$ the tangent space $T_p\mathbb{R}^n$ and the vectors x and y as $x_p,\ y_p \in T_p\mathbb{R}^n$. So, the inner product was added to the vector space structure of $T_p\mathbb{R}^n$. Having all these in mind at each $p \in \mathbb{R}^n$ we consider n differentiable vector fields $\{e_1, e_2, ..., e_n\}$ such that

$$\langle e_i, e_j\rangle_p = \delta_{ij}.$$

This is an orthonormal moving frame of $T\mathbb{R}^n$ such that each e_i is a smooth map on $\mathbb{R}^n$ to $\mathbb{R}$.

Since $(de_i)_p$ is also a linear map on $\mathbb{R}^n$ to $\mathbb{R}$, at each p and $v \in \mathbb{R}^n$ we can write

$$(de_i)_p(v) = \sum_{j=1}^{n} (\omega_{ij})_p(v) e_j.$$

where ω_{ij} are n^2 differential 1-forms called the *connection forms of* $\mathbb{R}^n$.

Let us keep in mind this relation in the form

$$de_i = \sum_{j=1}^{n} \omega_{ij} e_j.$$

From

$$\langle e_i, e_j\rangle_p = \delta_{ij}$$

it results in

$$\langle de_i, e_j\rangle_p + \langle e_i, de_j\rangle_p = \omega_{ij} + \omega_{ji} = 0,$$

i.e. the antisymmetry of the connection form,

$$\omega_{ij} = -\omega_{ji}.$$

We define at each $p \in \mathbb{R}^n$ the dual basis of $\{e_1, e_2, ..., e_n\}$, i.e. the 1-differential forms $\{(\omega_1)_p, (\omega_2)_p, ..., (\omega_n)_p\}$ such that

$$\omega_i(e_j) = \delta_{ij}.$$

Let us observe that any $x \in \mathbb{R}^n$ can be thought as an inclusion map on $\mathbb{R}^n$ to $\mathbb{R}^n$, that is to say that the 1-forms $\{\omega_i\}_{i=1,\dots,n}$ are dual to the moving frame $\{e_i\}_{i=1,\dots,n}$ is equivalent to saying that

$$dx = \sum_{i=1}^{n} \omega_i e_i.$$

Theorem 13.3.3 *(Elie Cartan) In the previous notations*

$$d\omega_i = \sum_{j=1}^{n} \omega_j \wedge \omega_{ji},$$

$$d\omega_{ij} = \sum_{k=1}^{n} \omega_{ik} \wedge \omega_{kj}.$$

Proof We start from

$$dx = \sum_{i=1}^{n} \omega_i e_i.$$

Since $d(dx) = 0$ we have

$$0 = d(dx) = \sum_{i=1}^{n} d\omega_i e_i - \sum_{i=1}^{n} \omega_i \wedge de_i = \sum_{i=1}^{n} d\omega_i e_i - \sum_{i=1}^{n} \omega_i \wedge \sum_{j=1}^{n} \omega_{ij} e_j =$$

$$= \sum_{j=1}^{n} \left(d\omega_j - \sum_{i=1}^{n} \omega_{ji} \wedge \omega_i \right) e_j.$$

The second equality is proved in the same way starting from

$$de_i = \sum_{j=1}^{n} \omega_{ij} e_j.$$

□

These structure equations above can be applied to the special case of the surfaces of the Euclidean space $\mathbb{E}^3$.

The immersion $x : \mathbb{R}^n \to \mathbb{R}^n$ must be replaced by $x : \mathbb{R}^2 \to \mathbb{R}^3$ which plays the role of a surface.

The inner product at each point $p \in \mathbb{R}^2$ is understood exactly as we saw in the case of surfaces, i.e.

$$\langle v_1, v_2 \rangle := \langle dx_p(v_1), dx_p(v_2) \rangle_{x(p)},$$

the second inner product being in $\mathbb{E}^3$.

Of course, the image of the surface x can be only the image of an open set U of $\mathbb{R}^2$, therefore we have the same notations as in the case of surfaces except f, which now is denoted by x.

For $p \in U$, we have $x(p) \in x(U) \subset \mathbb{E}^3$. It is possible to choose in the tangent space of $x(U)$ two vectors $\{e_1, e_2\}$ and perpendicular to them a vector e_3, such that the set $\{e_1, e_2, e_3\}$ is a moving frame in the sense discussed above.

Associated to these vectors there are the 1-differential forms $\omega_1, \omega_2, \omega_3$ which are the dual basis and nine connection forms ω_{ij} which satisfy $\omega_{ij} = \omega_{ji}, i, j \in \{1, 2, 3\}$.

Let us write the structure equations:

$$d\omega_1 = \omega_2 \wedge \omega_{21} + \omega_3 \wedge \omega_{31}$$

$$d\omega_2 = \omega_1 \wedge \omega_{12} + \omega_3 \wedge \omega_{32}$$

$$d\omega_3 = \omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23}$$

$$d\omega_{12} = \omega_{13} \wedge \omega_{32}$$

$$d\omega_{13} = \omega_{12} \wedge \omega_{23}$$

$$d\omega_{23} = \omega_{21} \wedge \omega_{13}$$

Since dx depends only on e_1 and e_2, it results that $\omega_3 = 0$. Then

$$d\omega_3 = \omega_1 \wedge \omega_{13} + \omega_2 \wedge \omega_{23} = 0,$$

therefore we obtain

$$\omega_{13} = h_{11}\omega_1 + h_{12}\omega_2$$

and

$$\omega_{23} = h_{21}\omega_1 + h_{22}\omega_2$$

with $h_{12} = h_{21}$.

Let us show that the matrix (h_{ij}) is the Weingarten matrix associated to the surface x.

We observe that our choice related to the moving frame, that is,

$$de_3 = \omega_{31}e_1 + \omega_{32}e_2,$$

leads to

$$de_3(v) = -\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

for any vector $v = a_1 e_1 + a_2 e_2$. It results that e_3 is the Gauss map. Taking into account that the Gaussian curvature K at each point is the determinant of the previous matrix, a simple computation shows that

$$d\omega_{12} = \omega_{13} \wedge \omega_{32} = -(h_{11}h_{22} - h_{12}h_{21})\omega_1 \wedge \omega_2,$$

therefore

$$d\omega_{12} = -K\omega_1 \wedge \omega_2.$$

In conclusion, we have another perspective offered by the differential forms via the structure equations for the geometry of surfaces. We can discuss about the structure equations in a given Minkowski space or, more generally, in the case of a coordinate system endowed with a metric.

13.4 Affine Connections, Torsion, and Curvature

Definition 13.4.1 A $\mathbb{R}$-bilinear map $\nabla : \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M)$, such that

$$\nabla_{fX} Y = f \nabla_X Y$$

and

$$\nabla_X fY = f \nabla_X Y + X(f)\, Y,$$

for any $X, Y \in \mathcal{X}(M)$, and $f \in \mathcal{F}(M)$ is called an affine connection on M.

Let us observe that the formula

$$\nabla_{fX} Y = f \nabla_X Y$$

implies

$$\nabla_{(fX+gX')} Y = f \nabla_X Y + g \nabla_{X'} Y,$$

for any $f, g \in \mathcal{F}(M)$, $X, X', Y \in \mathcal{X}(M)$, that is the affine connection is $\mathcal{F}(M)$-linear in the first variable. The second formula of the definition is the Leibniz rule in the second variable.

The different behaviour of the two variables makes clear why the notation $\nabla_X Y$ is preferred instead of the usual notation $\nabla(X, Y)$.

We denote by $\mathcal{C}(M)$ the set of affine connections of M.

Proposition 13.4.2 *If for any $Y \in \mathcal{X}(M)$ there are the vector fields $X_1, X_2, ..., X_n \in \mathcal{X}(M)$ such that*

$$Y = \sum_{i=1}^{n} Y^i X_i,$$

then the formula

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i) X_i$$

defines an affine connection on M.

Proof The $\mathbb{R}-$bilinearity is obvious. We have both

$$\nabla_{fX} Y = \sum_{i=1}^{n} (fX)(Y^i) X_i = f \sum_{i=1}^{n} X(Y^i) X_i = f \nabla_X Y$$

and

$$\nabla_X fY = \sum_{i=1}^{n} X(fY^i) X_i = f \sum_{i=1}^{n} X(Y^i) X_i + X(f) \sum_{i=1}^{n} Y^i X_i = f \nabla_X Y + X(f)\, Y$$

properties fulfilled, therefore the formula

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i) X_i$$

defines an affine connection on M. □

Let us observe that, in a system of coordinates of M, say $\left(x^1, ..., x^n\right)$, the set $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$ has the property coming from the previous proposition, i.e. any $Y \in \mathcal{X}(M)$ can be written in the form:

$$Y = \sum_{i=1}^{n} Y^i \frac{\partial}{\partial x^i}.$$

Therefore the formula

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i) \frac{\partial}{\partial x^i}$$

defines an affine connection on M.

Consider two affine connections on M, $\nabla_X Y$ and $\overline{\nabla}_X Y$. According to their properties, the difference

$$S(X, Y) := \nabla_X Y - \overline{\nabla}_X Y$$

is a $\mathcal{F}(M)-$bilinear map. Therefore we have an infinity of affine connections on M because if we add any $S : \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M)$, $\mathcal{F}(M)$-bilinear map to the affine connection given by the formula

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i)\frac{\partial}{\partial x^i},$$

we obtain a new affine connection. The geometry of M depends on the properties of the connections on M. And the properties of the connections on M depend on two special maps

$$T : \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M),$$
$$R : \mathcal{X}(M) \times \mathcal{X}(M) \times \mathcal{X}(M) \longrightarrow \mathcal{X}(M),$$

defined, respectively, by the formulas

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

T is called torsion, while R is called curvature of the affine connection ∇.

In what follows, we use the Einstein notation for the sum.

In coordinates, we denote by Γ^k_{ij}, the components of the affine connection ∇, which appear from

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} := \Gamma^k_{ij}\frac{\partial}{\partial x^k}.$$

The components of T and R are denoted, respectively, by T_{ij} and R^i_{jkl}. They appear from

$$T\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) := T^k_{ij}\frac{\partial}{\partial x^k}$$
$$R\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right)\frac{\partial}{\partial x^l} := R^i_{ljk}\frac{\partial}{\partial x^i}.$$

We have

$$T\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = T^k_{ij}\frac{\partial}{\partial x^k} = \nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^j}}\frac{\partial}{\partial x^i} = \Gamma^k_{ij}\frac{\partial}{\partial x^k} - \Gamma^k_{ji}\frac{\partial}{\partial x^k} =$$
$$= \left(\Gamma^k_{ij} - \Gamma^k_{ji}\right)\frac{\partial}{\partial x^k},$$

therefore

$$T^k_{ij} = \Gamma^k_{ij} - \Gamma^k_{ji}.$$

In the same way

$$R\left(\frac{\partial}{\partial x^j}, \frac{\partial}{\partial x^k}\right)\frac{\partial}{\partial x^l} = R^i_{ljk}\frac{\partial}{\partial x^i} = \nabla_{\frac{\partial}{\partial x^j}}\nabla_{\frac{\partial}{\partial x^k}}\frac{\partial}{\partial x^l} - \nabla_{\frac{\partial}{\partial x^k}}\nabla_{\frac{\partial}{\partial x^j}}\frac{\partial}{\partial x^l} =$$

$$= \nabla_{\frac{\partial}{\partial x^j}} \left(\Gamma^s_{kl} \frac{\partial}{\partial x^s} \right) - \nabla_{\frac{\partial}{\partial x^k}} \left(\Gamma^s_{jl} \frac{\partial}{\partial x^s} \right) =$$

$$= \Gamma^s_{kl} \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^s} + \frac{\partial \Gamma^s_{kl}}{\partial x^j} \cdot \frac{\partial}{\partial x^s} - \Gamma^s_{jl} \nabla_{\frac{\partial}{\partial x^k}} \frac{\partial}{\partial x^s} - \frac{\partial \Gamma^s_{jl}}{\partial x^k} \cdot \frac{\partial}{\partial x^s} =$$

$$= \left(\frac{\partial \Gamma^i_{kl}}{\partial x^j} - \frac{\partial \Gamma^i_{jl}}{\partial x^k} + \Gamma^s_{kl} \Gamma^i_{js} - \Gamma^s_{jl} \Gamma^i_{ks} \right) \frac{\partial}{\partial x^i},$$

therefore

$$R^i_{ljk} = \frac{\partial \Gamma^i_{kl}}{\partial x^j} - \frac{\partial \Gamma^i_{jl}}{\partial x^k} + \Gamma^s_{kl} \Gamma^i_{js} - \Gamma^s_{jl} \Gamma^i_{ks}.$$

Let us note that the last formula is the same that we considered when we studied surfaces with $\Gamma^i_{jk} = \Gamma^i_{kj}$. In this case, it is $T_{jk} = 0$. Such affine connections are called *torsion-free connections.*. They are also called *symmetric connections*. Therefore an affine connection is a symmetric connection if its torsion is null.

The property $\Gamma^i_{jk} = \Gamma^i_{kj}$ of the torsion-free connections reminds us a property of the Christoffel symbols and the above formula, written for a torsion-free affine connection, i.e.

$$R^i_{ljk} = \frac{\partial \Gamma^i_{lk}}{\partial x^j} - \frac{\partial \Gamma^i_{lj}}{\partial x^k} + \Gamma^s_{lk} \Gamma^i_{sj} - \Gamma^s_{lj} \Gamma^i_{sk}$$

reminds us the Riemann curvature mixed tensor formula.

Next, we show how the covariant derivative is determined by an affine connection $\nabla \in \mathcal{C}(M)$. Parallel transport and geodesics appear in the same way as we saw in the chapter of basic Differential Geometry.

13.5 Covariant Derivative, Parallel Transport, and Geodesics

Definition 13.5.1 Consider both a vector field $X \in \mathcal{X}(M)$ and an affine connection $\nabla \in \mathcal{C}(M)$. The map

$$\nabla_X : \mathcal{X}(M) \longrightarrow \mathcal{X}(M), \text{ defined by } \nabla_X(Y) := \nabla_X Y,$$

is called a covariant derivative of the vector field Y along the vector field X.

The covariant derivative is a generalization of the directional derivative. In coordinates, if $X = X^i \frac{\partial}{\partial x^i}$ and $Y = Y^j \frac{\partial}{\partial x^{ji}}$, we obtain successively

$$\nabla_X Y = \nabla_{X^i \frac{\partial}{\partial x^i}} Y^j \frac{\partial}{\partial x^{ji}} = X^i \nabla_{\frac{\partial}{\partial x^i}} Y^j \frac{\partial}{\partial x^{ji}} = X^i \left(Y^j \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} + \frac{\partial Y^j}{\partial x^i} \cdot \frac{\partial}{\partial x^j} \right) =$$
$$= X^i \left(Y^j \left(\Gamma^k_{ij} \frac{\partial}{\partial x^k} \right) + \frac{\partial Y^k}{\partial x^i} \cdot \frac{\partial}{\partial x^k} \right) = \left(X^i \frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} X^i Y^j \right) \frac{\partial}{\partial x^k}.$$

Therefore the covariant derivative formula is

$$\nabla_X Y = X^i \left(\frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} Y^j \right) \frac{\partial}{\partial x^k}.$$

Consider the smooth curve $c : I \subset \mathbb{R} \longrightarrow M$ described in our coordinate system as

$$c(t) := \left(x^1(t), ..., x^n(t) \right).$$

The tangent vector $\dot{c}(t) := \frac{dc(t)}{dt} = \left(\dot{x}^1(t), ..., \dot{x}^n(t) \right)$ belongs, at each point of the curve, to the corresponding tangent space, i.e. $\dot{c}(t) \in T_{c(t)}M$. Replacing X^i by $\dot{x}^i(t)$ in the covariant derivative formula, we have

$$\nabla_{\dot{c}(t)} Y = \dot{x}^i(t) \left(\frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} Y^j \right) \frac{\partial}{\partial x^k}.$$

Definition 13.5.2 The vector field Y is parallel transported along $c \in M$ if $\nabla_{\dot{c}(t)} Y(t) = 0$.

In coordinates

$$\frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} Y^j = 0, \ k = 1, ..., n.$$

Definition 13.5.3 $c \subset M$ is called a geodesic of M if $\nabla_{\dot{c}(t)} \dot{c}(t) = 0$.

This means, in coordinates,

$$\ddot{x}^k + \Gamma^k_{ij} \dot{x}^i \dot{x}^j = 0, \quad \text{for } k = 1, ..., n.$$

Let us observe that the initial conditions $c(0) = p$ and $\dot{c}(0) = v_p$ determine a unique geodesic.

The formulas obtained in this abstract case both for parallel transport and geodesics are similar to the formulas seen above in the chapter on basic Differential Geometry.

The next theorem is important.

Theorem 13.5.4 *For any given affine connection $\nabla \in \mathcal{C}(M)$ there exists an affine connection $\nabla^1 \in \mathcal{C}(M)$ with the properties:*

(*i*) ∇^1 *is a torsion-free connection;*
(*ii*) ∇^1 *has the same geodesics as the initial affine connection* ∇.

Proof Consider the affine connection

$$\nabla^1_X Y := \nabla_X Y - \frac{1}{2} T_\nabla (X, Y),$$

where $T_\nabla (X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$ is the torsion of the initial affine connection ∇.

In order to prove the symmetry of the affine connection ∇^1, we first construct the torsion of it.

$$T_{\nabla^1} (X, Y) = \nabla^1_X Y - \nabla^1_Y X - [X, Y].$$

Then, the computations lead to

$$T_{\nabla^1} (X, Y) = \nabla_X Y - \frac{1}{2} T (X, Y) - \nabla_Y X + \frac{1}{2} T (Y, X) - [X, Y] = \nabla_X Y - \frac{1}{2} \nabla_X Y +$$
$$+ \frac{1}{2} \nabla_Y X + \frac{1}{2} [X, Y] + \frac{1}{2} [X, Y] - \nabla_Y X + \frac{1}{2} \nabla_Y X - \frac{1}{2} \nabla_X Y - \frac{1}{2} [Y, X] - [X, Y] = 0.$$

Therefore ∇^1 is a torsion-free connection.

Now, from

$$\nabla^1_{\dot{c}(t)} \dot{c} (t) = \nabla_{\dot{c}(t)} \dot{c} (t) - \frac{1}{2} T (\dot{c} (t), \dot{c} (t)) = \nabla_{\dot{c}(t)} \dot{c} (t)$$

it results that the two affine connections have the same geodesics.

This ends the proof. □

13.6 A Geometric Description of Riemann Curvature Mixed Tensor via Parallel Transport

The Riemann symbols, in the case of surfaces, were obtained by considering the partial derivative of Gauss formulas. The way we introduced the parallel transport of contravariant vectors, without using an extra dimension, allows us to think at a way to introduce the Riemann mixed curvature tensor without an extra dimension.

The key of the geometric description of the curvature of an affine connection is the parallel transport of vectors along infinitesimal vectors.

Suppose the infinitesimal vector is $A = (\delta x^0, \delta x^1, ..., \delta x^n)$ and let V be the vector we parallel transport along the infinitesimal vector A. If we act in a Euclidean space where $\Gamma^k_{ij} = 0$, at the end we have only the vector $A + V$ of components $A^k + V^k$. But, in general, $\Gamma^k_{ij} \neq 0$ and the parallel transport highlights the vector of components $A^k + V^k + \delta V^k$, where

$$\delta V^k = -\Gamma^k_{ij} V^j \delta x^i.$$

If the components of V are dx^k, i.e. $V = (dx^0, dx^1, .., dx^n)$, the previous formula becomes

$$\delta(dx^k) = -\Gamma^k_{ij} dx^j \delta x^i$$

and each component of the parallel transported vector V along A becomes $\delta x^k + dx^k + \delta(dx^k)$, that is,

$$\delta x^k + dx^k - \Gamma^k_{ij} dx^j \delta x^i .$$

If we consider the parallel transport of the infinitesimal vector A along the infinitesimal vector V, at the end we have the components $V^k + A^k + dA^k$, where

$$dA^k = -\Gamma^k_{ij} A^j dx^i .$$

Therefore, at the end of the parallel transport of the vector A along the vector V, we obtain

$$dx^k + \delta x^k - \Gamma^k_{ij} \delta x^j dx^i .$$

In the Euclidean space, the condition $A^k + V^k = V^k + A^k,\ k \in \{0, 1, .., n\}$ describes a parallelogram. Here, the parallelogram is described by the condition

$$A^k + (V^k + \delta V^k) = V^k + (A^k + dA^k),\ k \in \{0, 1, .., n\},$$

that is,

$$\cancel{\delta x^k} + \cancel{dx^k} - \Gamma^k_{ij} dx^j \delta x^i = \cancel{dx^k} + \cancel{\delta x^k} - \Gamma^k_{ij} \delta x^j dx^i .$$

This condition may be written in the form

$$\Gamma^k_{ij} dx^j \delta x^i = \Gamma^k_{ji} \delta x^i dx^j .$$

The last equality is true if and only if the connection coefficients symbols are symmetric, i.e. $\Gamma^k_{ij} = \Gamma^k_{ji}$. Therefore if the torsion of the connection is null.

Let us take into account the parallelogram considered above and a vector $W = (W^0, W^1, .., W^n)$. We consider the parallel transport of W along the first two sides. Along A we first obtain the vector X of coordinates $X^k := A^k + (W^k + \delta W^k)$. Then, this vector is parallel transported along V. We obtain the vector of coordinates $V^k + (X^k + dX^k)$. Therefore the parallel transport of W along the first two sides leads to the vector of coordinates

$$T^k_1 := V^k + [A^k + (W^k + \delta W^k) + d(A^k + (W^k + \delta W^k))] \quad (1)$$

The parallel transport of W along V leads to the vector Y of coordinates

$$Y^k := V^k + (W^k + dW^k)$$

and the parallel transport of Y along A leads to the vector of coordinates

$$A^k + (Y^k + \delta Y^k).$$

Therefore the parallel transport of W along the other two sides leads to the vector of coordinates

$$T_2^k := A^k + [V^k + (W^k + dW^k) + \delta(V^k + (W^k + dW^k))] \quad (2)$$

We continue to work in coordinates. The relation which allows us to consider the initial parallelogram is $dA^k = \delta V^k$. If we denote by

$$R := T_2^k - T_1^k,$$

it results in

$$R = \delta(dW^k) - d(\delta W^k).$$

If we compute

$$-\delta(dW^k) = \delta(\Gamma^k_{ij} W^i dx^j) = \delta\Gamma^k_{ij} W^i dx^j + \Gamma^k_{ij}(\delta W^i)dx^j + \Gamma^k_{ij} W^i \delta(dx^j),$$

we obtain

$$-\delta(dW^k) = \frac{\partial \Gamma^k_{ij}}{\partial x^l}\delta x^l W^i dx^j - \Gamma^k_{ij}\Gamma^i_{ab} W^a \delta x^b dx^j - \Gamma^k_{ij}\Gamma^j_{ab} W^i dx^a \delta x^b.$$

Arranging the indexes

$$\delta(dW^k) = \left(-\frac{\partial \Gamma^k_{ij}}{\partial x^l} + \Gamma^k_{sj}\Gamma^s_{il} + \Gamma^k_{si}\Gamma^s_{jl}\right) W^i dx^j \delta x^l,$$

and in the same way

$$-d(\delta W^k) = \left(\frac{\partial \Gamma^k_{il}}{\partial x^j} - \Gamma^k_{sl}\Gamma^s_{ij} - \Gamma^k_{si}\Gamma^s_{lj}\right) W^i dx^j \delta x^l.$$

Therefore, after cancelling $\Gamma^k_{si}\Gamma^s_{lj}$, one obtains

$$\delta(dW^k) - d(\delta W^k) = \left(\frac{\partial \Gamma^k_{il}}{\partial x^j} - \frac{\partial \Gamma^k_{ij}}{\partial x^l} + \Gamma^k_{sj}\Gamma^s_{il} - \Gamma^k_{sl}\Gamma^s_{ij}\right) W^i dx^j \delta x^l = R^k_{ijl} W^i dx^j \delta x^l.$$

The curvature formula of a torsion-free connection in coordinates is recovered. If the vector $R = \delta(dW^k) - d(\delta W^k)$ is 0, the two vectors are coincident, it happens in the Euclidean Geometry. If not, a curvature of M appears.

13.7 The Levi-Civita Connection

We consider a metric tensor g in the sense described in the chapter dedicated to the basic Differential Geometry.

Definition 13.7.1 An affine connection $\nabla \in \mathcal{C}(M)$ which fulfills the relation

$$X(g(Y,Z)) - g(\nabla_X Y, Z) - g(\nabla_X Z, Y) = 0$$

for all $X, Y, Z \in \mathcal{X}(M)$, is called a metric connection of M.

The previous condition can be written in the simplified form $\nabla_X g(Y,Z) = 0$. Therefore

$$\nabla_X g(Y,Z) := X(g(Y,Z)) - g(\nabla_X Y, Z) - g(\nabla_X Z, Y)$$

is in fact an abstract covariant derivative formula of the metric tensor. At the same time, this formula is a compatibility condition between the metric and the connection.

Theorem 13.7.2 *For each metric tensor g, a unique torsion-free metric connection $\nabla \in \mathcal{C}(M)$ exists such that $\nabla_X g(Y,Z) = 0$ for all $X, Y, Z \in \mathcal{X}(M)$.*

Proof Let us first look at the uniqueness of the torsion-free metric connection asserted by the previous statement.

The torsion-free condition may be written in the form $\nabla_X Z = \nabla_Z X + [X, Z]$ and replaced in the formula above. We obtain

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(\nabla_Z X, Y) + g([X,Z], Y).$$

Consider $X \to Y \to Z \to X$ and the corresponding relations

$$Y(g(Z,X)) = g(\nabla_Y Z, X) + g(\nabla_X Y, Z) + g([Y,X], Z);$$

$$Z(g(X,Y)) = g(\nabla_Z X, Y) + g(\nabla_Y Z, X) + g([Z,Y], X).$$

Adding the first two and subtracting the last it results

$$\begin{aligned} 2g(\nabla_X Y, Z) = {} & X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) - g([X,Z], Y) \\ & + g([Z,Y], X) - g([Y,X], Z). \end{aligned}$$

If, by absurdum, there exist two connections ∇, ∇^1 which fulfill the previous formula, on subtracting, we obtain

$$2g(\nabla_X Y, Z) - 2g\left(\nabla^1_X Y, Z\right) = 0, \quad \text{for all } X, Y, Z \in \mathcal{X}(M),$$

which means that g is non degenerate and Z is arbitrary, i.e. $\nabla_X Y = \nabla^1_X Y$.

The existence is related to the above formula,

$$g\left(\nabla_X Y, Z\right) := \frac{1}{2}\left[X\left(g\left(Y, Z\right)\right) + Y\left(g\left(Z, X\right)\right) - Z\left(g\left(X, Y\right)\right)\right] - \\ -\frac{1}{2}\left[g\left(\left[X, Z\right], Y\right) - g\left(\left[Z, Y\right], X\right) + g\left(\left[Y, X\right], Z\right)\right].$$

The so-defined ∇ is a torsion-free affine metric connection, and this is shown by simple computations. □

The previous torsion-free metric connection is known as the *Levi-Civita connection.*

In coordinates, if

$$X := \frac{\partial}{\partial x^i}, Y := \frac{\partial}{\partial x^j}, Z := \frac{\partial}{\partial x^k},$$

$$\nabla_X Y = \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} := \Gamma^r_{ij} \frac{\partial}{\partial x^r},$$

$$g_{ij} := g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right),$$

then

$$2g\left(\Gamma^r_{ij}\frac{\partial}{\partial x^r}, \frac{\partial}{\partial x^k}\right) = \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}.$$

We recognize the Christoffel symbol's definition and it is important to mention that Christoffel obtained the same results as Levi-Civita even before. He published them in *Crelle's Journal* in 1869. The results of Levi-Civita regarding the parallel transport and the covariant derivative were published in 1917 in *Rendiconti del Circolo Matematico di Palermo*. In the context of the Levi-Civita connection, all tensors considered in the abstract Differential Geometry can be described via the Christoffel symbols seen here.

13.8 Coordinate Changes for Geometric Objects Generated by the Levi-Civita Connection

We use the same notations as we used in the two Differential Geometry chapters of the book. According to the previous subsection, the Levi-Civita connection will produce, in coordinates, the first and second kind of Christoffel symbols. The Christoffel symbols of the second kind are involved in the description of the second Riemann symbols. This can be also seen as the description in coordinates of the curvature of an affine connection, while the metric will be again implied to define the Riemann symbols of the first kind, the Ricci symbols, the Ricci mixed symbols, and the Ricci scalar. After the next theorems we will find out that Christoffel symbols are not tensors, while Riemann symbols are. The relation with the two previous chapters of Differential Geometry will be fully realized.

Theorem 13.8.1 *A change of coordinates* $x^r = x^r(\overline{x}^h), r \in \{0, 1, ..., n\}, h \in \{0, 1, ..., n\}$ *transforms the Christoffel symbols of first kind under the rule*

$$\overline{\Gamma}_{ij,k} = \Gamma_{rs,p} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \frac{\partial x^p}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k}$$

Proof Let us start from

$$\Gamma_{ij,k} = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right)$$

and

$$\overline{g}_{jk} = g_{rs} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k}.$$

We have

$$\frac{\overline{g}_{jk}}{\partial \overline{x}^i} = \frac{g_{rs}}{\partial \overline{x}^p} \frac{\partial x^p}{\partial \overline{x}^i} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k} + g_{rs} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial^2 x^s}{\partial \overline{x}^i \partial \overline{x}^k},$$

$$\frac{\overline{g}_{ki}}{\partial \overline{x}^j} = \frac{g_{rs}}{\partial \overline{x}^p} \frac{\partial x^p}{\partial \overline{x}^j} \frac{\partial x^r}{\partial \overline{x}^k} \frac{\partial x^s}{\partial \overline{x}^i} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^j \partial \overline{x}^k} \frac{\partial x^s}{\partial \overline{x}^i} + g_{rs} \frac{\partial x^r}{\partial \overline{x}^k} \frac{\partial^2 x^s}{\partial \overline{x}^j \partial \overline{x}^i},$$

$$\frac{\overline{g}_{ij}}{\partial \overline{x}^k} = \frac{g_{rs}}{\partial \overline{x}^p} \frac{\partial x^p}{\partial \overline{x}^k} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^k \partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + g_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial^2 x^s}{\partial \overline{x}^k \partial \overline{x}^j}.$$

Since

$$r \to s \to p \to r \implies \frac{g_{rs}}{\partial \overline{x}^p} \frac{\partial x^p}{\partial \overline{x}^i} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k} = \frac{g_{sp}}{\partial \overline{x}^r} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^p}{\partial \overline{x}^k} \frac{\partial x^s}{\partial \overline{x}^j},$$

$$r \to p \to s \to r \implies \frac{g_{rs}}{\partial \overline{x}^p} \frac{\partial x^p}{\partial \overline{x}^j} \frac{\partial x^r}{\partial \overline{x}^k} \frac{\partial x^s}{\partial \overline{x}^i} = \frac{g_{pr}}{\partial \overline{x}^s} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^p}{\partial \overline{x}^k} \frac{\partial x^s}{\partial \overline{x}^j},$$

after we add, considering the first two equalities with plus and the last one with minus, we obtain

$$\overline{\Gamma}_{ij,k} = \Gamma_{rs,p} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \frac{\partial x^p}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k}.$$

□

Theorem 13.8.2 *A change of coordinates* $x^r = x^r(\overline{x}^h), r \in \{0, 1, , ..., n\}, h \in \{0, 1, ..., n\}$ *transforms the Christoffel symbols of second kind under the rule*

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\Gamma^k_{rs} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\Gamma}^r_{ij} \frac{\partial x^k}{\partial \overline{x}^r}.$$

Proof We start from the equalities

$$\begin{cases} \overline{g}^{ij}\overline{g}_{jk} = \delta^i_k \\ \overline{g}_{jk} = g_{rs} \dfrac{\partial x^r}{\partial \overline{x}^j} \dfrac{\partial x^s}{\partial \overline{x}^k} \end{cases} \Longrightarrow \overline{g}^{ij} g_{rs} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k} = \delta^i_k$$

which are multiplied by $g^{pq} \dfrac{\partial \overline{x}^k}{\partial x^p}$. It results in

$$\overline{g}^{ij} g_{rs} \frac{\partial x^r}{\partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k} g^{pq} \frac{\partial \overline{x}^k}{\partial x^p} = \delta^i_k g^{pq} \frac{\partial \overline{x}^k}{\partial x^p},$$

i.e.

$$\overline{g}^{ij} g_{rs} \frac{\partial x^r}{\partial \overline{x}^j} g^{pq} \frac{\partial x^s}{\partial x^p} = g^{pq} \frac{\partial \overline{x}^i}{\partial x^p}.$$

The left side makes sense only if $s = p$, therefore we have

$$\overline{g}^{ij} g^{pq} g_{pr} \frac{\partial x^r}{\partial \overline{x}^j} = g^{pq} \frac{\partial \overline{x}^i}{\partial x^p},$$

that is

$$\overline{g}^{ij} \frac{\partial x^q}{\partial \overline{x}^j} = g^{pq} \frac{\partial \overline{x}^i}{\partial x^p}.$$

Multiplying

$$\overline{\Gamma}_{ij,k} = \Gamma_{rs,p} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \frac{\partial x^p}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \frac{\partial x^s}{\partial \overline{x}^k}$$

by $\overline{g}^{mk}$, we obtain

$$\overline{\Gamma}^m_{ij} = \overline{g}^{mk}\overline{\Gamma}_{ij,k} = \Gamma_{rs,p} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \cdot \overline{g}^{mk} \frac{\partial x^p}{\partial \overline{x}^k} + g_{rs} \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j} \cdot \overline{g}^{mk} \frac{\partial x^s}{\partial \overline{x}^k},$$

that is

$$\overline{\Gamma}^m_{ij} = \Gamma_{rs,p} g^{pq} \frac{\partial \overline{x}^m}{\partial x^q} \cdot \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + g_{rs} \cdot g^{sq} \frac{\partial \overline{x}^m}{\partial x^q} \cdot \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j},$$

and finally

$$\overline{\Gamma}^m_{ij} = \Gamma^q_{rs} \frac{\partial \overline{x}^m}{\partial x^q} \cdot \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \frac{\partial \overline{x}^m}{\partial x^r} \cdot \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j}.$$

