A Modern View of Relativity (A Rigorous Introduction for Mathematicians)

J.D. Vélez, Camilo Arias, Alexander Quintero, Dany Gómez

2019

Contents

Conventions **ix**

1 INTRODUCTION XXXXX **3**

I Mathematical Background **5**

2 Multilinear Algebra **7**
- 2.1 Bases and Matrices . . . 7
- 2.2 The Dual of a Vector Space . . . 8
- 2.3 Bilinear Functions . . . 8
- 2.4 Tensor Products . . . 10
  - 2.4.1 Construction of the Tensor Product . . . 10
  - 2.4.2 Tensors . . . 13
  - 2.4.3 Change of Bases . . . 14
  - 2.4.4 Contraction of Tensors. . . . 15
  - 2.4.5 Raising and Lowering Indices . . . 16

3 Geometry **19**
- 3.1 Differentiable Manifolds . . . 19
- 3.2 The Tangent Space . . . 24
- 3.3 Vector Bundles . . . 28
- 3.4 The Tangent Bundle and Vector Fields . . . 31
- 3.5 Vector Fields and Flows . . . 40
- 3.6 The Cotangent Bundle . . . 49
- 3.7 Tensor Fields . . . 50
- 3.8 Pull-backs and Lie Derivatives . . . 51

4 How the Metric Determines the Geometry **53**
4.1 The Metric Tensor . . . 53
4.2 Examples . . . 55
4.3 Length of a Curve . . . 58
4.4 Isometries and Killing Vector Fields . . . 59

5 Connections, Parallel transport and Geodesics **65**
5.1 Connections . . . 65
5.2 The Levi-Civita Connection . . . 67
5.3 Pullback of Bundles and Connections . . . 73
5.4 Parallel Transport . . . 76
5.5 Geodesics . . . 77

6 Curvature **81**
6.1 The Riemann Curvature Tensor . . . 81
6.2 The Ricci tensor, scalar curvature and the Einstein tensor . . . 85
6.3 Sectional Curvature . . . 88
6.4 Curvature as Geodesic Deviation . . . 93

7 Differential Forms **95**
7.1 Differential Forms . . . 95
7.2 DeRham Cohomology . . . 96
7.3 The De Rham operator in $\mathbb{R}^3$ . . . 99

8 Integration and Stokes Theorem **101**
8.1 Manifolds with Boundary . . . 101
8.2 Orientations . . . 103
8.3 Integration . . . 105
8.4 Stokes' Theorem . . . 108
8.5 Classical Stokes Theorem . . . 111
8.6 Conservation of Mass . . . 115
8.7 The Hodge Star Operator . . . 116
8.8 Lie Groups . . . 118
8.9 Partitions of Unity . . . 121
8.10 The Picard-Lindelöf theorem . . . 122

II A Brief Introduction to Electromagnetism 125

9 Introduction to Electromagnetism 127
9.1 Coulomb's Law . . . 128
9.2 Electrostatics: Charges at Rest . . . 128
9.2.1 Electric Potential . . . 130
9.3 Electrodynamics: Moving Charges . . . 130
9.4 The Law of Biot-Savart . . . 131
9.5 The Divergence of B . . . 133
9.6 Magnetostatics . . . 134
9.7 Varying Fields . . . 138
9.8 Faraday's Law of Induction . . . 139
9.9 Maxwell's Equations . . . 142
9.9.1 Maxwell's Equations under Galilean Transformations . 145
9.10 Maxwell's Equations in Minkowski spacetime . . . 148

10 Special Relativity 151
10.1 The Michelson-Morley Experiment . . . 151
10.1.1 Geometric Units for Time . . . 157
10.2 Lorentz Transformations . . . 158
10.3 Perception of Time and Space . . . 162
10.3.1 Muon Decay: An Experimental Test for Special Relativity . . . 167
10.4 Observers in Minkowski Spacetime . . . 168
10.4.1 Proper Time . . . 169
10.4.2 Velocities under Lorentz Transformations . . . 170
10.5 4-Momentum, 4-Force and Proper Acceleration . . . 172
10.6 Maxwell's Equations Revisited . . . 174
10.7 Equivalence of Mass and Energy . . . 177
10.8 4-Momentum . . . 180
10.8.1 Particles with Nonzero Rest Mass. . . . 180
10.8.2 Particles with Zero Rest Mass . . . 182
10.9 Conservation of Momentum . . . 183
10.10 Accelerated Observers . . . 183
10.10.1 Time Dilation: A Journey to Kepler 22-b . . . 185
10.11 Redshift and blueshift . . . 187

III How Energy and Matter Curve Space-Time 191

11 Foundations of the General Theory of Relativity 193
11.1 Foundations . . . 193
11.2 Observers and Physical Coordinates . . . 196
11.3 Light Cones . . . 208

12 Examples of Space-Time Manifolds 211
12.1 Gravitational Waves . . . 211
12.2 Rotating Universes . . . 214
12.2.1 Rotating observer . . . 214
12.3 Rotating Universes of Godel XXXXX . . . 216
12.4 The FRLW Metric . . . 216
12.4.1 Null Geodesic and Causality . . . 217
12.4.2 Cosmological Red Shift and Hubble's Law . . . 220
12.4.3 Red Shift Factor for Nearby Celestial Objects . . . 225
12.4.4 Age and Diameter of the Visible Universe . . . 227

13 Tensor of Energy-Momentum 229
13.1 Flows of Dust . . . 229
13.2 Perfect Fluids . . . 233
13.3 The Energy Conservation Law . . . 235
13.3.1 Interpretation of The Conservation Law . . . 235

14 The Field Equations 239
14.1 The Gravitational Potential . . . 239
14.2 Poisson's Equation . . . 240
14.3 Tidal Forces . . . 243
14.4 The Equivalence Principle . . . 245
14.5 Estimate of g_{00} . . . 249
14.6 Geodesics in a Weak Gravitational Field . . . 251
14.7 The Astonishing Analogy . . . 253
14.8 Einstein's Equations in Plain English . . . 256
14.8.1 Geometric Units . . . 259
14.9 EINSTEIN'S EQUATIONS FROM A VARIATIONAL VIEWPOINT XXXXX . . . 260

IV Solutions to Einstein's Equation **261**

15 The Schwarzschild Metric **263**
15.1 Static Space-Time Manifolds . . . 264
15.1.1 Fermat's Principle . . . 265
15.2 Spherical symmetry . . . 267
15.3 Birkhoff's Theorem . . . 271
15.4 The Exterior of a Non Rotating Star . . . 273
15.5 Redshift in Schwarzschild Space-Time . . . 275
15.6 Spherical Shells . . . 276
15.6.1 Interior of a Uniformly Dense Star . . . 280
15.7 Geometry Inside a Spherical Empty Cavity . . . 281
15.8 Time Machines . . . 282
15.9 Planetary Motion . . . 286
15.10Motion of Nonzero Mass Particles . . . 289
15.11Orbits in Schwarzschild's Space-Time . . . 291
15.12Radial Plunge Orbits . . . 295
15.13Precession of Mercury's Perihelion . . . 298
15.14Motion of Massless Particles . . . 303
15.15Gravitational Bending of Light . . . 306

16 Black Holes **315**
16.1 Introduction . . . 315
16.2 Eddigton-Finkelstein Coordinates . . . 316
16.3 Kruskal-Szekeres Coordinates . . . 321
16.4 Hawking Radiation . . . 325

17 ROTATING BLACK HOLES XXXXX **329**

18 Cosmological Models **331**
18.1 The FRLW Metric . . . 331
18.2 A FURTHER SECTION ON COSMOLOGY XXXXX . . . 342

19 HAWKING AND PENROSE SINGULARITIES THEOREM XXXXX **345**

Conventions

- Manifolds are denoted by letters $M, N, S, \ldots$.
- Smooth functions are denoted by $f, h, g \ldots$.
- Vector fields (and also tangent vectors at a point) will be denoted by $X, Y, Z, \ldots$.
- In local coordinates we use the following notations for vector fields:

$$X = \sum_a f^a \frac{\partial}{\partial x^a} = \sum_a f^a \partial_{x^a} = \sum_a f^a \partial_a.$$

  We use Latin small letters to denote a set of indices for a Lorentzian manifold (space-time coordinates), that will run from 0 to 3, and $i, j, k, ...$ to denote indices for $\mathbb{R}^3$, or in general for local coordinates of a manifold.

- Capital Latin letters will denote ordered multiindices: $I = (i_1, \ldots, i_k)$, $i_1 < i_2 < \cdots < i_k$. Differential forms are denoted by greek letters $\omega, \theta, \eta, \ldots$ In local coordinates a differential form takes the form:

$$\omega = \sum_I f_I dx^I = \sum_I f_I \; dx^{i_1} \wedge \cdots \wedge dx^{i_k}.$$

- A tensor written in local coordinates is denoted by:

$$T = \sum_{(a),(b)} T^{(a)}{}_{(b)} \; v_{(a)} \otimes \omega^{(b)}$$

  where $v_{(a)} = v_{a_1} \otimes \cdots \otimes v_{a_r}$, $\omega^{(b)} = \omega^{b_1} \otimes \cdots \otimes \omega^{b_s}$ $T^{(a)}{}_{(b)} = T^{a_1 \ldots a_r}{}_{b_1 \cdots b_s}$, and the sum runs over all multiindices $(a) = (a_1, \ldots, a_r)$, $(b) = (b_1, \ldots, b_s)$.

- A chart in a manifold is denoted by letters $\varphi, \phi, \ldots, \varphi : U \to V; \quad \varphi = (x^1, \ldots, x^m)$.
- Open sets are denoted by $U, V, W, \ldots$.
- The metric is denoted by g.
- The tangent space at a point p is denoted by T_pM.
- The derivative of a function $f : M \to N$ is denoted by $Df(p)$

Preface

It has been more than a century since Einstein's published his extraordinary Theory of Relativity. However, there are still a few textbooks that present Einstein's theory using the coordinate-free language of modern differential geometry.

Even though the classical notation is of great help in many situations, it is useless when it comes to machine-aid computations since every programing language demands objects to be defined using linear algebra. On the other hand, for the student accustomed to the language of modern mathematics, purely mechanical manipulation of symbols could be conceptually quite confusing. In fact, coordinates are sometimes used just as labels for space-time events, and some other times they represent actual measurements carried out by a particular observer. I this book we try to keep a balance between the classical and the more modern language. We hope this book may serve as a bridge between the formal language that is common among mathematicians [32], and the classical terminology used in most textbooks in physics.

This book is aimed at advanced undergraduate and graduate math students who want to understand the fundamentals of Einstein's theory of relativity (Special and General). Our aim is to give a rigorous introduction to the Theory without falling into the error of developing the Theory as some sort of long appendix to a book on differential geometry. On the contrary, we have put special care into motivating the fundamental mathematical concepts from a clear a precise physical viewpoint. With this in mind, we have included a terse introduction to electromagnetism, with the aim of formulating Maxwell's equations, historically the main incentive for the development of Einstein's theory of relativity.

We will assume some familiarity with the basic notions of multilinear algebra and differential geometry. Nevertheless, in the first part of the book, we offer a summary of the main topological and geometric notions that are necessary to develop Einstein's theory. By itself, this material could be used for a short course in differential geometry, having the theory of General Relativity as one of its prime objectives. We have also included an appendix with a more comprehensive presentation of some other topics that, even though they are nor used in this book, they are needed in a more advanced course in relativity. There are many excellent references for this material. The reader may refer to [14], [24], [35] (just to cite a few) for gentle and concise introductions to the modern language of differential geometry.

There are many very good books at the introductory level for Einstein's Special Theory of Relativity (SR) ([33], [31]). As for the General Theory (GR), [11] and [4] are superb introductions. More advanced material may be consulted in [38], [40], [12]. A concise and very good introduction is given in [1]. The classical, and probably most complete reference for the subject, is the thick and quite comprehensive book by Misner, Wheeler and Thorne [26].

We intend to develop in a unified manner both the General and Special Theory of Relativity (the latter being just the spacial case in which the metric is the global Minkowski metric). For this purpose, it is convenient to introduce from the beginning the general setting in which Einstein's theory is developed. The traditional approach is to present SR first, and only then (GR) is developed. In our opinion, the Special Theory is, in some sense, "too special": Many concepts with a rather clear physical content in SR are sometimes not so easily translated into GR. This very often becomes a cause for misunderstandings, and a reason for wrong intuitions and general confusion. This is why we believe it would be better to introduce the subject in such a way that from the beginning every physical notion is introduced in the general setting of GR. Good physical motivation can always be provided by resorting to SR.

Chapter 1

INTRODUCTION XXXXX

Part I

Mathematical Background

Chapter 2

Multilinear Algebra

In this section we review the main notions of multilinear algebra and introduce the notation for tensor products and its operations that we will used along these notes. A full treatment of the subject, including complete proofs for all statements, is available in any standard text in linear algebra. We recommend [19].

2.1 Bases and Matrices

Let V and V' be two real vector spaces of dimensions n and m, respectively. Each choice of bases $B = \{v_a\}$ and $B' = \{v'_a\}$ for V and V' gives rise to a representation of the elements of V and V' as column vectors: For each $u = \sum_a l^a v_a$ in V (sums will always run over the indices $a = 0, \ldots, n-1$, where we enumerate the entries of a vector starting with zero, as it is customary in most textbook in Relativity) the *transposed* of the row vector $u = (l^a)$ is the column vector with the same entries, denoted by $[u]_B$. If $f : V \to V'$ is a linear transformation, by $F = [f]_{B'B}$ we denote *the matrix associated to f in the bases B and B'*, defined as follows: The b-th column of F is the column vector with entries $l_{0b}, \ldots, l_{m-1b}$, which are the coefficients of the vector $f(v_b)$ expressed in the basis B': That is, $f(v_b) = \sum_a l_{ab} v'_a$. In particular, if $V = V'$, and $f = Id$ is the identity, the matrix $[Id]_{B'B}$ is the bases-change matrix from B to B'. Using this notation it readily follows that $[u]_{B'} = [Id]_{B'B}[u]_B$.

The correspondence between linear transformations and matrices respects the composition of functions. That is, if $f : V \to V'$ and $g : V' \to V''$ are linear transformations, and B, B' and B'' are bases for V, V' and V'', respectively, then, as it is shown in the elementary linear algebra courses, $[g \circ f]_{B''B} = [g]_{B''B'}[f]_{B'B}$.

If V and V' are vector spaces, $\mathrm{Hom}_{\mathbb{R}}(V, V')$ will denote the vector space of linear maps from V to V'. It has dimension mn, where n denotes the dimension of V and m denotes the dimension of V'. Moreover, any choice of bases B and B' determines an isomorphism between $\mathrm{Hom}_{\mathbb{R}}(V, V')$ and $\mathrm{Mat}_{m\times n}(\mathbb{R})$, the space of $m \times n$ matrices with entries in $\mathbb{R}$, given by the linear map:

$$\mathrm{Hom}_{\mathbb{R}}(V, V') \to \mathrm{Mat}_{m\times n}(\mathbb{R}), \;\; f \mapsto [f]_{B'B}.$$

2.2 The Dual of a Vector Space

We recall that *the dual* of a vector space V, denoted by V^*, is the vector space of all linear functional, i.e., the space $\mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$. Any linear map $f : V \to W$ induces, canonically, another linear transformation $f^* : W^* \to V^*$, by sending each functional $\omega \in W^*$ into $\omega \circ f \in V^*$. For each choice of basis $B = \{v_a\}$ for V we denote by $B^* = \{\omega^a\}$ its *dual basis*, where ω^a is the functional that takes the value 1 when evaluated at v_a, and the value zero when evaluated at any other vector v_b, with $b \neq a$. If $F = [f]_{\widetilde{B}B}$ represents f with respect to the bases B and $\widetilde{B}$, then the linear map f^* is represented (in the respective dual bases) by *the transpose of* F (the matrix obtained from F by interchanging rows and columns). That is, $F^* = [f^*]_{B^*\widetilde{B}^*}$.

2.3 Bilinear Functions

We recall that a map $g : V \times V \to \mathbb{R}$ is called *bilinear* if it is linear in each of the two arguments separately. Once we fix a basis B for V, the map g can be uniquely represented in matrix form as: $g(v, u) = [v]_B^* \; G \; [u]_B$, where each entry g_{ab} of the matrix G is given by $g_{ab} = G(v_a, v_b)$ ($[v]_B^*$ denotes the transpose of the vector $[v]_B$). The matrix G will be called *the associated matrix of* g in the basis B, and will be denoted by $[g]_{BB}$.

If B_1 and B_2 are any two bases for V, one has $[v]_{B_2} = [Id]_{B_2B_1}[v]_{B_1}$, and

$[u]_{B_2} = [Id]_{B_2B_1}[u]_{B_1}$. Hence, we see that

$$\begin{aligned} g(v,u) &= [v]^*_{B_2}[g]_{B_2B_2}[u]_{B_2} \\ &= \left([v]^*_{B_1}[Id]^*_{B_2B_1}\right)[g]_{B_2B_2}\left([Id]_{B_2B_1}[u]_{B_1}\right) \\ &= [v]^*_{B_1}\left([Id]^*_{B_2B_1}[g]_{B_2B_2}[Id]_{B_2B_1}\right)[u]_{B_1}. \end{aligned}$$

From the uniqueness of Equation (2.3) it follows that the product of matrices inside the parenthesis must be equal to $[g]_{B_1B_1}$, from which one obtains:

$$[g]_{B_1B_1} = [Id]^*_{B_2B_1}[g]_{B_2B_2}[Id]_{B_2B_1}. \tag{2.1}$$

The bilinear map g is called *symmetric* if $g(v,u) = g(u,v)$. The condition of being symmetric implies that if $G = [g]_{B_2B_1}$ is a representation of g in any fixed pair of bases B_1 and B_2 then $G^* = G$. The map g is called *non-degenerated* if $g(v,-)$ is the zero function only if $v = 0$. A symmetric and non-degenerated bilinear map g is called an *inner product*, hereby, denoted by $g(v,u)$, and also by $\langle v,u\rangle_g$.This product is called *positive-definite* if it also satisfies $g(v,v) \geq 0$, for all vectors v. The norm of a vector v is defined as $|v| = \sqrt{\varepsilon\,\langle v,v\rangle_g}$, with $\varepsilon = 1$, if $\langle v,v\rangle_g \geq 0$, and $\varepsilon = -1$, if $\langle v,v\rangle_g < 0$.

In general, a space V and it dual V^* are not isomorphic in a canonical way. However, any inner product $\langle -,-\rangle_g$ in V defines a canonical isomorphism between V and V^* that sends $v \in V$ into the functional $\lambda^v : V \to \mathbb{R}$ defined as: $\lambda^v(u) = \langle v,u\rangle_g$. In any vector space V endowed with an inner product g there is always a basis C that puts G in diagonal form (see [19], Chapter XIV). That is, such that:

$$[g]_{C,C} = \mathrm{diag}[-1,\ldots,-1,1,\ldots,1], \tag{2.2}$$

where $r \geq 0$ entries in the diagonal equal to -1, and $s \geq 0$ entries equal to 1. The $(r+s)$-tuple $(-1,\ldots,-1,1,\ldots,1)$, called the *signature* of g, only depends on g and not on the particular choice of C. One such basis is called an *orthonormal basis*.

Starting from any basis $B = \{v_a\}$ in a vector space edowed with a bilinear form with signature as in (2.2), there is a standard procedure to produce an orthonormal basis $\{e_a\}$ from B (the *Gram-Schmidt algorithm*)): One starts by defining e_0 as $e_0 = v_0/|v_0|$. Then, the vectors $e_1,\ldots,e_{n-1}$ are constructed inductively as $e_k = e_k/|\mathfrak{e}_k|$, where

$$\mathfrak{e}_k = v_k - \varepsilon_k \sum_{i=0}^{k-1} \langle v_k, e_i\rangle_g\, e_i, \tag{2.3}$$

with $\varepsilon_0 = \cdots = \varepsilon_{r-1} = -1$, and $\varepsilon_r = \cdots = \varepsilon_{n-1} = +1$.

2.4 Tensor Products

Let V and $\widetilde{V}$ be two vector spaces. We want to construct a vector space $V \otimes \widetilde{V}$ and a bilinear map $\varepsilon : V \times \widetilde{V} \to V \otimes \widetilde{V}$ with the following *universal property*: Given any bilinear map $B : V \times \widetilde{V} \to Z$ into another vector space Z there exists a unique linear transformation $L_B : V \otimes \widetilde{V} \to Z$ such that the following diagram commutes:

$$\begin{array}{ccc} V \times \widetilde{V} & \stackrel{B}{\to} & Z \\ \varepsilon \downarrow & \nearrow & L_B \\ V \otimes \widetilde{V} & & \end{array}$$

It is not hard to see that if $(V \otimes \widetilde{V}, \varepsilon)$ exist, then it must be unique up to isomorphism. This means that if (U, ε') is another pair satisfying the same universal property, then there is an isomorphism $h : V \otimes \widetilde{V} \to U$ making the following diagram commute:

$$\begin{array}{ccc} V \times \widetilde{V} & & \\ \varepsilon \downarrow & \searrow & \varepsilon' \\ V \otimes \widetilde{V} & \stackrel{h}{\to} & U \end{array}$$

2.4.1 Construction of the Tensor Product

Let V, $\widetilde{V}$ be real vector spaces, and let F denote the vector space of all real linear combinations of elements of the set $B = \{e(v, \widetilde{v}) : v \in V, \widetilde{v} \in \widetilde{V}\}$:

$$F = \{f : f = a_1 e(v_1, \widetilde{v}_1) + \cdots + a_n e(v_n, \widetilde{v}_n), \text{ with } v_i \in V, \, \widetilde{v}_j \in \widetilde{V}\}.$$

One declares two elements of F as equal iff the coefficients a_i in the expression above are the same. The set F is given a real vector space structure in a natural way: if $f = \sum a_i e(v_i, \widetilde{v}_i)$ and $g = \sum b_i e(v_i, \widetilde{v}_i)$ are elements of F one defines: $f + g = \sum (a_i + b_i) e(v_i, \widetilde{v}_i)$, $\alpha f = \sum \alpha a_i e(v_i, \widetilde{v}_i)$, $\alpha \in \mathbb{R}$. With this structure, B is, a priory, a basis for F. Let H be the subspace generated by vectors of one of the following four forms:

1. $e(v_1 + v_2, \widetilde{v}) - e(v_1, \widetilde{v}) - e(v_2, \widetilde{v})$

2. $e(\alpha v, \widetilde{v}) - \alpha e(v, \widetilde{v})$

3. $e(v, \widetilde{v}_1 + \widetilde{v}_2) - e(v, \widetilde{v}_1) - e(v, \widetilde{v}_2)$

4. $e(v, \alpha\widetilde{v}) - \alpha e(v, \widetilde{v})$,

for all $v, v_i \in V, \widetilde{v}, \widetilde{v}_i \in \widetilde{V}$, and $\alpha \in \mathbb{R}$.

We define *the tensor product* of V and $\widetilde{V}$, denoted by $V \otimes \widetilde{V}$, as the quotient vector space F/H. The equivalent class of $e(v, \widetilde{v})$ is denoted by $v \otimes \widetilde{v}$.

It is apparent from this definition that each element in the tensor product can be written (not necessarily in a unique way) as a sum of the form $\sum_i \alpha_i \, v_i \otimes \widetilde{v}_i$. Since the classes of $e(v, \widetilde{v}_1 + \widetilde{v}_2)$ and $e(v, \widetilde{v}_1) + e(v, \widetilde{v}_2)$ are the same, it is obvious that $\otimes$ satisfies that: $v \otimes (\widetilde{v}_1 + \widetilde{v}_2) = v \otimes \widetilde{v}_1 + v \otimes \widetilde{v}_2$. Similarly,

$$\begin{aligned}(\widetilde{v}_1 + \widetilde{v}_2) \otimes v &= \widetilde{v}_1 \otimes v + \widetilde{v}_2 \otimes v, \\ (\widetilde{v}_1 + \widetilde{v}_2) \otimes v &= \widetilde{v}_1 \otimes v + \widetilde{v}_2 \otimes v, \text{ and} \\ \alpha(v \otimes \widetilde{v}) &= (\alpha v) \otimes \widetilde{v} = v \otimes (\alpha\widetilde{v}).\end{aligned}$$

We define $\varepsilon : V \times \widetilde{V} \to V \otimes \widetilde{V}$ as $\varepsilon(v, \widetilde{v}) = v \otimes \widetilde{v}$. This function is clearly bilinear. It is easy to see that $(V \otimes \widetilde{V}, \varepsilon)$ satisfies the universal property that characterizes the tensor product construction. The following propositions are standard ([19], Chapter XVI).

Proposition 2.4.1. Let V, $\widetilde{V}$, $V_1, \ldots, V_n$, $\widetilde{V}_1, \ldots, \widetilde{V}_m$, Z be vector spaces. There exist canonical isomorphisms of vector spaces:

1. $V \otimes \widetilde{V} \simeq \widetilde{V} \otimes V$, where $v \otimes \widetilde{v}$ is sent into $\widetilde{v} \otimes v$.

2. $V \otimes (\widetilde{V} \otimes Z) \simeq (V \otimes \widetilde{V}) \otimes Z$, where $v \otimes (\widetilde{v} \otimes z)$ is sent into $(v \otimes \widetilde{v}) \otimes z$.

3. $\mathbb{R} \otimes V \simeq V$, where $\alpha \otimes v$ is sent into αv.

4. $(V_1 \oplus V_2 \oplus \cdots \oplus V_n) \otimes \widetilde{V} \simeq (V_1 \otimes \widetilde{V}) \oplus (V_2 \otimes \widetilde{V}) \oplus \cdots \oplus (V_n \otimes \widetilde{V})$, where $(v_1, \ldots, v_n) \otimes \widetilde{v}$ is sent into $(v_1 \otimes \widetilde{v}, \ldots, v_n \otimes \widetilde{v})$.

5. If $B = \{v_a\}$ and $\widetilde{B} = \{\widetilde{v}_a\}$ are bases for V and $\widetilde{V}$, respectively then $B \otimes \widetilde{B} = \{v_a \otimes \widetilde{v}_b\}$ is a basis for $V \otimes \widetilde{V}$. If $\dim(V) = m$, and $\dim(\widetilde{V}) = n$ one has that $\dim(V \otimes \widetilde{V}) = mn$.

Proposition 2.4.2. Let $f : V \to V'$ and $g : Z \to Z'$ be linear transformations.

1. The maps f and g induce a linear map $f \otimes g : V \otimes Z \to V' \otimes Z'$ that sends each $v \otimes z$ into $(f \otimes g)(v \otimes z) = f(v) \otimes g(z)$.

2. Let B and B' be bases for V and V', and let C and C' be bases for Z and Z', respectively. If $F = [f]_{B'B}$ and $G = [g]_{Z'Z}$ are the matrices that represent f and g in these bases, the matrix $H = [f \otimes g]_{(B'\otimes Z')\,(B\otimes Z)}$ represents $f \otimes g$ in the bases $B \otimes Z$, and $B' \otimes Z'$. The matrix H is known as the Kronecker product of F and G. If $A = [\alpha_{ij}]$ and $B = [\beta_{ij}]$ are matrices of sizes $p \times n$ and $q \times m$, respectively, the Kronecker product is the $pq \times mn$-matrix whose block form is:

$$A \otimes B = \begin{bmatrix} \alpha_{11}B & \cdots & \alpha_{1n}B \\ \vdots & & \vdots \\ \alpha_{p1}B & \cdots & \alpha_{pn}B \end{bmatrix}$$

We leave the following assertions as exercises.

Exercise 2.4.3. Show that the map $V^* \otimes Z \to \mathrm{Hom}_{\mathbb{R}}(V, Z)$ that sends each generators $\lambda \otimes z \in V^* \otimes Z$ into the linear transformation $h_{\lambda\otimes z} : V \to \widetilde{V}$, with $h_{\lambda\otimes z}(u) = \lambda(u)z$, is a well defined isomorphism of vector spaces.

Exercise 2.4.4. *I*n general, the tensor product of vector spaces $V_1, \dots, V_r$ can be defined as a pair $(V_1 \otimes \cdots \otimes V_r, \varepsilon)$, where $\varepsilon : V_1 \times V_2 \times \cdots \times V_r \to V_1 \otimes \cdots \otimes V_r$ is a multilinear map (linear in each factor) that satisfies the following universal property: Given a multilinear map T, there is a unique linear transformation L_T making the following diagram commutative:

$$\begin{array}{ccc} V_1 \times \cdots \times V_r & \xrightarrow{T} Z \\ \varepsilon \downarrow & \nearrow L_T \\ V_1 \otimes \cdots \otimes V_r & \end{array}$$

Prove the following assertions:

1. The pair $(V_1 \otimes \cdots \otimes V_r, \varepsilon)$ exists and is unique, up to canonical isomorphisms.

2. The vector space $\mathrm{Mult}(V \times \cdots \times V, Z)$ of all multilinear functions from $V \times \cdots \times V$ into Z (with the natural operations) is canonically isomorphic to $\mathrm{Hom}_{\mathbb{R}}(V \otimes \cdots \otimes V, Z)$.

3. If $f_i : V_i \to Z_i$ are linear transformations, there exists a linear map

$$f_1 \otimes \cdots \otimes f_r : V_1 \otimes \cdots \otimes V_r \to Z_1 \otimes \cdots \otimes Z_r,$$

that sends each generator $v_1 \otimes \cdots \otimes v_r$ into $f_1(v_1) \otimes \cdots \otimes f_r(v_r)$.

4. Let $B_j = \{v_1^j, \ldots, v_{n_j}^j\}$ and $Z_j = \{z_1^j, \ldots, z_{m_j}^j\}$ be bases for V_j and Z_j, respectively, the set $B = B_1 \otimes \cdots \otimes B_r$ of all elements of the form

$$B = \left\{ v_{j_1}^1 \otimes \cdots \otimes v_{j_s}^s \otimes \cdots \otimes v_{j_r}^r : v_{j_s}^s \in B_s \right\},$$

is a basis for $V_1 \otimes \cdots \otimes V_r$. The matrix for $f_1 \otimes f_2 \otimes \cdots \otimes f_r$ in the bases B, $Z = \left\{ z_{j_1}^1 \otimes \cdots \otimes z_{j_s}^s \otimes \cdots \otimes z_{j_r}^r : z_{j_s}^s \in Z_s \right\}$ is the Kronecker product $A_1 \otimes A_2 \otimes \cdots \otimes A_r$, where $A_j = [f_j]_{Z_j B_j}$.

2.4.2 Tensors

Let V be a vector space, and let V^* be its dual. *A tensor of type* (p, q) is an element of the vector space $T^{(p,q)}(V) = V^{\otimes_p} \otimes (V^*)^{\otimes_q}$, where $V^{\otimes_p}$ denotes the tensor product of p copies of V, and $(V^*)^{\otimes_q}$ denotes the tensor product of q copies of V^*. Any linear map $f : V \to V$ induces in a natural way another linear map, denoted by $T^{(p,q)}(f)$, defined as the tensor product $(f \otimes \cdots \otimes f) \otimes (f^* \otimes \cdots \otimes f^*)$ of p copies of f and q copies of f^*. This map sends each element $v_1 \otimes \cdots \otimes v_p \otimes \omega^1 \otimes \cdots \otimes \omega^q$ into $f(v_1) \otimes \cdots \otimes f(v_p) \otimes f^*(\omega^1) \otimes \cdots \otimes f^*(\omega^q)$.

Let $B = \{e_a\}$ be a basis for V, and let $B^* = \{\omega^a\}$ be its dual basis for V^*. We will denote by (a) an ordered p-multi-index (allowing repetitions) $1 \leq a_1 \leq \cdots \leq a_p \leq n$. If (a) and (b) are a p-multi-index and a q-multi-index, respectively, by $e_{(a)}$ and $\omega^{(b)}$ we denote the vectors $e_{a_1} \otimes \cdots \otimes e_{a_p}$ and $\omega^{b_1} \otimes \cdots \otimes \omega^{b_q}$, respectively. With respect to the basis for $T^{(p,q)}(V)$ given by $B^{(p,q)} = \{e_{(a)} \otimes \omega^{(b)}\}$ any (p, q)-tensor can be written as $T = \sum_{(a),(b)} T^{(a)}{}_{(b)} e_{(a)} \otimes \omega^{(b)}$. Explicitly

$$T = \textstyle\sum_{(a),(b)} T^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}\ e_{a_1} \otimes \cdots \otimes e_{a_p} \otimes \omega^{b_1} \otimes \cdots \otimes \omega^{b_q}.$$

Once the basis $B^{(p,q)}$ is fixed, it is customary to refer to T by its components $T^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}$.

With this notation, the $(p+r, q+s)$ *outer product* of the (p,q)-tensor T and the (r,s)-tensor S can be written as the tensor whose components are $R^{a_1\ldots a_p c_1 \ldots c_r}{}_{b_1\ldots b_q d_1\ldots d_s} = T^{a_1\ldots a_p}{}_{b_1\ldots b_q}\ S^{c_1\ldots c_r}{}_{d_1\ldots d_s}$, where $T^{a_1\ldots a_p}{}_{b_1\ldots b_q}$ and $S^{c_1\ldots c_r}{}_{d_1\ldots d_s}$ are the components of T and S. More precisely,

$$T \otimes S = \textstyle\sum_{(a),(b),(c),(d)} T^{(a)}{}_{(b)}\ S^{(c)}{}_{(d)}\ e_{(a)} \otimes e_{(c)} \otimes \omega^{(b)} \otimes \omega^{(d)}.$$

Any tensor $T = \sum_{(a),\ (b)} T^{a_1\ldots a_p}{}_{b_1\ldots b_q}\ e_{a_1} \otimes \cdots \otimes e_{a_p} \otimes \omega^{b_1} \otimes \cdots \otimes \omega^{b_q}$ can be regarded *as a multilinear map* $T : (V^*)^{\otimes p} \otimes (V)^{\otimes q} \to \mathbb{R}$ (this is the usual definition in most textbooks) that sends each element $\varepsilon = \lambda_1 \otimes \cdots \otimes \lambda_p \otimes u_1 \otimes \cdots \otimes u_q$ into the scalar

$$T(\varepsilon) = \textstyle\sum_{(a),(b)} T^{a_1\ldots a_p}{}_{b_1\ldots b_q}\ \lambda_1(e_{a_1}) \cdots \lambda_p(e_{a_p})\ u_1^{b_1} \cdots u_q^{b_q}.$$

It is useful to consider both viewpoints. In these notes we regard T as an element of $T^{(p,q)}$, and also as a multilinear map.

2.4.3 Change of Bases

Let $B = \{e_a\}$ and $\widetilde{B} = \{\widetilde{e}_a\}$ be two bases for V. Let $B^* = \{\omega^b\}$ and $\widetilde{B}^* = \{\widetilde{\omega}^b\}$ be the corresponding dual bases for V^*. We know that

$$B^{(p,q)} = \left\{ e_{(c)} \otimes \omega^{(d)} : (c),\ (d) \text{ are } p \text{ and } q\text{-multi-indices, respectively} \right\},$$

and

$$\widetilde{B}^{(p,q)} = \{\widetilde{e}_{(a)} \otimes \widetilde{\omega}^{(b)} : (a) \text{ and } (b),\ \text{ are } p \text{ and } q\text{-multi-indices, respectively}\}$$

are bases for $T^{(p,q)}(V)$. Let $L = [Id]_{\widetilde{B}B}$ be the bases-change matrix form B to $\widetilde{B}$. We know that L^* (its transpose) is the bases-change matrix from $\widetilde{B}^*$to B^*, i.e., $L^* = [Id^*]_{B^*\widetilde{B}^*}$. Thus, $R = (L^*)^{-1} = [Id^*]_{\widetilde{B}^* B^*}$ is the bases-change matrix from B^* to $\widetilde{B}^*$. Let's see how to compute the bases-change matrix from $B^{(p,q)}$ to $\widetilde{B}^{(p,q)}$; that is, the matrix $A^{(p,q)} = [Id]_{\widetilde{B}^{(p,q)} B^{(p,q)}}$. By definition, the entries l_{ac} of the c-th column of A satisfy $e_c = \sum_a l_{ac}\widetilde{e}_a$. If r_{bd} denote the entries of R, then $\omega^d = \sum_b r_{bd}\widetilde{\omega}^b$. Hence, $e_{(c)} \otimes \omega^{(d)}$ is equal to

$$\begin{aligned}(\textstyle\sum_a l_{a\ c_1}\widetilde{e}_a) \otimes \cdots \otimes (\sum_a l_{a\ c_p}\widetilde{e}_a) \otimes (\sum_b r_{b\ d_1}\widetilde{\omega}^b) \otimes \cdots \otimes (\sum_b r_{b\ d_q}\widetilde{\omega}^b)\\ = \textstyle\sum_{(a),(b)} l_{a_1\ c_1} l_{a_2\ c_2} \cdots l_{a_p\ c_p}\ r_{b_1\ d_1} \cdots r_{b_q\ d_q}\ \widetilde{e}_{(a)} \otimes \widetilde{\omega}^{(b)}.\end{aligned} \tag{2.4}$$

Let T be an arbitrary element of $T^{(p,q)}(V)$. Writing

$$T = \sum_{(c),(d)} T^{c_1 \ldots c_p}{}_{d_1 \ldots d_q}\ e_{(c)} \otimes \omega^{(d)} \text{ and } T = \sum_{(a),(b)} \widetilde{T}^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}\ \widetilde{e}_{(a)} \otimes \widetilde{\omega}^{(b)}$$

in the bases $B^{(p,q)}$and $\widetilde{B}^{(p,q)}$, respectively, Equation (2.4) shows that the coefficients of T in the bases $B^{(p,q)}$ and $\widetilde{B}^{(p,q)}$are related by the following formula:

$$\widetilde{T}^{a_1 \ldots a_p}{}_{b_1 \ldots b_q} = \sum_{(c),(d)} l_{a_1 c_1} l_{a_2 c_2} \cdots l_{a_p c_p}\ r_{b_1 d_1} \cdots r_{b_q d_q}\ T^{c_1 \ldots c_p}{}_{d_1 \ldots d_q} \tag{2.5}$$

2.4.4 Contraction of Tensors.

Let T be a (p,q)-tensor (viewed as a multilinear map). With respect to a fixed basis, *the contraction of T with respect to the indices* $1 \leq i \leq p$ and $1 \leq j \leq q$ is the multilinear map defined as $C^i{}_j T(\cdots) = \sum_v T(\cdots \underset{i}{\omega^v}, \cdots \underset{j}{e_v}, \cdots)$. Writing $T = \sum_{(a),\,(b)} T^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}\ e_{(a)} \otimes \omega^{(b)}$, the contraction with respect to the indices i and j is the $(p-1, q-1)$-tensor $C^i{}_j T$ given in that same basis by:

$$(C^i{}_j T)^{a_1 \ldots \widehat{a_i} \ldots a_p}{}_{b_1 \ldots \widehat{b_j} \ldots b_q} = \sum_v T^{a_1 \ldots v \ldots a_p}{}_{b_1 \ldots v \ldots b_q},$$

where the hat $\widehat{}$ means that the corresponding index has been deleted. In fact, with respect to the basis $e_{(a)} \otimes \omega^{(b)}$ the tensor $C^i{}_j T$ can be written as:

$$C^i{}_j T(\cdots) = \sum_v \sum_{(a),(b)} T^{(a)}{}_{(b)}\ e_{a_1} \otimes \cdots \otimes e_{a_i}(\omega^v) \otimes \cdots e_{a_p} \otimes \omega^{b_1} \otimes \cdots \otimes \omega^{b_j}(e_v) \otimes \cdots \otimes \omega^{b_q}.$$

For each fixed v, the only nonzero terms in the latter sum would be

$$\sum_{\substack{(a_1, \ldots, \underset{i}{v} \ldots a_p) \\ (b_1, \ldots, \ldots, \underset{j}{v}, \ldots b_q)}} T^{a_1 \ldots v \ldots a_p}{}_{b_1 \ldots v \ldots b_q} T^{(a)}{}_{(b)}\ e_{a_1} \otimes \cdots \otimes \widehat{e_v} \otimes \cdots e_{a_p} \otimes \omega^{b_1} \otimes \cdots \otimes \widehat{\omega^v} \otimes \cdots \otimes \omega^{b_q}$$

From this, one immediately sees that:

$$(C^i{}_j T)^{a_1 \ldots \widehat{a_i} \ldots a_p}{}_{b_1 \ldots \widehat{b_j} \ldots b_q} = \sum_v T^{a_1 \ldots v \ldots a_p}{}_{b_1 \ldots v \ldots b_q}.$$

Now we verify that $C^i{}_j T$ is well defined, i.e., does not depend on the chosen basis. Let us express $e_v = \sum_s l_{sv} \widetilde{e}_s$, and $\omega^v = \sum_t r_{tv} \widetilde{\omega}^t$, as in Section 2.4.3, where $L = [Id]_{\widetilde{B}B} = [l_{sv}]$, and $R = (L^*)^{-1} = [r_{tv}]$ are the change of base matrices. Since $R^* = L^{-1}$, the dot product of the s-th row of $L = [Id]_{\widetilde{B}B}$ and the t-th column of $R^* = L^{-1}$, equal to the sum $\sum_v l_{sv} r_{tv}$, must be δ_{st}, where δ_{st} denotes the delta function of Kronecker. Taking this into account we have:

$$\begin{aligned} C^i{}_j T(\cdots) &= \textstyle\sum_v T(\cdots, \underset{i}{\omega^v}, \cdots, e_v, \cdots) = \sum_v T(\cdots \sum_t r_{tv} \widetilde{\omega}^t, \cdots, \sum_s l_{sv} \widetilde{e}_s, \cdots) \\ &= \textstyle\sum_{s,t} \sum_v l_{sv} r_{tv} T(\cdots, \widetilde{\omega}^t, \cdots, \widetilde{e}_s, \cdots) = \sum_{s,t} \delta_{st} T(\cdots \widetilde{\omega}^t, \cdots \widetilde{e}_s, \cdots) \\ &= \textstyle\sum_s T(\cdots \widetilde{\omega}^s, \cdots \widetilde{e}_s, \cdots), \end{aligned}$$

and therefore the contraction of a tensor, as defined above, does not depend on the chosen basis.

2.4.5 Raising and Lowering Indices

Let V be a vector space endowed with an inner product $g = \langle -, - \rangle$. This bilinear form allows us to identify V and V^* canonically by sending a vector v into the functional λ^v defined as $\lambda^v(-) = \langle v, - \rangle$. Let's fix a basis $B = \{e_a\}$ for V, and let $B^* = \{\omega^a\}$ be the corresponding dual basis. Let $G = [g_{ab}]$ denote the matrix $[g]_{BB}$, and let G^{-1} denote its inverse. It is customary to write the entries of G^{-1} as g^{ab}. If $v = \sum_a v^a e_a$ is a vector, then

$$\lambda^v(e_b) = \textstyle\sum_a v^a \langle e_a, e_b \rangle = \sum_a v^a g_{ab}.$$

Thus, λ^v can be written in the dual basis $\{\omega^a\}$ as $\lambda^v = \sum_b \mu_b \omega^b$, with $\mu_b = \sum_s g_{sb} v^s$. Physicists prefer to denote this as $\omega_b = v^a g_{ab}$, where *Einstein's notation for summation* is implied, and where ω_a denotes the coefficient μ_a. The dual vector ω_b is said to have been obtained from v^a by *lowering its index.*

On the other hand, if $\mu = \sum_b \mu_b \omega^b$ is given, then it is easy to verify that if $v^a = \sum_s g^{as} \mu_s$ then $v = \sum_a v^a e_a$ is the unique vector with the property that $\lambda^v = \mu$. In most textbook this is written as: $v^a = g^{ab} \omega_b$, where $\omega_b = \mu_b$. The vector v is said to have been obtained by *raising the index* of the $(0,1)$-tensor ω_b.

The canonical isomorphism $V \to V^*$ that sends v into λ^v can be readily extended to an identification between $T^{(p,q)}$ and $T^{(p-1,q+1)}$, and to an identification between $T^{(p,q)}$ and $T^{(p+1,q-1)}$:

Definition 2.4.5. Suppose T is a (p,q)-tensor, regarded as a multilinear map. The tensor obtained by lowering the i-th index, hereby denoted by $T_{(i)}$, is the $T^{(p-1,q+1)}$ tensor defined to be the multilinear map

$$T_{(i)} : V^{\otimes i-1} \otimes V^* \otimes V^{\otimes p-i} \otimes (V^*)^{\otimes q} \to \mathbb{R}$$

given by

$$T_{(i)}(\lambda^1, \dots, \lambda^{i-1}, w, \lambda^{i+1}, \dots, \lambda^p,\ \underline{v}) = T(\lambda^1, \dots, \lambda^{i-1}, \lambda^w, \lambda^{i+1}, \dots, \lambda^p,\ \underline{v}),$$

where $\underline{v}$ denotes the q-tuple $v_1, \dots, v_q$.

Similarly, the tensor $T^{(j)} : V^{\otimes p} \otimes (V^*)^{\otimes j-1} \otimes V \otimes (V^*)^{\otimes q-j} \to R$ obtained by raising the j-index is defined as:

$$T^{(j)}(\underline{\lambda},\ v_1, \dots, v_{i-1}, \mu, v_{i+1}, \dots, v_q) = T(\underline{\lambda},\ v_1, \dots, v_{i-1}, u, v_{i+1}, \dots, v_q),$$

where u is the unique vector satisfying $\lambda^u = \mu$, and $\underline{\lambda}$ denotes the p-tuple $\lambda^1, \dots, \lambda^p$

The coefficients of $T_{(i)}$, written in a fixed basis, will be denoted by $T^{a_1 \dots a_{i-1}}{}_{a_i}{}^{a_{i+1} \dots a_p}{}_{b_1 \dots b_q}$. One can readily verify that

$$T^{a_1, \dots, a_{i-1}}{}_{a_i}{}^{a_{i+1}, \dots, a_p}{}_{b_1 \dots b_q} = \textstyle\sum_s g_{s a_i} T^{a_1 \dots a_{i-1}\ s\ a_i \dots a_p}{}_{b_1 \dots b_q}$$

The coefficients of the tensor $T^{(j)}$ will be denoted by $T^{a_1, \dots, a_p}{}_{b_1 \dots b_{j-1}}{}^{b_j}{}_{b_{j+1} \dots b_q}$. Similarly, one can readily verify that

$$T^{a_1, \dots, a_p}{}_{b_1 \dots b_{j-1}}{}^{b_j}{}_{b_{j+1}, \dots, b_q} = \textstyle\sum_s g^{s b_j} T^{a_1 \dots a_p}{}_{b_1 \dots b_{j-1}\ s\ b_j \dots b_q}.$$

Example 2.4.6. Let $T^{abc}{}_{de}$ be a $(3,2)$-tensor. Then $T^a{}_b{}^{cde}$ denotes the tensor obtained by lowering and raising indices as follows:

$T^{abc}{}_d{}^e = \sum_s g^{es} T^{abc}{}_{ds}$ (raising e);
$T^{abcde} = \sum_u g^{du} T^{abc}{}_u{}^e$ (raising d);
$T^a{}_b{}^{cde} = \sum_w g_{bw} T^{awcde}$ (lowering b).

Chapter 3

Geometry

3.1 Differentiable Manifolds

Definition 3.1.1. A topological manifold of dimension n is a Hausdorff, second countable topological space which is locally isomorphic to $\mathbb{R}^n$. That is, given any point $p \in M$ there exists an open neighborhood U_p that contains p and a homeomorphism $\varphi : U_p \to V$, for some open subset $V \subseteq \mathbb{R}^n$.

Since we will be interested in using calculus over the manifold M, it is necessary to endow M with an additional smooth structure, which is defined in terms of an atlas. Let M be a topological manifold. A *chart* for M is a pair (U, φ), where U is an open set in M and $\varphi : U \to V$ is a homeomorphism onto an open set V of $\mathbb{R}^n$. The chart (U, φ) allows us to assign local coordinates to each point $p \in U$, $x^a : U \to \mathbb{R}$, defined by $x^a(p) = u^a(\varphi(p))$, where the functions u^a denote the standard coordinates in $\mathbb{R}^n$.

Definition 3.1.2. A smooth atlas for M is a family $\{(U_\alpha, \varphi_\alpha)\}_{\alpha \in A}$of charts that satisfies the following properties:

- The open sets $\{U_\alpha\}_{\alpha \in A}$ cover M.
- For any $\alpha, \beta \in A$, the change of coordinates function

  $$h_{\beta\alpha} = \varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$$

  is a smooth function.

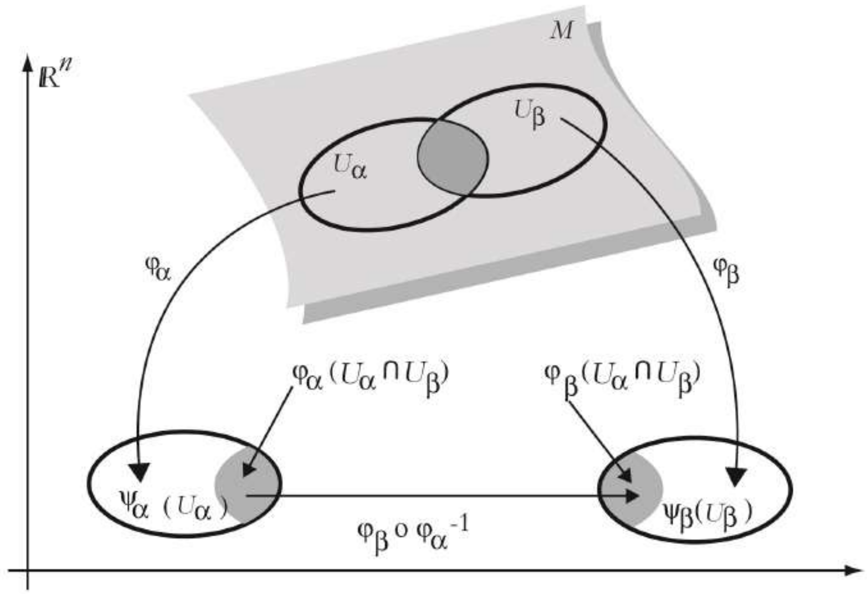


Figure 3.1: Charts

Definition 3.1.3. A smooth atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ for M is maximal if it is not properly contained in another smooth atlas for M.

Exercise 3.1.4. Show that any atlas is contained in a unique maximal atlas.

Definition 3.1.5. A manifold is a topological manifold together with a maximal atlas.

In view of Exercise 3.1.4, any atlas (not necessarily maximal) gives a topological manifold M the structure of a manifold. We will say that a chart (U, φ) is smooth if it is contained in the maximal atlas defining M.

Definition 3.1.6. A function $f : M \to \mathbb{R}$ is smooth if for any smooth chart (U, φ), the map $f \circ \varphi^{-1} : V \to \mathbb{R}$ is smooth. The space of all smooth functions $f : M \to \mathbb{R}$ is denoted by $C^\infty(M)$ an has the structure of a commutative ring with respect to pointwise multiplication.

Definition 3.1.7. Let M, N be smooth manifolds. A function $f : M \to N$ is called smooth if for each $p \in M$ there exist charts (U, φ) and (W, ψ) around p and $f(p)$, respectively, such that $f(U) \subseteq W$, and $\psi \circ f \circ \varphi^{-1} : \varphi(U) \to \psi(W)$ is smooth .

The function $\psi \circ f \circ \varphi^{-1}$ is called a *local representation* of f with respect to the charts (U, φ) and (V, ψ). It is a simple exercise to show that the composition of smooth functions is smooth and that the identity function is smooth. We conclude that manifolds together with their smooth functions are naturally organized into a category.

Definition 3.1.8. A smooth function $f : M \to N$ is called a diffeomorphism if it is invertible and its inverse is smooth. The function f is called a local diffeomorphism if for each point $p \in M$ there is an open neighborhood U of p such that the restriction of f to U is a diffeomorphism onto its image.

Example 3.1.9. The topological space $\mathbb{R}^n$ is a manifold of dimension n with respect to the atlas given by the identity map $(\mathbb{R}^n, \mathsf{id})$.

Example 3.1.10. The sphere of dimension n, denoted S^n, is the topological subspace of $\mathbb{R}^{n+1}$:

$$S^n = \{(x^1, \dots, x^{n+1}) \in \mathbb{R}^{n+1} : (x^1)^2 + \cdots + (x^{n+1})^2 = 1\}.$$

S^n inherits the topology from $\mathbb{R}^{n+1}$ and becomes a Hausdorff second countable space. We may endow S^n with a smooth structure by means of the stereographic projections. Let

$$N = (0, \dots, 0, 1)\,,\ S = (0, \dots, 0, -1)$$

be the north and south poles of the sphere, and set

$$U_S = S^n - \{N\}\,,\ U_N = S^n - \{S\}\,.$$

Define $\varphi_S : U_S \to \mathbb{R}^n$ and $\varphi_N : U_N \to \mathbb{R}^n$ by:

$$\varphi_N\left(x^1, \dots, x^{n+1}\right) = \left(\frac{x^1}{1 - x^{n+1}}, \dots, \frac{x^n}{1 - x^{n+1}}\right),$$

and

$$\varphi_S\left(x^1, \dots, x^{n+1}\right) = \left(\frac{x^1}{1 + x^{n+1}}, \dots, \frac{x^n}{1 + x^{n+1}}\right).$$

Geometrically, $\varphi_N\left(x^1, \dots, x^{n+1}\right)$ is the point of intersection of the straight line that passes through N and $(x^1, \dots, x^{n+1})$ with the plane $x^{n+1} = 0$. It is clear that φ_S and φ_N are continuous functions, and it can be easily proved

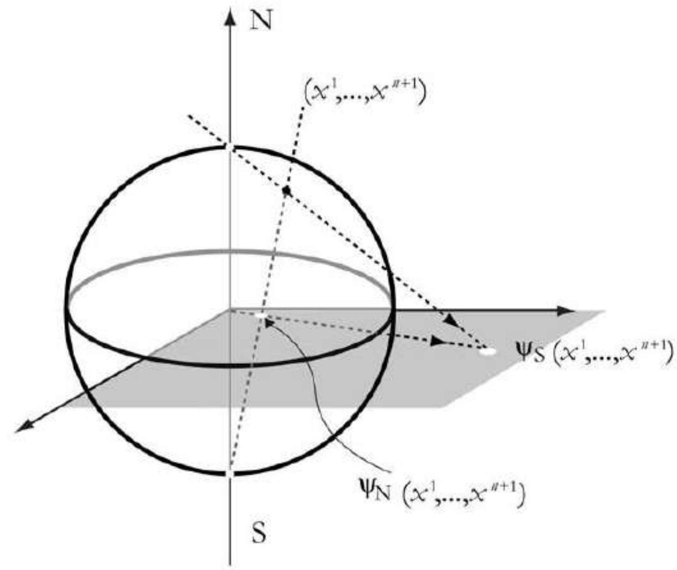


Figure 3.2: Stereographic projection

that they are bijective with continuous inverses. In fact, the inverse maps are given by

$$\varphi_S^{-1}\left(y^1,\ldots,y^n\right)=\left(1+|y|^2\right)^{-1}\left(2y^1,\ldots,2y^n,|y|^2-1\right)$$

and

$$\varphi_N^{-1}\left(y^1,\ldots,y^n\right)=\left(1+|y|^2\right)^{-1}\left(2y^1,\ldots,2y^n,1-|y|^2\right)$$

for each $y=(y^1,\ldots,y^n)\in\mathbb{R}^n$. Let us show that $\{(U_S,\varphi_S),(U_N,\varphi_N)\}$ is an atlas. Obviously, $U_S\cup U_N=S^n$. The transition map:

$$\varphi_S\circ\varphi_N^{-1}:\mathbb{R}^n-\{0\}\to\mathbb{R}^n-\{0\}$$

is given by

$$\varphi_S\circ\varphi_N^{-1}(y)=\frac{(y^1,\ldots,y^n)}{(y^1)^2+\cdots+(y^n)^2}.$$

By symmetry, the map $\varphi_N\circ\varphi_S^{-1}$ is also smooth and we conclude that S^n is a manifold.

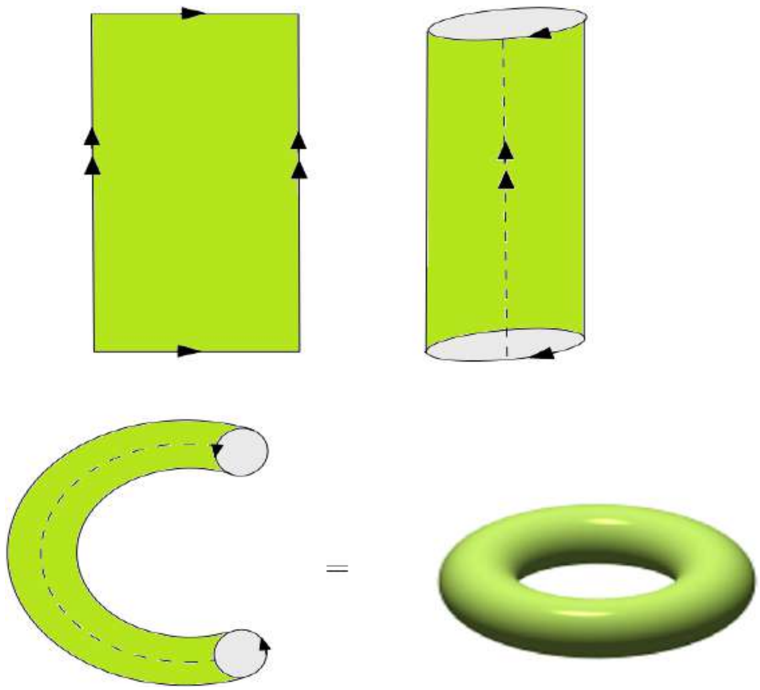


Figure 3.3: Torus

Example 3.1.11. Let M and N be smooth manifolds of dimensions m and n. The cartesian product $M \times N$ can be endowed with the product topology, and with a natural atlas $C = \{(U_\alpha \times V_\beta, \varphi_\alpha \times \psi_\beta)\}$ induced by two fixed atlases $A = \{(U_\alpha, \varphi_\alpha)\}$ and $B = \{(V_\beta, \psi_\beta)\}$, for M and N respectively.

Example 3.1.12. The additive group $\mathbb{Z}^2$ acts on $M = \mathbb{R}^2$ by translations: $(m,n) \cdot (x,y) = (x+m, y+n)$, for $m,n \in \mathbb{Z}$, and $x,y \in \mathbb{R}$. The map $f : \mathbb{R}^2 \to S^1 \times S^1$ defined by $f(x,y) = (e^{2\pi i x}, e^{2\pi i y})$ is a local diffeomorphism. Moreover $f(x,y) = f(x',y')$ if and only if $(x,y) - (x',y') \in \mathbb{Z}^2$. We conclude that f induces a homeomorphism from $\mathbb{R}^2/\mathbb{Z}^2$ to $S^1 \times S^1$.

Exercise 3.1.13. The Mobius strip M is topological space defined as the quotient $\mathbb{R}^2/G$, where G denotes the subgroup of diffeomorphism generated by the map $\alpha(x,y) = (x+1, -y)$. Let $\pi : \mathbb{R}^2 \to M$ be the canonical map to the quotient. Show that M admits a unique smooth structure such that π is a local diffeomorphism.

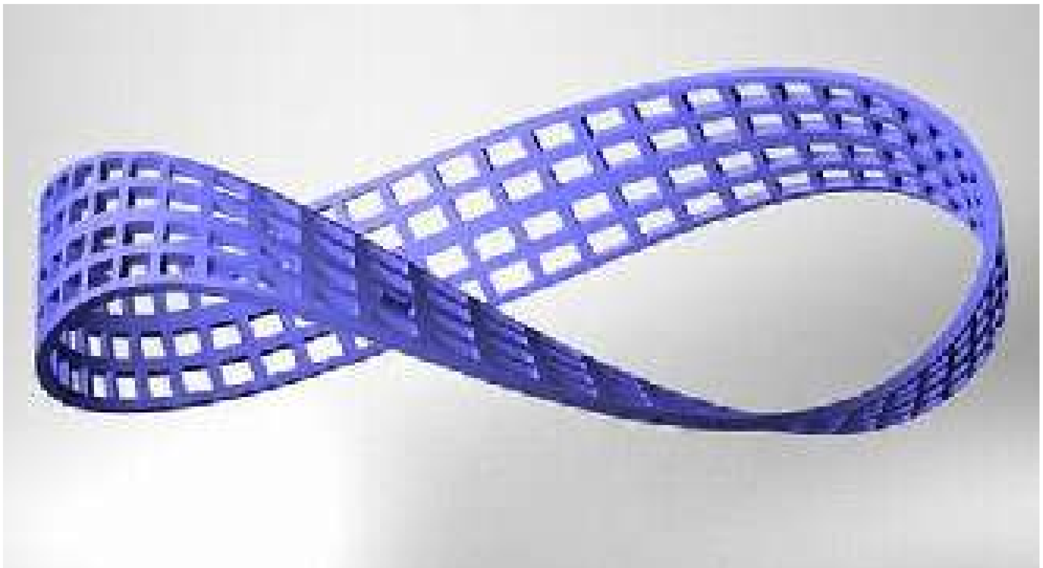


Figure 3.4: Moebius strip

Example 3.1.14. The group $G = \mathbb{Z}/2\mathbb{Z}$ acts on the sphere S^2 by the antipodal map $(x, y) \mapsto (-x, -y)$. The quotient space $P = S^2/G$ is called the real projective plane.

Example 3.1.15. Let G be the group of diffeomorphisms of the plane generated by ϕ and ρ, where $\phi(x, y) = (x + 1, y)$ and $\rho(x, y) = (-x, y + 1)$. The quotient space $K = \mathbb{R}^2/G$ is called the Klein bottle. Show that K admits a unique smooth structure such that the quotient map $\pi : \mathbb{R}^2 \to K$ is a local diffeomorphism.

3.2 The Tangent Space

A smooth structure on a topological manifold M can be used to define the tangent space at each point $p \in M$. This is a fundamental construction that allows the use of the methods of calculus over M. Before discussing the general construction, let us consider an example.

Example 3.2.1. The tangent space of the sphere at $p \in S^2$ is the set of all vectors that are perpendicular to p :

$$T_pS^2 = \{v \in \mathbb{R}^3 \mid \langle v, p\rangle = 0\}.$$

Figure 3.5: Klein bottle

Note that T_pS^2 is a vector space of dimension two.

Intuitively, the tangent space at a point $p \in M$ is the space that parametrizes all the possible velocities of an object moving in M that passes through the point p.

Definition 3.2.2. Let $p \in M$ be a point in M. A curve through p is a smooth function $\gamma : I \to M$ such that $\gamma(0) = p$, where I will denote an interval (a, b) that contains 0.

There exists a natural equivalence relation on the set of all functions that pass through $p \in M$. We say that two curves γ and β are equivalent if and only if $(\varphi \circ \gamma)'(0) = (\varphi \circ \beta)'(0)$, for any choice of coordinates $\varphi : U \to V$

Definition 3.2.3. The tangent space at the point $p \in M$, denoted by T_pM, is the set of equivalence classes of curves through p.

Proposition 3.2.4. The set T_pM has a natural structure of a vector space of dimension $m = \mathsf{dim}(M)$.

Proof. Let us fix coordinates $\varphi : U \to V$. This choice determines a function:

$$F_\varphi : T_pM \to \mathbb{R}^m; \;\; \gamma \mapsto (\varphi \circ \gamma)'(0).$$

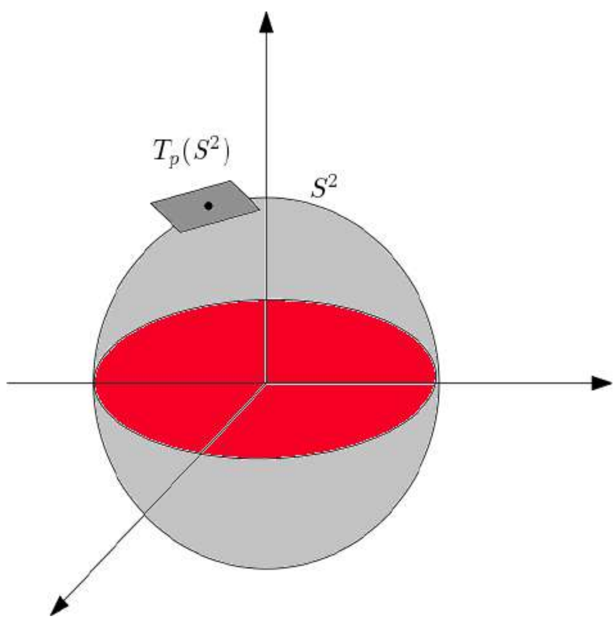


Figure 3.6: Tangent space of a sphere

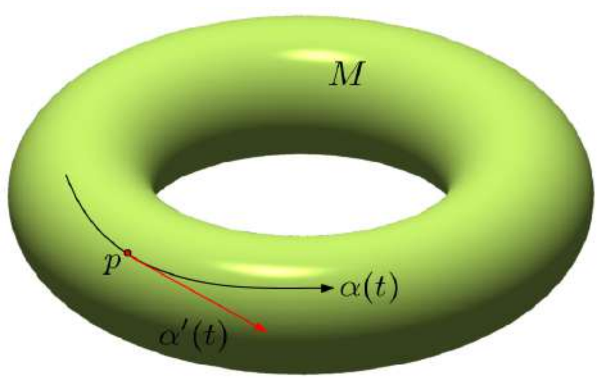


Figure 3.7: Smmoth curve and its tangent

The function F_φ is bijective, with inverse given by:

$$(F_\varphi)^{-1} : \mathbb{R}^m \to T_pM;\ v \mapsto (F_\varphi)^{-1}(v) = [\gamma]$$

where

$$\gamma(t) = \varphi^{-1}(tv + \varphi(p)).$$

This bijection gives T_pM the structure of a vector space. It remains to show that this structure is independent of the coordinates. It suffices to show that if $\phi : U \to W$ is another choice of coordinates then $F_\varphi \circ F_\phi^{-1}$ is a linear isomorphism. For this we compute:

$$F_\varphi \circ F_\phi^{-1}(v) = (\varphi \circ F_\phi^{-1}(v))'\Big|_{t=0} = (\varphi \circ \phi^{-1}(tv + \phi(p)))'\Big|_{t=0} = D(\varphi \circ \phi^{-1})(v).$$

We conclude that the vector space structure on T_pM is independent of the choice of coordinates. □

Definition 3.2.5. Let $f : M \to N$ be a smooth function. Given $p \in M$, the derivative of f at p, denoted $Df(p)$, is the lineal function

$$Df_p : T_pM \to T_{f(p)}N;\ [\gamma] \mapsto [f \circ \gamma].$$

Exercise 3.2.6. Show that the derivative of f is a well defined linear map.