After multiplying by $\dfrac{\partial x^k}{\partial \overline{x}^m}$, it results in

$$\overline{\Gamma}_{ij}^m \frac{\partial x^k}{\partial \overline{x}^m} = \Gamma_{rs}^q \frac{\partial x^k}{\partial \overline{x}^m} \frac{\partial \overline{x}^m}{\partial x^q} \cdot \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \frac{\partial \overline{x}^m}{\partial x^r} \frac{\partial x^k}{\partial \overline{x}^m} \cdot \frac{\partial^2 x^r}{\partial \overline{x}^i \partial \overline{x}^j},$$

which can be arranged in the form

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\Gamma_{rs}^k \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\Gamma}_{ij}^r \frac{\partial x^k}{\partial \overline{x}^r}.$$

□

Theorem 13.8.3 *A change of coordinates transforms the Riemann symbols according to the formulas*

$$1)\ \overline{R}_{ijl}^m \frac{\partial x^k}{\partial \overline{x}^m} = R_{rsp}^k \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} \frac{\partial x^p}{\partial \overline{x}^l}.$$

$$2)\ \overline{R}_{ebgd} = R_{rjkl} \frac{\partial x^r}{\partial \overline{x}^e} \frac{\partial x^j}{\partial \overline{x}^b} \frac{\partial x^k}{\partial \overline{x}^g} \frac{\partial x^l}{\partial \overline{x}^d}.$$

Proof For 1), we consider the partial derivative $\dfrac{\partial}{\partial \bar{x}^l}$ of the expression

$$\frac{\partial^2 x^k}{\partial \overline{x}^i \partial \overline{x}^j} = -\Gamma_{rs}^k \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \overline{\Gamma}_{ij}^r \frac{\partial x^k}{\partial \overline{x}^r}.$$

It results in

$$\frac{\partial^3 x^k}{\partial \overline{x}^l \partial \overline{x}^i \partial \overline{x}^j} =$$

$$= -\frac{\partial \Gamma_{rs}^k}{\partial x^p} \frac{\partial x^p}{\partial \overline{x}^l} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} - \Gamma_{rs}^k \left(\frac{\partial^2 x^r}{\partial \overline{x}^l \partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^j} + \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial^2 x^s}{\partial \overline{x}^l \partial \overline{x}^j} \right) + \frac{\partial \overline{\Gamma}_{ij}^r}{\partial \overline{x}^l} \frac{\partial x^k}{\partial \overline{x}^r} + \overline{\Gamma}_{ij}^r \frac{\partial^2 x^k}{\partial \overline{x}^l \partial \overline{x}^r}.$$

We commute l and j indexes in the previous formula

$$\frac{\partial^3 x^k}{\partial \overline{x}^j \partial \overline{x}^i \partial \overline{x}^l} =$$

$$= -\frac{\partial \Gamma_{rs}^k}{\partial x^p} \frac{\partial x^p}{\partial \overline{x}^j} \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^l} - \Gamma_{rs}^k \left(\frac{\partial^2 x^r}{\partial \overline{x}^j \partial \overline{x}^i} \frac{\partial x^s}{\partial \overline{x}^l} + \frac{\partial x^r}{\partial \overline{x}^i} \frac{\partial^2 x^s}{\partial \overline{x}^j \partial \overline{x}^l} \right) + \frac{\partial \overline{\Gamma}_{il}^r}{\partial \overline{x}^j} \frac{\partial x^k}{\partial \overline{x}^r} + \overline{\Gamma}_{il}^r \frac{\partial^2 x^k}{\partial \overline{x}^j \partial \overline{x}^r}.$$

Since

$$\frac{\partial^3 x^k}{\partial \overline{x}^l \partial \overline{x}^i \partial \overline{x}^j} = \frac{\partial^3 x^k}{\partial \overline{x}^j \partial \overline{x}^i \partial \overline{x}^l}$$

we put the two equalities together and separate the quantities having bar on second kind Christoffel symbols by the ones without bar. We also cancel the equal quantities and the left member, here denoted as LM, becomes

$$LM = \frac{\partial \Gamma^k_{rs}}{\partial x^p}\frac{\partial x^p}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^l} - \frac{\partial \Gamma^k_{rs}}{\partial x^p}\frac{\partial x^p}{\partial \overline{x}^l}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j} + \Gamma^k_{rs}\frac{\partial^2 x^r}{\partial \overline{x}^j \partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\frac{\partial^2 x^r}{\partial \overline{x}^l \partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j}.$$

We divide the left member LM in two parts. In the first part, we interchange p and s. Then

$$\frac{\partial \Gamma^k_{rs}}{\partial x^p}\frac{\partial x^p}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^l} - \frac{\partial \Gamma^k_{rs}}{\partial x^p}\frac{\partial x^p}{\partial \overline{x}^l}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j} = \frac{\partial \Gamma^k_{rp}}{\partial x^s}\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^p}{\partial \overline{x}^l} - \frac{\partial \Gamma^k_{rs}}{\partial x^p}\frac{\partial x^p}{\partial \overline{x}^l}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j} =$$

$$= \left(\frac{\partial \Gamma^k_{rp}}{\partial x^s} - \frac{\partial \Gamma^k_{rs}}{\partial x^p}\right)\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^p}{\partial \overline{x}^l}.$$

In the second part of the left member, we use the formulas which explain how the second-type Christoffel symbols are transformed under a change of coordinates:

$$\Gamma^k_{rs}\frac{\partial^2 x^r}{\partial \overline{x}^j \partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\frac{\partial^2 x^r}{\partial \overline{x}^l \partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j} =$$

$$= \Gamma^k_{rs}\left(-\Gamma^r_{ab}\frac{\partial x^a}{\partial \overline{x}^j}\frac{\partial x^b}{\partial \overline{x}^i} + \overline{\Gamma}^a_{ji}\frac{\partial x^r}{\partial \overline{x}^a}\right)\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\left(-\Gamma^r_{ab}\frac{\partial x^a}{\partial \overline{x}^l}\frac{\partial x^b}{\partial \overline{x}^i} + \overline{\Gamma}^a_{li}\frac{\partial x^r}{\partial \overline{x}^a}\right)\frac{\partial x^s}{\partial \overline{x}^j} =$$

$$= -\Gamma^k_{rs}\Gamma^r_{ab}\frac{\partial x^a}{\partial \overline{x}^j}\frac{\partial x^b}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^l} + \Gamma^k_{rs}\Gamma^r_{ab}\frac{\partial x^a}{\partial \overline{x}^l}\frac{\partial x^b}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j} + \Gamma^k_{rs}\overline{\Gamma}^a_{ji}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\overline{\Gamma}^a_{li}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^j}.$$

We rearrange the dummy indexes such that the product of the three ratios becomes

$$\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^p}{\partial \overline{x}^l}.$$

The left member, denoted here by LM, becomes

$$\left(\frac{\partial \Gamma^k_{rp}}{\partial x^s} - \frac{\partial \Gamma^k_{rs}}{\partial x^p} + \Gamma^k_{as}\Gamma^a_{pr} - \Gamma^k_{ap}\Gamma^a_{rs}\right)\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^p}{\partial \overline{x}^l} + \Gamma^k_{rs}\overline{\Gamma}^a_{ji}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\overline{\Gamma}^a_{li}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^j}.$$

The final form of the left member is

$$LM = R^k_{rsp}\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^p}{\partial \overline{x}^l} + \Gamma^k_{rs}\overline{\Gamma}^a_{ji}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^l} - \Gamma^k_{rs}\overline{\Gamma}^a_{li}\frac{\partial x^r}{\partial \overline{x}^a}\frac{\partial x^s}{\partial \overline{x}^j}.$$

In same way, the right member, denoted here by RM, becomes

$$RM = \frac{\partial \overline{\Gamma}^r_{il}}{\partial \overline{x}^j}\frac{\partial x^k}{\partial \overline{x}^r} - \frac{\partial \overline{\Gamma}^r_{ij}}{\partial \overline{x}^l}\frac{\partial x^k}{\partial \overline{x}^r} + \overline{\Gamma}^r_{il}\frac{\partial^2 x^k}{\partial \overline{x}^j \partial \overline{x}^r} - \overline{\Gamma}^r_{ij}\frac{\partial^2 x^k}{\partial \overline{x}^l \partial \overline{x}^r} =$$

$$\left(\frac{\partial \overline{\Gamma}^r_{il}}{\partial \overline{x}^j} - \frac{\partial \overline{\Gamma}^r_{ij}}{\partial \overline{x}^l}\right)\frac{\partial x^k}{\partial \overline{x}^r} + \overline{\Gamma}^r_{il}\left(-\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^j}\frac{\partial x^b}{\partial \overline{x}^r} + \overline{\Gamma}^s_{jr}\frac{\partial x^k}{\partial \overline{x}^s}\right) + \overline{\Gamma}^r_{ij}\left(-\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^l}\frac{\partial x^b}{\partial \overline{x}^r} + \overline{\Gamma}^s_{lr}\frac{\partial x^k}{\partial \overline{x}^s}\right)$$

$$= \left(\frac{\partial \overline{\Gamma}^s_{il}}{\partial \overline{x}^j} - \frac{\partial \overline{\Gamma}^s_{ij}}{\partial \overline{x}^l} + \overline{\Gamma}^r_{il}\overline{\Gamma}^s_{jr} - \overline{\Gamma}^r_{ij}\overline{\Gamma}^s_{lr}\right)\frac{\partial x^k}{\partial \overline{x}^s} - \overline{\Gamma}^r_{il}\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^j}\frac{\partial x^b}{\partial \overline{x}^r} + \overline{\Gamma}^r_{ij}\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^l}\frac{\partial x^b}{\partial \overline{x}^r} =$$

$$= \overline{R}^s_{ijl}\frac{\partial x^k}{\partial \overline{x}^s} - \overline{\Gamma}^r_{il}\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^j}\frac{\partial x^b}{\partial \overline{x}^r} + \overline{\Gamma}^r_{ij}\Gamma^k_{ab}\frac{\partial x^a}{\partial \overline{x}^l}\frac{\partial x^b}{\partial \overline{x}^r}.$$

Comparing the final forms of the left and right members after reducing the equal terms, we obtain the formula

$$R^k_{rsp}\frac{\partial x^r}{\partial \overline{x}^i}\frac{\partial x^s}{\partial \overline{x}^j}\frac{\partial x^p}{\partial \overline{x}^l} = \overline{R}^m_{ijl}\frac{\partial x^k}{\partial \overline{x}^m}.$$

2) In 1), we multiply the right member by $g_{ri}\dfrac{\partial x^r}{\partial \bar{x}^e}$ and the left member by the same quantity written in the form $\overline{g}_{se}\dfrac{\partial \overline{x}^s}{\partial x^i}$.

This is possible because the formula $g_{ri}\dfrac{\partial x^r}{\partial \overline{x}^e} = \overline{g}_{se}\dfrac{\partial \overline{x}^s}{\partial x^i}$ comes from

$$\overline{g}_{es} = \overline{g}_{se} = g_{ri}\frac{\partial x^r}{\partial \overline{x}^e}\frac{\partial x^i}{\partial \overline{x}^s},$$

formula which was proved before. Then

$$\overline{g}_{se}\overline{R}^a_{bgd}\frac{\partial x^i}{\partial x^a}\frac{\partial \overline{x}^s}{\partial x^i} = g_{ri}R^i_{jkl}\frac{\partial x^r}{\partial \overline{x}^e}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d}.$$

In the left member, $s := a$ leads to

$$\overline{R}_{ebgd} = R_{rjkl}\frac{\partial x^r}{\partial \overline{x}^e}\frac{\partial x^j}{\partial \overline{x}^b}\frac{\partial x^k}{\partial \overline{x}^g}\frac{\partial x^l}{\partial \overline{x}^d}.$$

We may say that Riemann symbol of second kind is a mixed curvature $(1, 3)$-type tensor and the Riemann symbol of first kind is a curvature covariant $(0, 4)$-type tensor. □

13.9 Some Remarks on the Mathematical Language of Metric-Affine Gravity

Before developing the Metric-Affine Theories of Gravity, in particular the Teleparallel Gravity, we want to show that the above mathematical tools can be perfectly adopted for a formulation more general than the standard metric one. In this section, we summarize some useful results and formulas which will be fully developed in the next chapter.

Let us start from the Levi-Civita connection. In coordinates, if

$$X := \frac{\partial}{\partial x^i}, Y := \frac{\partial}{\partial x^j}, Z := \frac{\partial}{\partial x^k},$$

$$\nabla_X Y = \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} := \Gamma^r_{ij} \frac{\partial}{\partial x^r},$$

$$g_{ij} := g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right),$$

then

$$2g\left(\Gamma^r_{ij} \frac{\partial}{\partial x^r}, \frac{\partial}{\partial x^k}\right) = \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ik}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k},$$

that is, the Christoffel symbols emerge. Let us modify this statement in the new notation. Instead of

$$X := \frac{\partial}{\partial x^i}$$

we write

$$X := \partial_i.$$

Therefore we have:
if

$$X := \partial_i, \; Y := \partial_j, \; Z := \partial_k,$$

$$\nabla_X Y = \nabla_{\partial_i} \partial_j := \Gamma^r_{ij} \partial_r,$$

$$g_{ij} := g\left(\partial_i, \partial_j\right),$$

then

$$2g\left(\Gamma^r_{ij} \partial_r, \partial_k\right) = \partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}.$$

Another example is

$$\nabla_X Y = X^i \left(\frac{\partial Y^k}{\partial x^i} + \Gamma^k_{ij} Y^j \right) \frac{\partial}{\partial x^k}.$$

This formula is written in the new form as

$$\nabla_X Y = X^i \left(\partial_i Y^k + \Gamma^k_{ij} Y^j \right) \partial_k.$$

The Riemann curvature tensor formula

$$R^i_{ljk} = \frac{\partial \Gamma^i_{lk}}{\partial x^j} - \frac{\partial \Gamma^i_{lj}}{\partial x^k} + \Gamma^s_{lk}\Gamma^i_{sj} - \Gamma^s_{lj}\Gamma^i_{sk}$$

is written now in a simpler way

$$R^i_{ljk} = \partial_j \Gamma^i_{lk} - \partial_k \Gamma^i_{lj} + \Gamma^s_{lk}\Gamma^i_{sj} - \Gamma^s_{lj}\Gamma^i_{sk}.$$

It is

$$Y = \sum_{i=1}^{n} Y^i \frac{\partial}{\partial x^i}$$

and

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i) \frac{\partial}{\partial x^i}$$

are replaced by

$$Y = \sum_{i=1}^{n} Y^i \partial_i = Y^i \partial_i$$

and

$$\nabla_X Y := \sum_{i=1}^{n} X(Y^i)\partial_i = X(Y^i)\partial_i,$$

if we take into consideration Einstein's summation formula.

13.9.1 From Latin to Greek Indexes and Vice Versa

An important remark is in order at this point. In the following chapter, we will consider Greek indexes for coordinates and Latin indexes for tetrads. This formalism allows us to pass from space-time (holonomic) coordinates to tetradic (anholonomic) coordinates. The physical reasons for this change is a wide field of debate which we will take into account in the next chapter. For a detailed discussion see [9, 45].

With this perspective in mind, let us define objects in space-time coordinates with Greek indexes. Accordingly, the metric invariant, given by the metric tensor $g_{\mu\nu}$, is written in the form

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu .$$

The covariant derivative ∇_ν acts on contravariant vector fields, denoted here by A^μ, as

$$\nabla_\nu A^\mu := \partial_\nu A^\mu + \Gamma^\mu_{\ \nu\alpha} A^\alpha$$

and for covariant vector fields, i.e. for differential 1-forms, as

$$\nabla_\nu A_\mu := \partial_\nu A_\mu - \Gamma^\alpha_{\ \nu\mu} A_\alpha .$$

Covariant derivative ∇ acts on a $(1, 1)$ tensor $A^\alpha_{\ \beta}$ in the following way:

$$\nabla_\mu A^\alpha_{\ \beta} := \partial_\mu A^\alpha_{\ \beta} + \Gamma^\alpha_{\ \mu\rho} A^\rho_{\ \beta} - \Gamma^\rho_{\ \mu\beta} A^\alpha_{\ \rho} .$$

Other covariant derivatives can be written accordingly.

An important result is related to the commutator of the covariant derivatives. We will prove here that this is a way to introduce the curvature tensor, that is, the geometric curvature for our set of coordinates, and also the torsion of the connection. It is

$$[\nabla_\nu, \nabla_\mu]V^\alpha = \nabla_\nu\nabla_\mu V^\alpha - \nabla_\mu\nabla_\nu V^\alpha = \nabla_\nu(\partial_\mu V^\alpha + \Gamma^\alpha_{\ \mu\beta}V^\beta) - \nabla_\mu(\partial_\nu V^\alpha + \Gamma^\alpha_{\ \nu\beta}V^\beta) =$$

$$= \partial_\nu(\partial_\mu V^\alpha + \Gamma^\alpha_{\ \mu\beta}V^\beta) + \Gamma^\alpha_{\nu\rho}(\partial_\mu V^\rho + \Gamma^\rho_{\ \mu\beta}V^\beta) - \Gamma^\rho_{\nu\mu}(\partial_\rho V^\alpha + \Gamma^\alpha_{\ \rho\beta}V^\beta) -$$

$$- \partial_\mu(\partial_\nu V^\alpha + \Gamma^\alpha_{\ \nu\beta}V^\beta) - \Gamma^\alpha_{\mu\rho}(\partial_\nu V^\rho + \Gamma^\rho_{\ \nu\beta}V^\beta) + \Gamma^\rho_{\mu\nu}(\partial_\rho V^\alpha + \Gamma^\alpha_{\ \rho\beta}V^\beta) =$$

$$= R^\alpha_{\mu\nu\beta}V^\beta - T^\rho_{\nu\mu}\nabla_\rho V^\alpha .$$

The four terms containing the quantities $\Gamma^\alpha_{\nu\rho}\partial_\mu V^\rho,\ \Gamma^\alpha_{\mu\rho}\partial_\nu V^\rho$ appear both with alternate sign and then they cancel out. It is: $\partial_\nu\partial_\mu V^\alpha - \partial_\mu\partial_\nu V^\alpha = 0$.

The formula we obtained is

$$[\nabla_\nu, \nabla_\mu]V^\alpha = R^\alpha_{\mu\nu\beta}V^\beta - T^\rho_{\nu\mu}\nabla_\rho V^\alpha .$$

We have an equivalent formula for 1-forms,

$$[\nabla_\nu, \nabla_\mu]V_\alpha = -R^\beta_{\mu\nu\alpha}V_\beta - T^\rho_{\nu\mu}\nabla_\rho V_\alpha$$

as it can be easily derived. Here, the curvature tensor is

$$R^{\alpha}_{\beta\gamma\delta} = \partial_\gamma \Gamma^{\alpha}_{\beta\delta} - \partial_\delta \Gamma^{\alpha}_{\beta\gamma} + \Gamma^{s}_{\beta\delta}\Gamma^{\alpha}_{s\gamma} - \Gamma^{s}_{\beta\gamma}\Gamma^{\alpha}_{s\delta}$$

and a simple computation leads to

$$R^{\alpha}_{\beta\gamma\delta} = -R^{\alpha}_{\beta\delta\gamma}.$$

In the same way, we can derive the Bianchi identities. In case of symmetric connection, we have the first Bianchi identity:

$$R^{\alpha}_{\beta\gamma\delta} + R^{\alpha}_{\gamma\delta\beta} + R^{\alpha}_{\delta\beta\gamma} = 0\,,$$

and the second Bianchi identity:

$$\nabla_\alpha R^{\mu}_{\nu\beta\gamma} + \nabla_\beta R^{\mu}_{\nu\gamma\alpha} + \nabla_\gamma R^{\mu}_{\nu\alpha\beta} = 0\,,$$

as presented in the Differential Geometry chapter. Clearly, in presence of torsion, the right-hand side of the second Bianchi identity is not zero but it results in a combination of Riemann and torsion tensors. It is

$$\nabla_\lambda R^{\alpha}_{\beta\mu\nu} + \nabla_\mu R^{\alpha}_{\beta\nu\lambda} + \nabla_\nu R^{\alpha}_{\beta\lambda\mu} = -T^{\rho}_{\mu\lambda} R^{\alpha}_{\beta\rho\nu} - T^{\rho}_{\nu\lambda} R^{\alpha}_{\beta\rho\mu} - T^{\rho}_{\nu\mu} R^{\alpha}_{\beta\rho\lambda}\,,$$

as it can be derived with some algebra. This exercise is left to the reader and will be reconsidered in the next chapter.

We can rise and lower indexes using the metric tensor. From $A^{\alpha}{}_{\beta}$, we get the (2, 0) tensor $A^{\gamma\alpha}$ by the rule

$$A^{\gamma\alpha} = g^{\gamma\lambda} A^{\alpha}{}_{\lambda},$$

or the (0, 2) tensor $A_{\gamma\alpha}$ by the rule

$$A_{\gamma\alpha} = g_{\gamma\lambda} A^{\lambda}{}_{\alpha}.$$

The case of objects as $A_{\alpha\beta\gamma}$ is more interesting. We can rise each index but we have to take care of its position. Let us see some examples.

$$A^{\mu}{}_{\beta\gamma} := g^{\mu\alpha} A_{\alpha\beta\gamma},$$

while in the case of the second index, we have

$$A_{\alpha}{}^{\mu}{}_{\gamma} := g^{\mu\beta} A_{\alpha\beta\gamma}.$$

The same, in the case of the third index,

$$A^{\mu}_{\alpha\beta} := g^{\mu\gamma} A_{\alpha\beta\gamma}.$$

Therefore an expression like

$$A^{\mu}_{\alpha\ \gamma} - A^{\mu}_{\alpha\gamma}$$

is not necessary 0 in the case when the initial $A_{\alpha\beta\gamma}$ has supplementary properties as we will see below.

We can proceed in the same way for the commutator of covariant derivatives. We obtain

$$[\nabla_{\mu}, \nabla_{\nu}]V_{\alpha} = R_{\alpha\beta\mu\nu}V^{\beta} - T^{\beta}_{\mu\nu}\nabla_{\beta}V_{\alpha}$$

and

$$[\nabla_{\mu}, \nabla_{\nu}]V_{\alpha} = -R_{\beta\alpha\mu\nu}V^{\beta} - T^{\beta}_{\mu\nu}\nabla_{\beta}V_{\alpha},$$

so we can deduce

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu}.$$

Since the antisymmetry is relevant in presence of torsion, this property can play a very important role in Metric-Affine Theories of Gravity, as we will see in the next chapter.

In these theories, the components of the Levi-Civita connection are denoted by

$$\left\{ \begin{matrix} \rho \\ \mu\nu \end{matrix} \right\} := \frac{1}{2}g^{\rho\lambda}(\partial_{\mu}g_{\lambda\nu} + \partial_{\nu}g_{\mu\lambda} - \partial_{\lambda}g_{\mu\nu}).$$

Let us remember that, if the connection is Levi-Civita, we have

$$\nabla_{\mu}g_{\nu\rho} = 0.$$

This property is the isometry already discussed above in another framework. Isometry is related, in this way, to the Equivalence Principle. If isometry does not hold, that is if the connection ∇ is arbitrary and not compatible with the metric, we can define the non-metricity tensor

$$Q_{\mu\nu\rho} := \nabla_{\mu}g_{\nu\rho} = \partial_{\mu}g_{\nu\rho} - \Gamma^{\lambda}_{\nu\mu}g_{\lambda\rho} - \Gamma^{\lambda}_{\rho\mu}g_{\nu\lambda}.$$

We can adopt the notation:

$$A_{(\mu\nu)} = A_{\mu\nu} + A_{\nu\mu}$$

for symmetric objects and

$$A_{[\mu\nu]} = A_{\mu\nu} - A_{\nu\mu}$$

for anti-symmetric objects. Using the first formula, we can write the non-metricity tensor as

$$Q_{\mu\nu\rho} := \nabla_{\mu}g_{\nu\rho} \equiv \partial_{\mu}g_{\nu\rho} - \Gamma^{\lambda}{}_{(\nu|\mu}g_{\rho)\lambda}.$$

Using the second formula, one can write the torsion tensor $T^{\alpha}_{\beta\gamma} = \Gamma^{\alpha}_{\beta\gamma} - \Gamma^{\alpha}_{\gamma\beta}$ in the form

$$T^{\alpha}_{\beta\gamma} = \Gamma^{\alpha}_{[\beta\gamma]}.$$

An important remark is in order at this point. If torsion and non-metricity are different from zero, the Levi-Civita connection is just a part of a more general connection which is

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2}g^{\alpha\beta}(\partial_\mu g_{\nu\beta} + \partial_\nu g_{\mu\beta} - \partial_\beta g_{\mu\nu}) + \frac{1}{2}(T_{\mu}{}^{\alpha}{}_{\nu} + T_{\nu}{}^{\alpha}{}_{\mu} - T^{\alpha}_{\mu\nu}) - \frac{1}{2}(Q^{\rho}{}_{\mu\nu} - Q_{\mu}{}^{\rho}{}_{\nu} - Q_{\nu}{}^{\rho}{}_{\mu}).$$

In the next chapter, we will discuss in detail this extension.

Another important notion we are going to discuss is that of the *tetrad fields* or the *vierbein*, that is, the "four-legs" [9]. They are the fundamental variables of Teleparallel Gravity.

We can suppose that, at each point of the tangent space of a given 4-manifold, there are four vector fields (e^{μ}_{a}), with $\mu, a \in \{0, 1, 2, 3\}$, that is, $\{e^{\mu}_{0}, e^{\mu}_{1}, e^{\mu}_{2}, e^{\mu}_{3}\}$ which form an orthonormal basis with respect to the metric tensor $g_{\mu\nu}$. This means the following relation is fulfilled:

$$g_{\mu\nu}e^{\mu}_{a}e^{\nu}_{b} = \eta_{ab}.$$

Here $\eta^{\alpha\beta} = \eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$ are the components of the flat Minkowski metric of signature $(- + ++)$. If we define by (e^{a}_{μ}), the inverse of the vector (e^{μ}_{a}), we can compute the metric components by the formulas

$$g_{\mu\nu} = e^{a}_{\mu}e^{b}_{\nu}\eta_{ab}, \quad g^{\mu\nu} = e^{\mu}_{a}e^{\nu}_{b}\eta^{ab}.$$

Consider a Lorentz rotation. Therefore consider a matrix who keeps invariant the Minkowski metric written in the form

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2.$$

This Lorentz rotation transforms the tetrad field according to the rule

$$e^{a}_{\mu} \to \Lambda^{a}_{b}e^{b}_{\mu}$$

because this is the only way we can preserve the orthonormal basis. This suggests that, if we have a matrix with two kinds of indexes, Latin (anholonomic) and Greek (holonomic), the tetrad field helps us to construct tensor quantities with the same kind of indexes. An example is

$$A^{a_1a_2\ldots a_n}{}_{b_1b_2\ldots b_n} = e^{a_1}_{\alpha_1}e^{a_2}_{\alpha_2}\ldots e^{a_n}_{\alpha_n}A^{\alpha_1\alpha_2\ldots\alpha_n}{}_{\beta_1\beta_2\ldots\beta_n}e^{\beta_1}_{b_1}e^{\beta_2}_{b_2}\ldots e^{\beta_n}_{b_n}.$$

These arguments will be reconsidered in detail in the forthcoming chapter.

Finally, we can introduce new coefficients $\omega^{a}_{\mu b}$ to extend the parallel transport to more general quantities. We adopt the following notation. For Greek indexes, we use Γ coefficients and for Latin indexes, we use ω coefficients. We define ω the *spin connection*. An example is the following:

$$\nabla_\mu A^{a\nu} = \partial_\mu A^{a\nu} + \Gamma^\nu_{\mu\rho} A^{a\rho} + \omega^a_{\mu b} A^{b\nu}.$$

Tetrad field can change the index type. If we want to commute with respect to the covariant derivative of a tensor, we get the condition:

$$\partial_\mu e^a_\nu + \omega^a_{\mu b} e^b_\nu - \Gamma^\rho_{\mu\nu} e^a_\rho = 0,$$

which means the vanishing of the new covariant derivative of the tetrad.

In other words, we can define the *Lorentz covariant derivative* with respect to the Latin indexes only, denoted by $\mathcal{D}_\mu$. It is

$$\Gamma^\rho_{\mu\nu} = e^\rho_a (\partial_\mu e^a_\nu + \omega^a_{\mu b} e^b_\nu) \equiv e^\rho_a \mathcal{D}_\mu.$$

The spin connection can be defined directly by the formula

$$\omega^a_{\mu b} = e^\rho_a \Gamma^\rho_{\mu\nu} e^\nu_b - e^\nu_b \partial_\mu e^a_\nu.$$

We have now two curvatures tensors, one related to Γ and one related to ω:

$$R^\alpha_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^s_{\beta\delta} \Gamma^\alpha_{s\gamma} - \Gamma^s_{\beta\gamma} \Gamma^\alpha_{s\delta},$$

$$R^\alpha_{\beta\gamma\delta}(\omega) = \partial_\gamma \omega^\alpha_{\beta\delta} - \partial_\delta \omega^\alpha_{\beta\gamma} + \omega^s_{\beta\delta} \omega^\alpha_{s\gamma} - \omega^s_{\beta\gamma} \omega^\alpha_{s\delta}\,.$$

They are related to each other by

$$R^\alpha_{\beta\gamma\delta}(\Gamma) = e^\alpha_a R^a_{b\gamma\delta}(\omega) e^b_\beta.$$

In particular, if one finds the spin connection ω corresponding to the Levi-Civita connection $\Gamma(G)$ of a given metric g, a tetrad description of General Relativity is obtained.

To conclude, we have now all the ingredients to deal with Metric-Affine Theories of Gravity which we will discuss in detail in the next chapter.

Chapter 14
Metric-Affine Theories of Gravity

The development of Physics, like the development of any science, is a continuous one.

Owen Chamberlain

General Relativity is not the end of the story. Several issues, ranging from Quantum Gravity to the Dark Side of the Universe need to be addressed in a self-consistent theory. Here we want to summarize some of these approaches. In particular, we want to show that the same theory, General Relativity, can be represented in different ways. This fact is questioning the basic foundations of the theory like the Equivalence Principle, the Lorentz Invariance, and so on. According to this picture, new possibilities can be taken into account in view of further theoretical and experimental developments. Topics of this chapter are more advanced with respect to the rest of the book and could be considered for some short advanced lectures.

14.1 A Survey on Theories of Gravity

In the nineteenth century, Newtonian mechanics was considered as the best theory to describe gravity, since it was successfully exploited in everyday life and capable of describing the motion of planets and stars. However, in this period, there was a great cultural ferment around *non-Euclidean geometries* starting from fundamental works by Gauss, Lobachevsky, Bolyai, Riemann, Bianchi, Ricci-Curbastro, and several others [29]. The Euclidean framework, the arena for classical Physics, was overtaken by the formulation of elliptical and hyperbolic geometries, stemming out from a rigorous axiomatic reformulation of the geometry foundations. Indeed, two

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_14

approaches were more and more emerging from these studies: (i) *affine geometry*, introduced by Euler in 1748, deriving from the Latin word *affinis*, meaning related, and after promoted by Möbius, Klein, and Weyl. It essentially focuses on the study of *parallel lines*, based on the validity or redefinition of the fifth Euclid postulate, and on the *affine transformations* [24]; (ii) *metric geometry*, introduced by Fréchet and Hausdorff, relies on a *metric function* defining the concept of distance between any two points, members of a non-empty set [43].

Einstein, inspired by this line of nonconformist ideas, arrived, in 1915, to the formulation of General Relativity (GR) [91]. This new vision of gravitational interaction, ruled by the space–time curvature, took time to be comprehended and accepted by the scientific community owed to the outcoming effects, retained to be too small to be measured and observed at that time. The well-known subsequent astronomical confirmations constituted the success of GR [195].

Although GR was not yet validated, some authors were however eager to advance proposals to extend it with the aim to fulfil more general purposes. In 1918, Weyl started to study the question on how to connect gravity and electromagnetism in a single and coherent geometric theory. To achieve this objective, he took into account an additional gauge field, which singles out a unique *length connection*, whose four additional degrees of freedom (DoFs) are associated to the electromagnetic potentials. In the Weyl geometry, besides the GR connection, there is also an additional length connection, which is symmetric, metric incompatible, and gauge invariant. The consequence is that, during a parallel transport, both direction and length of vectors vary [192, 193]. However, Weyl's theory revealed to be in conflict not only with some experiences (for example, the frequency of spectral lines of atomic clocks depends on the location and past histories of the atoms), but even in a more fundamental way with Quantum Mechanics (e.g. masses of particles rest on their past histories).

In 1930, along the same line of thinking, Einstein himself proposed some modifications to his theory. Fascinated by teleparallelism and tetrad formalism, he initiated a prolific and extensive correspondence mainly with Cartan, Weitzenböck, and Lanczos [109, 110, 188]. Indeed, since the tetrad fields posses sixteen independent components, he associated ten of them to the metric tensor, whereas the other six were believed to be linked to a separate connection, entrusted to model electromagnetic potentials. Unfortunately, he failed in his attempt, but his studies shed new light on the importance of additional DoFs, which theoretically belong to the Lorentz group and physically are a consequence of the local Lorentz invariance.

In 1922, Cartan concentrated on a different direction, since he considered a natural extension of GR constituted not only by the Levi-Civita connection but also by the *torsion tensor* (essentially the antisymmetric part of a metric compatible affine connection). Given these premises, he developed all the ensuing geometric formulation [74], where he suggested that the torsion can be physically related to the intrinsic (quantum) angular momentum of matter and it vanishes as soon as vacuum regions are considered [69–73].