Definition 3.2.7. A function $f : M \to N$ is a submersion if for all $p \in M$, $Df(p)$ is surjective. A function $f : M \to N$ is an immersion if for all $p \in M$, $Df(p)$ is injective.

Example 3.2.8. Inclusions and projections are the canonical examples of immersions and submersions:

- The function $i : \mathbb{R}^k \to \mathbb{R}^{k+m}$ defined by $(x_1, \dots, x_k) \mapsto (x_1, \dots, x_k, 0 \dots, 0)$ is an immersion.

- The function $\pi : \mathbb{R}^{k+m} \to \mathbb{R}^k$ defined by $(x_1, \dots, \dots, x_{k+m}) \mapsto (x_1, \dots, x_k)$ is a submersion.

3.3 Vector Bundles

We have already seen that if M is a manifold then for each point $p \in M$ there is a tangent vector space T_pM. This means that the tangent space construction provides a family of vector spaces parametrised by the manifold M. This situation is naturally organised in the concept of a vector bundle, which we now introduce.

Definition 3.3.1. A rank k vector bundle over M is a manifold E together with a smooth map $\pi : E \to M$ such that:

- For all $p \in M$, the set $E_p = \pi^{-1}(\{p\})$ is a vector space of dimension k.
- There exists an open cover $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ of M and diffeomorphisms
$$\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$$
such that the following diagram commutes:
$$\begin{array}{ccc} \pi^{-1}(U_\alpha) & \xrightarrow{\phi_\alpha} & U_\alpha \times \mathbb{R}^k \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle p} \\ U_\alpha & \xrightarrow{\mathsf{id}} & U_\alpha, \end{array}$$
Here p denotes the natural projection.
- The restriction of ϕ_α to each fiber is a linear isomorphism, that is, the function
$$\phi_\alpha\Big|_{\pi^{-1}\{p\}} : \pi^{-1}\{p\} \to \{p\} \times \mathbb{R}^k,$$
is a linear isomorphism.

Given a vector bundle $\pi : E \to M$, the vector space $\pi^{-1}\{p\}$ is usually denoted by E_p and called the fiber over p.

Example 3.3.2. The manifold $M \times \mathbb{R}^k$ together with the natural projection is a vector bundle over M called the trivial vector bundle.

Exercise 3.3.3. Let E be the Möbius strip, regarded as the quotient space $E = [0,1] \times \mathbb{R}/\sim$, where one identifies $(0, v)$ with $(1, -v)$. Consider the map $\pi : E \to S^1 \subseteq \mathbb{C}$ given by $(t, r) \mapsto e^{2\pi i t}$. Show that $\pi : E \to S^1$ is a vector bundle over the circle.

Definition 3.3.4. Let $\pi : E \to M$ be a vector bundle. A section of E is a smooth map $\alpha : M \to E$ such that $\pi \circ \alpha = \mathsf{id}_M$. We will denote by $\Gamma(E)$ the set of all sections of E.

Observation 1. The set $\Gamma(E)$ has the structure of a module over the ring $C^\infty(M)$ with respect to the following operations:

- $(f * \alpha)(p) = f(p)\alpha(p) \in E_p$,
- $(\alpha + \beta)(p) = \alpha(p) + \beta(p)$.

Definition 3.3.5. Let $\pi : E \to M$ and $\pi : E' \to M$ be vector bundles over M. An isomorphism from E to E' is a diffeomorphism $\varphi : E \to E'$ such that:

- The following diagram commutes:

$$\begin{array}{ccc} E & \xrightarrow{\varphi} & E' \\ \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{\mathsf{id}} & M \end{array}$$

- The map $\varphi|_{E_p} : E_p \to E'_p$ is a linear isomorphism.

Clearly, if φ is an isomorphism from E to E' then φ^{-1} is an isomorphism from E' to E. We will say that E and E' are isomorphic if there is an isomorphism between them.

Let $\pi : E \to M$ be a vector bundle and $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ an open cover of M, such that for each α there are local trivializations

$$\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k.$$

For each pair of indices α, β, there are isomorphisms:

$$\phi_\beta \circ \phi_\alpha^{-1} : U_\beta \cap U_\alpha \times \mathbb{R}^k \to U_\beta \cap U_\alpha \times \mathbb{R}^k.$$

That is, for each $p \in U_\beta \cap U_\alpha$ we obtain a linear automorphism of $\mathbb{R}^n$. This defines smooth functions:

$$f_{\beta,\alpha} : U_\beta \cap U_\alpha \to GL(n, \mathbb{R}),$$

which satisfy the following conditions:

$$f_{\alpha\alpha} = \mathsf{id},$$
$$f_{\gamma\beta} \circ f_{\beta\alpha} = f_{\gamma\alpha}.$$

It turns out that these functions determine the vector bundle E.

Definition 3.3.6. A family of cocycles is an open cover $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ together with smooth functions $f_{\beta\alpha} : U_\alpha \cap U_\beta \to GL(\mathbb{R}^k)$ such that

1. $f_{\alpha\alpha} = \mathsf{id}$,
2. $f_{\gamma\beta} \circ f_{\beta\alpha} = f_{\gamma\alpha}$.

A family of cocycles $f_{\beta\alpha}$ determines a vector bundle E as follows. As a set, one defines the total space as the disjoint union of the sets

$$\coprod_\alpha U_\alpha \times \mathbb{R}^k,$$

modulo the equivalence relation $\sim$ generated by $(p, v) \sim (p, f_{\beta\alpha}(v))$. That is $E = \coprod_\alpha U_\alpha \times \mathbb{R}^k \,/ \sim$. The map $\pi : E \to M$ is the projection onto the first factor. The topology and the smooth structure on E are characterized by the property that for each α, the natural function $U_\alpha \times \mathbb{R}^k \to \pi^{-1}(U_\alpha)$ is a diffeomorphism.

Exercise 3.3.7. Let E be a vector bundle and $\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$ a family of local trivializations for E with corresponding cocycles $f_{\beta\alpha}$. Show that the vector bundle associated to the family of cocycles $f_{\beta\alpha}$ is naturally isomorphic to E.

The natural functors of linear algebra such as taking duals and exterior powers can be used to construct new vector bundles out of given ones, as we now explain.

Remark 3.3.8. Let E, F be vector bundles over M with local trivializations $\phi_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^k$, and $\lambda_\alpha : \pi^{-1}(U_\alpha) \to U_\alpha \times \mathbb{R}^m$, respectively. Let us denote by $f_{\beta\alpha}$ and $g_{\beta\alpha}$ the corresponding families of cocycles. Then:

- The family of cocycles $h_{\beta\alpha}$ given by

  $$h_{\beta\alpha}(p) = (f_{\alpha\beta}(p))^*$$

  defines a vector bundle E^* whose fiber over p is $(E_p)^*$.

- The family of cocycles $h_{\beta\alpha}$ given by

  $$h_{\beta\alpha}(p) = f_{\beta\alpha} \oplus g_{\beta\alpha}$$

  defines a vector bundle $E \oplus F$ whose fiber over p is $E_p \oplus F_p$.

- The family of cocycles $h_{\beta\alpha}$ given by

$$h_{\beta\alpha}(p) = f_{\beta\alpha} \otimes g_{\beta\alpha}$$

defines a vector bundle $E \otimes F$ whose fiber over p is $E_p \otimes F_p$.

- For each $k \in \mathbb{N}$, the family of cocycles $h_{\beta\alpha}$ given by

$$h_{\beta\alpha}(p) = \Lambda^k(f_{\beta\alpha})$$

defines a vector bundle $\Lambda^k(E)$ whose fiber over p is $\Lambda^k(E_p)$.

- For each $k \in \mathbb{N}$, the family of cocycles $h_{\beta\alpha}$ given by

$$h_{\beta\alpha}(p) = T^{\otimes k}(f_{\beta\alpha})$$

defines a vector bundle $T^{\otimes k}(E)$ whose fiber over p is $T^{\otimes k}(E)$.

Exercise 3.3.9. Show that the constructions defined in Remark 3.3.8 are independent of the choice of local trivializations for the original vector bundles E and F. That is, show that the resulting vector bundles E^*, $E \oplus F$, $E \otimes F$, $\Lambda^k(E)$, $T^{\otimes k}(E)$ depend, up to isomorphism, only on E and F and not in the choice of cocycles for them.

3.4 The Tangent Bundle and Vector Fields

We will now describe the fundamental example of a vector bundle, the tangent bundle. Let us show that the set

$$TM = \coprod_{p \in M} T_pM,$$

has a natural structure of a vector bundle over M. The projection $\pi : TM \to M$ is given by $\pi[\gamma] = p$ if $[\gamma] \in T_pM$. Let $\varphi : U \to V \subseteq \mathbb{R}^m$ be a coordinate chart, then φ induces a bijection

$$\psi : \pi^{-1}(U) \to U \times \mathbb{R}^m; \; \psi([\gamma]) = (\pi([\gamma]), (\varphi \circ \gamma)'(0)).$$

Let us show that there exists a unique topology on TM such that π is continuous and for any choice of coordinates φ, the bijection ψ is a homeomorphism. Since π should be continuous we know that the sets $\pi^{-1}(U)$ should be open.

Since M can be covered with open sets that are the domain of coordinate charts, it suffices to show that if $\varphi : U \to V$ and $\varphi' : U' \to V'$ are two charts then the topologies induced on $\pi^{-1}(U \cap U')$ are the same. It is enough to prove that the function:

$$\psi'^{-1} : (U \cap U') \times \mathbb{R}^m \to (U \cap U') \times \mathbb{R}^m$$

is a homeomorphism. This function is given by:

$$(p, v) \to (p, D(\varphi'^{-1})(\varphi(p))(v)).$$

Since (φ'^{-1}) is a diffeomorphism, we conclude that ψ'^{-1} is a homeomorphism and indeed a diffeomorphism which is linear in the fibers. We define an atlas on TM by declaring that the functions $\psi : \pi^{-1}(U) \to U \times \mathbb{R}^m$ are smooth. It only remains to show that TM is a Hausdorff second countable space. We will first show that it is Hausdorff. Let us take $[\gamma], [\theta] \in TM$. If $p = \pi([\gamma]) \neq \pi([\theta]) = q$ then since M is Hausdorff there exists disjoint open sets $q \in U_q, p \in U_p$ and since π is continuous, the open sets $\pi^{-1}(U_p)$ and $\pi^{-1}(U_q)$ separate $[\gamma]$ and $[\theta]$. In case $p = q$ we consider the homeomorphism $\pi^{-1}(U) \simeq U \times \mathbb{R}^m$ induced by the choice of local coordinates. Since U is Hausdorff, this shows that $[\gamma]$ and $[\theta]$ can be separated in TM. Finally, let us show that TM is second countable. Consider a countable basis $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ for M such that each element of the basis is the domain of a coordinate chart and therefore $\pi^{-1}(U_\alpha) \simeq U_\alpha \times \mathbb{R}^m$. For each α we take a countable basis $\left\{W^\alpha_\beta\right\}$ of $\pi^{-1}(U_\alpha)$ so that $\left\{W^\alpha_\beta\right\}$ is a countable basis for TM.

Definition 3.4.1. The tangent bundle of a manifold M is the vector bundle TM. A vector field over M is a section of the tangent bundle. The set of all vector fields over M is denoted by $\mathfrak{X}(M) = \Gamma(TM)$.

Notation 3.4.2. We have seen that given a chart $\varphi : U \to V \subset \mathbb{R}^n$ there exists an identification

$$D\varphi : TM\Big|_U \simeq TU \to U \times \mathbb{R}^m,$$

which induces an isomorphism at the level of sections:

$$\mathfrak{X}(U) \simeq \Gamma(U \times \mathbb{R}^m) \simeq C^\infty(M, \mathbb{R}^m).$$

It is usual to denote by $\frac{\partial}{\partial x^i}$ the vector field that corresponds to the constant function with value e_i under this isomorphism. Thus, we see that a vector field over U can be written uniquely in the form:

$$X = \sum_i f^i \frac{\partial}{\partial x^i}.$$

We will also use the following shorthand notations for vector fields in local coordinates:

$$\frac{\partial}{\partial x^i} = \partial_{x^i} = \partial_i.$$

Geometrically, a vector field is a choice of a direction of movement in each point in M. We have seen that, in general, the set $\Gamma(E)$ is a module over the ring $C^\infty(M)$. In the case $E = TM$, the space of sections has an additional algebraic structure: $\mathfrak{X}(M)$ is a Lie algebra.

Definition 3.4.3. A Lie algebra is a vector space $\mathfrak{g}$ together with a bilinear map

$$[\,,\,] : \mathfrak{g} \otimes \mathfrak{g} \to L$$

such that:

1. $[\,,\,]$ is skew symmetric, that is:

$$[a, b] + [b, a] = 0.$$

2. $[\,,\,]$ satisfies the Jacobi identity, that is:

$$[a, [b, c]] + [c, [a, b]] + [b, [c, a]] = 0.$$

A subalgebra of a Lie algebra is a vector subspace that is closed with respect to the bracket.

Example 3.4.4. If A is an associative algebra then there exists a Lie algebra, $\mathsf{Lie}(A)$, defined as follows. As a vector space $\mathsf{Lie}(A) = A$. The bracket is given by the commutator, $[a, b] = ab - ba$.

Example 3.4.5. Let V be a vector space. Then the space of endomorphisms of V, $\mathsf{End}(V)$ is an associative algebra and therefore $\mathsf{Lie}(\mathsf{End}(V))$ is a Lie algebra.

Definition 3.4.6. A derivation D of an associative algebra A is a linear function $D : A \to A$ such that $D(ab) = D(a)b + aD(b)$. We denote by $\mathsf{Der}(A)$ the space of all derivations of A.

Proposition 3.4.7. $\mathsf{Der}(A) \subseteq \mathsf{End}(A)$ is a Lie subalgebra.

Proof. Let us take $D, D' \in \mathsf{Der}(A)$ and show that $[D, D'] = DD' - D'D$ is a derivation of A. Indeed:

$$\begin{aligned}
(DD' - D'D)(ab) &= D(D'(ab)) - D'(D(ab)) \\
&= D(D'(a)b + aD'(b)) - D'(D(a)b + aD(b)) \\
&= D(D'(a))b + D'(a)D(b) + D(a)D'(b) \\
&+ aDD'(b) - D'(D(a))b - D(a)D'(b) \\
&- D'(a)D(b) - aD'(D(b)) \\
&= [D, D'](a)\, b + a[D, D'](b).
\end{aligned}$$

□

We will now show that the space $\mathfrak{X}(M)$ of vector fields on M admits an algebraic description as the space of derivations of the algebra $C^\infty(M)$.

Lema 3.4.8. Let $U \subseteq M$ be an open subset. Then there exists a unique linear map:

$$\rho : \mathsf{Der}(C^\infty(M)) \to \mathsf{Der}(C^\infty(U))$$

with the property that for any $\delta \in \mathsf{Der}(C^\infty(M))$ and $g \in C^\infty(U)$:

$$\rho(\delta)(g)(p) = \delta(\tilde{g})(p)$$

for any function $\tilde{g} \in C^\infty(M)$ which coincides with g in a neighborhood of p. Moreover, the homomorphism ρ is a morphism of Lie algebras.

Proof. First we will show that given $g \in C^\infty(U)$ and $p \in U$ there exists an open $p \in W \subseteq U$ and a function $\tilde{g} \in C^\infty(M)$ such that

$$\tilde{g}\big|_W = g\big|_W.$$

Choose an open $p \in U' \subseteq U$ and a chart $\varphi : U' \to \mathbb{R}^m$. Fix a function $\lambda \in C^\infty(\mathbb{R}^m)$ such that

$$\lambda(x) = \begin{cases} 1, & \text{if } x \in [-1, 1]^m \\ 0, & \text{if } x \notin [-2, 2]^m \end{cases}$$

and we set:

$$\tilde{g}(x) = \begin{cases} g(x)\lambda(\varphi(x)), & \text{if } x \in U' \\ 0, & \text{if } x \notin U' \end{cases}$$

The function $\tilde{g}$ is smooth and coincides with g on $W := \varphi^{-1}\big((-1,1)^m\big)$. Now we define:

$$\rho(\delta)(g)(p) := \delta(\tilde{g})(p).$$

Let us see that the definition is independent of $\tilde{g}$. It suffices to show that if $h \in C^\infty(M)$ is such that $h\big|_W = 0$ then $\delta(h)\big|_W = 0$. Fix a point $p \in W$. As before, there exists a smooth function $z \in C^\infty(M)$ such that $z(x) = 1$ if $x \notin W$ y $z(x) = 0$ in a neighbourhood of p. Then $h = zh$ and:

$$\delta(h)(p) = \delta(zh)(p) = (\delta(z)h + \delta(h)z)(p) = 0.$$

We conclude that ρ is well defined. Let us prove that it is a morphism of Lie algebras:

$$\begin{aligned} \rho([\delta,\delta'])(g)(p) &= [\delta,\delta'](\tilde{g})(p) \\ &= \delta(\delta'(\tilde{g}))(p) - \delta'(\delta(\tilde{g}))(p) \\ &= \rho(\delta)(\rho(\delta')(g))(p) - \rho(\delta')(\rho(\delta)(g))(p) \\ &= [\rho(\delta),\rho(\delta')](g)(p). \end{aligned}$$

□

Lema 3.4.9. There exists a linear map $L : \mathfrak{X}(M) \longrightarrow \mathsf{Der}(C^\infty(M))$ given by $X \mapsto L_X$, where:

$$L_X(f)(p) = \frac{d}{dt}\bigg|_{t=0} f \circ \gamma(t),$$

for a curve γ such that $X(p) = [\gamma] \in T_pM$. Moreover, L is a morphism of $C^\infty(M)$-modules.

Proof. We need to prove that the map L is well defined. It suffices to observe that

$$L_X(f)(p) = Df(X(p))$$

with

$$Df(p) : T_pM \longrightarrow T_{f(p)}\mathbb{R} \cong \mathbb{R}.$$

Let us now see that L_X is a derivation:

$$L_X(fg)(p) = \frac{d}{dt}\bigg|_{t=0} (fg)\circ\gamma(t) = \frac{d}{dt}\bigg|_{t=0} f(\gamma(t))g(\gamma(t))$$

$$= \frac{d}{dt}\bigg|_{t=0} f(\gamma(t))g(p) + f(p)g(\gamma(t)),$$

therefore
$$L_X(fg)(p) = (L_X(f)g + fL_X(g))(p).$$
In order to show that L is linear on functions we compute:
$$\begin{aligned} L(fX)(g)(p) &= Dg(p)(fX(p)) \\ &= D(g)(p)(f(p)X(p)) \\ &= f(p)Dg(p)(X(p)) \\ &= (fL_X)(g)(p). \end{aligned}$$
□

Lema 3.4.10. The homomorphism $L : \mathfrak{X}(M) \to \mathsf{Der}(C^\infty(M))$ commutes with restrictions, i.e. for any open $U \subseteq M$ the following identity holds $\rho \circ L = L \circ \rho$.

Proof. On the one hand we have:
$$L(\rho(X))(g)(p) = dg(p)(X(p)).$$
On the other hand:
$$\rho(L_X)(g)(p) = L_X(\tilde{g})(p) = d(\tilde{g})(p)(X(p)) = dg(p)(X(p)).$$
□

Lema 3.4.11. If there exists an open cover $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ of M such that $L^\alpha : \mathfrak{X}(U_\alpha) \to \mathsf{Der}(C^\infty(U_\alpha))$ is an isomorphism for all $\alpha \in \mathcal{A}$ then $L : \mathfrak{X}(M) \to \mathsf{Der}(C^\infty(M))$ is an isomorphism.

Proof. Let us prove that L is inyective. If $L_X = 0$, then $L_X\big|_{U_\alpha} = 0$ for all α. Therefore
$$L_{(X|_{U_\alpha})} = 0.$$
Since L^α is inyective we conclude that
$$X\ |_{U_\alpha} = 0;\ \forall \alpha \in \mathcal{A}.$$
This implies that $X = 0$. Let us now prove surjectivity. Consider a derivation $\delta \in \mathsf{Der}(C^\infty(M))$ and set $\delta^\alpha = \delta\big|_{U_\alpha}$. By assumption there exist vector fields X_α such that $\delta^\alpha = L_{X_\alpha}$. We define X by:
$$X(p) := X_\alpha(p),$$
for any α such that $p \in U_\alpha$. It is easy to check that X is well defined and $L_X = \delta$. □

Theorem 3.4.12. The linear map $L : \mathfrak{X}(M) \longrightarrow \mathsf{Der}(C^\infty(M))$ is an isomorphism of $C^\infty(M)$-modules.

Proof. In view of Lemma 3.4.11 it is enough to consider the case $M = \mathbb{R}^m$. We have seen that in this case any vector field can be written uniquely in the form

$$X = \sum_i f^i \frac{\partial}{\partial x^i}; \qquad f^i \in C^\infty(\mathbb{R}^m),$$

Moreover:

$$L_X(g) = \sum_i f^i \frac{\partial g}{\partial x^i}.$$

Let us show that L is injective: If $L_X = 0$, then

$$\sum_i f^i \frac{\partial g}{\partial x^i} = 0$$

for any function $g \in C^\infty(M)$. This implies that each function $f^i = 0$ and therefore $X = 0$. Let us now show that L is surjective. For a derivation $\delta \in \mathsf{Der}(C^\infty(\mathbb{R}^m))$ we want to show that $\delta = L_Y$. Note that

$$L_X(x^j) = \sum_i f^i \frac{\partial x^j}{\partial x^i} = f^j.$$

Let us set $h^i := \delta(x^i)$ and

$$Y := \sum_i h^i \frac{\partial}{\partial x^i}.$$

We claim that $L_Y = \delta$. Let us fix a function g, a point $p \in M$ and a path $\gamma(t) = (1-t)p + tx$. Using the fundamental theorem of calculus we compute:

$$\int_0^1 g(\gamma(t))' dt = g(x) - g(p),$$

therefore:

$$g(x) = g(p) + \int_0^1 (g(tx + (1-t)p))' dt.$$

Expanding the derivative we obtain:

$$\begin{aligned} g(x) &= g(p) + \sum_i \int_0^1 \frac{\partial g}{\partial x^i}(\gamma(t))(x^i - p^i)dt \\ &= g(p) + \sum_i (x^i - p^i) \int_0^1 \frac{\partial g}{\partial x^i}(\gamma(t))dt. \end{aligned}$$

Applying δ on both sides, we obtain:

$$\begin{aligned} \delta(g)(x) &= \sum_i \delta\Big((x^i - p^i) \int_0^1 \frac{\partial f}{\partial x^i}(xt + (1-t)p)dt\Big) \\ &= \sum_i h^i \int_0^1 \frac{\partial g}{\partial x^i}(\gamma(t))dt + \sum_i (x^i - p^i)\delta\Big(\int_0^1 \frac{\partial g}{\partial x^i}(\gamma(t))dt\Big). \end{aligned}$$

Finally, we evaluate at $x = p$ to obtain:

$$\delta(g)(p) = \sum_i h^i(p) \int_0^1 \frac{\partial g}{\partial x^i}(p)dt = \sum_i h^i(p)\frac{\partial g}{\partial x^i}(p) = L_Y(g)(p).$$

This completes the proof. □

Corollary 3.4.13. The isomorphism $L : \mathfrak{X}(M) \to \mathsf{Der}(C^\infty(M))$ gives the vector space $\mathfrak{X}(M)$ the structure of a Lie algebra.

The natural question arises of describing the bracket of vector fields more explicitly. This can be done as follows:

Lema 3.4.14. The bracket of vector fields in $\mathbb{R}^m$ is characterized by the following properties:

1. $\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0,$
2. $[X, fY] = f[X, Y] + L_X(f)Y.$

Proof. Given two vector fields

$$X = \sum_i a^i \frac{\partial}{\partial x^i};\ Y = \sum_j b^j \frac{\partial}{\partial x^i},$$

the conditions above imply:

$$\begin{aligned}
[X,Y] &= \sum_{i,j}\left[a^i\frac{\partial}{\partial x^i}, b^j\frac{\partial}{\partial x^j}\right] \\
&= \sum_{i,j} b^j\left[a^i\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] + a^i\frac{\partial b^j}{\partial x^i}\frac{\partial}{\partial x^j} \\
&= \sum_{i,j} -b^j\frac{\partial a^i}{\partial x^j}\frac{\partial}{\partial x^i} + a^i\frac{\partial b^j}{\partial x^i}\frac{\partial}{\partial x^j} \\
&= \sum_i\Big(\sum_j a^j\frac{\partial b_i}{\partial x^j} - b^j\frac{\partial a^i}{\partial x^j}\Big)\frac{\partial}{\partial x^i}.
\end{aligned}$$

This shows uniqueness. For existence, it suffices to show that the bracket induced by the isomorphism L satisfies the conditions above. The first condition is verified because partial derivatives commute. For the second equation we compute:

$$\begin{aligned}
[L_X, L_{fY}](g) &= L_X\circ L_{fY}(g) - L_{fY}\circ L_X(g) \\
&= L_X\circ(fL_Y(g)) - f(L_Y\circ L_X(g)) \\
&= L_X(f)L_Y(g) + fL_X\circ(L_Y(g)) - f(L_Y\circ L_X(g)) \\
&= L_X(f)L_Y(g) + f[L_X, L_Y](g).
\end{aligned}$$

□

Exercise 3.4.15. Prove directly that the formula:

$$[X,Y] = \sum_i\Big(\sum_j a^j\frac{\partial b^i}{\partial x^j} - b^j\frac{\partial a^i}{\partial x^j}\Big)\frac{\partial}{\partial x^i},$$

where $X = \sum_i a^i\frac{\partial}{\partial x^i}$, $Y = \sum_j b^j\frac{\partial}{\partial x^j}$ defines a Lie algebra structure in $\mathfrak{X}(\mathbb{R}^m)$.

We have seen that in $U \subseteq \mathbb{R}^m$ vector fields can be written in the form $X = \sum_i a^i\frac{\partial}{\partial x^i}$ with $a^i \in C^\infty(U)$. Suppose that M is a manifold with coordinates $\varphi_\alpha : U_\alpha \to V_\alpha$, $\varphi_\beta : U_\beta \to V_\beta$, denoted by $\varphi_\alpha = (x^1,\dots,x^m)$ and $\varphi_\beta = (y^1,\dots,y^m)$.

A vector field $X \in \mathfrak{X}(U_\alpha \cap U_\beta)$ can be written in two different ways:

$$X = \sum_i a^i \frac{\partial}{\partial x^i} = \sum_j b^j \frac{\partial}{\partial y^j}.$$

It is natural to ask then what the relationship is between the functions a^i and b^j. The chain rule implies:

$$\frac{\partial}{\partial x^i} = \sum_j \frac{\partial y^j}{\partial x^i} \frac{\partial}{\partial y^j}.$$

Substituting in the equality above we obtain

$$X = \sum_i a^i \frac{\partial}{\partial x^i} = \sum_i a^i \Big(\sum_j \frac{\partial y^j}{\partial x^i} \frac{\partial}{\partial y^j} \Big) = \sum_j \Big(\sum_i a^i \frac{\partial y^j}{\partial x^i} \Big) \frac{\partial}{\partial y^j}.$$

Therefore: $b^j = \sum_i a^i \dfrac{\partial y^j}{\partial x^i}$.

3.5 Vector Fields and Flows

Definition 3.5.1. A flow on M is an action of the group $\mathbb{R}$ on M, that is, a smooth function

$$\begin{array}{rcl} H : \mathbb{R} \times M & \to & M, \\ (t, p) & \mapsto & H(t, p) \end{array}$$

such that

- $H(0, p) = p$.
- $H(s + t, p) = H(s, H(t, p))$.

Definition 3.5.2. Given a flow H on M, the vector field X induced by H is

$$X(p) = \frac{d}{dt}\Big|_{t=0} H(t, p).$$

The vector field X is called the infinitesimal generator of H.

Not every vector field is the infinitesimal generator of a flow. For instance, take $M = \mathbb{R} - \{0\}$. Seen as a vector field on $\mathbb{R}$, X generates the flow:

$$H(t, x) = x + t.$$

This implies that $H(1, -1) = 0 \notin M$. In this situation one says that the solution goes to infinity in finite time. It turns out that this is the only way in which a vector field may fails to generate a flow. In general, a vector field does generate a local flow.

Definition 3.5.3. A local flow on M is an open subset $\Omega \subseteq \mathbb{R} \times M$ that contains $\{0\} \times M$ and intersects each $\mathbb{R} \times \{p\}$ in an interval, together with a map

$$\begin{aligned} H : \Omega &\to M \\ (t, p) &\mapsto H(t, p) \end{aligned}$$

such that

- $H(0, p) = p$
- $H(s + t, p) = H(s, H(t, p))$, when both sides are defined.

The infinitesimal generator of a local flow is defined in the same way as that of a flow. The Picard Lindelöf theorem discussed in 8.10 implies the following:

Proposition 3.5.4. If H and Γ are two local flows which have the same infinitesimal generator then they coincide in the intersection of their domains.

Definition 3.5.5. A local flow H generated by X is called maximal if any other local flow generated by X has domain contained in that of H.

By Proposition 3.5.4, any local flow contained in a unique maximal local flow.

Theorem 3.5.6. The function that assigns to a maximal local flow its infinitesimal generator is a bijection between maximal local flows and vector fields.

Proof. By Proposition 3.5.4 the correspondence is injective. It remains to show that any vector field generates a local flow. Since this is a local statement, it suffices to prove it for an open subset U of $\mathbb{R}^m$. Let us consider a vector field X on U. By the Picard-Lindelöf theorem, there exists an open covering $\{U_p\}_{p\in U}$ and numbers $\epsilon^p > 0$ such that for all $a \in U_p$ there exists a unique solution

$$\gamma_a : (-\epsilon^p, \epsilon^p) \longrightarrow U,$$

to the equations $\gamma_a(0) = a$ and $\dot{\gamma_a} = X(\gamma_a)$. Define H as follows:

$$\Omega := \bigcup_{p\in U}(-\epsilon^p, \epsilon^p) \times U_p;\ H(t, a) = \gamma_a(t).$$

The function H is well defined by the uniqueness part of the Picard-Lindelöf theorem. It remains to show that H is a local flow. Clearly:

$$H(0, a) = \gamma_a(0) = a.$$

It remains to show that

$$H(s + t, a) = H(s, H(t, a)).$$

Fix t, a and consider the following functions of s:

$$\eta(s) = H(s + t, a));\ \omega(s) = H(s, H(t, a)).$$

We want to show that η, ω are integral curves of X with the same initial conditions. On the one hand:

$$\eta(0) = H(t, a) = \omega(0).$$

Next, we compute the derivatives:

$$\frac{d}{ds}\omega(s) = \frac{d}{ds}H(s, H(t, a)) = \frac{d}{ds}\gamma_{H(t,a)}(s) = X(\gamma_{H(t,a)}(s)) = X(\omega(s)),$$

$$\frac{d}{ds}\eta(s) = \frac{d}{ds}H(s{+}t, a) = \frac{d}{ds}\gamma_a(s{+}t) = X(\gamma_a(t{+}a)) = X(H(s{+}t, a)) = X(\eta(s)).$$

□

Example 3.5.7. Take $M = \mathbb{R}^2$. Consider the vector fields $E, X \in \mathfrak{X}(M)$ given by

$$E = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y},$$

$$X = \frac{\partial}{\partial \theta}.$$

Recall that $x = rcos(\theta)$ y $y = rsen(\theta)$ and therefore:

$$\frac{\partial}{\partial \theta} = \frac{\partial x}{\partial \theta}\frac{\partial}{\partial x} + \frac{\partial y}{\partial \theta}\frac{\partial}{\partial y} = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}.$$

We now compute $[E, V]$:

$$[E,V] = [x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}, -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}]$$

$$= -[x\frac{\partial}{\partial x}, y\frac{\partial}{\partial x}] + [x\frac{\partial}{\partial x}, x\frac{\partial}{\partial y}] - [y\frac{\partial}{\partial y}, y\frac{\partial}{\partial x}] + [y\frac{\partial}{\partial y}, x\frac{\partial}{\partial y}]$$

$$= y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x} - x\frac{\partial}{\partial y} = 0.$$

This can also be computed as follows. One observes that $E = r\dfrac{\partial}{\partial r}$, so that

$$[V,E] = \Big[\frac{\partial}{\partial \theta}, r\frac{\partial}{\partial r}\Big] = r\Big[\frac{\partial}{\partial \theta}, \frac{\partial}{\partial r}\Big] + \frac{\partial r}{\partial \theta}\frac{\partial}{\partial r} = 0.$$

Let us consider the flows associated to V and E. Using the identification $\mathbb{R}^2 \simeq \mathbb{C}$ we set $H : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$ by $H(t, z) := e^t z$, for each $(t, z) \in R \times \mathbb{C}$. Observe that:

$$H(0,t) = e^0 z = z,$$
$$H(t, H(s, z)) = e^t e^s z = e^{t+s} z,$$

and therefore H is a flow on $\mathbb{C} \simeq \mathbb{R}^2$. Let us compute the generator of H:

$$\frac{d}{dt}\Big|_{t=0} e^t z = e^0 z = z.$$

We conclude that H is generated by E. Let us also set $\varphi : \mathbb{R} \times \mathbb{C} \to \mathbb{C}$; $\varphi(t, z) := e^{it} z$. The flow φ is generated by:

$$\frac{d}{dt}\Big|_{t=0} \varphi(t, z) = \frac{d}{dt}\Big|_{t=0} e^{it} z = iz.$$

We conclude that the vector field X generates φ. Notice that φ y H commute, that is:

$$\varphi_s H_t = H_t \varphi_s, \forall s, t \in \mathbb{R}.$$

This is not a coincidence, we will see that given two vector fields X and Y, the corresponding local flows commute precisely when $[X, Y] = 0$.

Definition 3.5.8. Let $\phi : M \to N$ be a diffeomorphism and $X \in \mathfrak{X}(M)$. The push forward of X with respect to ϕ, denoted $\phi_*(X)$, is the vector field on N defined by-:

$$\phi_*(X)(q) = Df\left(\phi^{-1}(q)\right) X\left(\phi^{-1}(q)\right).$$

Given a vector field $W \in \mathfrak{X}(N)$, we define the pull-back as follows:

$$\phi^*(W) = (\phi^{-1})_*(W) \in \mathfrak{X}(M).$$

Definition 3.5.9. Sea $\phi : M \to N$ un difeomorfismo y $X \in \mathfrak{X}(M)$. El push-forward de X con respecto a ϕ, $\phi_*(X)$, es el campo vectorial en N definido así-:

$$\phi_*(X)(q) = Df\left(\phi^{-1}(q)\right) X\left(\phi^{-1}(q)\right).$$

Given a smooth function $f \in C^\infty(M)$, the push forward of f with respecto to ϕ is the function:

$$\phi_*(f) := f \circ \phi^{-1} \in C^\infty(N).$$

For a smooth function $g \in C^\infty(N)$ we define the pull-back by:

$$\phi^*(f) = (\varphi^{-1})_*(f) = f \circ \phi \in C^\infty(M).$$

The pull-back of a function is defined for an arbitrary smooth function ϕ which is not necessarily a diffeomorphism.

Lema 3.5.10. Let $\phi : M \to N$ be a diffeomorphism, $X \in \mathfrak{X}(M)$ and $f \in C^\infty(N)$. Then:

$$\varphi_*(X)(f) = \varphi_*(X(\varphi^*(f))) \in C^\infty(N).$$

Proof. Evaluating the left hand side at $p \in N$ one obtains:

$$\begin{aligned}\phi_*(X)(f)(p) &= df(p)\left(\phi_*(X)(p)\right)\\ &= df(p) \circ D\phi\left(\phi^{-1}(p)\right) X\left(\phi^{-1}(p)\right).\end{aligned}$$

On the other hand, the right hand side at $p \in N$ is:

$$\begin{aligned}\phi_*\left(X(\phi^*(f))\right)(p) &= X\left(\phi^*(f)\right)\left(\phi^{-1}(p)\right)\\ &= D\left(\phi^*(f)\right)\left(\phi^{-1}(p)\right) V\left(\phi^{-1}(p)\right)\\ &= D(f \circ \phi)\left(\phi^{-1}(p)\right) X\left(\phi^{-1}(p)\right)\\ &= df(p) \circ D\phi\left(\phi^{-1}(p)\right) X(\phi^{-1}(p)).\end{aligned}$$

□

Definition 3.5.11. Let $\phi : M \to N$ be a diffeomorphism and $\delta \in \mathsf{Der}(C^\infty(M))$ a derivation of the algebra of functions on M. The push-forward of δ, $\phi_*(\delta) \in \mathsf{Der}(C^\infty(N))$ is given by:

$$\phi_*(\delta)(g) := \phi_*(\delta(\phi^*(g))); \quad \forall g \in C^\infty(N).$$

We will now show that the identification between vector fields and derivations is compatible with the push-forward operation.

Lema 3.5.12. Let $\phi : M \to N$ be a diffeomorphism and $X \in \mathfrak{X}(M)$. Then:

$$\phi_*(L_X) = L_{\phi_*(X)}.$$

Proof. Let us evaluate both sides of the equation on $g \in C^\infty(N)$. On the one hand:

$$\begin{aligned}\phi_*(L_X)(g)(q) &= \phi_*(L_X(\phi^*(g)))(q)\\ &= L_X(\phi^*(g))(\phi^{-1}(q))\\ &= d(\phi^*(g))(\phi^{-1}(q))(X(\phi^{-1}(q)))\\ &= d(g \circ \phi)(\phi^{-1}(q))(X(\phi^{-1}(q)))\\ &= dg(q) \circ D\phi(\phi^{-1}(q)))(X(\phi^{-1}(q))).\end{aligned}$$

On the other hand:

$$\begin{aligned}L_{\phi_*(X)}(g)(q) &= dg(q)\phi_*(X)(q)\\ &= dg(q) \circ D\phi(\phi^{-1}(q))(X(\phi^{-1}(q))).\end{aligned}$$

□

Lema 3.5.13. Let X and Y be vector fields which generate local flows H y Γ, respectively, and let $f \in C^\infty(M)$ be a smooth function. Then:

$$\frac{d}{dt}\Big|_{t=0} (H_t)^*(f)(p) = X(f)(p), \tag{3.1}$$

$$\frac{d}{dt}\Big|_{t=0} (H_t)^*(Y)(p) = [X, Y](p). \tag{3.2}$$

Proof. For the first statement we compute:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0} (H_t)^*(f)(p) &= \frac{d}{dt}\Big|_{t=0} f(H(t,p)) \\ &= df(p) \frac{d}{dt}\Big|_{t=0} H(t,p) \\ &= df(p)X(p) = X(f)(p). \end{aligned}$$

For the second statement it suffices to show that he two vector fields induce the same derivation. Take a function $g \in C^\infty(M)$ and compute:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=0} (H_t^*)(Y)(g) &= \frac{d}{dt}\Big|_{t=0} (H_t^{-1})_*(Y)(g) \\ &= \frac{d}{dt}\Big|_{t=0} (H_{-t})_*(Y)(g) \\ &= \frac{d}{dt}\Big|_{t=0} (H_{-t})_* \left(Y\left(H_{-t}^*(g)\right)\right) \\ &= \frac{d}{dt}\Big|_{t=0} Y\left(H_{-t}^*(g)\right) + \frac{d}{dt}\Big|_{t=0} (H_{-t})_*(Y(g)) \\ &= \frac{d}{dt}\Big|_{t=0} Y H_{-t}^*(g) + \frac{d}{dt}\Big|_{t=0} H_t^*(Y(g)) \\ &= -YX(g) + XY(g) = [X, Y](g). \end{aligned}$$

□

Lema 3.5.14. If H is the local flow generated by X then:

$$\frac{d}{dt}\Big|_{t=t_0} H(t,p) = X(H(t_0, p)).$$

Proof. Let us compute:

$$\begin{aligned}
\frac{d}{dt}\Big|_{t=t_0} H(t,p) &= \frac{d}{dt}\Big|_{t=t_0} H(t-t_0, H(t_0,p)) \\
&= \frac{d}{ds}\Big|_{s=0} H(s, H(t_0,p)) \\
&= X(H(t_0,p)).
\end{aligned}$$

□

Lema 3.5.15. If H is the local flow generated by X and $[X,Y]=0$, then:

$$(H_t)^*(Y) = Y.$$

Proof. It suffices to show that $\frac{d}{dt}\Big|_{t=t_0}(H_t)^*(Y)(p) = 0$. In fact:

$$\begin{aligned}
\frac{d}{dt}\Big|_{t=t_0}(H_t)^*(Y)(p) &= \frac{d}{dt}\Big|_{t=t_0}(H_{t-t_0}\circ H_{t_0})^*(Y)(p) \\
&= \frac{d}{dt}\Big|_{t=t_0}(H_{t_0})^*\left(H^*_{t-t_0}(Y)\right)(p) \\
&= \frac{d}{ds}\Big|_{s=0}(H^*_{t_0})\left(H^*_s(Y)(p)\right) \\
&= D(H_{-t_0})\frac{d}{ds}\Big|_{s=0} H^*_s(Y)(p) \\
&= D(H_{-t_0})[X,Y](p) = 0.
\end{aligned}$$

□

Theorem 3.5.16. Let H y Γ be the local flows generated by X y Y respectively. Then $[X,Y]=0$ if and only if:

$$H_t\Gamma_s = \Gamma_s H_t,$$

for all $s,t\in\mathbb{R}$ where both sides are defined.

Proof. Let us first assume that the flows commute and compute:

$$\begin{aligned}
[X,Y](p) &= \frac{d}{dt}\Big|_{t=0} H_t^*(Y)(p) \\
&= \frac{d}{dt}\Big|_{t=0} (H_{-t})_*(Y)(p) \\
&= \frac{d}{dt}\Big|_{t=0} (DH_{-t})(H_t(p))Y(H_t(p)) \\
&= \frac{d}{dt}\Big|_{t=0} (DH_{-t})(H_t(p))\frac{d}{ds}\Big|_{s=0} \Gamma_s(H_t(p)) \\
&= \frac{d}{dt}\Big|_{t=0} \frac{d}{ds}\Big|_{s=0} H_{-t}\Gamma_s H_t(p) \\
&= \frac{d}{dt}\Big|_{t=0} \frac{d}{ds}\Big|_{s=0} \Gamma_s(p) = 0.
\end{aligned}$$

Let us now consider the other direction. In the computation above we showed that:

$$[X,Y](p) = \frac{d}{dt}\Big|_{t=0} \frac{d}{ds}\Big|_{s=0} H_{-t}\Gamma_s H_t(p).$$

Let us assume that the vector fields commute. We fix p and need to show that:

$$H_{-t}\Gamma_s H_t \Gamma_{-s}(p) = p. \tag{3.3}$$

It is enough to show that:

$$\frac{d}{ds}\Big|_{s=s_0} H_{-t}\Gamma_s H_t \Gamma_{-s}(p) = 0.$$

We now compute:

$$\frac{d}{ds}\Big|_{s=s_0} H_{-t}\Gamma_s H_t \Gamma_{-s}(p) = \underbrace{\frac{d}{ds}\Big|_{s=s_0} H_{-t}\Gamma_{s_0} H_t \Gamma_{-s}(p)}_{A} + \underbrace{\frac{d}{ds}\Big|_{s=s_0} H_{-t}\Gamma_s H_t \Gamma_{-s_0}(p)}_{B}.$$

We will see that $A+B=0$. Observe that, since $[X,Y]=0$, then $(H_{-t})_*(Y) = Y$. On the one hand:

$$\begin{aligned}
B &= DH_{-t}\left(\frac{d}{ds}\Big|_{s=s_0} \Gamma_s H_t \Gamma_{-s_0}(p)\right) = DH_{-t}\left(Y(\Gamma_{s_0} H_t \Gamma_{-s_0})(p)\right) \\
&= Y\left(H_{-t}\Gamma_{s_0} H_t \Gamma_{-s_0}(p)\right).
\end{aligned}$$

Let us now compute A:

$$\begin{aligned} A &= DH_{-t} \circ D\Gamma_{s_0} \circ DH_t \frac{d}{ds}\Big|_{s=s_0} \Gamma_{-s}(p) \\ &= -DH_{-t} \circ D\Gamma_{s_0} \circ DH_t(Y(\Gamma_{-s_0}(p))) \\ &= -(H_{-t})_* \circ (\Gamma_{s_0})_* \circ (H_t)_* Y(H_{-t}\Gamma_{s_0} H_t \Gamma_{-s_0}(p)) \\ &= -Y(H_{-t}\Gamma_{s_0} H_t \Gamma_{-s_0}(p)) = -B. \end{aligned}$$

□

3.6 The Cotangent Bundle

Definition 3.6.1. The cotangent bundle of M, T^*M, is the vector bundle dual to the tangent bundle. A section of the cotangent bundle is called a one form. We will denote by $\Omega^1(M)$ the space of all differential one forms on M.

Example 3.6.2. If $f \in C^\infty(M)$ is a function the its derivative Df is a one form on M. We will also denote this one form by df.

A choice of local coordinates $\varphi : U \to V$, induces a trivialization of the cotangent bundle. If we write $\varphi = (x^1, \ldots, x^m)$ there is an isomorphism of vector bundles:

$$\psi : T^*U \to U \times \mathbb{R}^m$$

given by

$$\sum_i f_i dx^i \mapsto (f_1, \ldots, f_m).$$

This means that any one form $\eta \in \Omega^1(U)$ can be written uniquely in the form

$$\eta = \sum_i f_i dx_i.$$

Again, one would like to know how the functions f_i change for different choices of coordinates. Let $\varphi_\alpha = (x^1, \ldots, x^m)$ and $\varphi_\beta = (y^1, \ldots, y^m)$ be charts on M. Then, over $U_\alpha \cap U_\beta$:

$$dx^i = \sum_j \frac{\partial x^i}{dy^j} dy^j, \tag{3.4}$$

and therefore, if $\eta = \sum_i f_i dx^i$, then:

$$\eta = \sum_i f_i \Big(\sum_j \frac{\partial x^i}{dy_j} dy^j \Big) = \sum_j \Big(\sum_i f_i \frac{\partial x^i}{dy^j} \Big) dy^j. \tag{3.5}$$

3.7 Tensor Fields

Definition 3.7.1. A tensor field T of type (p, q) is a section of the vector bundle $(TM)^{\otimes p} \otimes (T^*M)^{\otimes q}$.

We will often call T simply a tensor instead of a tensor field. As we have seen before, the choice of local coordinates induces local trivializations on the vector bundles TM and T^*M. Therefore, in local coordinates a tensor T can be written uniquely in the form:

$$T = \sum_{(a),(b)} T^{(a)}{}_{(b)}\ \partial_{x^{a_1}} \otimes \cdots \otimes \partial_{x^{a_p}} \otimes dx^{b_1} \otimes \cdots \otimes dx^{b_q}.$$

The functions $T^{(a)}{}_{(b)} = T^{a_1 \ldots a_p}{}_{b_1 \ldots b_q}$ are called the local components of the tensor T in the coordinates $\varphi = (x^a)$. The set of all tensor fields of type (p, q) will be denoted by $T^{(p,q)}(M)$:

$$T^{(p,q)}(M) = \Gamma((TM)^{\otimes p} \otimes (T^*M)^{\otimes q}).$$

Example 3.7.2. Tensors of type $(0, 0)$ are smooth functions, tensors of type $(1, 0)$ are vector fields, and tensors of type $(0, 1)$ are 1-forms.

As usual, we would like to have a formula that describes how the local components of a tensor transform for different choices of coordinates.

Consider $\varphi = (x^a)$ and $\phi = (y^a)$ coordinates in an open set $U \subseteq M$ and a tensor T of type (p, q) on U. Then we know that T can be written:

$$T = \sum_{(a),(b)} T^{(a)}{}_{(b)} \partial_{x^{a_1}} \otimes \cdots \otimes \partial_{x^{a_p}} \otimes dx^{b_1} \otimes \cdots \otimes dx^{b_q},$$

and also in the form:

$$T = \sum_{(a)(b)} L^{(a)}{}_{(b)} \partial_{y^{a_1}} \otimes \cdots \otimes \partial_{y^{a_p}} \otimes dy^{b_1} \otimes \cdots \otimes dy^{b_q}.$$

We have seen that:

$$\partial_{x^a} = \sum_c \frac{\partial y^c}{\partial x^a} \partial_y{}^c, \text{ and } dx^a = \sum_c \frac{\partial x^a}{\partial y^c} dy^c.$$

From which we conclude that:

$$L^{(a)}{}_{(b)} = \sum_{(c)(d)} T^{(c)}{}_{(d)} \frac{\partial y^{a_1}}{\partial x^{c_1}} \cdots \frac{\partial y^{a_p}}{\partial x^{c_p}} \frac{\partial x^{d_1}}{\partial y^{b_1}} \cdots \frac{\partial x^{d_p}}{\partial y^{b_q}}. \tag{3.6}$$

Remark 3.7.3. Given a finite dimensional vector space V there is a natural isomorphism:

$$\lambda : (V^*)^{\otimes k} \cong (V^{\otimes k})^*$$

given by:

$$\lambda(\phi_1 \otimes \cdots \otimes \phi_k)(v_1 \otimes \cdots \otimes v_k) := \phi_1(v_1) \cdots \phi_k(v_k).$$

This implies that the vector bundle $T^*M^{\otimes k}$ is naturally isomorphic to $(TM^{\otimes k})^*$. Using this natural identification it is common to regard a tensor of type $(0, k)$ evaluated at $p \in M$ as a linear map from $T_pM^{\otimes k}$ to $\mathbb{R}$.

Appendix 2 contains a thorough discussion of the linear algebra involved in computing with tensor fields.

3.8 Pull-backs and Lie Derivatives

As we have seen above, diffeomorphisms act naturally on functions and vector fields. More generally, diffeomorphisms act on arbitrary tensor fields. We will now describe how this action is defined in the case of tensors of type $(0, k)$. This will be enough for our purposes. Consider a tensor $T \in \Gamma(T^*M^{\otimes k})$ of type $(0, k)$ and a diffeomorphism $\phi : M \to N$. The push forward of T along ϕ is the $(0, k)$ tensor on N given by:

$$\phi_*(T)(q)(X_1 \otimes \cdots \otimes X_k) := T(p)(D\phi^{-1}(X_1) \otimes \cdots \otimes D\phi^{-1}(X_k)),$$

for $q = \phi(p)$. The pull-back operation is defined as the push-forward by the inverse map, that is:

$$\phi^*(T_*'^{-1})(T').$$

As in the case of functions, the pull-back operation on tensors of type $(0, k)$ is defined for functions ϕ which are not necessarily diffeomorphisms.

Definition 3.8.1. Let $X \in \mathfrak{X}(M)$ be a vector field on M and T a tensor of type $(0,k)$ then the Lie derivative of T in the direction of X is defined by:

$$L_X(T) = \frac{d}{dt}\Big|_{t=0} (H_t)^*(T),$$

where H denotes the local flow determined by X.

Lema 3.8.2. Let X be vector field on M with local flow H. Given a tensor field T of type $(0,k)$, the following are equivalent:

- $L_X(T) = 0$
- $(H_t)^*(T) = T$ for all t where the local flow is defined.

Proof. Clearly, the second condition implies the first. Let us assume that $L_X(T) = 0$. It suffices to show that:

$$\frac{d}{dt}\Big|_{t=t_0} (H_t)^*(T) = 0.$$

For this we compute:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=t_0} (H_t)^*(T) &= \frac{d}{ds}\Big|_{s=0} (H_{s+t_0})^*(T) \\ &= \frac{d}{ds}\Big|_{s=0} (H_{t_0})^*(H_s)^*(T) \\ &= (H_{t_0})^* \frac{d}{ds}\Big|_{s=0} (H_s)^*(T) \\ &= (H_{t_0})^*(L_X(T)) = 0. \end{aligned}$$

□

Exercise 3.8.3. Let U be an open subset of $\mathbb{R}^m$ and consider a vector field

$$X = \sum_{i=1}^{m} f^i \frac{\partial}{\partial x_i}$$

and a tensor:

$$T = \sum_{(b)} T_{(b)} dx^{b_1} \cdots dx^{b_k}.$$

Show that:

$$L_X(T) = \sum_{(b)} L_X(T_{(b)}) dx^{b_1} \cdots dx^{b_k} + \sum_{(b)} \sum_{i=1}^{k} \sum_{j=1}^{m} T_{(b)} \frac{\partial f^{b_i}}{\partial x^j} dx^{b_1} \cdots (dx^j) \cdots dx^{b_k}$$

Chapter 4

How the Metric Determines the Geometry

4.1 The Metric Tensor

So far we have studied only topological properties of smooth manifolds which, by themselves, are flexible objects without any specific geometric structure. In order to study geometric properties such as angles, distances and volumes, additional structure is necessary: the metric.

Let V be a finite dimensional vector space. A bilinear form $g : V \otimes V \to \mathbb{R}$ is symmetric if $g(v, w) = g(w, v)$. It is non degenerated if $g(v, w) = 0$ for all $w \in V$ implies $v = 0$. As explained in Appendix 2, given a non degenerated symmetric bilinear form g there exists a basis $\{e_1, \ldots, e_k\}$ and a natural number $p \leq k$ such that:

$$g(e_i, e_j) = \begin{cases} 0 & \text{if } i \neq j; \\ 1 & \text{if } i = j \leq p \\ -1 & \text{if } i = j > p \end{cases}$$

Moreover, the number p is well defined. The signature of g is the pair of numbers (p, q) where $p + q = \mathsf{dim}(V)$.

Definition 4.1.1. A pseudo-Euclidean structure in a finite dimensional vector space V is a symmetric bilinear form:

$$g : V \otimes V \to \mathbb{R},$$

which is symmetric and non degenerated. A Euclidian structure on V is a pseudo-Euclidian structure of signature $(d, 0)$. A Lorentzian structure on V is a pseudo-Euclidian structure of signature $(d-1, 1)$.

Definition 4.1.2. A pseudo-Riemannian metric on M is a section $g \in \Gamma((TM \otimes TM)^*)$ such that for each $p \in M$ the bilinear form:

$$g(p) : T_pM \otimes T_pM \to \mathbb{R}$$

is a pseudo-Euclidian structure on T_pM.

Definition 4.1.3. A pseudo-Riemannian metric g is Riemannian if for all $p \in M$ the bilinear form $g(p)$ is a Euclidian structure on T_pM. A pseudo-Riemannian metric g is Lorentzian if for all $p \in M$ the bilinear form $g(p)$ is a Lorentzian structure on T_pM. A Riemannian manifold is a manifold together with a Riemannian metric. A Lorentzian manifold is a manifold together with a Lorentzian metric.

Most of differential geometry is concerned with Riemannian manifolds. However, we will focus mainly on the Lorentzian case because in general relativity, space-time is modelled by a Lorentzian manifold. The asymmetry between directions of positive and negative length squared distinguishes space and time.

In local coordinates a metric g can be written in the form

$$g = \sum_{ij} g_{ij} dx^i \otimes dx^j,$$

where the functions g_{ij} are determined by the property:

$$g_{ij}(p) = g(p)\big(\partial_{x^i}(p), \partial_{x^j}(p)\big).$$

Example 4.1.4. The standard Riemannian metric on the manifold $\mathbb{R}^m$ which gives each tangent space $T_p(\mathbb{R}^m) \cong \mathbb{R}^m$ the usual inner product, is given by

$$g = \sum_i dx^i \otimes dx^i.$$

Example 4.1.5. Minkowski spacetime is the Lorentzian manifold $\mathbb{R}^4$ with metric:

$$-dx^0 \otimes dx^0 + \textstyle\sum_i dx^i \otimes dx^i.$$

In general, tensors of type $(0,2)$ can be locally represented in coordinates $x=(x^a)$ as

$$T=\sum_{a,b} g_{ab}dx^a\otimes dx^b.$$

Given two systems of coordinates $x=(x^a)$ and $y=(y^a)$, one obtains two different expressions for T:

$$T=\sum_{a,b} g_{ab}dx^a\otimes dx^b=\sum_{a,b} h_{ab}dy^a\otimes dy^b$$

which, by formula (3.6) are related by:

$$h_{ab}=\sum_{c,d}\frac{\partial x^c}{\partial y^a}\frac{\partial x^d}{\partial y^b}g_{cd}.$$

4.2 Examples

Example 4.2.1. Consider the manifold $M=\mathbb{H}^n_+$ defined by

$$\mathbb{H}^n_+=\{(x^1,\dots,x^n)\in\mathbb{R}^n : x^n>0\},$$

with metric

$$g=\frac{1}{(x^n)^2}\sum_a dx^a\otimes dx^a.$$

This manifold is called the hyperbolic n-dimensional space.

Definition 4.2.2. Let $\iota: S\to M$ be an immersion and g be a pseudo-Riemannian metric on M. The pull-back bilinear form $\iota^*(g)\in\Gamma((TS\otimes TS)^*)$ is defined by:

$$\iota^*(g)(p)(V,W)=g(\iota(p))(D\iota(p)(V),D\iota(p)(W)),$$

for $p\in S$ and $V,W\in T_pS$. The bilinear form $\iota^*(g)(p)$ is symmetric but in general it may fail to be non degenerated.

Let us consider local coordinates $\varphi=(y^a)$ around $p\in S$ and $\phi=(x^b)$ around $\iota(p)\in M$, so that we can write $g=\sum_{c,d} g_{cd}dx^c\otimes dx^d$. The local expression of the pullback form in a neighborhood of p is:

$$\iota^*(g)=\sum_{ab}\left(\sum_{cd}\iota^*(g_{cd})\frac{\partial x^c}{\partial y^a}\frac{\partial x^d}{\partial y^b}\right)dy^a\otimes dy^b. \tag{4.1}$$

Example 4.2.3. Let $\iota : S^2 \hookrightarrow \mathbb{R}^3$ be the standard embedding of the sphere in $\mathbb{R}^3$ and $g = \sum_i dx^i \otimes dx^i$ the euclidean metric. Let S be the southern hemisphere and consider coordinates $\varphi : S \to \mathbb{R}^2$ given by:

$$\varphi(x^1, x^2, x^3) = \left(\frac{x^1}{1 - x^3}, \frac{x^2}{1 - x^3} \right)$$

which are defined by the stereographic projection from the north pole. The inverse function $\varphi^{-1} : \mathbb{R}^2 \to S$, with $(y^1, y^2) \mapsto (x^1, x^3, x^3)$ is given by:

$$\begin{aligned} x^1(y^1, y^2) &= \frac{2y^1}{(y^1)^2 + (y^2)^2 + 1}, \\ x^2(y^1, y^2) &= \frac{2y^2}{(y^1)^2 + (y^2)^2 + 1}, \\ x^3(y^1, y^2) &= \frac{(y^1)^2 + (y^2)^2 - 1}{(y^1)^2 + (y^2)^2 + 1}. \end{aligned}$$

Thus, $\iota = \varphi^{-1}$ is an embedding of $\mathbb{R}^2$ into $\mathbb{R}^3$. The Jacobian of ι is:

$$D\iota = \frac{2}{((y^1)^2 + (y^2)^2 + 1)^2} \begin{bmatrix} -(y^1)^2 + (y^2)^2 + 1 & -2y^1y^2 \\ -2y^1y^2 & (y^1)^2 - (y^2)^2 + 1 \\ 2y^1 & 2y^2 \end{bmatrix}$$

And we conclude that the pullback metric is:

$$\iota^*(g) = \frac{4}{((y^1)^2 + (y^2)^2 + 1)^2} \left(dy^1 \otimes dy^1 + dy^2 \otimes dy^2 \right).$$

Exercise 4.2.4. Show that in spherical coordinates (θ, ϕ) for a sphere S^2 of fixed radius $r > 0$ the embedding $\iota : S^2 \to \mathbb{R}^3$ takes the form:

$$\begin{aligned} x^1 &= r \operatorname{sen} \phi \cos \theta, \quad 0 < \phi < \pi \\ x^2 &= r \operatorname{sen} \phi \operatorname{sen} \theta, \quad 0 < \theta < 2\pi \\ x^3 &= r \cos \phi. \end{aligned}$$

and the metric is:

$$i^*(g) = r^2 \operatorname{sen}^2 \phi \; d\theta \otimes d\theta + r^2 d\phi \otimes d\phi. \tag{4.2}$$

Example 4.2.5. Consider an embedding $\iota : S \hookrightarrow \mathbb{R}^3$ of surface in $\mathbb{R}^3$ and let $\varphi = (y^1, y^2)$ denote local coordinates for S. In matrix notation, the induced metric $\iota^*(g)$ is given by the product $\iota^*(g) = (D\iota)^* D\iota$ where $D\iota$ is the Jacobian matrix. It is common to use the notation:

$$\iota^*(g) = \begin{bmatrix} E & F \\ F & G \end{bmatrix},$$

where:

$$E = \sum_i \left(\frac{\partial x^i}{\partial y^1} \right)^2, \; G = \sum_i \left(\frac{\partial x^i}{\partial y^2} \right)^2$$
$$F = \frac{\partial x^1}{\partial y^1} \frac{\partial x^1}{\partial y^2} + \frac{\partial x^2}{\partial y^1} \frac{\partial x^2}{\partial y^2} + \frac{\partial x^3}{\partial y^1} \frac{\partial x^3}{\partial y 2},$$

It is also common to write:

$$\iota^* g = E\, dy^1 \otimes dy^1 + F\, dy^1 \otimes dy^2 + F\, dy^2 \otimes dy^1 + G\, dy^2 \otimes dy^2.$$

Example 4.2.6. Let $N \subset \mathbb{R}^4$ be the submanifold:

$$N = \{(x^a) \in \mathbb{R}^4 : (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 = 0, \; x^0 > 0\}.$$

We fix local coordinates (t, θ, ϕ) for N such that the inclusion $\iota : N \to \mathbb{R}^4$ takes the form:

$$\begin{aligned} x^0 &= t, \\ x^1 &= t \operatorname{sen} \phi \cos \theta, \\ x^2 &= t \operatorname{sen} \phi \operatorname{sen} \theta, \\ x^3 &= t \cos \phi. \end{aligned}$$

For $t > 0$, $0 < \theta < 2\pi$, and $0 < \phi < \pi$. If g is the Minkowski metric then the induced bilinear form expressed in coordinates $(t, \theta, \phi) = (y^0, y^1, y^2)$ is

$$\iota^*(g) = \sum_{a,b} \left(g_{00} \frac{\partial x^0}{\partial y^a} \frac{\partial x^0}{\partial y^b} + g_{11} \frac{\partial x^1}{\partial y^a} \frac{\partial x^1}{\partial y^b} + g_{22} \frac{\partial x^2}{\partial y^a} \frac{\partial x^2}{\partial y^b} + g_{33} \frac{\partial x^3}{\partial y^a} \frac{\partial x^3}{\partial y^b} \right) dy^a \otimes dy^b.$$

which in this case simplifies to:

$$\begin{aligned} i^*(g) &= dt \otimes dt - dt \otimes dt + t^2 \operatorname{sen}^2 \phi \; d\theta \otimes d\theta + t^2 d\phi \otimes d\phi \\ &= t^2 d\phi \otimes d\phi + t^2 \operatorname{sen}^2 \phi \; d\theta \otimes d\theta. \end{aligned}$$

This form is degenerate and therefore does not define a metric on N.

4.3 Length of a Curve

In a pseudo-Riemannian manifold not all directions are equal: some have positive norm squared and others have negative norm squared. This asymmetry allows for the distinction between different types of curves.

Definition 4.3.1. Let (M, g) be a pseudo-Riemannian manifold. A curve $\gamma : I \to M$ is said to be spacelike if $g(\gamma'(t), \gamma'(t)) > 0$ for all $t \in I$. It is said to be timelike if $g(\gamma'(t), \gamma'(t)) < 0$.

Definition 4.3.2. The length of a spacelike curve $\gamma : [a, b] \to M$ is

$$L(\gamma) = \int_a^b |\gamma'(s)|\, ds,$$

where the norm of $\gamma'(s)$ is

$$|\gamma'(t)| = \sqrt{g_{\gamma(s)}(\gamma'(s), \gamma'(s))}.$$

The length of a timelike curve $\gamma : [a, b] \to M$ is

$$L(\gamma) = \int_a^b |\gamma'(s)|\, ds,$$

where the norm of $\gamma'(s)$ is

$$|\gamma'(t)| = \sqrt{-g_{\gamma(s)}(\gamma'(s), \gamma'(s))}.$$

Choosing local coordinates in M we obtain the following formula for the length of a spacelike curve.

$$L(\gamma) = \int_a^b \left(\sum_{a,b} g_{ab}(\gamma(s)) \, \frac{d\gamma^a}{ds} \frac{d\gamma^b}{ds} \right)^{1/2} ds.$$

Exercise 4.3.3. A reparametrization of a curve γ is a curve $\gamma \circ \sigma$, where $\sigma : [c, d] \to [a, b]$ is a diffeomeorphism. Show that the length of a curve is invariant under reparametrization i.e.: $L(\gamma) = L(\gamma \circ \sigma)$.