Around 1960, Kibble and Sciama revisited the theory formulating it within the gauge theory of the Poincaré group [119, 123, 170]. This approach can be extended to the more general affine group, leading thus to the metric-affine gauge gravity [115].

There have been other proposals and experiments to probe the fundamental nature of gravitation, in particular, to establish its geometric structure. In this vein, it was growing the awareness that affinity and metricity could be considered as two different and independent concepts, where the affine connection could not respect a priori the metric postulate. This perspective is considered into the so-called *Palatini approach*, where GR is constituted by a metric tensor and an affine connection, considered as two different geometric structures. Varying the Einstein–Hilbert action with respect to the metric, the Einstein field equations are recovered; whereas, varying it with respect to the affine connection, the metric compatibility condition is naturally obtained and the Levi-Civita connection is restored [157]. This shows that GR structure entails metric compatibility, and the affine connection can be considered as a true dynamical field. As it is well known, this coincidence does not work for extensions of GR as $f(R)$ [10].

These considerations led to the development of theories of gravity beyond the Einstein picture, where the field equations, besides the scalar curvature, can be formulated in terms of other geometric invariants. Furthermore, the affine connections were not considered anymore with an ancillary role with respect to the metric tensor, but, contrarily, they assumed a dynamical fundamental role. These approaches gave rise to the current variegated realm of the *Extended* and *Alternative Theories of Gravity* (see e.g. [45, 51, 55, 61, 78, 147, 148, 150, 172]).

In any case, GR revealed to be extraordinarily successful because passed several astrophysical and cosmological observational tests like the Solar System tests [85, 146, 195], the direct detection of gravitational waves [2–4, 19, 168], the recent black hole imaging [6, 94–99, 156], and other robust confirmations [76, 111, 128, 174].

Despite these achievements, the theory exhibits various pathological issues, still matter of debate, suggesting that approaches beyond Einstein gravity should be pursued [55]. For example, from galaxies to cosmic evolution, the infrared behaviour of gravitational field presents several shortcomings mainly related to the *Dark Matter* [15, 48, 65, 139] and *Dark Energy* problems [51, 68, 80], and the tensions in cosmological parameters like H_0 [11, 87]. At ultraviolet scales, the lack of renormalizability and unitarity of gravitational field points out that a final, self-consistent theory of Quantum Gravity is not at hand [63, 106, 107, 153, 164, 180, 181].

In general, the formulation of a new theory of gravity to solve the above issues is not a simple task. There are principles, constraints, mathematical consistencies, and the agreement with observations that any novel approach must necessarily fulfill before being accepted as a self-consistent picture. This is one of the thorny theoretical challenges of modern physics.

In this perspective, we want to focus our attention on GR and its dynamically equivalent formulations, in view to put in evidence similarities and differences towards a unified view of gravitational interaction.

This chapter is organized as follows: in Sect. 14.2, we describe the general framework, represented by the metric-affine theories of gravity, in which the so-called *Geometric Trinity of Gravity* [23] can be formulated (Sect. 14.3). In Sect. 14.4, we provide the mathematical tools necessary for the formulation of any theory of gravity. In Sect. 14.5, we discuss the Geometric Trinity of Gravity in terms of their Lagrangian equivalence. Section 14.6 is devoted to the field equations derived from the second Bianchi identity. In Sect. 14.7, we analyse the spherically symmetric solutions in the three equivalent formulations, recovering, in all of them, the Schwarzschild metric and the Birkhoff theorem. Finally, Sect. 14.8 is devoted to the conclusions and to the discussion of some crucial issues necessary for any self-consistent formulation of gravity.

Notations. We adopt the metric signature $(-, +, +, +)$. Greek indexes take values 0, 1, 2, 3, while the lowercase Latin ones 1, 2, 3. Capital Latin letters indicate tetrad indexes The flat metric is indicated by $\eta^{\alpha\beta} = \eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$. The determinant of the metric $g_{\mu\nu}$ is denoted by g. Round (square) brackets around a pair of indexes stand for symmetrization (antisymmetrization) procedure, i.e. $A_{(ij)} = A_{ij} + A_{ji}$ $(A_{[ij]} = A_{ij} - A_{ji})$.

In this chapter, we will number the formulas to avoid to report them several times along the discussion.

14.2 Metric-Affine Theories of Gravity

A first extension of Einstein gravity starts by generalizing the affine connections which cannot be strictly Levi-Civita. A metric-affine theory is defined by the following triplet $\{\mathcal{M}, g_{\mu\nu}, \Gamma^{\rho}{}_{\mu\nu}\}$, where $\mathcal{M}$ is a four-dimensional space–time manifold, $g_{\mu\nu}$ is a rank-two symmetric tensor (with 10 independent components), and $\Gamma^{\rho}{}_{\mu\nu}$ is the affine connection (endowed with 64 independent components). *A priori* there is no relation between the metric and the affine connection, where the former is responsible to describe the *casual structure*, whereas the latter deals with the *geodesic structure*. As it is well known, the structures coincide if the Equivalence Principle is the basic foundation of the theory [55, 195].

Let us consider now a system of coordinates $\{x_0, x_1, x_2, x_3\}$ defined on $\mathcal{M}$, where x_0 labels the time and $\{x_1, x_2, x_3\}$ the space. The metric $g_{\mu\nu}$ defines the line element $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}$. The notion of covariant derivative ∇ acts on a generic $(1, 1)$ tensor in the following way [190]:

$$\nabla_{\mu} A^{\alpha}{}_{\beta} := \partial_{\mu} A^{\alpha}{}_{\beta} - \Gamma^{\rho}{}_{\beta\mu} A^{\alpha}{}_{\rho} + \Gamma^{\alpha}{}_{\rho\mu} A^{\rho}{}_{\beta}. \tag{14.1}$$

The components of the general affine connection $\Gamma^{\rho}{}_{\mu\nu}$ can be uniquely decomposed as follows [18]:

$$\Gamma^{\rho}{}_{\mu\nu} := \begin{Bmatrix} \rho \\ \mu\nu \end{Bmatrix} + K^{\rho}{}_{\mu\nu} + L^{\rho}{}_{\mu\nu}, \tag{14.2}$$

where $\begin{Bmatrix} \rho \\ \mu\nu \end{Bmatrix}$ is the Levi-Civita connection, $K^{\rho}{}_{\mu\nu}$ is the *contortion tensor*, and $L^{\rho}{}_{\mu\nu}$ is the *disformation tensor*, whose explicit expressions are [18]

$$\begin{Bmatrix} \rho \\ \mu\nu \end{Bmatrix} := \frac{1}{2} g^{\rho\lambda}(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}), \tag{14.3a}$$

$$K^{\rho}{}_{\mu\nu} := \frac{1}{2}(T_{\mu}{}^{\rho}{}_{\nu} + T_{\nu}{}^{\rho}{}_{\mu} - T^{\rho}{}_{\mu\nu}), \tag{14.3b}$$

$$L^{\rho}{}_{\mu\nu} := \frac{1}{2}(Q^{\rho}{}_{\mu\nu} - Q_{\mu}{}^{\rho}{}_{\nu} - Q_{\nu}{}^{\rho}{}_{\mu}). \tag{14.3c}$$

Notice that, while the Levi-Civita part is non-tensorial, the contortion and disformation terms are tensors under changes of coordinates. The three main geometric objects (related to the dynamics) are the *curvature tensor* $R^{\mu}{}_{\nu\alpha\beta}$, the *torsion tensor* $T^{\rho}{}_{\mu\nu}$, and the *non-metricity tensor* $Q_{\rho\mu\nu}$. Their explicit expressions in terms of metric and connections are [18]

$$R^{\mu}{}_{\nu\rho\sigma} := \partial_\rho \Gamma^{\mu}{}_{\nu\sigma} - \partial_\sigma \Gamma^{\mu}{}_{\nu\rho} + \Gamma^{\mu}{}_{\tau\rho}\Gamma^{\tau}{}_{\nu\sigma} - \Gamma^{\mu}{}_{\tau\sigma}\Gamma^{\tau}{}_{\nu\rho}, \tag{14.4a}$$

$$T^{\mu}{}_{\nu\rho} := \Gamma^{\mu}{}_{[\rho\nu]} \equiv \Gamma^{\mu}{}_{\rho\nu} - \Gamma^{\mu}{}_{\nu\rho}, \tag{14.4b}$$

$$Q_{\mu\nu\rho} := \nabla_\mu g_{\nu\rho} \equiv \partial_\mu g_{\nu\rho} - \Gamma^{\lambda}{}_{(\nu|\mu} g_{\rho)\lambda} = \partial_\mu g_{\nu\rho} - \Gamma^{\lambda}{}_{\nu\mu} g_{\lambda\rho} - \Gamma^{\lambda}{}_{\rho\mu} g_{\nu\lambda} \neq 0. \tag{14.4c}$$

These tensors show the following symmetries:

$$R^{\mu}{}_{\nu\rho\sigma} = -R^{\mu}{}_{\nu\sigma\rho}, \tag{14.5a}$$

$$T^{\mu}{}_{\nu\rho} = -T^{\mu}{}_{\rho\nu}, \tag{14.5b}$$

$$Q_{\mu\nu\rho} = Q_{\mu\rho\nu}. \tag{14.5c}$$

The above geometric quantities, differently affect the parallel transport of a vector on a manifold. We have that:

- *curvature* manifests its presence when a vector is parallel transported along a closed curve on a non-flat background and comes back to its starting point forming a non-null angle with its initial position;
- *torsion* entails a rotational geometry, where the parallel transport of two vectors is antisymmetric by exchanging the transported vectors and the direction of transport. This property results in the non-closure of parallelograms;
- *non-metricity* is responsible to alter the length of the vectors along the transport.

In a generic metric-affine theory, all these effects can work together and could have also further meanings corresponding to physical quantities (e.g. the torsion tensor is linked to the spin in the Einstein–Cartan theory [119]).

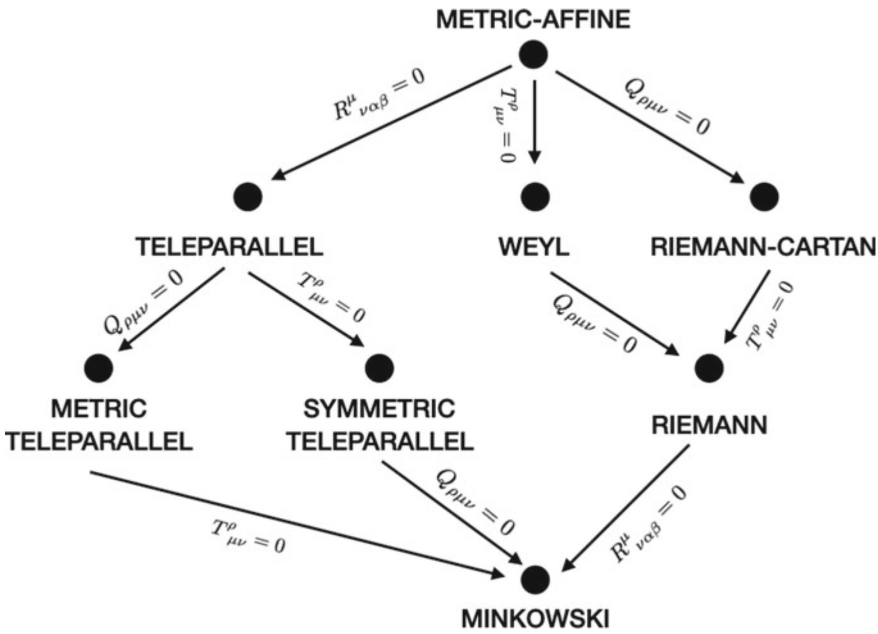


Fig. 14.1 A possible classification of theories emerging from metric-affine gravity

In general, the following *Bianchi identities* hold [18]:

$$R^\mu{}_{[\nu\rho\sigma]} = \nabla_{[\nu} T^\mu{}_{\rho\sigma]} + T^\mu{}_{\alpha[\nu} T^\alpha{}_{\rho\sigma]}, \tag{14.6a}$$

$$\nabla_{[\alpha} R^\mu{}_{|\nu|\rho\sigma]} = -R^\mu{}_{\nu\tau[\alpha} T^\tau{}_{\rho\sigma]}, \tag{14.6b}$$

which involve only curvature and torsion tensors.

Metric-affine theories are a broad class of theories whose dynamics can be related to the tensors $R^\mu{}_{\nu\rho\sigma}$, $T^\mu{}_{\nu\rho}$, and $Q_{\mu\nu\rho}$ which can be grossly classified as in Fig. 14.1.

(1) The *Riemann–Cartan geometry* is expressed in terms of metric compatible curvature and torsion tensors. It is also known in the literature as U_4 or Einstein–Cartan–Sciama–Kibble theory, where the role of the torsion is deputed to model the quantum spin effects present in the matter [116–119].
(2) The *Weyl geometry* is constructed by vanishing the torsion, where curvature and non-metricity are the only surviving geometric objects. This theory has interesting implications and, moreover, it represents also the origin of the $U(1)$ gauge theory [194].
(3) *Teleparallel geometries* are curvatureless and are based on the concept of *Fernparallelismus* or *parallelism at distance*, because two vectors can be immediately seen whether they are parallel or not, since the parallel transport of vectors becomes independent of the path [9]. They admit two special subclasses, represented by

(3.1) *metric teleparallel theories* expressed only in terms of the torsion tensor;
(3.2) *symmetric teleparallel theories* described only by the non-metricity tensor.

(4) The *Riemannian geometry* represents the first arena within which Einstein framed his theory, constructed only upon the curvature tensor [140, 162].
(5) The *Minkowski geometry* is obtained by setting curvature, torsion, and non-metricity to zero, where the flat metric $\eta_{\mu\nu}$, as well as zero affine connections, are adopted. This is the arena of Special Relativity [140, 162].

14.3 The Geometric Trinity of Gravity

Among the possible metric-affine gravity theories, Riemannian and teleparallel models are particularly interesting. GR is an example of Riemannian geometry, whereas the so-called metric teleparallel equivalent of GR (TEGR) and symmetric teleparallel equivalent of GR (STEGR) are examples of teleparallel geometries. See Fig. 14.2. These three theories constitute the so-called *Geometric Trinity of Gravity*.

A fundamental property of TEGR is that torsion replaces curvature for dynamics and it is able to provide the same descriptions of the gravitational interaction under a different perspective. In GR, the geometric curvature is entrusted to model the gravitational force, whereas geodesics coincide with the free-falling test particle's trajectories. On the other hand, in TEGR, the gravitational interaction emerges through the torsion tensor and acts as a (gauge) force. This is the reason why, in the teleparallel framework, the concept of geodesics is replaced by force equations, analogously to what happens in electrodynamics where the Lorentz force is present. STEGR shares several similar properties with TEGR. In this theory, one requires that curvature and torsion are both zero, and gravitational dynamics is attributed to the non-metricity tensor.

GR is described in terms of the metric $g_{\mu\nu}$; TEGR in terms of the tetrads $e^A_{\ \mu}$ (accounting for the dynamical description of gravity) and spin connection $\omega^A_{\ B\mu}$ (flat

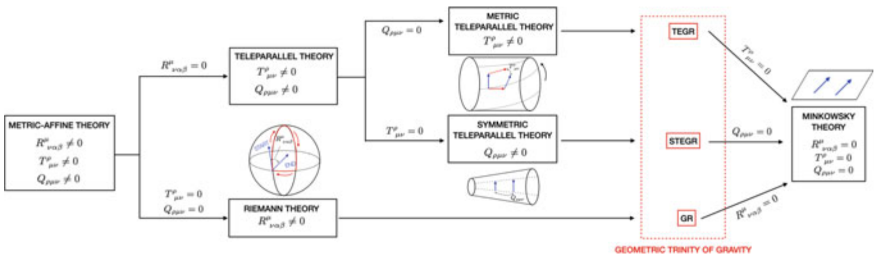


Fig. 14.2 The Geometric Trinity of Gravity framework and the dynamical role of tensor invariants. *Curvature* rules how the tangent space rolls a curve on a manifold; *torsion* how the tangent space twists around a curve when we parallel transport two vectors along each other; *non-metricity* encodes the variation of vectors' length when they moved along a curve [18]

connection outlining inertial effects); STEGR embodies the Palatini idea where metric $g_{\mu\nu}$ and affine connection $\Gamma^{\mu}_{\ \alpha\beta}$ are two separated dynamical structures. Like other fundamental interactions in Nature, gravitation can be reformulated as a gauge theory through TEGR and STEGR. The most peculiar property of gravitation seems to be its *universal character* that all objects, regardless of their internal structure, feel this force, which is encoded in the *Equivalence Principle* of GR. In the teleparallel formulations, the Equivalence Principle is sometimes claimed to be not valid in the literature, instead, we will underline how it can be recovered in such theories, also if it does not lie at their foundation. This fact is extremely relevant because, if the Equivalence Principle were shown to be violated at some fundamental level, the final theory of gravitation could be non-metric.

In these equivalent pictures, we can define alternative ways of representing the gravitational field, accounting for the same DoFs, related to specific geometric invariants: the Ricci curvature scalar R, the torsion scalar T, and the non-metricity scalar Q. In this sense, GR, TEGR, and STEGR give rise to the Geometric Trinity of Gravity.

Similarly to GR where we can extend to $f(R)$ gravity, $f(T)$ and $f(Q)$ gravity are the extensions of TEGR and STEGR, respectively. It is worth noticing that, in general, the equivalence among the three representations is not valid anymore among the extensions, because they give rise to dynamics with different DoFs (see Fig. 14.3). In particular, in $f(R)$ gravity to fourth-order field equations, in metric representation, whereas $f(T)$ and $f(Q)$ still remains of second-order [45, 55, 160]. In addition, in $f(T)$ and $f(Q)$, we cannot choose, in general, a gauge to simplify the calculations, as in the cases of TEGR and STEGR. On the contrary, we have to consider field equations for the spin connection in $f(T)$ and affine field equations for $f(Q)$ [45, 83, 160]. In the following, we shall develop these points in detail.

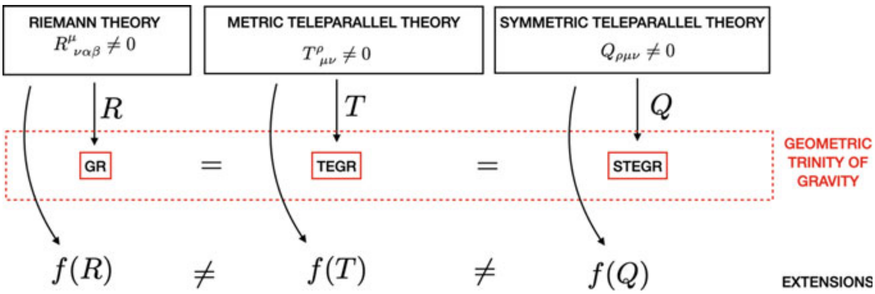


Fig. 14.3 The Geometric Trinity of Gravity and related extensions. The equivalence holds only for theories linear in the scalar invariants. Extensions can involve further degrees of freedom which lead to the breaking of equivalence among different representations of gravity. It can be restored identifying correct boundary terms

14.4 Tetrads and Spin Connection

Before going into details of Trinity Gravity, some considerations on the mathematical structure are in order. To define a theory of gravity, we need to fix the underlying geometry, the transformation properties, and the set of observables. GR is based on the metric tensor from which we can construct the Levi-Civita connection, and finally the curvature, which encodes the gravitational dynamics. The possibility to relate the metric and the geodesic structure, which essentially coincide, relies on the validity of the Equivalence Principle [55]. However, GR can be reformulated also in terms of tetrad [9, 144, 160] and spin connection formalisms [9, 129], giving rise to the teleparallel equivalent GR. In Sect. 14.4.1, we describe in detail the tetrad formalism, whereas, in Sect. 14.4.2, we introduce the spin connection.

14.4.1 The Tetrad Formalism

The geometric setting of any theory of gravity occurs in the tangent bundle, a natural construction always present in any smooth space–time. In fact, at each point of the space–time, it is possible to construct the tangent space attached to it, which is a vector (fibre bundle) space. On the respective domains of definition, any vector or covector can be expressed in terms of a general linear orthonormal frame called tetrads.

A tetrad field is a geometric construction, which permits to easily carry out the calculations on the tangent space. Physically, they represent the standard laboratory apparatus of the observer for carrying out the measurements in space and time. Using a tetrad field means to adopt a *Lagrangian point of view*, which entails to follow an individual fluid parcel as it moves through space and time. A tetrad field establishes a relationship between the manifold and its tangent spaces. This geometric structure is always present, independently of any prior gravity-model assumption. The theoretical framework intervenes to characterize the gravitational effects occurring in this frame.

We first introduce the definition and properties of the tetrads (see Sect. 14.4.1), and then we present their anhonolonomy structure (see Sect. 14.4.1) and its importance in the first Cartan structure equation (see Sect. 14.4.1). Finally, we describe preferred frames represented by the inertial class and trivial tetrads (see Sect. 14.4.1).

Tetrads: Definition and Properties

Let us assign a general metric space–time $(\mathcal{M}, g_{\mu\nu})$, being $\mathcal{M}$ a four-dimensional differential manifold of class C^∞, whose tangent spaces $T_p\mathcal{M}$, at each point $p \in \mathcal{M}$, are Minkowski space–times with metric η_{AB}, and $g_{\mu\nu}$ the symmetric metric tensor. In these hypotheses, there exists a *compatible atlas of charts* $\mathcal{A}$, being an open covering of $\mathcal{M}$. Therefore, for each $p \in \mathcal{M}$, there exists a chart $(\mathcal{U}, \varphi)$ of domain $\mathcal{U}$, being an open neighbourhood of p, and a coordinate map $\varphi : \mathcal{U} \to \varphi(\mathcal{U}) \subseteq \mathbb{R}^4$ (being a homeomorphism). In addition, for all $(\mathcal{U}, \varphi), (\mathcal{V}, \psi) \in \mathcal{A}$, the map $\psi \circ \varphi^{-1} : \varphi(\mathcal{U} \cap \mathcal{V}) \to \psi(\mathcal{U} \cap \mathcal{V})$ is a C^∞-diffeomorphism called *coordinate transformation*. Therefore, to each point $p \in \mathcal{M}$, we can associate its *coordinates* by $(x^0, x^1, x^2, x^3) := \varphi(p) \in \mathbb{R}^4$ [162]. Defined the coordinate x^μ-axes in $\mathbb{R}^4$, it is possible to construct the related coordinate curves γ_{x^μ} on $\mathcal{M}$ via the use of the charts. Therefore, all the parallel curves to coordinate axes in $\mathbb{R}^4$ forms the related grid on $\mathcal{M}$, which permits to uniquely identify the space–time location of all points.

A natural differentiable basis or *holonomic basis* of each tangent bundle $T_p\mathcal{M}$ is given by a sets of vectors tangent to the coordinate lines at each point p, i.e.

$$\partial_\mu := \left(\frac{\partial}{\partial x^\mu}\right)_p, \tag{14.7}$$

as well as for covector fields defined on the cotangent bundle $T_p^*\mathcal{M}$ (set of all linear maps $\alpha : T_p\mathcal{M} \to \mathbb{R}$) we have the following basis $\{\mathrm{d}x^\mu\}$ applied to the point $p \in \mathcal{M}$, which satisfies the orthonormality condition

$$\mathrm{d}x^\mu \partial_\nu = \delta^\mu_\nu. \tag{14.8}$$

The tangent $T_p\mathcal{M}$ and cotangent $T_p^*\mathcal{M}$ bundles in $p \in \mathcal{M}$ are related through the metrics $g_{\mu\nu}$ and η_{AB}.

Every vector or covector applied to a point $p \in \mathcal{M}$ can be expressed in terms of the natural basis. Therefore, we can define a set of orthonormal vectors and covectors, which can be related to the natural basis through [9]

$$e_A := e_A^{\ \mu}\partial_\mu, \qquad e^A := e^A_{\ \mu}\mathrm{d}x^\mu, \tag{14.9}$$

where the set of coefficients $\left\{e_A^{\ \mu}\right\}$ are called *tetrads* and belong to the linear group of all real 4×4 invertible matrices $GL(4, \mathbb{R})$. The tetrads act as a *soldering agent* between the general manifold (Greek indexes) and the Minkowski space–time (capital Latin indexes[1]) as follows:

$$g_{\mu\nu} = \eta_{AB} e^A_{\ \mu} e^B_{\ \nu}, \qquad \eta_{AB} = g_{\mu\nu} e_A^{\ \mu} e_B^{\ \nu}. \tag{14.10}$$

Therefore, a tetrad field is a linear frame gluing together the coordinate charts on $\mathcal{M}$ to the preferred orthonormal basis e_A on the tangent space, where calculations can be carried out in a considerably simplified manner. As $g_{\mu\nu}$ varies from point to point on the manifold $\mathcal{M}$, the vierbein $e_A^{\ \mu}$ do the same. Calculating the determinant of (14.10), we obtain $-g = e^2$, where e denotes the determinant of $e_A^{\ \mu}$ and it is negative owed

[1] Sometimes, capital Latin indexes, referring to local coordinate indexes, are also indicated by an over hat on the Greek indexes, i.e. $e^A_{\ \mu} = e^{\hat{\nu}}_{\ \mu}$.

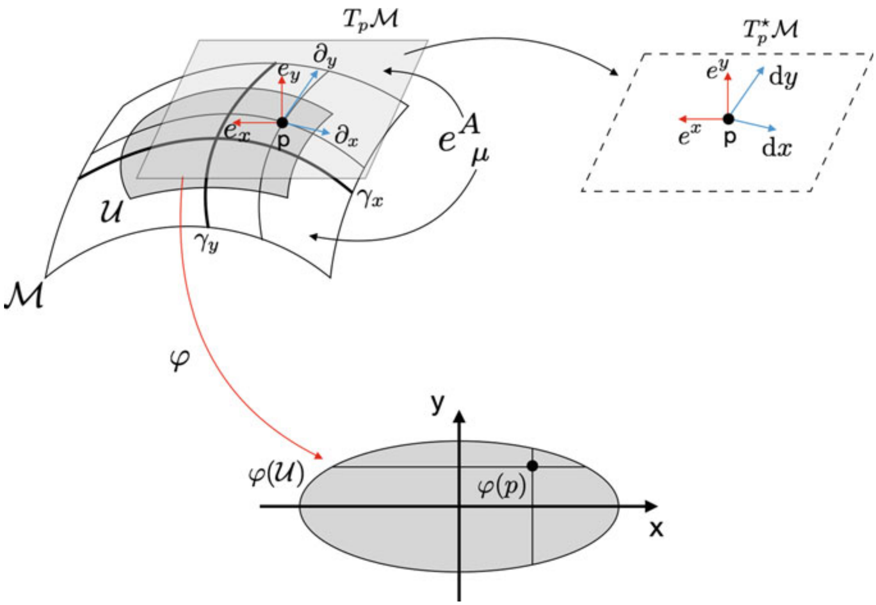


Fig. 14.4 Two-dimensional picture to explain the tetrad formalism. Tetrads $e^A_{\ \mu}$ solder the coordinate chart $(\mathcal{U}, \varphi)$ on the manifold $\mathcal{M}$ to the orthonormal basis $\{e_x, e_y\}$ in the tangent bundle $T_p\mathcal{M}$. They represent also the coefficients in the natural (holonomic) basis $\{\partial_x, \partial_y\}$. The coordinate map φ assign at each point $p \in \mathcal{U} \subseteq \mathcal{M}$ the coordinates $\varphi(p) = (x, y) \in \varphi(\mathcal{U}) \subseteq \mathbb{R}^4$. Passing from $T_p\mathcal{M}$ to the cotangent bundle $T_p^*\mathcal{M}$ through $g_{\mu\nu}$ and η_{AB}, the natural basis $\{dx, dy\}$ is transformed into the orthonormal basis $\{e^x, e^y\}$ through the use of tetrads $e_A^{\ \mu}$

to the signature of η_{AB}. Generally speaking, we note that the vierbein represents the square root of the metric. In Fig. 14.4, the tetrads together with their properties are displayed.

Anholonomy of Tetrad Frames

Let us now analyse one of the consequences in using of the tetrad fields. A general tetrad basis $\{e_A\}$ (cf. Eq. (14.9)) satisfies the commutation relation [9, 129]

$$\begin{aligned}
[e_A, e_B] &:= e_A e_B - e_B e_A \\
&= (e_A^{\ \mu}\partial_\mu)(e_B^{\ \nu}\partial_\nu) - (e_B^{\ \nu}\partial_\nu)(e_A^{\ \mu}\partial_\mu) \\
&= \left[e_A^{\ \mu} e^C_{\ \nu}(\partial_\mu e_B^{\ \nu}) - e_B^{\ \nu} e^C_{\ \mu}(\partial_\nu e_B^{\ \mu})\right] e_C \\
&= e_A^{\ \mu} e_B^{\ \nu}\left[\partial_\nu e^C_{\ \mu} - \partial_\mu e^C_{\ \nu}\right] e_C \\
&= f^{\ C}_{AB} e_C,
\end{aligned} \tag{14.11}$$

where we have set

$$f^{C}_{\ AB} := e_A^{\ \mu} e_B^{\ \nu} \left[\partial_\nu e^C_{\ \mu} - \partial_\mu e^C_{\ \nu} \right], \tag{14.12}$$

which are known as *structure constants* or *coefficients of anholonomy* related to the frame $\{e_A\}$. They quantify the failure of the parallelogram closure generated by the vectors e_A and e_B. In general, when $f^{C}_{\ AB} \neq 0$, the tetrad basis is *anholonomic* or *non-trivial*, and the coefficients of anholonomy specify how much they depart from being holonomic. This approach reveals important properties of the underlying geometric framework on which we are working. In GR, they have been used in the *Bianchi classification*, which leads to 11 possible different space–times, useful to develop cosmological models [130, 166, 177].

The First Cartan Structure Equation

Given a 1-form ω and defined $\mathrm{d}\omega$ as the exterior derivative, it can be written in components as

$$\mathrm{d}\omega = \partial_\mu \omega_\nu \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu, \tag{14.13}$$

where $\wedge$ is the external product defined as

$$\mathrm{d}x^\mu \wedge \mathrm{d}x^\nu = \mathrm{d}x^\mu \otimes \mathrm{d}x^\nu - \mathrm{d}x^\nu \otimes \mathrm{d}x^\mu, \tag{14.14}$$

with $\otimes$ the tensorial product. Due to the antisymmetry of the exterior product and the Schwarz theorem, we have $\mathrm{d}^2\omega = 0$ thanks to the *Poincaré lemma* [144].

We consider the 2-form $\mathrm{d}\omega$ applied to two vectors $u = u^\mu \partial_\mu,\ v = v^\nu \partial_\nu$, which can be written as [86]

$$\mathrm{d}\omega(u, v) = u\omega(v) - v\omega(u) - \omega([u, v]_{\mathcal{L}}), \tag{14.15}$$

where

$$\mathrm{d}\omega(u, v) := \partial_\mu \omega_\nu (u^\mu v^\nu - u^\nu v^\mu), \tag{14.16a}$$

$$u\omega(v) := u^\mu v^\nu \partial_\mu \omega_\nu + u^\mu \omega_\nu \partial_\mu v^\nu, \tag{14.16b}$$

$$\omega([u, v]_{\mathcal{L}}) := \omega_\nu (u^\mu \partial_\mu v^\nu - v^\mu \partial_\mu u^\nu), \tag{14.16c}$$

with $\omega = \omega_\mu \mathrm{d}x^\mu$ and $[u, v]_{\mathcal{L}} \equiv (\mathcal{L}_u v) := (u^\mu \partial_\mu v^\nu - v^\mu \partial_\mu u^\nu)\partial_\nu$. It is the Lie bracket or the Lie derivative of the vector field v with respect to the vector field u. It is important to note that $\mathrm{d}\omega(u, v)$ produces a scalar.