4.4 Isometries and Killing Vector Fields

We have discussed above how vector fields generate flows, which are actions of the group $\mathbb{R}$ by diffeomorphisms. If the manifold M is endowed with a pseudo-riemannian metric, is often interesting to consider diffeomorphisms that preserve the metric.

Definition 4.4.1. Let (M, g) and (N, h) be pseudo-Riemannian manifolds. A diffeomorphism $\varphi : M \to N$ is an isometry if the derivative map $D\varphi(p) : T_pM \to T_{\varphi(p)}N$ preserves the pseudo-Euclidian structure for all $p \in M$.

Example 4.4.2. Let $M = \mathbb{R}^m$ with the standard Riemannian metric. For any $v \in \mathbb{R}^m$, the translation map $x \mapsto x + v$ is an isometry

Example 4.4.3. Let $M = S^2$ with the standard Riemannian structure and consider a matrix $A \in GL(3, \mathbb{R})$ such that $AA^t = \mathsf{id}$. Then the map $\hat{A} : S^2 \to S^2$ given by $x \mapsto Ax$ is an isometry. Indeed, take a point $p \in S^2$ and two tangent vectors $v, w \in T_pS^2 \subset \mathbb{R}^3$. Then:

$$\langle D\hat{A}(p)(v), D\hat{A}(p)(w)\rangle = \langle Av, Aw\rangle = (Av)^t \cdot Aw = v^tA^tAw = v^tw = \langle v, w\rangle.$$

Example 4.4.4. The hyperbolic plane $\mathbb{H}_2^+$ is the manifold

$$H_2^+ := \{z = x + iy \in \mathbb{C} : y > 0\}$$

with Riemannian metric:

$$g = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy).$$

Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

be a matrix with $\mathsf{det}(A) = 1$. The map:

$$M_A : \mathbb{H}_2^+ \to \mathbb{H}_2^+ ; z \mapsto \frac{az + b}{cz + d}$$

is an isometry of the hyperbolic plane. Let us first show that $M_A(z)$ belongs to the upper half plane.

$$\begin{aligned} M_A(z) &= \frac{az + b}{cz + d} = \frac{(az + b)(c\overline{z} + d)}{|cz + d|^2} \\ &= \frac{ac|z|^2 + adz + bc\overline{z} + bd}{|cz + d|^2} \end{aligned}$$

We conclude that:

$$\mathsf{Re}(M_A(z)) = \frac{ac|z|^2 + adx + bcx + bd}{|cz+d|^2}$$

and

$$\mathsf{Im}(M_A(z)) = \frac{ady - bcy}{|cz+d|^2} = \frac{y}{|cz+d|^2}.$$

In particular $\mathsf{Im}(M_A(z)) > 0$. Let us show that:

$$M_A \circ M_B = M_{AB}.$$

Take

$$B = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix}$$

and compute:

$$M_A \circ M_B z = M_A \left(\frac{az+b}{cz+d} \right) = \frac{(aa'+bc')z + ab' + bd')}{(ca'+dc')z + cb' + dd'}.$$

On the other hand:

$$AB = \begin{bmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+dd' \end{bmatrix}$$

and therefore:

$$M_{AB} = \frac{(aa'+bc')z + ab' + bd'}{(ca'+dc')z + cb' + dd'}.$$

This implies that M_A is a diffeomorphism with inverse $M_{A^{-1}}$. It remains to show that the derivative of M_A preserves the inner product. Notice that:

$$D(M_A)(z) = \frac{1}{(cz+d)^2}.$$

Take vectors $v = \alpha + \beta i$ and $w = \alpha' + \beta' i$ in $T_z(\mathbb{H}^+_2) \cong \mathbb{C}$. In terms of the complex structure, the inner product can be computed as follows:

$$\langle v, w \rangle_z = \frac{\mathsf{Re}(v\overline{w})}{y^2} = \frac{\alpha\alpha' + \beta\beta'}{y^2}.$$

On the other hand:

$$\begin{aligned}\langle D(M_A)(z)(v), D(M_A)(z)(w)\rangle_{M_A(z)} &= \langle \frac{v}{(cz+d)^2}, \frac{w}{(cz+d)^2}\rangle_{M_A(z)} \\ &= \frac{1}{\mathsf{Im}^2(M_A(z))}\mathsf{Re}\left(\frac{v\overline{w}}{(cz+d)^2(c\overline{z}+d)^2}\right) \\ &= \frac{|cz+d|^4}{y^2}\frac{\mathsf{Re}(v\overline{w})}{|cz+d|^4} \\ &= \frac{\alpha\alpha' + \beta\beta'}{y^2}.\end{aligned}$$

One concludes that M_A is an isometry.

Example 4.4.5. Recall that Minkowski spacetime is the manifold $\mathbb{R}^4$ with the metric:

$$g = -dx^0 \otimes dx^0 + \sum_{i=1}^{3} dx^i \otimes dx^i.$$

Given a number $0 < u < 1$ we set $l = \frac{1}{\sqrt{1-u^2}}$, and define the matrix:

$$B = \begin{bmatrix} l & -lu & 0 & 0 \\ -lu & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The map $x \mapsto Bx$ is an isometry of Minkowski spacetime. Indeed, if we view the metric as a matrix:

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Then a simple computation shows that:

$$B^t g B = g.$$

Take $v, w \in T_x(\mathbb{R}^4) \cong \mathbb{R}^4$ and compute:

$$\langle DB(x)(v), DB(x)(w)\rangle = \langle Bv, Bw\rangle = v^t B^t g B w = v^t g w = \langle v, w\rangle.$$

We conclude that B is an isometry of Minkowski spacetime. These transformations are known as Lorentz boosts.

Definition 4.4.6. Let M be a Riemannian manifold with metric g. A vector field $X \in \mathfrak{X}(M)$ is called a Killing vector field if $L_X(g) = 0$.

In local coordinates where:

$$X = \sum_k \frac{\partial}{\partial x^k}$$

and

$$g = \sum_{ij} g_{ij} dx^i \otimes dx^j,$$

the Lie derivative can be computed as follows:

$$\begin{aligned}
L_X(g) &= \sum_{ij} L_X\left(g_{ij} dx^i \otimes dx^j\right) \\
&= \sum_{ij} \Big(L_X(g_{ij}) dx^i \otimes dx^j + g_{ij} L_X(dx^i) \otimes dx^j + g_{ij} dx^i \otimes L_X(dx^j) \Big) \\
&= \sum_{ij} \Big(L_X(g_{ij}) dx^i \otimes dx^j + g_{ij} d(L_X(x^i)) \otimes dx^j + g_{ij} dx^i \otimes d(L_X(x^j)) \Big) \\
&= \sum_{ijk} \Big(f_k \frac{\partial g_{ij}}{\partial x^k} dx^i \otimes dx^j + g_{ij} d(f_k \frac{\partial x^i}{\partial x^k}) \otimes dx^j + g_{ij} dx^i \otimes d(f_k \frac{\partial x^j}{\partial x^k}) \Big) \\
&= \sum_{ijk} \Big(f_k \frac{\partial g_{ij}}{\partial x^k} dx^i \otimes dx^j \Big) + \sum_{ij} \Big(g_{ij} df_i \otimes dx^j + g_{ij} dx^i \otimes df_j \Big) \\
&= \sum_{ijk} \Big(f_k \frac{\partial g_{ij}}{\partial x^k} dx^i \otimes dx^j + g_{ij} \frac{\partial f_i}{\partial x^k} dx^k \otimes dx^j + g_{ij} \frac{\partial f_j}{\partial x^k} dx^i \otimes dx^k \Big).
\end{aligned}$$

We conclude that:

$$L_X(g) = \sum_{ij} h_{ij} dx^i \otimes dx^j,$$

where:

$$h_{ij} = \sum_k \Big(f_k \frac{\partial g_{ij}}{\partial x^k} + g_{kj} \frac{\partial f_k}{\partial x^i} + g_{ik} \frac{\partial f_k}{\partial x^j} \Big).$$

Therefore, the equations for a vector field to be a Killing vector field are:

$$\sum_k \Big(f_k \frac{\partial g_{ij}}{\partial x^k} + g_{kj} \frac{\partial f_k}{\partial x^i} + g_{ik} \frac{\partial f_k}{\partial x^j} \Big) = 0. \tag{4.3}$$

Lema 4.4.7. Let X and Y be Killing vector fields in a Riemannian manifold M. Then $[X,Y]$ is also a Killing vector field.

Proof. The statement is a consequence of the following:

$$L_{[X,Y]}(g) = L_X(L_Y(g)) - L_Y(L_X(g)) = 0.$$

□

Proposition 4.4.8. Let X be a vector field in a Riemannian manifold and denote by H_t its local flow. The vector field X is Killing if and only if H_t is an isometry for all t.

Proof. Recall that the Lie derivative is defined by:

$$L_X(g) := \frac{d}{dt}\Big|_{t=0} (H_t)^*(g).$$

If H_t is an isometry then $(H_t)^*(g) = g$ so that $L_X(g) = 0$. For the converse let us assume that:

$$\frac{d}{dt}\Big|_{t=0} (H_t)^*(g) = 0$$

and compute:

$$\begin{aligned} \frac{d}{dt}\Big|_{t=t_0} (H_t)^*(g) &= \frac{d}{dt}\Big|_{s=0} (H_{s+t_0})^*(g) \\ &= \frac{d}{dt}\Big|_{s=0} (H_s \circ H_{t_0})^*(g) \\ &= \frac{d}{dt}\Big|_{s=0} (H_{t_0})^* \circ (H_s)^*(g) \\ &= (H_{t_0})^*\Big(\frac{d}{dt}\Big|_{s=0} (H_s)^*(g)\Big) = 0. \end{aligned}$$

From this one concludes that $(H_t)^*(g)$ is independent of t. On the other hand, $(H_0)^*(g) = g$. One concludes that H_t is an isometry for all t. □

Example 4.4.9. Let us consider the manifold $\mathbb{R}^2$ with the standard Riemannian metric. The equations for a vector field:

$$X = f_1\frac{\partial}{\partial x^1} + f_2\frac{\partial}{\partial x^2}$$

to be a Killing vector field are:

$$\frac{\partial f_1}{\partial x^1} = \frac{\partial f_2}{\partial x^2} = \frac{\partial f_1}{\partial x^2} + \frac{\partial f_2}{\partial x^1} = 0. \tag{4.4}$$

For arbitrary constants $a, b, c \in \mathbb{R}$, the vector field:

$$X = a\frac{\partial}{\partial x^1} + b\frac{\partial}{\partial x^2} + c(x^2\frac{\partial}{\partial x^1} - x^1\frac{\partial}{\partial x^2}) \tag{4.5}$$

is a Killing vector field. Let us show that these are all Killing vector fields in the plane. Differentiating Equation (4.4) with respect to x^1 one obtains:

$$0 = \frac{\partial^2 f_1}{\partial x^2 \partial x^1} + \frac{\partial^2 f_2}{\partial x^1 \partial x^1} = \frac{\partial^2 f_2}{\partial x^1 \partial x^1}.$$

This implies that f_2 is a linear function of x^1 and, by symmetry, f_1 is a linear function of x^2. Using Equation (4.4) again one sees that the vector field has the form (4.5).

Example 4.4.10. Let us consider Minkowski spacetime $\mathbb{R}^4$ with the metric:

$$g = -dx^0 \otimes dx^0 + \sum_{i=1}^{3} dx^i \otimes dx^i.$$

The equations for a vector field:

$$X = \sum_{i=0}^{3} f_i \frac{\partial}{\partial x^i}$$

to be Killing are:

$$\frac{\partial f_0}{\partial x^0} = \frac{\partial f_a}{\partial x^a} = \frac{\partial f_a}{\partial x^b} + \frac{\partial f_b}{\partial x^a} = \frac{\partial f_1}{\partial x^b} - \frac{\partial f_b}{\partial x^1} = 0, \text{for } a, b \geq 1.$$

The constant vector fields are Killing and generate translations. For $a, b \geq 1$ the vector field:

$$X = x^a\frac{\partial}{\partial x^b} - x^b\frac{\partial}{\partial x^a}$$

is Killing and generate rotations. The vector field:

$$Y = x^0\frac{\partial}{\partial x^a} + x^a\frac{\partial}{\partial x^0}$$

is Killing and generates Lorentz boosts.

Chapter 5

Connections, Parallel transport and Geodesics

5.1 Connections

Given a smooth function $f = (f^1, \dots, f^m) : M \to \mathbb{R}^m$ and a vector field $X \in \mathfrak{X}(M)$ it makes sense to consider the derivative of the function f in the direction of X:

$$X(f)(p) = (X(f^1)(p), \dots, X(f^m)(p)) = Df(p)(X(p)) \in T_{f(p)}\mathbb{R}^m = \mathbb{R}^m.$$

On the other hand, if $\alpha \in \Gamma(E)$ is a section of a vector bundle E, there is no natural way to differentiate α in the direction of a vector field. A connection on a vector bundle E is a rule that prescribes such a differentiation rule.

Definition 5.1.1. Let $\pi : E \to M$ be a vector bundle. A connection ∇ on E is a linear map:

$$\nabla : \mathfrak{X}(M) \otimes \Gamma(E) \to \Gamma(E); \quad (X, \alpha) \mapsto \nabla_X \alpha$$

such that for any smooth function $f \in C^\infty(M)$, $X \in \mathfrak{X}(M)$ and $\alpha \in \Gamma(E)$ the following two conditions are satisfied:

1. $\nabla_{fX}\alpha = f\nabla_X\alpha$.
2. $\nabla_X (f\alpha) = (X(f))\,\alpha + f\nabla_X\alpha$.

Exercise 5.1.2. Show that, in the case where $E = M \times \mathbb{R}^m$ is the trivial bundle, the directional derivative described above is a connection on E.

Exercise 5.1.3. Use partitions of unity to show that any vector bundle $\pi : E \to M$ admits a connection.

Definition 5.1.4. If E is a vector bundle with connection, we will say that a section $\alpha \in \Gamma(E)$ is covariantly constant if $\nabla_X(\alpha) = 0$ for any vector field $X \in \mathfrak{X}(M)$.

Let us now consider the case $E = TM$ and describe how a connection is expressed in local coordinates $\varphi = (x^1, \dots, x^m)$. The Christoffel symbols

$$\Gamma^k_{ij} : M \to \mathbb{R}$$

are smooth functions determined by the condition:

$$\nabla_{\partial_i}\partial_j = \sum_k \Gamma^k_{ij}\partial_k.$$

Let us see that the connection ∇ is determined by the Christoffel symbols. In fact, given vector fields $X = \sum_i a^i\partial_i$, $Y = \sum_j b^j\partial_j$, one computes:

$$\begin{aligned}
\nabla_X Y &= \sum_i a^i \nabla_{\partial_i}\left(\sum_j b^j\partial_j\right) = \sum_{i,j} a^i\nabla_{\partial_i}\left(b^j\partial_j\right) \\
&= \sum_{i,j} a^i \left(\frac{\partial b^j}{\partial x^i}\partial_j + b^j\nabla_{\partial_i}\partial_j\right) \\
&= \sum_{i,j} a^i \left(\frac{\partial b^j}{\partial x^i}\partial_j + b^j\sum_k \Gamma^k_{ij}\partial_k\right) \\
&= \sum_{i,j} a^i\frac{\partial b^j}{\partial x^i}\partial_j + \sum_{i,j} a^i b^j \sum_k \Gamma^k_{ij}\partial_k \\
&= \sum_k \left(\sum_i \frac{\partial b^k}{\partial x^i}\partial_k + \sum_{i,j}\Gamma^k_{ij}b^j a^i\right)\partial_k.
\end{aligned}$$

As we have seen before, given a vector bundle E one can construct new bundles by the usual operations of linear algebra such as taking duals and tensor products. It turns out that a connection ∇ on E induces connections on all the bundles naturally associated to E. This is the content of the following exercise.

Exercise 5.1.5. Let ∇, ∇' be connections on the vector bundles $\pi : E \to M$ and $\pi : E' \to M$, respectively. Show that:

- There is a connection on the bundle $E \oplus E'$ given by:

  $$\nabla_X(\alpha + \beta) = \nabla_X(\alpha) + \nabla_X(\beta),$$

  for $X \in \mathfrak{X}(M)$, $\alpha \in \Gamma(E)$ and $\beta \in \Gamma(E')$.

- There is a connection on the bundle $E \otimes E'$ given by:

  $$\nabla_X(\alpha \otimes \beta) = \nabla_X(\alpha) \otimes \beta + \alpha \otimes \nabla'_X(\beta),$$

  for $X \in \mathfrak{X}(M)$, $\alpha \in \Gamma(E)$ and $\phi \in \Gamma(E')$.

- There is a connection on the dual bundle E^* given by:

  $$(\nabla_X \phi)(\alpha) = X(\phi(\alpha)) - \phi(\nabla_X(\alpha)),$$

  for $X \in \mathfrak{X}(M)$, $\alpha \in \Gamma(E)$ and $\phi \in \Gamma(E^*)$.

- There exists a connection on the bundle $E^{\otimes k}$ given by:

  $$\nabla_X(\alpha_1 \otimes \cdots \otimes \alpha_k) = \sum_i \alpha_1 \otimes \cdots \otimes \nabla_X(\alpha_i) \otimes \cdots \otimes \alpha_k,$$

  for $X \in \mathfrak{X}(M)$ and $\alpha_i \in \Gamma(E)$.

- There exists a connection in the bundle $\Lambda^k(E)$ given by:

  $$\nabla_X(\alpha_1 \wedge \cdots \wedge \alpha_k) = \sum_i \alpha_1 \wedge \cdots \wedge \nabla_X(\alpha_i) \wedge \cdots \wedge \alpha_k,$$

  for $X \in \mathfrak{X}(M)$ and $\alpha_i \in \Gamma(E)$.

5.2 The Levi-Civita Connection

A pseudo-Riemannian metric g on a manifold M induces a connection, called the *Levi-Civita Connection*, on the tangent bundle TM.

Definition 5.2.1. Let ∇ be a connection on TM. The torsion of ∇ is the function

$$T : \mathfrak{X}(M) \otimes \mathfrak{X}(M) \to \mathfrak{X}(M); \quad (X, Y) \mapsto \nabla_X Y - \nabla_Y X - [X, Y].$$

It is easy to verify that given vector fields $X, Y, Z \in \mathfrak{X}(M)$, the torsion satisfies:

- Linearity with respect to functions:

$$T(fX, Y) = fT(X, Y); \quad T(X, fY) = fT(X, Y).$$

- Skewsymmetry:

$$T(X, Y) + T(Y, X) = 0.$$

Exercise 5.2.2. Let $E_1, \ldots, E_k$ be vector bundles over M and consider a section $\phi \in \Gamma(E_1^* \otimes \cdots \otimes E_k^*)$.

1. Show that ϕ determines a function:

   $$\hat{\phi} : \Gamma(E_1) \otimes \cdots \otimes \Gamma(E_k) \to C^\infty(M)$$

   given by:

   $$\hat{\phi}(\alpha_1 \otimes \cdots \otimes \alpha_k)(p) = \phi(p)(\alpha_1(p) \otimes \cdots \otimes \alpha_k(p)).$$

   which is linear with respect to functions in each variable, i.e.:

   $$\hat{\phi}(\alpha_1 \otimes \cdots \otimes f\alpha_i \otimes \cdots \otimes \alpha_k) = f\hat{\phi}(\alpha_1 \otimes \cdots \otimes \alpha_k)$$

2. Show that any function $\mu : \Gamma(E_1) \otimes \cdots \otimes \Gamma(E_k) \to C^\infty(M)$, which is linear with respect to functions in each variable is of the form $\mu = \hat{\phi}$ for a unique $\phi \in \Gamma(E_1^* \otimes \cdots \otimes E_k^*)$.

Observation 2. In view Exercise 5.2.2, we can identify the torsion with a section:

$$T \in \Omega^2(M, TM) = \Gamma(\Lambda^2(T^*M) \otimes TM),$$

defined by:

$$T(p)(v, w) = \nabla_X Y(p) - \nabla_Y X(p) - [X, Y](p),$$

for any choice of vector fields X, Y such that $X(p) = v$ and $Y(p) = w$.

Definition 5.2.3. A connection on TM is called symmetric if its torsion is zero.

Exercise 5.2.4. Show that a connection ∇ is symmetric if and only if for any choice of coordinates, the Christoffel symbols satisfy $\Gamma^k_{ij} = \Gamma^k_{ji}$.

Definition 5.2.5. A connection on a pseudo-Riemannian manifold (M, g) is compatible with the metric if g is covariantly constant i.e., $\nabla_X(g) = 0$, for all $X \in \mathfrak{X}(M)$. Here g is seen as a section of $(TM \otimes TM)^*$ which we regard as having the connection induced by ∇. The condition that g is covariantly constant is equivalent to:

$$X(g(Y,Z)) = g\left(\nabla_X Y,\ Z\right) + g\left(Y,\ \nabla_X Z\right),$$

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

Theorem 5.2.6 (Levi-Civita)**.** Let (M, g) be a pseudo-Riemannian manifold. There exists a unique torsion free connection ∇ which is compatible with the metric. Moreover, this connection satisfies:

$$\begin{aligned} g\left(Z, \nabla_Y X\right) = \frac{1}{2}&\left(Xg\left(Y,Z\right) + Yg\left(Z,X\right) - Zg\left(X,Y\right)\right. \\ &\left.-g\left(\left[X,Z\right],Y\right) - g\left(\left[Y,Z\right],X\right) - g\left(\left[X,Y\right],Z\right)\right). \end{aligned} \tag{5.1}$$

Proof. Any connection compatible with the metric satisfies:

$$\begin{aligned} Xg\left(Y,Z\right) &= g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right), \\ Yg\left(Z,X\right) &= g\left(\nabla_Y Z, X\right) + g\left(Z, \nabla_Y X\right) \\ Zg\left(X,Y\right) &= g\left(\nabla_Z X, Y\right) + g\left(X, \nabla_Z Y\right), \end{aligned}$$

Adding the first two equations, subtracting the third and using the symmetry one obtains:

$$\begin{aligned} &Xg\left(Y,Z\right) + Yg\left(Z,X\right) - Zg\left(X,Y\right) \\ &= g\left(\left[X,Z\right],Y\right) + g\left(\left[Y,Z\right],X\right) + g\left(\left[X,Y\right],Z\right) + 2g\left(Z, \nabla_Y X\right), \end{aligned}$$

which implies:

$$\begin{aligned} g\left(Z, \nabla_Y X\right) = \frac{1}{2}&\left(Xg\left(Y,Z\right) + Yg\left(Z,X\right) - Zg\left(X,Y\right)\right. \\ &\left.-g\left(\left[X,Z\right],Y\right) - g\left(\left[Y,Z\right],X\right) - g\left(\left[X,Y\right],Z\right)\right). \end{aligned}$$

Since the metric is nondegenerate, this implies uniqueness. In order to prove existence we define $\nabla_Y X$ to be the unique vector field that satisfies Equation

(5.1). In order to prove that ∇ defined in this way is a connection, the only nontrivial statement is:

$$\nabla_X(fY) = f\nabla_X Y + X(f)Y.$$

For this we compute:

$$\begin{aligned} g\left(Z, \nabla_Y\left(fX\right)\right) = \frac{1}{2}&\left(fXg\left(Y,Z\right) + Yg\left(Z,fX\right) - Zg\left(fX,Y\right)\right. \\ &\left.-g\left(\left[fX,Z\right],Y\right) - g\left(\left[Y,Z\right],fX\right) - g\left(\left[fX,Y\right],Z\right)\right). \end{aligned}$$

Using the equations

$$\begin{aligned} Yg\left(Z,fX\right) &= \left(Yf\right)g\left(Z,X\right) + fYg\left(Z,X\right), \\ Zg\left(fX,Y\right) &= \left(Zf\right)g\left(X,Y\right) + fZg\left(X,Y\right), \\ g\left(\left[fX,Z\right],Y\right) &= fg\left(\left[X,Z\right],Y\right) - \left(Zf\right)g\left(X,Y\right) \\ g\left(\left[fX,Y\right],Z\right) &= fg\left(\left[X,Y\right],Z\right) - \left(Yf\right)g\left(X,Z\right) \end{aligned}$$

we obtain:

$$\begin{aligned} g\left(Z,\nabla_Y\left(fX\right)\right) &= fg\left(Z,\nabla_Y X\right) + \frac{1}{2}\left(2\left(Yf\right)g\left(Z,X\right)\right) \\ &= g\left(Z, f\nabla_Y X + \left(Yf\right)X\right). \end{aligned}$$

We leave it as an exercise to the reader to prove that ∇ is symmetric and compatible with the metric. $\square$

Definition 5.2.7. The connection defined by Theorem 5.2.6 is called the Levi-Civita connection on (M, g).

Considering Equation (5.1) in local coordinates $\varphi = (x^1, \ldots, x^m)$ and taking $Z = \partial_k$, $Y = \partial_j$ and $X = \partial_i$, we see that

$$\sum_l \Gamma^l_{ij} g_{lk} = \frac{1}{2}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right). \tag{5.2}$$

Since the matrix $[g_{ij}]$ is invertible we can write $[g^{ij}]$ for its inverse. Then:

$$\sum_{k,l} \Gamma^l_{ij} g_{lk} g^{kn} = \frac{1}{2}\sum_k g^{kn}\left(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}\right).$$

Each of the terms in parenthesis is called the Christoffel symbol of the first kind and is denoted by:

$$\Gamma_k^{ij} = \frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k}.$$

On the other hand:

$$\sum_l \left(\Gamma_{ij}^l \sum_k g_{lk} g^{kn} \right) = \Gamma_{ij}^n.$$

Therefore:

$$\Gamma_{ij}^n = \frac{1}{2} \sum_k g^{kn} \Gamma_k^{ij} = \frac{1}{2} \sum_k g^{kn} \Big(\frac{\partial g_{jk}}{\partial x^i} + \frac{\partial g_{ki}}{\partial x^j} - \frac{\partial g_{ij}}{\partial x^k} \Big). \tag{5.3}$$

This useful formula expresses the Christoffel symbols, and therefore the connection, in terms of the metric.

Remark 5.2.8. By Exercise 5.1.5 we know that the Levi-Civita connection induces connections on the vector bundles $(TM)^{\otimes p} \otimes (T^*M)^{\otimes q}$. The metric g also induces an isomorphism:

$$g^\sharp : TM \to T^*M; \; X \mapsto g^\sharp(X) \in \Omega^1(M),$$

where $g^\sharp(X)(Y) = g(X, Y)$. Since the metric g is covariantly constant we know that:

$$g^\sharp(\nabla_Z(X))(Y) = g(\nabla_Z X, Y) = Zg(X, Y) - g(X, \nabla_Z Y) = \nabla_Z(g^\sharp(X))(Y).$$

Thus we conclude that:

$$g^\sharp(\nabla_Z(X)) = \nabla_Z(g^\sharp(X)),$$

that is, the isomorphism $g^\sharp$ preserves the connection. We also conclude that the inverse map:

$$g_\sharp = (g^\sharp)^{-1} : T^*M \to TM$$

preserves the connection.

Exercise 5.2.9. If $p, q > 0$ then for each pair $i \leq p$, $j \leq q$ there is a contraction map:

$$\tau_j^i : (TM)^{\otimes p} \otimes (T^*M)^{\otimes q} \to (TM)^{\otimes p-1} \otimes (T^*M)^{\otimes q-1}$$

given by:

$$X_1 \otimes \cdots \otimes X_p \otimes \omega^1 \otimes \cdots \otimes \omega^q \mapsto \omega_j(X_i) X_1 \otimes \cdots \otimes \hat{X}_i \otimes \cdots \otimes X_p \otimes \omega^1 \otimes \cdots \otimes \hat{\omega}^j \otimes \cdots \otimes \omega^q.$$

Show that the map τ_j^i commutes with the connection, i.e.

$$\nabla_X \circ \tau_j^i = \tau_j^i \circ \nabla_Z.$$

Exercise 5.2.10. If $p > 0$ then for each $i \leq p$ there is a map:

$$(g^\sharp)_i : (TM)^{\otimes p} \otimes (T^*M)^{\otimes q} \to (TM)^{\otimes p-1} \otimes (T^*M)^{\otimes q+1}$$

given by:

$$X_1 \otimes \cdots \otimes X_p \otimes \omega^1 \otimes \cdots \otimes \omega^q \mapsto X_1 \otimes \cdots \otimes \hat{X}_i \otimes \cdots \otimes X_p \otimes (g^\sharp)(X_i) \otimes \omega^1 \otimes \cdots \otimes \omega^q.$$

Show that the map $(g^\sharp)^i$ commutes with the connection, i.e.

$$\nabla_X \circ (g^\sharp)_i = (g^\sharp)_i \circ \nabla_Z.$$

Of course, the analogous map in the other direction using $g_\sharp$ also preserves the connection.

Notation 5.2.11. Consider a tensor

$$T = \sum_{(a)(b)} T^{(a)}{}_{(b)}\ \partial_{a_1} \otimes \cdots \otimes \partial_{a_p} \otimes dx^{b_1} \otimes \cdots \otimes dx^{b_q}.$$

In physics books it is often written

$$\nabla_X T^{a_1,\dots,a_p}_{b_1,\dots,b_q}$$

to denote the corresponding component of the tensor $\nabla_X(T)$. This should not be confused with the derivative of the function $T^{a_1,\dots,a_p}{}_{b_1,\dots,b_q}$ in the direction of the vector field X. In order to avoid any confusion we will often write

$$\nabla_X(T)^{a_1,\dots,a_p}_{b_1,\dots,b_q}$$

for the components of $\nabla_X(T)$.

Exercise 5.2.12. Show that

$$2\frac{\partial g_{ij}}{\partial x^k} = \Gamma_i^{jk} + \Gamma_j^{ik},$$

and conclude that $\Gamma_{ij}^k(p) = 0$, for all i, j, k, if and only if

$$\frac{\partial g_{ij}}{\partial x^k}(p) = 0,$$

for all i, j, k.

Example 5.2.13. Let $\varphi = (r, \theta)$ be the standard polar coordinates in $\mathbb{R}^2$, and let $x^1 = r\cos\theta$, $x^2 = r\sin\theta$ be the standard euclidean coordinates. The Jacobian matrix for the change of variables is

$$J = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$$

Hence, in polar coordinates, the Euclidian metric is given by

$$g = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}$$

Which can also be written as $g = dr \otimes dr + r^2 d\theta \otimes d\theta$. Using formula (5.3) we see that the Christoffel symbols for the standard connection ∇ on $\mathbb{R}^2$ are:

$$\Gamma_{\theta\theta}^r = -r;\ \Gamma_{r\theta}^\theta = \Gamma_{\theta r}^\theta = 1/r,$$

and all the other coefficients are all zero.

5.3 Pullback of Bundles and Connections

We will show that the space $\mathsf{Conn}(E)$ of all connections on the vector bundle E is an affine space modelled on an infinite dimensional vector space.

Observation 3. Let ∇, ∇' connections on E. Then there exists a differential form

$$\theta \in \Omega^1(M, \mathsf{End}(E)) = \Gamma(T^*M \otimes \mathsf{End}(E)),$$

defined by:

$$\theta(X, \alpha) = \nabla_X(\alpha) - \nabla'_X\alpha.$$

On the other hand, for any $\theta \in \Omega^1(M, \mathsf{End}(E))$ the expression

$$\nabla'_X\alpha = \nabla_X\alpha + \theta(X, \alpha)$$

defines a connection on E. We conclude that the space $\mathsf{Conn}(E)$ has the structure of an affine space modelled over the vector space $\Omega^1(M, \mathsf{End}(E))$.

In local coordinates $\varphi = (x^1, \ldots, x^m)$ where the bundle is trivialized with a frame of sections $\{\alpha_1, \ldots, \alpha_k\}$ there is a natural connection determined by the condition

$$\nabla'_{\partial_i}\alpha_j = 0.$$

Therefore any other connection ∇ on $E|_U$ is determined by a differential form $\theta \in \Omega^1(U, \mathsf{End}(E))$ such that $\nabla_X(\alpha) = \nabla'_X(\alpha) + \theta(X)(\alpha)$.

Exercise 5.3.1. Let $f : N \to M$ a smooth function and $\pi : E \to M$ a vector bundle. Show that the set $f^*(E) = \coprod_{p\in M} E_{f(p)}$, admits a unique structure of a vector bundle over M such that:

1. The map $\tilde{f} : f^*(E) \to E; \quad v \in E_{f(p)} \mapsto v \in E_{f(p)}$ is smooth.
2. The projection $\pi : f^*(E) \to M$ is given by $v \in E_{f(p)} \mapsto p$.
3. The diagram:

$$\begin{array}{ccc} f^*(E) & \xrightarrow{\tilde{f}} & E \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ M & \xrightarrow{f} & N \end{array}$$

   commutes and is a linear isomorphism on each fiber.

4. Prove that if $h : S \to M$ is another smooth map then there is a natural isomorphism $h^*(f^*(E)) \cong (f \circ h)^*(E)$.

The vector bundle $f^*(E)$ is called the pullback bundle along f. We will now see that a connection on E induces one on $f^*(E)$.

Proposition 5.3.2. Let ∇ be a connection on $\pi : E \to M$ and $f : N \to M$ a smooth function. Then there exists a unique connection $f^*(\nabla)$ on $f^*(E)$ such that for any $\alpha \in \Gamma(E)$, $X \in \mathfrak{X}(N)$ and $Y \in \mathfrak{X}(M)$ satisfying

$$Df(p)(X(p)) = Y(f(p))$$

the following holds:

$$f^*(\nabla)_X(f^*(\alpha))(p) = \nabla_Y(\alpha)(f(p)). \tag{5.4}$$

Proof. Since connections are local operators, it suffices to consider the case of a trivializable bundle. Let $\{\alpha_1, \dots, \alpha_k\}$ be a frame of local sections on E. This defines a new connection ∇' in E determined by:

$$\nabla'_Y \alpha_i = 0.$$

We define the connection:

$$f^*(\nabla) := \nabla'' + f^*(\theta),$$

where ∇'' is the connection on $f^*(E)$ determined by:

$$\nabla''_X(f^*(\alpha_i)) = 0,$$

and θ is the differential form:

$$\theta = \nabla - \nabla'.$$

Let us verify that the connection $f^*(\nabla)$ satisfies Equation (5.4). Since $\{\alpha_1, \dots, \alpha_k\}$ are a frame for E, it is enough to consider a section α of the form $\alpha = h\alpha_i$. Then we compute:

$$\begin{aligned} f^*(\nabla)_X(f^*(h\alpha_i))(p) &= f^*(\nabla)_X(f^*(h)f^*(\alpha_i))(p) \\ &= X(f^*(h))f^*(\alpha_i)(p) + f^*(h)f^*(\nabla_i)_X(f^*(\alpha_i))(p) \\ &= X(f^*(h))f^*(\alpha_i)(p) + h(f(p))\theta(f(p))(Df(p)(X(p), \alpha_i(f(p)) \\ &= Y(h)(\alpha_i)(p) + h(f(p))\theta(f(p))(Y(f(p)), \alpha_i(f(p))) \\ &= \nabla_Y(h\alpha_i)(f(p)). \end{aligned}$$

It remains to prove the uniqueness of $f^*(\nabla)$. Let $\tilde{\nabla}$ be other connection on $f^*(E)$ with the required properties. Since $\{f^*(\alpha_1), \dots, f^*(\alpha_k)\}$ is a frame for $f^*(E)$ it suffices to show that:

$$f^*(\nabla)_X(f^*(\alpha_i))(p) = \tilde{\nabla}_X(f^*(\alpha_i))(p).$$

This is the case because Equation (5.4) guarantees that both sides are equal to $\nabla_Y(\alpha_i)(p)$. □

Exercise 5.3.3. Show that the pullback of connections is compatible with composition of functions. That is, if ∇ is a connection on $\pi : E \to N$, and $f : M \to N$ and $h : S \to M$ are smooth functions then $(f \circ h)^*(\nabla) = h^*(f^*(\nabla))$.

5.4 Parallel Transport

Recall that we say that a section $\alpha \in \Gamma(E)$ of a vector bundle with connection is covariantly constant if $\nabla_X(\alpha) = 0$, for any vector field $X \in \mathfrak{X}(M)$. By imposing this conditions on vector bundles over an interval one obtains the notion of parallel transport along a path.

Proposition 5.4.1. Let ∇ be a connection on a vector bundle $\pi : E \to I$, where $I = [a, b]$ is an interval. Given a vector $v \in E_a$ there exists a unique covariantly constant section $\alpha \in \Gamma(E)$ such that $\alpha(a) = v$. Moreover, the function $P_a^b : E_a \to E_b$ given by $P_a^b(v) = \alpha(b)$ is a linear isomorphism. The function P_a^b is called the parallel transport of the connection ∇.

Proof. Since all vector bundles over an interval are trivializable, we may choose a frame $\{\alpha_1, \dots, \alpha_k\}$ for E. There exists a one form $\theta \in \Omega^1(I, \mathsf{End}(E))$ such that:

$$\nabla_X(\alpha_i) = \theta(X, \alpha_i).$$

Let us fix $v = \sum_i \lambda_i \alpha_i(a) \in E_a$. A section $\alpha = \sum_i f_i \alpha_i$ is covariantly constant if it satisfies the differential equation:

$$\sum_i \nabla_{\partial_t}(f_i \alpha_i) = 0,$$

which is equivalent to:

$$\sum_i \frac{\partial f_i}{\partial t}\alpha_i + f_i \theta(\partial_t, \alpha_i) = 0.$$

The Picard-Lindelöf theorem, see Appendix **??**, guarantees the existence and uniqueness of a solution of this equation. In order to show that P_a^b is linear it is enough to observe that if α and β are covariantly constant, so is $\alpha + \beta$. It remains to show that P_a^b is an isomorphism. Suppose that $v \in E_a$ is such that $P_a^b(v) = 0$. By symmetry we know that there exists a unique section $\alpha \in \Gamma(E)$ such that $\alpha(b) = 0$. This section is the zero section and we conclude that $v = 0$. □

Definition 5.4.2. Let ∇ be a connection on $\pi : E \to M$ and $\gamma : [a, b] \to M$ a smooth curve. The parallel transport along γ with respect to ∇ is the linear isomorphism:

$$P_\nabla(\gamma) : E_{\gamma(a)} \to E_{\gamma(b)}; \quad P_\nabla(\gamma)(v) = P_a^b(v),$$

where P_a^b denotes the parallel transport associated with the vector bundle $\gamma^*(E)$ over the interval $I = [a, b]$ with respect to the connection $\gamma^*(\nabla)$.

Lema 5.4.3. Let $\gamma : [a, c] \to M$ be a curve and $b \in (a, c)$. Set $\mu = \gamma|_{[a,b]}$; $\sigma = \gamma|_{[b,c]}$. Then $P_\nabla(\gamma) = P_\nabla(\sigma) \circ P_\nabla(\mu)$.

Proof. It is enough to observe that if $\alpha \in \Gamma(\gamma^*(E))$ is covariantly constant then $\alpha|_{[a,b]}$ and $\alpha|_{[b,c]}$ are also covariantly constant. □

Lema 5.4.4. Parallel transport is parametrization invariant. That is, if ∇ is a connection on $\pi : E \to M$, $\gamma : [a, b] \to M$ is a curve and $\varphi : [c, d] \to [a, b]$ is an orientation preserving diffeomorphism then $P_\nabla(\gamma) = P_\nabla(\gamma \circ \varphi)$.

Proof. In view of Exercise 5.3.3 we know that:

$$(\gamma \circ \varphi)^*(\nabla) = \varphi^*(\gamma^*(\nabla)).$$

Observe also that if $\alpha \in \Gamma(\gamma^*(E))$ is covariantly constant then $\varphi^*(\alpha) \in \Gamma(\varphi^*(\gamma^*(E))) = \Gamma((\gamma \circ \varphi)^*(E))$ is also covariantly constant. □

5.5 Geodesics

Definition 5.5.1. Let (M, g) be a pseudo-Riemannian manifold and ∇ the Levi-Civita connection. A curve $\gamma : [a, b] \to M$ is called a geodesic if the section $\gamma' \in \Gamma(\gamma^*(TM))$ is covariantly constant with respect to the connection $\gamma^*(\nabla)$.

In local coordinates $\varphi = (x^1, \ldots, x^m)$ where $\gamma = (u_1, \ldots, u_m)$ and ∇ has Christoffel symbols Γ_{ij}^k one has $\gamma'(t) = \sum_i u_i'(t)\partial_i$, and the geodesic equation takes the form:

$$\begin{aligned}
\gamma^*(\nabla)_{\partial_t}(\gamma'(t)) &= \sum_i \gamma^*(\nabla)_{\partial_t}(u_i'(t)\partial_i) \\
&= \sum_i \Big(u_i''(t)\partial_i + u_i'(t)\gamma^*(\nabla)_{\partial_t}\partial_i \Big) \\
&= \sum_i \Big(u_i''(t)\partial_i + u_i'(t) \sum_j u_j'(t)\nabla_{\partial_j}\partial_i \Big) \\
&= \sum_i \Big(u_i''(t)\partial_i + u_i'(t) \sum_{j,k} u_j'(t)\Gamma_{ij}^k\partial_k \Big).
\end{aligned}$$

We conclude that γ is a geodesic precisely when it satisfies the system of differential equations:

$$u_i''(t) + \sum_{j,k} u_j'(t)u_k'(t)\Gamma^i_{jk} = 0, \tag{5.5}$$

for $i = 1, \ldots, m$.

Example 5.5.2. On Euclidian space $\mathbb{R}^m$ the Christoffel symbols are $\Gamma^k_{ij} = 0$, and therefore the differential equation for a geodesic is just $u_i''(t) = 0$. We conclude that geodesics in euclidean space are straight lines. The same is true on Minkowski spacetime.

Theorem 5.5.3. Let ∇ be the Levi-Civita connection on TM. Given $v \in T_pM$, there exists an interval $(-\epsilon, \epsilon)$ for which there is a unique geodesic $\gamma : (-\epsilon, \epsilon) \to M$ such that $\gamma(0) = p$ and $\gamma'(0) = v$.

Proof. Let $\varphi = (x^1, \ldots, x^m)$ be local coordinates such that $\varphi(p) = 0$. We write $\gamma(t) = (u_1(t), \ldots, u_m(t))$ and want to solve the system of equations:

$$u_i''(t) + \sum_{j,k} u_j'(t)u_k'(t)\Gamma^i_{jk} = 0.$$

This is a second order ordinary differential equation. The existence and uniqueness of solutions is guaranteed by the Pickard-Lindelöf theorem discussed in Apendix **??**. □

Definition 5.5.4. Let M be a pseudo-Riemannian manifold and $\gamma : [a, b] \to M$ a curve that is either timelike or spacelike. We say that γ is parametrized by arclength if

$$\int_a^s |\gamma'(t)| dt = s - a.$$

Here, as before, $|\gamma'(t)| = \sqrt{g(\gamma'(t), \gamma'(t))}$, if the curve is spacelike, and $|\gamma'(t)| = \sqrt{-g(\gamma'(t), \gamma'(t))}$ if the curve is timelike.

Exercise 5.5.5. Show that if $\gamma : I \to M$ is a geodesic then $g(\gamma'(t), \gamma'(t))$ is a constant function. Conclude that if γ is either spacelike or timelike then it can be parametrized by arclength.

Example 5.5.6. The hyperbolic plane is the Riemannian manifold

$$\mathbb{H}^2_+ = \{(x, y) \in \mathbb{R}^2 : y > 0\},$$

with metric

$$g = \frac{dx \otimes dx + dy \otimes dy}{y^2}.$$

The components of the metric are

$$g_{11} = g_{22} = \frac{1}{y^2}, \quad g_{12} = g_{21} = 0.$$

The components of the inverse matrix are:

$$g^{11} = g_{22} = y^2; \quad g^{12} = g^{21} = 0.$$

Using Equation (5.3) we obtain:

$$\Gamma^1_{11} = \Gamma^1_{22} = \Gamma^2_{12} = 0; \quad \Gamma^1_{12} = \Gamma^2_{22} = \frac{-1}{y}; \quad \Gamma^2_{11} = \frac{1}{y}.$$

The equations for a geodesic take the form

$$\ddot{x}y = 2\dot{x}\dot{y}; \quad \ddot{y}y = \dot{y}^2 - \dot{x}^2.$$

The first of these equations is equivalent to

$$\frac{d}{dt}\Big(\frac{\dot{x}}{y^2}\Big) = 0,$$

and we conclude that

$$\dot{x} = cy^2. \tag{5.6}$$

If $c = 0$, then x is constant and one obtains geodesic that are vertical lines. In case $c \neq 0$, if we assume that the curve is parametrized by arclength, we obtain $(\dot{x}^2 + \dot{y}^2)/y^2 = 1$. Using Equation 5.6 we get:

$$\frac{dy}{dx} = \frac{\dot{y}}{\dot{x}} = \sqrt{\frac{y^2 - c^2y^4}{c^2y^4}}.$$

This implies

$$dx = cydy/\sqrt{1 - c^2y^2},$$

which has as solution

$$c(x - a) = -\sqrt{1 - c^2y^2}.$$

We conclude that geodesics in the hyperbolic plane are vertical lines, as well as half circles centered at the x axis.

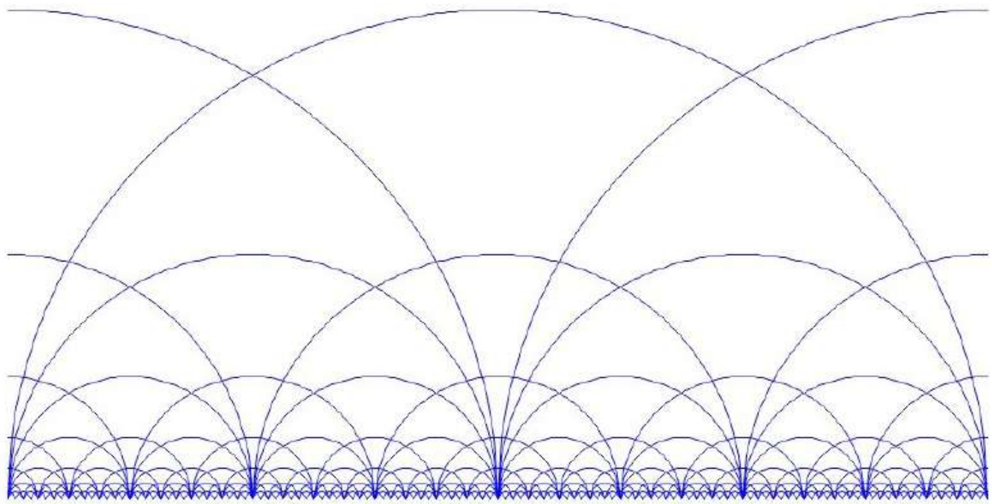


Figure 5.1: Half Plane

Exercise 5.5.7. Let $\gamma : I \to M$ be a curve. We define $\theta \in \gamma^*(T^*M))$ by the formula $\theta(s)(v) = g(\gamma'(s), v)$, for any $v \in T_{\gamma(s)}M$. Show that:

1. The curve γ is a geodesic if and only if θ is covariantly constant, i.e. $\nabla_X(\theta) = 0$.

2. In local coordinates, the condition for θ to be covariantly constant is:

$$\frac{d\theta_k}{ds} = \frac{1}{2} \sum_{i,j} \frac{\partial g_{ij}}{\partial x^k} v^i v^j. \tag{5.7}$$

   Here the functions v^l are the coefficients of $\gamma'(s)$: $\gamma'(s) = \sum_l v^l \partial_l$.

Chapter 6

Curvature

6.1 The Riemann Curvature Tensor

Definition 6.1.1. Let ∇ be a connection on a vector bundle $\pi : E \to M$. The curvature of ∇ is the function:

$$R : \mathfrak{X}(M) \otimes \mathfrak{X}(M) \otimes \Gamma(E) \to \Gamma(E)$$

Defined by:

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

It is more common to write $R\,(X, Y)\,(\alpha)$ instead of $R\,(X, Y, \alpha)$.

Exercise 6.1.2. Show that:

- The curvature is skew symmetric on X and Y.
- The curvature is linear with respect to functions in each of the variables .

Use Exercise 5.2.2 to conclude that the curvature R is a tensor:

$$R \in \Omega^2(M, \mathsf{End}(E)) = \Gamma(\Lambda^2(T^*M) \otimes \mathsf{End}(E)).$$

Proposition 6.1.3. Let ∇ be a connection on TM and $X, Y, Z \in \mathfrak{X}(M)$. If $\sum_{cyc}$ denotes the sum over cyclic permutations then:

$$\sum_{cyc} R(X,Y)Z = \sum_{cyc} (\nabla_X T)(Y,Z) + T(T(X,Y),Z).$$

$$\sum_{cyc} (\nabla_X R)(Y,Z) + R(T(X,Y),Z) = 0.$$

Here, T denotes the torsion of the connection ∇, which is defined by:

$$T(X,Y) := \nabla_X Y - \nabla_Y X - [X,Y].$$

Proof. In order to prove the first identity we observe that:

$$(\nabla_X T)(Y,Z) = \nabla_X (T(Y,Z)) - T(\nabla_X Y, Z) - T(Y, \nabla_X Z).$$

From the definition of T obtain:

$$\begin{aligned} T(T(X,Y),Z) &= T(\nabla_X Y - \nabla_Y X - [X,Y], Z) \\ &= T(\nabla_X Y, Z) + T(Z, \nabla_Y X) - T([X,Y], Z). \end{aligned}$$

Which implies:

$$\sum_{cyc} T(T(X,Y),Z) = \sum_{cyc} (\nabla_X (T(Y,Z)) - (\nabla_X T)(Y,Z) - T([X,Y],Z)).$$

Therefore:

$$\begin{aligned} &\sum_{cyc} ((\nabla_X T)(Y,Z) + T(T(X,Y),Z)) = \sum_{cyc} (\nabla_X (T(Y,Z)) - T([X,Y],Z)) \\ &= \sum_{cyc} \big(\nabla_X \nabla_Y Z - \nabla_X \nabla_Z Y - \nabla_X [Y,Z] - \nabla_{[X,Y]} Z + \nabla_Z [X,Y] + [[X,Y],Z]\big) \\ &= \sum_{cyc} \big(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z\big) = \sum_{cyc} R(X,Y)Z. \end{aligned}$$

For the second identity we compute:

$$\begin{aligned} \sum_{cyc} R(T(X,Y),Z) &= \sum_{cyc} R(\nabla_X Y - \nabla_Y X - [X,Y], Z) \\ &= \sum_{cyc} (R(\nabla_X Y, Z) + R(Z, \nabla_Y X) - R([X,Y],Z)). \end{aligned}$$

Also:

$$\sum_{cyc} (\nabla_X R)(Y, Z) = \sum_{cyc} (\nabla_X (R(Y, Z)) - R(\nabla_X Y, Z) - R(Y, \nabla_X Z) - R(Y, Z)\nabla_X).$$

Therefore:

$$\sum_{cyc} ((\nabla_X R)(Y, Z) + R(T(X, Y), Z)) = \sum_{cyc} (\nabla_X (R(Y, Z)) - R(Y, Z)\nabla_X - R([X, Y], Z))$$

$$= \sum_{cyc} \big(\nabla_X \nabla_Y \nabla_Z - \nabla_X \nabla_Z \nabla_Y - \nabla_X \nabla_{[Y,Z]} - \nabla_Y \nabla_Z \nabla_X + \nabla_Z \nabla_Y \nabla_X$$
$$+ \nabla_{[Y,Z]} \nabla_X - \nabla_{[X,Y]} \nabla_Z + \nabla_Z \nabla_{[X,Y]} + \nabla_{[[X,Y],Z]}\big) = 0.$$

□

Remark 6.1.4. Let (M, g) be a pseudo-Riemannian manifold. The curvature R of the Levi-Civita connection is called the Riemann curvature tensor. From Proposition 6.1.3 and the fact that the Levi-Civita connection is torsion free we obtain

$$R(X, Y)Z + R(Z, X)Y + R(Y, Z)X = 0 \tag{6.1}$$

and

$$(\nabla_X R)(Y, Z) + (\nabla_Z R)(X, Y) + (\nabla_Y R)(Z, X) = 0. \tag{6.2}$$

These relations are known as the first and second Bianchi identities, respectively.

Proposition 6.1.5. Let (M, g) be a pseudo-Riemannian manifold and $X, Y, Z, V \in \mathfrak{X}(M)$. Then:

$$g(R(X, Y)Z, V) + g(R(Z, X)Y, V) + g(R(Y, Z)X, V) = 0 \tag{6.3}$$
$$g(R(X, Y)Z, V) + g(R(Y, X)Z, V) = 0 \tag{6.4}$$
$$g(R(X, Y)Z, V) + g(R(X, Y)V, Z) = 0 \tag{6.5}$$
$$g(R(Z, X)Y, V) - g(R(Y, V)Z, X) = 0. \tag{6.6}$$

Proof. Property (14.2) follows from the first Bianchi identity. Equation (6.4) holds because R is skewsymmetric in the first to variables. Property (6.5) is equivalent to

$$g(R(X, Y)Z, Z) = 0.$$

Which can be proved as follows:

$$\begin{aligned}
g\left(R\left(X,Y\right)Z,Z\right) &= g\left(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, Z\right)\\
&= g\left(\nabla_X\nabla_Y Z, Z\right) - g\left(\nabla_Y\nabla_X Z, Z\right) - g\left(\nabla_{[X,Y]}Z, Z\right)\\
&= Xg\left(\nabla_Y Z, Z\right) - g\left(\nabla_Y Z, \nabla_X Z\right) - Yg\left(\nabla_X Z, Z\right)\\
&+ g\left(\nabla_X Z, \nabla_Y Z\right) - \frac{1}{2}\left[X,Y\right]g\left(Z,Z\right)\\
&= \frac{1}{2}YXg\left(Z,Z\right) - \frac{1}{2}XYg\left(Z,Z\right) - \frac{1}{2}\left[X,Y\right]g\left(Z,Z\right)\\
&= \frac{1}{2}\left(YX - YX - [X,Y]\right)g\left(Z,Z\right) = 0.
\end{aligned}$$

To prove (6.6), we observe that (14.2) implies:

$$\begin{aligned}
g\left(R\left(X,Y\right)Z,V\right) + g\left(R\left(Z,X\right)Y,V\right) + g\left(R\left(Y,Z\right)X,V\right) = 0,\\
g\left(R\left(Y,Z\right)V,X\right) + g\left(R\left(V,Y\right)Z,X\right) + g\left(R\left(Z,V\right)Y,X\right) = 0,\\
g\left(R\left(Z,V\right)X,Y\right) + g\left(R\left(X,Z\right)V,Y\right) + g\left(R\left(V,X\right)Z,Y\right) = 0,\\
g\left(R\left(V,X\right)Y,Z\right) + g\left(R\left(Y,V\right)X,Z\right) + g\left(R\left(X,Y\right)V,Z\right) = 0.
\end{aligned}$$

Adding the identities above and using (6.5), we find:

$$2g\left(R\left(Z,X\right)Y,V\right) + 2g\left(R\left(Y,V\right)X,Z\right) = 0,$$

using (6.5) again:

$$g\left(R\left(Z,X\right)Y,V\right) = g\left(R\left(Y,V\right)Z,X\right).$$

□

Given local coordinates $\varphi = (x^1, \ldots, x^m)$ we define the functions R^l_{ijk} by the property

$$R(\partial_j, \partial_k)(\partial_i) = \sum_l R^l_{ijk}\partial_l.$$

We also define $R_{lijk} = g\left(R\left(\partial_j, \partial_k\right)\partial_i, \partial_l\right)$. One can directly compute:

$$\begin{aligned}
R\left(\partial_i, \partial_j\right)\partial_k &= \nabla_{\partial_i}\nabla_{\partial_j}\partial_k - \nabla_{\partial_j}\nabla_{\partial_i}\partial_k - \nabla_{[\partial_i,\partial_j]}\partial_k\\
&= \nabla_{\partial_i}\left(\sum_l \Gamma^l_{jk}\partial_l\right) - \nabla_{\partial_j}\left(\sum_l \Gamma^l_{ik}\partial_l\right)
\end{aligned}$$

$$= \sum_l \left(\frac{\partial \Gamma^l_{jk}}{\partial x^i} \partial_l + \Gamma^l_{jk} \sum_n \Gamma^n_{il} \partial_n \right) - \sum_l \left(\frac{\partial \Gamma^l_{ik}}{\partial x^j} \partial_l + \Gamma^l_{ik} \sum_n \Gamma^n_{jl} \partial_n \right)$$

$$= \sum_n \left(\frac{\partial \Gamma^n_{jk}}{\partial x^i} - \frac{\partial \Gamma^n_{ik}}{\partial x^j} + \sum_l \Gamma^l_{jk} \Gamma^n_{il} - \sum_l \Gamma^l_{ik} \Gamma^n_{jl} \right) \partial_n.$$

We conclude that:

$$R^l_{ijk} = \frac{\partial \Gamma^l_{ik}}{\partial x^j} - \frac{\partial \Gamma^l_{ji}}{\partial x^k} + \sum_n \Gamma^l_{jn} \Gamma^n_{ik} - \sum_n \Gamma^l_{kn} \Gamma^n_{ij},$$

$$R_{nijk} = \sum_l R^l_{ijk} g_{ln}.$$

The Bianchi identities are equivalent to:

$$R^l_{ijk} + R^l_{kij} + R^l_{jki} = 0 \tag{6.7}$$

$$(\nabla_{\partial_i} R)^n_{jkl} + (\nabla_{\partial_l} R)^n_{jik} + (\nabla_{\partial_k} R)^n_{jli} = 0 \tag{6.8}$$

Example 6.1.6. Since for Euclidian space $\mathbb{R}^m$ the Cristoffel symbols vanish we conclude that $R = 0$.

Example 6.1.7. Compute the Riemann tensor for the hyperbolic plane and for the two dimensional sphere in the coordinates provided by the stereographic projection.

6.2 The Ricci tensor, scalar curvature and the Einstein tensor

Definition 6.2.1. Let (M, g) be a pseudo-Riemannian manifold. The Ricci tensor, denoted by $\mathrm{Ric} \in \Gamma((TM \otimes TM)^*)$ is the tensor defined by:

$$\mathrm{Ric}(X, Y)(p) = \mathsf{tr}\big(R(p)(X(p), -)(Y(p))\big).$$

Here $X, Y \in \mathfrak{X}(M)$ are vector fields on M and $R(p)(X(p), -)(Y(p))$ is the function from T_pM to T_pM defined by

$$Z \mapsto R(p)(X(p), Z)(Y(p)).$$

Exercise 6.2.2. Show that $\mathsf{Ric}(X,Y)=\mathsf{Ric}(Y,X)$. Prove that the functions

$$\mathsf{Ric}_{ij}=\sum_l R^l_{jli},$$

are the components of the Ricci tensor.

By raising the indices one obtains the $(0,2)$ tensor $\overline{\mathsf{Ric}}$ with components

$$\overline{\mathsf{Ric}}^{ij}=\sum_{kl} g^{ik}g^{jl}\mathsf{Ric}_{kl}.$$

Exercise 6.2.3. Prove the following identities:

1. $R_{ijkl}=-R_{jikl}=-R_{ijlk}$.
2. $R_{ijkl}=R_{klij}$.
3. $\mathsf{Ric}_{ij}=\mathsf{Ric}_{ji}$.

Definition 6.2.4. The scalar curvature of a pseudo-Riemannian manifold (M,g), denoted by $S\in C^\infty(M)$ is the function:

$$S(p)=\mathsf{tr}\Big(\mathsf{Ric}^g(p)\Big),$$

where $\mathsf{Ric}^g(p):T_pM\to T_pM$ is the linear function characterized by:

$$g(p)\Big(\mathsf{Ric}^g(p)(X),Y\Big)=\mathsf{Ric}(p)(X,Y).$$

In local coordinates the scalar curvature is given by:

$$\mathrm{S}=\sum_{ij} g^{ij}\mathsf{Ric}_{ij}.$$

Definition 6.2.5. Einstein's tensor G is the $(2,0)$ tensor defined by

$$\mathrm{G}=\mathsf{Ric}-\frac{g\mathrm{S}}{2}.$$

By raising the indices one obtains a tensor $\overline{\mathrm{G}}$ of type $(0,2)$ with components

$$\overline{\mathrm{G}}^{ij}=\sum_{kl} g^{ik}g^{jl}G_{kl}.$$

Example 6.2.6. Recall that the hyperbolic plane is the Riemannian manifold

$$\mathbb{H}^2_+ = \{(x,y) \in \mathbb{R}^2 : y > 0\},$$

with metric

$$g = \frac{dx \otimes dx + dy \otimes dy}{y^2}.$$

Compute the Riemann curvature tensor, the Ricci tensor and the scalar curvature of the hyperbolic plane.

Proposition 6.2.7. The following identities hold:

$$\sum_s (\nabla_{\partial_s} R)^s_{kjl} + (\nabla_{\partial_l} \mathsf{Ric})_{kj} - (\nabla_{\partial_j} \mathsf{Ric})_{kl} = 0, \tag{6.9}$$

$$2 \sum_s (\nabla_{\partial s} \mathsf{Ric})^s_j - \nabla_{\partial j} \mathrm{S} = 0, \tag{6.10}$$

$$\sum_s (\nabla_{\partial_s} \overline{G})^{si} = 0. \tag{6.11}$$

Proof. By Exercise 5.2.9 we know that contracting indices commutes with covariant differentiation and therefore

$$(\nabla_{\partial_i} \mathsf{Ric})_{jk} = \textstyle\sum_s (\nabla_{\partial_i} R)^s_{jsk}.$$

On the other hand, the second Bianchi identity gives:

$$(\nabla_{\partial_i} R)^s_{kjl} + (\nabla_{\partial_l} R)^s_{kij} + (\nabla_{\partial_j} R)^s_{kli} = 0.$$

Using the skew-symmetry of the Riemann tensor and summing over $i = s$ one obtains:

$$\sum_s (\nabla_{\partial_s} R)^s_{kjl} + \sum_s (\nabla_{\partial_l} R)^s_{ksj} - \sum_s (\nabla_{\partial_j} R)^s_{ksl} = 0.$$

which is precisely:

$$\sum_s (\nabla_{\partial_s} R)^s_{kjl} + (\nabla_{\partial_l} \mathsf{Ric})_{kj} - (\nabla_{\partial_j} \mathsf{Ric})_{kl} = 0,$$

as required. Let us now prove the second identity. Multiplying Equation (6.9) by g^{kr} and summing over k one obtains:

$$\sum_{s,k} (g^{kr} \nabla_{\partial_s} R)^s_{kjl} + \sum_k g^{kr} (\nabla_{\partial_l} \mathsf{Ric})_{kj} - \sum_k g^{kr} (\nabla_{\partial_j} \mathsf{Ric})_{kl} = 0.$$

This can also be written:

$$\sum_s (\nabla_{\partial_s} R)^{sk}_{jl} + (\nabla_{\partial_l} \mathsf{Ric})^k_j - (\nabla_{\partial_j} \mathsf{Ric})^k_l = 0.$$

We now contract the indices j and k to obtain:

$$\nabla_{\partial_l}(S) - 2\sum_s (\nabla_{\partial_s} \mathsf{Ric})^s_l,$$

which is equivalent to (6.10). Finally, in order to prove (6.11) we multiply (6.10) by g^{kj} and sum over j to obtain:

$$2\sum_{s,j} g^{jk} (\nabla_{\partial s} \mathsf{Ric})^s_j - \sum_j g^{jk} \nabla_{\partial j} S = 0.$$

This is the same as:

$$2\sum_s (\nabla_{\partial s} \mathsf{Ric})^{sk} - \sum_j \nabla_{\partial j}(S) g^{jk} = 0.$$

Which can also be written:

$$2\sum_s (\nabla_{\partial s} \mathsf{Ric})^{sk} - \sum_s \nabla_{\partial s}(S\overline{g})^{sj} = 2\sum_s (\nabla_{\partial_s} \overline{G})^{sj} = 0.$$

□

6.3 Sectional Curvature

We will now describe another local invariant of a pseudo-Riemannian manifold: the sectional curvature. Let (M, g) be a pseudo-Riemannian manifold and $\Pi \subset T_pM$ a two dimensional vector subspace of the tangent space at p such that the metric g restricted to Π is non-degenerated. The sectional curvature K of (M, g) evaluated at Π is the number:

$$K(p)(\Pi) = \frac{\langle R(X, Y)(Y), X\rangle}{\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2},$$

where the vectors X and Y generate Π. Note that the hypothesis that the metric is nondegenerated on Π implies that the denominator is nonzero. Let us show that the right hand side depends only on the vector subspace Π. The Bianchi identities imply that the numerator is symmetric on X and Y. One concludes that the whole expression also is. It is also clear that the number does not change if X or Y are multiplied by a nonzero scalar. Finally, the antisymmetry of the Riemann tensor implies that the right hand side does not change if X is replaced by $X' = X + \lambda Y$. The quantity K is known as the sectional curvature of (M, g). A pseudo-Riemannian manifold is said to have constant sectional curvature if $K(p)(\Pi)$ is a constant quantity.