If we consider the tetrad basis $\{e_A\}$ and take $\omega = e^A$, then we have the following relation [86]:

$$\begin{aligned}\left\{\mathrm{d}e^{C}(e_A, e_B)\right\} e_C &= \left\{e_A[e^C(e_B)] - e_B[e^C(e_A)]\right.\\ &\quad \left.-e^C([e_A, e_B]_{\mathcal{L}})\right\} e_C\\ &= -e^C([e_A, e_B]^L_{\mathcal{L}} e_L)e_C\\ &= -[e_A, e_B]_{\mathcal{L}}. \end{aligned} \tag{14.17}$$

Assigned a general metric-compatible affine connection $\Gamma^{\lambda}_{\ \alpha\beta}$, and the associated covariant derivative ∇, we have

$$\nabla_{e_A} e_B = \gamma^{C}_{\ AB} e_C, \tag{14.18}$$

where $\gamma^{C}_{\ AB}$ are the *Ricci rotation coefficients*, which measure the rotation of all frame tetrads when moved in various directions, encoding thus gravitational and non-inertial effects [86, 144]. When we use the natural basis, they reduce to the affine connection $\Gamma^{C}_{\ AB}$. It is important to note that such coefficients arise also in a flat space–time when, generally, non-linear coordinates are exploited, since they give rise to non-inertial effects. In particular, in the considered tetrad basis, they assume the following expression and symmetries [144]:

$$\begin{aligned}\gamma_{\lambda\nu\mu} &:= e^{A}_{\ \mu} e^{B}_{\ \lambda} \nabla_A (e_\nu)_B\\ &= -e^{A}_{\ \mu}(e_\nu)_B \nabla_A e^{B}_{\ \lambda}\\ &= -e^{A}_{\ \mu} e^{B}_{\ \nu} \nabla_A (e_\lambda)_B = -\gamma_{\nu\lambda\mu}, \end{aligned} \tag{14.19}$$

where we have used the compatibility condition in the last equality. $\gamma^{C}_{\ AB}$ can be seen as the action of the *connection 1-forms* $\omega^{C}_{\ B}$ on the tetrad basis e_A, i.e. [86]

$$\gamma^{C}_{\ AB} = \omega^{C}_{\ B}(e_A) \quad \Leftrightarrow \quad \omega^{C}_{\ B} = \gamma^{C}_{\ AB} e^A. \tag{14.20}$$

Since we know that $\nabla_\mu \partial_\nu = \Gamma^{\lambda}_{\ \mu\nu}\partial_\lambda$, if we consider the commutator of ∇_μ and ∂_ν, we obtain

$$\begin{aligned}[\nabla_\mu, \partial_\nu] &= \nabla_\mu \partial_\nu - \nabla_\nu \partial_\mu\\ &= \left(\Gamma^{\lambda}_{\ \mu\nu} - \Gamma^{\lambda}_{\ \nu\mu}\right)\partial_\lambda\\ &= T^{\lambda}_{\ \mu\nu}\partial_\lambda, \end{aligned} \tag{14.21}$$

where $T^{\lambda}_{\ \mu\nu}$ is the torsion tensor measuring the antisymmetry of the affine connections. In a coordinate-independent approach, the torsion T (associated to the covariant derivative ∇) is a $(1, 2)$-type tensor, which acts on pairs of vectors (v, u) to give another vector according to the following relation [86, 140]:

$$T(v, u) := \nabla_v u - \nabla_u v - [v, u]_{\mathcal{L}}. \tag{14.22}$$

Applying Eq. (14.22) to $\{e_A\}$, exploiting Eq. (14.17), and considering $\omega^C_{\ B}(e_A) = (\omega^C_{\ D} \otimes e^D)(e_A, e_B)$, we obtain

$$\begin{aligned} T(e_A, e_B) &= \nabla_{e_A} e_B - \nabla_{e_B} e_A - [e_A, e_B]_{\mathcal{L}} \\ &= \left[\omega^C_{\ B}(e_A) - \omega^C_{\ A}(e_B) + \mathrm{d}e^C(e_A, e_B)\right] e_C \\ &= \left[(\omega^C_{\ D} \wedge e^D + \mathrm{d}e^C)(e_A, e_B)\right] e_C . \end{aligned} \tag{14.23}$$

Defined $\Omega^C := \omega^C_{\ D} \wedge e^D + \mathrm{d}e^C$ as the *torsion differential 2-form*, Eq. (14.23) can be written as [86, 144]

$$T = \Omega^C \otimes e_C, \tag{14.24}$$

which is the *first Cartan structure equation*. In the case of Riemann geometry, namely when the torsion vanishes, Eq. (14.24) becomes [9, 129]

$$\begin{aligned} de^C &:= -\omega^C_{\ A} \wedge e^A \\ &= -\frac{1}{2}\left(\gamma^C_{\ AB} - \gamma^C_{\ BA}\right) e^A \wedge e^B \\ &= -\frac{1}{2} e_A^{\ \mu} e_B^{\ \nu} (\partial_\nu e^C_{\ \mu} - \partial_\mu e^C_{\ \nu}) e^A \wedge e^B \\ &= -\frac{1}{2} f^C_{\ AB} e^A \wedge e^B , \end{aligned} \tag{14.25}$$

where the anhonolonomy coefficients emerge as an antisymmetric combination of the Ricci rotation coefficients. They are also related to the curls of the tetrad vector derivatives, as occur to the components of a differential 2-form [140, 144].

Inertial Frames and Trivial Tetrads

Among the different frames, a special class is represented by the *inertial frames*, which can be denoted by $\{e'_A\}$, whose coefficients of anhonolonomy $f'^C_{\ AB}$ locally satisfy the condition $f'^C_{\ AB} = 0$. For Eq. (14.25) we have $\mathrm{d}e'^A = 0$, which is locally exact and can be written as $e'^A = \mathrm{d}x'^A$ and therefore it is holonomic. Therefore, all coordinate bases belong to this family. It is worth noting that *this is not a local property, but it holds everywhere for all frames being part of this inertial class* [9].

In absence of gravitation, the anholonomy is only caused by inertial forces present in these frames. The metric $g_{\mu\nu}$ reduces to the Minkowski metric $\eta_{\mu\nu}$. In all coordinate systems, $\eta_{\mu\nu}$ is a function of the space–time point, and independently of whether $\{e_A\}$ is holonomic (inertial) or not. In this case, tetrads always relate the tangent Minkowski space to a Minkowski space–time

$$\eta_{AB} = \eta_{\mu\nu} e_A^{\ \mu} e_B^{\ \nu}. \tag{14.26}$$

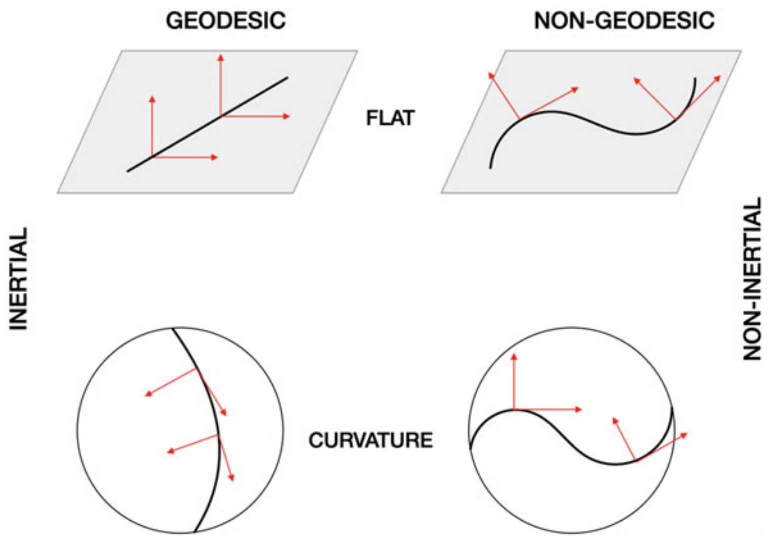


Fig. 14.5 This figure shows how tetrads behave in terms of inertial and gravitational effects. When no gravity is present, and we consider inertial effects only (i.e. we move along geodesics), we obtain trivial (holonomic) tetrads, whereas, when non-inertial contributions take place (i.e. following non-geodesic orbits), the tetrads become anholomic. The situation is analogue when gravitation is switched on. Along geodesic, we obtain inertial frames, whereas along non-geodesic trajectories we have the most general anholomic frames

These are the frames appearing in Special Relativity, which are usually called *trivial frames* or *trivial tetrads*. They are very useful when we deal with spaces involving torsion [129]. Of course, in the absence of inertial forces, the class of inertial frames is, consequently, represented by vanishing structure coefficients. These concepts are sketched in Fig. 14.5.

14.4.2 The Spin Connection

The spin connection plays a fundamental role when we deal with tetrads, because it encodes the inertial effects occurring in the considered frame. Let us briefly recall the fundamental properties of the Lorentz group (see Sect. 14.4.2), then we discuss the associated Lorentz algebra as well as its properties (see Sect. 14.4.2). Lorentz connections will be first introduced under a mathematical point of view (see Sect. 14.4.2) together with the fundamental tetrad postulate (see Sect. 14.4.2), and then the same subject will be considered under a physical perspective (see Sect. 14.4.2).

The Lorentz Group

Electromagnetism is framed under the standard of Special Relativity by postulating [162]:

(1) *the optical isotropy principle:* all inertial frames are optically isotropic, i.e. the light propagates in these frames with velocity $c = 1/\sqrt{\epsilon_0 \mu_0}$ in any direction;
(2) *the principle of relativity:* the laws of physics assume the same form in all inertial reference frames.

Given two inertial frames and assuming that one is moving with respect to the other with uniform velocity $\boldsymbol{v} := (v^1, v^2, v^3)$, the *Lorentz transformation* is a linear (affine) map relating the temporal and spatial coordinates of the two inertial observers [162]

$$\Lambda^{\mu}{}_{\nu} : x^{\mu} \longrightarrow x'^{\mu} = \Lambda^{\mu}{}_{\nu}(x) x^{\nu}, \tag{14.27}$$

which leaves invariant the following quadratic form:

$$\eta_{\mu\nu} x^{\mu} x^{\nu} = -t^2 + x^2 + y^2 + z^2. \tag{14.28}$$

A general Lorentz transformation is given by [191]

$$\Lambda^{\alpha}{}_{\beta} = \mathcal{G} \cdot \begin{bmatrix} \gamma & -\gamma \mathcal{R}^i{}_j \frac{v^j}{c} \\ -\gamma \mathcal{R}^i{}_j \frac{v^j}{c} & \mathcal{R}^i{}_j \left(\delta^i_j + (\gamma - 1) \frac{v^i v^j}{v^2} \right) \end{bmatrix}, \tag{14.29}$$

where $v := \sqrt{(v^1)^2 + (v^2)^2 + (v^3)^2}$ is the modulus of the spatial velocity $\boldsymbol{v}$, $\gamma := (1 - \frac{v^2}{c^2})^{-1/2}$ is the Lorentz factor, $\mathcal{R}^i{}_j$ is a rotation matrix, and $\mathcal{G}$ is one of the following operators

$$1 := \text{diag}(1, 1, 1, 1),\ P := \text{diag}(1, -1, -1, -1),\ T := \text{diag}(-1, 1, 1, 1),\ P \cdot T.$$

They are the unitary, parity, and time reversal operators, respectively. The expression of $\Lambda^{\alpha}{}_{\beta}$ shows that a Lorentz transformation is defined in terms of six parameters: three related to the rotation angles and the other three to the components of the spatial velocity v.

The set of all Lorentz transformations of Minkowski space–time forms the *(homogeneous) Lorentz orthogonal group* $O(1, 3)$. The requirement (14.28), together with (14.27), entails, in matrix notation, that $\eta = \Lambda^T \eta \Lambda$. This gives rise to $\det^2 \Lambda = 1$, namely *proper* ($\det \Lambda = 1$) and *improper* ($\det \Lambda = -1$) Lorentz transformations, which can be further subdivided (cf. Eq. (14.29)) in *orthochronous* ($\Lambda^0{}_0 \geq 1$) and *non-orthochronous* ($\Lambda^0{}_0 \leq -1$) [134, 191]. The proper orthochronous Lorentz transformations form the *restricted Lorentz special orthogonal group* $SO^+(1, 3)$. Therefore, *the Lorentz group is a six-dimensional, non-compact, non-Abelian, and real Lie group endowed with four connected components* [134, 191]. The Lorentz group is closely involved in all known fundamental laws of Nature describing the related symmetries of space and time. In particular, in GR, we consider the *local Lorentz invariance*, because in every small enough region of space–time, thanks to the Equivalence Principle, the gravitational effects can be neglected, i.e. this occurs in the *local inertial frame* (LIF), which permits to recover the Special Relativity physics.

At each point of a Riemannian space–time, the metric $g_{\mu\nu}$ determines a tetrad up to the local Lorentz transformations in the tangent space. In other words, a tetrad vector (covector) base $\{e_A\}$ ($\left\{e^A\right\}$) is not unique, because it is always possible to find another base $\{\bar{e}_A\}$ ($\left\{\bar{e}^A\right\}$) by performing a local Lorentz transformation, namely

$$\bar{e}^A_{\ \mu} = \Lambda^A_{\ B} e^B_{\ \mu}, \tag{14.30}$$

such that

$$g_{\mu\nu} = \eta_{AB}\bar{e}^A_{\ \mu}\bar{e}^B_{\ \nu} \qquad \eta_{AB} = g_{\mu\nu}\bar{e}_A^{\ \mu}\bar{e}_B^{\ \nu}. \tag{14.31}$$

The Lorentz Algebra

Another important feature of the Lorentz group is that it admits a *Lorentz algebra* $\mathfrak{L}$ [134, 191]. If we consider an infinitesimal transformation in $SO^+(1,3)$, we have

$$\Lambda^\alpha_{\ \beta} = \delta^\alpha_\beta + \omega^\alpha_{\ \beta} + \mathcal{O}[(\omega^\alpha_{\ \beta})^2]. \tag{14.32}$$

Applying $\eta = \Lambda^T\eta\Lambda$, at linear order in $\omega^\alpha_{\ \beta}$, $\omega_{\mu\nu} = -\omega_{\nu\mu}$ is an antisymmetric 4×4 matrix with six independent indexes. Therefore, we can associate six generators to the Lorentz algebra labelled by J_{AB}, with $J_{AB} = -J_{BA}$ [134], where each of them can be expressed in the *4-vector representation* by a 4×4 matrix as follows:

$$(J_{AB})^C_{\ D} := 2i\eta_{[B|D}\delta^C_{A]} = i(\eta_{BD}\delta^C_A - \eta_{AD}\delta^C_B). \tag{14.33}$$

Each element of the Lorentz group can be written as [134]

$$\Lambda = e^{\frac{i}{2}\omega_{AB}J^{AB}}. \tag{14.34}$$

The Derivation of Lorentz Connection

Some geometric objects with an established behaviour may lose the covariant character under point-dependent transformations, e.g. ordinary derivative of covariant objects. In order to supply for this defective behaviour, it is fundamental to introduce *connections* ω_μ fulfilling the following properties: (i) they behave like vectors in the space–time indexes; (ii) they act as non-tensor in the algebraic indexes to compensate this effect and to reestablish the correct tensorial trend. The *linear connections* fulfilling these requirements belong to the subgroup $SO^+(1,3)$ of $GL(4,\mathbb{R})$, and they are dubbed *Lorentz connections*. It is worth noticing that all Lorentz connections exhibit the presence of torsion (see Ref. [9], and discussions therein).

A Lorentz connection, also known as *spin connection*, ω_μ is a 1-form acting in the Lorentz algebra, namely

$$\omega_\mu : J_{AB} \in \mathcal{L} \longrightarrow \omega_\mu := \frac{1}{2}\omega^{AB}{}_\mu J_{AB}, \tag{14.35}$$

where $\omega^{AB}{}_\mu$ are the spin connection coefficients, which are antisymmetric in the AB indexes owed to the antisymmetry of J_{AB}, i.e. $\omega^{AB}{}_\mu = -\omega^{BA}{}_\mu$. This permits to introduce the *Fock–Ivanenko covariant derivative* [9, 60]

$$\mathcal{D}_\mu := \partial_\mu - \omega_\mu = \partial_\mu - \frac{i}{2}\omega^{AB}{}_\mu J_{AB}. \tag{14.36}$$

where J_{AB} is the generator in the appropriate representation of the Lorentz group. The right member of Eq. (14.36) acts only on tangent (algebraic) space indexes. If we apply Eq. (14.33) to the field e^C, we obtain

$$\begin{aligned} \mathcal{D}_\mu e^C &= \partial_\mu e^C - \frac{i}{2}\omega^{AB}{}_\mu \left[i(\eta_{BD}\delta^C_A - \eta_{AD}\delta^C_B)\right] e^D \\ &= \partial_\mu e^C + \frac{1}{2}\left[\omega^A{}_{D\mu}\delta^C_A + \omega^B{}_{D\mu}\delta^C_B\right] e^D \\ &= \partial_\mu e^C + \omega^C{}_{D\mu} e^D. \end{aligned} \tag{14.37}$$

Considering Eq. (14.37) and splitting e^A by Eq. (14.9), we obtain the following expressions :

$$\begin{aligned} \mathcal{D}_\mu(e^C_\lambda dx^\lambda) &= \mathcal{D}_\mu(e^C_\lambda)\mathrm{d}x^\lambda + e^C_\lambda \mathcal{D}_\mu(\mathrm{d}x^\lambda) \\ &= \mathcal{D}_\mu(e^C_\lambda)\mathrm{d}x^\lambda + e^C_\lambda(\delta^\lambda_\mu + e_E^\lambda e^D_\mu \omega^E{}_{D\rho}\mathrm{d}x^\rho) \\ &= \mathcal{D}_\mu(e^C_\lambda)\mathrm{d}x^\lambda + e^C_\mu, \end{aligned} \tag{14.38a}$$

$$\begin{aligned} \mathcal{D}_\mu(e^C_\lambda dx^\lambda) &= \partial_\mu(e^C_\lambda dx^\lambda) + \omega^C{}_{D\mu} e^D_\lambda \mathrm{d}x^\lambda \\ &= \partial_\mu(e^C_\lambda)dx^\lambda + e^C_\mu + \omega^C{}_{D\mu} e^D_\lambda \mathrm{d}x^\lambda. \end{aligned} \tag{14.38b}$$

Equating Eq. (14.38a) with (14.38b), we obtain

$$\mathcal{D}_\mu(e^C_\lambda) = \partial_\mu(e^C_\lambda) + \omega^C{}_{D\mu} e^D_\lambda. \tag{14.39}$$

The Tetrad Postulate

In non-coordinate bases $\{e_A\}$, the covariant derivative $\tilde{\nabla}$ of an algebraic (1,1) tensor X^A_B can be written in terms of the spin connection as

$$\tilde{\nabla}_\mu X^A_{\ B} := \partial_\mu + \omega^A_{\ C\mu} X^C_{\ B} - \omega^C_{\ B\mu} X^A_{\ C}. \tag{14.40}$$

Instead, the covariant derivative of a vector V, considered in the coordinate bases $\{\partial_\mu\}$, is

$$\begin{aligned} \nabla V &= (\nabla_\mu V^\nu) \mathrm{d}x^\mu \otimes \partial_\nu \\ &= (\partial_\mu V + \Gamma^\nu_{\ \mu\lambda} V^\lambda) \mathrm{d}x^\mu \otimes \partial_\nu. \end{aligned} \tag{14.41}$$

If we consider now the same vector V written in a mixed basis, tetrad and coordinate basis, gives

$$\begin{aligned} \tilde{\nabla} V &= (\tilde{\nabla}_\mu V^A) \mathrm{d}x^\mu \otimes e_A \\ &= (\partial_\mu V^A + \omega^A_{\ B\mu} V^B) \mathrm{d}x^\mu \otimes e_A \\ &= [\partial_\mu (e^A_{\ \lambda} V^\lambda) + \omega^A_{\ B\mu} e^B_{\ \lambda} V^\lambda] \mathrm{d}x^\mu \otimes (e_A^{\ \nu} \partial_\nu) \\ &= [\partial_\mu V^\nu + (e_A^{\ \nu} \partial_\mu e^A_{\ \lambda} + \omega^A_{\ B\mu} e_A^{\ \nu} e^B_{\ \lambda}) V^\lambda] \mathrm{d}x^\mu \otimes \partial_\nu \\ &= [\partial_\mu V^\nu + (e_A^{\ \nu} \mathcal{D}_\mu e^A_{\ \lambda}) V^\lambda] \mathrm{d}x^\mu \otimes \partial_\nu. \end{aligned} \tag{14.42}$$

This is a crucial point, because the operations (14.41) and (14.42) are in principle distinct. However, it is reasonable to assume $\nabla \equiv \tilde{\nabla}$, because the same covariant derivative of a vector cannot change in terms of which type of basis one chooses. This is the so-called *tetrad postulate*, which is valid for any affine connection, defined on a smooth manifold $\mathcal{M}$, and no metric is involved.

Therefore, it implies (cf. Eqs. (14.41) and (14.42))

$$\Gamma^\lambda_{\ \mu\nu} \equiv e_A^{\ \lambda} \mathcal{D}_\mu e^A_{\ \nu}. \tag{14.43}$$

This identity entails several significant implications on the spin connections: (i) since it does not possess a tensorial character, it acquires a non-homogeneous term under the Fock–Ivanenko covariant derivative owed to the affine connection [9]; (ii) a spin connection is naturally induced by the affine connection; (iii) it can be also regarded as the gauge field generated by local Lorentz transformations; (iv) inverting Eq. (14.43) with respect to the spin connection, we obtain [9]

$$\omega^A_{\ B\mu} = e^A_{\ \lambda} e_B^{\ \nu} \Gamma^\lambda_{\ \mu\nu} + e^A_{\ \sigma} \partial_\mu e_B^{\ \sigma} \equiv e^A_{\ \nu} \nabla_\mu e_B^{\ \nu}; \tag{14.44}$$

(v) according to Eq. (14.44), the connection 1-form $\omega^C_{\ B}$ (cf. Eqs. (14.20), (14.19)) can be written as

$$\omega^{AB} = \omega^{AB}_{\ \ \ \mu} \mathrm{d}x^\mu, \tag{14.45}$$

and *the Ricci rotation coefficients are the space–time indexes of the spin connection components*; (vi) the covariant derivative of the tetrad, expressed in terms of the affine and spin connections, vanishes identically (cf. Eq. (14.44)), namely

$$\nabla_\mu e^A_\nu = \partial_\mu e^A{}_\nu - \Gamma^\lambda{}_{\mu\nu} e^A{}_\lambda + \omega^A{}_{B\mu} e^B{}_\nu = 0; \tag{14.46}$$

(vii) we note that ∇_μ is the covariant derivative linked to the connection $\Gamma^\lambda_{\mu\nu}$ when acts on external indexes and can be defined for tensorial fields, whereas the Fock–Ivanenko derivative $\mathcal{D}_\mu$ acts on internal indexes and can be defined for all tensorial and spinorial fields [9]; $(viii)$ from the metric compatibility condition, we obtain a sort of consistency check given by (cf. Eqs. (14.39) and (14.43))

$$\begin{aligned} 0 = \nabla_\lambda g_{\mu\nu} &= \partial_\lambda g_{\mu\nu} - \Gamma^\sigma{}_{\lambda\mu} g_{\sigma\nu} - \Gamma^\sigma{}_{\lambda\nu} g_{\mu\sigma} \\ &= \partial_\lambda (e^A{}_\mu e^B{}_\nu \eta_{AB}) - e_A^{\ \sigma} g_{\sigma\nu} \mathcal{D}_\lambda e^A{}_\mu - e_A^{\ \sigma} g_{\mu\sigma} \mathcal{D}_\lambda e^A{}_\nu \\ &= -e^A{}_\nu e^D{}_\mu (\omega_{AD\lambda} - \omega_{DA\lambda}), \end{aligned} \tag{14.47}$$

which implies $\omega_{AB\mu} = -\omega_{BA\mu}$, i.e. $\omega_{AB\mu}$ is Lorentzian. If the metric postulate (14.47) is not valid, the corresponding spin connection cannot assume values in the Lorentz algebra, because it is not a Lorentz connection [9]. Therefore, we have this equivalence: *metric compatibility holds if and only if we choose a Lorentz connection.*

Physical Considerations on the Lorentz Connection

We have seen how the tetrads transform under local (point-dependent) Lorentz transformations $\Lambda^A{}_B(x)$ (cf. Eq. (14.30)), and now let us apply the same transformations to the spin connections. Let us first consider the inertial frames (see Sect. 14.4.1) $\{e'^A{}_\mu\}$, which, in general coordinates $\{x'^\mu\}$, can be written in the holonomic form $e'^A{}_\mu = \partial_\mu x'^A$, where $x'^A = x'^A(x^\mu)$ is a point-dependent vector. Under a local transformation $x^A = \Lambda^A{}_B(x) x'^B$, we have $e^A_\mu = \Lambda^A{}_B(x) e'^B{}_\mu$ by transforming the vectors x^A and x'^A in the coordinate base $\{\partial_\mu\}$.

Let us evaluate $\partial_\mu x'^A$, which gives ($\partial'_A \equiv \partial/\partial x'^A$). It is

$$\begin{aligned} \partial_\mu x'^A &= \partial_\mu (\Lambda^A{}_B(x) x^B) \\ &= (\partial_\mu x^B) \Lambda^A{}_B(x) + x^B (\partial_\mu \Lambda^A{}_B(x)), \end{aligned} \tag{14.48}$$

$$\partial_\mu x'^A = e'^C{}_\mu \partial'_C x'^A = e'^A{}_\mu = e^C{}_\mu \Lambda^A{}_C(x). \tag{14.49}$$

Therefore, gathering together the above results, we have (using Eq. (14.37) and $\mathcal{D}_\mu x^A = e^A{}_\mu$)

$$e^A{}_\mu = \partial_\mu x^A + \overset{\bullet}{\omega}{}^A{}_{B\mu} x^B \equiv \mathcal{D}_\mu x^A, \tag{14.50}$$

where

$$\overset{\bullet}{\omega}{}^A{}_{B\mu} := \Lambda^A{}_C(x) \partial_\mu \Lambda^C{}_B(x) \tag{14.51}$$

is defined as a *purely inertial spin connection*, because it physically manifests the inertial effects occurring in the Lorentz rotated frame $e^A{}_\mu$. From Eq. (14.51), we

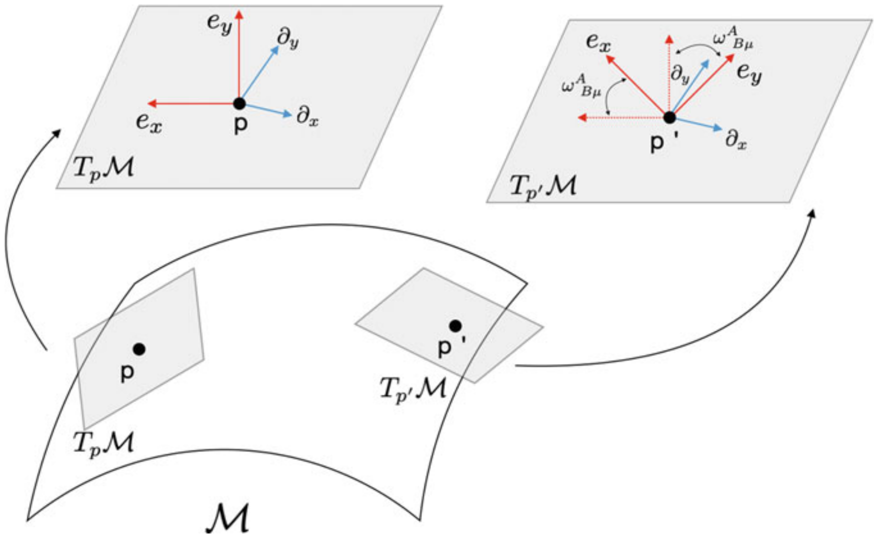


Fig. 14.6 Two-dimensional picture displaying the role of the spin connection $\omega^A{}_{B\mu}$. It translates the inertial effects present in the tetrad anhonolomic frame $\{e_x, e_y\}$. When we pass from $p \in \mathcal{M}$ to $p' \in \mathcal{M}$, the related tetrads in $T_p\mathcal{M}$ and $T_{p'}\mathcal{M}$ exhibit a rotation, modelled by the spin connection. Instead, the inertial holonomic frame $\{\partial_x, \partial_y\}$ does not undergo any rotation, because it admits vanishing spin connection

learn that *the Lorentz connections physically represent the inertial effects present in a given frame. In the inertial frames (i.e. $e'^A{}_\mu = \partial_\mu x'^A$), these effects are absent since the Lorentz connections vanish, $\omega'^{AB}{}_\mu = 0$ for Eq.* (14.50) [129].

To better understand these results, let us consider the transformation of the spin connection under local Lorentz transformations, which leads to [9, 129]

$$\omega^A{}_{B\mu} = \underbrace{\Lambda^A{}_C(x)\omega'^C{}_{D\mu}\Lambda^D{}_C}_{\text{non inertial}} + \underbrace{\Lambda^A{}_C\partial_\mu\Lambda^C{}_B(x)}_{\text{inertial}}. \tag{14.52}$$

When we pass from a frame to another one, there are two distinct contributions: (1) *non-inertial effects* connected with the new frame; and (2) *inertial contributions* due to the rotation of the new frame with respect to the previous one. Therefore, starting from inertial frames ($\omega'^{AB}{}_\mu = 0$), it is possible to obtain a class of non-inertial frames (cf. Eq. (14.52)) via local Lorentz transformations. It is important to note that all these infinite frames are related through global (point-independent) Lorentz transformations $\Lambda^A{}_B = \text{const}$ [129]. In Fig. 14.6, we display the spin connection mechanism.

From Eqs. (14.44) and (14.51), the coefficients of anholonomy (14.11) can be written as ($\overset{\bullet}{\omega}{}^{A}{}_{BC} = \overset{\bullet}{\omega}{}^{A}{}_{B\mu} e^{\mu}_{C}$) [9, 129]

$$f^{C}_{AB} = \overset{\bullet}{\omega}{}^{C}{}_{BA} - \overset{\bullet}{\omega}{}^{C}{}_{AB}. \tag{14.53}$$

From this relation, we can define the spin connection in terms of the structure constants as

$$\overset{\bullet}{\omega}{}^{A}{}_{BC} = \frac{1}{2}(f_{B}{}^{A}{}_{C} + f_{C}{}^{A}{}_{B} - f^{A}_{BC}). \tag{14.54}$$

Let us show now other two important implications of the purely inertial connection. Inserting its expression (14.51) into the definitions of curvature and torsion tensors (cf. Eqs. (14.4a) and (14.4b)), we obtain the following relations [9, 129]:

$$R^{A}{}_{B\mu\nu} = \partial_{\nu}\overset{\bullet}{\omega}{}^{A}{}_{B\mu} - \partial_{\mu}\overset{\bullet}{\omega}{}^{A}{}_{B\nu} + \overset{\bullet}{\omega}{}^{A}{}_{E\nu}\overset{\bullet}{\omega}{}^{E}{}_{B\mu} - \overset{\bullet}{\omega}{}^{A}{}_{E\mu}\overset{\bullet}{\omega}{}^{E}{}_{B\nu} \equiv 0, \tag{14.55a}$$

$$T^{A}_{\nu\mu} = \partial_{\nu}e^{A}_{\mu} - \partial_{\mu}e^{A}_{\nu} + \overset{\bullet}{\omega}{}^{A}{}_{E\nu}e^{E}_{\mu} - \overset{\bullet}{\omega}{}^{A}{}_{E\mu}e^{E}_{\nu}. \tag{14.55b}$$

To prove that Eq. (14.55a) is identically vanishing, we have used the property $\Lambda^{E}_{C}\partial_{\mu}\Lambda^{C}_{E} = -\Lambda^{C}_{E}\partial_{\mu}\Lambda^{E}_{C}$. This result physically tells that inertial effects cannot generate curvature effects, but it is possible to produce only non-null torsional effects, see Eq. (14.55b). However, if we consider trivial tetrads (i.e. $e^{A}_{\mu} = \partial_{\mu}x^{a}$ and $\overset{\bullet}{\omega}{}^{A}{}_{B\mu} = 0$), we can further nullify also the torsion tensor.

14.5 Equivalent Representations of Gravity: The Lagrangian Level

Let us consider now the Geometric Trinity of Gravity, taking into account its mathematical and physical aspects [49]. We discuss first the formulation of gravity according to GR in Sect. 14.5.1. Gravity under the standard of gauge description is considered in Sect. 14.5.2). In Sect. 14.5.3, the basic concepts of GR, TEGR, and STEGR are compared and discussed.

The notations we are going to use are the following: *over-circles* refer to quantities built up on the Levi-Civita connection (i.e. $\mathring{A}^{\mu}{}_{\nu}$), *over-hats* denote quantities related to the teleparallel connection (i.e. $\hat{A}^{\mu}{}_{\nu}$), and *over-diamonds* denote quantities involving non-metricity (i.e. $\overset{\diamond}{A}{}^{\mu}{}_{\nu}$).

14.5.1 Metric Formulation of Gravity: The Case of General Relativity

The GR is the first geometric formulation of gravity in curved space–times. We first recall its basic principles (Sect. 14.5.1), and implications related to the geodesic equations (see Sect. 14.5.1). The fundamental geometric object is the metric tensor, which allows to define uniquely the Levi-Civita connection, which, in turn, determines the Riemann curvature tensor (Sect. 14.5.1, for the description of its properties and symmetries). Then, Lagrangian and field equations of GR are presented in Sect. 14.5.1. Finally, we discuss the tetrad formalism in GR (see Sect. 14.5.1).

Principles of General Relativity

Einstein theory is essentially based on the following pillar ideas, which can be stated as follows [55, 140, 162]:

(1) *Relativity Principle*: there are no preferred inertial frames, i.e. all frames are good for Physics;
(2) *General Covariance Principle:* the basic laws of Physics can be formulated in tensor form in any smooth four-dimensional manifold $\mathcal{M}$. This means that field equations must be "covariant" in form, i.e. they must be invariant under the action of space–time diffeomorphisms;
(3) *Equivalence Principle:* in any smooth four-dimensional manifold $\mathcal{M}$, it is possible to consider a small space–time region $\mathcal{W}$ where spatial and temporal gravitational changes are negligible. Therefore, there always exists a LIF where gravitational effects can be nullified. In other words, inertial effects are locally indistinguishable from gravitational effects (which means the equivalence between the inertial and the gravitational masses).
(4) *Causality Principle:* each point of space–time has to admit a universally valid notion of past, present, and future.

The first two principles are strictly related. They configure the extension of Relativity Principle of Special Relativity to any reference frame independently of the acceleration state. In other words, they figure out a sort of *democracy principle for all observers*, i.e. all observers have the same right to describe the physical reality [162].

Regarding the third principle, it permits to locally recover the Physics of Special Relativity. Geometrically, it translates in determining the tangent plane in every point of a smooth manifold. Furthermore, gravity is the only interaction that cannot be switched off in *absolute*, as instead it occurs for electromagnetic and other fields. Therefore, the gravitational field can be defined as what remains when we have deactivated the other interactions in an absolute way and independently from the observer. It can be only locally nullified in the LIFs, physically coinciding with local free-falling frames. Due to the underlying Riemannian geometric description, LIF is defined by the *Riemann theorem* for every $p \in \mathcal{M}$ in a local chart $(\mathcal{U}, \varphi)$ of p as [162]

$$g_{\mu\nu}(\varphi(p)) = \eta_{\mu\nu}, \qquad \partial_\lambda g_{\mu\nu}(\varphi(p)) = 0. \tag{14.56}$$

This holds if we assume that inertial and gravitational mass coincide (see Refs. [162], for more details). This is the *(weak) equivalence principle* or also known as *university of free fall*, stating that the trajectory of a point mass in a gravitational field depends only on its initial position and velocity and it is independent of its composition and structure. Therefore, the inertial effects may be globally eliminated by an appropriate choice of the reference frame (see Sect. 14.4.2), whereas the gravitational field can be only locally disregarded not eliminated [162].

The fourth principle is needed to ensure the uniqueness of the time notion despite of space–time deformations and singularities. As it is well known, several issues of modern physics are questioning the Causality Principle but we will not go into this discussion again here.