Example 6.3.1. Let (Σ, g) be a Riemannian manifold of dimension $d = 2$. At each point $p \in \Sigma$ there is a unique two dimensional subspace of $T_p\Sigma$, namely the whole tangent space. Therefore, in this case, the sectional curvature is a smooth function:

$$K : \Sigma \to \mathbb{R}.$$

Let us see that in this case K is one half of the scalar curvature, $K = S/2$. This quantity is also known as the Gaussian curvature of the surface. If X and Y are an orthonormal basis for $T_p\Sigma$ then:

$$S(p) = \mathsf{tr}(\mathsf{Ric}) = \mathsf{Ric}(X, X) + \mathsf{Ric}(Y, Y) = 2\langle R(X, Y)(Y), X\rangle = 2K(p).$$

Example 6.3.2. The m dimensional sphere of raduis R:

$$S^m = \{v \in \mathbb{R}^{m+1} : |v| = R\}$$

has constant sectional curvature $K = 1/R^2$.

Exercise 6.3.3. Recall that the m dimensional hyperbolic space $\mathbb{H}^m$ is the manifold:

$$\mathbb{H}^m = \{(x^1, \cdots, x^m) \in \mathbb{R}^{m+1} : x^m > 0\}$$

with metric:

$$g = \frac{dx^1 \otimes dx^1 + \cdots + dx^m \otimes dx^m}{(x^m)^2}.$$

Show that hyperbolic space has constant curvature $K = -1$.

Exercise 6.3.4. Let $\mathbb{R}^{1,4}$ be the 5 dimensional Minkowski space. That is, the smooth manifold $\mathbb{R}^5$ with metric:

$$g = -dx^0 \otimes dx^0 + \sum_{i=1}^{4} dx^i \otimes dx^i.$$

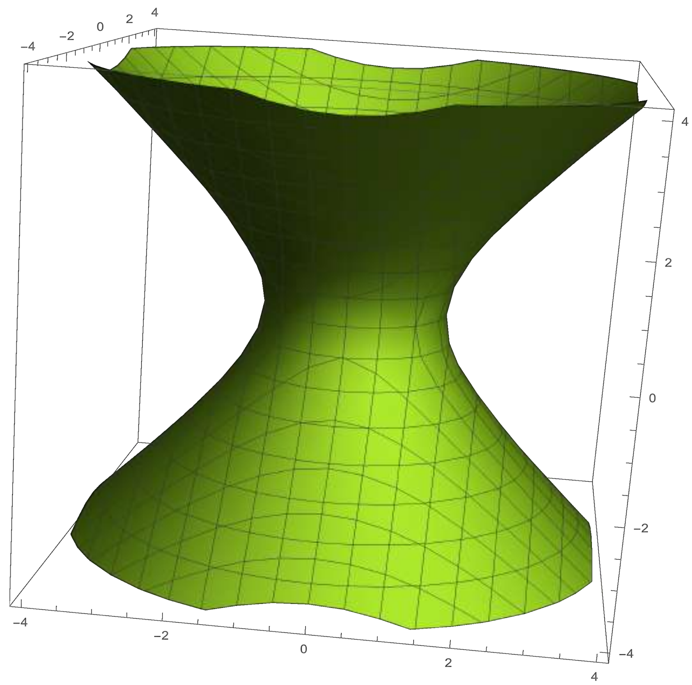


Figure 6.1: Tow dimensional de Sitter space

De Sitter space is the submanifold:

$$dS_4 = \Big\{(x^0, \cdots, x^4) \in \mathbb{R}^{1,4} : -(x^0)^2 + \sum_{i=1}^{4}(x^i)^2 = \alpha^2\Big\}.$$

Show that de Sitter space is diffeomorphic to $\mathbb{R} \times S^3$ and that the Minkowski metric induces a metric of Lorentz signature on dS_4. Show that the Riemann tensor satisfies:

$$R_{abcd} = \frac{1}{\alpha^2}\big(g_{ac}g_{bd} - g_{ad}g_{bc}\big).$$

Prove also that the Ricci tensor is proportional to the metric:

$$\mathsf{Ricc} = \frac{3g}{\alpha^2}$$

and the sectional curvature of de Sitter space is $K = 1/\alpha^2$.

Exercise 6.3.5. Let $\mathbb{R}^{2,3}$ be the smooth manifold $\mathbb{R}^5$ with metric:

$$g = -dx^0 \otimes dx^0 - dx^1 \otimes dx^1 + \sum_{i=2}^{4} dx^i \otimes dx^i.$$

Anti-de Sitter space is the submanifold:

$$AdS_4 = \Big\{(x^0, \cdots, x^4) \in \mathbb{R}^{1,4} : -(x^0)^2 - (x^1)^2 + \sum_{i=2}^{4}(x^i)^2 = -\alpha^2\Big\}.$$

Show that Anti-de Sitter space is diffeomorphic to $\mathbb{R}^3 \times S^1$ and that the Minkowski metric induces a metric of Lorentz signature on AdS_4. Show that the Riemann tensor satisfies:

$$R_{abcd} = -\frac{1}{\alpha^2}\big(g_{ac}g_{bd} - g_{ad}g_{bc}\big).$$

Prove also that the Ricci curvature is proportional to the metric:

$$\mathsf{Ric} = \frac{-3g}{\alpha^2},$$

and the sectional curvature of Anti-de Sitter space is $K = -1/\alpha^2$.

A fundamental result is that the sectional curvature in all tangent planes at a point determines the curvature tensor at that point.

Theorem 6.3.6. Let (M, g) be a Riemannian manifold of dimension ≥ 3. Fix $p \in M$. If $K(\Pi)$ is known for all planes Π at p then the value of the Riemann tensor at p is determined.

The result follows from a direct application of the antisymmetric property of the Riemann tensor and the Bianchi identities. For a proof we refer to [3], Theorem 3.5, page 381.

We will say that a Riemannian manifold (M, g) is locally isotropic at $p \in M$ if for every pair of unitary tangent vectors $u, v \in T_pM$ there exist open subsets $U, V \subseteq M$ and an isometry $\varphi : U \to V$ such that $\varphi(p) = p$ and $D\varphi(p)(v) = w$.

Proposition 6.3.7. Let (M, g) be a 3-dimensional Riemannian manifold which is locally isotropic at $p \in M$. Then M has constant sectional curvature at p. This means that $K(p)(\Pi) = K(p)(\Pi')$ for any two planes $\Pi, \Pi' \subset T_pM$.

Proof. Given Π and Π' consider unitary vectors v, v' which are orthogonal to Π and Π' respectively. Fix a local isometry such that $D\varphi(p)(v) = v'$. This implies that $D\varphi(p)(\Pi) = \Pi'$. Therefore:

$$K(p)(\Pi) = K(\varphi(p))(D\varphi(H)) = K(p)(\Pi').$$

□

The following result, whose proof can be found in [17], states that if the sectional curvature is constant at p then it is locally constant.

Theorem 6.3.8 (Schur's Lemma)**.** Let (M, g) be a connected Riemannian manifold of dimension ≥ 3. If there exists a function $f : M \to \mathbb{R}$ such that $f(p) = K(p)(\Pi)$, for all $\Pi \in T_pM$, then f is constant.

Note that the condition that $d \geq 3$ is necessary. In dimension $d = 2$ the statement is obviously false since the Gaussian curvature is typically not constant.

Proposition 6.3.9. Let (M, g) be a 3-dimensional Riemannian manifold with constant curvature, so that its sectional curvature can be seen as a function of $p \in M$. Then the Riemann tensor satisfies

$$R(X, Y)(Z) = K(\langle Z, Y\rangle X - \langle Z, X\rangle Y).$$

Hence, the it can be written in coordinates as follows:

$$R_{dcab} = \kappa_p(H)(h_{ad}h_{bc} - h_{ac}h_{bd}).$$

Proof. Both sides of the equation are tensors with the same symmetry properties. Set:

$$R'(X, Y)(Z) = K(\langle Z, Y\rangle X - \langle Z, X\rangle Y),$$

and define:

$$K'(\Pi) = \frac{\langle X, R'(X, Y)(Y)\rangle}{\langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2},$$

where X, Y generate Π. In view of Theorem 6.3.6, it suffices to show that $K = K'$. Choose an orthonormal basis ∂_i and linearly independent vectors

$X = \sum_i f^i \partial_i$ and $Y = \sum_i h^i \partial_i$ generating the plane Π and compute:

$$\begin{aligned} K'(\Pi) &= \frac{\langle X, R'(X,Y)(Y)\rangle}{\langle X,X\rangle\langle Y,Y\rangle - \langle X,Y\rangle^2} \\ &= \frac{\sum_{ijkl} f^i f^j h^k h^l \langle \partial_i, R'(\partial_j,\partial_k)(\partial_l)\rangle}{\sum_{ijkl} f^i f^j h^k h^l g_{ij}g_{kl} - g_{ik}g_{jl}} \\ &= K(\Pi)\frac{\sum_{ijkl} f^i f^j h^k h^l g_{lk}g_{ij} - g_{lj}g_{ik}}{\sum_{ijkl} f^i f^j h^k h^l g_{ij}g_{kl} - g_{ik}g_{jl}} = K(\Pi). \end{aligned}$$

□

A Riemannian manifold (M, g) is called geodesically complete if the domain of every geodesic can be extended to the whole real line. The following remarkable theorem provides a classification for manifolds of constant curvature.

Theorem 6.3.10 (Killing-Hopf)**.** Every geodesically complete, connected, simply connected Riemannian manifold (M, g) of constant curvature $K \in \{0, 1, -1\}$ is isometric to one of the following $L_0 = \mathbb{R}^n$, $L_1 = S^n$ or $L_{-1} = \mathbb{H}^n$, respectively.

Proof. See [3], Theorem 6.3, Page 401, or [14], Theorem 19, page 168. □

Exercise 6.3.11. Let (M, g) be a Riemannian manifold with sectional curvature K and $a > 0$ a positive constant. Show that the sectional curvature K_{ag} of the manifold (M, ag) is given by:

$$K_{ag} = K/a.$$

The previous exercise implies that any geodesically complete, connected, simply connected Riemannian manifold of constant curvature $K = a$ is isometric to on of the following $(L_k, \frac{1}{a}g_k)$, where the manifolds (L_k, g_k) are the standard models above.

6.4 Curvature as Geodesic Deviation

In Eucledian space, straight lines which are parallel stay parallel. This does not happen in arbitrary spaces. The curvature tensor can be interpreted as a measure of the deviation between geodesics.

Let us fix a pseudo-Riemannian manifold (M, g). A family of geodesics in M is an embedding $\sigma : I \times I \to M$, where each curve $\sigma_\lambda(s) = \sigma(\lambda, s)$ is a geodesic. One can define vector fields X, Y on The surface $S = \mathrm{im}(\sigma)$ by

$$Y = \sigma_* \partial_\lambda; \quad X = \sigma_* \partial_s.$$

Since each of the curves $\sigma_\lambda(s)$ is a geodesic we know that

$$\nabla_X(X) = 0.$$

Moreover, since the vector fields ∂_λ and ∂_s commute, we have

$$[Y, X] = 0.$$

Therefore, since ∇ is torsion free $\nabla_X(Y) = \nabla_Y(X)$. This implies that the curvature satisfies:

$$R(X, Y)(X) = \nabla_X \nabla_Y(X) = \nabla_X \nabla_X(Y).$$

Thus, the curvature is the second derivative of the vector Y in the direction of the geodesics. If we choose local coordinates and we write:

$$\nabla_X \nabla_X Y = \sum_i A^i \partial_i; \quad X = \sum_j X^j \partial_j; \quad Y = \sum_k Y^k \partial_k.$$

Then:

$$A^i = \sum_{j,k,l} X^j Y^k X^l R^i_{ljk}. \tag{6.12}$$

Chapter 7

Differential Forms

7.1 Differential Forms

Definition 7.1.1. A k-form η on M is a section of the k-th exterior power of the cotangent bundle $\Lambda^k(T^*M)$. The space of all k-forms on M is denoted by $\Omega^k(M)$. That is:

$$\Omega^k(M) = \Gamma\left(\Lambda^k T^*M\right).$$

A 0-form on M is a smooth function, i.e.,

$$\Omega^0(M) = C^\infty(M).$$

Given local coordinates, for each point $p \in U$ there is a basis $\{dx^1(p), \ldots, dx^m(p)\}$ for $T^*M(p)$. Therefore, the set

$$\{dx^{i_1}(p) \wedge \cdots \wedge dx^{i_k}(p) \,:\, 1 \leq i_1 < \cdots < i_k \leq m\}$$

is a basis for $\Lambda^k(T^*M)(p)$.

A differential form $\eta \in \Omega^k(M)$ can be written uniquely in the form,

$$\eta = \sum_I f_I dx^{i_1} \wedge \cdots \wedge dx^{i_k},$$

where the sum runs over all multiindices $I = (i_1, \ldots, i_k)$, $i_1 < \cdots < i_k$.

Definition 7.1.2. A differential form on M is a section of $\Lambda(T^*M)$. We will write:

$$\Omega(M) = \Gamma\left(\Lambda(T^*M)\right) = \Gamma\left(\bigoplus_k \Lambda^k(T^*M)\right) = \bigoplus_k \Omega^k(M).$$

Observation 4. The vector space $\Omega(M) = \bigoplus_k \Omega^k(M)$ is a graded algebra with product given by

$$(\eta \wedge \omega)(p) = \eta(p) \wedge \omega(p) \in \Lambda(T_p^* M).$$

This algebra is commutative in the graded sense, i.e. if $\eta \in \Omega^k(M)$ and $\omega \in \Omega^p(M)$ then $\omega \wedge \eta = (-1)^{pk} \eta \wedge \omega$.

7.2 DeRham Cohomology

We have seen that the derivative of a 0-form $f \in \Omega^0(M)$ is a 1-form $df \in \Omega^1(M)$. More generally, one can define the derivative $d\eta$ of a k-form η which turns out to be a $k+1$-form. This derivation operation, called the DeRham operator, plays a central role in topology.

Proposition 7.2.1. Let M be a manifold and $\Omega(M)$ the algebra of differential forms on M. There exists a unique degree 1 derivation d in $\Omega(M)$ such that:

1. For a function $f \in \Omega^0(M)$, df is the usual derivative.
2. $d^2 = 0$.

The derivation d is called the DeRham operator. The proof of Proposition 7.2.1 is a formal consequence of the following two lemmas.

Lema 7.2.2. Proposition 7.2.1 holds for $M = U \subseteq \mathbb{R}^m$.

Proof. Since the algebra $\Omega(U)$ is generated by smooth functions and the elements $dx^1, \dots, dx^m$, there is at most one derivation satisfying the conditions of the proposition. Let us define the operator d as follows. Given

$$\omega = \sum_I f_I \, dx^{i_1} \wedge \dots \wedge dx^{i_l},$$

set:

$$\begin{aligned} d\omega &= \sum_I df_I \wedge dx^{i_1} \wedge \dots \wedge dx^{i_l} \\ &= \sum_{I,j} \frac{\partial f_I}{\partial x^j} \wedge dx^j \wedge dx^{i^1} \wedge \dots \wedge dx^{i_l}. \end{aligned}$$

We leave it as an exercise to the reader to show that the operator d is a derivation that squares to zero. □

Lema 7.2.3. Let M be a manifold and $\{U_\alpha\}_{\alpha\in\mathcal{A}}$ an open cover of M. Then:

1. If D is a derivation of $\Omega(M)$, for each $\alpha \in \mathcal{A}$ there exists a unique derivation D^α of $\Omega(U_\alpha)$ such that
$$D(\omega)(p) = D^\alpha(\omega\big|_{U_\alpha})(p),$$
for any $\omega \in \Omega(M)$ and $p \in U_\alpha$. The restricted derivation D^α is also denoted $D\big|_{U\alpha}$.

2. $D = 0$ if and only if $D^\alpha = 0$ for all $\alpha \in \mathcal{A}$.

3. Given a family of derivations δ_α en $\Omega(U_\alpha)$ such that
$$\delta_\alpha\big|_{U_\alpha\cap U_\beta} = \delta_\beta\big|_{U_\alpha\cap U_\beta},$$
there exists a unique derivation δ on $\Omega(M)$ such that $\delta_\alpha = \delta^\alpha$.

Proof. By the argument in the proof of Lemma 3.4.8 one can show that given $\omega \in \Omega(U_\alpha)$ and $p \in U_\alpha$ there exists an open neighborhood of W of p and a form $\tilde{\omega} \in \Omega(M)$ such that
$$\omega\big|_W = \tilde{\omega}\big|_W.$$
We define:
$$D^\alpha(\omega)(p) := D(\tilde{\omega})(p),$$
and it is easy to verify that D^α is well defined and satisfies the required conditions. Let us show the second statement. Obviously, $D = 0$ implies $D^\alpha = 0$. On the other hand, suppose that $D^\alpha = 0$ for all $\alpha \in \mathcal{A}$. Since $p \in U_\alpha$ for some $\alpha \in \mathcal{A}$, we know that:
$$D(\omega)(p) = D^\alpha(\omega\big|_{U_\alpha})(p) = 0.$$
We conclude that $D = 0$. It remains to prove the last statement. For this one defines:
$$\delta(\omega)(p) := \delta_\alpha(\omega\big|_{U_\alpha})(p),$$
for any α such that $p \in U_\alpha$. We leave it as an exercise to the reader to show that this defines a derivation with the required properties. □

Corollary 7.2.4. $\Omega(M)$ is a commutative differential graded algebra.

Definition 7.2.5. The DeRham cohomology of a manifold M is the graded commutative algebra $\mathrm{H}(M) = \mathrm{H}^\bullet(\Omega(M), d)$.

A smooth map $f : M \to N$ induces an algebra homomorphism:

$$f^* : C^\infty(N) \to C^\infty(M);\ g \mapsto f^*(g) = g \circ f.$$

The homomorphism f^* is called the pull-back map. This homomorphism can be extended naturally to a homomorphism between the differential graded algebras of differential forms . Let us first consider the situation in local coordinates. Consider U, V open subsets of $\mathbb{R}^k$ and $\mathbb{R}^m$ respectively, and $f : U \to V$ a smooth function. We define the homomorphism $f^* : \Omega(V) \to \Omega(U)$ by the condition that is should be a morphism of differential graded algebras, as follows. Given

$$\omega = \sum_I g_I dx^{i_1} \wedge \cdots \wedge dx^{i_l} \in \Omega^l(V),$$

we set:

$$\begin{aligned} f^*(\omega) &= f^*\Big(\sum_I g_I dx^{i_1} \wedge ... \wedge dx^{i_l}\Big) \\ &= \sum_I f^*(g_I) \wedge f^*(dx^{i_1}) \wedge ... \wedge f^*(dx^{i_1}) \\ &= \sum_I f^*(g_I) \wedge df^*(x^{i_1}) \wedge ... \wedge df^*(x^{i_1}) \\ &= \sum_I f^*(g_I) \wedge df^{i_1} \wedge ... \wedge df^{i_l} \\ &= \sum_{I, j_1, \dots, j_l} (g_I \circ f) \left(\frac{\partial f^{i_1}}{\partial y^{j_1}} \cdots \frac{\partial f^{i_1}}{\partial y^{j_1}} \right) dy^{j_1} \wedge \cdots \wedge dy^{j_l}. \end{aligned}$$

Exercise 7.2.6. Show that given a smooth function $f : M \to N$ there exists a unique homomorphism of differential graded algebras, $f^* : \Omega(N) \to \Omega(M)$, such that for any $g \in \Omega^0(M)$, $f^*(g) = g \circ f$. Prove that $\mathsf{id}^* = \mathsf{id}$. Show that if $h : N \to S$ is another map then $(h \circ f)^* = f^* \circ h^*$.

Example 7.2.7. Consider $M = \mathbb{R}$. Then

$$\Omega(\mathbb{R}) = \Omega^0(\mathbb{R}) \oplus \Omega^1(\mathbb{R}) \cong C^\infty(\mathbb{R}) \oplus (C^\infty(\mathbb{R}) \otimes \mathbb{R}dt).$$

Therefore:

$$\begin{aligned} \mathrm{H}^0(\mathbb{R}) &= \mathsf{ker}(d : \Omega^0(\mathbb{R}) \to \Omega^1(\mathbb{R})) \\ &= \{f \in C^\infty(\mathbb{R}); df = 0\} \\ &= \{f \in C^\infty(\mathbb{R}); f = k\} \cong \mathbb{R}. \end{aligned}$$

Let us now show that $\mathrm{H}^1(\mathbb{R}) = 0$. It suffices to show that

$$d : \Omega^0(\mathbb{R}) \to \Omega^1(\mathbb{R}); \quad f \mapsto \frac{\partial f}{\partial t} dt$$

is surjective. Take $\eta = f dt \in \Omega^1(\mathbb{R})$ and define $g(t) = \int_0^t f(x)dx$. Then

$$dg = \frac{\partial g}{\partial t} dt = f dt = \eta.$$

We conclude that $\mathrm{H}^1(\mathbb{R}) = 0$.

7.3 The De Rham operator in $\mathbb{R}^3$

For the manifold $M = \mathbb{R}^3$, the spaces of differential forms can be naturally identified with vector fields and smooth functions. Under these identifications the de Rham operator corresponds to the gradient, divergence and curl. Let us recall the definitions of these operations.

Definition 7.3.1. Let f be a smooth function in $\mathbb{R}^3$ and

$$X = f^1 \frac{\partial}{\partial x^1} + f^2 \frac{\partial}{\partial x^2} + f^3 \frac{\partial}{\partial x^3}$$

a vector field. Then:

1. The gradient of f is the vector field:

$$\mathsf{grad}(f) = \frac{\partial f}{\partial x^1}\frac{\partial}{\partial x^1} + \frac{\partial f}{\partial x^2}\frac{\partial}{\partial x^2} + \frac{\partial f}{\partial x^2}\frac{\partial}{\partial x^3}.$$

2. The divergence of X is the smooth function:

$$\mathsf{div}(X) = \nabla \cdot X = \frac{\partial f^1}{\partial x^1} + \frac{\partial f^2}{\partial x^2} + \frac{\partial f^3}{\partial x^3}$$

3. The curl of X is the vector field:

$$\mathsf{curl}(X) = \nabla \times X = \left(\frac{\partial f^3}{\partial x^2} - \frac{\partial f^2}{\partial x^3}\right)\frac{\partial}{\partial x^1} + \left(\frac{\partial f^1}{\partial x^3} - \frac{\partial f^3}{\partial x^1}\right)\frac{\partial}{\partial x^2} + \left(\frac{\partial f^2}{\partial x^1} - \frac{\partial f^1}{\partial x^2}\right)\frac{\partial}{\partial x^3}.$$

There are natural identifications between differential forms and vector fields defined as follows. The isomorphism $\tau_1 : \mathfrak{X}(\mathbb{R}^3) \to \Omega^1(\mathbb{R}^3)$ is given by:

$$\tau_1(f^1\frac{\partial}{\partial x^1} + f^2\frac{\partial}{\partial x^2} + f^3\frac{\partial}{\partial x^3}) = f^1dx^1 + f^2dx^2 + f^3dx^3.$$

The isomorphism $\tau_2 : \mathfrak{X}(\mathbb{R}^3) \to \Omega^2(\mathbb{R}^3)$ is given by:

$$\tau_2(f^1\frac{\partial}{\partial x^1} + f^2\frac{\partial}{\partial x^2} + f^3\frac{\partial}{\partial x^3}) = f^3dx^1 \wedge dx^2 - f^2dx^1 \wedge dx^3 + f^1dx^2 \wedge dx^3.$$

The isomorphism $\tau_3 : C^\infty(\mathbb{R}^3) \to \Omega^3(\mathbb{R}^3)$ is given by

$$\tau_3(g) = gdx^1 \wedge dx^2 \wedge dx^3.$$

Exercise 7.3.2. Show that under the identifications above, the gradient, curl and divergence correspond to the DeRham operator, i.e. the following diagram commutes:

$$\begin{array}{ccccccc}
\Omega^0(\mathbb{R}^3) & \xrightarrow{d} & \Omega^1(\mathbb{R}^3) & \xrightarrow{d} & \Omega^2(\mathbb{R}^3) & \xrightarrow{d} & \Omega^3(\mathbb{R}^3 \\
\uparrow{\scriptstyle \mathsf{id}} & & \uparrow{\scriptstyle \tau_1} & & \uparrow{\scriptstyle \tau_2} & & \uparrow{\scriptstyle \tau_3} \\
\Omega^0(\mathbb{R}^3) & \xrightarrow{\mathsf{grad}} & \mathfrak{X}(\mathbb{R}^3) & \xrightarrow{\mathsf{curl}} & \mathfrak{X}(\mathbb{R}^3) & \xrightarrow{\mathsf{div}} & \Omega^0(\mathbb{R}^3)
\end{array}$$

Chapter 8

Integration and Stokes Theorem

8.1 Manifolds with Boundary

Definition 8.1.1. The upper half-space of dimension k is defined as

$$\mathbb{H}^k = \{(x^1, \dots, x^k) \in \mathbb{R}^k : x^k \geq 0\}.$$

The boundary of $\mathbb{H}^k$, denoted by $\partial\mathbb{H}^k$ is the subspace:

$$\partial\mathbb{H}^k = \{(x^1, \dots, x^k) \in \mathbb{R}^k : x^k = 0\}.$$

Definition 8.1.2. For an arbitrary subset $X \subseteq \mathbb{R}^k$ we say that $f : X \to M$ is smooth if for each $x \in X$ there exists an open subset $U_x \subseteq \mathbb{R}^k$ and a smooth function $\tilde{f}_x : U_x \to M$ such that $\tilde{f}|_{U_x \cap X} = f|_{U_x \cap X}$.

If X, Y are subsets of $\mathbb{R}^k$, a function $f : X \to Y$ is called a diffeomorphism if it is smooth, invertible and its inverse is smooth.

Lema 8.1.3. Let $U, V \subseteq \mathbb{H}^k$ be open subsets and $\varphi : U \to V$ a diffeomorphism. Then $\varphi(U \cap \partial\mathbb{H}^k) \subseteq \partial\mathbb{H}^k$.

Proof. Suppose that there exists $p = (x^1, \dots, x^{k-1}, 0) \in U$ such that

$$\varphi(p) = (y_1, \dots, y_k),$$

and $y_k > 0$. Consider the inverse function:

$$\varphi^{-1}\big|_W : W \to U$$

where $W \subseteq V$ is an open in $\mathbb{R}^k$ with $\varphi(p) \in W$. Since φ^{-1} is a diffeomorphism, its image is open in $\mathbb{R}^k$. On the other hand $p \in \varphi^{-1}(W) \subseteq U$. This is imposible because any open in $\mathbb{R}^k$ that contains p also contains points whose last coordinate is negative. □

Definition 8.1.4. A manifold with boundary M of dimension m is a Hausdorff second countable topological space together with an atlas $(U_\alpha, \varphi_\alpha)_{\alpha \in \mathcal{A}}$, where $\{U_\alpha\}_{\alpha \in A}$ is an open cover of M and $\varphi_\alpha : U_\alpha \to V_\alpha \subseteq \mathbb{H}^m$ are homeomorphisms such that the transition functions

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta)$$

are diffeomorphisms.

The interior of a manifold with boundary M is the subspace:

$$M^o = \{p \in M : \varphi_\alpha(p) \notin \partial(\mathbb{H}^m) \text{ for some chart } \varphi_\alpha\}.$$

The boundary of M is:

$$\partial M = \{p \in M : \varphi_\alpha(p) \in \partial(\mathbb{H}^m) \text{ for some chart } \varphi_\alpha\}.$$

Exercise 8.1.5. Let M be a manifold with boundary of dimension m. Show that:

1. $M^o \cap \partial M = \emptyset$.
2. M^o is a manifold of dimension m.
3. ∂M is a manifold of dimension $m - 1$.

Example 8.1.6. Then closed disk:

$$D^k := \{x \in \mathbb{R}^k : |x| \leq 1\}$$

is a manifold with boundary. The cylinder:

$$C^k := \{x \in \mathbb{R}^k : \frac{1}{2} \leq |x| \leq 1\}$$

is also a manifold with boundary.

8.2 Orientations

Definition 8.2.1. Let V be a real vector space of dimension $k < \infty$. The vector space $\Lambda^k(V)$ has dimension $d = 1$ and therefore as a topological space $\Lambda^k(V) - \{0\}$ has two connected components. An orientation of the vector space V is a choice of one of these connected components.

Exercise 8.2.2. An ordered basis $\{v_1, \dots, v_k\}$ for V determines an orientation: the connected component of $v_1 \wedge \cdots \wedge v_k \in \Lambda^k(V)$. Show that two basis induce the same orientation if and only if the change of base matrix has positive determinant.

Remark 8.2.3. An orientation on a one dimensional vector space L determines an orientation on L^* by the condition that if $v \in L$ and $\phi \in L^*$ are oriented then $\phi(v) > 0$.

Exercise 8.2.4. Show that if V has dimension $k < \infty$ then there is a natural isomorphism:

$$\Lambda^k(V^*) \cong \big(\Lambda^k(V)\big)^*,$$

given by:

$$\big(\phi_1 \wedge \cdots \wedge \phi_k\big)(v_1 \wedge \cdots \wedge v_k) = \frac{1}{k!} \sum_{\sigma \in S_k} \phi_{\sigma(1)}(v_1) \dots . \phi_{\sigma(k)}(v_k),$$

where S_k denotes the symmetric group. Conclude that an orientation on V induces naturally an orientation on V^*.

Definition 8.2.5. An orientation on a manifold M is a choice of an orientation on each tangent space T_pM which is locally constant in the following sense. For each point $p \in M$ there exist coordinates

$$\varphi = (x^1, \dots, x^m) : U \to V,$$

such that for all $q \in U$ the orientation in T_qM is given by

$$\frac{\partial}{\partial x^1}(q) \wedge \cdots \wedge \frac{\partial}{\partial x^m}(q).$$

A manifold is orientable if it admits an orientation. An oriented manifold is a manifold together with a choice of orientation.

Definition 8.2.6. An atlas $(U_\alpha, \varphi_\alpha)_{\alpha\in\mathcal{A}}$ in M is said to be oriented if for all $\alpha, \beta \in \mathcal{A}$ the transition functions:

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \to \varphi_\beta(U_\alpha \cap U_\beta),$$

satisfy the condition $\mathsf{det}(D(\varphi_\beta \circ \varphi_\alpha^{-1})(q)) > 0$, for all $q \in \varphi_\alpha(U_\alpha \cap U_\beta)$.

Definition 8.2.7. A volume form on a manifold M of dimension m is a differential form $\omega \in \Omega^m(M)$ such that $\omega(p) \neq 0$ for all $p \in M$.

Proposition 8.2.8. Let M be a manifold. Then:

1. An oriented atlas $(U_\alpha, \varphi_\alpha)$ induces an orientation on M. All orientations are induced by an oriented atlas.

2. A volume form ω induces and orientation on M. All orientations are induced by a volume form.

Proof. Let $(U_\alpha, \varphi_\alpha)$ be an oriented atlas. This defines an orientation on M by declaring that at each point $p \in M$:

$$\partial_{x^1}(p) \wedge \cdots \wedge \partial_{x^m}(p)$$

is oriented for any coordinate $\varphi = (x^1, \ldots, x^m)$ in the atlas. The change of basis matrix is:

$$D(\varphi_\beta \circ \varphi_\alpha^{-1})(q),$$

which has positive determinant since the atlas is oriented. By Exercise 8.2.2 we conclude that this orientation is well defined. Conversely, given an orientation $\mathcal{O}$ on M one can choose an oriented subatlas of the maximal atlas inducing the orientation $\mathcal{O}$ by requiring the condition that:

$$\partial_{x^1}(p) \wedge \cdots \wedge \partial_{x^m}(p)$$

is is oriented. Let us now prove the second claim. Consider a volume form $\omega \in \Omega^m(M)$. We define an orientation on each cotangent space T^*M by declaring that:

$$\omega(p) \in \Lambda^m(T_p^*M)$$

lies in the positive connected component. Let us show that any orientation $\mathcal{O}$ can be defined in this manner. We consider an oriented atlas $(U_\alpha, \varphi_\alpha)$

inducing $\mathcal{O}$ and a partition of unity ρ_α subordinate to the cover $\{U_\alpha\}$. Then we define a volume form $\omega \in \Omega^m(M)$ by:

$$\omega(p) := \sum_\alpha \rho_\alpha(p) dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^m.$$

Here the sum is over all indices α such that $p \in U_\alpha$. Since the partition of unity is locally finite, the sum is well defined. Since the atlas is oriented we know that:

$$dx_\alpha^1 \wedge \cdots \wedge dx_\alpha^m = \lambda dx_\beta^1 \wedge \cdots \wedge dx_\beta^m$$

for some $\lambda > 0$ and therefore $\omega(p) \neq 0$. □

Exercise 8.2.9. Show that the Klein bottle K (3.1.14) does not admit an orientation.

8.3 Integration

Definition 8.3.1. The support of a differential form $\omega \in \Omega^k(M)$ is the set:

$$\mathsf{supp}(\omega) = \overline{\{p \in M : \omega(p) \neq 0\}}.$$

The set of differential forms with compact support will be denoted by:

$$\Omega_c(M) = \{\Omega \in \omega(M) : \mathsf{supp}(\omega) \text{ is compact}\}.$$

Definition 8.3.2. Let $U \subseteq \mathbb{R}^m$ be an open set and $\eta \in \Omega_c^m(U)$ a form with compact support. Then η can be written uniquely as $\eta = f dx^1 \wedge \cdots \wedge dx^m$. The integral of η over U is defined as:

$$\int_U \eta = \int_U f|dx^1 \wedge \cdots \wedge dx^m|,$$

where the right hand side denotes the Riemann integral of the function f.

Lema 8.3.3. Let $\varphi : U \to V$ be an orientation preserving diffeomorphism between open subsets of $\mathbb{H}^m$ and $\eta \in \Omega_c^m(V)$. Then

$$\int_U \varphi^*(\eta) = \int_V \eta.$$

Proof. We write $\eta = f dx^1 \wedge \cdots \wedge dx^m$ and use the change of variable formula to compute:

$$\begin{aligned}\int_U \varphi^*(\eta) &= \int_U (f \circ \varphi)\mathsf{det}(D\varphi) dy_1 \wedge \cdots \wedge dy_m = \int_U (f \circ \varphi)|\mathsf{det}(D\varphi)||dy_1 \wedge dy_m| \\ &= \int_V f|dx^1 \wedge \cdots \wedge dx^m| = \int_V \eta.\end{aligned}$$

□

In view of the previous lemma, the following definition makes sense.

Definition 8.3.4. Let V be an oriented manifold which is diffeomorphic to an open subset of $\mathbb{H}^m$ and $\eta \in \Omega_c^m(V)$. We define

$$\int_V \eta = \int_U \varphi^*(\eta),$$

for any diffeomorphism $\varphi : U \subseteq \mathbb{H}^m \to V$ that preserves the orientation.

Definition 8.3.5. Let M an m dimensional oriented manifold with boundary and $\omega \in \Omega_c(M)$. We define

$$\int_M \omega = \sum_\alpha \int_{U_\alpha} \rho_\alpha \omega,$$

where $(U_\alpha, \varphi_\alpha)_{\alpha \in \mathcal{A}}$ is a finite atlas for an open subset $W \subseteq M$ that contains the support of ω and ρ_α is a partition of unity.

We will show that the definition above is independent of the choices of atlas and partition of unity. For a different atlas (V_β, ψ_β) and partition of unity τ_β, we have:

$$\begin{aligned}\sum_\alpha \int_{U_\alpha} \rho_\alpha \omega &= \sum_\alpha \int_{U_\alpha} \rho_\alpha \left(\sum_\beta \tau_\beta\right) \omega = \sum_{\alpha,\beta} \int_{U_\alpha} \rho_\alpha \tau_\beta \omega \\ &= \sum_{\alpha,\beta} \int_{U_\alpha \cap V_\beta} \rho_\alpha \tau_\beta \omega = \sum_\beta \int_{V_\beta} \tau_\beta \omega.\end{aligned}$$

Lema 8.3.6. Let $U, V \subseteq \mathbb{H}^k$ be open subsets, $p \in \partial U$ and $\varphi : U \to V$ a diffeomorphism. The derivative matrix $D(\varphi)(p)$ has the form:

$$D(\varphi)(p) = \begin{pmatrix} * & * & \dots & * \\ \vdots & \vdots & \dots & \vdots \\ 0 & \dots & 0 & \frac{\partial x^k}{\partial y^k} \end{pmatrix}$$

where $\partial x^k / \partial y^k(p) > 0$.

Proof. We need to prove that

$$\frac{\partial x^k}{\partial y^i}(p) = 0,$$

for $i < k$, and that

$$\frac{\partial x^k}{\partial y^k}(p) > 0.$$

For $i < k$ the vector $\partial_{y^i}(p)$ is tangent to the boundary, and since φ preserves the boundary, so is $D\varphi(p)(\partial_{y^i}(p))$. We conclude that

$$\frac{\partial x^k}{\partial y^i}(p) = 0.$$

On the other hand, we know that $D\varphi(p)$ is not singular and therefore

$$\frac{\partial x^k}{\partial y^k}(p) \neq 0.$$

But

$$\frac{\partial x^k}{\partial y^k} = \frac{d}{dt}\Big|_{t=0} x^k(\varphi(p + ty_k))$$

is a nonnegative number becuase $x^k(\varphi(p + ty_k)) > 0$ for $t > 0$. □

Let M be a manifold with boundary, $p \in \partial M$ and $v \in T_pM$ such that $v \notin T_p\partial M$. We say that v points inside if for any choice of coordinates

$$\varphi_\alpha : U_\alpha \to V_\alpha,$$

the last component of $D\varphi_\alpha(p)(v)$ is positive. We say that v points outside if it does not point inside. Note that, in view of the previous lemma, if $\varphi_\alpha : U_\alpha \to V_\alpha$ and $\varphi_\beta : U_\beta \to V_\beta$ are coordinates then the last coordinate of $D\varphi_\alpha(p)(v)$ has the same sign as the last coordinate of $D\varphi_\beta(p)(v)$.

Definition 8.3.7. Let M be an oriented manifold with boundary. The manifold ∂M acquires an orientation defined by the following rule. An ordered basis $\{v_1, \dots, v_{k-1}\}$ of $T_p\partial M$ is oriented if and only if the ordered basis $\{e, v_1, \dots, v_{k-1}\}$ for T_pM is oriented, for any vector $e \in T_pM$ that points outside.

Exercise 8.3.8. Show that the definition above does not depend on the vector $e \in T_pM$.

Exercise 8.3.9. Show that if $dx^1 \wedge \cdots \wedge dx^n$ is an oriented volume form in $\mathbb{H}^m$ then $(-1)^m dx^1 \wedge \cdots \wedge dx^{m-1}$ is an oriented volume form in $\partial\mathbb{H}^n$.

8.4 Stokes' Theorem

Let us now discuss the higher dimensional generalization of the fundamental theorem of calculus: Stokes' theorem. Slightly abusing the notation, for $\alpha \in \Omega(M)$ we will often write:

$$\int_{\partial M} \alpha := \int_{\partial M} \iota^* \alpha,$$

where $\iota : \partial M \to M$ is the natural inclusion.

Theorem 8.4.1 (Stokes' Theorem)**.** Let M be an m dimensional oriented manifold with boundary and $\alpha \in \Omega_c^{m-1}(M)$. Then

$$\int_{\partial M} \alpha = \int_M d\alpha.$$

Proof. We will divide the proof in three steps:

1. $M = \mathbb{R}^n$.
2. $M = \mathbb{H}^n$.
3. The previous cases imply the general statement.

In the first case, the differential form α can be written in the form:

$$\alpha = \sum_i f^i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n.$$

By linearity and symmetry it suffices to prove the statement for

$$\beta = f dx^1 \wedge \cdots \wedge dx^{n-1}.$$

Take $L \gg 0$ such that $\mathsf{supp}(\beta) \subseteq \mathbb{R}^{n-1} \times [-\frac{L}{2}, \frac{L}{2}]$. Then:

$$\begin{aligned}
\int_{\mathbb{R}^n} d\beta &= \int_{\mathbb{R}^n} \frac{\partial f}{\partial x^n} dx^n \wedge dx^1 \wedge \cdots \wedge dx^{n-1} \\
&= \pm \int_{\mathbb{R}^{n-1}} \int_{-L}^{L} \frac{\partial f}{\partial x^n} |dx^1 \wedge \cdots \wedge dx^n| \\
&= \pm \int_{\mathbb{R}^{n-1}} (f(\ldots, L) - f(\ldots, -L)) |dx^1 \ldots dx^n| \\
&= \pm \int_{\mathbb{R}^{n-1}} 0 = 0.
\end{aligned}$$

On the other hand, since $\partial \mathbb{R}^n = \emptyset$:

$$\int_{\partial \mathbb{R}^n} \beta = \int_{\emptyset} \beta = 0.$$

Let us look now at the second case. It suffices to consider two types of forms:

$$\begin{aligned}
\beta &:= f dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n, \ i < n, \\
\gamma &:= f dx^1 \wedge \cdots \wedge dx^{n-1}.
\end{aligned}$$

For the form β we compute:

$$\int_{\partial \mathbb{H}^n} \beta = \int_{\partial \mathbb{H}^n} \iota^*(f dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n) = \int_{\partial \mathbb{H}^n} 0 = 0.$$

On the other hand, taking $L \gg 0$ sufficiently large:

$$\begin{aligned}
\int_{\mathbb{H}^n} d\beta &= \int_{\mathbb{H}^n} \frac{\partial f}{\partial x^i} dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\
&= \pm \int_{\mathbb{H}^{n-1}} \int_{-L}^{L} \frac{\partial f}{\partial x^i} |dx^1 \ldots dx^n| \\
&= \pm \int_{\mathbb{H}^{n-1}} (f(\ldots, L, \ldots) - f(\ldots, -L, \ldots)) |dx^1 \ldots dx^n| \\
&= \pm \int_{\mathbb{H}^{n-1}} 0 = 0.
\end{aligned}$$

Let us now consider the form γ:

$$\begin{aligned}
\int_{\mathbb{H}^n} d\gamma &= \int_{\mathbb{H}^n} \frac{\partial f}{\partial x^n} dx^n \wedge dx^1 \wedge \cdots \wedge dx^{n-1} \\
&= (-1)^{n-1} \int_{\mathbb{H}^n} \frac{\partial f}{\partial x^n} dx^1 \wedge \cdots \wedge dx^n \\
&= (-1)^{n-1} \int_{\mathbb{H}^n} \frac{\partial f}{\partial x^n} |dx^1 \dots dx^n| \\
&= (-1)^{n-1} \int_{\mathbb{R}^{n-1}} \int_0^L \frac{\partial f}{\partial x^n} |dx^1 \dots dx^n| \\
&= (-1)^{n-1} \int_{\mathbb{R}^{n-1}} (f(\dots, L) - f(\dots, 0)) |dx^1 \dots dx^{n-1}| \\
&= (-1)^n \int_{\mathbb{R}^{n-1}} f(\dots, 0) |dx^1 \dots dx^{n-1}| \\
&= \int_{\partial\mathbb{H}^n} f dx^1 \wedge \cdots \wedge dx^{n-1} = \int_{\partial\mathbb{H}^n} \gamma.
\end{aligned}$$

Finally, let us prove the last claim. Let W be an open subset that contains the support of α and (U_i, ρ_i) a finite partition of unity in W such that each of the open subsets U_i is isomorphic either to $\mathbb{R}^n$ or to $\mathbb{H}^n$. Then:

$$\begin{aligned}
\int_M d\alpha &= \sum_i \int_M d(\rho_i \alpha) = \sum_i \int_{U_i} d(\rho_i \alpha) = \sum_i \int_{\partial U_i} \rho_i \alpha \\
&= \sum_i \int_{\partial M} \rho_i \alpha = \int_{\partial M} \alpha.
\end{aligned}$$

□

Example 8.4.2. Consider the interval $M = [a, b] \subseteq \mathbb{R}$. A 0-form is a smooth function on $[a, b]$. Taking into account the orientation induced on the boundary $\partial M = \{a, b\}$, Stokes' theorem states that:

$$\int_a^b \frac{\partial f}{\partial t} dt = \int_{[a,b]} df = \int_{\partial[a,b]} f = f(b) - f(a).$$

This is of course the fundamental theorem of calculus.

Example 8.4.3. Given $a, b > 0$, the area of the ellipse

$$M = \{(x, y) \in \mathbb{R}^2 : ax^2 + by^2 \leq 1\}$$

is then given by $A = \int_M dx \wedge dy$. This integral can be computed using a change of variables

$$\varphi : D^2 \to M; \quad \varphi(p,q) = (\frac{p}{\sqrt{a}}, \frac{q}{\sqrt{b}}).$$

So that:

$$\begin{aligned} A = \int_M dx \wedge dy &= \int_{D^2} \varphi^*(dx) \wedge \varphi^*(dy) \\ &= \int_{D^2} \frac{dp}{\sqrt{a}} \wedge \frac{dq}{\sqrt{q}} = \frac{1}{\sqrt{ab}} \int_{D^2} dp \wedge dq = \frac{\pi}{\sqrt{ab}}. \end{aligned}$$

On the other hand, we observe that

$$\alpha = \frac{1}{2}(xdy - ydx)$$

satisfies $d\alpha = dx \wedge dy$, so that Stokes' theorem gives:

$$A = \int_M d\alpha = \int_{\partial M} \alpha.$$

Parametrising the boundary of the ellipse by the function:

$$\begin{array}{ccccc} \gamma & : & [0, 2\pi] & \longrightarrow & \partial M \\ & & t & \longmapsto & (\frac{\cos t}{\sqrt{a}}, \frac{\sin t}{\sqrt{b}}) \end{array}$$

one obtains:

$$\gamma^*(\alpha) = \frac{1}{2}(\gamma^*(xdy) - \gamma^*(ydx)) = \frac{1}{2}\left(\frac{\cos^2 t}{\sqrt{ab}} + \frac{\sin^2 t}{\sqrt{ab}}\right) dt = \frac{dt}{2\sqrt{ab}},$$

so that:

$$A = \int_{\partial M} \alpha = \int_0^{2\pi} \gamma^*(\alpha) = \int_0^{2\pi} \frac{dt}{2\sqrt{ab}} = \frac{\pi}{\sqrt{ab}}.$$

8.5 Classical Stokes Theorem

Recall from the previous chapter that, in $\mathbb{R}^3$, differential forms can be identified with vector fields and functions in such a way that the following diagram commutes:

$$\begin{array}{ccccccc}
\Omega^0(\mathbb{R}^3) & \xrightarrow{d} & \Omega^1(\mathbb{R}^3) & \xrightarrow{d} & \Omega^2(\mathbb{R}^3) & \xrightarrow{d} & \Omega^3(\mathbb{R}^3 \\
\uparrow \text{id} & & \uparrow \tau_1 & & \uparrow \tau_2 & & \uparrow \tau_3 \\
\Omega^0(\mathbb{R}^3) & \xrightarrow{\text{grad}} & \mathfrak{X}(\mathbb{R}^3) & \xrightarrow{\text{curl}} & \mathfrak{X}(\mathbb{R}^3) & \xrightarrow{\text{div}} & \Omega^0(\mathbb{R}^3)
\end{array} \tag{8.1}$$

Given a curve $\gamma : I \to \mathbb{R}^3$ and a vector field

$$X = f^1\partial_{x^1} + f^2\partial_{x^2} + f^3\partial_{x^3} \in \mathfrak{X}(U)$$

defined on an open set U that contains $\gamma(I)$, the integral of X along γ is defined by:

$$\int_\gamma X = \int_I \langle \gamma'(t), X(\gamma(t)\rangle dt.$$

This definition is compatible with the identification τ_1 in the sense that:

$$\int_\gamma X = \int_\gamma \tau_1(X). \tag{8.2}$$

Let us consider an oriented surface Σ embedded in $\mathbb{R}^3$. The area form $dA \in \Omega^2(\Sigma)$ is characterized by the property that given a local diffeomorphism that preserves the orientation

$$\varphi : U \to \Sigma, \quad \varphi(z, w) = (x^1, x^2, x^3),$$

the pull-back of the area form is given by:

$$\varphi^*(dA) = |\varphi_z \times \varphi_w| dz \wedge dw,$$

where

$$\varphi_z = \left(\frac{\partial x^1}{\partial z}, \frac{\partial x^2}{\partial z}, \frac{\partial x^3}{\partial z}\right); \quad \varphi_w = \left(\frac{\partial x^1}{\partial w}, \frac{\partial x^2}{\partial w}, \frac{\partial x^3}{\partial w}\right).$$

Exercise 8.5.1. Prove that the volume form dA described above is well defined. Moreover, prove that given a vector field $X \in \mathfrak{X}(U)$ where $\Sigma \subseteq U$ one has:

$$(X \cdot \mathbf{n})\, dA = \iota^*(\tau_2(X)),$$

where $\mathbf{n}$ denotes the unit vector normal to Σ such that if $\{Y, Z\}$ is an oriented basis in $T_p\Sigma$ then:$\{\mathbf{n}, Y, Z\}$ is an oriented basis in $T_p\mathbb{R}^3 \cong \mathbb{R}^3$.

In dimension $d = 2$ the theorem of Stokes becomes that of Green:

Theorem 8.5.2 (Green)**.** Let Σ be an oriented compact surface with boundary embedded in $\mathbb{R}^3$ and $X = f^1\partial_{x^1} + g^2\partial_{x^2} + f^3\partial_{x^3}$ be a vector field defined on an open region that contains Σ. Then:

$$\int_{\partial\Sigma} X = \int_{\Sigma} \big(\mathsf{curl}(X) \cdot \mathbf{n}\big) dA.$$

Proof. Using Equation (8.2), Exercise 8.5.1 and Stokes' theorem we compute:

$$\int_{\partial\Sigma} X = \int_{\partial\Sigma} \tau_1(X) = \int_{\Sigma} d\big(\tau_1(X)\big) = \int_{\Sigma} \tau_2\big(\mathsf{curl}(X)\big) = \int_{\Sigma} \big(\mathsf{curl}(X) \cdot \mathbf{n}\big) dA.$$

□

Theorem 8.5.3 (Gauss)**.** Let B be an oriented compact 3-manifold with boundary embedded in $\mathbb{R}^3$ and:

$$X = f^1\partial_{x^1} + f^2\partial_{x^2} + f^3\partial_{x^3}.$$

be a vector field defined on an open region that contains B. Then:

$$\int_{\partial B} \big(X \cdot \mathbf{n}\big) dA = \int_{B} \mathsf{div}(X) dx^1 \wedge dx^2 \wedge dx^3.$$

Proof. Using Exercise 8.5.1 and Stokes' theorem we compute:

$$\int_{\partial B} \big(X \cdot \mathbf{n}\big) dA = \int_{\partial B} \iota^*(\tau_2(X)) = \int_{B} d(\tau_2(X)) = \int_{B} \tau_3(\mathsf{div}X) = \int_{b} (\mathsf{div}X) dx^1 \wedge dx^2 \wedge dx^3.$$

□

Finally, let us see how to recover the usual Stokes' theorem in $\mathbb{R}^3$. We need to express the integral of 2-forms over surfaces in the language of vector calculus.

Lema 8.5.4. Let S be a compact oriented surface with boundary in $\mathbb{R}^3$ and X be vector field defined in an open neighborhood that contains S. Then:

$$\int_{S} \tau_2(X) = \int_{S} X \cdot \mathbf{n} dA.$$

Proof. It suffices to prove the statement in the case where S can be covered by one coordinate chart:

$$\varphi : U \to S; (z, w) \mapsto \varphi(z, w).$$

Suppose that the vector field X is given by:

$$X = f_1 \frac{d}{dx^1} + f_2 \frac{d}{dx^2} + f_3 \frac{d}{dx^3},$$

so that the form $\tau_2(X)$ is:

$$\tau_2(X) = f_1 dx^1 \wedge dx^2 - f_2 dx^1 \wedge dx^3 + f_3 dx^1 \wedge dx^2,$$

Let us first compute the left hand side:

$$\int_S \tau_2(X) = \int_U \varphi^*(\tau_2(X)) = \int_U (f_1, f_2, f_3) \cdot \left(\frac{d\varphi}{dz} \times \frac{d\varphi}{dw} \right) dz \wedge dw. \tag{8.3}$$

On the other hand, the right hind side is computed as follows:

$$\int_S X \cdot \mathbf{n} dA = \int_U (f_1, f_2, f_3) \cdot \frac{\left(\frac{d\varphi}{dz} \times \frac{d\varphi}{dw} \right)}{\left| \frac{d\varphi}{dz} \times \frac{d\varphi}{dw} \right|} \sqrt{\det(g_{ij})} dz \wedge dw. \tag{8.4}$$

Here g_{ij} are the components of the metric in the coordinates z, w. By Lagrange's identity, we know that:

$$\sqrt{\det(g_{ij})} = \left| \frac{d\varphi}{dz} \times \frac{d\varphi}{dw} \right|,$$

so that 8.4 is equal to 8.3. □

Theorem 8.5.5. Let S be an compact oriented surface with boundary embedded in $\mathbb{R}^3$ and X a vector field defined in an open that contains S. Then:

$$\int_{\partial S} X = \int_S \mathsf{curl} X \cdot \mathbf{n} dA.$$

Proof. This follows formally from the general Stokes' theorem, Lemma 8.5.4 and diagram 8.1:

$$\int_{\partial S} X = \int_{\partial S} \tau_1(X) = \int_S d(\tau_1(X)) = \int_S (\tau_2)^{-1} d(\tau_1(X)) \cdot \mathbf{n} dA = \int_S \mathsf{curl} X \cdot \mathbf{n} dA$$

□

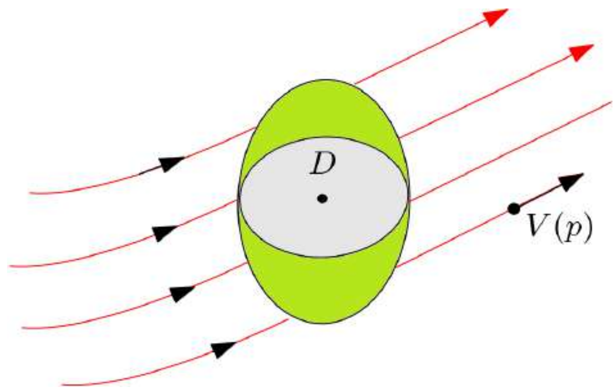


Figure 8.1: Motion of a fluid

8.6 Conservation of Mass

Let D be a domain with smooth boundary in $\mathbb{R}^3$ contained in a region filled with a fluid. For our immediate purposes, by a fluid we mean a continuous distribution of matter that traverses a well defined trajectory. Mathematically, the fluid is determined by a density function $\mu(x,t)$, and a a velocity vector field $v = \sum_k v^k(x,t)\partial_{x^k}$.

The density function has the property that, for each t, the total mass contained in D is equal to

$$m(t,D) = \int_D \mu(x,t)dx^1 \wedge dx^2 \wedge dx^3.$$

Hence, the rate of change of mass inside D is given by

$$\frac{dm}{dt} = \int_D \frac{d\mu(x,t)}{dt} dx^1 \wedge dx^2 \wedge dx^3.$$

On the other hand, the fluid flow rate across a small section of boundary ΔA is given approximately by $\mu(p,t)\mathrm{v}(p,t) \cdot \mathbf{n}\Delta A$.

Therefore, the total fluid crossing the boundary ∂D at time t is

$$\int_{\partial D} \mu(x,t)v(x,t) \cdot \mathbf{n} dA.$$

By the principle of conservation of mass, the total fluid crossing the boundary must be equal to the rate of change of mass, i.e.,

$$\int_{\partial D} \mu(x,t)v(x,t) \cdot \mathbf{n} dA = -\int_D \frac{d\mu(x,t)}{dt} dx^1 \wedge dx^2 \wedge dx^3.$$

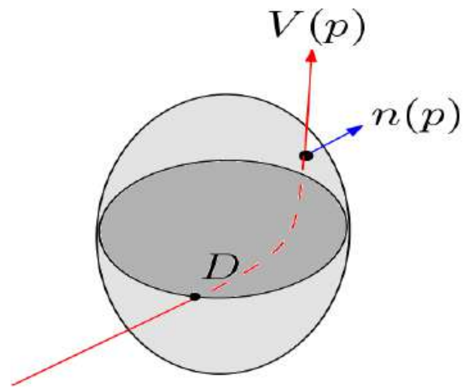


Figure 8.2: Flow through D

On the other hand, by Gauss' theorem:

$$\int_{\partial D} \mu(x,t)v(x,t)\cdot \mathbf{n}dA = \int_D \mathsf{div}(\mu v)dx^1\wedge dx^2\wedge dx^3.$$

Since these equations are valid on an arbitrary domain D, one concludes:

$$\frac{d\mu}{dt} + \mathrm{div}(\mu \mathrm{v}) = 0. \tag{8.5}$$

This equation is known as the continuity equation and expresses the conservation of mass for a fluid.

8.7 The Hodge Star Operator

Let V be a vector space of dimension m and g a pseudo-Euclidian structure on V. There is an induced pseudo-Euclidian structure on $\Lambda^k(V)$, also denoted by g, given by:

$$g(v_1\wedge\cdots\wedge v_k, w_1\wedge\cdots\wedge w_k) = \mathsf{det}([g(v_i,w_j)]).$$

Let us fix an orientation on V and denote by ω the unique element of $\Lambda^m(V)$ which is oriented and has unit norm. The Hodge star operator, denoted by $*$ is the linear isomorphism:

$$*:\Lambda^k(V)\to\Lambda^{m-k}(V),$$

characterized by the property that:

$$\alpha \wedge *(\beta) = g(\alpha, \beta)\omega.$$

More explicitly, if $B = \{v_1, \ldots, v_m\}$ is an oriented basis for V then the star operator is given by:

$$*(v_1 \wedge \cdots \wedge v_k) = \frac{1}{\sqrt{|\mathsf{det}(g_{ij})|}} v_{k+1} \wedge \cdots \wedge v_m.$$

Note that if V is an oriented vector space with a pseudo-Euclidian structure, so is V^* and therefore the Hodge star operator also induces isomorphisms:

$$* : \Lambda^k(V^*) \to \Lambda^{m-k}(V^*),$$

such that if if $B = \{\omega^1, \ldots, \omega^m\}$ is an oriented basis for V^* then the star operator is given by:

$$*(\omega^1 \wedge \cdots \wedge \omega^k) = \sqrt{|\mathsf{det}(g_{ij})|}\omega^{k+1} \wedge \cdots \wedge \omega^m.$$

Let M be an oriented pseudo-Riemannian manfiold. There is a unique volume form $\omega \in \Omega^m(M)$ which in local oriented coordinates can be written as:

$$\omega = \sqrt{|\mathsf{det}(g_{ij})|} dx^1 \wedge \otimes \wedge dx^m.$$

This form is well defined because the value $\omega(p)$ can be characterised as the unique vector in $\Lambda^m(T_p^*M)$ which is oriented and has unit norm.

The Hodge star operator is the isomorphism:

$$* : \Omega^k(M) \to \Omega^{m-k}(M),$$

characterised by the property that:

$$\eta \wedge *(\theta) = g(\eta, \theta)\omega,$$

for any pair of forms $\eta, \theta \in \Omega^k(M)$. In local oriented coordinates the star operator takes the form:

$$*(f dx^1 \wedge \cdots \wedge dx^k) = f\sqrt{|\mathsf{det}(g_{ij})|} dx^{k+1} \wedge \cdots \wedge dx^m.$$

The formal adjoint

$$d^* : \Omega^k(M) \to \Omega^{k-1}(M)$$

of the de-Rham operator is defined by:

$$d^* = (-1)^k *^{-1} d * .$$

The Hodge Laplacian of M is the differential operator: $\Delta : \Omega(M) \to \Omega(M)$ defined by: $\Delta = dd^* + d^*d$.

8.8 Lie Groups

A *Lie group* G is a manifold endowed with a group structure such that the product map

$$(g, h) \mapsto gh$$

and the inversion map

$$g \mapsto g^{-1}$$

are smooth. Here are some of the examples that are occur in relativity.

- Let $G = \mathbb{R}$ be the *additive* group of real numbers. This is an abelian Lie group.
- Let $G = \mathbb{R}^+$ be the *multiplicative* abelian group of positive real numbers. Then G is also a Lie group.
- The circle $G = S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is an abelian Lie group with respect to the multiplication of complex numbers.
- If $G_1, \ldots, G_n$ are Lie groups, then $G = G_1 \times \cdots \times G_n$ is a Lie group with the group operations defined component wise. For instance the group

  $$T^n = \underbrace{S^1 \times \cdots \times S^1}_{n}$$

  is called the *n-torus*.
- The 3-sphere S^3, seen as the set of all quaternions of norm 1, is a Lie group with respect to quaternionic multiplication.
- The *general linear group*

  $$\mathsf{GL}(n, \mathbb{R}) = \{A \in \mathsf{Mat}_n(\mathbb{R}) : \mathsf{det}(A) \neq 0\}.$$

is a Lie group. Notice that since the determinant is a continuous map, $\mathrm{GL}(n,\mathbb{R})$ is an open subset of the space of all matrices. The product and inverse functions are algebraic and therefore smooth.

We call *matrix Lie groups* the subgroups of $\mathrm{GL}(n,\mathbb{R})$ that are also smooth submanifolds.

- The *special linear group*

$$\mathrm{SL}(n,\mathbb{R}) = \{A \in \mathsf{Mat}_n(\mathbb{R}) | \det(A) = 1\}.$$

The special linear group is a matrix Lie group. To see this, it is enough to prove that $\mathrm{SL}(n,\mathbb{R})$ is a smooth submanifold of $\mathrm{GL}(n,\mathbb{R})$. Consider the determinant function

$$\det \colon \mathrm{GL}(n,\mathbb{R}) \to \mathbb{R}.$$

Then $\mathrm{SL}(n,\mathbb{R}) = \det^{-1}(1)$. We assert that for every matrix $A \in \mathrm{SL}(n,\mathbb{R})$ the $\mathbb{R}$-linear map

$$D(\det)_A \colon T_A \,\mathrm{GL}(n,\mathbb{R}) \to T_1\mathbb{R}$$

is surjective. This follows from the fact that every matrix X in $T_A \,\mathrm{GL}(n,\mathbb{R}) = \mathsf{Mat}_n(\mathbb{R})$ satisfies:

$$\begin{aligned} D(\det)_A(X) &= \frac{d}{dt}\Big|_{t=0} \det(A+tX) = \frac{d}{dt}\Big|_{t=0} \det[A(\mathsf{id}+tA^{-1}X)] \\ &= \frac{d}{dt}\Big|_{t=0} \det A \det(\mathsf{id}+tA^{-1}X) = \frac{d}{dt}\Big|_{t=0} \det(\mathsf{id}+tA^{-1}X) \\ &= \frac{d}{dt}\Big|_{t=0} [1 + t\mathsf{tr}(A^{-1}X) + O(t^2)] = \mathsf{tr}(A^{-1}X). \end{aligned}$$

Thus, $D(\det)_A(A) = \mathsf{tr}(A^{-1}A) = n$, it follows that $D(\det)_A \neq 0$. One concludes that $\mathrm{SL}(n,\mathbb{R}) = \det^{-1}(1)$ is a smooth submanifold of $\mathrm{GL}(n,\mathbb{R})$ of dimension n^2-1.

- The *orthogonal group* is

$$\varnothing(n) = \{A \in \mathrm{GL}(n,\mathbb{R}) | AA^t = A^tA = \mathsf{id}\}.$$

One can see that the group $\emptyset(n)$ is also a matrix Lie group. It is enough to prove that $\emptyset(n)$ is a smooth submanifold of $\mathrm{GL}(n, \mathbb{R})$. Let us consider the space of symmetric matrices

$$\mathrm{S}(n, \mathbb{R}) = \{M \in \mathsf{Mat}_n(\mathbb{R}) | M^{\mathsf{t}} = M\}$$

which is vector subspace of $\mathsf{Mat}_n(\mathbb{R})$ of dimension $\frac{n(n+1)}{2}$. Consider also the smooth map $F\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{S}(n, \mathbb{R})$ defined by $F(A) = A^{\mathsf{t}}A$. Then $\emptyset(n) = F^{-1}(\mathsf{id})$. We assert that for every matrix $A \in \emptyset(n)$ the linear map

$$DF_A : T_A \, \mathrm{GL}(n, \mathbb{R}) \to T_{F(A)} \, \mathrm{S}(n, \mathbb{R})$$

is surjective. Fix $M \in T_{F(A)} \, \mathrm{S}(n, \mathbb{R}) = \mathrm{S}(n, \mathbb{R})$. Then, for every $X \in T_A \, \mathrm{GL}(n, \mathbb{R}) = \mathsf{Mat}_n(\mathbb{R})$ we have that:

$$\begin{aligned} DF_A(X) &= \left.\frac{d}{dt}\right|_{t=0} F(A + tX) = \left.\frac{d}{dt}\right|_{t=0} (A + tX)^{\mathrm{T}}(A + tX) \\ &= X^{\mathrm{T}}A + A^{\mathrm{T}}X. \end{aligned}$$

This implies

$$DF_A(\tfrac{1}{2}AM) = \tfrac{1}{2}M^{\mathsf{t}}A^{\mathsf{t}}A + \tfrac{1}{2}A^{\mathsf{t}}AM = \tfrac{1}{2}M + \tfrac{1}{2}M = M.$$

One concludes that $\emptyset(n) = F^{-1}(\mathsf{id})$ is a smooth submanifold of $\mathrm{GL}(n, \mathbb{R})$ of dimension $n^2 - n(n+1)/2 = n(n-1)/2$. Note that $\emptyset(n)$ is a closed and bounded subspace of a vector space and therefore, it is compact.

- The *special orthogonal group* is

$$\mathrm{SO}(n) = \emptyset(n) \cap \mathrm{SL}(n, \mathbb{R}).$$

The group $\mathrm{SO}(n)$ is also a matrix Lie group. To see this, consider

$$\mathrm{GL}^{+}(n, \mathbb{R}) = \{A \in \mathsf{Mat}_n(\mathbb{R}) | \det(A) > 0\}.$$

Then $\mathrm{GL}^{+}(n, \mathbb{R}) = \det^{-1}[(0, \infty)]$ which means that it is open in $\mathrm{GL}(n, \mathbb{R})$. On the other hand, if $A \in \emptyset(n)$, then $A^{\mathsf{t}}A = \mathsf{id}$ and therefore

$$1 = \det(A^t A) = (\det(A))^2 \quad \text{which means} \quad \det(A) = \pm 1.$$

This implies that $\mathrm{SO}(n) = \O(n) \cap \mathrm{GL}^+(n, \mathbb{R})$. Then $\mathrm{SO}(n)$ is an open subset of $\O(n)$.

It is easily verified that SO(2) can be parametrized by $\theta \in [0, 2\pi)$ in the following way:

$$\mathrm{SO}(2) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in [0, 2\pi) \right\}.$$

Using this, it is easy to see that SO(2) is abelian. However, $\mathrm{SO}(n)$ is not abelian for $n > 2$.

- The *Lorentz group* is the group of all linear endomorphisms of $\mathbb{R}^4$ that preserve the Minkowski metric

$$\O(3, 1) = \{A \in \mathrm{GL}(4, \mathbb{R}) : \langle Av, Aw \rangle = \langle v, w \rangle \text{ for all } v, w \in \mathbb{R}^4\}.$$

Here the inner product is the one given by the Minkowski metric. It is easy to see that:

$$\O(3, 1) = \{A \in \mathrm{GL}(4, \mathbb{R}) : A^t g A = g\}.$$

where, as usual:

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

8.9 Partitions of Unity

A partition of unity is a technical concept which is very useful in proving existence results in differential geometry. Let $f : M \to \mathbb{R}$ be a smooth function on a manifold. The support of f is the set:

$$\mathsf{supp}(f) = \overline{\{p \in M : f(p) \neq 0\}}.$$

Definition 8.9.1. Let $\{U_\alpha\}$ be an open cover of a manifold M. A partition of unity subordinate to $\{U_\alpha\}$ is a family of smooth functions $\rho_\alpha : M \to \mathbb{R}$ such that:

1. The support of ρ_α is contained in U_α, $\mathsf{supp}(\rho_\alpha) \subseteq U_\alpha$.

2. Each function ρ_α takes only nonnegative values: $\rho_\alpha(p) \geq 0$ for all $p \in M$.

3. For each $p \in M$ there exists an open subset V that contains p such that $\rho_\alpha\big|_V \neq 0$ only for finitely many indices α. Moreover:

$$\sum_\alpha \rho_\alpha(q) = 1$$

for all $q \in V$.

The proof of the following results can be found in most texts in differential geometry, for instance [36].

Theorem 8.9.2. Let $\{U_\alpha\}$ be an open cover of M. There exists a partition of unity ρ_α subordinate to this open cover.

8.10 The Picard-Lindelöf theorem

One of the basic existence and uniqueness theorems for ordinary differential equations is the following:

Theorem 8.10.1. Let $U \in \mathbb{R} \times \mathbb{R}^n$ be an open subset and $F : U \to \mathbb{R}^n$ a smooth function. Given $(t_0, p) \in U$ there exists an open interval (a, b) containing t_0 and a smooth function:

$$y : (a, b) \to \mathbb{R}^n$$

such that $y(t_0) = p$, and $y'(t) = F(t, y(t))$. Moreover, if $z : (c, d) \to \mathbb{R}^n$ satisfies the same equations then z and y coincide in the intersection of their domains.

The Picard-Lindelöf theorem guarantees existence and uniqueness of solutions of first order ordinary differential equations. On the other hand, a higher order differential equation:

$$y^{(k)}(t) = F(t, y(t), \ldots, y^{(k-1)}(t))$$

with initial conditions:

$$y(t_0) = p_1,\ y'(t_0) = p_2, \ldots,\ y^{(k-1)}(t_0) = p_k,$$

can be rewritten as a system of first order equations:

$$\begin{aligned} z_1'(t) &= z_2(t); \\ z_2'(t) &= z_3'(t), \\ &\vdots \\ z_{k-1}'(t) &= F(t, z_1(t), \ldots, z_k(t)). \end{aligned}$$

with initial conditions:

$$z_1(t_0) = p_1,\ z_2(t_0) = p_2, \ldots,\ z_k(t_0) = p_k.$$

Therefore, the Picard-Lindelöf theorem implies the following.

Theorem 8.10.2. Let $U \in \mathbb{R} \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ be an open subset and

$$F : U \to \mathbb{R}^n$$

a smooth function. Given $(t_0, p_1, \ldots, p_k) \in U$ there exists an open interval (a, b) containing t_0 and a smooth function $y : (a, b) \to \mathbb{R}^n$ such that:

$$y(t_0) = p_1,\ y'(t_0) = p_2, \ldots,\ y^{(k-1)}(t_0) = p_k,$$

and

$$y^{(k)}(t) = F(t, y(t), y'^{(k-1)}(t)).$$

Moreover, if $z : (c, d) \to \mathbb{R}^n$ satisfies the same equations then z and y coincide in the intersection of their domains.

Part II

A Brief Introduction to Electromagnetism

Chapter 9

Introduction to Electromagnetism

The ancient Greeks observed that rods of amber, when rubbed with a cat's fur, attracted light objects like little pieces of dry leaves or bits of straw. This observation led them to conjecture the existence of charged particles, divided accordingly into two classes, positive and negative, depending on specific properties of the object being rubbed. Since the Greek word for amber was *elektron*, these forces were later called *electric forces*.

For millennia, electricity would remain little more than a curiosity. It was not until the 17th century that the English scientist William Gilbert made a scientific study of electricity and magnetism. The Latin word *electricus* (like amber) was coined by him to refer to the property of attracting small objects. The words *electric* and *electricity*, on the other hand, appeared for the first time in print in Thomas Brown's Pseudodoxia Epidemica, in 1646.

Magnetism was also known since time immemorial. Centuries before the Greeks, people had noticed that lodestones (pieces of magnetite) could attract certain metals. However, the first analysis of magnetism was due to the philosopher Thales of Miletus, who lived from 625 BC to 545 BC while the first truly scientific study of the phenomenon was carried out for the first time in the 18th century, by Hans Christian Oersted and Andre-Marie Ampere. These scientists observed that when a charged particle moves near a magnet, it becomes subject to forces of a different nature, which they gave the name *magnetic*. Electricity and magnetism (and therefore light) were definitively linked in a series of remarkable equations in 1862, by James Clerk Maxwell, in his work *On Physical Lines of Force*.

9.1 Coulomb's Law

Coulombs's law describes the electric force between two charged particles: A particle of charge q (measured in Coulombs) located at a place $x \in \mathbb{R}^3$ exerts over a particle of charge Q located at $y \in \mathbb{R}^3$ a force (in Newtons) given by:

$$F = \frac{1}{4\pi\epsilon_0}\frac{qQ(y-x)}{|y-x|^3}, \qquad \text{(Coulomb's Law)}$$

where $\epsilon_0 = 8.854 \times 10^{-12}\mathrm{N}^{-1}\mathrm{m}^{-2}\mathrm{C}^2$ is a constant of nature known as the *permittivity of free space.*

The formula above takes the signs of the charges into account: If q and Q have opposite signs, then the force is attractive, while it is repulsive if they are equal. An *electric field* E defined on a region $U \subset \mathbb{R}^3$ is a vector field on U such that a particle of charge q Coulombs located at $x \in U$ is subject to an electric force:

$$F_E(x) = E(x)q.$$

The electric field E is measured in units of Newton per Coulomb (N/C).

Magnetic fields, on the other hand, exert forces on charged particles only if they move. A particle of charge q Coulombs at the place $x \in \mathbb{R}^3$, moving with velocity v, is subject to a magnetic force:

$$F_B = qv \times B(x).$$

The magnetic field B is usually measured in *Teslas*: T = (N × s)/(C × m). A magnetic field of magnitude one Tesla exerts on a particle of charge one Coulomb, moving with speed of one meter per second, a magnetic force of magnitude one Newton. We conclude that a moving charged particle in the presence of electric and magnetic fields is subject to a total force given by Lorentz's Law:

$$F = q(E(x) + v \times B(x)). \tag{9.1}$$

9.2 Electrostatics: Charges at Rest

Suppose there are particles with charges $q_1, \ldots, q_n$ located at fixed positions $x_1, \ldots, x_n$ in Euclidean space. Coulomb's Law implies that the total electric field at any point (y) is given by:

$$E(y) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} \frac{q_i(y - x_i)}{|y - x_i|^3}.$$

In the limit, where the charge is distributed according to a density function $\rho(x)$ which is independent of time, the total amount of charge in a region U can be calculated as

$$Q = \int_U \rho(x)dx.$$

In this case, the total electric field at a point (y) is also given by an integral:

$$E(y) = \frac{1}{4\pi\epsilon_0} \int_U \frac{\rho(x)(y - x)}{|y - x|^3} dx. \tag{9.2}$$

Let us consider a charge q located at a point x inside a region D of space (see figure below). We want to compute the total flux of the electric field across S, the boundary of D. For this, let C be a small ball of radius r centered at x. A simple computation shows that $\mathsf{div}(E) = 0$ outside C. If we set $V = D - C$, By applying Stokes' Theorem one obtains:

$$0 = \int_V \mathsf{div}\ E dV = \int_{\partial D} E \cdot \mathbf{n}\ dA - \int_{\partial C} E \cdot \mathbf{n}\ dA.$$

On the other hand:

$$\int_{\partial C} E \cdot \mathbf{n}\ dA = \frac{q}{4\pi\epsilon_0} \int_{\partial C} \frac{dA}{r^2} = \frac{q}{\epsilon_0}.$$

We then conclude that the total flux across S is proportional to the total charge inside D. By linearity, this also holds for an arbitrary number of charges inside this region. In the limit, where the charge is distributed according to a density function $\rho(x)$, one obtains the *law of Gauss* in integral form:

$$\int_S (E \cdot \mathbf{n}) dA = \frac{1}{\epsilon_0} \int_D \rho(x) dV = \frac{q}{\epsilon_0}. \tag{9.3}$$

Since this formula holds for an arbitrary region one gets that for any ball C:

$$\int_C \mathsf{div}\ E\ dV = \int_{\partial C} (E \cdot \mathbf{n})\ dA = \frac{1}{\epsilon_0} \int_C \rho(x) dV,$$

which implies:

$$\mathsf{div}\ E(x) = \frac{\rho(x)}{\epsilon_0}. \tag{9.4}$$

This equation is known as *Gauss's Law in differential form.*

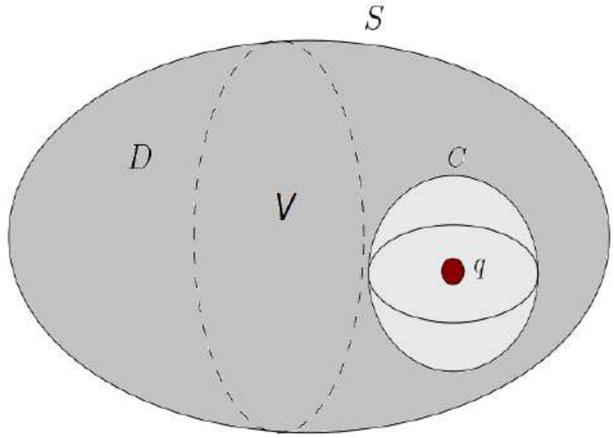


Figure 9.1: A charge in the interior of a region D

9.2.1 Electric Potential

When the electric field E is generated by a time-independent charge density function $\rho(y)$ one can compute E easily in terms of the gradient of a function $\phi(y)$, called the *electric potential*, given by:

$$\phi(y) = \frac{1}{4\pi\epsilon_0}\int_U \frac{\rho(x)}{|y-x|}dx. \tag{9.5}$$

In fact, a simple computation shows that $\mathsf{grad}(\phi)(y) = E(y)$. Since the electric field E generated by static charges is the gradient of a function, we know in this case that $\mathsf{Rot}\ E = 0$. But since $\mathsf{div}(\mathsf{grad}(\phi)) = \nabla^2\phi$, then for a static field the electric potential satisfies the equation:

$$\frac{\partial^2\phi}{\partial x_1^2} + \frac{\partial^2\phi}{\partial x_2^2} + \frac{\partial^2\phi}{\partial x_3^2} = \frac{\rho(x)}{\epsilon_0}, \qquad \text{(Poisson's equation)}$$

known as *Poisson's equation.*

9.3 Electrodynamics: Moving Charges

Moving charges are known as *currents* and can be described mathematically by a vector field

$$J(x,t) = \rho(x,t)v(x,t),$$

where $\rho(x,t)$ measures the density of the flux, and $v(x,t)$, a time-dependent vector field, describes the current's velocity. Consider an oriented surface $S \subset U$. The current passing through S at time t_0 is defined by:

$$I(t_0) = \int_S J(x,t_0) \cdot \mathbf{n} \, dA.$$

It is customary to measure the current in Coulombs per second, a unit known as *Ampere*: 1A= 1C/s. Consider a region $D \subset U$ and define the function $Q(t)$ as the total amount of charge inside D at time t, that is:

$$Q(t) = \int_D \rho(x,t) \; dV.$$

It is a fundamental law of nature that *the total amount of charge is conserved*. Thus, the charge leaving the region D must be equal to the change of $Q(t)$, and therefore:

$$\int_{\partial D} J(x,t_0) \cdot \mathbf{n} \, dA = -\frac{dQ}{dt} = -\int_D \frac{\partial \rho}{\partial t} dV.$$

Stoke's Theorem allows us to compute the integral on the left as $\int_D \mathsf{div}\; J \; dV$. Hence,

$$\int_D \mathsf{div}\; J \; dV = -\int_D \frac{\partial \rho}{\partial t} dV.$$

Since this equation holds for any arbitrary region D, one can write it in differential form as:

$$\mathsf{div}\; J = -\frac{\partial \rho}{\partial t}. \tag{9.6}$$

This equation is known as the *equation of conservation of charge*.

9.4 The Law of Biot-Savart

André-Marie Ampère, Jean-Baptiste Biot, and Félix Savart, among other famous physicists, were the first to give a mathematical description of the magnetic phenomena. By 1820, Oersted had discovered that magnetic fields could be generated by making an electric current circulating through a conductor. The law of Biot-Savart describes the magnetic field induced by a stationary flow of charged particles. The term *stationary* means that the current is a constant function of time, i.e., $J(x,t) = \rho(x)v(x)$.

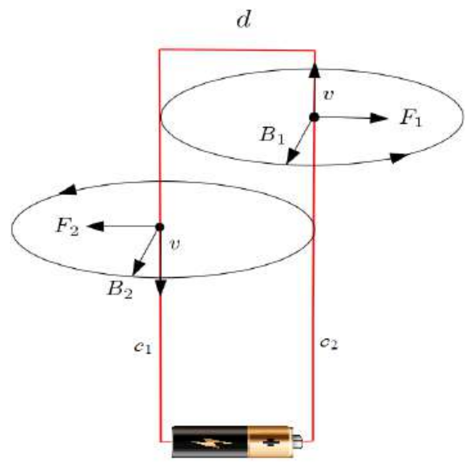


Figure 9.2: Ampere's experiment

We want to start discussing a classic experiment: A wire of constant cross-section A is connected to a battery so that a stationary current is made to circulate through it. Two segments c_1 and c_2 are separated by a distance d, as shown in Figure 9.2. One observes (for small batteries this effect is barely noticeable) that c_1 and c_2 repel each other, indicating the presence of two induced magnetic fields B_1, B_2 which, according to Lorentz' law, would exert a force on the other segment of wire. Once this force is measured, one discovers that its intensity diminishes proportionally to $1/d^2$, where d is the distance between the two cables.

These experiments motivate a general law of nature: Let $J(x)$ be a stationary current that circulates inside some specific region U. We know that the magnetic field B that this current induces at any point p, with coordinates (y), is given at p by the sum of all the contributions $\Delta B(x)$, each corresponding to every small region $\Delta R(x)$ inside U. Notice that the total amount of charge inside $\Delta R(x)$ is equal to $\Delta q(x) = \rho(x)\mathrm{Vol}(\Delta R(x))$, and that each element of field $\Delta B(x)$ must point in the direction of the unit vector $u = (y-x)/\left|y-x\right|$. As noticed by Ampère, it must have a magnitude equal to:

$$|\Delta B(x)| = \mathrm{K}_m \frac{\Delta q(x) \ \ |v \times u|}{|y-x|^2} = \mathrm{K}_m \frac{\Delta q(x) \ \ |v \times (y-x)|}{|y-x|^3},$$

for a suitable constant K_m. For historical reasons one writes K_m as $\mu_0/4\pi$,

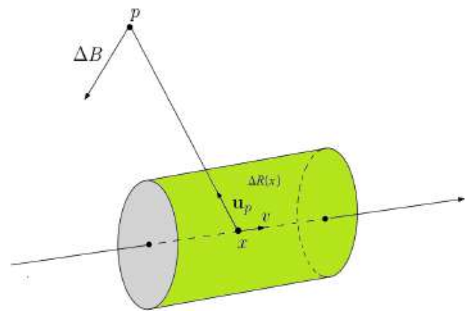


Figure 9.3: Biot Savart's Law

where

$$\mu_0 = 4\pi \times 10^{-7} \frac{\text{Kg} \times \text{m}}{\text{C}^2}$$

is another constant of nature called the *permeability of the vacuum.* Therefore, the natural choice for $\Delta B(x)$ would be:

$$\begin{aligned}\Delta B(x) &= \frac{\mu_0}{4\pi} \frac{\Delta q(x)\ v \times (y-x)}{|y-x|^3} \\ &= \frac{\mu_0}{4\pi} \frac{\rho(x)\ v \times (y-x)}{|y-x|^3} \text{Vol}(\Delta R(x)) \\ &= \frac{\mu_0}{4\pi} J(x) \times (y-x) \frac{\text{Vol}(\Delta R(x))}{|y-x|^3}.\end{aligned}$$

Thus, the sum of all the contributions ΔB inside U is given by:

$$B(y) = \frac{\mu_0}{4\pi} \int_U \frac{J(x) \times (y-x)}{|y-x|^3} dx. \tag{9.7}$$

Equation (9.7) is known as *the law of Biot-Savart* for the magnetic field induced by a stationary current.

9.5 The Divergence of B

We will now see that a magnetic field B induced by a *stationary* current, as described by the law of Biot-Savart, satisfies the equation $\mathsf{div}\ B = 0$. This fact follows from the existence of a *vector potential* for B. Let us define

$$A(y) = \frac{\mu_0}{4\pi} \int_U \frac{J(x)}{|y-x|} dx,$$

An elementary calculation shows that:

$$\mathsf{Rot}\ \frac{J(x)}{|y-x|} = J(x) \times \frac{y-x}{|y-x|^3}.$$

Thus,

$$\mathsf{Rot}\ A(y) = \frac{\mu_0}{4\pi} \int_U \mathsf{Rot}\ \frac{J(x)}{|y-x|} dx = \frac{\mu_0}{4\pi} \int_U J(x) \times \frac{y-x}{|y-x|^3} dx = B(y).$$

Since the divergence of the curl of any vector field is zero, one obtains:

$$\mathsf{div}\ B(y) = \mathsf{div}(\mathsf{Rot}\ A(y)) = 0. \tag{9.8}$$

The universality of this law for any magnetic field is a fundamental law of nature since no one as yet has ever observed a *monopole*, the magnetic equivalent of one charged particle. This leads us to postulate the following equation for any magnetic field:

$$\mathsf{div}\ B = 0. \tag{Maxwell 2}$$

9.6 Magnetostatics

In this section, we will assume that the electric and magnetic fields, the charge density, and the current density, are independent of time. Let us consider a closed circuit C determined by a wire of constant cross-section A and a stationary current I circulating through it. Suppose C' is a closed curve that is linked to C, as shown below: Fix parametrizations $\alpha(t)$, $\beta(s)$ for C and C', respectively. If the current density of I is given by $J = \rho(x)v(x)$ then in a small segment of C of length Δl the element of charge contained in it would be $\Delta q = \rho(x)A\Delta l$, where $x = \alpha(t)$. On the other hand, the total amount of current passing through a cross-section of C at any time is given by

$$I = A\ J \cdot \frac{v(x)}{|v(x)|} = A\rho(x)\,|v(x)|\,.$$

Thus, $\Delta q = I\ \Delta l / |v(x)|$. Using the law of Biot-Savart one can compute the magnetic field J induces at any point p with coordinates $y = y(p)$ as the sum of all fields ΔB, where:

$$\Delta B(p) = \frac{\mu_0}{4\pi |y-x|^3} \Delta q(v(x) \times (y-x)) = \frac{\mu_0 I}{4\pi} \left(\frac{v(x)}{|v(x)|} \times \frac{(y-x)}{|y-x|^3} \right) \Delta l.$$

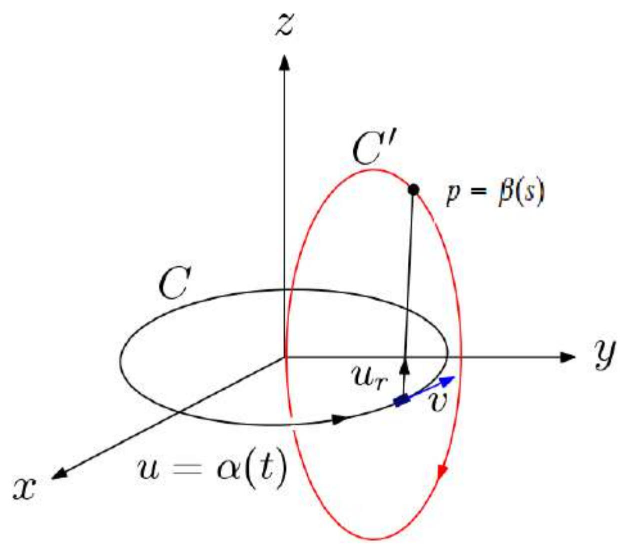


Figure 9.4:

If $p = \beta(s)$ is a point in C' one sees that

$$\begin{aligned}\Delta B(\beta(s)) &= \frac{\mu_0 I}{4\pi}\left(\frac{\alpha'(t)}{|\alpha'(t)|} \times \frac{(\beta(s)-\alpha(t))}{|\beta(s)-\alpha(t)|^3}\right)|\alpha'(t)|\,\Delta t \\ &= \frac{\mu_0 I}{4\pi}\alpha'(t) \times \frac{(\beta(s)-\alpha(t))}{|\beta(s)-\alpha(t)|^3}\Delta t.\end{aligned}$$

From this one obtains:

$$B(\beta(s)) = \frac{\mu_0 I}{4\pi}\int_a^b \alpha'(t) \times \frac{(\beta(s)-\alpha(t))}{|\beta(s)-\alpha(t)|^3}dt. \tag{9.9}$$

One can compute the *circulation of B along* C' as the integral:

$$L = \int_c^d B(\beta(s)) \cdot \beta'(s)\; ds.$$

Using Formula (9.9) for the magnetic field at any p in C' one can write L as follows:

$$L = \frac{\mu_0 I}{4\pi}\int_c^d \int_a^b \alpha'(t) \times \frac{(\beta(s)-\alpha(t))}{|\beta(s)-\alpha(t)|^3} \cdot \beta'(s)\; dt\; ds.$$

Recall that if γ, μ, ν are arbitrary vectors, then:

$$(\gamma \times \mu) \cdot \nu = \det(\gamma, \mu, \nu).$$

From this, we see that the term inside the latter integral is equal to:

$$\frac{\det(\alpha'(t), \beta(s) - \alpha(t), \beta'(s))}{|\beta(s) - \alpha(t)|^3}.$$

Consequently,

$$L = \mu_0 \; I \int_c^d \int_a^b \frac{\det(\alpha'(t),\; \beta(s) - \alpha(t),\; \beta'(s))}{4\pi \, |\beta(s) - \alpha(t)|^3} dt ds.$$

In general, if C and C' are closed 1-manifolds embedded in $\mathbb{R}^3$, the integral

$$\mathsf{Link}(C, C') = \int_c^d \int_a^b \frac{\det(\alpha'(t),\; \beta(s) - \alpha(t),\; \beta'(s))}{4\pi \, |\beta(s) - \alpha(t)|^3} dt ds$$

corresponds to the *linking number* $\mathsf{Link}(C, C')$ of C and C'. This number has interesting topological properties some of which we go into more detail in the Appendix (§**??**). In terms of the linking number we can write the circulation along C' as:

$$L = \mu_0 I \; \mathsf{Link}(C, C').$$

Let S be a surface whose boundary is C'. The flux of current through S is $I = \int_S J \cdot \mathbf{n} \, dA$, where $\mathbf{n}$ denotes the normal exterior vector to S. Let us now assume that C' is a small circle which is *simply linked to* C, i.e., $\mathsf{Link}(C, C') = 1$. Then:

$$L = \int_{C'} B \cdot d\beta = \mu_0 \int_S J \cdot \mathbf{n} \; dA. \tag{9.10}$$

By Stokes' theorem:

$$\int_{C'} B \cdot d\beta = \int_S \mathsf{Rot} \; B \cdot \mathbf{n} \, dA.$$

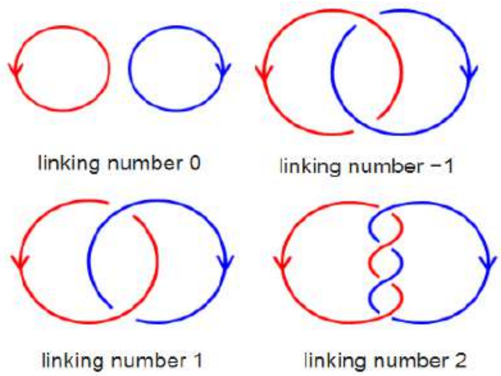


Figure 9.5: (Taken from https://en.wikipedia.org/wiki/Linkingnumber)

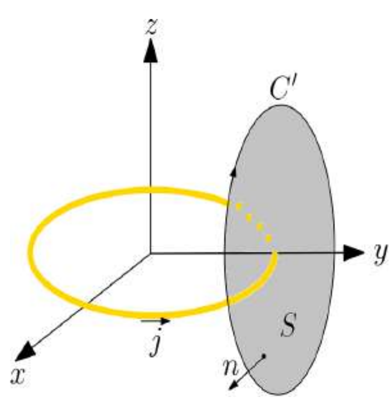


Figure 9.6: (Taken from https://en.wikipedia.org/wiki/Linkingnumber)

Therefore:

$$\int_S \mathsf{Rot}\ B \cdot \mathbf{n}\, dA = \mu_0 \int_S J \cdot \mathbf{n}\, dA. \tag{9.11}$$

Since S is an arbitrary surface we must have that for a stationary current J, the magnetic field B it induces satisfies:

$$\mathsf{Rot}\ B = \mu_0 J. \tag{9.12}$$

This equation is known as *Ampère's law for a stationary current.*

9.7 Varying Fields

One can see directly why Ampere's law cannot hold in general. For this, one can take the divergence on both sides of (9.12) and use (9.6) to obtain:

$$0 = \mathsf{div}(\mathsf{Rot}\ B) = \mu_0 \mathsf{div}\ J = -\mu_0 \frac{\partial \rho}{\partial t},$$

which cannot always be true. To correct this, we may assume that there is an extra field W so that

$$\mathsf{Rot}\ B = \mu_0 J + W.$$

Taking again the divergence on both sides one gets:

$$0 = \mathsf{div}(\mathsf{Rot}\ B) = -\mu_0 \frac{\partial \rho}{\partial t} + \mathsf{div}\ W,$$

and henceforth $\mathsf{div}\ W = \mu_0 \partial\rho/\partial t$. On the other hand, we already know that $\mathsf{div}\ E = \rho/\epsilon_0$ and therefore

$$\mathsf{div}\ \frac{\partial E}{\partial t} = \frac{\partial}{\partial t} \mathsf{div}\ E = \frac{1}{\epsilon_0} \frac{\partial \rho}{\partial t}.$$

Hence, $\mathsf{div}\ W = \mu_0 \epsilon_0 \mathsf{div}\ \partial E/\partial t$, or equivalently

$$\mathsf{div}(W - \mu_0 \epsilon_0 \partial E/\partial t) = 0.$$

One can then postulate, as Maxwell did, that $W = \mu_0 \epsilon_0 \partial E/\partial t$. The more general version of the equation, as corrected by Maxwell, then becomes:

$$\mathsf{Rot}\ B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}. \tag{9.13}$$

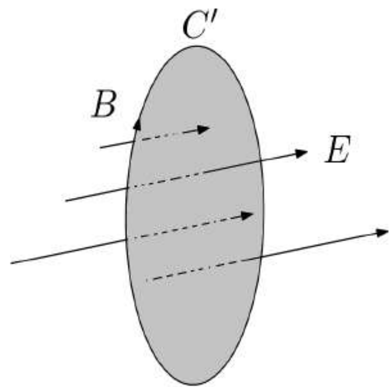


Figure 9.7: The flux of E and the circulation of B

In the absence of currents, Ampere's law states that a time-dependent electric field $E(x,t)$ induces a magnetic field $B(x,t)$ such that if S is a surface with boundary C then:

$$\int_C B \cdot d\beta = \mu_0 \epsilon_0 \frac{d}{dt} \int_S E(x,t) \cdot \mathbf{n}\, dA.$$

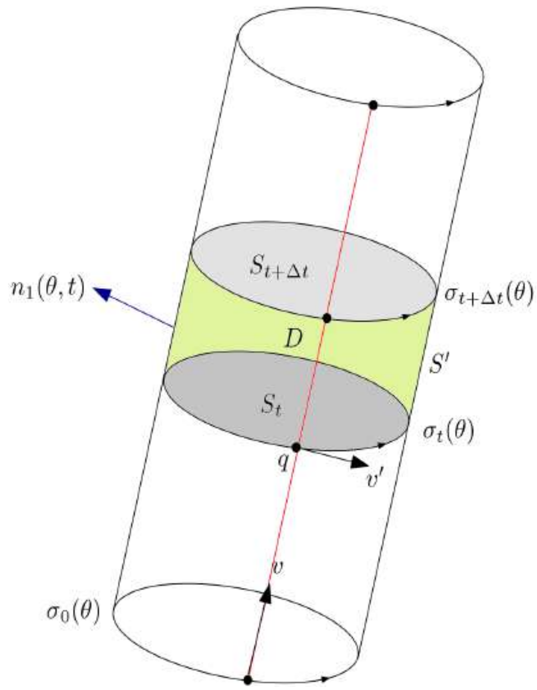
9.8 Faraday's Law of Induction

We have discussed how a stationary current or a variable electric field gives rise to a magnetic field B. In this section we will see that there is a symmetry in this situation: A changing magnetic field must generate an electric field.

Let us consider a closed wire moving at constant velocity v with respect to some reference frame O so that the map $\sigma(t,\theta) = \sigma_0(\theta) + tv$, with $\sigma_0(\theta) = (\cos(2\pi\theta), \sin(2\pi\theta), 0)$, gives its position at time t. Suppose there is a magnetic field B that O regards as static (doesn't change with time). Let us denote by S_t the surface whose boundary is the curve $\sigma_t(\theta) = \sigma(t,\theta)$, and denote by $\phi(t)$ the flux of B along S_t, i.e.,

$$\phi(t) = \int_{S_t} B \cdot \mathbf{n}\, dA.$$

We want to compute the rate of change of $\phi(t)$ with respect to t. Let us first estimate $\phi(t + \Delta t) - \phi(t)$ for a small increment Δt. We know by our



previous discussion that div $B = 0$, hence the Divergence Theorem implies that:

$$0 = \int_D \mathsf{div}\ B\ dV = \phi(t + \Delta t) - \phi(t) + \int_{S'} B \cdot \mathbf{n}_1\, dA,$$

where D is the region between S_t and $S_{t+\Delta t}$, and S' is the lateral part of the boundary. This last integral can be computed as follows:

$$\begin{aligned} B(\sigma_t(\theta)) \cdot \mathbf{n}_1\ dA &= B(\sigma_t(\theta)) \cdot \left(\frac{\partial \sigma}{\partial \theta} \times \frac{\partial \sigma}{\partial t}\right) d\theta\ dt \\ &= \left(\frac{\partial \sigma}{\partial t} \times B(\sigma_t(\theta))\right) \cdot \frac{\partial \sigma}{\partial \theta} d\theta dt \\ &= (v \times B(\sigma_t(\theta))) \cdot \frac{\partial \sigma}{\partial \theta}\ d\theta dt. \end{aligned}$$

Therefore:

$$\int_{S'} B \cdot \mathbf{n}_1\ dA = \int_0^{2\pi} \int_t^{t+\Delta t} (v \times B(\sigma_t(\theta))) \cdot \frac{\partial \sigma}{\partial \theta}\ d\theta\ dt.$$

Since Δt is small we can approximate this integral as:

$$\Delta t \int_0^{2\pi} (v \times B(\sigma_t(\theta))) \cdot \sigma_t'(\theta) d\theta.$$

Thus:

$$\frac{\phi(t+\Delta t) - \phi(t)}{\Delta t} \approx - \int_0^{2\pi} (v \times B(\sigma_t(\theta))) \cdot \sigma_t'(\theta) d\theta.$$

Taking the limit as $\Delta t \to 0$, we can express the negative change in the magnetic flux as the line integral along the curve $C = \sigma_t(\theta)$:

$$\frac{\partial \phi}{\partial t} = - \int_C (v \times B) \cdot d\sigma_t. \tag{9.14}$$

Let us consider a test particle q in the wire with charge 1 Coulomb moving inside with velocity v'. For the observer O, the magnetic field B exerts on the particle a force $F = B \times (v + v')$, according to Lorentz law. The work done by F in making q go once around the wire is given by:

$$W = \int_C F \cdot d\sigma_t = \int_C (v \times B) \cdot d\sigma_t + \int_C (v' \times B) \cdot d\sigma_t. \tag{9.15}$$

Since v' and α' are vectors pointing in the same direction, the second integral in the right-hand side must be zero. Therefore,

$$W = \int_C (v \times B) \cdot d\sigma_t = -\frac{\partial \phi}{\partial t}. \tag{9.16}$$

Let us now analyze the situation from the viewpoint of an observer O' that moves along with the wire. Since the charge q moves with velocity v' in his frame of reference, the magnetic field B (no longer static from O' standpoint) cannot account for the motion of q, since we know $\int_C (v' \times B) \cdot d\sigma_t = 0$, and therefore B cannot do any work on q. Hence W must come from the presence of some electric field E, ie.

$$W = \int_C E \cdot d\sigma_t.$$

We already know that

$$W = -\partial\phi/\partial t = -\frac{d}{dt}\int_S B \cdot \mathbf{n}\, dA,$$

from which we conclude Faraday's law: *That a change in the flux of a magnetic field B across a surface S induces an electric field E satisfying the equation*:

$$\int_C E \cdot d\sigma_t = -\frac{d}{dt}\int_S B \cdot \mathbf{n}\, dA, \tag{9.17}$$

Stokes' theorem allows us to write Faraday's law in differential form as:

$$\mathsf{Rot}\ E = -\frac{\partial B}{\partial t}. \tag{9.18}$$

9.9 Maxwell's Equations

The relations between the electric and magnetic fields discussed above can be summarized in the following equations, known as Maxwell's equations:

$$\begin{aligned}
\mathsf{div}\ E &= \frac{\rho}{\epsilon_0},\\
\mathsf{div}\ B &= 0\\
\mathsf{Rot}\ E &= -\frac{\partial B}{\partial t}\\
\mathsf{Rot}\ B &= \mu_0 J + \epsilon_0\mu_0\frac{\partial E}{\partial t}
\end{aligned}$$

In the particular case where there are no charges, so that $J = \rho = 0$, the equations are known as *Maxwell's equations in the vacuum.*

Let us fix a system of coordinates $u = (x, y, z)$, and look for solutions of the vacuum Maxwell. For simplicity, we assume $E = (0, E_2, 0)$ and $B =$

$(0, 0, B_3)$. In this case, the equations become:

$$\frac{\partial E_2}{\partial y} = 0,$$
$$\frac{\partial B_3}{\partial z} = 0,$$
$$\mathsf{Rot}\ E = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ \partial_1 & \partial_2 & \partial_3, \\ 0 & E_2 & 0 \end{bmatrix} = \left(-\frac{\partial E_2}{\partial z}, 0, \frac{\partial E_2}{\partial x}\right) = \left(0, 0, \frac{-\partial B_3}{\partial t}\right)$$
$$\mathsf{Rot}\ B = \det \begin{bmatrix} e_1 & e_2 & e_3 \\ \partial_1 & \partial_2 & \partial_3 \\ 0 & 0 & B_3 \end{bmatrix} = \left(\frac{\partial B_3}{\partial y}, \frac{-\partial B_3}{\partial x}, 0\right) = \epsilon_0 \mu_0 \left(0, \frac{\partial E_2}{\partial t}, 0\right).$$

We can write these equations as follows:

$$\frac{\partial E_2}{\partial y} = \frac{\partial E_2}{\partial z} = \frac{\partial B_3}{\partial z} = \frac{\partial B_3}{\partial y} = 0, \tag{9.19}$$
$$\frac{\partial E_2}{\partial x} = \frac{-\partial B_3}{\partial t}, \tag{9.20}$$
$$\frac{-\partial B_3}{\partial x} = \epsilon_0 \mu_0 \frac{\partial E_2}{\partial t}. \tag{9.21}$$

We observe that both E_2 and B_3 are functions of x and t, and do not depend on the other variables. Differentiating Equation (9.20) with respect to x one obtains:

$$\frac{\partial^2 E_2}{\partial x^2} = -\frac{\partial^2 B_3}{\partial x \partial t},$$

and differentiating Equation (9.21) with respect to t one gets:

$$-\frac{\partial^2 B_3}{\partial t \partial x} = \epsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2}.$$

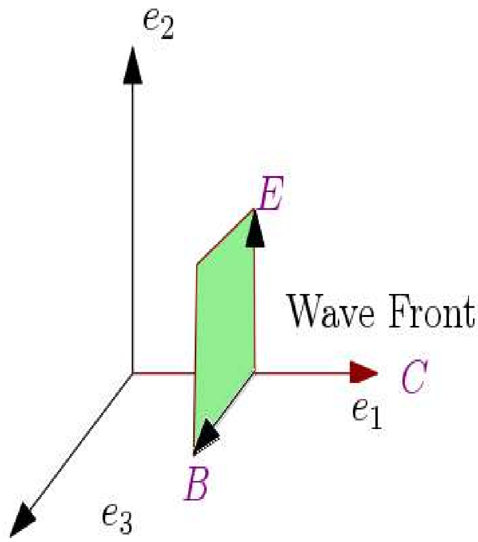
Hence,

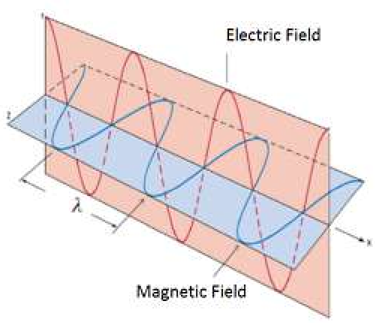
$$\frac{\partial^2 E_2}{\partial x^2} = \frac{1}{\mathrm{c}^2} \frac{\partial^2 E_2}{\partial t^2}, \text{ where } \mathrm{c} = 1/\sqrt{\epsilon_0 \mu_0}. \tag{9.22}$$

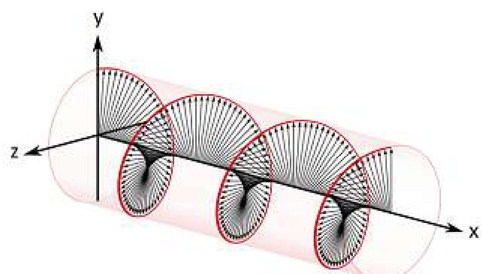
Similarly,

$$\frac{\partial^2 B_3}{\partial x^2} = \frac{1}{\mathrm{c}^2} \frac{\partial^2 B_3}{\partial t^2}. \tag{9.23}$$

This last equation shows that the electromagnetic field satisfies the wave equation of a *plane wave* that moves in the direction of the x-axis at speed







$c = 1/\sqrt{\epsilon_0 \mu_0} \approx 3 \times 10^8$ m/s. It is easy to see that a solution of (9.23) is given by:

$$\begin{aligned} E_2 &= \mathrm{E}_0 \sin(kx - \omega t), \\ B_3 &= -\mathrm{B}_0 \cos(kx - \omega t), \end{aligned}$$

where E_0 and B_0 are constants such that $\mathrm{E}_0 = \mathrm{cB}_0$ and k and ω are such that $\lambda = 2\pi/k$ is the *wavelength,* and $f = \omega/2\pi$ is *the period* in the direction of x of the electromagnetic wave. The general wave equation is derived from Maxwell's equations as follows: For any vector field E, the Laplacian identity states that

$$\mathsf{Rot}(\mathsf{Rot}\ E) = \text{grad}(\mathsf{div}\ E) - \nabla^2 E.$$

Since

$$\mathsf{div}\ E = 0, \text{ and } \mathsf{Rot}\ E = -\frac{\partial B}{\partial t},$$

one concludes that

$$-\frac{\partial}{\partial t}\mathsf{Rot}\ B = -\nabla^2 E.$$

On the other hand,

$$\mathsf{Rot}\ B = \epsilon_0\mu_0\frac{\partial E}{\partial t}.$$

Therefore,

$$-\nabla^2 E + \epsilon_0\mu_0\frac{\partial^2 E}{\partial t^2} = 0.$$

Or equivalently,

$$\nabla^2 E_i = \frac{1}{\mathrm{c}^2}\frac{\partial^2 E_i}{\partial t^2}, \text{ where } E = (E_i). \qquad \text{(Wave Equation)}$$

Similarly, one obtains for the component of the magnetic field:

$$\nabla^2 B_i = \frac{1}{\mathrm{c}^2}\frac{\partial^2 B_i}{\partial t^2}, \text{ where } B = (B_i).$$

One example of a solution of (Wave Equation) corresponds to circularly polarized waves. For such a wave the electric field E rotates in the yz-plane while B stays in the xz-plane. For instance, the fields

$$E = (\mathrm{E}_0\cos(kx - \omega t), \mathrm{E}_0\sin(kx - \omega t), 0),$$

and

$$B = (\mathrm{B}_0\sin(kx - \omega t), 0, \mathrm{B}_0\cos(kx - \omega t)$$

satisfy (Wave Equation).

9.9.1 Maxwell's Equations under Galilean Transformations

Maxwell's equations describe the conditions that electric and magnetic fields must satisfy. Since these fields describe the forces experienced by charged particles, one would expect them to take the same form in any two frames related by a Galilean transformation. We will show in this section that this is not the case. This exhibits a fundamental incompatibility between the laws of electromagnetism and Newtonian mechanics, one of Einsteins' primary motivations for introducing his Special Theory of Relativity.

Let us recall what a Galilean transformation is. Denote by (x) the standard Cartesian coordinate system for an observer O at rest at the origin. If an observer O' moves at constant velocity v with respect to O, the coordinate system $(\overline{x})$ for $\overline{O}$ is related to O's by:

$$\overline{x} = x - vt.$$

We have seen above that a *moving* charge Q induces a magnetic field. However, this is a problematic principle, since the term *moving* is not a well-defined concept. If the charge is moving with respect to an observer O, the principle predicts the existence of both electric and magnetic fields. However, for an observer O' moving along with Q there would only be an electric field present.

Let us look at this situation in more detail. Let us consider a charge Q that moves with constant speed in the direction of the x^1-axis, as measured by an observer O. From O's viewpoint, the charge Q induces a *time-dependent* electric field at any point p with coordinates (x) given by:

$$E(x,t) = \frac{Q}{4\pi\epsilon_0}\frac{x - vt}{|x - vt|^3}.$$

It also induces a time-dependent magnetic field which O measures to be equal to:

$$B(x,t) = \frac{Q\mu_0}{4\pi}\frac{v \times (x - vt)}{|x - vt|^3}.$$

Henceforth, a test particle of charge q which at time $t = 0$ is located at a point p, and with velocity w with respect to O, will experiment a total force (as measured by O) given by (Lorentz's law):

$$F = q(E(p) + w \times B(p)) = \frac{Qq}{4\pi\epsilon_0}\frac{x(p)}{|x(p)|^3} + \frac{\mu_0}{4\pi}\frac{Qq}{|x(p)|^3} w \times v \times x(p). \quad (9.24)$$

We can consider the same situation from the standpoint of $\overline{O}$ for whom the charge Q looks static at the origin. In this description the charge would only generate a (static) electric field given at p by:

$$E(\overline{x}(p)) = \frac{Q}{4\pi\epsilon_0}\frac{\overline{x}(p)}{|\overline{x}(p)|^3}.$$

Hence, for $\overline{O}$, the test particle (moving with velocity $w - v$) would only experience a force due exclusively to this electric field and given by:

$$\overline{F} = \frac{qQ}{4\pi\epsilon_0}\frac{\overline{x}(p)}{|\overline{x}(p)|^3}. \tag{9.25}$$

At $t = 0$ one has $x(p) = \overline{x}(p)$, hence Equation (9.24) differs from (9.25) in one additional term that, in general, does not vanish. Since O' moves with constant velocity with respect to O, both observers must coincide in the measurement of accelerations, and therefore they must measure the same force. This argument shows that Maxwell's equations cannot be invariant under Galilean transformations of coordinates. However, when the magnitudes of v and w are small, this additional term is quite tiny ($\mu_0/4\pi \approx 10^{-7}$) and henceforth F and $\overline{F}$ are almost identical, as it is observed in the classic experiments.

Now, let us discuss the invariance of Maxwell's equations in complete generality. Suppose that with respect to a reference frame O the electric and magnetic fields are represented in these coordinates by vector fields $E(x,t)$, $B(x,t)$. Let us consider another observer $\overline{O}$ moving with constant velocity v with respect to O, let us say in the positive direction of the x^1-axis. A test particle of charge q which moves with velocity w with respect to O will experience a force (as measured by O):

$$F = q(E + w \times B). \tag{9.26}$$

Now, let $\overline{E}(\overline{x},t)$, $\overline{B}(\overline{x},t)$ be the expressions for the electric and magnetic field with respect to the reference frame $\overline{O}$. In these coordinates, the test particle is moving with velocity $w-v$ and therefore experiences a force (as measured by O') given by

$$\overline{F} = q(\overline{E} + (w-v) \times \overline{B}). \tag{9.27}$$

Since O and $\overline{O}$ move with respect to each other at a constant velocity, the force measured by both observers must be the same. In the particular case where $w = v$, the fact that $F = \overline{F}$ forces the equation:

$$\overline{E} = E + v \times B. \tag{9.28}$$

Equating (9.26) and (9.27) one obtains:

$$(w - v) \times B = (w - v) \times \overline{B}.$$

Since w is arbitrary, we conclude that $B = \overline{B}$.

One expects that $\overline{E}$ and $\overline{B}$ should also satisfy Maxwell's equations. In particular,

$$\mathsf{Rot}\ \overline{E} = -\frac{\partial \overline{B}}{\partial t}.$$

Since $\mathsf{Rot}\ E = -\partial B/\partial t$ and $B = \overline{B}$, one must have $\mathsf{Rot}\ \overline{E} = \mathsf{Rot}\ E$. This equation, together with (9.28) implies that:

$$\mathsf{Rot}\ (v \times B) = 0.$$

From this, we get that

$$v \times \mathsf{Rot}\ B = 0. \tag{9.29}$$

Since v is arbitrary, from (9.29) we deduce $\mathsf{Rot}\ B = 0$. Therefore, from Ampere's equation, O would regard any electric field E time-independent, a conclusion that contradicts an empirical fact since some electric fields do change with time. This argument shows that *there is a fundamental inconsistency between Maxwell's equations and the principles of Newtonian mechanics.*

9.10 Maxwell's Equations in Minkowski spacetime

Recall that Minkowski spacetime is the Lorentzian manifold $\mathbb{R}^4$ with the metric:

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

We will show that the vacuum Maxwell equations can be written in terms of the geometry of Minkowski spacetime. This geometric rewriting makes manifest that the equations are preserved by the isometries of Minkowski spacetime.

Consider varying electric and magnetic fields defined in the standard coordinates (x) of $\mathbb{R}^3$ as:

$$E(t,x) = \sum E_i(t,x)\partial_{x_i},\ B(t,x) = \sum B_i(t,x)\partial_{x_i}.$$

We define the *Electromagnetic Tensor* on Minkowski spacetime as the $(0,2)$-tensor F :

$$F = \frac{1}{2}F_{ab}(dx^a \wedge dx^b),$$

where

$$F_{ab} = \begin{bmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{bmatrix}.$$

Explicitly,

$$\begin{aligned} F &= B_1 dx_2 \wedge dx_3 + B_2 dx_3 \wedge dx_1 + B_3 dx_1 \wedge dx_2 \\ &\quad + E_1 dx_1 \wedge dt + E_2 dx_2 \wedge dt + E_3 dx_3 \wedge dt. \end{aligned} \tag{9.30}$$

Let us see that the electric and magnetic fields satisfy Maxwell's equations in the vacuum precisely when F satisfies:

$$dF = d(*F) = 0. \tag{9.31}$$

Here $*$ denotes the Hodge operator (Section 8.7).

First, we observe that:

$$\begin{aligned} *F &= B_1 dt \wedge dx_1 + B_2 dt \wedge dx_2 + B_3 dt \wedge dx_3 \\ &\quad + E_1 dx_2 \wedge dx_3 + E_2 dx_3 \wedge dx_1 + E_3 dx_1 \wedge dx_2. \end{aligned}$$

A straightforward computations shows that $dF = 0$ is equivalent to the equations:

$$-\frac{dB}{dt} = \mathsf{Rot}\ E,\ \mathsf{div}(B) = 0.$$

Also, one can easily check that $d(*F) = 0$ is equivalent to:

$$\frac{dE}{dt} = \mathsf{Rot}\ B,\ \mathsf{div}(E) = 0.$$

Equations (9.31) have been generalized vastly to the context of connections on principal fiber bundles. In this general form, they are known as the *Yang-Mills equations* and play a prominent role in modern physics as well as in geometry.

The fact that Maxwell's equations can be written in terms of the *geometry* of Minkowski spacetime is a fundamental observation. Once the electric and magnetic fields are combined into a two form F in Minkowski spacetime, it is manifest that the equations are invariant under the isometry group of Minkowski spacetime: *The Lorentz group*. Suppose that $(\bar{t}, \overline{x}^i)$ are new coordinates for Minkowski spacetime related to the standard ones by an isometry $\varphi : \mathbb{R}^4 \to \mathbb{R}^4$. Then, in the new coordinates the electromagnetic tensor is given by $\varphi^*(F)$. Assuming that F satisfies (9.31) one can compute:

$$
\begin{aligned}
d\varphi^*(F) &= \varphi^*(dF) = 0;\\
d(*\varphi^*(F)) &= d(\varphi^*(*F)) = \varphi^*(d(*(F))) = 0.
\end{aligned}
$$

Thus, *the equations of electromagnetism are naturally invariant under the symmetries of Minkowski spacetime.* This is an indication of the fact that spacetime has a specific geometry which plays a fundamental role in the laws of physics. We observed before that Maxwell's equations could not be made compatible with Galilean transformations. In contrast, they are manifestly invariant with respect to Lorentz transformations. This was one of the main motivations for Einstein to introduce his Special Theory of Relativity.

Chapter 10

Special Relativity

10.1 The Michelson-Morley Experiment

In classical electrodynamics, a magnet at rest is interpreted quite differently from one in motion, as we discussed in §9.8. Imagine a wire C that moves at constant velocity v with respect to an observer O who measures a *static* magnetic field B (we may think, for instance, that O keeps a steady magnet with him). From O's point of view, any charged particle p moving inside the wire experiences a force due to B (Lorentz Law) that accounts for its motion along C. However, for an observer O' moving along with the wire, it is an electric field E, and not the magnetic field B, the one responsible for the motion of p. From this viewpoint the situation looks somewhat paradoxical: It is as if an electric field is not there for O while O' would claim there must be one present. Later in his life, Einstein described how this realization affected him quite profoundly. In Einstein's own words:

> The idea that these were two, in principle different cases, was unbearable for me. The difference between the two, I was convinced, could only be a difference in choice of viewpoint and not a real difference [...]. Thus the existence of the electric field was a relative one, according to the state of motion of the coordinate system used, and only the electric and magnetic field together could be ascribed a kind of objective reality, apart from the state of motion of the observer or the coordinate system. The phenomenon of magneto-electric induction compelled me to postulate the (special) principle of relativity.

On the other hand, there is a particular feature of the general wave equation (9.23) that did not pass unnoticed to Maxwell and other eminent physicists. It was quite surprising that the speed of the electromagnetic field in vacuum $c = 1/\sqrt{\varepsilon_0 \mu_0}$ was a constant, since this velocity only depended on ε_0 and μ_0, and not on the observer! It was suggested, initially, that this could only be explained by the existence of some natural medium, the *luminous ether*. This unidentified substance, permeating space, was supposed to be some strange fluid that would "vibrate" in the presence of an electromagnetic field so that light would propagate at a constant speed c with respect to an observer at rest in the ether, the same way sound propagates at 340 m/s with respect to the surrounding air. Even if the nature of this mysterious agent would be difficult to establish, the relative motion of the Earth with respect to this medium should be observed. It would be impossible for the Earth to stay at rest with respect to the ether at one particular point in its orbit around the Sun and also at the point it will occupy six months later when it would be traveling in the opposite direction.

In one of the most famous experiments in the history of physics, two American physicists, Albert A. Michelson and Edward Morley, attempted to measure the relative velocity of the Earth with respect to the ether at various points of the Earth's orbit around the Sun. To accomplish this, Michelson and Morley constructed an interferometer (see Figure 10.1). The

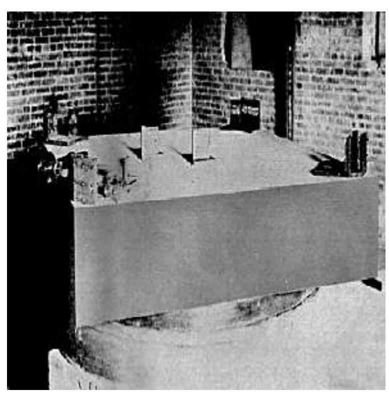


Figure 10.1: Michelson and Morley interferometer

apparatus consisted of a source of light, three mirrors, and an ocular lens. A beam of monochromatic light would split at a central half-silvered mirror into two beams traveling at right angles to equally distant mirrors E_1 and E_2, respectively. The light was then reflected on each mirror, and recombined at

E, where it was then directed towards an observer O, as shown schematically in Figure 10.2. Now, if the whole apparatus moved with respect to the ether

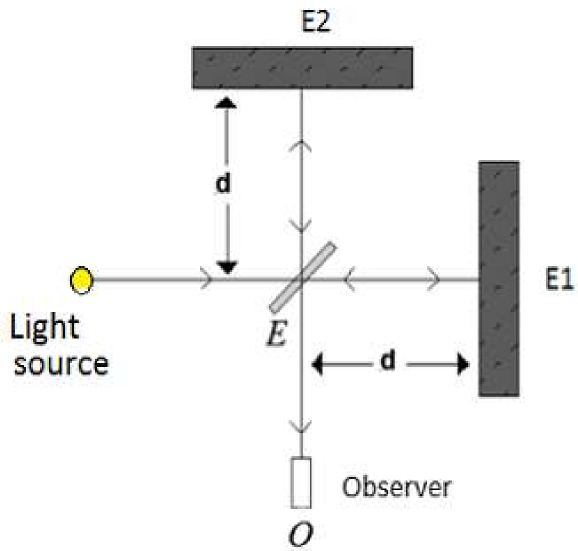
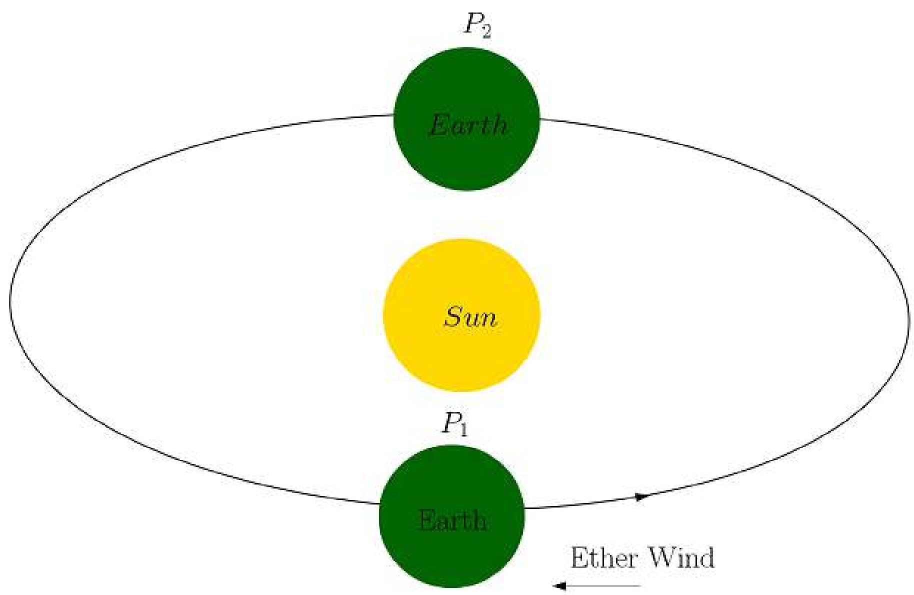


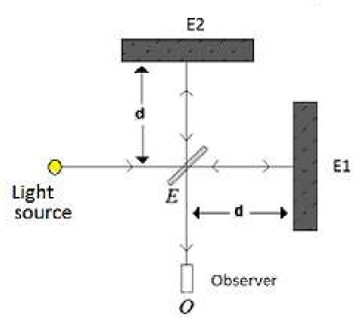
Figure 10.2: Michelson and Morley interferometer

an interference pattern must be observed since the time it would take light to travel both paths (of equal length d) would have to be different as the "ether's wind" would drag any photon traveling towards E_1, making it move at a velocity higher than c.

More precisely, suppose the luminous ether fills empty space (dotted lines in the figure below) and that the solar system moves through this medium at some (unknown) speed ω. If v_1 is the Earth's velocity in its orbit around the Sun (approximately 30 km/s), there must be at least one point in its orbit where the velocity of our planet, relative to the luminous ether, is not zero, since velocities at diametrically opposed points should be $\omega + v_1$, and $\omega - v_1$, respectively.

Let us analyze what would occur at a point P_1 on Earth's orbit where its velocity relative to the luminous ether was not zero. Let us denote this velocity by v. An observer O_p on Earth whose interferometer is oriented in the opposite direction of motion (the segment connecting the light source to the mirror E_1 would point contrary to the Earth's direction of motion) could assume that his laboratory is at rest while the ether would be moving in the

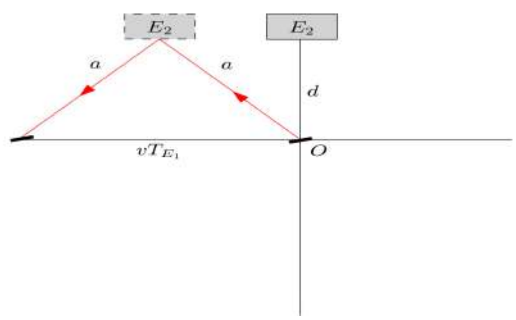

P_2
$Earth$
Sun
P_1
Earth
Ether Wind


E2
d
E1
Light
source
E
d
Observer
O

opposite direction. Then, this "ether wind" would "drag" light coming from the source, so that the velocity of a beam of light traveling towards E_1 would have to be $\mathrm{c} + v$. Henceforth, the time it takes a photon to move from E to E_1 would be $d/(\mathrm{c}+v)$. Similarly, the time it takes the photon to travel "against the flow of ether," i.e., from E_1 to E would be $d/(\mathrm{c}-v)$. Hence, the total time to go from E to E_1 and back must be:

$$T_1 = \frac{d}{\mathrm{c}+v} + \frac{d}{\mathrm{c}-v} = \frac{2\mathrm{c}d}{\mathrm{c}^2 - v^2} = \frac{2d/\mathrm{c}}{1-(v/\mathrm{c})^2}.$$

On the other hand, let us denote by T_2 the time it takes light to go from E to E_2 and back. It is "obvious," at least in classical physics, that this time interval should be the same regardless of the observer (as we shall see, the universality of time is no longer valid in Einstein's theory). Hence, we may compute T_2 from the standpoint of an observer O_e who is at rest relative to the ether. For him, the trajectory of the beam of light from the half-silvered mirror E up to E_2 and back would look like this:



Then, with respect to O_e, each photon should take $T_2 = 2a/\mathrm{c}$ to travel back and forth. On the other hand:

$$a^2 = (\frac{vT_2}{2})^2 + d^2 = \frac{v^2T_2^2}{4} + d^2.$$

Substituting $T_2 = 2a/\mathrm{c}$ into this equation one obtains $a^2 = 4v^2a^2/4\mathrm{c}^2 + d^2$, and solving for a we see that $a = \mathrm{c}d/\sqrt{(\mathrm{c}^2 - v^2)}$. Thus,

$$T_2 = \frac{2\mathrm{c}d}{\mathrm{c}\sqrt{(\mathrm{c}^2 - v^2)}} = \frac{2d/\mathrm{c}}{\sqrt{1 - v^2/\mathrm{c}^2}}.$$

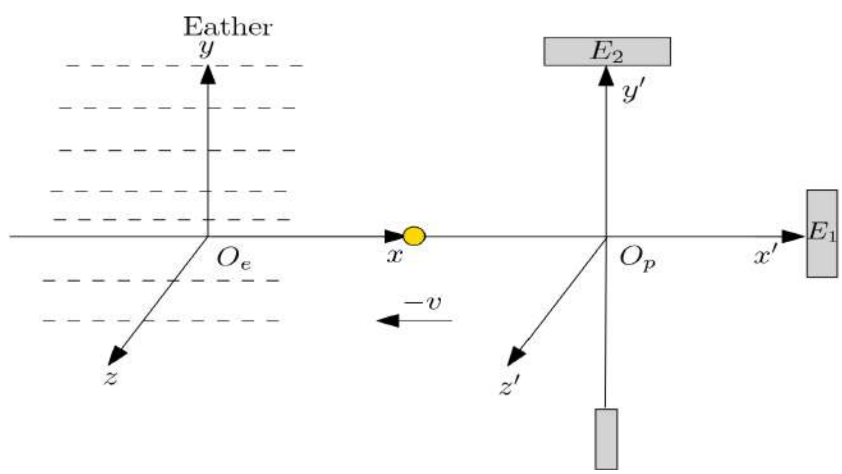
Now, the time difference

$$T_1 - T_2 = \frac{2d/\mathrm{c}}{1 - v^2/\mathrm{c}^2} - \frac{2d/\mathrm{c}}{\sqrt{1 - v^2/\mathrm{c}^2}}$$

(notice that if $v < c$ then $T_1 > T_2$) would manifest itself as a pattern of interference, one which should be observed by O_p. However, this interference pattern *was never observed* by Michelson and Morley, although measurements were performed at different points opposed along the Earth's orbit.

Modern experiments, like that performed by Brillet and Hall in 1978, have corroborated the result of Michelson-Morley on a scale of a tenth of a million ([11], page 48). A law of nature is thus demonstrated: *the constancy of the speed of light, regardless of the observer.* Based on this fundamental fact, Einstein embarked on a seven-year intellectual journey that culminated in his Special Theory of Relativity.

Let us analyze the previous experiment from another point of view. As before, let O_e be an observer at rest with respect to the luminous ether whose spatial coordinates we will denote by $(\overline{x}, \overline{y}, \overline{z})$. Similarly, let (x, y, z) be the frame of reference for O_p, an observer on Earth who moves along with the interferometer. Let us assume both frames of reference coincide at the half-silvered mirror E, at time $t = 0$, and that O_e is moving at constant velocity $u > 0$ in the x-axis with respect to O_p. In classical Galilean mechanics, the relationship between the two frames of reference would be $x = \overline{x} + ut$, $y = \overline{y}$, $z = \overline{z}$. Of course, O_p would be moving at velocity $-u$, with respect to O_e :First, in either frame of reference the speed of a photon that travels



from a point in space $p_1 = (x_1, y_1, z_1)$, at time t_1, to a point $p_2 = (x_2, y_2, z_2)$, at time t_2, would be measured as the quotient

$$\text{speed} = \frac{|p_1 - p_2|}{t_2 - t_1} = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}}{t_2 - t_1}.$$

Now we consider the following sequence of events to which O_p will assign four numbers to each one, a time coordinate followed by the three spatial coordinates:

1. At $t = 0$ a photon starts its journey from E towards E_1. This event has (time) and spatial coordinates $\varepsilon_0 = (0, 0, 0, 0)$.

2. At t_1, that same photon reaches E_1. This event corresponds to the point $\varepsilon_1 = (t_1, d, 0, 0)$.

The Galilean transformations tell us that O_e would record these events in his frame of reference as:

1. $\varepsilon_0 = (0, 0, 0, 0)$

2. $\varepsilon_1 = (t_1, d - ut_1, 0, 0)$

Since for O_e the velocity of light is always c when he measures the velocity of a ray moving for E to E_1 he finds $\mathrm{c} = (d+ut_1)/t_1$, from which $t_1 = d/(\mathrm{c}+u)$, as we had obtained before. However, O_p would measure $d/t_1 = \mathrm{c} + u$ for the speed of that same photon.

Thus, we are compelled to accept the following conclusion: In the absence of an ideal medium, and hence, if no preferred system of coordinates exists, as the Michelson-Morley experiment suggests, *the speed of light measured by any two observers O and O' who move at relative constant velocity should be the same*. This velocity, as we will see, is the maximum possible velocity for any signal in the universe. Moreover, any particle possessing mass must travel at a velocity $< \mathrm{c}$.