Geodesic Equations

Starting from the universality of free fall postulate in LIF via the coordinates $\{\xi^\mu\}$, a test particle will draw a straight line, whose equation of motion is given by

$$\frac{\mathrm{d}^2\xi^\alpha}{\mathrm{d}s^2} = 0, \tag{14.57}$$

where $\mathrm{d}s^2 = \eta_{\alpha\beta} d\xi^\alpha d\xi^\beta$ is the line element. Since in such a frame it is not possible to experience the existence of gravitational effects, we perform a change of coordinates $\xi^\alpha = \xi^\alpha(x^\mu)$, with $\{x^\mu\}$ the new coordinates. Applying this transformation to Eq. (14.57), we obtain

$$\frac{\mathrm{d}^2x^\lambda}{\mathrm{d}s^2} + \mathring{\Gamma}^\lambda{}_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}s}\frac{\mathrm{d}x^\nu}{\mathrm{d}s} = 0, \tag{14.58}$$

where $\mathring{\Gamma}^\lambda{}_{\mu\nu}$ is the affine connection responsive of the geodesic space–time structure, which arises from the gravitational force acting on the test particle and being responsible for the departure from the straight trend. Its expression is now given by

$$\mathring{\Gamma}^\lambda{}_{\mu\nu} := \frac{\partial x^\lambda}{\partial \xi^\sigma}\frac{\partial^2 \xi^\sigma}{\partial x^\mu \partial x^\nu}, \tag{14.59}$$

which explicitly shows that it is not a tensor. Physically they are the apparent forces acting on the body due to the curved geometric background induced by gravity.

Therefore, assigned the metric tensor $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$, in a generic coordinate system $\{x^\mu\}$, the *geodesic equation* is described by Eq. (14.58). In a metric compatible and torsion-free space–time, we have that the unique affine symmetric connection is the *Levi-Civita one* via the *Levi-Civita theorem* [140, 162]. The condition $\mathring{\nabla}_\lambda g_{\mu\nu} = 0$ gives $\mathring{\Gamma}^\lambda{}_{\mu\nu} \equiv \begin{Bmatrix} \lambda \\ \mu\nu \end{Bmatrix}$ (see Eq. (14.3a)).

The Riemann Curvature Tensor

We have seen the effect of geometric curvature in the geodesic equation, but to quantify it as a field we have to introduce the *Riemann curvature tensor* $\mathring{R}^\alpha{}_{\beta\mu\nu}$ (see Eq. (14.4a) with $\Gamma^\lambda{}_{\mu\nu} = \mathring{\Gamma}^\lambda{}_{\mu\nu}$) arising from the commutation of covariant derivatives on a generic vector v^α,that is

$$[\mathring{\nabla}_\mu, \mathring{\nabla}_\nu]v^\alpha = \mathring{R}^\alpha{}_{\beta\mu\nu} v^\beta. \tag{14.60}$$

The above equation is telling us that the Schwarz theorem, applied to covariant derivatives, does not hold; otherwise, we have a flat space–time (i.e. $\mathring{R}^\alpha{}_{\beta\mu\nu} = 0$). The gravitational field is fully encoded in this tensor.

The Riemann tensor maintains the symmetry (14.5a) in a generic metric-affine theory. However in GR (due to the symmetries of the Levi-Civita connection) it acquires the following further symmetries [140]:

$$\mathring{R}_{\mu\nu\alpha\beta} = -\mathring{R}_{\nu\mu\alpha\beta}, \tag{14.61a}$$

$$\mathring{R}_{\mu\nu\alpha\beta} = \mathring{R}_{\alpha\beta\mu\nu}. \tag{14.61b}$$

The two Bianchi identities (14.6) have both the right members equal to zero, since GR is torsion-free. Due to the symmetries (14.61a), we can define the symmetric Ricci tesor $\mathring{R}_{\alpha\beta} = \mathring{R}^\mu{}_{\alpha\mu\beta}$ and the scalar curvature $\mathring{R} = \mathring{R}^\alpha{}_\alpha$.

Let us consider now a one-parameter family of geodesics $\gamma_s(t)$, where t is the affine parameter along the geodesic, and $s \in [a, b] \subset \mathbb{R}$ labels the curves. We assume that the collection of these curves defines a smooth two-dimensional surface $x^\mu(t, s)$ embedded in $\mathcal{M}$. Provided that this family of geodesics forms a congruence, the parameters t and s are the coordinates on this surface.

A natural vector basis adapted to the coordinate system is given by $\{T^\mu, S^\mu\}$, whose expressions are [162]

$$T^\mu = \frac{\partial x^\mu}{\partial t}, \qquad S^\mu = \frac{\partial x^\mu}{\partial s}. \tag{14.62}$$

Then, we define the relative velocity V^μ and acceleration A^μ along the geodesics as follows:

$$V^\mu = T^\nu \mathring{\nabla}_\nu T^\mu, \tag{14.63a}$$

$$A^\mu = T^\nu \mathring{\nabla}_\nu V^\mu. \tag{14.63b}$$

We then obtain the *geodesic deviation equation* [162]

$$A^\mu = \mathring{R}^\mu{}_{\lambda\alpha\beta} T^\lambda T^\alpha S^\beta, \tag{14.64}$$

where the relative acceleration between two close geodesics is proportional to the Riemann curvature tensor, which characterizes the behaviour of a one-parameter family of neighbouring geodesics.

Lagrangian Formalism and Field Equations

The GR dynamics is derived from the *Hilbert–Einstein action*, whose expression is given by [81]

$$S_{\rm GR} := \frac{c^4}{16\pi G} \int \mathrm{d}^4x \sqrt{-g}\, (\mathscr{L}_{\rm GR} + \mathscr{L}_{\rm m})\,, \tag{14.65}$$

where $\mathscr{L}_{\rm GR} := \mathring{R}(g)$ is the Einstein–Hilbert Lagrangian, coinciding with the Ricci curvature scalar, and $\mathscr{L}_{\rm m}$ is the matter Lagrangian. In this case, the fundamental object is the metric, as underlined in the curvature scalar $\mathring{R}(g)$. The total DoFs are represented by the ten independent components of the metric tensor, from which we must subtract the four-parameter diffeomorphisms underlying the invariance (gauge symmetries' freedom) and other four by a suitable choice of the coordinates (gauge fixing) [140, 197]. Therefore, the gravitational dynamical DoFs becomes two, corresponding thus to the *graviton*, massless spin-2 particle, related to the X and $+$ polarizations of gravitational waves [55].

Applying the principle of least action to Eq. (14.65), we derive the GR field equations in the presence of matter

$$\mathring{G}_{\mu\nu} := \mathring{R}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathring{R} = \frac{8\pi G}{c^4} T_{\mu\nu}, \tag{14.66}$$

where $\mathring{G}_{\mu\nu}$ is the Einstein tensor and

$$T^{\mu\nu} = -\frac{1}{2\sqrt{-g}} \frac{\delta \mathscr{L}_m}{\delta g_{\mu\nu}} \tag{14.67}$$

is the (second-order) energy–momentum tensor which is symmetric with respect to the conservation equations $\mathring{\nabla}_\mu T^{\mu\nu} = 0$, and physically represents the source of gravitational field.

Particular consideration has to be devoted to matter fields and gravity, because some subtleties can arise. For example, (1) ambiguity in the matter coupling; (2)

treatment of bosonic and fermionic fields. In GR, it is clear that a point particle follows the geodesic equations according to the Levi-Civita part of the connection. More problematic issues are linked to bosons (coupling only to the metric) and fermions (coupling with metric and connection). Therefore, when matter fields are taken into account, one must either consider minimally coupled fields or formulate consistent theories in metric-affine formalism. For example in GR, the presence of fermions requires the introduction of tetrads and spin connection.

Tetrad Formalism in General Relativity

GR conceives the gravitational interaction as a change in the geometry of space–time itself, where we pass from the Minkowski $\eta_{\mu\nu}$ to the Riemannian metric $g_{\mu\nu}$, and from partial ∂ to covariant derivatives ∇. The metric plays the role of the fundamental field, which is defined everywhere. In order to study how gravitation couples with other fields, we have to introduce the tetrads to deal with spinors in curved space–times. In addition, tetrads encode the Equivalence Principle since they are locally defined, as gravitation is locally equivalent to an accelerated frame. Therefore, to obtain the effects of gravitation on general sources (particles or fields), we need to: (i) write all the related equations in the Minkowski space–time in general coordinates, represented by trivial tetrads; (ii) replace the holonomic tetrads with the anholonomic tetrads, keeping the same formulae. The resulting equations hold in GR. *Einstein's vierbein theory becomes thus a gauge field theory for gravity.*

Once we assign a general (anhonolomic) tetrad $\left\{e^{A}_{\ \mu}\right\}$, we can rewrite the Riemann tensor according to the Cartan structure equations (see Sect. 14.4.1) as [86]

$$\mathrm{d}e^{C} + \overset{\circ}{\omega}{}^{A}_{\ B} \wedge e^{B} = 0, \tag{14.68a}$$

$$\overset{\circ}{\omega}_{AB} + \overset{\circ}{\omega}_{BA} = \mathrm{d}g_{AB}, \tag{14.68b}$$

$$\mathrm{d}\overset{\circ}{\omega}{}^{A}_{\ B} + \overset{\circ}{\omega}{}^{A}_{\ C} \wedge \overset{\circ}{\omega}{}^{C}_{\ B} = \frac{1}{2}\overset{\circ}{r}{}^{A}_{\ BCD}e^{C} \wedge e^{D}, \tag{14.68c}$$

where $\overset{\circ}{r}{}^{A}_{\ BCD}$ is the Riemann curvature tensor in the tetrad frame, with

$$\overset{\circ}{\omega}{}^{A}_{\ B\mu} := e^{A}_{\ \nu}\overset{\circ}{\nabla}_{\mu}e_{B}^{\ \nu}, \tag{14.69a}$$

$$\overset{\circ}{f}{}^{A}_{\ BC} := \overset{\circ}{\gamma}{}^{A}_{\ BC} - \overset{\circ}{\gamma}{}^{A}_{\ CB}, \tag{14.69b}$$

$$\mathrm{d}g_{AB} = \partial_{C}\, g_{AB}e^{C}, \tag{14.69c}$$

$$\overset{\circ}{\gamma}{}^{A}_{\ BC} = \frac{1}{2}(\overset{\circ}{f}{}^{A}_{\ BC} - g_{CL}g^{AM}\overset{\circ}{f}{}^{L}_{\ BM} - g_{BL}g^{AM}\overset{\circ}{f}{}^{L}_{\ CM}) + \overset{\circ}{\Gamma}{}^{A}_{\ BC}. \tag{14.69d}$$

It is important to note that we can uniquely associate the Lorentz connection to the Levi-Civita connection via Eq. (14.44). In addition, if we consider the natural basis, then we have $\mathring{\omega}^A{}_{BC} = 0$ and therefore $\mathring{\gamma}^A{}_{BC} \equiv \mathring{\Gamma}^A{}_{BC}$. Using the above-cited equations, it is possible to extract the components of $\mathring{r}^A{}_{BCD}$, which are [86]

$$\mathring{r}^A{}_{BCD} = \partial_D \mathring{\gamma}^A{}_{BC} - \partial_C \mathring{\gamma}^A{}_{BD} + \mathring{\gamma}^A{}_{CM} \mathring{\gamma}^M{}_{DB} - \mathring{\gamma}^A{}_{DM} \mathring{\gamma}^M{}_{CB} - \mathring{\gamma}^A{}_{MB} \mathring{\gamma}^M{}_{CD}. \tag{14.70}$$

Also in this case, in the natural basis, we re-obtain the standard definition of the Riemann curvature tensor (14.60).

14.5.2 Gauge Formulation of Gravity: The Case of Teleparallel Gravity

A gauge formulation of gravity is possible in the teleparallel gravity theory. We first show that this general theory can be seen as a translation gauge theory (see Sect. 14.5.2), then we analyse the concepts of geodesics and autoparallel curves in this new framework (see Sect. 14.5.2). We finally concentrate on two important teleparallel subtheories: the metric teleparallel gravity, (in Sect. 14.5.2) and the symmetric teleparallel gravity, (see Sect. 14.5.2). Two important realizations of these approaches are the Teleparallel Equivalent General Relativity (TEGR) and the Symmetric Teleparallel Equivalent General Relativity (STEGR) respectively.

Translation Gauge Theory

In a modern vision of physics, it is very important to settle theories in a gauge framework [26]. In Sect. 14.5.1, we have seen that also GR can be converted in a gauge theory. Let us now sketch how GR can be formulated as a *gauge theory of translations* [9, 26].

This picture of GR can be achieved by both invoking the Nöther theorem and recalling that the source of the gravitational field is given by the energy and momentum. Indeed, provided that gravitational Lagrangian is invariant under space–time translations, the energy–momentum current is covariantly conserved. We will see that a metric teleparallel theory is more suitable to express gravity in this context, because it entails more benefits, and the introduction of tetrads reveals to be more natural.

This approach was first proposed by Lasenby, Doran, and Gull in 1998 [131]. Its geometric setting is the tangent bundle, where the gauge transformations take place. Let us first introduce $\{x^\mu\}$ and $\{x^A\}$ as the coordinates on $\mathcal{M}$ and $T_p\mathcal{M}$, respectively. Now, let us consider the following infinitesimal local translation

$$x^A \longrightarrow \bar{x}^A = x^A + \varepsilon^A(x^\mu), \tag{14.71}$$

where $\varepsilon^A(x^\mu)$ are the infinitesimal parameters of the transformation. The set of translations forms the *translation Lie group* $O(1,3)$, whose generators are

$$P_A := \partial_A. \tag{14.72}$$

They generate the *Abelian translation algebra*, because they satisfy the following trivial commutation rules:

$$[P_A, P_B] \equiv [\partial_A, \partial_B] = 0. \tag{14.73}$$

The infinitesimal transformation, written in terms of the generators, has the following expression:

$$\delta\bar{x}^A = \varepsilon(x^\mu)^B \partial_B x^A = \varepsilon(x^\mu)^A. \tag{14.74}$$

A general source field $\Psi = \Psi(\bar{x}^A(x^\mu))$ transforms under the map (14.71) as follows [9, 129]:

$$\delta_\varepsilon \Psi = \epsilon^A(x^\mu)\partial_A \Psi. \tag{14.75}$$

Let $\varepsilon^A =$ constant be a global translation, then the ordinary derivative $\partial_\mu \Psi$ transforms covariantly, because

$$\partial_\varepsilon(\partial_\mu \Psi) = \varepsilon^A \partial_A(\partial_\mu \Psi). \tag{14.76}$$

For a local translational transformation $\varepsilon^A(x^\mu)$, $\partial_\mu \Psi$ does not transform covariantly, because [9, 129]

$$\partial_\varepsilon(\partial_\mu \Psi) = \underbrace{\varepsilon^A(x^\mu)\partial_A(\partial_\mu \Psi)}_{\text{correct}} + \underbrace{(\partial_\mu \varepsilon^A(x^\mu))\partial_A \Psi}_{\text{spurious}}, \tag{14.77}$$

where the spurious term spoils the translational gauge covariance. However, in order to save this gauge covariance, we follow the praxis exploited in all other gauge theories [134]. Like in the electromagnetic case, where we include the gauge potential field A_μ to guarantee the covariance of the theory, also here we have to set forth the *translational gauge potential 1-form* B_μ, assuming values in the Lie algebra of the translation group, to guarantee the covariance of the gravity theory. Therefore, we introduce the following *gauge covariant derivative* (see Sect. 14.4.1):

$$e'_\mu \Psi \equiv \partial_\mu \Psi = \partial_\mu + B^A_{\ \mu} \partial_A \Psi, \tag{14.78}$$

which holds in the class of Lorentz inertial frames (see Sect. 14.4.1). To recover the gauge covariance, we require that the gauge potential B_μ transforms according to

$$\delta_\varepsilon B^A_{\ \mu} = -\partial_\mu \varepsilon^A(x^\mu). \tag{14.79}$$

Indeed, now $e_\mu \Psi$ transforms covariantly

$$\partial_\varepsilon (e'_\mu \Psi) = \underbrace{\varepsilon^A(x^\mu)\partial_A(\partial_\mu \Psi)}_{\text{correct}}, \tag{14.80}$$

since the potential (14.79) equals the spurious term in Eq. (14.76), cancelling it out. The above construction is based on trivial tetrads. However, for a general non-trivial tetrad field, it has the following expression:

$$e_\mu = \Psi = e^A{}_\mu \partial_A \Psi, \qquad e^A{}_\mu = \partial_\mu x^A + B^A{}_\mu, \tag{14.81}$$

where $B^A{}_\mu \neq -\partial_\mu \varepsilon^A(x^\mu)$ and $e'^A{}_\mu \neq \partial_\mu x^A$. Now let us consider a Lorentz transformation (14.27), and let us assume that the gauge potential $B^A{}_\mu$ transforms as a Lorentz vector in the algebraic index, namely it satisfies

$$B^A{}_\mu \longrightarrow \Lambda^A{}_B(x) B^B{}_\mu. \tag{14.82}$$

Therefore, the generalization of Eq. (14.78) becomes

$$e_\mu \Psi = \partial_\mu + \overset{\bullet}{\omega}{}^A{}_{B\mu} x^B \partial_A \Psi + B^A{}_\mu \partial_A \Psi, \tag{14.83}$$

where

$$e^A{}_\mu = \partial_\mu x^A + \overset{\bullet}{\omega}{}^A{}_{B\mu} x^B + B^A{}_\mu = \overset{\bullet}{\mathcal{D}}_\mu x^A + B^A{}_\mu. \tag{14.84}$$

For general non-trivial tetrads, we need to upgrade the gauge potential (14.79) as follows:

$$\delta_\varepsilon B^A{}_\mu = -\overset{\bullet}{\mathcal{D}}_\mu \varepsilon^A(x^\mu). \tag{14.85}$$

In the context of teleparallel gravity, we have applied the following *translation coupling prescription*

$$e'^A{}_\mu \longrightarrow e^A{}_\mu, \tag{14.86}$$

from which, the *gravitational coupling prescription*, assumed in GR, naturally emerges

$$\eta_{\mu\nu} \longrightarrow g_{\mu\nu}. \tag{14.87}$$

It is important to stress that the local Lorentz invariance is a fundamental symmetry respected by all physical laws in Nature, therefore, we must impose that our new theory be locally Lorentz invariant. Such a requirement requires the additional *Lorentz gravitational coupling prescription*, which is a direct consequence of the strong Equivalence Principle [9, 129]. Indeed, this prescription is based on the General Covariance Principle, which can be seen as an active version of the strong Equivalence Principle, namely given an equation valid in the presence of gravitation, the corresponding special relativistic equation is locally recovered (at a point or along a trajectory), i.e.

$$\partial_\mu \Psi \rightarrow \mathcal{D}'_\mu \Psi = \partial_\mu \Psi + \frac{1}{2} e'^A{}_\mu \left(f'_B{}^C{}_A + f'_A{}^C{}_B - f'^C{}_{BA} \right) S^B_C \Psi. \tag{14.88}$$

where Ψ is a general field, and S^B_C are the generators of the Lorentz group in the same representation to which Ψ belongs. However, in the presence of gravitation, we obtain

$$\partial_\mu \Psi \rightarrow \mathcal{D}_\mu \Psi = \partial_\mu \Psi + \frac{1}{2} e^A{}_\mu \left(f_B{}^C{}_A + f_A{}^C{}_B - f^C{}_{BA} \right) S^B_C \Psi, \tag{14.89}$$

which represents the *full (Lorentz plus translational) gravitational coupling prescription in teleparallel gravity*. We have therefore the following scheme:

$$\underbrace{\begin{Bmatrix} e'^A{}_\mu \longrightarrow e^A{}_\mu \\ \partial_\mu \longrightarrow \mathcal{D}_\mu \end{Bmatrix}}_{\text{grav. coupling prescription in TG}} \quad \Leftrightarrow \quad \underbrace{\eta_{\mu\nu} \longrightarrow g_{\mu\nu}}_{\text{grav. coupling prescription in GR}} \quad . \tag{14.90}$$

Autoparallels and Geodesics

Let us consider the equation of motion of a free test particle first described in the inertial frames $e'^A{}_\mu$, i.e. [129]

$$\frac{\mathrm{d}u'^A}{\mathrm{d}\sigma} = 0, \tag{14.91}$$

where u'^A is the anholonomic 4-velocity of the test particle and $\mathrm{d}\sigma$ is the Minkowskian line element $\mathrm{d}\sigma^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$. We note that Eq. (14.91) is written in a particular class of reference frames, and under a local Lorentz transformation (14.27), it is non-covariant since

$$\frac{\mathrm{d}u'^A}{\mathrm{d}\sigma} = \underbrace{\Lambda^A{}_B(x)\frac{\mathrm{d}u^B}{\mathrm{d}\sigma}}_{\text{correct}} + \underbrace{\frac{\mathrm{d}\Lambda^A{}_B(x)}{\mathrm{d}\sigma} u^B}_{\text{spurious}} \ . \tag{14.92}$$

This is an apparent failure of the covariance, because if we consider the anhonolomic frame $e^A{}_\mu$, associated to $e'^A{}_\mu$ through local Lorentz transformation (cf. Eq.(14.30)), we immediately recover the covariance, because

$$\frac{\mathrm{d}u'^A}{\mathrm{d}\sigma} = 0 \longrightarrow \frac{\mathrm{d}u^B}{\mathrm{d}\sigma} + \overset{\bullet}{\omega}{}^A{}_{B\mu} u^B u^\mu = 0. \tag{14.93}$$

In Sect. 14.5.1, we have defined the geodesic equation (14.58) in GR. This notion must be revised in the parallel framework. Let us consider a chart $(\mathcal{U}, \varphi)$ on the manifold $\mathcal{M}$ and let $\gamma^\mu(\tau)$ be the parametric equation of a curve γ contained in $\mathcal{U}$, where τ is the affine parameter along γ. The tangent vector $\dot{\gamma}$ to γ, in the natural basis $\{\partial_\mu\}$ along γ, is given by the following expression [162]:

$$\dot{\gamma}(\tau) := \frac{\mathrm{d}\gamma^\mu}{\mathrm{d}\tau}\partial_\mu. \tag{14.94}$$

A vector $Y^\mu(\tau)$ is defined to be *parallel transported along* γ if it fulfils the following request:

$$\frac{\mathrm{d}Y^\mu}{\mathrm{d}\tau} := \nabla_\gamma Y^\mu \equiv \frac{\mathrm{d}Y^\mu}{\mathrm{d}\tau} + \Gamma^\mu{}_{\alpha\beta} Y^\alpha \frac{\mathrm{d}\gamma^\beta}{\mathrm{d}\tau} = 0, \tag{14.95}$$

where, for the moment, we do not specify $\Gamma^\mu{}_{\alpha\beta}$. Eq. (14.95) represents a system of first-order differential equations in the unknown $Y^\mu(\tau)$, which admits a unique solution once the initial condition $Y_0^\mu := Y^\mu(\tau_0)$ has been provided. It is important to note that $Y^\mu(\gamma(\tau))$ depends on the curve γ. Therefore, a curve $\gamma(\tau)$ is said to be *autoparallel* if its tangent vector $\dot{\gamma}(\tau)$ satisfies [162]

$$\nabla_\gamma \dot{\gamma} \equiv \frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = 0, \tag{14.96}$$

or, in other words, if it remains parallel to itself along $\gamma(\tau)$, where x^μ are the coordinates of $\gamma(\tau)$ in the chart $(\mathcal{U}, \varphi)$. Eq. (14.96) is a system of second-order differential equations, which admit a unique solution once initial position and velocity have been assigned. It is worth noticing that, in GR, autoparallels and geodesic equations coincide, whereas, in teleparallel gravity, they give rise to two different structures, because the autoparallels are related to the affine connection, whereas the geodesic to the concept of metric, since it measures the minimal lengths between two or more points. In the teleparallel framework, Eq. (14.96) becomes (cf. Eq. (14.2))

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \mathring{\Gamma}^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = -K^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau}, \tag{14.97a}$$

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \mathring{\Gamma}^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = -L^\mu{}_{\alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau}\frac{\mathrm{d}x^\beta}{\mathrm{d}\tau}. \tag{14.97b}$$

Therefore, Eqs. (14.97) recover a new aspect of GR, seen not anymore geometrically as a minimal distance path, but in the gauge paradigm as a sort of *Lorentz force-like interaction* for the contortion tensor and *kinetic energy-like interaction* regarding the disformation tensor, acting on the test particle [18].

Another fundamental implication of autoparallels in teleparallel gravity is that they are sensitive to parameter changes, because it is possible to obtain another curve, although we do not alter the locus of its points. Therefore, if $\gamma(\lambda)$ is autoparallel, then $\mu(\tau) \equiv \gamma(\lambda(\tau))$ might be not autoparallel. This change of parameterization $\lambda = \lambda(\tau)$ entails that Eq. (14.96) becomes [162]

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tau^2} + \Gamma^\mu_{\ \alpha\beta} \frac{\mathrm{d}x^\alpha}{\mathrm{d}\tau} \frac{\mathrm{d}x^\beta}{\mathrm{d}\tau} = -\left(\frac{\mathrm{d}\lambda}{\mathrm{d}\tau}\right)^2 \frac{\mathrm{d}^2\tau}{\mathrm{d}\lambda^2} \frac{\mathrm{d}\gamma^\mu}{\mathrm{d}\tau}. \tag{14.98}$$

We immediately see that the autoparallel character of the curve $\gamma(\lambda)$ is conserved under the parameter change $\lambda = \lambda(\tau)$ if and only if $\tau = a\lambda + b$, with a, b being real arbitrary constants. Here λ, μ are called *canonical parameters*.

Metric Teleparallel Gravity

Metric (or torsional) teleparallel gravity (TG), known also as simply teleparallel gravity, is obtained by assuming the metric compatibility. The theory is geometrically described only by the torsion tensor. In Sect. 14.5.2, we have already seen that tetrads $e^A_{\ \mu}$ and spin connection $\omega^A_{\ B\mu}$ play a fundamental role in describing gravity. Indeed, GR can be recast as a translational gauge theory, where the related gravitational field strength arises from the commutation relation of the covariant derivatives, see Eqs. (14.21) and (14.83), namely[2]

$$[e_\mu, e_\nu] = \hat{T}^A_{\ \nu\mu}\partial_A, \tag{14.99}$$

where the torsion (antisymmetric in the indexes $\mu\nu$)

$$\hat{T}^A_{\ \mu\nu} = \partial_\nu B^A_{\ \mu} - \partial_\mu B^A_{\ \nu} + \dot{\omega}^A_{\ B\nu} B^B_{\ \mu} - \dot{\omega}^A_{\ B\mu} B^B_{\ \nu} \tag{14.100}$$

represents the field strength. Adding the vanishing term

$$\dot{\mathcal{D}}_\mu(\dot{\mathcal{D}}_\nu x^A) - \dot{\mathcal{D}}_\nu(\dot{\mathcal{D}}_\mu x^A) \equiv 0 \tag{14.101}$$

to Eq. (14.100), it becomes

$$\hat{T}^A_{\ \mu\nu} = \partial_\nu e^A_{\ \mu} - \partial_\mu e^A_{\ \nu} + \dot{\omega}^A_{\ B\nu} e^B_{\ \mu} - \dot{\omega}^A_{\ B\mu} e^B_{\ \nu}. \tag{14.102}$$

[2] We define the torsion tensor as minus of that defined in Eq. (14.23), for having the signs in agreement when compared to those of GR.

Exploiting Eqs. (14.43) and (14.102), we have that

$$\hat{T}^{\lambda}{}_{\mu\nu} = e_{A}{}^{\lambda}\hat{T}^{\lambda}{}_{\mu\nu} := \Gamma^{\lambda}{}_{\nu\mu} - \Gamma^{\lambda}{}_{\mu\nu}. \tag{14.103}$$

The spin connection is linked to the inertial effects present in the tetrad frame, it is covariant under both diffeomorphisms and local Lorentz transformations (see Sect. 14.4.2), assuring the same properties also for the torsion tensor. It is important to *associate at each tetrad the related spin connection*; therefore, in TG, we have always to provide the couple $\{e^{A}{}_{\mu}, \overset{\bullet}{\omega}{}^{A}{}_{B\mu}\}$ [129]. There exist frames in TG where the related spin connection vanishes, which are called *proper frames* $\{e^{A}{}_{\mu}, 0\}$. This definition leads to the *Weitzenböck gauge*, which produces the *Weitzenböck connection* $\hat{\Gamma}^{\lambda}{}_{\nu\mu} = e_{A}{}^{\lambda}\partial_{\mu}e^{A}{}_{\nu}$, being the *distant parallelism condition* from where TG takes its name.

A natural question spontaneously arises: *given a tetrad frame, how do we operatively associate the related spin connection?* The simplest solution is tho choose proper frames, but, *a priori*, we do not know which are the related tetrads. Therefore, we have to find a strategy to answer this question. As one can verify, determining them from the field equations is, in general, not a simple task (see Ref. [129], for details). The method we propose relies on first determining the inertial effects in the trivial tetrad frame and then associating the related spin connection (see Ref. [120], for another method). In this approach, let us first introduce the concept of *reference tetrad* $e^{A}_{(\mathrm{r})^{-}}$, in which gravity is switched off, that it

$$e^{A}_{(\mathrm{r})^{-}} := \lim_{G\to 0} e^{A}{}_{\mu}. \tag{14.104}$$

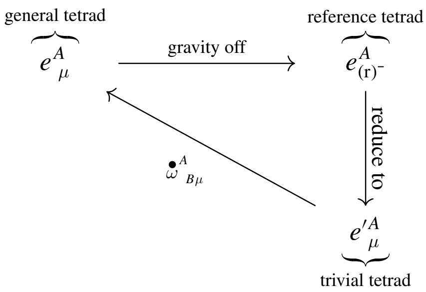
Through this process, we are basically exploiting the Equivalence Principle or the inverse translational coupling prescription (14.86). This has the effect to consider a trivial tetrad, where the anhonomaly coefficients are zero (see Sect. (14.4.1)), and therefore the torsion tensor vanishes. In formulae, this can be written as (cf. Eq. (14.53))

$$\hat{T}^{A}{}_{BC}(e^{A}{}_{\mu}, \overset{\bullet}{\omega}{}^{A}{}_{B\mu}) = \overset{\bullet}{\omega}{}^{A}{}_{BC} - \overset{\bullet}{\omega}{}^{A}{}_{BC} - f^{A}{}_{BC}(e_{(\mathrm{r})}) = 0, \tag{14.105}$$

from which we have

$$\overset{\bullet}{\omega}{}^{A}{}_{BC} = \frac{1}{2}e^{C}_{(\mathrm{r})^{-}}\left[f_{B}{}^{A}{}_{C}(e_{(\mathrm{r})}) + f_{C}{}^{A}{}_{B}(e_{(\mathrm{r})}) - f^{A}{}_{BC}(e_{(\mathrm{r})})\right]. \tag{14.106}$$

Since they differ only by the gravitational content, they represent the gravitational effects inside the tetrad frame. This approach can be schematized as follows:



(14.107)

The coefficients of anhonolonomy (14.12), in the presence of torsion, read as (cf. Eq. (14.23))

$$\overset{\bullet}{\omega}{}^C{}_{AB} - \overset{\bullet}{\omega}{}^C{}_{BA} = f^C{}_{AB} + T^C{}_{AB}. \tag{14.108}$$

This expression can be recombined as follows:

$$\frac{1}{2}\left(f_B{}^C{}_A + f_A{}^C{}_B - f^C{}_{BA}\right) = \overset{\bullet}{\omega}{}^C{}_{BA} - \hat{K}{}^C{}_{BA}, \tag{14.109}$$

where the *contortion tensor*

$$\hat{K}{}^C{}_{BA} = \frac{1}{2}\left(\hat{T}_B{}^C{}_A + \hat{T}_A{}^C{}_B - \hat{T}{}^C{}_{BA}\right), \tag{14.110}$$

has been introduced. Eq. (14.109) is a further development of Eq. (14.89). Using the fundamental identity of the theory of Lorentz connections, we obtain (cf. Eq. (14.69a)) [125, 129][3]

$$\overset{\bullet}{\omega}{}^C{}_{B\mu} - \hat{K}{}^C{}_{B\mu} = \overset{\circ}{\omega}{}^C{}_{B\mu}, \tag{14.111}$$

which joins together GR and TG in a single compact expression. We remark that this combined coupling prescription has been obtained from the General Covariance Principle, and it is thus consistent with the strong Equivalence Principle. In Eq. (14.111), there is $\overset{\circ}{\omega}{}^C{}_{B\mu}$ in GR, enclosing both gravitation and inertial effects in an indistinct form, whereas in TG, $\overset{\bullet}{\omega}{}^C{}_{B\mu}$ describes the inertial effects and $\hat{K}{}^C{}_{B\mu}$ represents only the gravitation. *This is a new and elegant perspective to see the strong Equivalence Principle in TG*. Therefore, in a local frame where the GR spin connection vanishes, we obtain the identity $\overset{\bullet}{\omega}{}^C{}_{B\mu} = \hat{K}{}^C{}_{B\mu}$, where inertial effects compensate gravitation [129], resembling the free-falling cabin' situation.

[3] Equation (14.111) is very important, but its derivation is also not trivial at all. Here, we provide an intuitive proof, although a more rigorous demonstration can be found in Sect. II.6 of Ref. [125]. Let us suppose to have the tetrads $\overset{\circ}{e}{}^A_\mu$ in GR and $\hat{e}{}^A_\mu$ in TG such that they have the same coefficients of anholonomy $\overset{\circ}{f}{}^A_{BC} = \hat{f}{}^A_{BC}$, guaranteed by the fact that there exists an isomorphism assuring this property. This implies $\overset{\circ}{\mathfrak{D}}_\mu = \overset{\wedge}{\mathfrak{D}}_\mu$, which then gives Eq. (14.111).