Einstein's theory of Special Relativity (SR) provides a correct mathematical frame where this principle holds. Historically, SR was the first fundamental contribution to our modern understanding of time and space.

10.1.1 Geometric Units for Time

In what follows, it is convenient to measure time in appropriated units so that $\mathrm{c} = 1$. Unless we specify otherwise, from now on, we will measure time in a more convenient unit, defined as *the time it takes light to travel a distance of one meter.* In one second light travels $\mathrm{c} \approx 3 \times 10^8$ meters; hence it takes c^{-1} seconds to travel one meter. *Meters* would be the natural units for this measure of time, although we will refer to it (this is by no means standard terminology) as *short seconds*, denoted by ss. Hence, 1s = c ss, consequently, any formula involving time in ss, velocity, acceleration, etc., v, t, a, ...can be transformed in standard units m/s by substituting t by ct, v by v/c, a by a/c^2, etc.

10.2 Lorentz Transformations

Suppose that an observer $\overline{O}$ in Euclidean space moves with constant velocity v with respect to an observer O that stays motionless at the origin, and whose frame of reference is given by the standard Cartesian coordinates (x, y, z) of Euclidean space. In Newtonian mechanics, one assumes that O and $\overline{O}$ share a universal time t and uses the Galilean transformations $\overline{x} = x - tv$ to define $\overline{O}$'s frame of reference. However, we have seen above that this prescription is incompatible with Maxwell's equations. On the other hand, we also observed that Maxwell's equations are preserved under isometries of Minkowski spacetime. This suggests that instead of a Galilean change of coordinates, one should look for a transformation induced by some isometry of Minkowski spacetime. A very simple one, known as *Lorentz boost*, is given by the linear transformation of $\mathbb{R}^4$ defined by the matrix

$$A = \begin{bmatrix} l & lu & 0 & 0 \\ lu & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where $0 < u < 1$ is a scalar. A straightforward computation shows that $A\eta A^* = \eta$ so that A is, in fact, an isometry. The change of coordinates associated to A will then be given by:

$$\begin{aligned} &\bar{t} = l(t - ux),\ \overline{x} = l(x - ut),\ l = \frac{1}{\sqrt{1 - u^2}}, \\ &\overline{y} = y,\ \overline{z} = z. \end{aligned} \tag{10.1}$$

These formulas must be interpreted as the transformation between the frame of reference (t, x, y, z) of the inertial observer O that stays at rest at the origin and the frame $(\bar{t}, \overline{x}, \overline{y}, \overline{z})$ of an observer $\overline{O}$ that moves in the positive direction of O's x-axis at constant speed u. We are assuming both observers carry clocks that are adjusted at time zero when they meet at the origin of coordinates (we will derive 10.1 in the next chapter from a different perspective). We notice that, by symmetry, since O moves with velocity $-u$ with

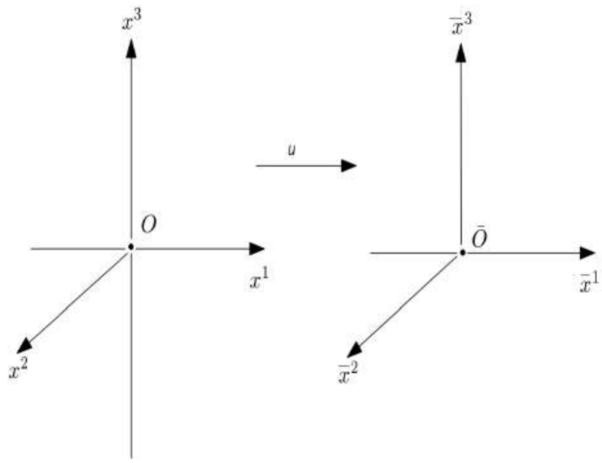


Figure 10.3: Lorentz boost

respect to $\overline{O}$ the inverse transformations should be given by:

$$t = l(\bar{t} + u\overline{x}), \quad x = l(\overline{x} + u\bar{t}), \; l = \frac{1}{\sqrt{1-u^2}}, \tag{10.2}$$
$$y = \overline{y}, \; z = \overline{z}.$$

In many texts, time in measured in seconds (instead of short seconds), and hence in 10.1 one must replace u by u/c and t by ct . Therefore, in classic units 10.1 is written as:

$$\bar{t} = l(t - u/\mathrm{c}^2 x), \; \overline{x} = l(x - ut), \; l = \frac{1}{\sqrt{1-(u/\mathrm{c})^2}}, \tag{10.3}$$
$$\overline{y} = y, \; \overline{z} = z.$$

Similarly,

$$t = l(\bar{t} + u/\mathrm{c}^2 \overline{x}), \; x = l(\overline{x} + u\bar{t}), \; l = \frac{1}{\sqrt{1-(u/\mathrm{c})^2}}, \tag{10.4}$$
$$y = \overline{y}, \; z = \overline{z}.$$

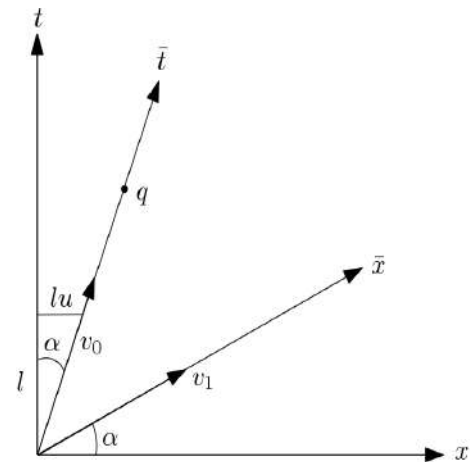


Figure 10.4: Lorentz transformations

If the velocity between the observers is much smaller than the speed of light, $l \approx 1$, the Galilean transformation approximates the spatial part of this Lorentz transformation very closely; hence both observers are led to think that there is a "universal time," as $t \approx \bar{t}$.

The coordinates of O and of $\overline{O}$ can be represented in a two-dimensional diagram that relates (t, x) to $(\bar{t}, \bar{x})$ (we ignore the other two coordinates since they coincide for both observers). Since $v_0 = [l, lu, 0, 0]$ and $v_1 = [lu, l, 0, 0]$, the world line of $\overline{O}$ can be represented by a straight-line at an angle α with respect to the t-axis, where $\tan(\alpha) = lu/l = u$. The coordinates $\bar{t}$ and $\bar{x}$ are therefore determined by the two vectors v_0 and v_1 that form angles α with respect to the t-axis, and the x-axis, respectively: The matrix A can be interpreted as a *generalized rotation in Minkowski spacetime*: Since $l \geq 1$, we can write l as $l = \cosh(r)$, for some unique real number $r \geq 0$. Since $\cosh(r)^2 - 1 = \sinh(r)^2$, then we must have

$$\sinh(r) = \sqrt{l^2 - 1} = u/\sqrt{1 - u^2} = lu.$$

Hence, the matrix A can be written as

$$A = \begin{bmatrix} \cosh(r) & \sinh(r) & 0 & 0 \\ \sinh(r) & \cosh(r) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and the change of coordinates (11.7) is then written as:

$$\begin{bmatrix} \bar{t} \\ \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \begin{bmatrix} \cosh(r) & -\sinh(r) & 0 & 0 \\ -\sinh(r) & \cosh(r) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}.$$

A more general isometry would be a Lorentz boost in the arbitrary direction of a spatial vector $v = (v^i) \in \mathbb{R}^3$, with $|v| < 1$. For any such $v \in \mathbb{R}^3$ we set:

$$A_v = \begin{bmatrix} l & -lv^1 & -lv^2 & -lv^3 \\ -lv^1 & 1 + \frac{(l-1)v^1v^1}{|v|^2} & \frac{(l-1)v^1v^2}{|v|^2} & \frac{(l-1)v^1v^3}{|v|^2} \\ -lv^2 & \frac{(l-1)v^1v^2}{|v|^2} & 1 + \frac{(l-1)v^2v^2}{|v|^2} & \frac{(l-1)v^2v^3}{|v|^2} \\ -lv^3 & \frac{(l-1)v^1v^3}{|v|^2} & \frac{(l-1)v^2v^3}{|v|^2} & 1 + \frac{(l-1)v^3v^3}{|v|^2} \end{bmatrix},$$

where $l = 1/\sqrt{1 - |v|^2}$.

A straightforward computation shows that $A\eta A^* = \eta$ so that A is, in fact, an isometry. Note that a Lorentz boost is only defined for vectors with norm less than $c = 1$. This corresponds to the fact that in Special Relativity it is assumed that observers move at a relative speed which is smaller than the speed of light.

In Special Relativity, the Euclidian geometry of space is replaced by the Lorentzian geometry of Minkowski spacetime. Galilean transformations, which are isometries of Euclidian space, are then replaced by Lorentz boosts which are *rotations* of Minkowski spacetime. It is easy to see that if an observer $\overline{O}$ moves with constant velocity v with respect to O their respective coordinates are related by:

$$\begin{bmatrix} \bar{t} \\ \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = \begin{bmatrix} l & -lv^1 & -lv^2 & -lv^3 \\ -lv^1 & 1 + \frac{(l-1)v^1v^1}{|v|^2} & \frac{(l-1)v^1v^2}{|v|^2} & \frac{(l-1)v^1v^3}{|v|^2} \\ -lv^2 & \frac{(l-1)v^1v^2}{|v|^2} & 1 + \frac{(l-1)v^2v^2}{|v|^2} & \frac{(l-1)v^2v^3}{|v|^2} \\ -lv^3 & \frac{(l-1)v^1v^3}{|v|^2} & \frac{(l-1)v^2v^3}{|v|^2} & 1 + \frac{(l-1)v^3v^3}{|v|^2} \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}.$$

Let us describe in general what the isometries of Minkowski spacetime look like.

A very simple possible isometry is obtained by an overall translation of the form $w' = w + v$, where $v \in \mathbb{R}^4$ is a constant vector. Another more exciting possibility is a change of coordinates of the form $w' = Aw$, where A is some constant matrix. Such a diffeomorphism will be an isometry of Minkowski spacetime precisely when A preserves the inner product in $\mathbb{R}^4$, that is, when $\langle Av, Aw\rangle = \langle v, w\rangle$, for all $v, w \in \mathbb{R}^4$, where the inner product is given by the Minkowski metric. This last condition is equivalent to $A\eta A^* = \eta$. The set of matrices satisfying these equation forms a six-dimensional Lie group called the *Lorentz group*, denoted $O(1,3)$. It turns out that these two types of transformations generate all the isometry group of Minkowski spacetime, as the following exercise shows.

Exercise 10.2.1. Let the *Poincaré group* be the group of diffeomorphisms $\varphi : \mathbb{R}^4 \to \mathbb{R}^4$ of the form:

$$\varphi(w) = Aw + w,$$

where $A \in O(1,3)$ and w is a constant vector. Show that the Poincaré group is the group of all isometries of Minkowski spacetime.

10.3 Perception of Time and Space

Let us consider observers O and $\overline{O}$ as in the previous section. We recall that O represents someone who stays at rest at the origin of coordinates while $\overline{O}$ represents an observer that moves in the positive direction of the x^1-axis of O, at constant speed u. We are assuming that they carry clocks, adjusted at time zero when they meet at the origin of coordinates.

One already knows that the coordinates of $\overline{O}$ are related to those of O by the equations:

$$\begin{aligned}&\bar{t} = l(t - ux),\ \overline{x} = l(x - ut)\\ &\overline{y} = y,\ \overline{z} = z.\end{aligned}$$

It is interesting to analyze how O and $\overline{O}$ differ in their perception of time and space. Consider two events p_0 and p_1 to which O assigns coordinates

$$\begin{aligned}&x(p_0) = 0,\ t(p_0) = t_0\\ &x(p_1) = ut_0,\ t(p_1) = t_0.\end{aligned}$$

We might imagine, for instance, that the event p_0 corresponds to O's clock standing at the origin and flashing on its display "time t_0." Event p_1, on the other hand, corresponds to $\overline{O}$ arriving at a point a on the x^1-axis, with

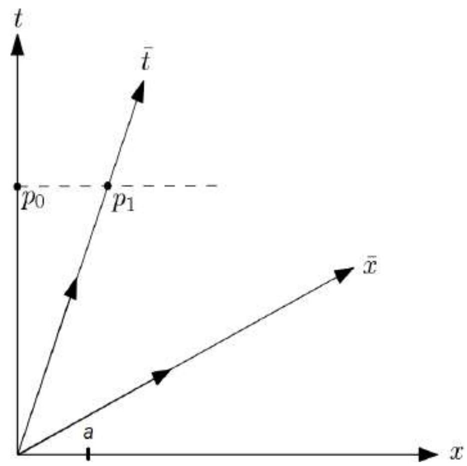


Figure 10.5: Time contraction

coordinate ut_0 (in O's frame of reference). From O's perspective this event occurs precisely when his clock marks time t_0 so that for O, the flashing of his clock and the arrival of $\overline{O}$ to the point a are simultaneous events.

On the other hand, the observer $\overline{O}$ will assign p_0 and p_1 coordinates:

$$\overline{x}(p_0) = -lut_0,\ \overline{t}(p_0) = lt_0;$$

$$\overline{x}(p_1) = 0,\ \overline{t}(p_1) = lt_0(1 - u^2) = \sqrt{1 - u^2}t_0.$$

Hence, the observer $\overline{O}$ will judge that p_1 occurs first than p_0, and therefore these events are not simultaneous from his viewpoint. Moreover, when $\overline{O}$ arrives at a, his watch will display an elapsed time T equal to $\sqrt{1 - u^2}t_0 < t_0$. Thus O will conclude that time is running slower for $\overline{O}$. By symmetry $\overline{O}$ has the same impression regarding O (in the formulas u is replaced by $-u$). One then concludes that *there is no absolute time in Special Relativity and that notions like simultaneity do not have absolute meaning.* Whether or not two events are simultaneous depends on the observer.

Let us now analyze the way in which the two observers O and $\overline{O}$ measure distances. Suppose a bar moves along with $\overline{O}$. This observer has put two flashing lights at the ends of the bar, and they are synchronized such that the two lights keep flashing simultaneously, as perceived by O. The world line of the bar is shown below: The length of the bar as measured by O is d, the difference Δx of the x coordinate corresponding to events $A = (0, 0)$ and $B = (0, d)$. Similarly, for $\overline{O}$, the length of the bar would correspond

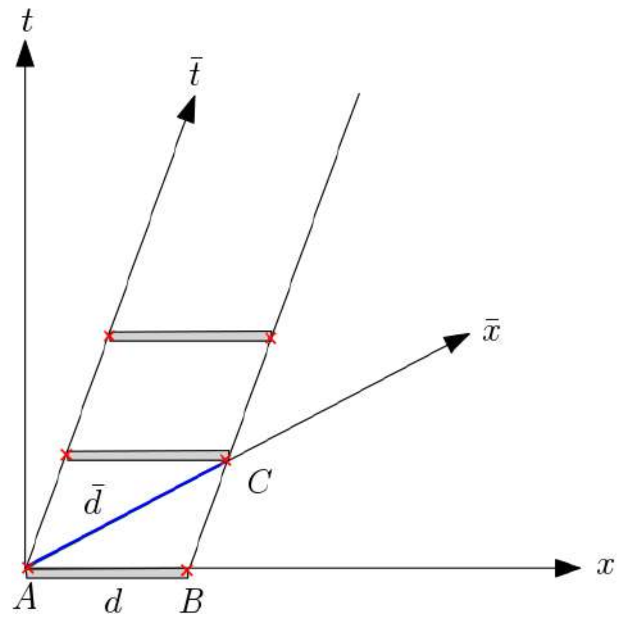


Figure 10.6: Length contraction

to the difference $\overline{d} = \Delta\overline{x}$, equal to the $\overline{x}$-difference between what he sees as simultaneous events, namely $A = (0, 0)$, and $C = (0, \overline{d})$, as measured in his frame of reference. In the frame of reference of $\overline{O}$, B is recorded as the event with coordinates $(-ldu, ld)$. In the figure above we see that the $\overline{x}$-coordinate of B is the same as the $\overline{x}$-coordinate of C, which is $\overline{d}$. Hence, $ld = \overline{d}$ or $d = \sqrt{1 - u^2}\,\overline{d} < \overline{d}$. One concludes that O measures a shorter length for the bar, as compared with the measurements performed by $\overline{O}$.

Although O measures a shorter moving bar, it is quite surprising, however, that *he actually sees it* as if it were as long as the bar $\overline{O}$ measures. To understand this apparent paradox, we have to clarify what we mean by *seeing instead of measuring.*

Imagine we have a large piece of photographic paper that acts as a projection screen. Parallel rays of light coming from an object B far away, impinging the paper perpendicularly, would cast a shadow leaving a photographic image on its surface. *The size of this picture is what we will call the apparent or observed size of* B For definiteness, consider a bar that moves at constant speed u in the xy plane of an inertial observer O, following the line $y = 1$ in the positive direction. We assume it has unit length, as measured by an observer $\overline{O}$ that moves along with it (we already know O would measure

Figure 10.7:

l^{-1} for its length, where, as before, we have set $l = 1/\sqrt{1-u^2}$). Let A and B be the events in flat spacetime corresponding to the emission of two parallel beams of light coming from the tail and the head of the bar, and moving downwards as perceived by $\overline{O}$ A and B have coordinates $\bar{t} = \bar{x} = 0$, $\bar{y} = 1$, and $\bar{t} = 0$, $\bar{x} = 1$, $\bar{y} = 1$, as measured by $\overline{O}$, which in turn correspond in O' s coordinates to $t = x = 0$, $y = 1$, for the event A, and $t = lu$, $x = l$, $y = 1$ for B (10.2). We notice they are not simultaneous events in O's watch. Let C and D label the events corresponding to the arrival of both rays of light at $\overline{O}$'s photographic paper. They both arrive at time $\bar{t} = 1$, with $\bar{x} = \bar{y} = 0$, and $\bar{x} = \bar{y} = 1$, respectively, so that $\overline{O}$ sees a bar of size one unit. For O, the event C has coordinates $t = l$, $x = l$, $y = 0$, and D has coordinates $t = l + lu$, $x = l + ul$, $y = 0$.

On the other hand, suppose O places a photographic plate *of unit length* at an angle $\alpha = \arctan(lu)$ with respect to the x-axis, where the approaching rays then come at an angle $\beta = \pi/2 - \alpha$ with respect to the x-axis (see picture). We claim *both rays reach the opposite sides of the plate at the same time, as measured by* O. In fact, the opposite side of α in the triangle cde has length lu; hence the segment be has length $l - lu$. However, $ac = bd$, since both rays are parallel and $ac = l$, because the speed of light is 1, and we know that the ray on the left arrives at the point c at $t = l$ (the time coordinate of the event C). Hence, the ray on the right takes $l - ul$ ss to reach e. Since it departs at time $t = lu$, it must reach the right side of the plate at time $t = l$, i.e., at the same time, the ray on the left reaches its left side, as claimed. That shows that both O's and $\overline{O}$'s photographs are identical.

If instead of a bar we place a cube with sides parallel to the spatial axis

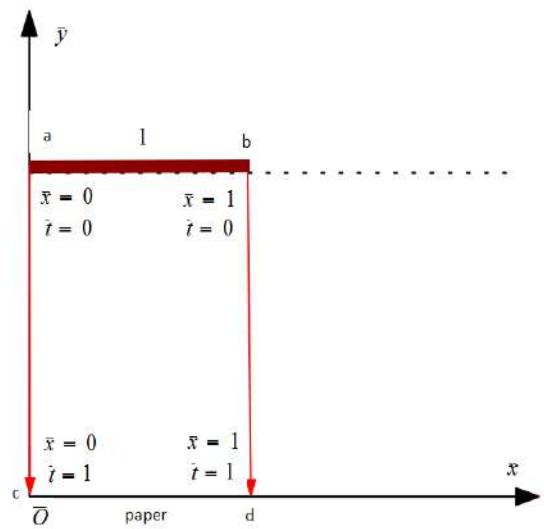


Figure 10.8:

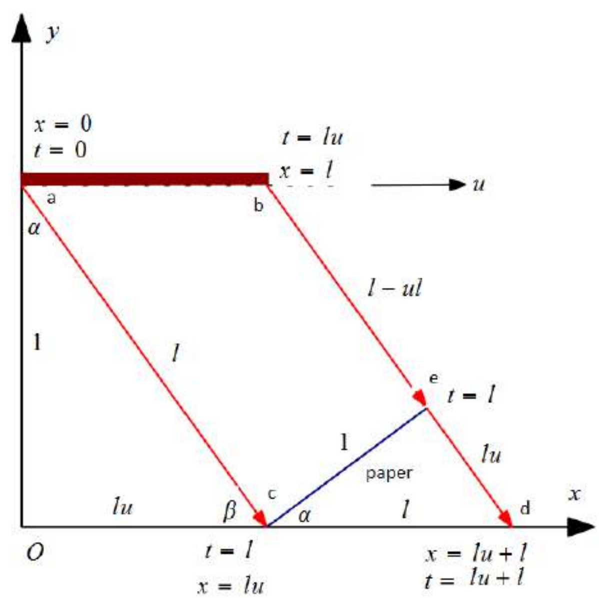


Figure 10.9:

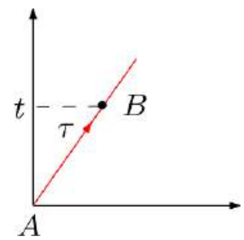


Figure 10.10: Muon Decay

x, y, z the observer O will see that cube as projected onto a plane at angle α (the photographic plate). Thus, O sees the cube the same size as $\overline{O}$'s but *rotated* an angle $\alpha = \arctan(lu)$. This angle is called the *angle of aberration*. This phenomenon implies, for example, that as Earth travels along its orbit, the fixed stars seem to trace small ellipses in the course of a year (see, for instance [41]).

10.3.1 Muon Decay: An Experimental Test for Special Relativity

One of the most dramatic examples of time dilation, as predicted by Einstein's Special Relativity, takes place at the subatomic level.

Muons are particles that decay into neutrinos and electrons after a period that is intrinsic to the particle, called its *lifetime*, which we denote by τ. A particle's lifetime is precisely its proper time between its birth and decay. At rest, the lifetime of a Muon is $\tau \approx 2.2 \times 10^{-6}$ seconds. In a series of famous experiments performed at CERN, in 1970 ([11], Page 65), Muons were accelerated to velocities of the order of $v = 0.9994\text{c}$. For these ultra-rapid Muons, scientists measured a lifetime equal to $t = 64.419 \pm 0.58 \times 10^{-6}$.

To understand this experiment, let A be the event corresponding to the crossing of the particle through the laboratory, and let B the event corresponding to its decay: Then, its lifetime, which is just its proper time τ, is related to t, the time it takes to decay from the standpoint of an observer at rest in the laboratory, with standard coordinates (t, x), is then given by:

$$t = \frac{1}{\sqrt{1 - v^2/\text{c}^2}}\tau \ = 28.871 \times 2.2 \times 10^{-6} = 63.51 \times 10^{-6},$$

a theoretical prediction in perfect agreement with the experiments!

10.4 Observers in Minkowski Spacetime

The observers O and $\overline{O}$ above can be thought of mathematically as a special class of curves $\gamma : I \to \mathbb{R}^4$ on Minkowski spacetime. The image of the curve γ represents the series of events that encompasses the whole history of a material particle, or, as it is usually called, the *worldline* of the observer. The fact that it cannot travel at speed higher than c = 1, the speed of light, is reflected in the fact that one must impose that $\gamma'(s)$ must be a *timelike vector*. If we write γ in the standard coordinates of $\mathbb{R}^4$, $\gamma(s) = (t(s), x^i(s))$, where we shall use $t(s)$ to represent the time coordinate of γ, and Latin superindices x^i, y^i, etc., $i = 1, 2, 3$, to denote the spatial coordinates of γ, by being timelike we mean that

$$\frac{dt}{ds} > \sqrt{\sum_i \left(\frac{dx^i}{ds}\right)^2}.$$

In fact, the observer at rest at the origin (in turn represented by the curve $O(s) = (s, 0, 0, 0)$) will measure the magnitude $|u(t)|$ of the velocity of the observer represented by $\gamma : I \to \mathbb{R}^4$ as:

$$\begin{aligned} |u(t)| &= \sqrt{\sum_i \left(\frac{dx^i}{dt}\right)^2} = \sqrt{\sum_i \left(\frac{dx^i}{ds}\frac{ds}{dt}\right)^2} \\ &= \frac{ds}{dt}\sqrt{\sum_i \left(\frac{dx^i}{ds}\right)^2} < 1 = \mathrm{c}. \end{aligned}$$

Thus, the condition that the tangent vectors to the worldline of an observer are timelike corresponds to the fact that observers do not move faster than the speed of light. Finally, we also demand that γ moves in the "future direction." By this, we mean that $t(s)$ is an increasing function of the parameter.

Definition 10.4.1. By an *observer* in Minkowski spacetime we mean a timelike curve that can be written in standard coordinates as

$$\gamma(s) = (t(s), x^i(s)),$$

with $t'(s) > 0$. The unitary tangent vector $\mathbf{u}(s_0) = \gamma'(s_0)/|\gamma'(s_0)|$ is called its *4-velocity* at $p = \gamma(s_0)$.

We should notice that the 4-velocity is a geometric property. However, the *3-velocity vector* $u(t) = \sum dx^i/dt\, \partial x^i$ *depends on which observer measures it.* The vector $u(t)$ is written in the standard coordinates (t, x^i) of Minkowski spacetime. These coordinates, in turn, are the natural frame of reference of the observer O that stays at rest at the origin.

Notice that being timelike is just the condition $\langle \gamma'(s), \gamma'(s)\rangle < 0$. In case the world line of γ is parametrized by arc length, i.e., if we choose as parameter $s = \int_0^s |\gamma'(\xi)|\, d\xi$, by taking derivatives on both sides one obtains $\langle \gamma'(s), \gamma'(s)\rangle = -1$ and in this case its 4-velocity $\mathbf{u}(s)$ is just the tangent vector $\gamma'(s)$.

10.4.1 Proper Time

Let us consider observers O and $\overline{O}$ as in Section 10.2. Denote by p_0 the event that corresponds to the meeting of O and $\overline{O}$ at the origin. Clearly, this event has null coordinates for both observers. In $\overline{O}$'s own clock the time elapsed between meeting O and arriving at a is the time she measures between p_0 and p_2 which is $T = \sqrt{1-u^2}t_0$. The series of events in between can be parametrized, in O's frame of reference, by $\beta(s) = (s, us, 0, 0)$, $0 \le s \le t_0$. The image of this curve represents the *world-line* of $\overline{O}$ and encompasses her whole history, from the initial meeting with O to her arrival at a. We can compute the total time elapsed in $\overline{O}$'s clock as:

$$T = \int_0^{t_0} |\beta'(s)|\, ds = \int_0^{t_0} \sqrt{1-u^2} ds = t_0\sqrt{1-u^2}.$$

Now, if $\overline{O}(s)$ is an arbitrary observer, and since small line segments can approximate any curve, it is reasonable to define its proper time between events p and q, as given by that same integral as above. More precisely:

Definition 10.4.2. Let $\gamma : I \to \mathbb{R}^4$ be an observer in Minkowski spacetime. The *proper time of* γ between two events $p = \gamma(s_0)$ and $q = \gamma(s_1)$ is defined as:

$$\tau(p, q) = \int_{s_0}^{s_1} |\gamma'(s)|\, ds = \int_{s_0}^{s_1} \sqrt{-\langle \gamma'(s), \gamma'(s)\rangle} ds. \tag{10.5}$$

Geometrically, this integral corresponds to the arc length between p and q in Minkowski spacetime.

We notice that the proper time is independent of the parametrization of the worldline: If $\phi : [c, d] \to [a, b]$ is an orientation-preserving diffeomorphism and if we set $\lambda = \gamma \circ \phi$ then:

$$\begin{aligned}\int_a^b \sqrt{-\langle \lambda'(t), \lambda'(t)\rangle} dt &= \int_a^b \phi'(t)\sqrt{-\langle \gamma'(\phi(t)), \gamma'(\phi(s))\rangle} dt \\ &= \int_a^b \sqrt{-\langle \gamma'(s), \gamma'(s)\rangle} ds = \tau(p, q).\end{aligned}$$

If one writes $\gamma(s) = (t(s), x^i(s))$ in the standard coordinates of Minkowski spacetime, the proper time between two events $p = \gamma(s_0)$ and $q = \gamma(s_1)$ is given by

$$\begin{aligned}\tau(p, q) &= \int_{s_0}^{s_1} \sqrt{\left(\frac{dt}{ds}\right)^2 - \sum_i \left(\frac{dx^i}{ds}\right)^2} ds \\ &= \int_{s_0}^{s_1} \left|\frac{dt}{ds}\right| \sqrt{1 - \sum_i \left(\frac{dx^i}{dt}\right)^2} ds \\ &= \int_{t_0}^{t_1} \sqrt{1 - |u(t)|^2} dt, \qquad (10.6)\end{aligned}$$

where we have changed the variable of integration, with $t_0 = t(s_0)$ and $t_1 = t(s_1)$, and where $u(t) = \sum_i dx^i/dt$ is the 3-velocity of $\overline{O}$ at time t.

10.4.2 Velocities under Lorentz Transformations

In Example 11.2.10 we found the relationship between the frames of reference of two observers O and $\overline{O}$ in Minkowski spacetime, where $\overline{O}$ moves in the direction of the x-axis of O at constant speed u. If these coordinates are (t, x^i) and $(\bar{t}, \overline{x}^i)$, respectively, then:

$$\begin{bmatrix} l & lu & 0 & 0 \\ lu & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{t} \\ \overline{x}^1 \\ \overline{x}^2 \\ \overline{x}^3 \end{bmatrix} = \begin{bmatrix} t \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} \qquad (10.7)$$

where $l = 1/\sqrt{1-u^2}$.

Let $\alpha : I \to \mathbb{R}^4$ be a timelike or a null curve that describes the world line of a particle P. Write

$$\alpha^0(s) = t(\alpha(s)),\ \alpha^i(s) = x^i(\alpha(s))$$

in O's frame of reference, and

$$\overline{\alpha}^0(s) = \overline{t}(a(s)),\ \overline{\alpha}^i(s) = \overline{x}^i(\alpha(s))$$

in $\overline{O}$'s frame. Hence, the 3-velocity of P at $s = s_0$, as measured by O, is given by $v = \sum_i v^i \partial x^i$, where

$$v^i = d\alpha^i/d\alpha^0\big|_{s=s_0} = (d\alpha^i/ds)/(d\alpha^0/ds)\big|_{s=s_0}.$$

Similarly, the 3-velocity of P measured by $\overline{O}$ would be: $\overline{v} = \sum_i \overline{v}^i \partial_{\overline{x}^i}$, with $\overline{v}^i = d\overline{\alpha}^i/d\overline{\alpha}^0\big|_{s=s_0} = (d\overline{\alpha}^i/ds)/(d\overline{\alpha}^0/ds)\big|_{s=s_0}$.On the other hand, equation (10.7) says that

$$\begin{bmatrix} l & lu & 0 & 0 \\ lu & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} d\overline{\alpha}^0/ds \\ d\overline{\alpha}^1/ds \\ d\overline{\alpha}^2/ds \\ d\overline{\alpha}^3/ds \end{bmatrix} = \begin{bmatrix} d\alpha^0/ds \\ d\alpha^1/ds \\ d\alpha^2/ds \\ d\alpha^3/ds \end{bmatrix}$$

Henceforth,

$$d\alpha^0/ds = l(d\overline{\alpha}^0/ds) + lu(d\overline{\alpha}^1/ds).$$

Also,

$$d\alpha^1/ds = lu(d\overline{\alpha}^0/ds) + l(d\overline{\alpha}^1/ds),$$

and consequently

$$v^1 = d\alpha^1/d\alpha^0 = \frac{d\alpha^1/ds}{d\alpha^0/ds} = \frac{lu(d\overline{\alpha}^0/ds) + l(d\overline{\alpha}^1/ds)}{l(d\overline{\alpha}^0/ds) + lu(d\overline{\alpha}^1/ds)}.$$

Dividing each term by $l(d\overline{\alpha}^0/ds)$ one obtains

$$v^1 = \frac{u + d\overline{\alpha}^1/d\overline{\alpha}^0}{1 + u(d\overline{\alpha}^1/d\overline{\alpha}^0)} = \frac{u + \overline{v}^1}{1 + u\overline{v}^1}. \tag{10.8}$$

Similarly one gets

$$v^2 = d\alpha^2/d\alpha^0 = \frac{d\alpha^2/ds}{d\alpha^0/ds} = \frac{d\overline{\alpha}^2/ds}{l(d\overline{\alpha}^0/ds) + lu(d\overline{\alpha}^1/ds)}$$
$$= \frac{d\overline{\alpha}^2/d\overline{\alpha}^0}{l(1 + u\ d\overline{\alpha}^1/d\overline{\alpha}^0)} = \frac{\overline{v}^2}{l(1 + u\overline{v}^1)}, \tag{10.9}$$

$$v^3 = d\alpha^3/d\alpha^0 = \frac{d\overline{\alpha}^3/d\overline{\alpha}^0}{l(1 + u\ d\overline{\alpha}^1/d\overline{\alpha}^0)} = \frac{\overline{v}^3}{l(1 + u\overline{v}^1)}. \tag{10.10}$$

Formulas (10.8,10.9 and 10.10) give the relationship between the components of the velocity of P as measured by O and $\overline{O}$.

In standard units of time (seconds) these formulas read

$$v^1 = \frac{u + \overline{v}^1}{1 + (u/\mathrm{c}^2)\overline{v}^1} \tag{10.11}$$
$$v^2 = \frac{\overline{v}^2}{L(1 + (u/\mathrm{c}^2)\overline{v}^1)},$$
$$v^3 = \frac{\overline{v}^3}{L(1 + (u/\mathrm{c}^2)\overline{v}^1)},$$
$$\text{with } L = 1/\sqrt{1 - (u/\mathrm{c})^2}.$$

We see that at non-relativistic velocities, that is if $u << \mathrm{c}$ one has $(u/\mathrm{c})^2 \approx 0$, and $L \approx 1$. Formulas (10.11) give the classical Galilean composition of velocities.

10.5 4-Momentum, 4-Force and Proper Acceleration

Associated to any observer $\gamma(s) = (t(s), x^i(s))$ there is a positive scalar m_0 called its *rest mass*. If we think of γ as representing the worldline of a particle P, we define its *4-momentum* at $p = \gamma(s)$ as $\mathbf{p}(s) = m_0\mathbf{u}(s)$.

Suppose γ is parametrized by arc length. *The proper acceleration of γ at p* is defined as the vector $\mathbf{a}(s) = \gamma''(s)$. In case γ corresponds to a particle with rest mass m_0, the *4-force* that account for such acceleration is defined as $\mathbf{f}(s) = m_0\mathbf{a}(s)$.

We notice that $\mathbf{f}(s) = \mathbf{p}'(s)$ is the change in momentum, as one would expect. On the other hand, since γ is assumed to be parametrize by arc length one has $|\gamma'(s)| = 1$ and therefore $\langle \gamma'(s), \gamma''(s) \rangle = 0$. Thus, $\mathbf{a}(s)$ $\mathbf{u}(s)$ are orthogonal vectors. On the other hand, $\langle \gamma'(s), \gamma'(s) \rangle = 1$ implies that

$$t'(s) = l(s) = 1/\sqrt{1 - |u(t(s))|^2}, \tag{10.12}$$

where $|u(t)|$ is the magnitude of the 3-velocity $u(t) = \sum dx^i/dt\ \partial x^i$ of P. As we noticed above, the 4-velocity is an intrinsic property while $u(t)$ is the 3-velocity measured by O, the observer at rest at the origin.

In O's frame of reference, the 4-momentum at q can be written as

$$\begin{aligned} \mathbf{p}(s) &= m_0 t'(s)\partial_t + m_0 \textstyle\sum_i \frac{dx^i}{dt}\frac{dt}{ds}\ \partial x_i \\ &= m_0 t'(s)(\partial_t + \textstyle\sum_i \frac{dx^i}{dt}\partial x_i) \\ &= m(s)\partial_t + m(s)u(t(s)), \end{aligned}$$

where $m(s) = m_0 l(s)$ corresponds to the *relativistic mass of* P. We shall see later that $m(s)$ corresponds to the actual mass that O would measure for the particle P. The 4-force acting on P at p is then given by

$$\mathbf{f}(s) = m'(s)\partial_t + \frac{d}{ds}(m(s)u(t(s))). \tag{10.13}$$

Then, the *spatial* force or 3-force acting on P, *as measured by* O, is equal to

$$f(s) = \frac{d}{ds}(m(s)u(t(s)) = \frac{d(m(s)u(t(s))}{dt}\frac{dt}{ds} = l(s)\ \frac{d(m(s)u(t(s))}{dt}. \tag{10.14}$$

When $u << 1$ one has that $l \approx 1$, $m \approx m_0$, $t(s) \approx s$ and henceforth

$$f(s) \approx m_0 \frac{du}{dt} = m_0 a(t),$$

which is Newton´s second law. Thus, *the spatial force coincides in the classic limit with the Newtonian force.*

From (10.13) and (10.14) we see that

$$\begin{aligned} 0 = \langle \mathbf{f}(s), \mathbf{u}(s) \rangle &= \langle m'(s)\partial_t + f(s), t'(s)(\partial_t + u) \rangle = \\ &= t'(s)(m'(s)\partial_t + f(s),\ \partial_t + u) = \\ &= t'(s)(-m'(s) + \langle f(s), u \rangle), \end{aligned}$$

from which one gets $m'(s) = \langle f(s), u \rangle$, and consequently that

$$\begin{aligned} \mathbf{f}(s) &= m'(s)\partial_t + \frac{d}{ds}(m(s)u(t(s)) \\ &= \langle f(s), u(t(s)) \rangle\, \partial_t + f(s). \end{aligned} \tag{10.15}$$

10.6 Maxwell's Equations Revisited

In Section 9.1 we discussed the force experienced by a moving charged particle in an electromagnetic field. Classically, the trajectory $\alpha(t) = (\alpha^i(t))$ of a particle P of charge q and mass m moving in this field satisfies the equation

$$qE(\alpha(t)) + q\alpha'(t) \times B(\alpha(t)) = m\alpha''(t). \tag{10.16}$$

Here, the left-hand side is the force described by Lorentz law, and the right-hand side is Newton's law, mass times acceleration.

We want to find the 4-force in Minkowski spacetime acting on P so that its corresponding 3-force, as measured by an observer O at rest at the origin of coordinates, corresponds to the left-hand side of (10.16).

Let us denote by $u(t) = \sum_i d\alpha^i/dt\ \partial x^i$ the 3-velocity of P measured by O, the observer at rest at the origin of coordinates. As we shall see later, the mass of P, as measured by O, depends on $|u(t)|$, and is equal to $m(t) = m_0 l(t)$, where m_0 is the inertial mass of P, and, as above, $l(t) = /\sqrt{1 - |u(t)|^2}$. On the other hand, *the 3-force acting on P must correspond the rate of change of momentum $d/dt(m(t)u(t))$*. Hence the relativistic form of Lorentz law (10.16) must be

$$qE(\alpha(t)) + qu(t) \times B(\alpha(t)) = \frac{d}{dt}(m(t)l(t)). \tag{10.17}$$

Let $\gamma(s) = (t(s), x^i(s))$ be the worldline of P in Minkowski spacetime, where we now assume γ is parametrized by arc length. As we observed above (10.12) γ being parametrized by arc length implies that $dt/ds = 1/\sqrt{1 - |u(t(s))|^2}$. Multiplying both sides of (10.17) by $l(t)$ one obtains

$$ql(t)\left[E(\alpha(t)) + u(t) \times B(\alpha(t))\right] = l(t)\ \frac{d(m(t)u(t)}{dt}$$

or equivalently

$$ql(t(s))\left[qE(x^i(s)) + u(t(s)) \times B(x^i(s))\right] = l(t(s))\ \frac{d(m(t(s))u(t(s))}{dt},$$

Since $\alpha^i(t(s)) = x^i(s)$. By (10.14), the right-hand side of the latter equation is precisely the 3-force $f(s)$ exerted on P, as measured by O. As we already know (10.15), the 4-force acting on P must be given by :

$$\mathbf{f}(s) = \langle f(s), u(t(s)) \rangle\, \partial_t + f(s).$$

from which one readily sees that

$$\mathbf{f}(s) = ql(s)\left[\langle E(\gamma(s), u(t(s)\rangle \partial_t + E(\gamma(s)) + u(t(s)) \times B(\gamma(s))\right], \tag{10.18}$$

If we write the 4-velocity as $\mathbf{u}(s) = t'(s)\partial_t + v(s)$, where $v(s)$ corresponds to the spatial part of the 4-velocity, using the chain rule one gets

$$\begin{aligned} v(s) &= \sum \frac{dx^i(s)}{ds}\partial x^i = \sum \frac{d\alpha^i(t(s))}{ds}\partial x^i \\ &= \sum \frac{dt}{ds}\frac{d\alpha^i(t)}{dt}\partial x^i = l(s)u(t). \end{aligned}$$

Henceforth, we may write (10.18) as

$$\mathbf{f}(s) = q\left\langle E(\gamma(s), v(s)\right\rangle \partial_t + ql(s)E(\gamma(s)) + qv(s) \times B(\gamma(s))$$

and the equation of motion of P would be

$$q\left\langle E(\gamma(s), v(s)\right\rangle \partial_t + ql(s)E(\gamma(s)) + qv(s) \times B(\gamma(s)) = m_0\gamma''(s). \tag{10.19}$$

Equation 10.19 can be written elegantly in terms of the Electromagnetic tensor (9.30)

$$\begin{aligned} F = B_1 dx_2 \wedge dx_3 + B_2 dx_3 \wedge dx_1 + B_3 dx_1 \wedge dx_2 \\ + E_1 dx_1 \wedge dt + E_2 dx_2 \wedge dt + E_3 dx_3 \wedge dt. \end{aligned}$$

for a particle P of charge q and rest mass m_0 the relativistic equation of motion can be written as:

$$-qF(\gamma'(s), X) = m_0\langle \gamma''(s), X\rangle, \tag{10.20}$$

for any vector $X \in T_{\gamma(t)}\mathbb{R}^4 \simeq \mathbb{R}^4$. Note that Equation (10.20) merely states that the covector $-qF(\gamma'(s), -)$ corresponds to the vector $m_0\gamma''(s)$ under the identification between the tangent and cotangent spaces induced by the Minkowski metric. On the vector $X = \partial x_i$, for instance, the left hand side of the equation is:

$$\begin{aligned} -qF(\gamma'(s), \frac{d}{dx_i}) &= -qF\left(t'(s)\partial t,\ \partial x^i\right) - q\sum_{j\neq i} F\left(\frac{dx^j}{ds}\partial x^j,\ \partial x^i\right) \quad (10.21) \\ &= ql(s)E_i(\gamma(s)) - ql(s)\sum_{jk} \epsilon_{kji}\frac{dx^j}{ds}B_k(\gamma(s)) \\ &= ql(s)\left[E_i(\gamma(s)) - u_i(t(s)) \times B_k(\gamma(s))\right]. \end{aligned}$$

(Here, ϵ_{kji} is equal to 1, if kji is an even permutation, equal to -1 if it is an odd permutation and 0 otherwise, and $u_i(t(s)) = dx^i/ds\, \partial x^i$ denotes the i-component of the 3-velocity of P, as measured by O.)

Notice that Equation 10.21 and Equation 10.19 coincide.

One expects that the equation of motion reduces to the classical one for velocities much smaller than the speed of light. Let us assume that $v << \mathrm{c} = 1$ so that $l \approx 1$. In this case

$$qF(\gamma'(s), \partial x^i) \approx qE_i - q\sum_{jk} \epsilon_{kji} \frac{dx^j}{dt} B_k,$$

which is the i-th component of the force according to Lorentz law. On the other hand, evaluating the right hand side on the vector $X = \partial x^i$ one obtains:

$$\begin{aligned} m_0\langle \gamma''(s), \partial x^i \rangle &= m_0 \langle \frac{d}{ds}\frac{dx^i}{ds}\partial x^i, \partial x^i \rangle \\ &= m_0 \frac{d}{ds}\frac{dx^i}{ds} \\ &= m_0 \frac{d}{ds}\left(\frac{dx^i}{dt} l(s)\right) \\ &= m_0 l \frac{d}{dt}\left(\frac{dx^i}{dt} l(s)\right) \\ &= m_0 l \left(\frac{d^2x^i}{dt^2} l(t(s) + \frac{dx^i}{dt}\frac{dl(t(s))}{dt}\right). \end{aligned}$$

When $l \approx 1$, Equation (10.16) becomes

$$qE_i - q\sum_{jk} \epsilon_{kji} \frac{dx^j}{dt} B_k = m_0 \langle \gamma''(s), \partial x^i \rangle \approx m_0 \frac{d^2x^i}{dt^2}.$$

We conclude that in the limit of small velocities, Equation (10.20) reduces to the classical equation of motion. It is worth mentioning that the relativistic equation of motion is manifestly coordinate independent. Indeed, it is a tensor equation which is written without reference to any coordinate system. Also, the electromagnetic field F is a two form on Minkowski spacetime and therefore transforms naturally with respect to changes of coordinates.

10.7 Equivalence of Mass and Energy

Let's start by analyzing the collision of two identical spheres B and $\overline{B}$ in flat space-time. We assume that associated to any particle P whose world-line is *timelike* there is a nonzero scalar called its *rest mass*, that we measure in kg, and that we will denote by $m_0(P)$. In relativistic mechanics the *total mass* of P is a scalar that depends on the observer, and that can be identified with the *total energy* of the particle, as measured by that particular observer. A precise definition can be given after we introduce the notion of 4-momentum. In this section, however, we will refer to the mass of a particle as a scalar $m(P)$ determined by each inertial observer, and which must coincide with $m_0(P)$ for all such observers when the particle is seen to be at rest.

The purpose of the thought experiment we will discuss next is to determine the mass that an inertial observer O would measure on a particle that moves along the x-axis at constant speed u. We will assuming the conservation of classical momentum when no energy is lost (elastic collisions).

Let's set up two inertial frames of reference: The first, denoted by (t, x, y, z), corresponds to an observer O for whom B stays at rest, i.e., at the origin of his coordinate system. The second, $(\bar{t}, \overline{x}, \overline{y}, \overline{z})$,corresponds to an observer $\overline{O}$ that moves in the direction of the positive x-axis at constant speed u, and one for which the particle $\overline{B}$ stays at rest in $\overline{O}$'s frame of reference: Suppose

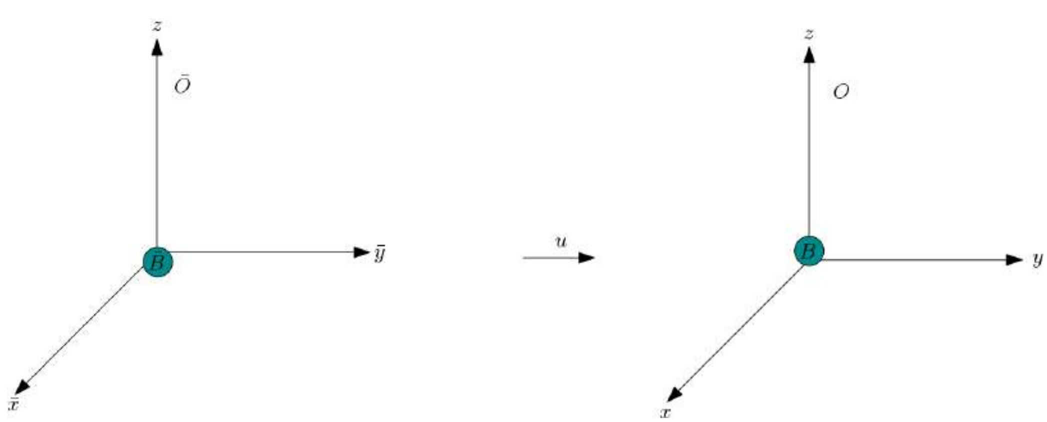


Figure 10.11: Collisions

that after B and $\overline{B}$ collide the two particles move in the x-z plane: By the principle of relativity the magnitude of the z-component of the velocities of B and $\overline{B}$ should be the same as measured by O, and by $\overline{O}$, respectively. Denote

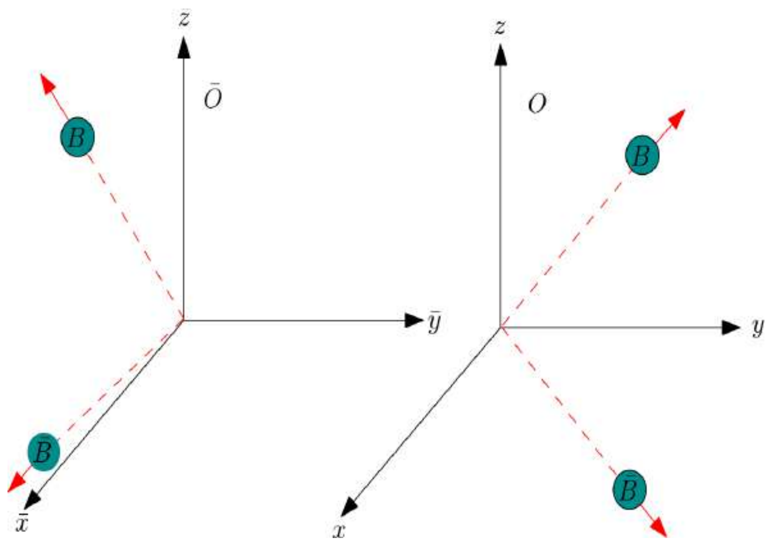


Figure 10.12: Collisions

these quantities by $v^3(B)$ and $\overline{v}^3(\overline{B})$, respectively. Hence, $v^3(B) = \overline{v}^3(\overline{B})$. On the other hand, by Formula (10.10) ones has:

$$v^3(\overline{B}) = \frac{\overline{v}^3(\overline{B})}{L(1 + (\mathrm{u}/\mathrm{c}^2)\overline{v}^1(\overline{B}))}, \text{ with } L = 1/\sqrt{1 - (\mathrm{u}/\mathrm{c})^2},$$

where $\overline{v}^1(\overline{B})$ denotes the $\overline{x}$-component of the velocity, as measured by $\overline{O}$.

Denote by $m_0(B)$ and $m_0(\overline{B})$ the pre-collision or rest masses of B and $\overline{B}$, as measured by O, and by $m(B)$ and $m(\overline{B})$ the post-collision masses of B and $\overline{B}$ as measured as well by O. By the classical conservation of momentum in the frame of reference of O one must have: $m(B)v^3(B) = m(\overline{B})v^3(\overline{B})$. Using (10.11), the right hand side of this equation can be written as

$$m(\overline{B})v^3(\overline{B}) = \frac{m(\overline{B})\overline{v}^3(\overline{B})}{L(1 + (\mathrm{u}/\mathrm{c}^2)\overline{v}^1(\overline{B}))}.$$

Thus, we may write:

$$m(B)v^3(B) = \frac{m(\overline{B})\overline{v}^3(\overline{B})}{L(1 + (\mathrm{u}/\mathrm{c}^2)\overline{v}^1(\overline{B}))}. \tag{10.22}$$

Since $v^3(B) = \overline{v}^3(\overline{B})$ one gets:

$$m(B) = \frac{m(\overline{B})}{L(1 + (\mathrm{u}/\mathrm{c}^2)\overline{v}^1(\overline{B}))}.$$

As we consider more and more glancing collisions, the quantity $\overline{v}^1(\overline{B})$ approaches zero while $v^1(\overline{B})$ approaches u. In the limit $\overline{B}$ and B will just touch tangentially, and henceforth the z-component of the velocities of both balls will be equal to zero. In the limit the velocity in the x-direction would be: $v^1(B) = \overline{v}^1(\overline{B}) = 0, v^1(\overline{B}) = \mathrm{u}$. Since in the limit B stays still from O's view point, he would deduce that: $m(B) = m_0(B)$. Henceforth, in the limit, Formula 10.22 becomes:

$$m_0(B) = \frac{m(\overline{B})}{L(1 + \mathrm{u}^2/\mathrm{c}^2 \times 0)} = \frac{m(\overline{B})}{L}. \tag{10.23}$$

Since B and $\overline{B}$ are identical at rest one must have $m_0(B) = m_0(\overline{B})$. From (10.23) one obtains: $m(\overline{B}) = Lm_0(B) = m_0(\overline{B})/\sqrt{1-(\mathrm{u}/\mathrm{c})^2}$.

Hence, from O's perspective mass increases with velocity by a factor of $1/\sqrt{1-(\mathrm{u}/\mathrm{c})^2}$. In fact, when $\mathrm{u} \to \mathrm{c}$ the post-collision mass of $\overline{B}$ approaches infinite, as measured by O. This implies that no particle with nonzero rest mass can ever reach the speed of light! From this last formula Einstein was able to deduce, in a way that is characteristic of his way of thinking, what is perhaps the most celebrated formula in all of physics.

First, the series $(1-x^2)^{-1/2}$ is convergent for $|x| < 1$, and the first two terms in Taylor's expansion around zero are $1 + x^2/2$. Henceforth, for $x = \mathrm{u}/\mathrm{c}$ this series converges. If u is very small compared with c one has: $L \simeq 1 + \frac{1}{2}\mathrm{u}^2/\mathrm{c}^2$, and

$$m(\overline{B}) \simeq m_0(\overline{B}) + \frac{m_0(\overline{B})}{2}\mathrm{u}^2/\mathrm{c}^2 = m_0(\overline{B}) + E/\mathrm{c}^2, \tag{10.24}$$

where $E_k = \frac{1}{2}m_0(\overline{B})\mathrm{u}^2$ is the kinetic energy of $\overline{B}$, as measured by O. From his point of view, if $\Delta m(\overline{B}) = m(\overline{B}) - m_0(\overline{B})$ denotes the mass increment, Formula (10.24) tells us that $E_k = \Delta m(\overline{B})\mathrm{c}^2$. Einstein observes that the mass increment $\Delta m(\overline{B})$ is indistinguishable from an increase in kinetic energy E_k. From this, he conjectures that mass and energy are just two manifestations of one single entity. Strictly speaking, the latter reasoning leads one to postulate the equivalence of mass and energy not as a theorem, but rather as an heuristic law. Once matter and energy are identified one may define:

Definition 10.7.1. The *total energy* of a particle with rest mass $m_0 \neq 0$, as measured by O, is defined to be $E = \mathrm{c}^2 m_0/\sqrt{1-(\mathrm{u}/\mathrm{c})^2}$. Its rest energy is defined as $E_0 = m_0\mathrm{c}^2$.

When $\mathrm{u} << \mathrm{c}$, the total energy E can be approximated as $E = E_k + m_0(\overline{B})\mathrm{c}^2 =$ Kinetic energy + Rest energy of $\overline{B}$.

10.8 4-Momentum

Now we introduce a fundamental concept in relativistic dynamics. First we deal with the case of particles with non zero rest mass. By this one means a positive constant scalar associated to each particle. In this section we use again geometric units (c = 1).

10.8.1 Particles with Nonzero Rest Mass.

Definition 10.8.1. Suppose $\beta : I \to M$ is the world-line in space-time (M, g) of a particle P. Then its 4-*momentum at a point* $q = \beta(s_0)$ is defined to be the vector $\mathbf{p}_0 = m_0\mathbf{v}_0$, where m_0 is the rest mass of P and, and $\mathbf{v}_0$ denotes its 4-velocity at q, i.e., the normalized tangent vector: $\mathbf{v}_0 = \beta'(s_0)/\left|\beta'(s_0)\right|$.

Let O be an observer that measures the momentum of P at the point $q = \beta(s_0)$. Choose $x = (x^a)$ a Lorentz frame for O at q and let $b^a(s) = x^a(\beta(s))$ be the coordinates of the world line of P in this frame of reference. Parametrize O's world-line by arc length so that its four velocity u_0 is equal to ∂_{x^0}: In

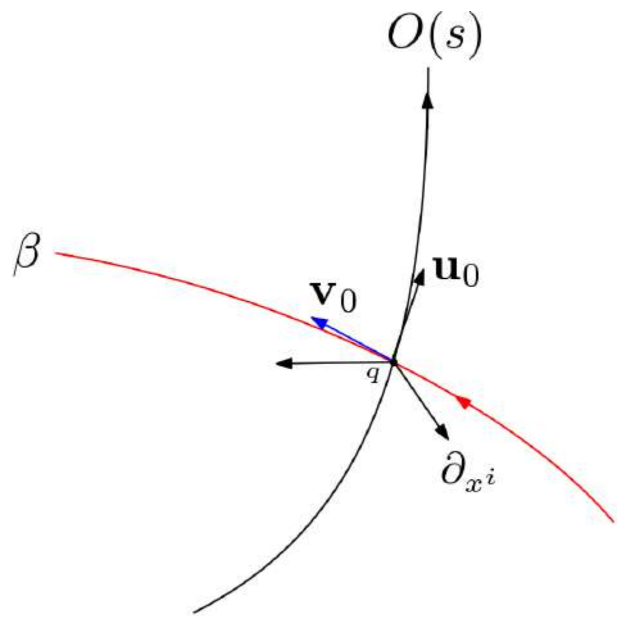


Figure 10.13: 4-momentum

the basis $\{\partial_{x^0} = \mathbf{u}_0, \, \partial_{x^i}\}$ one can write $\beta'(s) = (db^0/ds)\mathbf{u}_0 + \sum_i (db^i/ds)\partial_{x^i}$.

Then $|\beta'(s)| = \sqrt{(db^0/ds)^2 - \sum_i (db^i/ds)^2}$. Hence, p^0, the zero component of $\mathbf{p}_0 = m_0\beta'(s_0)/\left|\beta'(s_0)\right|$, is equal to (derivatives evaluated at $s = s_0$):

$$\begin{aligned} p^0 &= \frac{m_0(db^0/ds)}{\sqrt{(db^0/ds)^2 - \sum_i (db^i/ds)^2}} = \\ &= \frac{m_0}{\sqrt{1 - \sum_i (ds/db^0)^2 (db^i/ds)^2}} \\ &= \frac{m_0}{\sqrt{1 - \sum_i (db^i/db^0)^2}} = \frac{m_0}{\sqrt{1 - |\mathrm{v}|^2}}, \end{aligned} \tag{10.25}$$

where $\mathrm{v} = \sum_i v^i \partial_{x^i}$,$v^i = db^i/db^0|_{s=s_0}$ is the 3-velocity of B, and $|\mathrm{v}|$ denotes its norm (the speed of B, as measured by O). Similarly,

$$\begin{aligned} p^i &= \frac{m_0(db^i/ds)}{\sqrt{(db^0/ds)^2 - \sum_i (db^i/ds)^2}} = \\ &= \frac{m_0(db^i/ds)/(db^0/ds)}{\sqrt{1 - \sum_i (ds/db^0)^2 (db^i/ds)^2}} \\ &= \frac{m_0(db^i/db^0)}{\sqrt{1 - \sum_i (db^i/db^0)^2}} = \frac{m_0 v^i}{\sqrt{1 - |\mathrm{v}|^2}} = p^0 v^i. \end{aligned} \tag{10.26}$$

In units where $\mathrm{c} = 1$, the total energy of P at q, as measured by O, (Definition 10.7.1) is equal to $E = m_0/\sqrt{1 - |\mathrm{v}|^2}$. Thus, $E = p^0$.

Write $p = p^0\mathbf{u}_0 + \sum_i p^i \partial_{x^i}$. Since $\{\mathbf{u}_0, \partial_{x^i}\}$ is an orthonormal frame at q one sees that:

$$E = p^0 = -\left\langle \mathbf{p}_0, \mathbf{u}_0 \right\rangle = -m_0 \left\langle \mathbf{v}_0, \mathbf{u}_0 \right\rangle. \tag{10.27}$$

On the other hand, $\langle \mathbf{p}_0, \mathbf{p}_0 \rangle = \langle m_0\mathbf{v}_0, m_0\mathbf{v}_0 \rangle = m_0^2 \langle \mathbf{v}_0, \mathbf{v}_0 \rangle = -m_0^2$. Thus,

$$-m_0^2 = \langle \mathbf{p}_0, \mathbf{p}_0 \rangle = -(p^0)^2 + \sum_i \left\langle p^i, p^i \right\rangle = -E^2 + |\mathrm{p}|^2,$$

where $|\mathrm{p}|^2$ denotes the norm of the 3-momentum of P, defined as $p = \sum_i p^i \partial_{x^i}$. From this one gets:

$$E = \sqrt{m_0^2 + |\mathrm{p}|^2}. \tag{10.28}$$

Notice that when the 3-momentum is zero one finds the rest energy of the particle, $E = m_0$, which in standard units is written as $E = m_0\mathrm{c}^2$.

Remark 10.8.2. Formula (10.27) only depends on the particle P and the 4-velocity $\mathbf{u}_0$ of O and it is independent of which system of coordinates O chooses to measure the energy of P.

Now we deal with particles like photons that move in null geodesics for which their rest mass is zero.

10.8.2 Particles with Zero Rest Mass

Suppose P is a particle with zero rest mass, for instance a photon. Another of Einstein's famous equations relates the energy of P with the Planck constant h. If $\widetilde{\omega}$ denotes the frequency of the wave associated to the photon (measured in Hertz), P's energy is given by $E = h\widetilde{\omega}$. In many applications it is natural to use *angular frequency* ω which is the frequency of the light wave expressed in radians per ss. That is, $\omega = 2\pi\widetilde{\omega}$. It is also customary to absorb the 2π factor into the Planck constant. The resulting constant is called the reduced Planck constant or Dirac constant, denoted $\hbar$ (pronounced "h bar"): $\hbar = h/(2\pi)$. With this notation $E = \lambda\widetilde{\omega} = \hbar\omega$.

Let P be a photon moving in space-time (M, g) whose world-line is given by $\beta : I \to M$. In Definition 10.8.1 one can parametrize β arbitrarily since all that matters is the rest mass of the particle, and its 4-velocity. For a photon, however, since β is a null geodesic one can always change the parametrization by any affine parameter $as + b$. The constant b is fixed by choosing a point q on the world-line of β, and then demanding that $\beta(s_0) = q$ for some particular value $s = s_0$. The constant a, however, remains arbitrary. If one expects (as in Definition 10.8.1) the 4-momentum $\mathbf{p}_0$ of P at q to be a multiple of $\beta'(s_0)$ then the definition of $\mathbf{p}_0$ would always depend on the parametrization one choose for β. That is, *by fixing a particular parametrization one is also fixing* p_0, defined as $\mathbf{p}_0 = \beta'(s_0)$. For an observer O whose world-line intersects β at q we would expect the energy of the photon to be $p^0 = -\langle \mathbf{p}_0, \mathbf{u}_0 \rangle$, where $\mathbf{u}_0$ is the 4-velocity of O at q. The following definition summarizes this discussion:

Definition 10.8.3. Suppose $\beta : I \to M$ is the world-line in space-time (M, g) of a particle P with zero rest mass (β entails some fixed parametrization). The 4-*momentum of* P at a point $q = \beta(s_0)$ on its world-line is defined just to be $\mathbf{p}_0 = \beta'(s_0)$. Suppose O is an observer whose world-line intersects β at q. The total energy of P, as measured by O, is $E = -\langle \mathbf{p}_0, \mathbf{u}_0 \rangle$. The frequency measured by O is defined as $\omega = E/\hbar$, and its 4-wave vector as $\mathbf{k}_0 = (1/\hbar)\mathbf{p}_0$, so that $\mathbf{p}_0 = \hbar\mathbf{k}_0$.

10.9 Conservation of Momentum

Suppose $B_1 : I \to M$ and $B_2 : I \to M$ are the world-lines of two particles in space-time with rest masses equal to m_1 and m_2, respectively, which collide at a certain point q on the world-line of an observer O. If $\mathbf{p}_1$, $\mathbf{p}_2$ are their corresponding momenta at q before collision and $\widetilde{\mathbf{p}}_1$, $\widetilde{\mathbf{p}}_2$ are their momenta at q afterwards, then a fundamental law of physics says that $\mathbf{p}_1 + \mathbf{p}_2 = \widetilde{\mathbf{p}}_1 + \widetilde{\mathbf{p}}_2$ in $T_q(M)$. This law is known as the conservation of momentum. Choose $x = (x^a)$ a Lorentz frame for O at q. Let $\mathbf{u}_0$ be its 4-velocity and $\{\mathbf{u}_0, \partial_{x^i}\}$ an orthonormal frame for O at q. As observed in Section 10.8, the 4-momentum $\mathbf{p}_0$ of a particle B_0 with rest mass m_0 can be decomposed as $\mathbf{p}_0 = p^0\mathbf{u} + \mathrm{p}$, where $\mathrm{p} = \sum_i p^i \partial_{x^i}$ denotes the 3-momentum of the B_0. Moreover, $p^0 = E$, $p^i = Ev^i$, with $E = m_0/\sqrt{1 - |\mathrm{v}|^2}$, the total energy of B_0,and v, its 3-velocity. The conservation of momentum $\mathbf{p}_1 + \mathbf{p}_2 = \widetilde{\mathbf{p}}_1 + \widetilde{\mathbf{p}}_2$ implies immediately that:

$$p_1^0 + p_2^0 = \widetilde{p}_1^0 + \widetilde{p}_2^0, \text{ i.e., } E_1 + E_2 = \widetilde{E}_1 + \widetilde{E}_2, \text{ and}$$
$$\mathrm{p}_1 + \mathrm{p}_2 = \widetilde{\mathrm{p}}_1 + \widetilde{\mathrm{p}}_2$$

where E_i and $\widetilde{E}_i$ denote the energy of the particle B_i before and after the collision, respectively. Thus, the conservation of 4-momentum is equivalent to the conservation of energy plus the conservation of the 3-momentum.