Another fundamental ingredient of TG theory is represented by the *superpotential*, whose expression is [50]

$$\hat{S}_A^{\ \mu\nu} := \hat{K}^{\mu\nu}{}_A - e_A^{\ \nu}\hat{T}^{\mu} + e_A^{\ \mu}\hat{T}^{\nu}, \tag{14.112}$$

where $\hat{T}^{\alpha\mu}{}_{\alpha} := \hat{T}^{\mu}$ is dubbed *torsion vector*. This permits then to introduce the *torsion scalar*

$$\begin{aligned}\hat{T} &:= \frac{1}{2}\hat{S}_A^{\ \mu\nu}\hat{T}^A_{\ \mu\nu} \\ &= \frac{1}{4}\hat{T}^{\rho}_{\ \mu\nu}\hat{T}_{\rho}^{\ \mu\nu} + \frac{1}{2}\hat{T}^{\rho}_{\ \mu\nu}\hat{T}^{\nu\mu}{}_{\rho} - \hat{T}_{\mu}\hat{T}^{\mu},\end{aligned} \tag{14.113}$$

which is quadratic in all the possible torsion tensor combinations. In particular, in the last equality, the first term resembles that of the usual Lagrangian of internal gauge theories, whereas the other two stem out from the tetrad soldered character allowing thus to set at the same level internal and external indexes [129].

Since TG is curvatureless we have that

$$\hat{R} = \overset{\circ}{R} + \hat{T} + \frac{2}{e}\partial_{\mu}\left(e\hat{T}^{\mu}\right) = 0, \tag{14.114}$$

from which we immediately derive

$$\overset{\circ}{R} = -\hat{T} - \underbrace{\frac{2}{e}\partial_{\mu}\left(eT^{\mu}\right)}_{\text{boundary term}}. \tag{14.115}$$

In Sect. 14.6, the above calculations will be derived in details. Therefore, a particular TG Lagrangian is

$$S_{\text{TEGR}} = -\frac{c^4}{16\pi G}\int d^4x\ e\underbrace{\mathscr{L}_{\text{TEGR}}}_{-\hat{T}} + \int d^4x e\ \mathscr{L}_m, \tag{14.116}$$

up to a boundary term, which gives no contributions, because at infinity the space–time is asymptotically flat and therefore the tetrads reduce to the trivial tetrads and the torsion is null. Eq.(14.114) is dynamically equivalent to that of GR (cf. Eq.(14.65)), namely $S_{\text{TEGR}} = S_{\text{GR}}$. This specific TG theory is called TEGR.

The related field equations are [9]

$$\hat{G}_{\mu\nu} := \frac{1}{e}\partial_{\lambda}(e\ \hat{S}_{\mu\nu}{}^{\lambda}) - \frac{4\pi G}{c^4}{}_{\mu\nu} = \frac{4\pi G}{c^4}T_{\mu\nu}, \tag{14.117}$$

where $\hat{G}_{\mu\nu}$ is the TG Einstein tensor and

$$_{\mu\nu} = \frac{c^4}{4\pi G}\hat{S}_{\lambda\nu}{}^{\rho}\Gamma^{\lambda}{}_{\rho\mu} - g_{\mu\nu}\frac{c^4}{16\pi G}\hat{T} \tag{14.118}$$

is the energy–momentum (pseudo) tensor of the gravitational field. This equation shows that Eq. (14.112) is linked to the gauge representation of the gravitational energy–momentum tensor, namely [9]

$$\hat{S}_A{}^{\mu\nu} = -\frac{8\pi G}{c^4 e}\frac{\partial \mathscr{L}_{\text{TEGR}}}{\partial(\partial_\nu e^A_{\ \mu})}. \tag{14.119}$$

The field equations (14.117) can be also equivalently written in a more explicit form as [50]

$$\begin{aligned}\hat{G}_{\mu\nu} := \frac{1}{e}e^A_{\ \mu}g_{\nu\rho}\partial_\sigma(e\hat{S}_A^{\ \rho\sigma}) - \hat{S}_B^{\ \sigma}{}_{\nu}\hat{T}^B_{\ \sigma\mu} \\ + \frac{1}{2}\hat{T}g_{\mu\nu} - e^A_{\ \mu}\omega^B{}_{A\sigma}\hat{S}_{B\nu}{}^{\sigma} = \frac{8\pi G}{c^4}T_{\mu\nu}.\end{aligned} \tag{14.120}$$

In Sect. 14.6, we will explicitly show that these field equations coincide with those of GR. An important issue is related to the matter couplings, because the presence of torsion introduces some difficulties when dealing with fermions and bosons. Indeed, they are very sensitive to the appearance of distortions in the connections, and the unique resolution of this problem consists in resorting to the Weitzenböck gauge (see Refs. [9, 152], for more details).

Looking at the torsion scalar expression (14.113), we see that it is possible to obtain new theories by considering the following general definition of torsion scalar:

$$\hat{T}_{\text{gen}} := -\frac{c_1}{4}\hat{T}_{\alpha\mu\nu}\hat{T}^{\alpha\mu\nu} - \frac{c_2}{2}\hat{T}_{\alpha\mu\nu}\hat{T}^{\mu\alpha\nu} + c_3\hat{T}_{\alpha}\hat{T}^{\alpha}, \tag{14.121}$$

where c_1, c_2, c_3 are some free real constants, whose explicit values characterize the gravity model known under the name of *three-parameter Hayashi–Shirafuji theory* [114]. The general torsion scalar (14.121) is invariant under both general coordinates and local Lorentz transformations, independently of the numerical values of the coefficients, because it relies only on the properties of the torsion tensor. On the contrary, the equivalence with GR, and then TEGR, is achieved only for $c_1 = c_2 = c_3 = 1$, which is naturally obtained within the TG gauge paradigm, without resorting to hypotheses related to GR [9, 129]. This crucial aspect makes TG a self-consistent theory.

The Nöther energy–momentum pseudotensor $_{\mu}^{\ \rho}$ entails $\partial_{\mu}{}_{\mu}^{\ \rho} = 0$ [134]. In addition, considering the ∂_μ derivative of Eq.(14.117), we obtain $\partial_\mu T^{\mu\nu} = 0$, which shows that the energy–momentum tensor is conserved under ordinary derivative, which implies that the space–time charges $\mathcal{Q}^\mu := \int e\mathrm{d}^3x\,T^{0\mu}$ are conserved. In addition, being the TG field equations symmetric, it is thus very easy to be compared with the GR ones [9, 129]. Therefore, the antisymmetric part of the energy–momentum tensor (14.67) is vanishing, namely

$$\hat{T}_{[\mu\nu]} = e^A{}_{[\mu} g_{\nu]\rho} \hat{T}_A{}^{\rho} = 0. \tag{14.122}$$

Another way to see this identity is through the invariance of the action under local Lorentz transformations [9, 129]. In TEGR, the covariance eliminates six of the sixteen equations, which means that we are able to determine the tetrads up to a local Lorentz transformation, which is equivalent to determine the metric tensor.

The role of spin connection is not dynamical in TEGR and we will show that it trivially satisfies the field equations. The same result is also confirmed by exploiting the constrained variational principle via the Lagrange multipliers (see Ref. [129] and references therein, for details).

Let us consider the following TG Lagrangians

$$\mathscr{L}_{\text{TEGR}}(e^A{}_\mu, 0), \qquad \mathscr{L}_{\text{TEGR}}(e^A{}_\mu, \overset{\bullet}{\omega}{}^A{}_{B\mu}), \tag{14.123}$$

which are both dynamically equivalent to the Hilbert–Einsten action. Therefore, the following identity holds

$$\begin{aligned} &\mathscr{L}_{\text{TEGR}}(e^A{}_\mu, \overset{\bullet}{\omega}{}^A{}_{B\mu}) + \partial_\mu \left[\frac{ec^4}{8\pi G} \hat{T}^\mu (e^A{}_\mu, \overset{\bullet}{\omega}{}^A{}_{B\mu}) \right] \\ &= \mathscr{L}_{\text{TEGR}}(e^A{}_\mu, 0) + \partial_\mu \left[\frac{ec^4}{8\pi G} \hat{T}^\mu (e^A{}_\mu, 0) \right], \end{aligned} \tag{14.124}$$

which explicitly reads as [129]

$$\hat{T}^\mu (e^A{}_\mu, \overset{\bullet}{\omega}{}^A{}_{B\mu}) = \hat{T}^\mu (e^A{}_\mu, 0) - \overset{\bullet}{\omega}{}^\mu. \tag{14.125}$$

Therefore, we arrive to the conclusion that

$$\mathscr{L}_{\text{TEGR}}(e^A{}_\mu, \overset{\bullet}{\omega}{}^A{}_{B\mu}) = \mathscr{L}_{\text{TEGR}}(e^A{}_\mu, 0) + \partial_\mu \left[\frac{ec^4}{8\pi G} \overset{\bullet}{\omega}{}^\mu \right]. \tag{14.126}$$

This proves that the spin connection enters the Lagrangian as a total derivative, justifying also the possibility to reduce the calculations in TEGR by adopting the Weitzenböck gauge in any case. In addition, if we vary the Lagrangian in Eq. (14.126) with respect to the spin connection, we obtain an identically vanishing equation. Therefore, the spin connection does not contribute to the TG field equations, representing non-dynamical DoFs. This fact shows also that TEGR can be considered as a *pure tetrad teleparallel gravity* [136], which assumes that for whatever tetrad one chooses, the spin connection is zero, treating these two objects as independent structures (see Ref. [129], for details and for its implications).

We have understood that the spin connection is not relevant, if we are interested in searching for the solutions of TEGR field equations. However, formally, its presence fulfils a paramount role, because: (i) it guarantees the covariance of the action under local Lorentz transformations and diffeomorphisms; (ii) it is endowed with a regularizing power, because it removes the divergent inertial effects from the Lagrangian, dubbed thus *renormalized action* (see Ref. [129], for details); (iii) it permits to obtain a regular field theory and naturally produces, in its action, a Gibbons–Hawking–York term, which permits to be coherently related to the formulation of a quantum gravity theory (see Refs. [7, 8, 155], for details).

Finally, analysing the DoFs of TEGR, we start from the vierbein $e^A_{\ \mu}$ with 16 components. We have to subtract 6 DoFs related to the inertial effects due to the spin connection and other 8 non-dynamical DoFs due to diffeomorphisms (the same as in GR). The result is 2 DoFs as in the case of GR. Also for this feature, TEGR is dynamically equivalent to GR.

Symmetric Teleparallel Gravity

Symmetric teleparallel gravity (STG) is a formulation of gravitational interaction described only in terms of non-metricity (14.4c). While TG theories have been extensively discussed, STG patterns have only recently received a growing attention, and there are still some crucial points to be disclosed and better understood. This theory can be either formulated in terms of metric tensors or tetrads, although the former is the most common presentation followed in the literature [145]. In STG, the symmetric affine connection (cf. Eq. (14.2)) assumes a fundamental dynamical role and represents an independent structure. This hypothesis is not trivial at all, because it requires considerable efforts in determining all the affine components already in the simplest cases both at astrophysical and cosmological levels (see e.g. Refs. [54, 83, 121], for more details).

The presence of non-metricity entails particular geometric effects, which gives rise to counterintuitive implications from those analysed in the previous theories. They can be summarized in the following points:

- raising up or lowering down indexes of vectors or tensors under the covariant derivative $\overset{\diamond}{\nabla}$, is not straightforward like in metric case, namely given a vector v^μ, we have
$$g_{\nu\lambda}\overset{\diamond}{\nabla}_\mu v^\lambda = \overset{\diamond}{\nabla}_\mu v_\nu - v^\lambda \overset{\diamond}{Q}_{\mu\nu\lambda}; \tag{14.127}$$

- non-metricity does not preserve the length of vectors; indeed given two vectors $v = v^\mu\partial_\mu$ and $w = w^\mu\partial_\mu$ parallel along a curve γ, their tangent vectors are $T = T^\mu\partial_\mu$ with $T^\mu \equiv \dot{\gamma}^\mu$, namely $T^\lambda\overset{\diamond}{\nabla}_\lambda v^\mu = 0$ and $T^\lambda\overset{\diamond}{\nabla}_\lambda w^\mu = 0$. Let us calculate the evolution of the scalar product of the vectors
$$T^\lambda\overset{\diamond}{\nabla}_\lambda v\cdot w = T^\lambda v^\mu w^\nu \overset{\diamond}{Q}_{\lambda\mu\nu}, \tag{14.128}$$
where $v\cdot w := g_{\mu\nu}v^\mu w^\nu$, which is not conserved, as well as the norm of a vector $|v| := \sqrt{v\cdot v}$, and therefore it is not possible to normalize it. It follows also that the angles, between two vectors, do not in general conserve, namely

$$T^{\lambda}\overset{\diamond}{\nabla}_{\lambda}\left(\frac{v\cdot w}{|v||w|}\right)\neq 0. \tag{14.129}$$

STG is, in general, not a *conformal theory*, but it is possible to reduce it to a conformal one (see Ref. [104], for details). The above results imply also the impossibility to define a proper time along a curve as in GR;

- Given a 4-velocity u^{μ}, we define

$$a^{\mu} := u^{\lambda}\overset{\diamond}{\nabla}_{\lambda}u^{\mu}, \tag{14.130a}$$

$$\tilde{a}_{\mu} := u^{\lambda}\overset{\diamond}{\nabla}_{\lambda}u_{\mu} = a_{\mu} + \overset{\diamond}{Q}_{\lambda\nu\mu}u^{\lambda}u^{\nu}, \tag{14.130b}$$

where a^{μ} is the *acceleration*, whereas $\tilde{a}_{\mu}$ is the *anomalous acceleration*. In particular, this implies that the 4-velocity is not anymore orthogonal to the 4-acceleration, because

$$\begin{aligned} u_{\mu}a^{\mu} &= u_{\mu}u^{\lambda}\overset{\diamond}{\nabla}_{\lambda}u^{\mu} \\ &= u^{\lambda}\overset{\diamond}{\nabla}_{\lambda}(u_{\mu}u^{\mu}) - u^{\mu}u^{\lambda}\overset{\diamond}{\nabla}_{\lambda}u_{\mu} \\ &= \overset{\diamond}{Q}_{\lambda\mu\nu}u^{\lambda}u^{\mu}u^{\nu} + 2u_{\mu}a^{\mu} - \tilde{a}_{\mu}u^{\mu}, \end{aligned} \tag{14.131}$$

from which we obtain

$$a^{\mu}u_{\mu} = \tilde{a}_{\mu}u^{\mu} - \overset{\diamond}{Q}_{\lambda\mu\nu}u^{\lambda}u^{\mu}u^{\nu}. \tag{14.132}$$

From Eq. (14.130b), we get

$$\left(\tilde{a}_{\mu} - a_{\mu}\right)u^{\mu} = \overset{\diamond}{Q}_{\lambda\mu\nu}u^{\lambda}u^{\mu}u^{\nu}. \tag{14.133}$$

Therefore, the non-metricity tensor expresses how much the anomalous acceleration deviates from the standard acceleration, and it is also responsible to depart the acceleration from the spatial hypersurface orthogonal to the 4-velocity;

- the acceleration of autoparallels (cf. Eq. (14.95)) in STG becomes

$$a^{\mu} = 0, \qquad \tilde{a}_{\mu} = \overset{\diamond}{Q}_{\lambda\nu\mu}u^{\lambda}u^{\nu}; \tag{14.134}$$

- in order to recover the length conservation (cf. Eq. (14.128)) and the autoparallel definition (cf. Eq. (14.134)), we have to impose

$$\overset{\diamond}{Q}_{(\lambda\mu\nu)} = 0, \qquad \overset{\diamond}{Q}_{(\lambda\mu)\nu} = 0, \tag{14.135}$$

but these two conditions are too strict constraints. These issues can be solved by resorting to the *Weyl conformal transformations* (see Ref. [194], for details).

Let us consider now the STG action, constituted by the most general quadratic Lagrangian [83]:

$$S_{\text{STEGR}} := \int \mathrm{d}^4x \, \sqrt{-g} \left[\frac{c^4}{16\pi G} \underbrace{\mathscr{L}_{\text{STEGR}}}_{\overset{\diamond}{Q}} + \mathscr{L}_m \right], \tag{14.136}$$

where $\overset{\diamond}{Q}$ is the so-called *non-metricity scalar*, whose expression is given by [50, 83]

$$\begin{aligned} \overset{\diamond}{Q} &:= g^{\mu\nu} \left(\overset{\diamond}{L}{}^{\alpha}{}_{\beta\mu} \overset{\diamond}{L}{}^{\beta}{}_{\nu\alpha} - \overset{\diamond}{L}{}^{\alpha}{}_{\beta\alpha} \overset{\diamond}{L}{}^{\beta}{}_{\mu\nu} \right) \\ &= \frac{1}{4} \left(\overset{\diamond}{Q}_{\alpha} \overset{\diamond}{Q}{}^{\alpha} - \overset{\diamond}{Q}_{\alpha\beta\gamma} \overset{\diamond}{Q}{}^{\alpha\beta\gamma} \right) \\ &+ \frac{1}{2} \left(\overset{\diamond}{Q}_{\alpha\beta\gamma} \overset{\diamond}{Q}{}^{\beta\alpha\gamma} - \overset{\diamond}{Q}_{\alpha} \overset{\diamond}{\bar{Q}}{}^{\alpha} \right), \end{aligned} \tag{14.137}$$

where $\overset{\diamond}{Q}_{\alpha} := \overset{\diamond}{Q}_{\alpha\lambda}{}^{\lambda}$ and $\overset{\diamond}{\bar{Q}}_{\alpha} := \overset{\diamond}{Q}{}^{\lambda}{}_{\lambda\alpha}$ represent two independent and non-vanishing traces of the non-metricity tensor. This gives rise to the STEGR theory, where it is possible to show the validity of the following formula (see Sect. 14.6.3) [83]:

$$\overset{\diamond}{Q} = \mathring{R} + \mathring{\nabla}_{\mu} (\overset{\diamond}{Q}{}^{\mu} - \overset{\diamond}{\bar{Q}}{}^{\mu}), \tag{14.138}$$

and using the following GR identity [140]

$$\mathring{\nabla}_{\mu} (\overset{\diamond}{Q}{}^{\mu} - \overset{\diamond}{\bar{Q}}{}^{\mu}) \equiv \frac{1}{\sqrt{-g}} \partial_{\mu} \left[\sqrt{-g} (\overset{\diamond}{Q}{}^{\mu} - \overset{\diamond}{\bar{Q}}{}^{\mu}) \right], \tag{14.139}$$

we see that the STEGR action is dynamically equivalent to GR up to a boundary term, which is vanishing because at infinity the metric is flat. Since the STEGR Lagrangian is quadratic in terms of the non-metricity tensor, the most general STG Lagrangian is [22]

$$\begin{aligned} \overset{\diamond}{Q}_{\text{gen}} &:= c_1 \overset{\diamond}{Q}_{\alpha\beta\gamma} \overset{\diamond}{Q}{}^{\alpha\beta\gamma} + c_2 \overset{\diamond}{Q}_{\alpha\beta\gamma} \overset{\diamond}{Q}{}^{\beta\alpha\gamma} + c_3 \overset{\diamond}{Q}_{\alpha} \overset{\diamond}{Q}{}^{\alpha} \\ &+ c_4 \overset{\diamond}{\bar{Q}}_{\alpha} \overset{\diamond}{\bar{Q}}{}^{\alpha} + c_5 \overset{\diamond}{Q}_{\alpha} \overset{\diamond}{\bar{Q}}{}^{\alpha}. \end{aligned} \tag{14.140}$$

where $c_1, \; c_2, \; c_3, \; c_4, \; c_5$ are real free constant parameters, and this gives rise to the *five-parameter family of quadratic theories* or the so-called *New GR* (see Ref. [27] and references therein).

We can introduce a *superpotential* or the *non-metricity conjugate* as [50, 83, 104]

$$\begin{aligned}\overset{\diamond}{P}{}^{\alpha}_{\mu\nu} &:= \frac{1}{2\sqrt{-g}}\frac{\partial(\sqrt{-g}\overset{\diamond}{Q})}{\partial \overset{\diamond}{Q}{}_{\alpha}^{\mu\nu}} \\ &= \frac{1}{4}\overset{\diamond}{Q}{}^{\alpha}{}_{\mu\nu} - \frac{1}{4}\overset{\diamond}{Q}{}_{(\mu}{}^{\alpha}{}_{\nu)} - \frac{1}{4}g_{\mu\nu}\overset{\diamond}{Q}{}^{\alpha\beta}{}_{\beta} \\ &+ \frac{1}{4}\left[\overset{\diamond}{Q}{}_{\beta}{}^{\beta\alpha}g_{\mu\nu} + \frac{1}{2}\delta^{\alpha}_{(\mu}\overset{\diamond}{Q}{}_{\nu)}{}^{\beta}{}_{\beta}\right].\end{aligned} \tag{14.141}$$

Through this definition, we can describe the non-metricity scalar (14.137) equivalently as

$$\overset{\diamond}{Q} := \overset{\diamond}{Q}{}_{\alpha\mu\nu}\overset{\diamond}{P}{}^{\alpha\mu\nu}. \tag{14.142}$$

We can introduce also the further quantity [50, 104]

$$\begin{aligned}\frac{1}{\sqrt{-g}}\overset{\diamond}{q}{}_{\mu\nu} &:= \frac{1}{\sqrt{-g}}\frac{\partial(\sqrt{-g}\overset{\diamond}{Q})}{\partial g^{\mu\nu}} - \frac{1}{2}\overset{\diamond}{Q}g_{\mu\nu} \\ &= \frac{1}{4}\left(2\overset{\diamond}{Q}{}_{\alpha\beta\mu}\overset{\diamond}{Q}{}^{\alpha\beta}{}_{\nu} - \overset{\diamond}{Q}{}_{\mu\alpha\beta}\overset{\diamond}{Q}{}_{\nu}{}^{\alpha\beta}\right) \\ &- \frac{1}{4}\left(2\overset{\diamond}{Q}{}_{\alpha}{}^{\beta}{}_{\beta}\overset{\diamond}{Q}{}^{\alpha}{}_{\mu\nu} - \overset{\diamond}{Q}{}_{\mu}{}^{\beta}{}_{\beta}\overset{\diamond}{Q}{}_{\nu}{}^{\beta}{}_{\beta}\right) \\ &- \frac{1}{2}\left(\overset{\diamond}{Q}{}_{\alpha\beta\mu}\overset{\diamond}{Q}{}^{\beta\alpha}{}_{\nu} - \overset{\diamond}{Q}{}_{\beta}{}^{\beta}{}_{\alpha}\overset{\diamond}{Q}{}^{\alpha}{}_{\mu\nu}\right).\end{aligned} \tag{14.143}$$

We have now all the elements to write the STEGR field equations (obtained varying the STEGR action with respect to the metric tensor), which reads as [50, 83]

$$\begin{aligned}\overset{\diamond}{G}{}_{\mu\nu} &:= -2\nabla_{\alpha}\left(\sqrt{-g}\overset{\diamond}{P}{}^{\alpha}_{\mu\nu}\right) \\ &+ \overset{\diamond}{q}{}_{\mu\nu} - \frac{\sqrt{-g}\overset{\diamond}{Q}}{2}g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu},\end{aligned} \tag{14.144}$$

where $\overset{\diamond}{G}{}_{\mu\nu}$ is the STEGR Einstein tensor. The variation of STEGR action with respect to the connection produces the *connection field equations* [50, 83]

$$\nabla_{\mu}\nabla_{\nu}\left(\sqrt{-g}\overset{\diamond}{P}{}_{\alpha}{}^{\mu\nu}\right) = 0, \tag{14.145}$$

representing a set of first-order differential equations for the affine connection.

Using the general results of Sect. 14.4.2, it is possible to recast the STEGR connection via the tetrads $e^{\alpha}{}_{\beta} \in \mathrm{GL}(4, \mathbb{R})$ and the curvatureless hypothesis in the following form (cf. Eq. (14.43))[4]

[4] In Eq. (14.146), we have used a different notation with respect to those employed previously. Here, it is important to underline the inverse tetrad matrix for the implication in (14.149).

$$\Gamma^{\alpha}_{\ \mu\nu} := (e^{-1})^{\alpha}_{\ \beta}\partial_{\mu}e^{\beta}_{\ \nu}. \tag{14.146}$$

Since STEGR is torsionless, we have (cf. Eq. (14.102))

$$T^{\alpha}_{\ \mu\nu} := (e^{-1})^{\alpha}_{\ \beta}\partial_{[\nu}e^{\beta}_{\ \mu]} = 0, \tag{14.147}$$

which implies

$$\partial_{\mu}e^{\beta}_{\ \nu} = \partial_{\nu}e^{\beta}_{\ \mu} \quad \Leftrightarrow \quad e^{\alpha}_{\ \beta} \equiv e'^{\alpha}_{\ \beta} := \partial_{\beta}\xi^{\alpha}, \tag{14.148}$$

where the tetrad is holonomic and, in addition, it can be parameterized by $\xi^{\alpha} = \xi^{\alpha}(x^{\mu})$. Therefore, Eq. (14.146) becomes [50, 83]

$$\Gamma^{\alpha}_{\ \mu\nu} = \frac{\partial x^{\alpha}}{\partial \xi^{\lambda}}\partial_{\mu}\partial_{\nu}\xi^{\lambda}. \tag{14.149}$$

This connection can be set globally to zero, by considering the following affine (gauge) roto-translational transformation of coordinates [83]

$$\xi^{\alpha} := M^{\alpha}_{\ \beta}x^{\beta} + \xi^{\alpha}_{0}, \tag{14.150}$$

where $M^{\alpha}_{\ \beta} \in O(1, 3)$ is an orthogonal matrix and ξ^{α}_{0} is a constant translational vector, which permits to have $\Gamma^{\alpha}_{\ \mu\nu} = 0$, which is the so-called *coincident gauge*. Physically, this means that the origin of the tangent space (expressed by ξ^{α}) is coincident with the space–time origin (given by x^{μ}). This gauge is defined up to a linear affine transformation $ax^{\mu} + b$ with a, b real constant values.

It is important to note that this residual global symmetry does not vanish at infinity, ensuing significant properties at the infrared structure of the theory. In addition, recalling that the strong Equivalence Principle states that gravitation is indistinguishable from acceleration, its effects can be locally neglected via a diffeomorphic change of coordinates (i.e. LIFs). From this perspective, we understand that the affine connection is an integrable translation. Therefore, the coincident gauge embodies and saves the strong Equivalence Principle of GR [126].

It is worth noticing that the STEGR affine connection is purely inertial and it does not contain any information about gravitation. Another important implication of the coincident gauge is the explicit breaking of diffeomorphism invariance due to the particular choice of coordinates, which does not occur in other frames [121]. The use or not of the coincident gauge affects only the boundary term (14.138), which has no influence on the ensuing dynamics and therefore neither on the evolution of the metric tensor.

This particular gauge form permits to considerably simplify the calculations. In addition, the affine field equations (14.145) are trivially satisfied. In TG, the local Lorentz transformations are gauged through the spin connection and the calculations are simplified via the Weitzenböck choice, whereas, in STG, the diffeomorphism of coordinates become the new gauge and the calculations are easily carried out through the coincident gauge. This concept is summarized in the following scheme

$$\begin{array}{lcl} & \overset{\text{TEGR}}{\nearrow} \ \text{loc. Lorentz trans.} \longrightarrow \text{Weitzenböck} & \\ \text{gauge} & & \\ & \underset{\text{STEGR}}{\searrow} \ \text{diff. of coord.} \longrightarrow \text{coincident} & \end{array} \tag{14.151}$$

There are also other two beneficial effects considering Eq. (14.2). It is

$$\overset{\diamond}{\nabla}_\mu = \partial_\mu, \qquad \overset{\diamond}{L}{}^{\lambda}{}_{\mu\nu} = -\overset{\circ}{\Gamma}{}^{\lambda}{}_{\mu\nu}. \tag{14.152}$$

It is worth noticing that, in generic STG theories, it is not possible to require that the coincident gauge holds a priori. More specifically, it is not possible, in general, to use a coordinate system which simultaneously simplifies metric and connection. When this can be achieved, it holds for a restricted set of geometries or it reduces the class of solutions (see Refs. [83], for further details).

In STEGR, we have that the total DoFs are encoded in the metric tensor, having 10 components from which we have to subtract 8 diffeomorphisms as in GR, having therefore again 2 DoFs as in GR. Here we have that the four diffeomorphisms of coordinates become the gauge diffeomorphism symmetries. While, in TEGR, metric and connection are related and, in STEGR, the connection becomes essentially a pure gauge and all the dynamics are enclosed in the metric, which is trivially connected. It is possible to introduce a close analogy between the fields ξ^α, parameterizing the connection, and the *Stückelberg fields*, related to the invariance of coordinates' transformation, and also between the coincident gauge and the *unitary gauge* (see Ref. [165] for details).

14.5.3 A Discussion on Trinity Gravity at Lagrangian Level

GR, TEGR, and STEGR constitute the so-called Geometric Trinity of Gravity, but, from the above discussion, it is clear that they are nothing else but particular cases of wide classes of theories. According to their formulation, they are three non-communicating theories, because they start from different hypotheses and different dynamical-geometric objects. GR is usually conceived as the geometric formulation of gravity, whereas TEGR and STEGR as the gauge approaches to gravity, albeit also GR can be formulated in a gauge way via the use of tetrads and spin connection.

Covariance and (strong) Equivalence Principles are at the foundations of GR. The former postulate has a more general character, which can be easily recognized also in TEGR and STEGR, whereas the latter hides some subtleties, which are sources of confusion in literature. For example, some papers state that such a principle does not hold in TG. More properly, it is not strictly required at the foundation of TG but it must hold to provide mathematical and physical coherence for TEGR and STEGR theories. Basically, Equivalence Principle guarantees the *equivalence* among GR, TEGR, and STEGR. This last point can be summarized as follows:

$$\text{strong EP} \begin{cases} \xrightarrow{\text{GR}} \text{foundation} \\ \xrightarrow{\text{TEGR}} \overset{\bullet}{\omega}{}^{C}{}_{B\mu} - \hat{K}^{C}{}_{B\mu} = \overset{\circ}{\omega}{}^{C}{}_{B\mu} \\ \xrightarrow{\text{STEGR}} \text{coincident gauge} \end{cases} \tag{14.153}$$

Up to now, we underlined only the equivalence among the three theories at Lagrangian level, pointing out the difference for a boundary term, namely

$$\begin{array}{ccc} & \overbrace{\mathscr{L}_{\text{GR}}}^{\overset{\circ}{R}} & \\ \swarrow\!\!\nearrow & & \nwarrow\!\!\searrow \\ \underbrace{\mathscr{L}_{\text{TEGR}}}_{-\hat{T}-\frac{2}{e}\partial_\mu\left(e\hat{T}^\mu\right)} & \longleftrightarrow & \underbrace{\mathscr{L}_{\text{STEGR}}}_{\overset{\diamond}{Q}-\overset{\circ}{\nabla}_\mu(\overset{\diamond}{Q}{}^\mu-\overset{\diamond}{\bar{Q}}{}^\mu)} \end{array} \tag{14.154}$$

As stated above, the equivalence does not for extensions like $f(R)$, $f(T)$, and $f(Q)$ because, in general, these extended theories differ for the DoFs (see [20] for a straightforward example). However, equivalence can be restored also in extensions considering appropriate boundary terms. For example, $f(R)$, $f(T, B)$, and $f(Q, B)$ can be compared as fourth-order theories when an appropriate boundary term B is defined in each gravity framework. In $f(T, B)$, this equivalence has been explicitly proved considering the boundary term as in Eq.(14.115) (see e.g. [16, 17, 47]). An analogue procedure shows that also $f(Q, B)$ theory can be dynamically reduced to $f(R)$ defining a suitable boundary term as in Eq.(14.138).

14.6 Field Equations in Trinity Gravity

In the above discussion, equivalent representations of gravity have been compared at the level of actions and Lagrangians. Here we want to develop the same comparison at the level of field equations.

Let us start with the Bianchi identities, having the pivotal role to link the field equations of a theory with the conservation laws of the gravity tensor invariants and with the energy–momentum tensor [140]. We start from the second Bianchi identity (14.6), whose more explicit expression is [18]

$$\begin{aligned}&\nabla_\lambda R^\alpha{}_{\beta\mu\nu} + \nabla_\mu R^\alpha{}_{\beta\nu\lambda} + \nabla_\nu R^\alpha{}_{\beta\lambda\mu} \\ &= T^\rho{}_{\mu\lambda} R^\alpha{}_{\beta\nu\rho} + T^\rho{}_{\nu\lambda} R^\alpha{}_{\beta\mu\rho} + T^\rho{}_{\nu\mu} R^\alpha{}_{\beta\lambda\rho},\end{aligned} \tag{14.155}$$

and we prove the equivalence among GR (see Sect. 14.6.1), TEGR (see Sect. 14.6.2), and STEGR (see Sect. 14.6.3) in terms of their field equations, which we show to be equal to those already presented in Sect. 14.5.

14.6.1 GR Field Equations

Since in GR we have $R^\alpha{}_{\beta\mu\nu} = \mathring{R}^\alpha{}_{\beta\mu\nu}$ and $T^\alpha{}_{\beta\gamma} = Q_{\alpha\beta\gamma} = 0$, the second Bianchi identity (14.155) reduces to

$$\mathring{\nabla}_\lambda \mathring{R}^\alpha{}_{\beta\mu\nu} + \mathring{\nabla}_\mu \mathring{R}^\alpha{}_{\beta\nu\lambda} + \mathring{\nabla}_\nu \mathring{R}^\alpha{}_{\beta\lambda\mu} = 0. \tag{14.156}$$

To simplify the calculations, thanks to the Covariance Principle, we can exploit the LIF's coordinates (cf. Eq. (14.56)), where second derivatives of the metric are not null. Contracting α and λ, Eq. (14.156) becomes

$$\partial_\lambda \mathring{R}^\lambda{}_{\beta\mu\nu} + \partial_\mu \mathring{R}^\lambda{}_{\beta\nu\lambda} + \partial_\nu \mathring{R}^\lambda{}_{\beta\lambda\mu} = 0. \tag{14.157}$$

Using the antisymmetry in the last two indexes of the Riemann tensor (cf. Eq. (14.5a)), we obtain

$$\partial_\lambda \mathring{R}^\lambda{}_{\beta\mu\nu} - \partial_\mu \mathring{R}^\lambda{}_{\beta\lambda\nu} + \partial_\nu \mathring{R}^\lambda{}_{\beta\lambda\mu} = 0. \tag{14.158}$$

Applying the metric to first raise up the index β and then contracting β and μ, we have

$$-\partial_\lambda \mathring{R}^\lambda{}_\nu - \partial_\beta \mathring{R}^\beta{}_\nu + \partial_\nu \mathring{R} = 0, \tag{14.159}$$

from which we immediately obtain

$$\partial_\mu \mathring{R}^\mu{}_\nu - \frac{1}{2}\partial_\nu \mathring{R} = 0. \tag{14.160}$$

Using again the metric tensor, Eq. (14.160) becomes

$$\partial_\mu(\mathring{R}^{\mu\nu} - \frac{1}{2}g^{\mu\nu}\hat{R}) = 0 \;\Rightarrow\; \mathring{\nabla}_\mu(\mathring{R}^{\mu\nu} - \frac{1}{2}g^{\mu\nu}\hat{R}) = 0, \tag{14.161}$$

where the partial derivative is in general replaced by the covariant one. This relation leads to the Einstein field equations in vacuum (cf. Eq. (14.66)). The Einstein tensor $\mathring{G}^{\mu\nu}$ is divergenceless: this fact also implies the conservation of the energy–momentum tensor [140], namely

$$\mathring{\nabla}_\mu \mathring{G}^{\mu\nu} = 0, \quad \Leftrightarrow \quad \mathring{\nabla}_\mu T^{\mu\nu} = 0. \tag{14.162}$$

14.6.2 TEGR Field Equations

Since in TEGR curvature and non-metricity vanish, Eq. (14.155) can be further simplified via the Weitzenböck gauge (see Sect. 14.5.2) as follows

$$\hat{\nabla}_\lambda R^\alpha{}_{\beta\mu\nu} + \hat{\nabla}_\nu R^\alpha{}_{\beta\lambda\mu} + \hat{\nabla}_\mu R^\alpha{}_{\beta\nu\lambda} = 0, \tag{14.163}$$

where $R^\alpha{}_{\beta\mu\nu} \equiv \mathring{R}^\alpha{}_{\beta\mu\nu} + \hat{\mathcal{K}}^\alpha{}_{\beta\mu\nu} = 0$ with

$$\begin{aligned}\hat{\mathcal{K}}^\alpha{}_{\beta\mu\nu} := {} & \mathring{\nabla}_\mu \hat{K}^\alpha{}_{\beta\nu} - \mathring{\nabla}_\nu \hat{K}^\alpha{}_{\beta\mu} \\ & + \hat{K}^\alpha{}_{\sigma\mu} \hat{K}^\sigma{}_{\beta\nu} - \hat{K}^\alpha{}_{\sigma\nu} \hat{K}^\sigma{}_{\beta\mu},\end{aligned} \tag{14.164}$$

including all torsion tensor contributions and having also the following symmetry properties (cf. Eq. (14.110)):

$$\hat{\mathcal{K}}^\alpha{}_{\beta\mu\nu} = -\hat{\mathcal{K}}_\beta{}^\alpha{}_{\mu\nu}, \qquad \hat{\mathcal{K}}^\alpha{}_{\beta\mu\nu} = -\hat{\mathcal{K}}^\alpha{}_{\beta\nu\mu}. \tag{14.165}$$

Contracting α and λ, Eq. (14.163) becomes

$$\begin{aligned}& \hat{\nabla}_\lambda \mathring{R}^\lambda{}_{\beta\mu\nu} + \hat{\nabla}_\mu \mathring{R}^\lambda{}_{\beta\nu\lambda} + \hat{\nabla}_\nu \mathring{R}^\lambda{}_{\beta\lambda\mu} \\ & + \hat{\nabla}_\lambda \hat{\mathcal{K}}^\lambda{}_{\beta\mu\nu} + \hat{\nabla}_\mu \hat{\mathcal{K}}^\lambda{}_{\beta\nu\lambda} + \hat{\nabla}_\nu \hat{\mathcal{K}}^\lambda{}_{\beta\lambda\mu} = 0.\end{aligned} \tag{14.166}$$

Applying the same strategy of GR (see Sect. 14.6.1) and using the metric compatibility of TEGR, we obtain

$$\hat{\nabla}_\mu (\mathring{R}^\mu{}_\nu + \hat{\mathcal{K}}^\mu{}_\nu) - \frac{1}{2} \hat{\nabla}_\nu (\mathring{R} + \hat{\mathcal{K}}) = 0, \tag{14.167}$$

where $\hat{\mathcal{K}}_{\mu\nu} := \hat{\mathcal{K}}^\lambda{}_{\mu\lambda\nu}$ and $\hat{\mathcal{K}} := \hat{\mathcal{K}}^\nu{}_\nu$, having a formally similar definition of Ricci tensor and scalar curvature of GR. Equation (14.167) entails twofold implications

$$\mathring{R}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \mathring{R} = -\hat{\mathcal{K}}_{\mu\nu} + \frac{1}{2} g_{\mu\nu} \hat{\mathcal{K}}, \tag{14.168a}$$

$$\hat{\mathcal{K}}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \hat{\mathcal{K}} = 0, \tag{14.168b}$$

where the former tells that TEGR field equations are equivalent to those of GR, whereas the latter, derived using the GR vacuum field equations, gives the TEGR field equations, which are divergenceless in terms of $\hat{\nabla}$.

Now, we prove that Eq. (14.168b) reproduces exactly Eq. (14.120). To this end, we first analyse $\hat{\mathcal{K}}_{\mu\nu}$, which gives

$$\begin{aligned}\hat{\mathcal{K}}_{\mu\nu} &= \mathring{\nabla}_\alpha \hat{K}^\alpha{}_{\mu\nu} - \mathring{\nabla}_\nu \hat{K}^\alpha{}_{\mu\alpha} + \hat{K}^\sigma{}_{\mu\nu}\hat{K}^\alpha{}_{\sigma\alpha} - \hat{K}^\sigma{}_{\mu\alpha}\hat{K}^\alpha{}_{\sigma\nu} \\ &= \mathring{\nabla}_\alpha \hat{K}^\alpha{}_{\mu\nu} + \mathring{\nabla}_\nu \hat{T}_\mu - \hat{K}_{\sigma\mu\nu}\hat{T}^\sigma - \hat{K}^\sigma{}_{\mu\alpha}\hat{K}^\alpha{}_{\sigma\nu} \\ &= \mathring{\nabla}_\alpha \hat{S}_\nu{}^\alpha{}_\mu + \mathring{\nabla}_\alpha \hat{T}^\alpha g_{\mu\nu} - \hat{K}^\alpha{}_{\sigma\nu}\hat{S}_\alpha{}^\sigma{}_\mu,\end{aligned} \tag{14.169}$$

where we have used (cf. Eqs. (14.110) and (14.112))

$$\hat{K}^\alpha{}_{\mu\alpha} = -\hat{T}_\mu, \tag{14.170a}$$

$$\hat{K}^\alpha{}_{\alpha\mu} = 0, \tag{14.170b}$$

$$\hat{K}^\mu{}_{\nu\lambda} = \hat{S}_\lambda{}^{\mu\nu} + \delta^\nu_\lambda \hat{T}^\mu - \delta^\mu_\lambda \hat{T}^\nu. \tag{14.170c}$$

Now, we can analyse $\hat{\mathcal{K}}$, which gives (cf. Eq. (14.115))

$$\hat{\mathcal{K}} = 2\mathring{\nabla}_\lambda \hat{T}^\lambda + \hat{T} = \frac{2}{e}\partial_\lambda(e\hat{T}^\lambda) + \hat{T}. \tag{14.171}$$

Substituting Eqs. (14.169) and (14.171) into Eq. (14.168b), we have

$$\mathring{\nabla}_\alpha \hat{S}_{\nu\mu}{}^\alpha + \hat{K}^\alpha{}_{\sigma\nu}\hat{S}_\alpha{}^\sigma{}_\mu + \frac{1}{2}g_{\mu\nu}\hat{T} = 0, \tag{14.172}$$

which can be shown to be equal to Eq. (14.120) by exploiting metric compatibility, and tetrad postulate.

14.6.3 STEGR Field Equations

Since STEGR is curvatureless and torsionless, Eq. (14.155) can be further simplified via the coincident gauge (see Sect. 14.5.2), leading to the following expression:

$$\partial_\lambda R^\alpha{}_{\beta\mu\nu} + \partial_\nu R^\alpha{}_{\beta\lambda\mu} + \partial_\mu R^\alpha{}_{\beta\nu\lambda} = 0, \tag{14.173}$$

where, in this case, $R^\alpha{}_{\beta\mu\nu} \equiv \mathring{R}^\alpha{}_{\beta\mu\nu} + \overset{\diamond}{\mathcal{L}}{}^\alpha{}_{\beta\mu\nu} = 0$. $\overset{\diamond}{\mathcal{L}}{}^\alpha{}_{\beta\mu\nu}$ is a function of the disformation tensor, namely

$$\overset{\diamond}{\mathcal{L}}{}^\alpha{}_{\beta\mu\nu} = \mathring{\nabla}_\mu \overset{\diamond}{L}{}^\alpha{}_{\beta\nu} - \mathring{\nabla}_\nu \overset{\diamond}{L}{}^\alpha{}_{\beta\mu} + \overset{\diamond}{L}{}^\alpha{}_{\sigma\mu}\overset{\diamond}{L}{}^\sigma{}_{\beta\nu} - \overset{\diamond}{L}{}^\alpha{}_{\sigma\nu}\overset{\diamond}{L}{}^\sigma{}_{\beta\mu}, \tag{14.174}$$

endowed with the following symmetry properties:

$$\overset{\diamond}{\mathcal{L}}{}^{\alpha}{}_{\beta\mu\nu} = -\overset{\diamond}{\mathcal{L}}{}^{\beta}{}_{\alpha\mu\nu}, \qquad \overset{\diamond}{\mathcal{L}}{}^{\alpha}{}_{\beta\mu\nu} = -\overset{\diamond}{\mathcal{L}}{}^{\alpha}{}_{\beta\nu\mu}, \tag{14.175a}$$

where we have used Eqs. (14.4c) and (14.5c).

Contracting α and λ and giving the explicit expression of the Riemann tensor, Eq. (14.173) becomes

$$\begin{aligned} &\partial_\lambda \overset{\circ}{R}{}^{\lambda}{}_{\beta\mu\nu} + \partial_\mu \overset{\circ}{R}{}^{\lambda}{}_{\beta\nu\lambda} + \partial_\nu \overset{\circ}{R}{}^{\lambda}{}_{\beta\lambda\mu} \\ &+ \partial_\lambda \overset{\diamond}{\mathcal{L}}{}^{\lambda}{}_{\beta\mu\nu} + \partial_\mu \overset{\diamond}{\mathcal{L}}{}^{\lambda}{}_{\beta\nu\lambda} + \partial_\nu \overset{\diamond}{\mathcal{L}}{}^{\lambda}{}_{\beta\lambda\mu} = 0. \end{aligned} \tag{14.176}$$

Following the same strategy adopted in GR (see Sect. 14.6.1), we finally obtain (cf. Eqs. (14.168))

$$\overset{\circ}{R}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \overset{\circ}{R} = -\overset{\diamond}{\mathcal{L}}_{\mu\nu} + \frac{1}{2} g_{\mu\nu} \overset{\diamond}{\mathcal{L}}, \tag{14.177a}$$

$$\overset{\diamond}{\mathcal{L}}_{\mu\nu} - \frac{1}{2} g_{\mu\nu} \overset{\diamond}{\mathcal{L}} = 0, \tag{14.177b}$$

where $\overset{\diamond}{\mathcal{L}}_{\mu\nu} := \overset{\diamond}{\mathcal{L}}{}^{\alpha}{}_{\mu\alpha\nu}$ and $\overset{\diamond}{\mathcal{L}} := \overset{\diamond}{\mathcal{L}}{}^{\mu}{}_{\mu}$, resembling formally the expression of Ricci tensor and scalar curvature, respectively. Let us note that, in the coincident gauge, $\overset{\diamond}{L}{}^{\alpha}{}_{\mu\nu} = -\overset{\circ}{\Gamma}{}^{\alpha}{}_{\mu\nu}$, which soon reveals that Eq. (14.177b) is equivalent to the GR field equations. Equation (14.177a) proves the equivalence between GR and STEGR field equations, whereas Eq. (14.177b) represents the STEGR field equations, which we will demonstrate to be equal to Eq. (14.144). Let us first analyse $\overset{\diamond}{\mathcal{L}}_{\mu\nu}$, which yields

$$\begin{aligned} \overset{\diamond}{\mathcal{L}}_{\mu\nu} &= \overset{\circ}{\nabla}_\alpha \overset{\diamond}{L}{}^{\alpha}{}_{\mu\nu} - \overset{\circ}{\nabla}_\nu \overset{\diamond}{L}{}^{\alpha}{}_{\mu\alpha} + \overset{\diamond}{L}{}^{\sigma}{}_{\mu\nu} \overset{\diamond}{L}{}^{\alpha}{}_{\sigma\alpha} - \overset{\diamond}{L}{}^{\sigma}{}_{\mu\alpha} L^{\alpha}{}_{\sigma\nu} \\ &= \overset{\circ}{\nabla}_\alpha \overset{\diamond}{L}{}^{\alpha}{}_{\mu\nu} + \frac{1}{2} \overset{\circ}{\nabla}_\nu \overset{\diamond}{Q}_\mu - \frac{1}{2} \overset{\diamond}{Q}_\alpha \overset{\diamond}{L}{}^{\alpha}{}_{\mu\nu} \\ &- \frac{1}{4} \left[\overset{\diamond}{Q}_\mu{}^{\sigma}{}_{\alpha} \overset{\diamond}{Q}_\nu{}^{\alpha}{}_{\sigma} + 2 \overset{\diamond}{Q}{}^{\alpha}{}_{\sigma\nu} (\overset{\diamond}{Q}{}^{\sigma}{}_{\alpha\mu} - \overset{\diamond}{Q}_\alpha{}^{\sigma}{}_{\mu}) \right], \end{aligned} \tag{14.178}$$

where we have used

$$\overset{\diamond}{L}{}^{\alpha}{}_{\mu\alpha} = -\frac{1}{2} \overset{\diamond}{Q}_\mu, \tag{14.179a}$$

$$\begin{aligned} \overset{\diamond}{L}{}^{\alpha}{}_{\mu\nu} &= 2 \overset{\diamond}{P}{}^{\alpha}{}_{\mu\nu} + \frac{1}{2} g_{\mu\nu} (\overset{\diamond}{Q}{}^{\alpha} - \overset{\diamond}{\bar{Q}}{}^{\alpha}) \\ &- \frac{1}{4} (\delta^\alpha_\mu \overset{\diamond}{Q}_\nu + \delta^\alpha_\nu \overset{\diamond}{Q}_\mu). \end{aligned} \tag{14.179b}$$

Therefore, the scalar $\overset{\diamond}{\mathcal{L}}$ is expressed by

$$
\begin{aligned}
\overset{\diamond}{\mathcal{L}} &= \overset{\circ}{\nabla}_\alpha(\overset{\diamond}{Q}{}^\alpha - \overset{\diamond}{\tilde{Q}}{}^\alpha) + \frac{1}{4}\overset{\diamond}{Q}_{\alpha\beta\gamma}\overset{\diamond}{Q}{}^{\alpha\beta\gamma} - \frac{1}{2}\overset{\diamond}{Q}_{\alpha\beta\gamma}\overset{\diamond}{Q}{}^{\gamma\beta\alpha} \\
&\qquad - \frac{1}{4}\overset{\diamond}{Q}_\alpha\overset{\diamond}{Q}{}^\alpha + \frac{1}{2}\overset{\diamond}{Q}_\alpha\overset{\diamond}{\tilde{Q}}{}^\alpha \\
&= \overset{\circ}{\nabla}_\alpha(\overset{\diamond}{Q}{}^\alpha - \overset{\diamond}{\tilde{Q}}{}^\alpha) - \overset{\diamond}{Q}.
\end{aligned} \tag{14.180}
$$

Gathering together Eqs. (14.178) and (14.180), using the following identity (cf. Eq. (14.179a))

$$
\partial_\alpha\overset{\diamond}{Q}{}^\alpha = \overset{\circ}{\nabla}_\alpha\overset{\diamond}{Q}{}^\alpha + \overset{\diamond}{L}{}^\alpha{}_{\sigma\alpha}\overset{\diamond}{Q}{}^\sigma = \overset{\circ}{\nabla}_\alpha\overset{\diamond}{Q}{}^\alpha - \frac{1}{2}\overset{\diamond}{Q}_\alpha\overset{\diamond}{Q}{}^\alpha, \tag{14.181}
$$

we then obtain

$$
\begin{aligned}
&2\partial_\alpha P^\alpha{}_{\mu\nu} + \frac{1}{2}\overset{\diamond}{Q}_{\alpha\mu\nu}(\overset{\diamond}{Q}{}^\alpha - \overset{\diamond}{\tilde{Q}}{}^\alpha) + \frac{1}{2}g_{\mu\nu}\partial_\alpha(\overset{\diamond}{Q}{}^\alpha - \overset{\diamond}{\tilde{Q}}{}^\alpha) \\
&+ \frac{1}{2}\overset{\diamond}{L}{}^\sigma{}_{\mu\nu}\overset{\diamond}{Q}_\sigma + \frac{1}{4}\overset{\diamond}{Q}_\mu{}^\alpha{}_\sigma\overset{\diamond}{Q}_\nu{}^\sigma{}_\alpha + \frac{1}{2}\overset{\diamond}{Q}{}^\alpha{}_{\sigma\mu}(\overset{\diamond}{Q}{}^\sigma{}_{\nu\alpha} - Q_\alpha{}^\sigma{}_\nu) \\
&- \frac{1}{2}g_{\mu\nu}\overset{\circ}{\nabla}_\alpha(\overset{\diamond}{Q}{}^\alpha - \overset{\diamond}{\tilde{Q}}{}^\alpha) + \frac{1}{2}g_{\mu\nu}\overset{\diamond}{Q} = 0,
\end{aligned} \tag{14.182}
$$

which is equal to Eq. (14.144) in an empty space–time, i.e.

$$
\frac{2}{\sqrt{-g}}\partial_\alpha(\sqrt{-g}\overset{\diamond}{P}{}^\alpha{}_{\mu\nu}) - \frac{1}{\sqrt{-g}}\overset{\diamond}{q}_{\mu\nu} + \frac{1}{2}g_{\mu\nu}\overset{\diamond}{Q} = 0. \tag{14.183}
$$

An important remark is in order at this point. As already discussed above in the case of Lagrangians, the equivalence holds only for the theories stemming out from the scalar invariants R, T, and Q. In these specific cases, we obtain second-order equations. This is not true for extensions implying non-linear functions of these invariants. This fact points out again that GR, and its equivalent representations, are very peculiar cases among the theories of gravity.

14.7 Solutions in Trinity Gravity

Clearly, the equivalence of GR, TEGR, and STEGR has to be proven also at the solution level. In Sect. 14.6, the same field equations have been obtained, and then the same exact solutions, under the same symmetries and boundary conditions, have to be achieved.

In this perspective, performing the calculations to settle the solutions in the three gravity scenarios is useful also in view of extensions of the theories. Recently, it has been proposed a *3+1 splitting formalism in the geometric trinity of gravity* [50] entailing the following advantages: (1) simplicity in carrying out numerical analyses; (2) solving some theoretical issues existing in the various formulations of GR at the fundamental level (e.g. canonical quantization); (3) broadening this methodology also in extended and alternative gravity frameworks.

Here, we focus the attention on one of the simplest GR solutions, represented by the *Schwarazschild space–time*. Soon after the publication of GR theory by Einstein, Schwarzschild determined the solution, describing the space–time metric outside a spherically symmetric mass–energy distribution. This result is perfect agreement with the weak field approximation [90].

Jebsen, in 1921, and Birkhoff, in 1923, independently proved that the Schwarzschild solution holds outside a spherically symmetric mass distribution, even if this varies over time. This is now known as the *(Jebsen) Birkhoff theorem*, and it can be stated as follows [25, 122]: *any spherically symmetric solution of the GR field equations in vacuum has to be static and asymptotically flat. In addition, the Schwarzschild solution is the unique solution satisfying these hypotheses.*

This claim entails several significant implications: (1) the uniqueness of the Schwarzschild solution in GR by imposing the spherical symmetry as starting hypothesis; (2) no emission of gravitational waves, which can be interpreted, similarly as in electromagnetism, that there exists no monopole (spherically symmetric) radiation; (3) the outcome of the Birkhoff theorem in GR gravity theory can be compared with the Gauss theorem implications in electromagnetism and in classical Newtonian gravity.

Let us start our considerations taking into account a generic spherically symmetric metric, whose line element, written in spherical coordinates $\{t, r, \theta, \varphi\}$, in the equatorial plane $\theta = \pi/2$, and in geometric units ($G = c = 1$), reads as [140, 162]

$$ds^2 = -e^{\nu(t,r)}dt^2 + e^{\lambda(t,r)}dr^2 + r^2 d\varphi^2, \tag{14.184}$$

where $\nu(t, r)$, $\lambda(t, r)$ are the two unknown functions to be determined. We supplement this general metric with the well-known weak field limit on the metric time component

$$-e^{\nu(t,r)} \approx -1 + \frac{2M}{r}, \tag{14.185}$$

where M is the compact object mass, being the origin of the gravitational field, and $\frac{2M}{r}$ is the Newtonian gravitational potential.

We want now to solve the field equations in vacuum ($T_{\mu\nu} = 0$, namely outside the gravitational source) in GR (Sect. 14.7.1), TEGR (Sect. 14.7.2), and STEGR (Sect. 14.7.3). We will observe how the Birkhoff theorem emerges also in TEGR and STEGR. Finally, we will recover the *Schwarzschild metric* in all three gravity theories, namely [140, 162]

$$ds^2 = -\Big(1 - \frac{2M}{r}\Big)dt^2 + \Big(1 - \frac{2M}{r}\Big)^{-1} dr^2 + r^2 d\varphi^2, \tag{14.186}$$

admitting $r_S := 2M$ as event horizon (coordinate singularity) and $r = 0$ as essential (physical) singularity.

Given a function $f(t, r)$, we use the following notations

$$\dot{f}(t, r) := \frac{d f(t, r)}{dt}, \qquad f'(t, r) := \frac{d f(t, r)}{dr}. \tag{14.187}$$

14.7.1 Spherically Symmetric Solutions in GR

The vacuum field Eqs. (14.66) can be recast also as

$$\mathring{G}_{\mu\nu} \equiv \mathring{R}_{\mu\nu} = 0, \tag{14.188}$$

where $\mathring{R} = 0$. Analysing $\mathring{G}_{tr}$ we obtain

$$\mathring{G}_{tr} \equiv \frac{\dot{\lambda}(t, r)}{r} = 0, \quad \Rightarrow \quad \lambda = \lambda(r). \tag{14.189}$$

The other independent field equations are

$$\mathring{G}_{rr} \equiv -e^{\lambda(r)} + r\nu'(t, r) + 1 = 0, \tag{14.190a}$$

$$\mathring{G}_{tt} \equiv e^{-\lambda(r)}\big(r\lambda'(r) - 1\big) + 1 = 0. \tag{14.190b}$$

From Eq. (14.190a), we conclude that $\nu = \nu(r)$. All the metric components are independent of the coordinate time t, and this proves that the metric is *static*.

From Eq. (14.190b), we obtain

$$\big[e^{-\lambda(r)} r\big]' = 1 \quad \Rightarrow \quad e^{-\lambda(r)} = 1 - \frac{C_1}{r}, \tag{14.191}$$

where C_1 is an integration constant. Multiplying Eq. (14.190b) by $e^{\lambda(r)}$ and summing it to Eq. (14.190a), we obtain

$$\lambda'(r) + \nu'(r) = 0, \quad \Rightarrow \quad \lambda(r) + \nu(r) = C_2, \tag{14.192}$$

where the integration constant C_2 has to be $C_2 = 0$ to achieve the *asymptotic flatness*. From the *weak field limit* consideration (14.185), we obtain

$$-e^{\nu(r)} = 1 - \frac{2M}{r}, \qquad e^{\lambda(r)} = \frac{1}{1 - \frac{2M}{r}}. \tag{14.193}$$

14.7.2 Spherically Symmetric Solutions in TEGR

For solving the TEGR field equations, we adopt the tetrad formalism. We know that each tetrad field must be associated to the related spin connection (see, for example, Ref. [129]). However, in TEGR, we can drastically simplify the calculations resorting to the Weitzenböck gauge. Therefore, we can choose the diagonal tetrad

$$e^A_{\ \mu} = \begin{pmatrix} \sqrt{-e^{\nu(r)}} & 0 & 0 & 0 \\ 0 & \sqrt{e^{\lambda(r)}} & 0 & 0 \\ 0 & 0 & r & 0 \\ 0 & 0 & 0 & r\sin\theta \end{pmatrix}. \tag{14.194}$$

Let us recall that this tetrad is related to the off-diagonal tetrad (where the spin connection is naturally vanishing [129]) through a local Lorentz transformation $\Lambda^A_{\ B}(x)$. However, they both describe the same metric.

The non-zero torsion tensor components are

$$\hat{T}^t_{\ tr} = -\frac{1}{2}\nu'(r) = -\frac{M}{r^2}\left(1-\frac{2M}{r}\right)^{-1}, \tag{14.195a}$$

$$\hat{T}^\varphi_{\ r\varphi} = \frac{1}{r}. \tag{14.195b}$$

It is worth noticing that, physically, $\hat{T}^t_{\ tr}$ represents the redshifted radial gravitational force, because it is calculated with respect to the coordinate time t; whereas $\hat{T}^\varphi_{\ r\varphi}$ is the classical centrifugal force occurring in the tetrad frame.

Another important object is the contortion tensor, whose non-zero components read as

$$\hat{K}_{ttr} = \frac{1}{2}e^{\nu(r)}\nu'(r) = \frac{M}{r^2}, \tag{14.196a}$$

$$\hat{K}_{\varphi r\varphi} = r, \tag{14.196b}$$

whose interpretation is closely related to that already provided for the torsion tensor (cf. Eq. (14.97a)).

The superpotential components read as

$$\hat{S}_{\hat{t}}^{\ tr} = \frac{2e^{-\lambda(r)}\sqrt{e^{-\nu(r)}}}{r} = \frac{2}{r}\sqrt{1-\frac{2M}{r}}, \tag{14.197a}$$

$$\hat{S}_{\hat{\varphi}}^{\ r\varphi} = -\frac{e^{-\lambda(r)}\left(r\nu'(r)+2\right)}{2r^2} = \frac{M-r}{r^3}. \tag{14.197b}$$

Finally, the torsion scalar is

$$\hat{T} = -\frac{2e^{-\lambda(r)}\left(r\nu'(r)+1\right)}{r^2} = -\frac{2}{r^2}, \tag{14.198}$$

which represents the dynamically active part of the scalar curvature, whereas the remaining part is included in the dynamically passive boundary term.

Combining these elements, it is easy to prove that $\hat{G}_{\mu\nu} \equiv \overset{\circ}{G}_{\mu\nu}$ (cf. Eq. (14.172)). Then applying the same procedure of GR (see Sect. 14.7.1), the Birkhoff theorem holds also in TEGR.

14.7.3 Spherically Symmetric Solutions in STEGR

Regarding the STEGR field equations, we adopt the coincident gauge to ease the calculations, where $\overset{\diamond}{\nabla} = \partial_\mu$ and $\overset{\circ}{\Gamma}{}^{\mu}{}_{\alpha\beta} = -\overset{\diamond}{L}{}^{\mu}{}_{\alpha\beta}$. In this case, it is immediate to get $\overset{\circ}{G}_{\mu\nu} \equiv \overset{\diamond}{G}_{\mu\nu}$. However let us calculate the fundamental terms occurring in Eq. (14.177b) for extracting the physical information.

The non-metricity tensor has the following expression:

$$\begin{aligned}\overset{\diamond}{Q}_{r\mu\nu} &= \begin{pmatrix} -e^{\nu(r)}\nu'(r) & 0 & 0 \\ 0 & e^{\lambda(r)}\lambda'(r) & 0 \\ 0 & 0 & 2r \end{pmatrix} \\ &= \begin{pmatrix} -\frac{2M}{r^2} & 0 & 0 \\ 0 & -\frac{2M}{r^2\left(1-\frac{2M}{r}\right)^2} & 0 \\ 0 & 0 & 2r \end{pmatrix},\end{aligned} \tag{14.199}$$

where the derivative of gravitational potential represents the gravitational force acting on the observer and producing the disformations. For a comparison, we have that TEGR gravitational force makes the tetrad frame rotating (see Sect. 14.7.2), whereas STEGR gravitational force causes expansions and contractions of the observer laboratory. Instead, the conjugate potential reads as

$$\overset{\diamond}{P}{}^{t}{}_{tr} = \frac{r\lambda'(r) - r\nu'(r) + 4}{8r} = \frac{1}{8}\left(\lambda'(r) + \nu'(r)\right), \tag{14.200a}$$

$$\overset{\diamond}{P}{}^{r}{}_{rr} = \frac{e^{\nu(r)-\lambda(r)}}{r} = \frac{1}{r}\left(1 - \frac{2M}{r}\right)^2, \tag{14.200b}$$

$$\overset{\diamond}{P}{}^{r}{}_{\varphi\varphi} = -\frac{1}{4}re^{-\lambda(r)}\left(r\nu'(r) + 2\right) = \frac{M-r}{2}, \tag{14.200c}$$

$$\overset{\diamond}{P}{}^{\varphi}{}_{r\varphi} = \frac{1}{8}\left(\lambda'(r) + \nu'(r)\right) = 0, \tag{14.200d}$$

while the other components are null. The last quantity, represented by the above $q_{\mu\nu}$, reads as

$$\frac{\overset{\diamond}{q}_{\mu\nu}}{\sqrt{-g}} = \begin{pmatrix} \frac{2e^{\nu(r)-\lambda(r)}\nu'(r)}{r} & 0 & 0 \\ 0 & \frac{2r\nu'(r)+2}{r^2} & 0 \\ 0 & 0 & -\frac{r\nu'(r)+2}{e^{\lambda(r)}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{4M}{r^3}\left(1-\frac{2M}{r}\right) & 0 & 0 \\ 0 & \frac{2}{r^2\left(1-\frac{2M}{r}\right)} & 0 \\ 0 & 0 & \frac{2M}{r}-2 \end{pmatrix}. \tag{14.201}$$

Substituting the above expressions in Eq. (14.177b), we recover the same differential equations of GR (cf. Eq. (14.190)). Also, in this case, we obtain the Schwazrschild solution and the validity of the Birkhoff theorem.

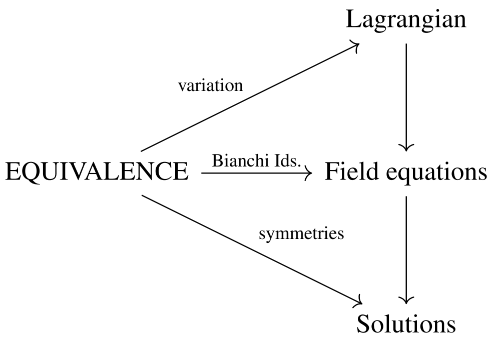
It is worth stressing that, also at this level, we cannot expect the same solutions for $f(R)$, $f(T)$, and $f(Q)$ extensions.

14.8 Discussion and Perspectives

We have gathered together basic concepts of Geometric Trinity of Gravity and derived the related dynamics pointing out analogies and differences between metric, affine, and non-metric approaches. We tried to give a self-consistent picture of the three representations of gravitational field. The main statement is that equivalence is strictly achieved for GR, TEGR, and STEGR and not for any extension of these theories.

Firstly, we introduced the geometric arena of metric-affine gravity, where metric tensor and affine connection are two separate and independent structures. After, we provided the fundamental geometric objects, that is tetrads and spin connection. The former represents the observer laboratory, which solders the tangent space to the space–time manifold. This procedure gives rise to anholonomic frames. These frames become holonomic when we are dealing with inertial frames, where a particular role is fulfilled by trivial tetrads of Special Relativity (see Sect. 14.4.1). The latter are intimately related with general tetrads, because they represent the inertial effects and they are generated by local Lorentz transformations. They form the Lorentz group, which, in turn, can be proved to give rise to a Lorentz algebra. This is a crucial aspect for defining the Fock–Ivanenko covariant derivative, useful to characterize the spin connection in terms of tetrads and to introduce the tetrad postulate (i.e. $\nabla_\mu e^A{}_\nu = 0$). This theoretical treatment can be interpreted from a physical point of view as discussed in Sect. 14.4.2.

These mathematical tools allow to describe the Geometric Trinity of Gravity. Specifically, a metric formulation (encoded in the Riemannian geometry), and a gauge approach (encoded in teleparallel gravity) are possible. GR, TEGR, and STEGR are dynamically equivalent from the variation of their Lagrangians up to a boundary term. Furthermore, starting from the second Bianchi identity, it is possible to infer the field equations, which are identical in the three representations. Finally, we analysed spherically symmetric solutions in the three theories deriving the Schwarzschild space–time and the Birkhoff theorem also in TEGR and STEGR. The approaches can be summarized as follows:



(14.202)

However, as pointed out above, also if mathematical results are equivalent, the physical interpretation can be different depending on the considered variables and observables. This fact opens several questions. Some of them can be listed as follows:

- Are there other equivalent formulations of gravity, outside of the Geometric Trinity? In other words, we can ask for the existence of other representations of gravity equivalent to GR within the metric-affine arena or, more in general, identifying other fundamental variables. The question implies also considering extended theories of gravity which can be "reduced" to GR (see, for example, [61, 124] for a discussion).
- From an observational point of view, what does it mean that these three theories are dynamically equivalent? This issue translates in extracting observables from each gravity theory and then interpreting them, from a physical viewpoint, finding out suitable transformation laws which make equivalent the set of variables of each theory.
- How can we construct observational apparatuses to test different theories dynamically equivalent to GR? This point is a direct consequence of the previous one. The question can be posed also in another way: Is it possible, if any, to discriminate different sets of observables for equivalent descriptions of gravity from an experimental point of view?
- STG theories are the less analysed among the three approaches. A general tetrad formulation is necessary in view of physical implications. In particular, the interpretation of gravity as a gauge theory could be particularly relevant to consider gravity under the same standard of other fundamental theories.