10.10 Accelerated Observers

Suppose as in Example 11.2.10 that O represents an observer in flat space time whose world line is the t axis, i.e., $O(t) = (t, \underline{0})$, where (t, x, y, z) are standard coordinates for flat space-time. Let $\overline{O}$ be an observer who moves in the direction of the x-coordinate with constant 3-acceleration. By this, one means that $\overline{O}$'s 3-acceleration, as measured in a Lorentz frame for $\overline{O}$ at any point p in his world line has constant magnitude a. In (**??**) we showed that if we parametrize $\overline{O}$ with arc length, the proper 4-acceleration and the 3-acceleration measured at $p = \overline{O}(s_0)$ *in this Lorentz frame* coincide: $\mathbf{A}_p = \mathrm{a}_p$. Notice that the 3-acceleration is not independent of the system of coordinates we use. However, regardless of which coordinate system we use we still have $|\mathbf{A}_p| = a$, for all $p \in \overline{O}$.

Let $\overline{O}(s) = t(s)\partial_t + x(s), 0, 0)$ be the arc length parametrization of $\overline{O}$ in the standard coordinates of $\mathbb{R}^4$, with $\overline{O}(0) = p$. For the vector $\overline{O}'(s) =$

$t'(s)\partial_t + x'(s)\partial_x$ one has:

$$\left|\overline{O}'(s)\right|^2 = -t'(s)^2 + x'(s)^2 = -1 \tag{10.29}$$

On the other hand, $\nabla_{\overline{O}'(s)}\overline{O}'(s)(s) = t''(s)\partial_t + x''(s)\partial_x$, since in flat space-time the coefficients of the connection vanish in the standard coordinates of $\mathbb{R}^4$. Thus,

$$a^2 = \langle \mathbf{A}_p, \mathbf{A}_p \rangle = -t''(s)^2 + x''(s)^2.$$

For convenience we may fix $p = (0, 1/a)$, so that $x(0) = 1/a$ and $t(0) = 0$. We also assume that the observer starts moving with zero velocity, i.e., $x'(0) = 0$. From 10.29 we see that $t'(0) = 1$.

These conditions determine the following system of differential equations:

$$\begin{aligned} t'(s)^2 - x'(s)^2 &= 1 \\ -t''(s)^2 + x''(s)^2 &= a^2 \\ t(0) &= 0 \\ t'(0) &= 1 \\ x'(0) &= 0. \end{aligned}$$

The solution of this system is given by $t(s) = a^{-1}\sinh(as)$, $x(s) = a^{-1}\cosh(as)$.

The world line of $\overline{O}$ can be pictured (if we forget the y and z axis) as the points (events) on the hyperbola $x^2 - t^2 = a^{-2}$: The 4-velocity of $\overline{O}$ at any point $q = \overline{O}(s_0)$ is therefore equal to

$$\mathbf{v}_0 = \overline{O}'(s_0) = \cosh(as_0)\partial_t + \sinh(as_0)\partial_x.$$

On the other hand, $\mathbf{v}_1 = \sinh(as_0)\partial_t + \cosh(as_0)\partial_x$ also has norm one, and $\langle \mathbf{v}_0, \mathbf{v}_1 \rangle = 0$. Hence, an orthonormal basis B for $T_p(\mathbb{R}^4)$ is given by $\{\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2 = e_2, \mathbf{v}_3 = e_3\}$. As in Proposition 11.2.8, the inverse of the matrix

$$A = \begin{bmatrix} \cosh(as_0) & \sinh(as_0) & 0 & 0 \\ \sinh(as_0) & \cosh(as_0) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

provides Lorentz coordinates (x^0, x^i) for $\overline{O}$ at q, given by

$$\begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix} = \begin{bmatrix} \cosh(as_0) & -\sinh(as_0) & 0 & 0 \\ -\sinh(as_0) & \cosh(as_0) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix} - V_p, \tag{10.30}$$

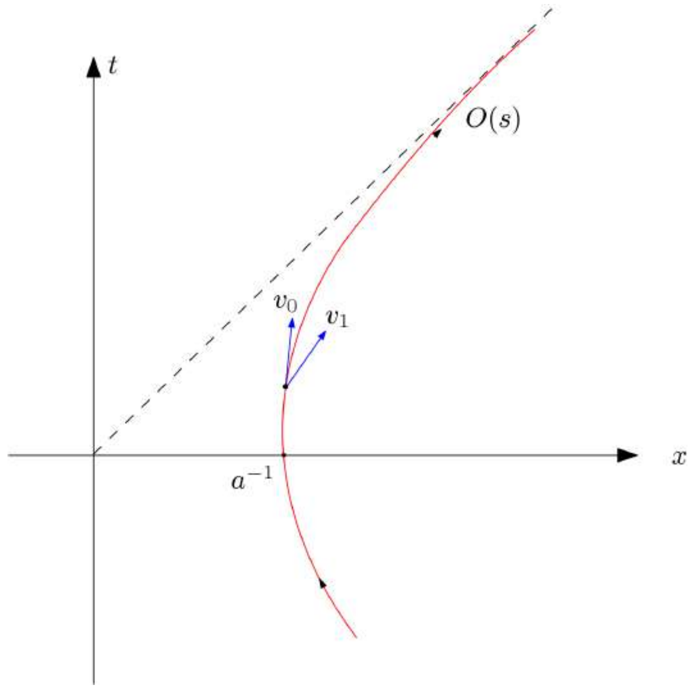


Figure 10.14: Accelerated observer

where we have subtracted the vector

$$V_p = \begin{bmatrix} \cosh(as_0) & -\sinh(as_0) & 0 & 0 \\ -\sinh(as_0) & \cosh(as_0) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a^{-1}\sinh(as_0) \\ a^{-1}\cosh(as_0) \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ a^{-1} \\ 0 \\ 0 \end{bmatrix}$$

so that $x^i(p) = 0$, and where we let $x^0(p)$ to be the proper time elapsed between $s = 0$ and $s = s_0$.The matrix equation 10.30 can be rewritten as:

$$\begin{aligned} x^0 &= \cosh(as_0)t - \sinh(as_0)x \\ x^1 &= -\sinh(as_0)t + \cosh(as_0)x - a^{-1} \\ x^2 &= y, \quad x^3 = z. \end{aligned} \tag{10.31}$$

10.10.1 Time Dilation: A Journey to Kepler 22-b

In a not too distant future humans may have developed the technology to explore outer space. We may imagine, many centuries from now, a scouting party in search of a new home for humanity. Among candidates, Kepler-22b, an exoplanet discovered in 2011 by the Kepler space telescope, an Earth like celestial body located 600 light years away from or planet, looks like an ideal place to go: Its sun, a yellow dwarf of the northern constellation of Cygnus, bathes the planet with a warm light. Its size, 2.5 times that of Earth's suggests that it almost certainly holds an atmosphere. According to density estimates, Kepler 22-b might also posses vast oceans of water. Its temperature is estimated between 22 and 27 degrees Celsius; years in that remote paradise are a little shorter: only 289 days.

We may imagine an adventurous exodus to that planet unraveling as follows: The crew starts their trip at some space station located $1/\mathrm{a} = \mathrm{c}^2/\mathrm{g} \simeq 0.97$ (g denotes the acceleration of gravity on the surface of the Earth) light years away from Earth. In a few minutes after departure, the spaceship reaches a velocity of several thousand kilometers per hour and continues to accelerate steadily at a rate of g = 9.8 m /s2. In the first hour the rocket will have gained a tremendous speed: around 120 000 km/h. Inside the probe the crew experiences a comfortable atmosphere. They appear to be motionless, everything is quite, and seems to be at rest. The astronauts experience no forces, beside a fictitious gravity that feels identical to that on Earth. According to plan, they will be reaching their maximum velocity, 99.99% the speed of light 68.6 terrestrial years after departure, which accounts to only *four years nine months and eighteen days*, as recorder in the spaceship logbook. By then, they will already be 67 light-years away from earth. At this moment the powerful engines fed with the little available hydrogen in interstellar space will stop, and will not be ignited again until the final approach to the planet.

During the next two years they will experience total weightlessness. At that speed, the sky, formerly saturated with millions of dim and remote lights looks now surreal, as if observed through a pair of goggles: At that fantastic speed, normal light coming from the stars now registers a frequency outside the visible spectrum. But infrared radiation, and other low frequency electromagnetic waves coming from approaching celestial objects has become visible. This is also the case for ultraviolet and other high frequency radiation coming from receding stars (see next section). But even at this incredible speed would not the travelers take more than six hundred years to reach its destination? Indeed, this would be the span of time recorded on terrestrial calendars, but not for them. Einstein's theory predicts that when arriving at the planet each crew member will have aged only about 6.8 years.

In Kepler-22b the explorers will remain for a decade building a space station, the foundations of a new home for human kind. Once the mission is completed, they must undertake the journey back home. When they finally arrive, the former young astronauts will be middle-aged adults after a long journey of 23.8 years, according to the spaceship's calendar. However, more than a millennium and two centuries will have elapsed, here on Earth!

Let us do the all the calculations. In the coordinates of the space center on Earth, (t, x, y, z), and in classical units, the equation of motion will be:

$$x^2 - \mathrm{c}^2 t^2 = (\mathrm{g}/\mathrm{c}^2)^{-2} = (\mathrm{c}^2/\mathrm{g})^2,$$

where $\mathrm{g} = 9.8\ \mathrm{m/s}^2$ is the acceleration of gravity. From this we obtain: $x(t) = \sqrt{\mathrm{c}^2 t^2 + (\mathrm{c}^2/\mathrm{g})^2}$. On the other hand, the spaceship's velocity measured from Earth is $\mathrm{v}(t) = dx/dt = \mathrm{c}^2 t/\sqrt{\mathrm{c}^2 t^2 + (\mathrm{c}^2/\mathrm{g})^2}$. Solving for t, one observes that the spaceship will reach a speed of 0.9999c, when $t_0 = 2.16 \times 10^9$ seconds, an equivalent to 68.63 years. However, the proper time for the crew will just be (Formula 10.6):

$$\int_0^{t_0} \sqrt{1 - (\mathrm{v}(s)/\mathrm{c})^2}\, ds \simeq 1.52 \times 10^8 \text{ seconds},$$

equivalent to 4.8 years in the spaceship's calendar. By then they will have traveled $x(t_0) = 6.4 \times 10^{17}$ m, the equivalent of 67.6 light-years. The total time t_1 it takes to reach Kepler-22b will approximately be 1.89×10^{10} seconds. That amounts to 600.96 years. But the proper time will just be

$$\int_0^{t_1} \sqrt{1 - (\mathrm{v}(s)/\mathrm{c})^2}\, ds \simeq 2.14 \times 10^8 \text{ seconds},$$

or 6.8 years. Hence, when they return to Earth, each member of the crew will be 23.8 years older. But here, in our planet, more than a millennium and two centuries will have elapsed.

10.11 Redshift and blueshift

It was claimed that during most part of the journey to Kepler 22-b (10.10.1) the sky would look surreal from the perspective of a passenger on the spaceship's bridge. To see why this must be true consider a photon P whose world-line is given by $\beta : I \to \mathbb{R}^4$, where $\beta(\tau) = (-b + \hbar\omega_0\tau, \hbar\omega_0\tau, 0, 0)$,

$b > 0$. The scalar ω_0 represents the *angular frequency,* as measured by an inertial observer on Earth $O(s) = (s, 0, 0, 0)$ at q_0: In fact, the energy measured by O is equal to $E_0 = -\langle \beta'(0), \mathbf{u}_0 \rangle = \hbar\omega_0$, where $\mathbf{u}_0 = \partial_t$, and hence the frequency must be $E/\hbar = \omega_0$.

On the other hand, the energy measured by the accelerated observer $\overline{O}$ at q would be $E_1 = -\langle \beta'(\tau_0), \mathbf{v}_0 \rangle$, where τ_0 is the value of the parameter for which $\beta(\tau_0) = q$,and $\mathbf{v}_0$ is the 4-velocity of $\overline{O}$ at q. Thus,

$$E_1 = \hbar\omega_0(\cosh(as_0) - \sinh(as_0)) = \hbar\omega_0 e^{-(as_0)},$$

with $\overline{O}(s_0) = q$. Thus the frequency measured at q by $\overline{O}$ would be $\omega_1 = E_1/\hbar = \omega_0 e^{-(\mathrm{a}s_0)}$. For $s_0 > 0$ this frequency is less that ω_0. Hence, light from a star that is moving away from the spaceship will look red shifted. On the contrary, when the spaceship approaches a star the light would be blue shifted, as $\omega_1 > \omega_0$.

The value of τ_0 can be determined by solving the system $-b + \hbar\omega_0\tau = a^{-1}\sinh(as)$, $\hbar\omega_0\tau = a^{-1}\cosh(as)$. For this, we let:

$$(\hbar\omega_0\tau)^2 - (-b + \hbar\omega_0\tau)^2 = a^{-2}(\cosh^2(as) - \sinh^2(as)) = a^{-2}.$$

From this one obtains

$$\tau_0 = \frac{1 + a^2b^2}{2a^2b\omega_0}, \quad s_0 = \frac{1}{a}\operatorname{arccos h}\left(\frac{a^2b^2 + 1}{2ab}\right) \tag{10.32}$$

The spectrum of frequencies of visible light varies in the range of $\omega_R = 3.8 \times 10^{14}$, the frequency of red light, and $\omega_V = 7.16 \times 10^{14}$, the frequency of violet light. As the astronauts travel farther away, the light of our sun will become dimmer, and each time more shifted toward the red side of the spectrum. It will remain visible while ω_V, the maximum frequency visible light, does not drop below the value ω_R, that is, while $\omega_V e^{-(\mathrm{a}s_0)} \geq \omega_R$, or equivalently, while $e^{-(\mathrm{a}s_0)} \geq 1/2$. From this, we need $s_0 \leq \ln(2)/\mathrm{a}$. For light coming from approaching stars we similarly see that $-s_0 \geq -\ln(2)/\mathrm{a}$. Since we are assuming $\mathrm{a} = 9.8/\mathrm{c}^2$ m/ss^2 we obtain from (10.32) that only while $|s_0| \leq 6.3 \times 10^{15}$ss the astronauts will be able to see the normal light of the spectrum. Measured in years, this corresponds to 0.67 years. Roughly after eight months of proper time travel in their way to Kepler 22-b, the sky will look phantasmagoric.

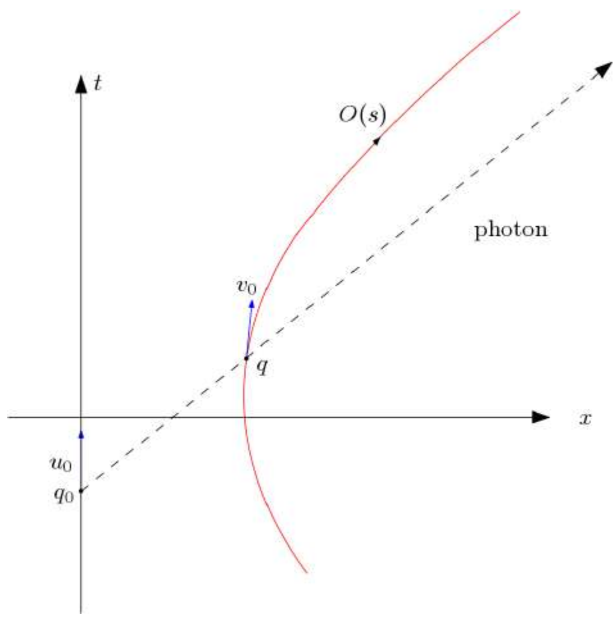


Figure 10.15: Redshift

Part III

How Energy and Matter Curve Space-Time

Chapter 11

Foundations of the General Theory of Relativity

11.1 Foundations

By (x, y, z) we denote the usual coordinates of euclidean space $\mathbb{R}^3$. By t we will denote "time", so that (t, x, y, z) will be coordinates for space-time in Newtonian physics, as well as in Special Relativity, the theater where physical phenomena occur. By e_a we will denote the vectors of the standard basis of $\mathbb{R}^3$. That is, at each point $p \in \mathbb{R}^3$, $e_0 = \partial_t|_p$, $e_1 = \partial_x|_p$, $e_2 = \partial_y|_p$, $e_3 = \partial_z|_p$.

The theory of General Relativity is a geometrical theory of gravitation. In Einstein's theory, *gravity is no longer regarded as a force but as a manifestation of the geometry of space-time.* The Field Equations determine a particular pseudo-riemannian metric, depending on the energy distribution in the universe. In the simplest case, where there is no energy present (Special Relativity), the structure of space-time is given by Minkowski's metric, a pseudo-riemannian metric conditioned by the fact that the inertial observers should measure the same value for the speed of light.

We start by modeling space-time as $\mathbb{R}^4$. Intuitively, each point p represent an *event*, recorded with four real numbers called *its local coordinates*, where by "coordinates" we mean any chart defined around p. Allowing this degree of freedom, a question naturally arises: What is the physical significance of these coordinates? We are assuming in disguised that x, y, z measure "distance" and t, measures "time", as in the Newtonian theory. But in the General Theory of Relativity, as we shall see, the meaning of these concepts is not so straightforward. In fact, after realizing how gravity could be incorporated within the framework of Special Relativity (his famous equivalence principle), it took Einstein seven more years to formulate his celebrated field equations. One of the main obstacles, in his own words, lied in the fact that *It is not so easy to free oneself from the idea that coordinates must have an immediate metrical meaning.*

As in the previous chapter, in general we call a vector $v \in T_p(M)$ *timelike* if $|v|^2 = g(v, v) < 0$; it is called *spacelike*, if $|v|^2 = g(v, v) > 0$; and it is called *null or lightlike*, if $|v|^2 = g(v, v) = 0$. A curve $\alpha : I \to M$ is called *timelike* (respectively*, spacelike or null*) if the tangent vector at any point $\alpha'(s)$ is a timelike vector (respectively, spacelike, lightlike).

Observers can be described by timelike curves $O : I \to \mathbb{R}^4$, where I will always denote some suitable open interval containing zero. An observer may be thought as a not "too massive" material particle that moves in spacetime at sub-light speed. The image of O is called its *world line.* This can be regarded as the collection of all events that encompass his complete history.

Interpreting O as a human observer, measurements performed by O are magnitudes are gauged in some frame of reference. By this we mean a suitable coordinate system $x = (x^0, x^i)$ defined in some neighborhood U_p of a particular event p in O's world line. In this frame, x^0 *represents a measurement of his wristwatch time (proper time),* while x^1, x^2, x^3 are used by O to determine spatial coordinates (we will make this precise in 11.2.2). Here we follow the convention that *Latin letters* i, j *denote spatial coordinates, and indices* $a, b, c, ..$will always be used to *denote all four coordinates.*

We expect an observer to move "forward in time", while he stays still with respect to his own coordinate system. Mathematically, this means that if we write O in the fame $x = (x^a)$, let's say $O(s) = (c^a(s))$, with $O(0) = p$, then $dc^0/ds|_{s=0} > 0$, and $dc^i/ds|_{s=0} = 0$. In fact, the first condition should hold for the world line of any particle, even one without mass (like photons), since nothing can travel back in time, as far as physicists know.

Now, let's assume that the curve $\gamma : I \to \mathbb{R}^4$ describes the trajectory of a photon that crosses O's laboratory. If s is any parameter for γ, and if we write $\gamma(s) = (c^a(s))$, with $dc^0/ds > 0$, and $\gamma(0) = 0$, then the velocity of the photon as measured by O in euclidean coordinates would be equal to the square root of $(dc^1/dc^0)^2 + (dc^2/dc^0)^2 + (dc^3/dc^0)^2$ (derivatives are taken at $s = 0$). In metric (nonstandard) units, O must register a speed equal to one unit, and consequently

$$(dc^0/ds)^2\left[\left(dc^1/dc^0\right)^2+\left(dc^2/dc^0\right)^2+\left(dc^3/dc^0\right)^2\right]=(dc^0/ds)^2\times 1.$$

Thus, $-(dc^0/ds)^2+\sum_i(dc^i/ds)^2=0$. If one expects the world-line of photons to be *null geodesics* $\gamma(s)$ (*one for which* $\langle\gamma'(s),\gamma'(s)\rangle=0$), then *the natural metric at* p, written in the coordinates $x=(x^a)$, *should be given by a matrix of the form:*

$$\eta=\begin{bmatrix}-1&&&\\&1&&\\&&1&\\&&&1\end{bmatrix},\tag{11.1}$$

with signature $(-1,1,1,1)$. Keeping this in mind we now formally set the stage for the magnificent theory of Einstein.

Axioms for the theory of General Relativity

1. In General Relativity (GR) space-time is modeled by a real smooth 4-manifold $(M,g=\langle -,-\rangle)$ endowed with a pseudo-riemannian metric of signature $(-1,1,1,1)$. The laws of physics should be expressed in a coordinate independent manner, that is, in terms of tensors, covariant derivatives and the like.

2. Each "particle" is characterized by its world line O, as well as a non-negative scalar m_0 associated to it, called its *rest mass*, in such a way that:

   i The worldline of any particle with nonzero mass will always be a *timelike curve*. If this curve is a geodesic, we say that the particle moves freely or that no "forces act upon it".

   ii Particles with *zero mass* (photons, neutrinos, etc.) have world lines that are *null geodesics*.

3. The distribution of masses (energy) in M is described mathematically by a tensor, the *energy-momentum tensor*. This tensor, T, is locally preserved: $D_X T=0$, for any vector field X in M (see Section 13.3).

4. The tensor of energy-momentum determines the geometry of space-time in accordance with Einstein's field equations: $\mathsf{Ric}-\frac{1}{2}\mathrm{R}g=(8\pi\mathrm{G}_N/\mathrm{c}^4)T$ (Chapter 14). This is the very heart of the theory.

Special Relativity (SR) is a particular case of GR, one in which $(M, g) = (\mathbb{R}^4, \langle -, - \rangle_\eta)$ and the metric is given by 11.1, in standard Cartesian coordinates. In classical notation, Minkowski's metric is written as $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$.

11.2 Observers and Physical Coordinates

In this section we rigorously introduce some fundamental notions of the Theory.

Definition 11.2.1. An *observer* O in space-time (M, g) is defined as a oriented 1-dimensional submanifold of M whose tangent space is timelike at any point. Around any point $p \in O$ we can always choose an orientation preserving parametrization $\alpha : I \to M$. For the most part we will identify α with O. The *observer's 4-velocity* at $p = \alpha(s_0)$ is defined to be the normalized vector $\mathbf{u}_0 = \alpha'(s_0)/\left|\alpha'(s_0)\right|$. Intuitively, O represents the world-line of a particle in space-time, i.e., O is the set of all events that encompass its whole existence. The fact that O is an oriented 1-manifold means that these events are "ordered" in one particular direction.

The physical meaning of coordinates in GR is a subtle issue, as we discussed in Section 11.1. Not every system of coordinates provides "true time and true spatial coordinates". In GR is not possible, in general, to define global systems of coordinates with "physical meaning". If O is any observer in space-time (M, g), in general one has to limit oneself to local frames of references at each point p in the world line of O. One would expect that such coordinate system (x^a) is such that x^0 represents O's "proper time" while x^i would be "legitimate" spatial coordinates. As such, each vector $u = \sum_i a^i \, \partial_{x^i}|_p$ in $T_p(M)$ should be spacelike. Moreover, by using these coordinates O should measure the speed of a photon as c $= 1$ (time is measured in short seconds). This motivates the following definition:

Definition 11.2.2. Let O be any observer in space-time $(M, g = \langle -, - \rangle)$. We will say that a system of coordinates $x = (x^a)$, defined in a neighborhood U_p of an event $p = O(s_0)$, is a *frame of reference for O at p with physical meaning* if:

1. The observer O moves in time but he does not move spatially with respect to this frame: This just means that $dc^0/ds|_{s=s_0} \neq 0$, and

$dc^i/ds|_{s=s_0} = 0$, where $c^a(s) = x^a(O(s))$ denote the coordinates of O in this frame. This, of course, is equivalent to saying that $O'(s_0)$ is a non zero multiple of $\partial_{x^0}|_p$. These conditions are clearly independent of the parametrization. We will always assume that O moves forward in time, i.e., for any observer we will always choose a local parametrization for O satisfying $dc^0(s)/ds|_{s=s_0} > 0$.

2. Each spatial vector $u = \sum_i u^i \, \partial_{x^i}|_p$ in $T_p(M)$ is spacelike, i.e., $\langle u, u\rangle > 0$.

3. O must register c = 1 for the speed of any photon that crosses O's laboratory in the direction of any spacelike vector u, when measured in this frame of reference. By such a photon we mean a null geodesic γ, with $\gamma(0) = p$, and such that it moves "forward in time", i.e., such that $dc^0/ds(0) > 0$, where $c^a(s) = x^a(\gamma(s))$ denote the coordinates of γ in O's frame of reference. On the other hand, by measuring its speed as c = 1 one means that the norm of the 3-vector $v = \sum_i dc^i/dc^0|_p \, \partial_{x^i}$ (its velocity, as measured by O) is equal to 1.

In choosing a frame of reference for O at p with physical meaning O is labeling his coordinates: x^0 represents a measurement of his proper time, while the other coordinates x^i are used to gauge spatial quantities.

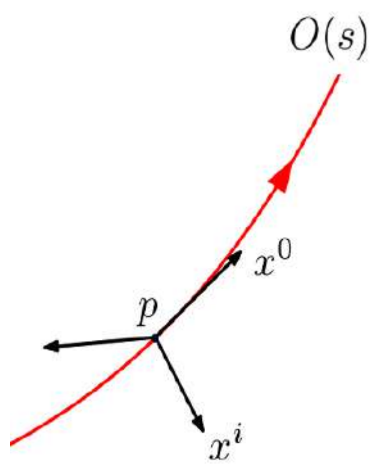


Figure 11.1: Coordinates with physical meaning

Remark 11.2.3.

1. In a frame with physical meaning, the 4-velocity of O has the direction of time, i.e., it is a multiple of ∂_{x_0}.

2. If O is an observer, and $x = (x^a)$ is a frame of reference for O at p with physical meaning, it is always possible for O to send a light signal in the spatial direction determined by any fixed spacelike vector $u = \sum_i u^i \partial_{x^i}$ of his choice. This means that there is a null geodesic γ(representing the world line of a photon) with $\gamma(0) = p$, and such that if $c^a(s) = x^a(\gamma(s))$, then $(dc^0/ds)(0) > 0$, and such that the 3-velocity vector of that photon, $\sum_i (dc^i/dc^0)(0)\partial_{x^i}$, is a scalar multiple of u.

Proof. The first assertion is an immediate consequence of (1). To see why the second assertion is true, we recall that given any null vector w, by Theorem **??** there is a unique null geodesic (up to affine parametrization) γ with $\gamma(0) = p$ and $\gamma'(0) = w$. Since $x = (x^a)$ is a frame of reference for O at p with physical meaning, condition (1) in (11.2.2) implies that $O'(s_0) = dx^0(O(s))/ds|_{s=s_0}\, \partial_{x^0}$. Since $O(s)$ is, by definition, a timelike curve, and $(dx^0(O)/ds)(s_0) > 0$, one must have $\langle \partial_{x^0}, \partial_{x^0} \rangle < 0$. Hence, it suffices to find a null vector of the form $w = \lambda\partial_{x^0} + \mathrm{u}$, with $\lambda > 0$ (this implies, of course, that the photon will move forward in time and that its 3-velocity would be $\lambda^{-1}\mathrm{u}$). We are looking for a positive real λ such that $\langle \lambda\partial_{x^0} + \mathrm{u},\ \lambda\partial_{x^0} + \mathrm{u} \rangle = 0$. This means that we must solve the following quadratic equation for λ:

$$\lambda^2 \langle \partial_{x^0}, \partial_{x^0} \rangle + 2\lambda \langle \partial_{x^0}, \mathrm{u} \rangle + |\mathrm{u}|^2 = 0. \tag{11.2}$$

But this is always possible, since

$$4 \langle \partial_{x^0}, \mathrm{u} \rangle^2 - 4 \langle \partial_{x^0}, \partial_{x^0} \rangle |\mathrm{u}|^2 = 4 \left(\langle \partial_{x^0}, \mathrm{u} \rangle^2 - \langle \partial_{x^0}, \partial_{x^0} \rangle |\mathrm{u}|^2 \right) > 0,$$

because $\langle \partial_{x^0}, \mathrm{u} \rangle^2 \geq 0 > \langle \partial_{x_0}, \partial_{x_0} \rangle |\mathrm{u}|^2$. Thus, (11.2) has two different real roots λ_1, λ_2 whose product must be equal to $|\mathrm{u}|^2 / \langle \partial_{x^0}, \partial_{x^0} \rangle$. Since this number is negative, λ_1 and λ_2 must have opposite signs. Hence, we may always choose a positive root for (11.2). □

Proposition 11.2.4. Let O be any observer in space-time (M, g). A system of coordinates (x^a) defined in a neighborhood U_p of an event $p = O(s_0)$ is a frame of reference for O at p with physical meaning if and only:

1. The metric at p in these coordinates takes the diagonal block form $g(p) = \begin{bmatrix} -1 & 0_{1\times 3} \\ 0_{3\times 1} & A_{3\times 3} \end{bmatrix}$, where A defines a Riemannian metric on the spatial directions. That is, for each spacelike vector $u = \sum_i u^i \, \partial_{x^i}|_p$ in $T_p(M)$ one has $\langle u, u \rangle_A > 0$.

2. $dx^0(O(s))/ds|_{s=s_0} > 0$, and $dx^i(O(s))/ds|_{s=s_0} = 0$.

Proof. $\Longrightarrow$) Let's suppose $x = (x^a)$ is a frame of reference at p with physical meaning. Fix $\mathrm{u}_1 = \partial_{x^1}|_p$; this vector is spacelike, by condition (2) in (11.2.2). By the previous remark, it is is always possible to send a light signal that crosses O in the direction of u_1. That is, there is a null geodesic $\gamma(s)$ corresponding to a photon moving forward in time. Let $\gamma(0) = p$, and $\gamma'(0) = \lambda \, \partial_{x^0}|_p + \mathrm{u}_1$. If $c^a(s) = x^a(\gamma(s))$ then $(dc^0/ds)(0) = \lambda > 0$, $(dc^1/ds)(0) = 1$, and $(dc^k/ds)(0) = 0$, for $k = 2, 3$. Denote by G the matrix that represents the metric at p in the coordinates $x = (x^a)$. Since $\gamma'(0) = \lambda \, \partial_{x^0}|_p + \partial_{x^1}|_p$ is a null vector one must have

$$0 = \langle \gamma'(0), \gamma'(0) \rangle_g = \gamma'(0)^* \; G \; \gamma'(0) = \lambda^2 g_{00} + 2\lambda g_{01} + g_{11}. \tag{11.3}$$

Now, let's see that $\lambda = |\mathrm{u}_1|$. Since the speed of this photon, as measured by O in the frame (x^a) is equal to one, we must have: $\left| dc^1/dc^0|_p \right| \left| \partial_{x^1}|_p \right| = 1$. Hence, $\left| dc^1/dc^0|_p \right| = 1/\left|\mathrm{u}_1\right|$. On the other hand, by the chain rule

$$\left.\frac{dc^1}{dc^0}\right|_p = \frac{(dc^1/ds)(0)}{(dc^0/ds)(0)} = \frac{1}{\lambda}.$$

Thus,

$$\lambda = \frac{1}{dc^1/dc^0|_p} = \frac{\pm 1}{1/\left|u_1\right|} = \pm \left|\mathrm{u}_1\right|.$$

Since $\lambda > 0$ we must have $\lambda = |\mathrm{u}_1|$. Similarly, a photon that crosses p in the spatial direction of $-\mathrm{u}_1$ moves in a unique (up to affine reparametrization) geodesic $\alpha(s)$, with $\alpha(0) = 0$, and $\alpha'(0) = \lambda' \, \partial_{x^0}|_p - \mathrm{u}_1$, with $\lambda' = |-u_1| = \lambda$. Hence, as in (18.11)

$$(\lambda')^2 g_{00} - 2\lambda' g_{01} + g_{11} = 0. \tag{11.4}$$

Subtracting 18.11 from 18.12 one obtains $g_{01} = g_{10} = 0$. Fixing $\mathrm{u}_j = -\left.\partial_{x^j}\right|_p$, for $j = 2, 3$,we obtain in a similar fashion that $g_{02} = g_{20} = 0$, and that

$g_{03} = g_{30} = 0$, respectively. This proves that the matrix for $g(p)$ takes the diagonal block form as above. Now we show that $g_{00} = -1$. Since $g_{01} = 0$, Equation 18.11 implies that $|u_1|^2 g_{00} + g_{11} = 0$. But $g_{11} = \langle \mathrm{u}_1, \mathrm{u}_1 \rangle = |\mathrm{u}_1|^2$, and consequently $|\mathrm{u}_1|^2 (g_{00} + 1) = 0$, from which we obtain $g_{00} = -1$. Finally, for each spacelike vector $\mathrm{u} = \sum_i a_i \, \partial_{x^i}|_p$ in $T_p(M)$, one has, by hypothesis, that $\langle \mathrm{u}, \mathrm{u} \rangle > 0$. But clearly $\langle \mathrm{u}, \mathrm{u} \rangle = \langle \mathrm{u}, \mathrm{u} \rangle_A$, and consequently A defines a Riemannian metric on the spatial directions. $\Longleftarrow$) Let's now suppose the metric $g(p)$ takes in the coordinates $x = (x^a)$ the diagonal block form as above. Suppose $\beta(s)$ parametrizes the trajectory of a photon that crosses O's laboratory at p in the spatial direction of a unit spacelike vector u. That is, if we denote $x^a(\beta(s))$ by $b^a(s)$, then $\beta(0) = p$, $\beta'(0) = db^0/ds(0)\partial_{x^0} + \sum_i db^0/ds(0)\partial_{x^i}$, with $db^0/ds(0) > 0$, and $\mathrm{u} = \sum_i db^0/ds(0)\partial_{x^i}$.Let's show that O must measures $c = 1$ for its speed. Since β is a null geodesic

$$0 = \langle \beta'(s), \beta'(s) \rangle = -(db^0/ds)^2 + \langle \mathrm{u}, \mathrm{u} \rangle_A \, ,$$

Henceforth, $(db^0/ds)^2(0) = |\mathrm{u}|^2$. Since $db^0/ds(0) > 0$ one sees that $(db^0/ds)(0) = |\mathrm{u}|$. The speed of the photon, as measured by O, is the norm of the vector $\mathrm{w} = \sum_i \, db^i/db^0|_p \, \partial_{x^i}$. By the chain rule

$$db^i/db^0\big|_p = (db^i/ds)(0)/(db^0/ds)(0) = \frac{1}{|\mathrm{u}|}(db^i/ds)(0).$$

Thus, $|\mathrm{w}| = (1/|\mathrm{u}|)\,|\mathrm{u}| = 1$. Finally, it is clear that $\langle \mathrm{u}, \mathrm{u} \rangle_g = \langle \mathrm{u}, \mathrm{u} \rangle_A > 0$, for any vector $\mathrm{u} = \sum_i a^i \, \partial_{x^i}|_p$. So we have proved conditions (1) and (3), while condition in (2) are given by hypothesis. □

The last proposition shows that if O is an observer, and if (y^a) are coordinates for space-time (M, g), then, in order to construct a frame of reference at p with physical meaning for O at a point $p = O(s_0)$ it suffices to find coordinates $x = (x^a)$ that put $g(p)$ in diagonal block form, and for which $O'(s_0) = \partial_{x^0}$.

Among coordinates with physical meaning, the *Lorentz frames* play an important role. These coordinates may be thought of as the instantaneous frame of reference of an observer in “free fall”, according to the famous principle of equivalence of Einstein. More precisely:

Definition 11.2.5. Let $O(s)$ be any observer in space-time (M, g). A *Lorentz frame* of reference for O at a point $p = O(s_0)$ (also called an *inertial frame* or a frame in free fall at p) on its world line is any choice of coordinates (x^a) in some neighborhood U_p of p such that:

1. ∂_{x^0} is the 4-velocity of O at p.

2. The set $\{\partial_{x^0},\ \partial_{x^i}\}$ is an orthonormal basis at p, and the metric $g(p)$ expressed in this basis takes the diagonal form: $\mathrm{diag}[-1,1,1,1]$.

3. In these coordinates all the Christoffel symbols $\Gamma^c_{ab}(p) = 0$ vanish at p. By Proposition **??** this is equivalent to the vanishing of all partial derivatives $\partial g_{ab}/\partial x^c(p) = 0$.

Remark 11.2.6.

1. Any Lorentz frame has physical meaning. This is clear, since the matrix that represents the metric at p is the Minkowski metric 11.1 and, by condition 1 above, $O'(s_0) = dx^0(O(s))/ds|_{s=s_0}\,\partial_{x^0}$ with $O'(s_0)/\left|O'(s_0)\right| = \partial_{x^0}$ which implies that $dx^0(O(s))/ds > 0$, and $dx^i(O(s))/ds|_{s=s_0} = 0$.

2. The condition $\partial g_{ab}/dx^c = 0$ in (iii) is equivalent to the vanishing at p of the Christoffel symbols $\Gamma^a_{b,c}(p)$ (Proposition **??**). So, a Lorentz frame is locally (this means only at p) the frame of reference of an "instantaneous" inertial observer. It is by no means clear that Lorentz frames exist. This will be proved in the next proposition. Moreover, we will see that any observer O can choose a smoothly varying collection of Lorentz frames of reference at each point of his world line.

Proposition 11.2.7. Let $O : I \to M$ be an observer in space-time (M, g). Then, at each point $p \in O$ there is a system of coordinates $x = (x^a)$ in an open neighborhood of p such that x is a Lorentz frame at p. These systems of coordinates can be chosen canonically so that they vary smoothly along the world line of O.

Moreover, if O is a geodesic, there is a smooth choice of bases $B_s = \{e_a(s)\}$ for each $T_{O(s)}(M)$ such that $\nabla_{\mathbf{u}_0} e_a(s) = 0$ along O, where $\mathbf{u}_0(O(s)) = O'(s)/\left|O'(s)\right|$ denotes the 4-velocity of O at each point. To each basis B_s there is a corresponding Lorentz frame $x_s = (x^a_s)$ at $O(s)$. The collection $\{x_s = (x^a_s)\}$ is called a *commoving Lorentz system of frames* for O.

Proof. Let u_p be the tangent vector of O at p, and let $e_0 = u_p/\left|u_p\right|$ be its normalization (O's 4-velocity). We choose vectors e_i such that $B = \{e_0, e_i\}$ is an orthonormal basis for $T_p(M)$. It is a well known fact (Proposition 9.2, [35]) that there exists a neighborhood of the origin W_0 in $T_p(M)$ and U_p in

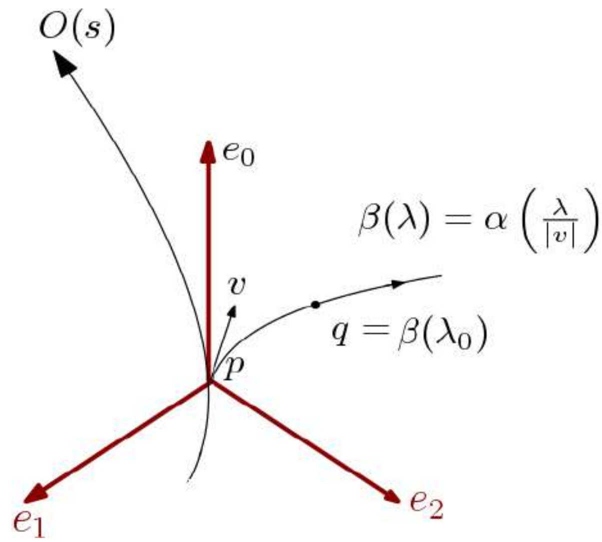


Figure 11.2: Geodesics and coordinates

M and a diffeomorphism $\sigma : W_0 \to U_p$ that sends each tangent vector $v \in U_p$ into the point $q \in U_p$ determined by the unique geodesic α that satisfies $\alpha(0) = p$, $\alpha'(0) = v$, and $\alpha(1) = q$. Hence, we may define coordinates for each point q in U_p by $x^a(q) = v^a$, where v^a are the components of v in the basis B, i.e., $v = \sum_a v^a e_a$. If $\beta(s) = \alpha(s/|v|)$ then β is a also a geodesic, this time with $\beta(0) = p$, $\beta'(0) = \mathbf{v}$, where $\mathbf{v} = v/|v|$, and $\beta(s_0) = q$, for $s_0 = |v|$. Therefore, each such geodesic emanating form p in the direction of v can be parametrized in the coordinates x^a as a linear map $\beta(s) = (s\mathbf{v}^a)$. Since β is a geodesic, it must satisfies the geodesic equations in any coordinate system. In particular, in the frame $x = (x^a)$. Thus, for all a, b, c:

$$\frac{d^2\beta^c}{ds^2} + \sum_{a,b} \Gamma^c_{ab}(\beta(s)) \frac{d\beta^a}{ds}\frac{d\beta^b}{ds} = 0.$$

But in these coordinates $d^2\beta^c/ds^2 = 0$, and $d\beta^a/ds|_{s=0} = \mathbf{v}^a$. Then, these equations can be read as $\sum_{a,b} \Gamma^c_{a,b}(p)\mathbf{v}^a\mathbf{v}^b = 0$, for arbitrary $\mathbf{v}^a, \mathbf{v}^b$. But this forces all the Christoffel symbols $\Gamma^c_{a,b}$ to vanish at p in the frame x. Finally, the matrix representing g at p in the frame x must have the form 11.1, since in these coordinates $\partial_a|_p = e_a$, and the e_a form an orthonormal system. Hence, $x = (x^a)$ is a Lorentz frame at p, and as such it is obvious that it satisfies the conclusion of Proposition 11.2.4, and, as such, it has physical meaning. In order to obtain a smoothly varying system of coordinates along O's worldline one chooses $e_0(p) = u_p/|u_p|$, and smoothly varying orthonormal frames $\{e_0(s), e_i(s)\}$ for $T_{O(s)}(M)$ along O. Then, one can take as a smoothly varying system of frames of reference the ones given by geodesic coordinates, as constructed above. When $O(s)$ is a geodesic, one first defines $e_0(s) = \mathbf{u}_0(s)$, the unitary tangent at the point $O(s)$. Next, at the point p, one chooses an orthonormal basis $\{e_0(p), e_i(p)\}$. Then we let $e_i(s)$ be the parallel transport of $e_i(p)$ along O. Since $\nabla_{\mathbf{u}_0} \langle e_a(s), e_b(s)\rangle = \langle \nabla_{\mathbf{u}_0} e_a(s),\ e_b(s)\rangle + \langle e_a(s),\ \nabla_{\mathbf{u}_0} e_b(s)\rangle = 0$ we see that $\langle e_a(s), e_b(s)\rangle =$ constant. Henceforth, $\langle e_a(s), e_b(s)\rangle = \langle e_a(p), e_b(p)\rangle = 0$, as wanted. Finally, each frame $x_s = (x_s^a)$ is constructed as in the paragraph above. $\square$

In flat space-time it is fairly simple to find a Lorentz frame for any observer O, at any point p in its world line. At $p = O(s_0)$, we may take $\mathbf{u}_0 = O'(s_0)/|O'(s_0)|$, and u_i, vectors in $T_p(\mathbb{R}^4)$ such that $B = \{u_0, u_i\}$ is an orthonormal basis for $T_p(\mathbb{R}^4)$. Denote by $x = (x^a)$ the standard coordinates of $\mathbb{R}^4$, and by $\varepsilon = \{e_a : e_a$ is the vector $\partial_{x^a}|_p\}$ the standard basis for $T_p(\mathbb{R}^4)$. Given a point $q \in \mathbb{R}^4$ the unique geodesic α with $\alpha(0) = p$, $\alpha(1) = q$ is a straight line from p to q, and henceforth if the geodesic coordinates of q are $z^a(q) = c^a$, then as vectors of $\mathbb{R}^4$, $q - p = \sum_a c^a u_a$. On the other hand, one can write $q - p = \sum_b x^b(q-p) e_b$. If w_{ab} denote the entries of the the matrix of bases change from ε to B, i.e., the matrix A^{-1}, where $A = [Id]_{\varepsilon B}$, then we may write $e_b = \sum_a w_{ab} u_a$, and therefore

$$\begin{aligned} q - p &= \textstyle\sum_b x^b(q-p) \sum_a w_{ab} u_a \\ &= \textstyle\sum_a \left(\sum_b x^b(q-p) w_{ab}\right) u_a. \end{aligned}$$

From this we immediately obtain $z^a(q) = \sum_b x^b(q-p) w_{ab}$.

It is sometimes convenient not to "reset" O's watch at ever point $p \in O$. This amounts to replacing the coordinate z^0, as constructed above, by $z^a(q) = \sum_b x^b(q) w_{ab}$, and where one chooses an arbitrary point p_0 in the world line of O at which one takes $z^0(p_0) = 0$, i.e., O sets his wristwatch at time zero at p_0, and starts measuring time from that moment one. These new coordinates have also physical meaning, since the metric expressed in the z^a's is the Minkowski matrix 11.1. Summarizing:

Proposition 11.2.8 (Notation as above)**.** Let O be an observer in flat space time. If $p = O(s_0)$ is a point on O, there is a Lorentz frame of reference for

O at p given by $z^a(q) = \sum_b x^b(q-p)w_{ab}$, for any $q \in \mathbb{R}^4$. One can also choose the coordinate z^0 in such a way that $z^0(q) = \sum_b w_{0b} x^b(q)$.

In general space-time (M, g) it is not always possible to find coordinates that are Lorentz in a whole region W of M, since this would imply the vanishing of the Christoffel symbols throughout U, and consequently the curvature tensor of ∇^g would be zero forcing W to be a flat region of space-time, which is never the case in the presence of gravity. However, in some situation there exist whole regions U in which is possible to define global coordinates that at least have physical meaning for full collection of observers that can "thoroughly map U". This involves the notion of a spacelike embedded hypersurface. By this we mean an embedding of a hypersurface $i : S \hookrightarrow M$ such that:

1. The metric induced in S by the pull back of i, denoted here by $\langle -, - \rangle_A$, is positive-definite.

2. At each point $p \in S$ the unitary vector $n_p \in T_p(M)$ orthogonal to $T_p(S)$ is timelike.

Proposition 11.2.9 (Gaussian Coordinates)**.** Let S be a spacelike embedded hypersurface in M. By α_p we denote the unique geodesic that starts at each point $p \in S$ in the direction of the unitary normal vector n_p to S, i.e., such that $\alpha_p(0) = p$, and $\alpha'_p(0) = n_p$. Then, there is an open region $U \subset S$ where $\mathrm{im}(\alpha_p) \cap \mathrm{im}(\alpha_{p'}) = \varnothing$, for all pairs $p \neq p'$ in U. Moreover, there is an open set $W \subset M$ such that $W \cap S = U$, and coordinates $x = (x^a)$ defined in W such that such that if $\mathrm{im}(\alpha_p) \cap W$ represents the world line of an observer $O_p(s)$ then:

1. In the coordinates $x = (x^a)$ the metric adopts a diagonal block form $g(q) = \begin{bmatrix} -1 & 0 \\ 0 & A_{3\times 3} \end{bmatrix}$ at each point q on the world line of any of the observers O_p.

2. In these coordinates, the observers $O_p(s)$ are *commoving*, which means that $O_p(s)$ can be written as $O_p(s) = (s, p)$, and therefore correspond to particles that move freely (under the action of no force), and that stay at the origin of their own system of coordinates at all time (their proper time). In particular, each set of coordinates has physical meaning.

3. The proper time register by any pair of observers O_p and O'_p between two events with the same x^0 coordinate is the same. Hence, they can synchronized their own watches so that they coincide with the universal time x^0.

Proof. According to (REF) it is always possible to choose an open region $W \subset M$, where $\text{im}(\alpha_p) \cap \text{im}(\alpha_q) = \varnothing$, for all pairs $p \neq q$ in $U = W \cap S$ (REF). For each point $p \in U$, let α_p be the unique geodesic with $\alpha(0) = p$, and $\alpha'(p) = n_p$. This curve is evidently timelike, since it is a geodesic (its tangent vector at any other point is a parallel transport of n_p). So it corresponds to an observer that we call O_p. Since $|n_p| = 1$, the curve $\alpha_p(s)$ must be parametrized by arc length. By making W smaller we may assume we can choose coordinates $y = (y^i)$ in $S \cap W$. We assign to each point $q \in W$ coordinates x^a in the following manner: First, we choose the only geodesic α_p where q lies. Then we assign $x^i(q) = y^i(p)$, and $x^0(q) = s$, where s is the arc length (proper time of O_p) from p to q along α_p. Hence, we have well defined vector fields in W, $Y_0 = \partial_{x^0}$, and $\partial_{x^i} = \partial_{y^i}$. Let's denote ∂_{y^i} by Y_i. We want to prove that $g_{00}(q) = -1$, and that $g_{0i}(q) = 0$.

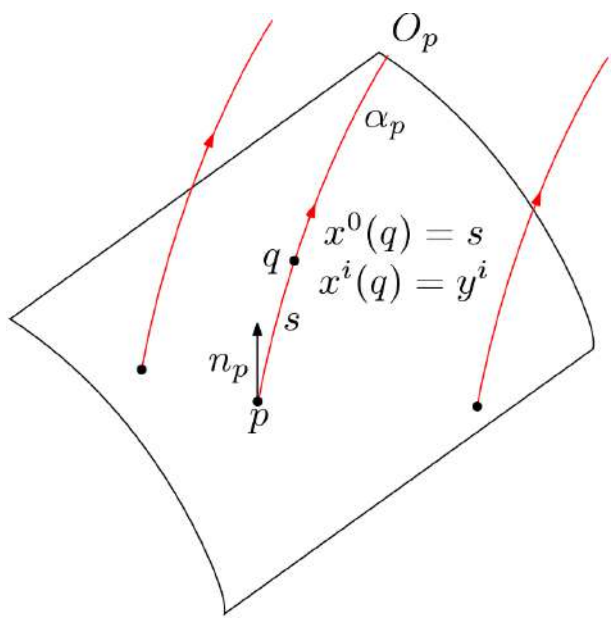


Figure 11.3: Gaussian coordinates

The first assertion follows from the fact that if $q \in \text{im}\alpha_p(s)$, then, since

$\alpha_p(s)$ is a geodesic, the magnitude of its tangent vector stays constant, and it is equal to -1, since this curve is a timelike geodesic. That is:

$$g_{00}(q) = \langle \alpha'(q), \alpha'(q) \rangle = \langle \alpha'(p), \alpha'(p) \rangle = -1.$$

For the second assertion we will show that $dg_{0i}(\alpha_p(s))/ds = 0$. Since $g_{0i}(\alpha_p(0)) = \langle n_p, Y_i \rangle = 0$, because n_p is orthogonal to all vectors in $T_p(S)$, one obtains $g_{0i}(\alpha_p(s)) = 0$, for all s, as wanted. Henceforth, we must prove that $dg_{0i}(\alpha_p(s))/ds = \nabla_{Y_0} \langle Y_0, Y_i \rangle$ is zero. Now,

$$\nabla_{Y_0} \langle Y_0, Y_i \rangle = \langle \nabla_{Y_0} Y_0, Y_i \rangle + \langle Y_0, \nabla_{Y_0} Y_i \rangle = \langle Y_0, \nabla_{Y_0} Y_i \rangle, \tag{11.5}$$

since $\nabla_{Y_0} Y_0 = \nabla_{\alpha'(s)} \alpha'(s) = 0$, because $\alpha_p(s)$ is a geodesic. We know the Levi-Civita connection is torsion free. Thus,

$$\nabla_{Y_0}(Y_i) - \nabla_{Y_i}(Y_0) = [Y_0, Y_i] = 0,$$

($[Y_0, Y_i] = 0$,because Y_0 and Y_i are coordinate vector fields). Hence, $\langle Y_0, \nabla_{Y_0} Y_i \rangle = \langle Y_0, \nabla_{Y_i} Y_0 \rangle$. On the other hand, since $\langle Y_0, Y_0 \rangle = -1$ along α_p, we must have $0 = \nabla_{Y_i} \langle Y_0, Y_0 \rangle = 2 \langle Y_0, \nabla_{Y_i} Y_0 \rangle$ and therefore the term on the right hand side of 11.5 is equal to zero and consequently $dg_{0i}(\alpha_p(s))/ds = 0$, as wanted. Now, let's determine the timelike geodesics in these coordinates. We claim that the curve $O_p(s) = (s, p)$ is a geodesic. In fact, let $l^a = x^a(O_p(s))$. Hence, $dl^0/ds = 1$, and $dl^i/ds = 0$. Hence, the systems of equations that a geodesic must satisfy reduces to a single equation $d^2 l^0/ds^2 = -\Gamma^c_{00}(O_p(s))$. But formula **??** implies that

$$\Gamma^c_{00} = \frac{1}{2} \sum_d g^{cd} \left(\frac{\partial g_{0d}}{\partial x^0} + \frac{\partial g_{0d}}{\partial x^0} - \frac{\partial g_{00}}{\partial x^d} \right) = 0,$$

since g_{00} and g_{0d} are constants. Consequently, the equation of a geodesic is just the equation of a straight line: $d^2 l^0/ds^2 = 0$, clearly satisfied by $O_p(s)$. On the other hand, $\mathfrak{n}_p = \partial_{x^0}$ satisfies $\langle \mathfrak{n}_p, \mathfrak{n}_p \rangle = -1$, and it is a vector orthogonal to any v in $T_p(S)$. Thus, $\mathfrak{n}_p$ must be equal to n_p. But a geodesic is determined by its initial condition, hence one must have $O_p(s) = \alpha_p(s)$. This proves the second assertion of the proposition. Finally, the third assertion is clear: $\alpha_p(s)$ being a geodesic satisfies $\left\langle \alpha'_p(s), \alpha'_p(s) \right\rangle = \langle n_p, n_p \rangle = -1$, for all s. Thus, the proper time for any observer O_p between events (t_0, p) and (t_1, p) is given by:

$$\int_{t_0}^{t_1} \sqrt{- \left\langle \alpha'_p(s), \alpha'_p(s) \right\rangle} ds = \int_{t_0}^{t_1} \sqrt{- \langle n_p, n_p \rangle} ds = t_1 - t_0,$$

and therefore does not depend on the observer O_p. □

Example 11.2.10.

1. In flat space-time, with standard coordinates $x = (x^a)$, the curve $O(s) = (s, \underline{0})$ represents an observer whose world line is the x^0- axis. A Lorentz frame at each $p \in O$ is given by $z^a = x^a - x^a(p)$. Any observer whose world line is a geodesic in flat space time, that is, a straight line in the standard coordinates, will be called an *inertial observer*.

2. Let us suppose that $\overline{O}$ is any observer in flat space-time that coincides with the standard inertial observer O at the origin, and at time $x^0 = 0$ (in O's watch) and moves in the direction of the x-axis at constant speed u. Its world line is represented by $\overline{O}(s) = (s, \mathrm{u}s, 0, 0)$ in the frame of reference of O. Let's determine a Lorentz frame $z = (z^a)$ for $\overline{O}$ at a point $p \in \mathrm{im}O(s_0)$. We denote the standard basis for $T_p(\mathbb{R}^4)$ by $E = \{e_a : e_a = \partial_{x^a}|_p\}$. In this basis $\overline{O}'(s) = e_0 + \mathrm{u}e_1$. In vector notation $\overline{O}'(s) = [1, \mathrm{u}, 0, 0]$. Now, $\left|\overline{O}'(s)\right| = \sqrt{-1 + \mathrm{u}^2}$, and therefore its normalization is equal to

$$\mathbf{v}_0 = [1, \mathrm{u}, 0, 0]/\sqrt{1 - \mathrm{u}^2} = [l, l\mathrm{u}, 0, 0],$$

with $l = 1/\sqrt{1 - \mathrm{u}^2}$. Define $v_1 = e_1 + \langle e_1, \mathbf{v}_0 \rangle \, v_0$. Since $\langle \mathbf{v}_0, \mathbf{v}_0 \rangle = -1$, it follows that $\langle v_1, \mathbf{v}_0 \rangle = 0$. We define $v_1 = v_1/|v_1|$, and $v_j = e_j$, for $j = 2, 3$. It is clear that the vectors v_a form an orthonormal basis B for $T_p(\mathbb{R}^4)$. A little calculation shows:

$$\begin{aligned} v_1 &= [0, 1, 0, 0] + l\mathrm{u}[l, l\mathrm{u}, 0, 0] = [l^2\mathrm{u}, 1 + l^2\mathrm{u}^2, 0, 0] \\ &= [l^2\mathrm{u}, l^2, 0, 0] = l^2[\mathrm{u}, 1, 0, 0], \end{aligned}$$

since $1 + l^2\mathrm{u}^2 = l^2$. Thus, $|v_1| = l$, and therefore $v_1 = v_1/|v_1| = l[\mathrm{u}, 1, 0, 0]$.

The bases-change matrix $[Id]_{EB}$ from the basis v_a to the standard basis e_a is given by

$$A = \begin{bmatrix} l & l\mathrm{u} & 0 & 0 \\ l\mathrm{u} & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{11.6}$$

Its inverse is the matrix:

$$A^{-1} = \begin{bmatrix} l & -l\mathrm{u} & 0 & 0 \\ -l\mathrm{u} & l & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This says that $e_0 = lv_0 - l\mathrm{u}v_1$, $e_1 = -l\mathrm{u}v_0 + lv_1$, $e_2 = v_2$, $e_3 = v_3$.

According to Proposition 11.2.8, Lorentz coordinates for $\overline{O}$ at any $p = \overline{O}(s_0)$ can be constructed by letting

$$\begin{aligned} z^0(q) &= l\left(x^0(q) - \mathrm{u}x^1(q)\right) \\ z^1(q) &= l(x^1(q) - \mathrm{u}x^0(q)) \\ z^2(q) &= x^2(q) \\ z^3(q) &= x^3(q) \end{aligned}$$

In more standard notation, if (t, x, y, z) and $(\bar{t}, \overline{x}, \overline{y}, \overline{z})$ denote the coordinates x and z, respectively, the formulas for the change of coordinates between O's frame and $\overline{O}$'s frame are:

$$\begin{aligned} &\bar{t} = l(t - \mathrm{u}x),\ \overline{x} = l(x - \mathrm{u}t),\ l = \frac{1}{\sqrt{1-\mathrm{u}^2}}. \\ &\overline{y} = y,\ \overline{z} = z. \end{aligned} \tag{11.7}$$

11.3 Light Cones

Let M be space-time with certain Lorentz metric $g = \langle -, - \rangle$. At any point $p \in M$, the timelike vectors can be separated into two classes: Two vectors v,w are in the same class $(v \sim w)$ if $\langle v, w \rangle < 0$.

Claim 11.3.1. The relation $\sim$ is an equivalence relation.

Proof. The only nontrivial assertion is the transitivity of $\sim$. We first notice that if v, w are two timelike vectors, then $\langle v, w \rangle \neq 0$. If the product were null, v and w would be orthogonal. Then, using the Gram-Schmidt algorithm (2.3) one could construct vectors in $T_p(M)$, $e_0 = v/|v|$, $e_1 = w/|w|$, e_2, e_3 such that $\langle e_a, e_b \rangle = 0$, for all $a \neq b$, and $\langle e_0, e_0 \rangle = \langle e_1, e_1 \rangle = -1$, while

$\langle e_k, e_k\rangle = 1$, for $k = 2, 3$. But this will imply that the signature of g would be $\{-1, -1, 1, 1\}$, which is a contradiction. Hence, if v, w are time like one must have $\langle v, w\rangle < 0$ or $\langle v, w\rangle > 0$. Now we prove that if v_0, v_1, v_2 are timelike vectors such that $\langle v_0, v_1\rangle < 0$ and $\langle v_1, v_2\rangle < 0$, then it can not happen that $\langle v_0, v_2\rangle > 0$. Without loss of generality, after dividing v_0 by its norm we may assume $\langle v_0, v_0\rangle = -1$. Let H denote the orthogonal subspace to v_0, and let v_1' be the projection of v_1 onto H. Define $e_1 = v_1'/|v_1'|$. Let e_2, e_3 be vectors in H such that if v_0 is denoted by e_0, the set $B = \{e_a : a = 0, 1, 2, 3\}$ is an orthonormal basis for $T_p(M)$. We can write v_1 as $v_1 = a_0e_0 + \widetilde{a}_1e_1$. Since $-a_0 = \langle v_0, v_1\rangle < 0$, then $a_0 > 0$. Hence, we may change v_1 by $\frac{1}{a_0}v_1$ without changing the relevant hypotheses. Thus, in the basis B one could write $v_0 = [1, 0, 0, 0]$, $v_1 = [1, a_1, 0, 0]$, with $a_1 = \widetilde{a}_1/a_0$, and $v_2 = [b_0, b_1, b_2, b_3]$. We want to prove that $\langle v_0, v_2\rangle = -b_0 < 0$. Now, by hypothesis

$$\begin{aligned}&\langle v_1, v_1\rangle = -1 + a_1^2 < 0, \quad \langle v_2, v_2\rangle = -b_0^2 + \textstyle\sum_i b_i^2 < 0,\\ &\langle v_1, v_2\rangle = -b_0 + a_1b_1 < 0.\end{aligned}$$

From this one obtains:

$$|a_1| < 1,\ b_0^2 > \textstyle\sum_i b_i^2 \geq b_1^2, \text{ hence } |b_0| > |b_1|\,, \text{ and } b_0 > a_1b_1. \tag{11.8}$$

Multiplying the first inequality by $|b_0|$ we see that $|b_0| > |a_1|\,|b_0|$. From the second inequality one obtains $|a_1|\,|b_0| > |a_1|\,|b_1|$. Thus, $|b_0| > |a_1b_1|$. Henceforth, the latter inequality implies that if had $b_0 < 0$, then a_1b_1 would be greater that b_0, contradicting the third inequality. Consequently $b_0 > 0$, as we wanted to prove. □

A vector that is timelike or null is called a *causal vector*. Causal vectors in $T_p(M)$ form a closed cone, i.e., if v is timelike or null, then so it is λv, for any $\lambda \neq 0$. This cone is divided into upper and lower cones, according to the partition induced by $\sim$. We arbitrarily choose one of these, and call it *the cone of future directed vectors*. The other one is called the *cone of past directed vectors*:

Definition 11.3.2.

1. A space-time manifold (M, g) is called *time-orientable* if there is a continuous choice of future directed and past directed vectors for all tangent spaces $T_p(M)$.

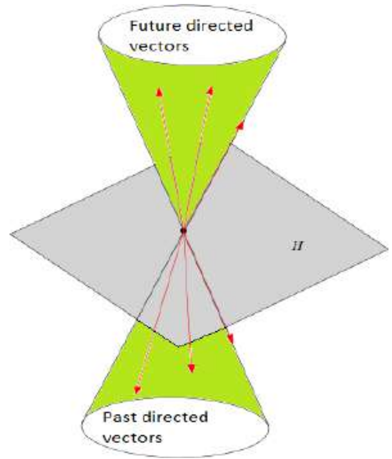


Figure 11.4: Causal cones

2. Let p be an event in a time-orientable space-time manifold (M, g). The *future cone* of p, $C^+(p)$, is defined as the set of all points q in M that can be reached from p along a causal curve whose tangent vector is future directed at all points. Similarly, we define the *past cone* of p, $C^-(p)$, as the set of all points q that can be reached from p along causal past directed curves.

Chapter 12

Examples of Space-Time Manifolds

12.1 Gravitational Waves

In flat space-time we consider the metric given in standard coordinates (t, x, y, z) by a perturbation of the Minkowski metric:

$$g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 + \mathrm{a} \sin \omega(t-z) & 0 & 0 \\ 0 & 0 & 1 - \mathrm{a} \sin \omega(t-z) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where a and ω are positive constants. Proposition 11.2.4 guarantees that this coordinates have physical meaning for an observer whose world line is $O(t) = (t, \underline{0})$, at any point $p \in O$. This example corresponds to a *plane gravitational wave that propagates along the z-axis.* The Christoffel symbols can be computed (for instance, with the program Maple) for this metric.

They vanish, except for

$$\begin{aligned}
\Gamma^0_{11} &= (\mathrm{a}/2)\cos\omega(z-t),\ \Gamma^0_{22} = -\frac{\mathrm{a}}{2}\cos\omega(z-t)\\
\Gamma^1_{01} &= \Gamma^1_{10} = \frac{1}{2}\frac{a\omega\cos\omega(t-z)}{1+\mathrm{a}\sin\omega(t-z)}\\
\Gamma^1_{13} &= \Gamma^1_{31} = \frac{\mathrm{a}}{2}\frac{-\omega\cos\omega(t-z)}{1+\mathrm{a}\sin\omega(t-z)}\\
\Gamma^2_{20} &= \Gamma^2_{02} = \frac{\mathrm{a}}{2}\frac{\omega\cos\omega(t-z)}{\mathrm{a}\sin\omega(t-z)-1}\\
\Gamma^2_{23} &= \Gamma^2_{32} = -\frac{\mathrm{a}}{2}\frac{\omega\cos\omega(t-z)}{\mathrm{a}\sin\omega(t-z)-1}\\
\Gamma^3_{11} &= \frac{\mathrm{a}}{2}\omega\cos\omega(t-z)\\
\Gamma^3_{22} &= -\frac{\mathrm{a}}{2}\omega\cos\omega(t-z).
\end{aligned}$$

Let P be a particle whose world line is given by $O(s) = (s, x(s), 0, 0)$ located originally at distance d_0 in the x-axis. This just means that $x(0) = \mathrm{d}_0$. We also assume it is a rest at time $t = 0$, that is $x'(0) = 0$, and *moves freely*, which just means that it moves in space-time following a geodesic.