- The Equivalence Principle (in its strong and weak formulations) is a fundamental aspect of GR [183]. It can be recovered in TEGR and STEGR also if it is not at the foundation of these theories. If it were violated at some level (e.g. at quantum level), would it be possible to state that TEGR and STEGR are more fundamental theories than GR because they do not require it as a basic principle?

The above ones are some of the open issues related to equivalent representations of gravity and, in particular, to Gravity Trinity. Besides the mathematical aspects, it emerges that systematic experimental and observational protocols are necessary to establish the set of fundamental variables. For example, questions if metric or connection are the "true" gravitational variables are still open. Non-metricity could have a main role in this discussion due to the fact that the stringent requirement of asking for Equivalence Principle could be relaxed. Forthcoming precision experiments [88], gravitational wave astronomy [5], and precision cosmology observations [45] could be the tools to answer these questions.

Chapter 15
Conclusions

This book is an "experiment" to demonstrate that, starting from simple arguments of Euclidean geometry, it is possible to arrive at the geometric formulation of physical theories: in the specific case, Special and General Relativity and, consequently, up to Relativistic Cosmology. Our attempt was aimed, above all, at undergraduate students, in particular those of our university courses, to demonstrate that, by a rigorous and extended mathematical development, theories deemed "difficult", such as General Relativity, can be understood and operationally used. We turned to students, and not to colleagues, to avoid falling into unnecessary technicalities that would have made the text unsuitable for a truly "basic" reading of Special and General Relativity. During the discussion, however, we introduced some advanced topics with the aim of stimulating the reader to further deepen and personal research. Our hope is not to have bored the reader but to have contributed something useful to the vast literature on the subject. Ours was a humble attempt, with no claim to completeness. We hope that our efforts have proved useful to someone eager to understand the wonderful book of Nature with the beautiful language of Mathematics which, as said by Leibniz, is "*the honor of the human spirit*".

We are all in the gutter but some of us are looking at the stars.

Oscar Wilde

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9_15

References

1. Abbott, B.P., et al.: [ligo-virgo collaboration], GW151226: observation of gravitational waves from a 22-solar-mass binary black hole coalescence. Phys. Rev. Lett. **116** (2016)
2. Abbott, B.P., et al.: Observation of gravitational waves from a binary black hole merger. Phys. Rev. Lett. **116**, 061102 (2016)
3. Abbott, B.P., et al.: Gw170817: observation of gravitational waves from a binary neutron star inspiral. Phys. Rev. Lett. **119**, 161101 (2017)
4. Abbott, R., et al.: The population of merging compact binaries inferred using gravitational waves through GWTC-3. Phys. Rev. X **13**, 011048 (2023)
5. Abedi, H., Capozziello, S.: Gravitational waves in modified teleparallel theories of gravity. Eur. Phys. J. C **78**(6), 474 (2018)
6. Akiyama, K., others, and EHT Collaboration: first sagittarius A* event horizon telescope results. III. Imaging of the galactic center supermassive black hole. Astrophys. J. Lett. **930**(2), L14 (2022)
7. Aldrovandi, R., Guillen, L.C.T., Pereira, J.G., Vu, K.H.: Bringing together gravity and the quanta. In: Alimi, J.M., Füzfa, A. (eds.) Albert Einstein Century International Conference, American Institute of Physics Conference Series, vol. 861, p. 277 (2006)
8. Aldrovandi, R., Pereira, J.G., Vu, K.H.: Gravity and the quantum: are they reconcilable?. In: Adenier, G., Khrennikov, A., . Nieuwenhuizen, T.M. (eds.) Quantum Theory: Reconsideration of Foundations, volume 810 of American Institute of Physics Conference Series, p. 217 (2006)
9. Aldrovandi, R., Pereira, J.G.: Teleparallel Gravity. Springer, Dordrecht (2013)
10. Allemandi, G., Capone, M., Capozziello, S., Francaviglia, M.: Conformal aspects of palatini approach in extended theories of gravity. General Relativ. Gravit. **38**, 33 (2006)
11. Aloni, D., Berlin, A., Joseph, M., Schmaltz, M., Weiner, N.: A Step in Understanding the Hubble Tension (2021). arXiv:2111.00014
12. Altschul, B., et al.: Quantum tests of the Einstein equivalence principle with the STE-QUEST space mission. Adv. Space Res. **55**, 501 (2015)
13. Amendola, L., Tsujikawa, S.: Dark Energy: Theory and Observations. Cambridge University Press, Cambridge (2015)
14. Arnold, V.I.: Mathematical Methods of Classical Mechanics. Springer, New York (1997)
15. Arun, K., Gudennavar, S.B., Sivaram, C.: Dark matter, dark energy, and alternate models: a review. Adv. Space Res. **60**(1), 166 (2017)
16. Bahamonde S., Capozziello, S.: Noether symmetry approach in $f(T, B)$ teleparallel cosmology. Eur. Phys. J. C **77**(2), 107 (2017)

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*,
UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9

17. Bahamonde, S., Böhmer, C.G., Wright, M.: Modified teleparallel theories of gravity. Phys. Rev. D **92**(10), 104042 (2015)
18. Bahamonde, S., Dialektopoulos, K.F., Escamilla-Rivera, C., Farrugia, G., Gakis, V., Hendry, M., Hohmann, M., Said, J.L., Mifsud, J., Di Valentino, E.: Teleparallel Gravity: From Theory to Cosmology (2021). arXiv:2106.13793
19. Bailes, M., et al.: Gravitational-wave physics and astronomy in the 2020 and 2030s. Nat. Rev. Phys. **3**, 344 (2021)
20. Bamba, K., Capozziello, S., De Laurentis, M., Nojiri, S., Sáez-Gómez, D.: No further gravitational wave modes in $F(T)$ gravity. Phys. Lett. B **727**, 194 (2013)
21. Banks, E.C.: Ernst Mach's World Elements. Kluwer Academic Pub., Dordrecht (2003)
22. Beltrán Jiménez, J., Heisenberg, L., Koivisto, T.: Coincident general relativity. Phys. Rev. D **98**(4), 044048 (2018)
23. Beltrán Jiménez, J., Heisenberg, L., Koivisto, T.S.: The geometrical trinity of gravity. Universe **5**(7), 173 (2019)
24. Bennett, M.K.: Affine and Projective Geometry. Wiley, New York (2011)
25. Birkhoff, G.D., Langer, R.E.: Relativity Modern Physics. Harvard University Press, Harvard (1923)
26. Blagojevic, M.: Gravitation and Gauge Symmetries, 1st edn. CRC Press (2001)
27. Blixt, D., Hohmann, M., Pfeifer, C.: Hamiltonian and primary constraints of new general relativity. Phys. Rev. D **99**(8), 084025 (2019)
28. Bobe, A., Boskoff, W.G., Ciuca, M.: Tzitzeica-type centro-affine invariants in Minkowski spaces. An. Stiint. Univ. Ovidius **20** (2), 27 (2012)
29. Bonola, R.: Non-euclidean Geometry: A Critical and Historical Study of Its Development. Dover Books on Mathematics Series, Dover Publications Inc. (1955)
30. Boskoff, W.G., Capozziello, S.: Cosmological constant and ADS spacetimes from Minkowski spheres. Int. J. Geometrics Methods Modern Phys. **19**(08), 2250114 (2022)
31. Boskoff, W.G., Capozziello, S.: Recovering the cosmological constant from affine geometry. Int. J. Geometrics Methods Modern Phys. **16**(10), 1950161 (2019)
32. Boskoff, W.G.: A simple Gödel-type universe. Int. J. Geometric Methods Modern Phys. **20**(1), 2350018 (2023)
33. Boskoff, W.G.: Constant gaussian curvature, FLRW conditions and a universe without matter. Int. J. Geometrics Methods Modern Phys. **19**(11), 2250171 (2022)
34. Boskoff, W.G., Capozziello, S.: Mathematical Journey to Relativity. Unitext for Physics, 1st edn. Springer, Berlin (2020)
35. Boskoff, W.G., Suceava, B.D.: Distances obtained by Barbilian's Metrization Procedure. Houston J. Math. **33**, 709 (2007)
36. Boskoff, W.G., Vajiac, A.: Discovering Geometry: An Axiomatic Approach. Matrix Rom Publishing House, Bucharest (2012)
37. Boskoff, W.G., Vajiac, A., Suceava, B.D.: Exploration of Hilbert's neutral geometry. Gazeta Matematica A **26**, 1 (2008)
38. Boskoff, W.G., Suceava, B.D., Ciuca, M.: Revisiting the foundations of Barbilian's Metrization procedure. Differ. Geometry Appl. **29**, 577 (2011)
39. Bronnikov, K.A.: Scalar-tensor theory and scalar charge. Acta Physica Polonica B **4**, 251 (1973)
40. Buchdahl, H.A.: Isotropic coordinates and schwarzschild metric. Int. J. Theor. Phys. **24**(7), 731 (1985)
41. Bunge, M.: Mach's Critique of Newtonian mechanics. Am. J. Phys. **34**, 585 (1966)
42. Bunge, M.: Quantum Theory and Reality. Springer, Berlin (1967)
43. Burago, D., Burago, Y., Ivanov, S.: A Course in Metric Geometry. Graduate Studies in Mathematics, pp. 33. American Mathematical Society, Providence, Rhode Island (2001)
44. Burke, W.: Spacetime, Geometry, Cosmology. University Science Books, Mill Valley (1980)
45. Cai, Y., Capozziello, S., De Laurentis, M., Saridakis, E.N.: f(T) teleparallel gravity and cosmology. Rep. Prog. Phys. **79**(10), 106901 (2016)

46. Callahan, J.: The Geometry of Spacetime: Special and General Relativity. Springer, New York (2000)
47. Capozziello, S., Capriolo, M., Transirico, M.: The gravitational energy-momentum pseudotensor: the cases of $f(R)$ and $f(T)$ gravity. Int. J. Geometric Methods Modern Phys. **15**(01), 1850164 (2018)
48. Capozziello, S., Consiglio, L., De Laurentis, M., De Rosa, G., Di Donato, C.: The Missing Matter Problem: from the Dark Matter search to Alternative Hypotheses (2011). arXiv:1110.5026(astro-ph)
49. Capozziello, S. , De Falco, V., Ferrara, C.: Comparing equivalent gravities: common features and differences. Eur. Phys. J. C **82**(10), 865 (2022)
50. Capozziello, S., Finch, A., Said, J.L., Magro, A.: The 3+1 Formalism in teleparallel and symmetric teleparallel gravity. Eur. Phys. J. C **81**(12), 1141 (2021)
51. Capozziello, S., Francaviglia, M.: Extended theories of gravity and their cosmological and astrophysical applications. General Relativ. Gravit. **40**(2–3), 357 (2008)
52. Capozziello, S., Lambiase, G.: A Comprehensive View of Cosmological Dark Side (2013). ArXiv:1304.5640
53. Capozziello, S.: Curvature quintessence. Int. J. Mod. Phys. D **11**(4), 483 (2002)
54. Capozziello, S., D'Agostino, R.: Model-independent reconstruction of f(Q) non-metric gravity. Phys. Lett. B **832**, 137229 (2022)
55. Capozziello, S., Faraoni, V.: Beyond Einstein Gravity: A Survey of Gravitational Theories for Cosmology and Astrophysics. Springer, Dordrecht (2011)
56. Capozziello, S., Faraoni, V.: Beyond Einstein Gravity. Springer, Dordrecht (2011)
57. Capozziello, S., Francaviglia, M.: Extended theories of gravity and their cosmological and astrophysical applications. General Relativ. Gravit. **40**, 357 (2008)
58. Capozziello, S., Funaro, M.: Introduzione alla Relativitá Generale con Applicazioni all'Astrofisica Relativistica e alla Cosmologia. Liguori, Napoli (2005)
59. Capozziello, S., Godani, N.: Non-local gravity wormholes. Phys. Lett. B **835**, 137572 (2022)
60. Capozziello, S., De Laurentis, M.: A review about invariance induced gravity: gravity and spin from local conformal-affine symmetry. Found. Phys. **40**, 867 (2010)
61. Capozziello, S., De Laurentis, M.: Extended theories of gravity. Phys. Rep. **509**(4), 167 (2011)
62. Capozziello, S., De Laurentis, M.: The dark matter problem from f(R) gravity viewpoint. Annalen Phys. **524**, 545 (2012)
63. Capozziello, S., Basini, G., De Laurentis, M.: Deriving the mass of particles from extended theories of gravity in LHC era. Eur. Phys. J. C **71**, 1679 (2011)
64. Capozziello, S., Darabi, F., Vernieri, D.: Equivalence between palatini and metric formalisms of $f(R)$ gravity by divergence free current. Mod. Phys. Lett. A **26**, 65 (2011)
65. Capozziello, S., Jovanović, P., Jovanović, V.B., Borka, D.: Addressing the missing matter problem in galaxies through a new fundamental gravitational radius. JCAP **06**, 044 (2017)
66. Capuzzo-Dolcetta, R.: Classical Newtonian Gravity. Springer, Dordrecht (2019)
67. Carlip, S.: General Relativity: A Concise Introduction. Oxford University Press, Oxford (2019)
68. Carroll, S.M.: The cosmological constant. Living Rev. Relativ. **4**, 1 (2001)
69. Cartan, E.: Sur les Équations de Structure des Espaces Généralisés et l'Éxpression Analytique du Tenseur d'Einstein. Gauthier-Villars (1922)
70. Cartan, E.: Sur les Espaces Conformes Généralisés et l'Univers Optique. Gauthier-Villars (1922)
71. Cartan, E.: Sur les Espaces Généralisés et la Théorie de la Relativité . Gauthier-Villars (1922)
72. Cartan, E.: Sur une Définition Géométrique du Tenseur d'Énergie d'Einstein. Gauthier-Villars (1922)
73. Cartan, E.: Sur une Généralisation de la Notion de Courbure de Riemann et les Espaces à Torsion. Gauthier-Villars (1922)
74. Cartan, E.: La Théorie des Groupes Finis et Continus et l'Analysis Situs. Gauthier-Villars, Mémorial des Sciences Mathématiques (1952)

75. Cheng, T.P.: Relativity, Gravitation, and Cosmology: A Basic Introduction. Oxford, New York (2005)
76. Ciufolini, I., Pavlis, E.C.: A confirmation of the general relativistic prediction of the lense-thirring effect. Nature **431**(7011), 958 (2004)
77. Clarke, C.: Elementary General Relativity. Edward Arnold, London, UK (1979); Wiley, New York, (1980)
78. Clifton, T.: Alternative Theories of Gravity. Ph.D. thesis (2006)
79. Cohen-Tonnoudji, C., Laloe, B.D.F.: Quantum Mechanics. Wiley-VCH, New York (2006)
80. Collaboration, Planck, Aghanim, N.,: Planck, et al.: Results. VI. Cosmological parameters. Astron. Astrophys. **641**(A6), 2020 (2018)
81. Corry, L., Renn, J., Stachel, J.: Belated decision in the Hilbert-Einstein priority dispute. Science **278**(5341), 1270 (1997)
82. Coxeter, H.S.M.: A geometrical background for de Sitter's world. Am. Math. Mon. **50**, 217 (1943)
83. D' Ambrosio, F., Fell, S.D.B., Heisenberg, L., Kuhn, S.: Black holes in f (Q) gravity. Phys. Rev. D **105**(2), 024042 (2022)
84. De Falco, V., Capozziello, S.: Static and spherically symmetric wormholes in metric-affine theories of gravity. Phys. Rev. D **108**, 104030 (2023)
85. De Marchi, F., Cascioli, G.: Testing general relativity in the solar system: present and future perspectives. Class. Quantum Gravity **37**(9), 095007 (2020)
86. Deruelle, N., Uzan, J.P., de Forcrand-Millard, P.: Relativity in Modern Physics. Oxford Graduate Texts, OUP Oxford (2018)
87. Di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., Mota, D.M., Riess, A.G., Silk, J.: In the realm of the hubble tension-a review of solutions. Class. Quantum Gravity **38**(15), 153001 (2021)
88. Di Virgilio A.D., et al.: Sensitivity limit investigation of a sagnac gyroscope through llinear regression analysis: sagnac gyroscope sensitivity limit. Eur. Phys. J. C **81**(5), 400 (2021). [Erratum: Eur. Phys. J. C **81**, 457 (2021)]
89. Droste, J.: The field of a single centre in Einstein's theory of the gravitation and the motion of a particle in that field. Proc. R. Netherlands Acad. Arts Sci. **19**(1), 197 (1917)
90. Einstein, A.: Die Feldgleichungen der gravitation. Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), p. 1 (1915)
91. Einstein, A.: Die Grundlage der Allgemeinen Relativitätstheorie. Annalen der Physik, **354**(7), 769 (1916)
92. Einstein, A.: Kosmologische Betrachtungen zur Allgemeine Relativitätstheorie. Sitzungsberichte der königlich Preussischen Akademie der Wissenschaften, p. 142 (1917)
93. Ellis, H.G.: Ether flow through a drainhole: a particle model in general relativity. J. Math. Phys.**14**(1), 104 (1973)
94. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. II. Array and instrumentation. Astrophys. J. Lett. **875**(1), L2 (2019)
95. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. III. Data processing and calibration. Astrophys. J. Lett. **875**(1), L3 (2019)
96. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. IV. Imaging the central supermassive black hole. Astrophys. J. Lett. **875**(1), L4 (2019)
97. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. VI. The shadow and mass of the central black hole. Astrophys. J. Lett. **875**(1), L6 (2019)
98. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. V. Physical origin of the asymmetric ring. Astrophys. J. Lett. **875**(1), L5 (2019)
99. Event Horizon Telescope Collaboration, Akiyama, K., et al.: First M87 event horizon telescope results. I. The shadow of the supermassive black hole. Astrophys. J. Lett. **875**(1), L1 (2019)
100. Faraoni, V.: Special Relativity. Springer, Dordrecht (2013)
101. Fermi, E.: Sopra i Fenomeni che Avvengono in Vicinanza di una Linea Oraria. Rendiconti dell'Accademia dei Lincei **31**, 21 (1922)

102. Feynman, P.P., Leighton, R.B., Sands, M.: The Feynman Lectures on Physics. Addison-Wesley, Reading (1963)
103. Frolov, V.P., Novikov, I.D.: Black Hole Physics. Kluwer Academic Pub., Dordrecht (1998)
104. Gakis, V., Krššák, M., Said, J.L., Saridakis, E.N.: Conformal gravity and transformations in the symmetric teleparallel framework. Phys. Rev. D **101**(6), 064024 (2020)
105. Gauss, K.F.: Disquisitiones Generales circa Superficies Curvas. Typis Dieterichianis, Göttingen (1828)
106. Giulini, D., Kiefer, C., Laemmerzahl, C.: Quantum Gravity: From Theory to Experimental Search. Springer, Berlin (2003)
107. Gleiser, M.: Book review: a first course in string theory/Cambridge U. Press 2004. Phys. Today **58**(9), 57 (2005)
108. Gödel, K.: An example of a new type of cosmological solution of Einstein's field equations of gravitation. Rev. Mod. Phys. **21**, 447 (1949)
109. Goenner, H.F.M.: On the history of unified field theories. Living Rev. Relativ. **7**(1), 2 (2004)
110. Goenner, H.F.M.: On the history of unified field theories. Part II. (ca. 1930-ca. 1965). Living Rev. Relativ. **17**(1), 5 (2014)
111. Gravity Collaboration, Abuter, R., et al.: Detection of the Schwarzschild precession in the orbit of the star S2 near the galactic centre massive black hole. Astron. Astrophys. **636**, L5 (2020)
112. Hartman, T.: Lecture Notes on Classical de Sitter Space (2017). www.harmanhep.net/GR
113. Hawking, S.W., Ellis, G.F.: The Large Scale Structure of Spacetime. Cambridge University Press, Cambridge (1973)
114. Hayashi, K., and Shirafuji, T.: New general relativity. Phys. Rev. D **19**, 3524 (1979). [Addendum: Phys. Rev. D **24**, 3312 (1982)]
115. Hehl, F.W., McCrea, J.D., Mielke, E.W., Ne'eman, Y.: Metric-affine gauge theory of gravity: field equations, noether identities, world spinors, and breaking of dilation invariance. Phys. Rep. **258**(1), 1 (1995)
116. Hehl, F.H.: How does one measure torsion of spacetime? Phys. Lett. A **36**(3), 225 (1971)
117. Hehl, F.H.: Spin and torsion in general relativity I: foundations. General Relativ. Gravit. **4**(4), 333 (1973)
118. Hehl, F.H.: Spin and torsion in general relativity II: geometry and field equations. General Relativ. Gravit. **5**(5), 491 (1974)
119. Hehl, F.W., von der Heyde, P., Kerlick, J.D., Nester, J.M.: General relativity with spin and torsion: foundations and prospects. Rev. Mod. Phys. **48**, 393 (1976)
120. Hohmann, M., Järv, L., Krššák, M., Pfeifer, C.: Modified teleparallel theories of gravity in symmetric spacetimes. Phys. Rev. D **100**(8), 084002 (2019)
121. Hohmann, M.: General covariant symmetric teleparallel cosmology. Phys. Rev. D **104** (12), 1 (2021)
122. Johansen, N.V., Ravndal, F.: On the discovery of Birkhoff's theorem. General Relativ. Gravit. **38**(3), 1 (2006)
123. Kibble, T.W.B.: Lorentz Invariance and the gravitational field. J. Math. Phys. **2**(2), 212 (1961)
124. Kijowski, J.: Universality of the Einstein theory of gravitation. Int. J. Geometric Methods Modern Phys. **13**(08), 1640008 (2016)
125. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, vol. 1. A Wiley Publication in Applied Statistics, Wiley (1996)
126. Koivisto, T.: An integrable geometrical foundation of gravity. Int. J. Geometric Methods Modern Phys. **15**(1), 1840006 (2018)
127. Kolb, E.W., Turner, M.S.: The Early Universe. Addison-Wesley Pub. Co. Inc. Redwood City (1990)
128. Kramer, M., Stairs, I.H., Manchester, R.N., Wex, N., Deller, A.T., Coles, W.A., Ali, M., Burgay, M., Camilo, F., Cognard, I., Damour, T., Desvignes, G., Ferdman, R.D., Freire, P.C.C., Grondin, S., Guillemot, L., Hobbs, G.B., Janssen, G., Karuppusamy, R., Lorimer, D.R., Lyne, A.G., McKee, J.W., McLaughlin, M., Münch, L.E., Perera, B.B.P., Pol, N., Possenti, A., Sarkissian, J., Stappers, B.W., Theureau, G.: Strong-field gravity tests with the double pulsar. Phys. Rev. X **11**(4), 041050 (2021)

129. Krššák, M., van den Hoogen, R.J., Pereira, J.G., Böhmer, C.G., Coley, A.A.: Teleparallel theories of gravity: illuminating a fully invariant approach. Class. Quantum Gravity **36**(18), 183001 (2019)
130. Landau, L.D., Lifschits, E.M.: The Classical Theory of Fields. Course of Theoretical Physics, vol. 2. Pergamon Press, Oxford (1975)
131. Lasenby, A., Doran, C., Gull, S.: Gravity, gauge theories and geometric algebra. Philos. Trans. R. Soc. Lond. Ser. A **356**(1737), 487 (1998)
132. Lopez, R.: Differential geometry of curves and surfaces in Lorentz-Minkowski space. Int. Electron. J. Geometry **7**, 44 (2014)
133. Lorentz, H.A., Einstein, A., Minkowski, H., Weyl, H.: The Principle of Relativity: A Collection of Original Memoirs on the Special and General Theory of Relativity. Dover Inc., New York (1952)
134. Maggiore, M.: A Modern Introduction to Quantum Field Theory. EBSCO ebook Academic Collection. Oxford University Press, Oxford (2005)
135. Maggiore, M.: Gravitational Waves. Oxford University Press, Oxford (2018)
136. Maluf, J.W.: The teleparallel equivalent of general relativity. Annalen der Physik **525**(5), 339 (2013)
137. Manasse, F.K., Misner, C.W.: Fermi normal coordinates and some basic concepts in differential geometry. J. Math. Phys. **4**(6), 735 (1963)
138. Martin, G.E.: The Foundations of Geometry and the Non-euclidean Plane. Springer, New York (1975)
139. Mayet, F., Green, A.M., Battat, J.B.R., Billard, J., Bozorgnia, N., Gelmini, G.B., Gondolo, P., Kavanagh, B.J., Lee, S.K., Loomba, D., Monroe, J., Morgan, B., O'Hare, C.A.J., Peter, A.H.G., Phan, N.S., Vahsen, S.E.: A review of the discovery reach of directional dark matter detection. Phys. Rep. **627**, 1 (2016)
140. Misner, C.W., Thorne, K.S., Wheeler, J.A.: Gravitation. W. H. Freeman, San Francisco (1973)
141. Misner, C., Thorne, K.S., Wheeler, J.A.: Gravitation. W.H. Freeman, San Francisco (1973)
142. Moore, T.: A General Relativity Workbook. University Science Books, Mill Valley (2013)
143. Morris, M., Thorne, K., Yurtsever, U.: Wormholes, time machines, and the week energy condition. Phys. Rev. Lett. **61**(13), 1446 (1988)
144. Nakahara, M.: Geometry, Topology and Physics. CRC Press, New York, London (2003)
145. Nester, J.m., Yo, H.J.: Symmetric teleparallel general relativity. Chinese J. Phys. **37**(2), 113 (1999)
146. Ni, W.: Solar-system tests of the relativistic gravity. Int. J. Modern Phys. D **25**(14), 1630003, 786 (2016)
147. Nojiri, S., Odintsov, S.D.: Introduction to modified gravity and gravitational alternative for dark energy. eConf C0602061, 6 (2006)
148. Nojiri, S., Odintsov, S.D.: Unified cosmic history in modified gravity: from F(R) theory to lorentz Non-invariant models. Phys. Rep. **505**(2), 59 (2011)
149. Nojiri, S., Odintsov, S.D.: Introduction to modified gravity and gravitational alternative for dark energy. Int. J. Geometrics Methods Modern Phys. **4**, 115 (2007)
150. Nojiri, S., Odintsov, S.D., Oikonomou, V.K.: Modified gravity theories on a nutshell: inflation, bounce and late-time evolution. Phys. Rep. **692**, 1 (2017)
151. Nomizu, K., Sasaki, T.: Affine Differential Geometry. Cambridge University Press, Cambridge, UK (1994)
152. Obukhov, Y.N., Pereira, J.G.: Lessons of spin and torsion: reply to "Consistent Coupling to Dirac Fields in Teleparallelism". Phys. Rev. D **69**(12), 128502 (2004)
153. Obukhov, Y.N.: Gravitational waves in poincaré gauge gravity theory. Phys. Rev. D **95**(8), 084028 (2017)
154. Ohanian, H.C., Ruffini, R.: Gravitation and Spacetime. Cambridge University Press, Cambridge (2013)
155. Oshita, N., Wu, Y.P.: Role of spacetime boundaries in a vierbein formulation of gravity. Phys. Rev. D **96**(4), 044042 (2017)

156. Ozel, F.: Black hole physics with the event horizon telescope. Am. Astron. Soc. Meeting Abstracts **227**, 200 (2016)
157. Palatini, A.: Deduzione Invariantiva delle Equazioni Gravitazionali dal Principio di Hamilton. Rendiconti del Circolo Matematico di Palermo **43**, 203 (1919)
158. Peacock, J.A.: Physical Cosmology. Cambridge University Press, Cambridge (1999)
159. Peebles, P.J.E.: Principles of Physical Cosmology. Princeton University Press, Princeton (1993)
160. Pereira, J.G.: Lorentz connections and gravitation. AIP Conf. Proc. **1483**, 239 (2012)
161. Poincaré, H.: La Science et l' Hypothése. E. Flammarion Editeur, Paris (1917)
162. Romano, A., Furnari, M.M.: The Physical and Mathematical Foundations of the Theory of Relativity: A Critical Analysis. Springer International Publishing (2019)
163. Rosen, N.: General Relativity and Flat Space. Phys. Rev. **57**, 147 (1940)
164. Rovelli, C.: Quantum Gravity. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge (2004)
165. Ruegg, H., Ruiz-Altaba, M.: The stueckelberg field. Int. J. Modern Phys. A **19**(20), 3265 (2004)
166. Ryan, M.P., Spatilly, L.C., Shepley, L.C.: Homogeneous Relativistic Cosmologies. Princeton Series in Physics. Princeton University Press (1975)
167. Ryan, M.P., Shepley, L.C.: Homogeneous Relativistic Cosmologies. Princeton University Press, Princeton (1975)
168. Sathyaprakash, B.S., Schutz, B.F.: Physics, astrophysics and cosmology with gravitational waves. Living Rev. Relativ. **12**(1), 2 (2009)
169. Schwarzschild, K.: Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften **7**, 189 (1916)
170. Sciama, D.W.: On the analogy between charge and spin in general relativity. In: Recent Developments in General Relativity, p. 415. Polish Scientific Publishers, Warsaw (1962)
171. Sormani, C., Denson Hill, C., Nurowski, P., Bieri, L., Garnfinkle, D., Yunes, N.: The mathematics of gravitational waves. Not. AMS **7**, 685 (2017)
172. Sotiriou, T.P., Faraoni, V.: f(R) Theories of gravity. Rev. Modern Phys. **82**(1), 451 (2010)
173. Spradlin, M., Strominger, A., Volovich, A.: Les Houches lectures on de sitter space. In: Bachas, C.P., Bilal, A., Douglas, M.R., Nekrasov, N.A., David, F. (eds.) Les Houches Session LXXVI, 31 July 30 August 2001, Unity from Duality: Gravity, Gauge Theory and Strings. Springer, Berlin (2002)
174. Stairs, I.H.: Testing general relativity with pulsar timing. Living Rev. Relativ. **6**(1), 5 (2003)
175. Starobinsky, A.A.: A new type of isotropic cosmological models without singularity. Phys. Lett. B **91**, 99 (1980)
176. Stephani, H.: General Relativity. Cambridge University Press, Cambridge (1990)
177. Stephani, H., Kramer, D., MacCallum, M., Hoenselaers, C., Herlt, E.: Exact Solutions of Einstein's Field Equations. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge (2009)
178. Straumann, N.: General Relativity. Springer, Dordrecht (2012)
179. Surhone, L.M., Timpledon, M.T., Marseken, S.F.: Poincaré Disk Model. Betascript Pub., Mauritius (2010)
180. 't Hooft, G., Veltman, M.: One-loop divergencies in the theory of gravitation. Annales de L'Institut Henri Poincare Section (A) Physique Theorique **20**(1), 69 (1974)
181. Thiemann, T.: Loop quantum gravity: an inside view. In: Stamatescu, I.O., Seiler, E. (eds.) Lecture Notes in Physics, vol. 721, p. 185. Springer, Berlin (2007)
182. Tino, G.M., et al.: SAGE: a proposal for a space atomic gravity explorer. Eur. Phys. J. D **73**, 228 (2019)
183. Tino, G.M., Cacciapuoti, L., Capozziello, S., Lambiase, G., Sorrentino, F.: Precision gravity tests and the Einstein equivalence principle. Prog. Particle Nuclear Phys. **112**, 1 (2020)
184. Tino, G.M., Cacciapuoti, L., Capozziello, S., Lambiase, G., Sorrentino, F.: Precision gravity tests and the Einstein equivalence principle. Prog. Particle Nuclear Phys. **112**, 103772 (2020)

185. Torretti, B.: Philosophy of Geometry from Riemann to Poincaré. Reidel Pub., Dordrecht (1978)
186. Tzitzeica, G.: Sur une Nouvelle Classe de Surfaces. Les Comptes Rendus de l'Académie des sciences **144**, 1257 (1907)
187. Tzitzeica, G.: Sur une Nouvelle Classe de Surfaces (la deuxième partie). Rendiconti del Circolo Matematico di Palermo **28**, 210 (1909)
188. Unzicker, A., Case, T.: Translation of Einstein's Attempt of a Unified Field Theory with Teleparallelism (2005). arXiv:physics/0503046
189. Varićak, V.: Über die Nichteuklidische Interpretation der Relativtheorie. Jahreisbericht der Deutschen Mathematiker-Vereinigung **21**, 103 (1912)
190. Vitagliano, V., Sotiriou, T.P., Liberati, S.:. The dynamics of metric-affine gravity. Ann. Phys. **326**(5), 1259 (2011)
191. Weinberg, S.: The Quantum Theory of Fields. Vol. 1: Foundations. Cambridge University Press (2005)
192. Weyl, H.: Elektron und gravitation. I, Zeitschrift für Physik **56**(5–6), 1 (1929)
193. Weyl, H.: *Space—Time—Matter*. Dover Publications Inc. (1952)
194. Wheeler, J.T.: Weyl geometry. General Relativ. Gravit. **50**(7), 80 (2018)
195. Will, C.M.: Theory and Experiment in Gravitational Physics. Cambridge University Press, Cambridge (1993)
196. Will, C.M.: Theory and Experiment in Gravitational Physics. Cambridge University Press, Cambridge, UK (1993)
197. Zakharov, A.F., Zinchuk, V.A., Pervushin, V.N.: Tetrad formalism and reference frames in general relativity. Phys. Particles Nuclei **37**(1), 104 (2006)
198. Zaroubi, S.: The epoch of reionization. In: Wiklind, T., Mobasher, B., Bromm, V. (eds.) The First Galaxies. Springer, Berlin (2013)
199. Zel'dovich, Y.B., Novikov, I.D.: The Structure and Evolution of the Universe. University of Chicago Pr, Chicago (1983)
200. Zwicky, F.: On the masses of nebulae and of clusters of nebulae. Astroph. J. **86**, 217 (1937)

Index

W.-G. Boskoff and S. Capozziello, *A Mathematical Journey to Relativity*, UNITEXT for Physics, https://doi.org/10.1007/978-3-031-54823-9