Since $\Gamma^c_{00} = 0$, we notice, as in the proof of Proposition 11.2.9, that $\gamma(s) = (s, \mathrm{d}_0, 0, 0)$ is the only geodesic with initial conditions $x^i(0) = (\mathrm{d}_0, 0, 0)$, and $\gamma'(0) = \partial_t$. But *even though the spatial coordinates of P remain fixed, the distance $d(t)$ from P to the origin varies with t*. In fact, at each time t the spatial part of the metric g is given by

$$A = \begin{bmatrix} 1+\mathrm{a}\sin\omega(t-z) & 0 & 0\\ 0 & 1-\mathrm{a}\sin\omega(t-z) & 0\\ 0 & 0 & 1\end{bmatrix}.$$

On each slice $t = t_0$ we may compute the distance $d_{p,q}$ between any two points $p = (t_0, 0, 0, 0)$, and $q = (t_0, \mathrm{b}, 0, 0)$: If we parametrize the segment from p to q as $L(z) = (t_0, z, 0, 0)$, with $0 \le z \le \mathrm{b}$, we see that $L'(z) = \partial_x$, and therefore

$$\begin{aligned}
d_{p,q} &= \textstyle\int_0^{\mathrm{b}} \sqrt{\langle L'(z), L'(z)\rangle}dz = \int_0^{\mathrm{b}}\sqrt{1+\mathrm{a}\sin(\omega t_0)}dz\\
&= \mathrm{b}(1+\mathrm{a}\sin(\omega t_0))^{1/2} \simeq \mathrm{b}(1+\frac{\mathrm{a}}{2}\sin(\omega t_0)).
\end{aligned}$$

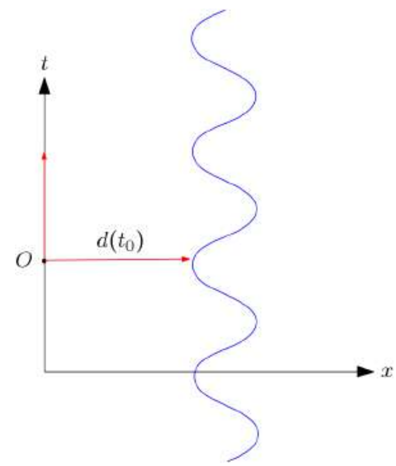


Figure 12.1: Gravitational Waves

Taking $\mathrm{b} = \mathrm{d}_0$, at $t = 0$, we then see that at any t the distance $d(t)$ form P to the origin decreases an increases cyclically with a period $\lambda = 2\pi/\omega$. If $\Delta d(t) = d(t) - \mathrm{d}_0$, then $\Delta d(t)/\mathrm{d}_0 = \mathrm{a}/2\sin(\omega t_0)$.

The first gravitational wave ever detected was measured on September 6, 2015, in the Laser Interferometer Gravitational-Wave Observatories, LIGO, located in Hanford, in the state of Washington, and in Livingston, Louisiana, USA. These interferometers can detect a change in the 4 km mirror spacing of less than a ten-thousandth the charge diameter of a proton, equivalent to measuring the distance to Proxima Centauri with an accuracy smaller than the width of a human hair! When a gravitational wave passes through the interferometer, the space-time in the local area is altered. Depending on the source of the wave and its polarization, this results in an effective change in length of one or both of the cavities. The effective length change between the beams will cause the light in the cavity to become slightly out of phase with respect to the incoming light. This results in a measurable signal. The actual length change measured in such experiments was of the order of $\Delta d(t) \simeq 10^{-21}\mathrm{d}_0$ meters. (https://en.wikipedia.org/wiki/LIGO#Observatories).

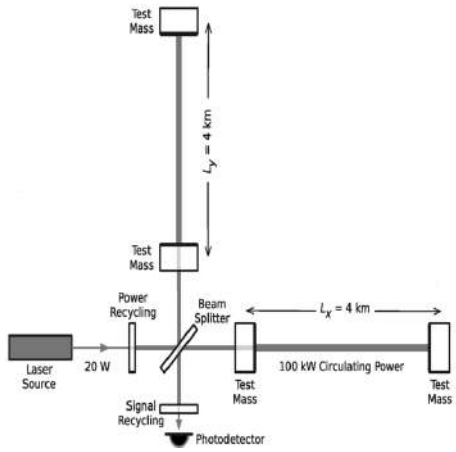


Figure 12.2: LIGO interferometer

12.2 Rotating Universes

12.2.1 Rotating observer

Let (t, x, y, z) be the standard coordinates of $\mathbb{R}^4$. Let O be an observer that moves in a circle of radius r_0 at a constant angular speed ω (in the positive direction) in the xy-plane, and that starts moving at a point with polar coordinates (r_0, θ_0). Let's denote by (t, r, θ, z) the standard polar coordinates for space-time $\mathbb{R}^4$.

The equations for change of coordinates are: $t = t$, $x = r\cos(\theta)$, $y = r\sin(\theta)$, $z = z$. Hence, the Minkowski metric in these coordinates will look in matrix form like $G = J^*MJ$, where J denotes the Jacobian matrix of change of coordinates:

$$J = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -r\sin(\theta) & 0 \\ 0 & \sin(\theta) & r\cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus,

$$G = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{12.1}$$

or in a more classical notation $ds^2 = -dt^2 + dr^2 + r^2 d\theta^2 + dz^2$.

In polar coordinates, any such observer O is given by the spiral curve

$$O(s) = (s,\ r_0, \theta_0 + \omega s,\ 0), \quad s \geq 0.$$

We define new coordinates for $\mathbb{R}^4 : (t, r, \widetilde{\theta}, z)$, where $\widetilde{\theta} = \theta - \omega t$. These

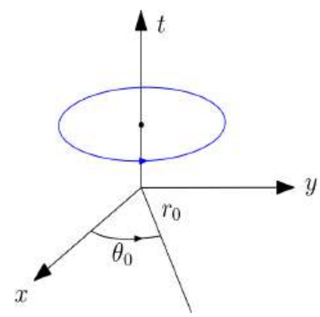


Figure 12.3: Rotation

coordinates have a clear geometrical meaning: A point p on the plane $t = t_0$ is given polar coordinates $r, \widetilde{\theta}, z$, where $\widetilde{\theta}$ is the angle $\widetilde{\theta}(p) = \theta(p) - \omega t_0$. This angle is the standard polar angle with respect to a plane that has been rotated an angle ωt in the positive direction (Figure below).It is clear that

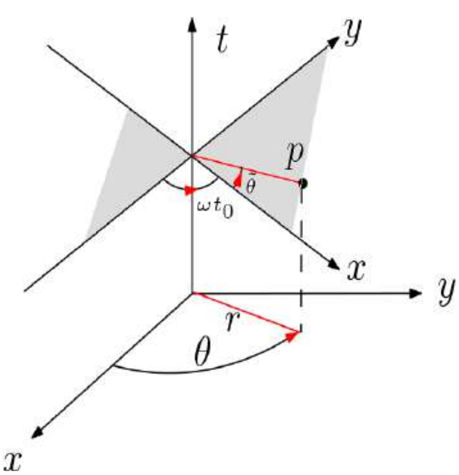


Figure 12.4: Rotation

we we omit the z-coordinate, the world line of O in these coordinates looks like a straight line that goes up in the direction of time, whereas the spatial coordinates remain fixed, that is: $O(s) = (s,\, r_0,\, \theta_0,\, 0)$.

A first computation shows that the Minkowski metric in the new coordinates $(t, r, \widetilde{\theta}, z)$ looks like

$$B = \begin{bmatrix} \omega^2 r^2 - 1 & 0 & \omega r^2 & 0 \\ \omega r^2 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

A Lorentz frame for O can be constructed as follows: XXXXX

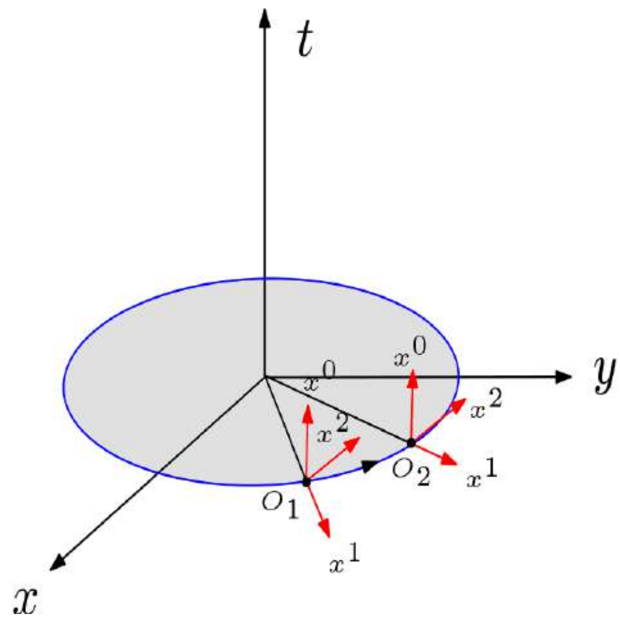


Figure 12.5: Rotating Universe

12.3 Rotating Universes of Godel XXXXX

12.4 The FRLW Metric

The Friedmann–Lemaître–Robertson–Walker (FLRW) metric corresponds to an exact solution of Einstein's field equations of GR for a homogeneous, isotropic, expanding or contracting universe. In Chapter (REF) we will define these notions in a precise manner, and will develop these ideas in a comprehensive manner. To illustrate some of the notions introduced so far, we consider in this section the case of a universe with flat spatial geometry. The flat spatial case is not only interesting by itself, but it also provides a good approximated model for the universe as we know it today.

In suitable local coordinates $x = (x^a)$ (defined only for the half space $x^0 > 0$) the flat FRLW metric is given by the following expression

$$g = -dx^0 \otimes dx^0 + A(x^0)^2(\textstyle\sum_i dx^i \otimes dx^i), \tag{12.2}$$

where $A(x^0)$ is a function called *the spatial dilation factor*. Functions of the form $A(x^0) = \varepsilon \times (x^0)^q$, with $0 < q < 1$, and ε a suitable constant, provide useful models for a flat universe. The exponent q, as we shall see, is determined by the distribution of matter and radiation in the universe. By *matter* cosmologists refer to particles with positive rest mass, like proton, neutron, electron..., while *radiation* refers to particles with zero rest mass, like photons and neutrinos. In a matter dominated universe, the value of $q = 2/3$ provides an approximated model. For a radiation dominated universe, q can be taken as $1/2$, as we will discuss later. The constant ε is introduced as a normalization, so that the dilation factor $A(x^0)$ takes a unit value at present time, defined as the spatial slice collection of all events with time coordinate equal to $x^0 = t_0$, with $t_0 = 1.29 \times 10^{26}$ss ($13.7 \times 10^9$ years). Of course, $A(t_0) = 1$ forces $\varepsilon = (t_0)^{-q}$. This particular value of t_0 corresponds to the estimated age of the universe, 13.7 billion years. The reason for choosing this particular value will become apparent.

12.4.1 Null Geodesic and Causality

The FRLW model provides a good example of the usefulness of Gaussian coordinates, as defined in Proposition 11.2.9. A computation with the program Maple (using Formulas **??**) shows that in these coordinates the connection coefficients are the following: $\Gamma^0_{00} = 0$; $\Gamma^0_{i0} = \Gamma^0_{0i} = 0$; $\Gamma^0_{ii} = A(x^0)A'(x^0) = \varepsilon^2 q(x^0)^{2q-1}$. And, $\Gamma^i_{00} = 0$; $\Gamma^i_{j0} = \Gamma^i_{0j} = A(x^0)/A'(x^0) = x^0/q$, if $i = j$, and zero otherwise. Finally, $\Gamma^i_{jk} = 0$ (A' denotes A's derivative with respect to x^0).

From this we see that if we fix a, b, c, the curve $\gamma(s) = (s, a, b, c)$,$s > 0$, defines a geodesic and consequently each $O_{(a,b,c)}(s) = (s, a, b, c)$ is a commoving observer. Clearly $x = (x^a)$ are coordinates with physical meaning for each one of these observers, since these are actually Gaussian coordinates. In this case we already know that the proper time measured by any two observers will be the same, and coincides with the x^0 coordinate. We call the coordinate x^0 the *universal time or cosmic time*. Now, let's determine the

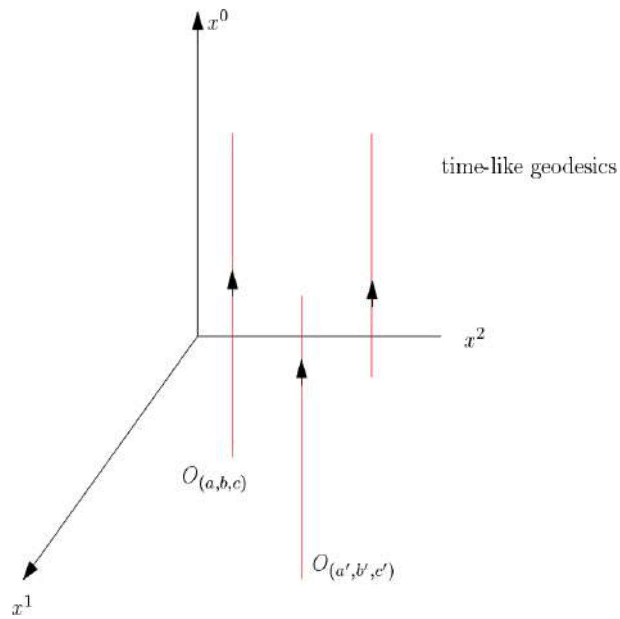


Figure 12.6: Observers in the FRLW space-time

equation of a general null geodesic $\gamma(s) = (u^a(s))$. We seek for a null geodesic that starts at a point $p = (0, a, 0, 0)$, and such that $u^2(s) = u^3(s) = 0$. The condition of being null means that

$$-\left(\frac{du^0}{ds}\right)^2 + \varepsilon^2 (u^0(s))^{2q} \left(\frac{du^1}{ds}\right)^2 = 0.$$

This is equivalent to $du^1/ds = \pm\varepsilon^{-1}(u^0)^{-q}\ du^0/ds$. On the other hand, by using the values of Γ^c_{ab} above one can see that the equations of a geodesic are the following:

$$\frac{d^2u^0}{ds^2} + \varepsilon^2 q (u^0)^{2q-1} (\frac{du^1}{ds})^2 = 0 \qquad (12.3)$$

$$\frac{d^2u^1}{ds^2} + 2\frac{u^0}{q}\frac{du^0}{ds}\frac{du^1}{ds} = 0. \qquad \text{(Geodesic equations for FRLW)}$$

By replacing $du^1/ds = \pm\varepsilon^{-1}(u^0)^{-q}du^0/ds$ into (12.3) we obtain:

$$\begin{aligned}&\frac{d^2u^0}{ds^2} + q(u^0)^{2q-1}(u^0)^{-2q}(\frac{du^0}{ds})^2\\ &= \frac{d^2u^0}{ds^2} + \frac{q}{u^0}(\frac{du^0}{ds})^2 = 0.\end{aligned} \tag{12.4}$$

On the other hand, integrating $du^1/ds = \pm\varepsilon^{-1}(u^0)^{-q}du^0/ds$ one obtains

$$u^1(s) = \pm\frac{u^0(s)^{1-q}}{\varepsilon(1-q)} + b, \tag{12.5}$$

from which one gets: $u^0(s) = (\varepsilon(1-q))^{1/(1-q)}(\pm u^1(s)-b)^{1/(1-q)}$. For $q = 2/3$, and $\varepsilon = (t_0)^{-2/3}$ this equation becomes: $u^0(s) = \pm\frac{\varepsilon^3}{27}(u^1(s)-b)^3$. The null geodesic that starts at $p = (0, b, 0, 0)$ is then given by

$$u^0(s) = \begin{cases} \frac{\varepsilon^3}{27}(u^1(s)-b)^3, & u^1 > b \\ \frac{-\varepsilon^3}{27}(u^1(s)-b)^3, & u^1 < b \end{cases} \tag{12.6}$$

where u^0, u^1 satisfy (12.4) and (12.5).

On the other hand, the future and past cones at each point $p = (t, a, 0, 0)$ are determined by lines at p with slopes $\pm A(t)$. In fact, a vector $v = \lambda_0\, \partial_{x^0}|_p + \lambda_1\, \partial_{x^1}|_p$ has norm zero iff $\lambda_0^2 = \lambda_1^2 A^2(t)$, i.e., iff $\lambda_0/\lambda_1 = \pm A(t)$:From this we see that events (like p and q in Figure above) that are too separated in space may not have any causal relation, since the past cones of both events are disjoint.

Let $q = (q_a)$ be the coordinates of an event q, expressed in the in global, Gaussian frame of reference in which we are writing the FRLW metric. The cosmic time coordinate $q_0 = t(q)$ of q, a notion that a priory depends on the chosen coordinates, would correspond to the time elapsed since the "beginning of the universe, the singularity $x^0 = 0$". However, let's see that we may recover q_0 intrinsically in terms of the metric: The coordinate q_0 coincides with the supremum of the proper times of all timelike observers lying in the past cone of q: Let's assume $\beta(s) = (c^a(s))$ is any timelike curve that ends at q. By letting $\tau = c^0(s)$ we may reparametrize β as $\beta(\tau) = (\tau, \underline{b^i(\tau)})$. Suppose $\beta(0) = p$ and $\beta(\tau_1) = q$, the proper time from event p to event q for this observer is then given by $T_\beta = \int_0^{\tau_1} \sqrt{1 - \sum_i (db^i/d\tau)^2}d\tau$. But $\sqrt{1 - \sum_i (db^i/d\tau)^2} \leq 1$, and therefore $T_\beta \leq \tau_1$. But $\tau_1 = q_0 = t(q)$, as we intended to show.

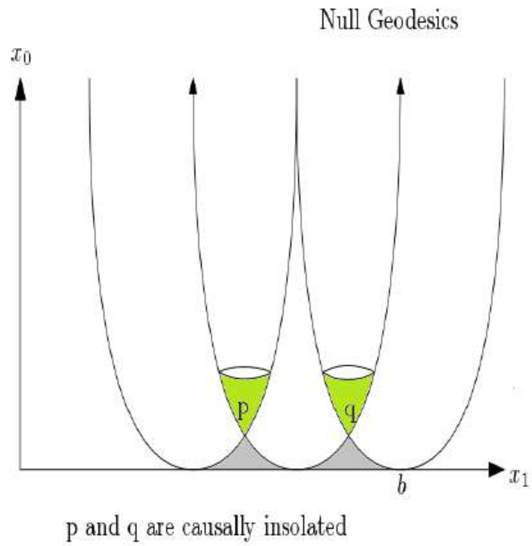
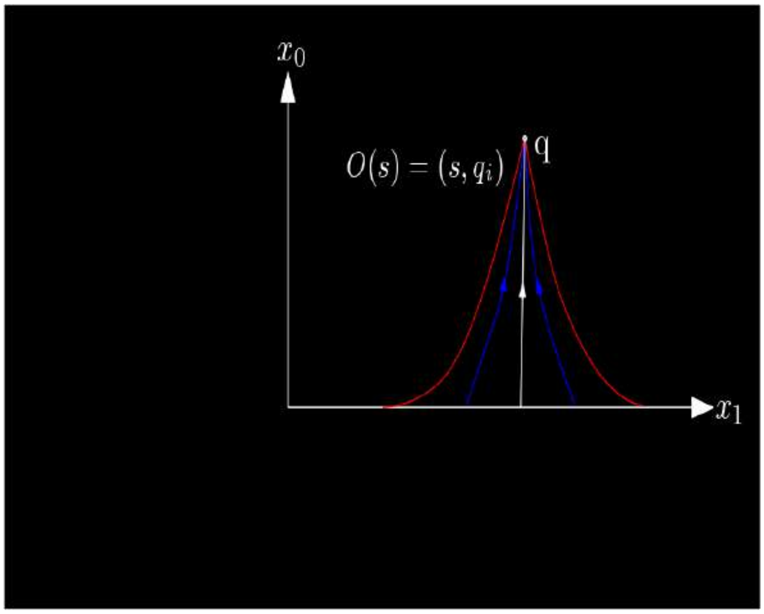


Figure 12.7: p and q are causally insolated

12.4.2 Cosmological Red Shift and Hubble's Law

Now we consider the case of a more general dilation function $A(x^0)$, that we assume to be some positive, smooth, increasing function. Let $O_b(s) = (s, b, 0, 0)$ and $O_a(s) = (s, a, 0, 0)$, with $b > a$, be two commoving observers. Suppose O_b emits a light signal at time t_1, and a short moment later, $t_2 = t_1 + h$, a second signal. Suppose O_a receives these two signals at times t'_1 and t'_2, respectively. We want to estimate $h' = t'_2 - t'_1$ as well as the quotient h/h'. We know that a null geodesic $\gamma = (u^0(s), u^1(s), 0, 0)$, with $\gamma(0) = (t_1, b, 0, 0)$, must satisfy (just for being a null curve) the equation: $(du^0/ds)^2 + A^2(u^0(s))(du^1/ds)^2 = 0$. Thus, $du^1/ds = \pm(du^0/ds)/A(u^0(s))$. Suppose that at $0 < s_1$, $\gamma(s_1) = (t'_1, a, 0, 0)$. Integrating both sides of the latter equation yields:

$$a - b = u^1(s_1) - u^1(0) = \pm \int_0^{s_1} \frac{du^1}{ds}(s)ds =$$
$$= \pm \int_0^{s_1} \frac{du^0/ds}{A(u^0(s))} ds = -\int_{t_1}^{t'_1} \frac{dz}{A(z)},$$



hence $b - a = \int_{t_1}^{t_1'} dz/A(z)$ (since $b - a > 0$, and $A(x^0) > 0$, one chooses the negative sign).

Similarly, let $\widetilde{\gamma}$ be a null geodesic such that $\widetilde{\gamma}(0) = (t_2, b, 0, 0)$, and $\widetilde{\gamma}(s_2) = (t_2', a, 0, 0)$. Integrating between $0 \leq s \leq s_2$ gives:

$$\int_0^{s_2} \frac{du^1}{ds}(s)ds = \pm \int_0^{s_2} \frac{du^0/ds}{A(u^0(s))}ds = \pm \int_{t_2}^{t_2'} \frac{dz}{A(z)},$$

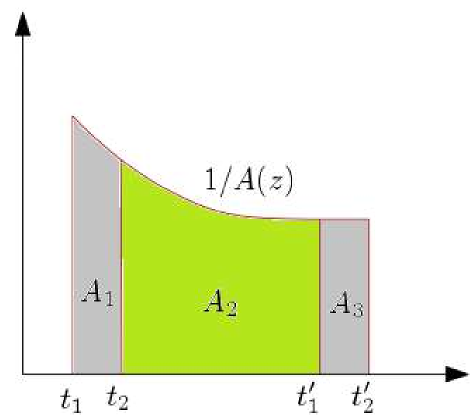
Henceforth, $b - a = \int_{t_2}^{t_2'} 1/A(z)dz$. This says that $A_1 + A_2 = A_2 + A_3$, and consequently $A_1 = A_3$. But in this example $h = t_1 - t_2$ and $h' = t_1' - t_2'$ will correspond to the period of some light signal, hence will be of the order of 10^{-6} ss, so one can use the Mean Value Theorem to estimate A_1 and A_3 as:

$$A_1 = \int_{t_1}^{t_2} \frac{dz}{A(z)} \simeq (t_2 - t_1)\frac{1}{A(t_1)} = \frac{h}{A(t_1)},$$

and

$$A_3 = \int_{t_1'}^{t_2'} \frac{dz}{A(z)} \simeq (t_2' - t_1')1/A(t_1') = \frac{h'}{A(t_1')}.$$

Thus, $h/A(t_1) = h'/A(t_1')$ (we could also have obtained $h/A(t_2) = h'/A(t_2')$). If h corresponds to the period of a circular wave, then its frequency at emission and reception will be $\omega_e = 2\pi/h$, and $\omega_0 = 2\pi/h'$ respectively. Since



$A(x^0)$ is an increasing function one obtains:

$$\frac{h'}{h} = \frac{\omega_e}{\omega_0} = \frac{A(t_1')}{A(t_1)} > 1, \tag{12.7}$$

and consequently $\omega_0 < \omega_e$. This phenomenon is the famous *cosmological redshift*, the fact that the wavelength of emitted radiation is lengthened due to the expansion of the universe that shifts visible light toward the red side of the spectrum. Galaxies and other celestial objects seem to be moving away as observed from any commoving observer O_b. The redshift is some times interpreted as a "Doppler effect of light", a useful analogy for many purposes, but essentially a misleading comparison.

The quotient in (12.7) is usually written as $z + 1$, with

$$z = \frac{h' - h}{h} = \frac{A(t_1')}{A(t_1)} - 1 = \frac{A(t_1') - A(t_1)}{A(t_1)}. \tag{12.8}$$

The quantity z is called the *redshift factor* corresponding to the celestial object represented by O_b.

Definition 12.4.1. Set $a = 0$, and think of O_a as an observer on Earth. The proper distance or commoving distance of an observer O_b, at time $t = x^0$, is define as $d(t) = A(t)b$.

This notion of distance should not to be confused with the concept of coordinate distance. However, since we have calibrated the dilation factor so that $A(t_0) = 1$, coordinate distance and proper distance coincide at present

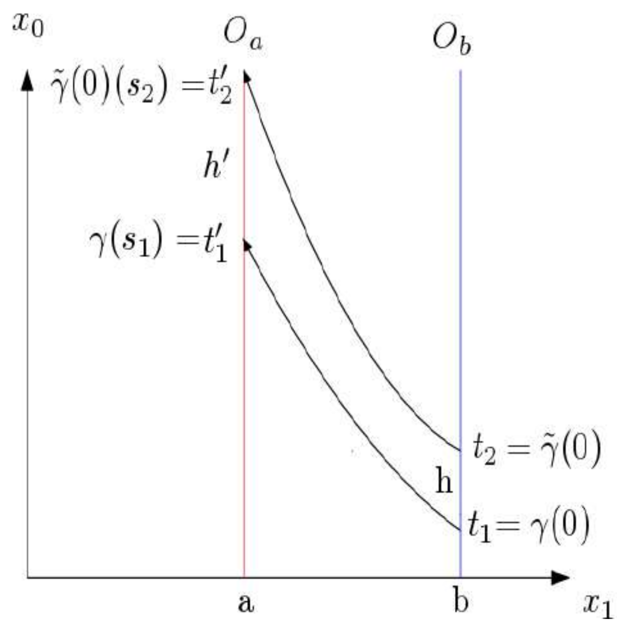


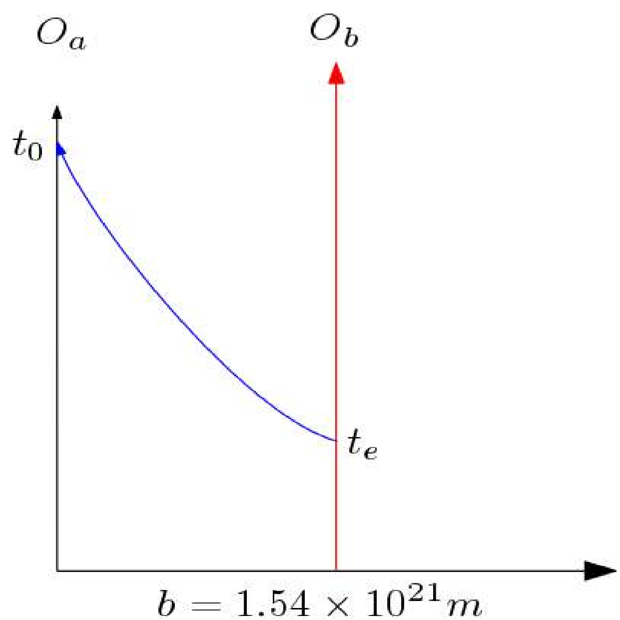
Figure 12.8: Cosmological Redshift

time. That is, $d_0 = d(t_0) = b$. Even the expression "proper distance" should be taken with a grain of salt: It would correspond to a measurable distance only if one could set an instantaneous system of roads and rules from O_a to O_b, a thought experiment that does not correspond to any real physical measurement like it would be, for instance, sending a light signal and then waiting for it to bounce back. This is why there are in cosmology many notions of distance between celestial objects. We will come back to that discussion later.

Let's interpret t_1' as the time coordinate of present time, that we denoted by $t_0 = 1.29 \times 10^{26}$ ss, and let's interpret t_1 as the coordinate of emission time (denote here by t_e) of a photon that is sent from O_b to O_a. Assume also that $a = 0$, and that O_a corresponds Earth:

Let's think for a moment that O_b corresponds to a celestial object known as RD1, discovered in 1988 in the constellation of Triangulum, a galaxy that until very recently held the title of the most distant galaxy in the universe [45]. For a dilation factor of the form $A(x^0) = \varepsilon(x^0)^q$, one has

$$d_0 = \int_{t_e}^{t_0} \frac{dz}{\varepsilon\ z^q} = \frac{1}{1-q}[(t_0)^{1-q} - (t_e)^{1-q}], \tag{12.9}$$



Its spectroscopic redshift has been measured to be $z = 5.34$. From Formula (12.8), and taking $q = 2/3$, we deduce that $A(t_0)/A(t_e) = 1 + 5.34 = 6.34$. Henceforth, $(t_0)^{2/3} = 6.34(t_e)^{2/3}$, and one obtains $t_e = t_0/(6.34)^{3/2} = 8.11 \times 10^{24}$ ss, approximately 858 millions of years. This means that the light we see today has taken about $t_0 - t_e = 1.21 \times 10^{26}$ ss, or 12.8 billion years to reach us. According to (12.9), the proper distance to the galaxy would be $d_0 = 3/\varepsilon((t_0)^{1/3} - (t_e)^{1/3}) = 2.34 \times 10^{26}$ m. But one light year is approximately 9.46×10^{15}m, henceforth, RD1 would be at a proper distance of 24.8×10^9 =24.8 billion light years. Its real proper distance is now estimated at about 26 billion light years, an error probably due to the fact that the exponent $q = 2/3$ may not be the exact value for the real distribution of matter and energy in our universe.

The apparent paradox of the existence of celestial objects whose distances in light years is much greater than t_0 is easily resolved if one takes into account the fact that the universe is expanding since the Big-Bang. In x^0-x^1coordinates (disregarding the coordinates x^2 and x^3) the FRLW metric could be visualized as the induced metric of a suitable bell like surface in $\mathbb{R}^3$, with a narrow neck near the Big-Bang singularity. A light signal traveling from O_b to O_a could be depicted as follows:From the diagram we see that once it reaches O_a, the proper distance has increased considerable to the

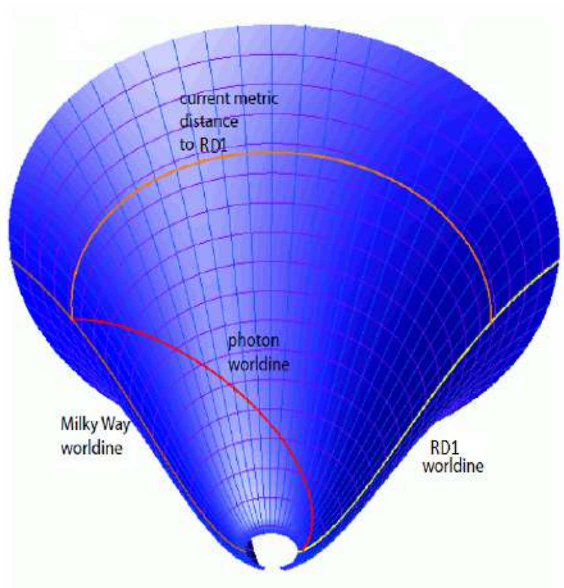


Figure 12.9: Coordinate and proper distance

value d_0.

12.4.3 Red Shift Factor for Nearby Celestial Objects

If we let $r = 1/(1-q)$, Taylor's expansion around $x = 0$ for the binomial function

$$(x+\delta)^r = x^r + r x^{r-1}\delta + \frac{r(r-1)}{2}\delta^2 x^{r-2} + \cdots$$

is everywhere convergent. When the coordinate distance b is small one may neglect terms of order ≥ 2. Using that $\varepsilon b = \int_{t_e}^{t_0} 1/z^q dz = (1/(1-q))[(t_0)^{1-q} - (t_e)^{1-q}]$ one gets

$$t_0 = [(1-q)\varepsilon b + (t_e)^{1-q}]^{1/(1-q)},$$

which can be approximated as:

$$\begin{aligned} t_0 &= t_e + \frac{1}{1-q}\varepsilon(t_e)^q(1-q)b + \cdots \\ &\simeq t_e + bA(t_e). \end{aligned}$$

That is, $t_0 - t_e \simeq bA(t_e)$ (since $\varepsilon \simeq 10^{-26}$ is very small, $b\varepsilon$ is also very small for nearby celestial objects).

We also notice that a dilation factor of the form $A(x^0) = \varepsilon(x^0)^q$, with $0 < q < 1$, has a very small derivative around $x^0 = t_0$, so that the variation of $A'(x^0)$ in the interval $t_e \leq \tau \leq t_0$ is negligible. In fact, in a scale where each unit of time is $\mathrm{Tm} = 10^{26}$ ss one has $|A'(T_0) - A'(T_e)| < 2 \times 10^{-5}$ Tm, when $|T_0 - T_e| < 10^{-4}$ Tm. For instance, for objects in the universe whose proper distance is at most 3.6 million light years (1 Mpc, as defined below), and so that $|t_0 - t_e| < 10^{22}$ ss, the variation of A' is less that 2×10^{-5}. Hence, we may use the Mean Value Theorem, $A(t_0) - A(t_e) = A'(\tau)(t_0 - t_e)$, for some $t_e \leq \tau \leq t_0$, to write (12.8) as

$$
\begin{aligned}
z &= \frac{A(t_0) - A(t_e)}{A(t_e)} \simeq \frac{A'(t_0)(t_0 - t_e)}{A(t_e)} \\
&= \frac{A'(t_0)bA(t_e)}{A(t_e)} = \frac{A'(t_0)}{A(t_0)}bA(t_0) = \frac{A'(t_0)}{A(t_0)}d_0,
\end{aligned} \tag{12.10}
$$

where d_0 is the proper distance to the galaxy at the present moment. The factor $\mathrm{H}(x^0) = A'(x^0)/A(x^0)$ is known as the *Hubble function*. Its value at present time is $\mathrm{H} = A'(t_0)/A(t_0)$, known as the *Hubble constant*. The apparent recession of galaxies and other objects in the universe was first measured by estimating the red shift factor z. Taking the derivative of $d(x^0) = A(x^0)b$ with respect to x^0 one sees that this "velocity of recession" is given by $v(x^0) = A'(x^0)b = d(x^0)\ A'(x^0)/A(x^0)$. Evaluating at $x^0 = t_0$ one obtains $v = \mathrm{H}d_0$, the famous Hubble's law, attributed to Edwin Hubble, although derived for the first time as a corollary of the equations of GR by Georges Lemaître, in 1927.

The current estimate for H is $70 + 2.4/ - 3.2$ (km/s)/(Mpc), where one megaparsec corresponds to 3.08×10^{22} m, or about 3.26 million light years. Notice that H has units of time^{-1}. Measured in ss, it is approximately equal to $\mathrm{H} = 7.56 \times 10^{-27}$ 1/ss.

One nearby celestial object O_b is the large Magellanic cloud:

(Ref: https://www.nasa.gov/) a galaxy "close" to Earth, with a proper distance of approximately 163 000 light years, equivalent to $d_0 = 1.54 \times 10^{21}$ m.

Its velocity of recession would be $v = \mathrm{H}d_0$, in units of m/ss, or $v = \mathrm{cH}d_0$ in m/s, equal in this example to 3.49 m/s.

Very distant objects may have a superluminal velocity of recession, i.e., greater than the speed of light! This does not contradict GR, since this "velocity of recession" is just a way of speaking, and rather corresponds to a

Figure 12.10: Magellanic Cloud

rate of change of the spatial part of the FRLW metric, and not to the speed at which an observer could send information.

12.4.4 Age and Diameter of the Visible Universe

On of the most striking consequences of the measurement of the Hubble's constant is that it provides a lower bound for the age of the universe. If we let $\mathrm{H} = A'(t_0)/A(t_0)$, then the tangent line to the curve $A(x^0)$ at t_0 intersects the x^0-axis at point $P = t_0 - 1/\mathrm{H}$ so that the distance between P and t_0 is $1/\mathrm{H}$, a constant known as the Hubble time, which corresponds to 1.32×10^{26} ss, approximately 13.9×10^9 years. Hence, as the picture shows, $1/\mathrm{H}$ provides a lower bound for the age of the universe, currently estimated to be $t_0 = 13.799 \pm 0.021$ billions of years. On the other hand, if we let $u^0(s) = t_0$, $u^1(s) = 0$ in equation $u^0(t) = \frac{-\varepsilon^3}{27}(u^1(s) - b)^3$, $u^1 < b$, one obtains the value $b = 3t_0$, which would be the radius of the visible universe: The value of b is 3.9×10^{26} m, or 41 billion light years. The current estimate, however, is 90 billion light years. This shows that our universe might not be matter dominated after all.

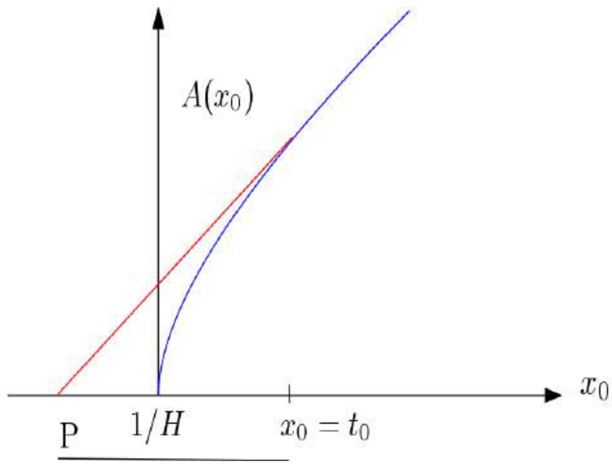


Figure 12.11: Hubble's constant

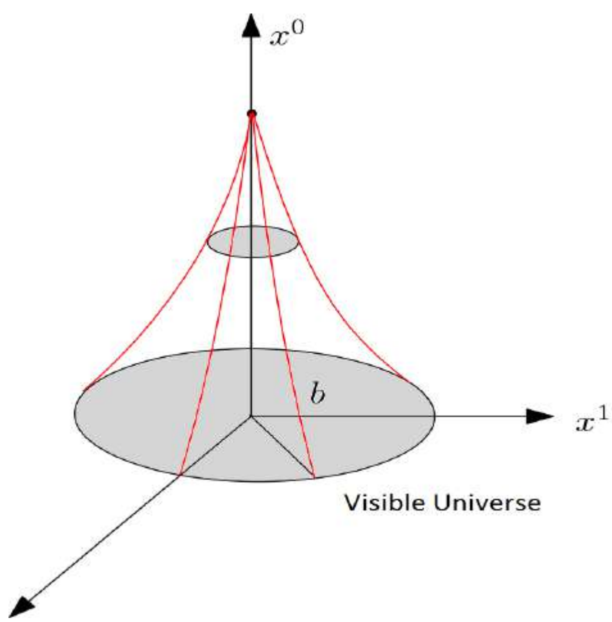


Figure 12.12: Hubble's constant

Chapter 13

Tensor of Energy-Momentum

13.1 Flows of Dust

In the last section we observed that the (mass $\equiv$ energy) of a particle is not an absolute quantity, but a scalar that depends on the measurement of a particular observer O (10.27). Therefore, if one wants to describe how energy is distributed in space-time one would need a mathematical gadget that takes the 4-velocity $\mathbf{u}_0$ of any particular observer, at some point p, and gives in return the energy density at that point, as measured by O. As we shall see, this can be achieved by means of a $(0, 2)$-tensor T with the property that $T_p(\mathbf{u}_0, \mathbf{u}_0)$ is precisely the energy density of the "fluid of energy" that flows through space-time at p, as measured by O (in units of $\rho = \text{energy}/\text{m}^3$).

To see how this tensor could be constructed we start by considering a fluid made of individual particles that flows in euclidean space $\mathbb{R}^3$ with constant 3-velocity $\mathrm{v} = \sum_i v^i e_i$. Let's look at a box B of total volume one, located at the origin of coordinates, where one of its sides is parallel to a vector v. We assume B contains N equal particles, each of rest mass m_0. For an observer $\overline{O}$ moving along with the box, the density of energy μ at some particular instant is just the sum of the rest energies of all the individual particles: $\mu = Nm_0$. We call this quantity *the rest energy density of the fluid*. Now, for the inertial observer O with standard coordinates $(t, \underline{x}^i)$ each particle would have mass $m_1 = m_0 l_{\mathrm{v}}$, where, as above, l_{v} denotes the factor $1/\sqrt{1 - |\mathrm{v}|^2}$. Due to length contraction in the direction of motion, the volume of the box, as measured by O, is equal to $\mathrm{vol}(B) = \sqrt{1 - |\mathrm{v}|^2}$. Hence, the energy density as measured by this observer would then be $\rho = Nm_1/\sqrt{1-|\mathrm{v}|^2} = \mu l_{\mathrm{v}}^2$.

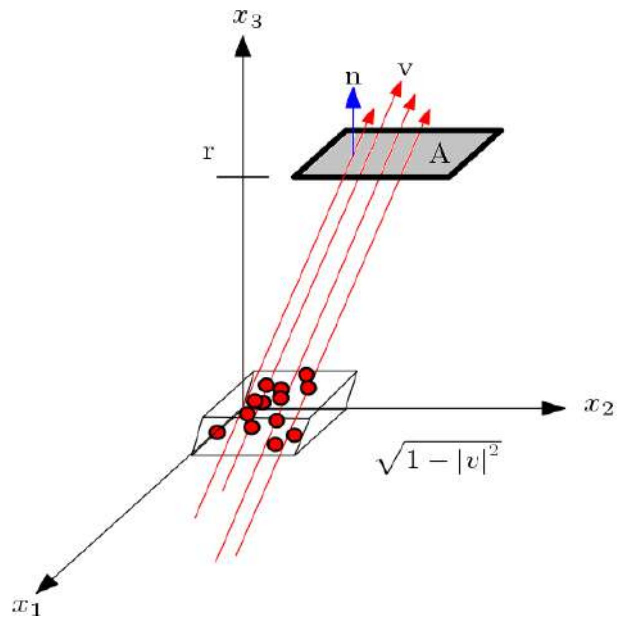


Figure 13.1: Energy-Momentum tensor

Now consider an element of area ΔA, parallel to the x^1x^2 plane, and located at $x^3 = r$. The number of particles that passes through this rectangular section in a time interval Δt is equal to $\rho \langle \mathrm{n}, \mathrm{v}\rangle \Delta A \Delta t = \rho v^3(\Delta A \Delta t)$, where $\mathrm{n} = \partial_{x^3}$ denotes the exterior unitary normal vector to ΔA. Hence, the flux of mass (energy) per unit area, per unit time (energy/m^2 $\times$ 1/ss) in the x^3-direction is just $\rho v^3 = \mu l_{\mathrm{v}}^2\, v^3$. Similarly, the flux of energy in any of the other coordinate directions x^1 and x^2 would be equal to $\rho v^1 = \mu l_{\mathrm{v}}^2\, v^1$, $\rho v^2 = \mu l_{\mathrm{v}}^2\, v^2$, respectively.

All this information, as we shall see in the next proposition, is contained in a $(2,0)$-tensor $\overline{T}$ (an upper bar will be used to distinguish it from the tensor T, obtained from $\overline{T}$ by lowering both indices) whose components in the inertial frame of O are given by:

$\overline{T}^{00} = \rho$, the *energy density*, as measured by O. (13.1)

$\overline{T}^{0i} = \rho v^i$, the *flux of energy* in the x^i-direction, as measured by O.

$\overline{T}^{ij} = \rho v^i v^j$.

We notice that the flux of energy in the x^i-direction, as measured by O, $\overline{T}^{0i} = \text{kg}/\text{m}^3 \times \text{m}/\text{ss}$, can also be interpreted as *momentum density*: $(\text{kg}\times\text{m}/\text{ss})/\text{m}^3$. On the other hand, how can we interpret the terms $\overline{T}^{ij}$? Let's take, for instance,$\overline{T}^{11} = \rho v^1 v^1$. If we look at its units we see these are

$$\begin{aligned}\overline{T}^{11} &= \frac{\text{kg}}{\text{m}^3}\times\frac{\text{m}}{\text{ss}}\times\frac{\text{m}}{\text{ss}} = (\frac{\text{kg}\times\text{m}}{\text{ss}^2})\times\frac{1}{\text{m}^2} \\ &= \frac{\text{Force}}{\text{m}^2} = \text{Pressure.}\end{aligned}$$

It is not difficult to see that $\overline{T}^{11}$ is precisely the force exerted by unit area (pressure) by the flux of particles against the surface located at $x^1 = r$, as shown in the figure above. Similarly, $\overline{T}^{22}$, $\overline{T}^{33}$ would represent the pressures by unit area exerted by the particles against surfaces with $x^2 = r$, and $x^3 = r$, respectively. The other entries $\overline{T}^{ij}$, with $i \neq j$, would measure stress forces in the mixed directions. This is why $\overline{T}$ is called the stress-energy-momentum tensor, or simply the energy-momentum tensor of the fluid. (As above, we put a bar above T only when there is room for confusion.)

A fluid of particles as discussed above can be modeled mathematically by a vector field S whose world lines have the form $\beta(s) = (s, sv^1, sv^2, sv^3)$. We think of each world line as representing a river of particles of density μ. Let $\mathbf{v} = \beta'(0)/|\beta'(0)| = l_{\text{v}}\partial_t + l_{\text{v}}\text{v}$ be the 4-velocity of β at the origin. The energy density $\overline{T}^{00}$ of this river, as measured by O at the origin, is then given by $\overline{T}^{00} = \mu l_{\text{v}}^2 = \mu v^0 v^0$ where $\mathbf{v} = \sum v^a \partial_{x^a}$. On the other hand, $\overline{T}^{0i} = \mu(l_{\text{v}})^2 v^i = \mu v^0 v^i$. This shows that the components of $\overline{T}$ should be given by $\overline{T}^{ab} = \mu v^a v^b$.

Proposition 13.1.1. Given a smooth function μ and a *unitary* timelike vector field S in space-time (M, g) it is possible to construct a symmetric $(2, 0)$-tensor $\overline{T}$ such that:

1. Fix $p \in M$, and let $\beta : I \to M$ be an integral curve of S, with $\beta(0) = p$. In an arbitrary coordinate system $y = (y^a)$ at p the tensor $\overline{T}$ takes the form $\overline{T} = \sum_{a,b}\overline{T}^{ab}\partial_{y^a} \otimes \partial_{y^b}$, with $\overline{T}^{ab} = \mu v^a v^b$, where $S = \sum_a v^a \partial_{y^a}$.

2. If the coordinates y correspond to a Lorentz frame for β at p, the components of $\overline{T}$ take the form $\overline{T}^{00} = \mu$, and $\overline{T}^{ab} = 0$, if $(a, b) \neq (0, 0)$.

Proof. Define $\overline{T} = \mu S \otimes S$. In a general coordinate system $y = (y^a)$, if one writes $S = \sum_a v^a \partial_{y^a}$, then $\overline{T} = \mu \sum_{a,b} v^a v^b \partial_{y^a} \otimes \partial_{y^b}$, and consequently $T^{ab} = v^a v^b$, as claimed. The second statement follows directly from the first. □

Definition 13.1.2.

1. The pair (μ, S) in space-time (M, g) is called a *fluid of dust.*

2. The tensor $\overline{T}$ is called the *energy-momentum tensor* of the fluid.

3. By T we denote the tensor obtained by lowering both indices of $\overline{T}$ (Section 2.4.5), a tensor that we will also call the *energy-momentum tensor of the flow.*

Definition 13.1.3. Let $O : I \to M$ be an observer in space time with 4-velocity $\mathbf{u}_0$.

1. The *energy density* of the flow, as measured by O at p, is defined as $T_p(\mathbf{u}_0, \mathbf{u}_0)$.

   If $x = (x^a)$ is a Lorentz frame for O at p, then one defines:

2. The *energy flux in the direction of* ∂_{x^i}, as $-T_p(\mathbf{u}_0, \partial_{x^i})$

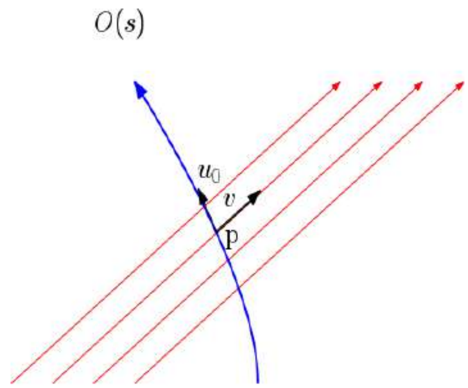
3. The *pressure of the fluid in the direction of* ∂_{x^i}, as $T_p(\partial_{x^i}, \partial_{x^i})$.

4. The *shear pressure (stress) of the fluid in the mixed directions,* as $T_p(\partial_{x^i}, \partial_{x^j})$.

This is a reasonable definition. In fact, fix any Lorentz frame for O at p, $x = (x^a)$, and let $\lambda^{\mathbf{u}_0} = \langle \mathbf{u}_0, - \rangle$. Then

$$\begin{aligned} T(\mathbf{u}_0, \mathbf{u}_0) &= \overline{T}(\lambda^{\mathbf{u}_0}, \lambda^{\mathbf{u}_0}) = \mu \textstyle\sum_{a,b} v^a v^b \partial x_a(\lambda^{\mathbf{u}}) \otimes \partial_{x^b}(\lambda^{\mathbf{u}_0}) \\ &= \mu \textstyle\sum_{a,b} v^a v^b \langle \mathbf{u}_0, \partial_{x^a} \rangle \ \langle \mathbf{u}_0, \partial_{x^b} \rangle = \mu \textstyle\sum_{a,b} v^a v^b \langle \partial_{x^0}, \partial_{x^a} \rangle \ \langle \partial_{x^0}, \partial_{x^b} \rangle \\ &= \mu v^0 v^0, \ \text{(The energy density of the flow)}, \end{aligned}$$

where the last two equalities hold because $x = (x^a)$ is a Lorentz frame. Similarly, one obtains $-T_p(\mathbf{u}_0, \partial_{x^i}) = \mu v^0 v^i$, and $T_p(\partial_{x^i}, \partial_{x^j}) = \mu v^i v^j$.



13.2 Perfect Fluids

Besides fluids of dust, there are some other cases of fluids of matter that are important in GR. These are fluids of matter that are only subjected to isotropic inner pressures, called *perfect fluids*, thus characterized by a tensor of energy-momentum that in the Lorentz rest of frame of an observer that moves along with the fluid takes the form

$$\overline{T} = \begin{bmatrix} \mu & 0 & 0 & 0 \\ 0 & \mathrm{P} & 0 & 0 \\ 0 & 0 & \mathrm{P} & 0 \\ 0 & 0 & 0 & \mathrm{P} \end{bmatrix}, \tag{13.2}$$

where μ gives its density, and P gives its inner pressure. Obviously, when $\mathrm{P} = 0$ the fluid corresponds to *dust, as* discussed above.

More precisely:

Proposition 13.2.1. Given two smooth function μ and P in space-time (M, g) and a unitary timelike vector field S one can construct a symmetric $(2,0)$-tensor $\overline{T}$ such that:

1. At any point $p \in M$, let $\beta : I \to M$ be the integral curve of S, with $\beta(0) = p$. Then in any Lorentz frame $z = (z^a)$ for β at p the components of tensor $\overline{T}$ take the form (13.2), i.e., $T^{00} = \mu$, $T^{ii} = \mathrm{P}$, and zero otherwise.

2. In an arbitrary coordinate system $y = (y^a)$ at p, the tensor $\overline{T}$ takes the form $\overline{T} = \sum_{a,b} T^{ab} \partial_{y^a} \otimes \partial_{y^b}$, with $T^{ab} = (\mu + \mathrm{P}) v^a v^b + g^{ab}\mathrm{P}$, where $S = \sum_a v^a \partial_{y^a}$.

Again, we denote by T the $(0,2)$-tensor obtained from $\overline{T}$ by lowering both indices.

3. In any Lorentz frame one has $T_{0i} = -T^{0i}$, and $T_{ab} = T^{ab}$ otherwise. In particular, in the frame $z = (z^a)$ of β one has $T_{00} = \mu$, $T_{ii} = P$, and zero otherwise.

Proof. It suffices to define $\overline{T} = (\mu + \mathrm{P})S \otimes S + \mathrm{P}g^{-1}$. Let $z = (z^a)$ be a Lorentz coordinate frame for β at p. Since S is unitary, the 4-velocity of β at p is just $\mathbf{v} = \partial_{z^0}$ and consequently $\overline{T}$ can be written in the frame z as $\overline{T} = (\mu + \mathrm{P})\partial_{z^0} \otimes \partial_{z^0} + \mathrm{P}g^{-1}$. But since z is a Lorentz frame

$$g^{-1} = \eta^{-1} = \eta = -\partial_{z^0} \otimes \partial_{z^0} + \textstyle\sum_i \partial_{z^i} \otimes \partial_{z^i}.$$

Thus, $T^{00} = \mu$, while $T^{ii} = P$, and zero otherwise. In an arbitrary coordinate system $y = (y^a)$ at p one has:

$$\begin{aligned} &\overline{T} = (\mu + \mathrm{P})S \otimes S + \mathrm{P}g^{-1}. \text{ Hence,} \\ &\overline{T} = (\mu + \mathrm{P})\textstyle\sum_{a,b} v^a v^b \partial_{y^a} \otimes \partial_{y^b} + \mathrm{P}\textstyle\sum_{a,b} g^{ab} \partial_{y^a} \otimes \partial_{y^b} \\ &\quad = \textstyle\sum_{a,b} (\mu + \mathrm{P}) v^a v^b + g^{ab}\mathrm{P}, \end{aligned}$$

as claimed in (1). The rest of the proof follows identical lines as the proof of Proposition (13.1.1). □

Definition 13.2.2. Let $O : I \to M$ be any observer with $p = O(s_0)$. Let $x = (x^a)$ be a Lorentz frame for O at p. If $\mathbf{u}_0$ is the 4-velocity of O at p we define:

1. The *energy density* measured by O at p to be $T_p(\mathbf{u}_0, \mathbf{u}_0)$.
2. The *flux of energy* at p in the direction x^i to be $-T_p(\mathbf{u}_0, \partial_{x^i})$.
3. The *pressure of the fluid* in the x^i direction as $T_p(\partial_{x^i}, \partial_{x^i})$.
4. The *shear pressure (stress)* of the fluid in the mixed directions as $T_p(\partial_{x^i}, \partial_{x^j})$.

Definition 13.2.3. The tensors $\overline{T}$ and T are both called the stress-energy-momentum tensor of the fluid. We notice that if $\mathrm{P} = 0$ the fluid is just a flow of dust.

13.3 The Energy Conservation Law

In space-time $(M, g = \langle -, - \rangle)$ the conservation law of energy and momentum can be summarized by just one equation. Denote by ∇ the Levi-Civita connection in M, and let $\overline{T}$ be a $(2,0)$-energy momentum tensor in space-time (M, g). Then the covariant derivative $\nabla_X \overline{T}$ is also a $(2,0)$-tensor. Define $L(A, B, X) = (\nabla_X \overline{T})(A, B)$ as the covariant derivative along X evaluated in two arbitrary $(0,1)$-tensor fields A and B (not to be confused with $\nabla_X(\overline{T}(A, B))$, the directional derivative in the direction of X of the smooth function $\overline{T}(A, B)$). The map $L : T^{(0,1)} \times T^{(0,1)} \times T^{(1,0)} \to \mathbb{R}$ is clearly $C^\infty(M)$-linear, and consequently defines a $(2,1)$-tensor in M. Denote its contraction ${C^1}_3\, L$ by $D\overline{T}$. This is clearly a $(1,0)$-tensor, called *the divergence* of $\overline{T}$.

With these preliminaries, the *local conservation law of energy in space-time* can be formulated as the vanishing of the divergence:

Definition 13.3.1. Let $\overline{T}$ be an energy momentum tensor of space-time (M, g).By the *local conservation of energy* we mean $D\overline{T} = 0$.

To understand what this law means, choose local coordinates $x = (x^a)$. Write $L = \sum_{r,s,t} {L^{r,s}}_t\ \partial_r \otimes \partial_s \otimes \omega^t$, and take $X = \partial_a$. Then $D\overline{T} = 0$ means that for each index b

$$D\overline{T}(\omega^b) = \textstyle\sum_a L(\omega^a, \omega^b, \partial_a) = \sum_a {L^{a,b}}_{\ a} = 0.$$

But ${L^{a,b}}_a = (\nabla_{\partial_a} \overline{T})^{ab}$. Thus the conservation of energy at each point $p \in M$ means that $\sum_a (\nabla_{\partial_a} \overline{T})^{ab} = 0$. In terms of T_{ab} this law can be formulated as $\sum_a (\nabla^a T)_{ab} = 0$, where the upper a-covariant derivative means $\nabla_{\partial_a} T^{(1)}$. Recall that $T^{(1)}$ denotes the tensor obtained from T after raising the first index. In local coordinates, its components are ${T^a}_b = \sum_s g^{as} T_{sb}$.Summarizing, the law of conservation of momentum can be written as:

$$\textstyle\sum_a (\nabla_{\partial_a} \overline{T})^{ab} = 0, \text{ or as } \sum_a (\nabla^a T)_{ab} = 0. \tag{13.3}$$

Physicists like to write these equations as $\nabla_\alpha T^{\alpha\beta} = 0$, and $\nabla^\alpha T_{\alpha\beta} = 0$, respectively.

13.3.1 Interpretation of The Conservation Law

We want interpret what the conservation law $\sum_a (\nabla_{\partial_a} \overline{T})^{ab} = 0$ means for a perfect fluid that moves in Minkowski space-time $(\mathbb{R}^4, \eta)$. Let's fix an arbitrary system of coordinates $x = (x^a)$ around a point p in the fluid, and write

$S = \sum v^a \partial x^a$. In these coordinates the components of the energy momentum tensor can be written as:

$$T^{ab}(x) = \textstyle\sum_{a,b}(\mu + \mathrm{P})(x)v^a(x)v^b(x) + \eta^{ab}(x)\mathrm{P}(x).$$

In flat space time the law of local conservation of energy becomes $\sum_a \partial T^{ab}/\partial x_a = 0$. Fixing the index b, and summing over the derivatives with respect to x^a one obtains:

$$E_b = \textstyle\sum_a \frac{\partial(\mu + \mathrm{P})}{\partial x^a} v^a v^b + \sum_a (\mu + \mathrm{P})[\frac{\partial v^a}{\partial x^a} v^b + v^a \frac{\partial v^b}{\partial x^a}] + \sum_a \eta^{ab} \frac{\partial \mathrm{P}}{\partial x^a} = 0. \quad (13.4)$$

Now assume $x = (x^a)$ are the standard coordinates of $\mathbb{R}^4$, (t, x^i). Equation 13.4 for $b = 0$ reads:

$$\textstyle\sum_a \frac{\partial(\mu + \mathrm{P})}{\partial x^a} v^a v^0 + \sum_a (\mu + \mathrm{P})[\frac{\partial v^a}{\partial x^a} v^0 + v^a \frac{\partial v^0}{\partial x^a}] + \sum_a \eta^{a0} \frac{\partial \mathrm{P}}{\partial x^a} = 0. \quad (13.5)$$

We now assume (μ, P, S) is a *classical* perfect fluid, that is a fluid that moves with a small 3-velocity. Since $|S| = 1$, if we write $S = v^0 \left(\partial_t + \sum_i (v^i/v^0)\partial_{x^i}\right)$, then $|v^0| \sqrt{1 - |\mathrm{v}|^2} = 1$, where $\mathrm{v} = \sum v^i/v^0 \partial x^i$ denotes the 3-velocity of the fluid. Since $|\mathrm{v}| << 1$ we may assume that v^0 is the constant function $v^0 \equiv 1$, and that $\mathrm{v} = \sum v^i \partial x^i$. Thus, Equation 13.5 becomes:

$$\begin{aligned}
&\textstyle\sum_a \frac{\partial(\mu + \mathrm{P})}{\partial x^a} v^a + \sum_a (\mu + \mathrm{P})[\frac{\partial v^a}{\partial x^a} + v^a \frac{\partial v^0}{\partial x^a}] - \frac{\partial \mathrm{P}}{\partial t} \\
&= \textstyle\sum_a \frac{\partial(\mu + \mathrm{P})}{\partial x^a} v^a + \sum_a (\mu + \mathrm{P}) \frac{\partial v^a}{\partial x^a} - \frac{\partial \mathrm{P}}{\partial t} \\
&= \textstyle\frac{\partial(\mu + \mathrm{P})}{\partial t} + \sum_i \frac{\partial(\mu + \mathrm{P})}{\partial x^i} v^i + \sum_i (\mu + \mathrm{P}) \frac{\partial v^i}{\partial x^i} - \frac{\partial \mathrm{P}}{\partial t} \\
&= \textstyle\frac{\partial \mu}{\partial t} + \sum_i \frac{\partial(\mu + \mathrm{P})}{\partial x^i} v^i + \sum_i (\mu + \mathrm{P}) \frac{\partial v^i}{\partial x^i} = 0.
\end{aligned}$$

where in the second and third line we used that $\partial v^0/\partial x^a = 0$, since $v^0 \equiv 1$ is a constant function. This last equation can be written as:

$$\frac{\partial \mu}{\partial t} + \textstyle\sum_i \frac{\partial((\mu + \mathrm{P})v^i)}{\partial x^i} = 0. \quad (13.6)$$

Since pressure comes from the random motion of individual particles, and here we are assuming the fluid is moving at very small velocities, we may assume P << μ so that $\mathrm{P} + \mu \simeq \mu$ and this equation becomes the classical equation of conservation of energy (Equation 8.5)

$$\frac{\partial \mu}{\partial t} + \mathrm{div}(\mu \mathrm{v}) = 0. \tag{13.7}$$

On the other hand, let's look at Equation $E_b = 0$ (13.5) for $b = i$. In this case one obtains:

$$\begin{aligned}
&\textstyle\sum_a \frac{\partial(\mu + \mathrm{P})}{\partial x^a} v^a v^i + \sum_a (\mu + \mathrm{P})[\frac{\partial v^a}{\partial x^a} v^i + v^a \frac{\partial v^i}{\partial x^a}] + \sum_a \eta^{ai} \frac{\partial \mathrm{P}}{\partial x^a} \\
&= \textstyle\sum_a \frac{\partial \mu}{\partial x^a} v^a v^i + \sum_a \mu \frac{\partial v^a}{\partial x^a} v^i + \sum_a \mu v^a \frac{\partial v^i}{\partial x^a} + \frac{\partial \mathrm{P}}{\partial x^i} \\
&= \left(\textstyle\sum_a \frac{\partial \mu}{\partial x^a} v^a + \sum_a \mu \frac{\partial v^a}{\partial x^a} \right) v^i + \sum_a \mu v^a \frac{\partial v^i}{\partial x^a} + \frac{\partial \mathrm{P}}{\partial x^i} = 0.
\end{aligned}$$

using again that $\mathrm{P} + \mu \simeq \mu$. Using one more time that $\partial v^0 / \partial x^a = 0$, this last equation can then be written as:

$$\left(\frac{\partial \mu}{\partial t} + \sum \frac{\partial (\mu v^j)}{\partial x^j} \right) v^i + \textstyle\sum_a \mu v^a \frac{\partial v^i}{\partial x^a} + \frac{\partial \mathrm{P}}{\partial x^i} = 0.$$

But the term inside the parenthesis is zero (Equation 13.7) thus:

$$\textstyle\sum_a \mu v^a \frac{\partial v^i}{\partial x^a} + \frac{\partial \mathrm{P}}{\partial x^i} = \mu \frac{\partial v^i}{\partial t} + \sum_j \mu v^j \frac{\partial v^i}{\partial x^j} + \frac{\partial \mathrm{P}}{\partial x^i} = 0.$$

The system of equations:

$$\begin{aligned}
\mu \frac{\partial v^1}{\partial t} + \textstyle\sum_j \mu v^j \frac{\partial v^1}{\partial x^j} + \frac{\partial \mathrm{P}}{\partial x^1} &= 0, \\
\mu \frac{\partial v^2}{\partial t} + \textstyle\sum_j \mu v^j \frac{\partial v^2}{\partial x^j} + \frac{\partial \mathrm{P}}{\partial x^2} &= 0, \\
\mu \frac{\partial v^3}{\partial t} + \textstyle\sum_j \mu v^j \frac{\partial v^3}{\partial x^j} + \frac{\partial \mathrm{P}}{\partial x^3} &= 0,
\end{aligned}$$

are the classical *Euler's equation in fluid mechanics for the motion of a fluid.* These are usually written in vector form as:

$$\mu[\partial_t \mathrm{v} + (\mathrm{v} \cdot \nabla)\mathrm{v}] = -\,\mathrm{grad}(\mathrm{P}),$$

([4]), Page 36, and ([22]), Chapter I.

Chapter 14

The Field Equations

14.1 The Gravitational Potential

In this chapter we derive Einstein's famous field equations $G = \mathsf{Ric} - \frac{1}{2}\mathrm{R}g = (8\pi \mathrm{G}_N/\mathrm{c}^4)T$, where G is the Einstein's tensor (**??**), T is the energy momentum tensor (13.3.1) and G_N is the universal gravitational constant, whose value is approximately 6.674×10^{-11} $\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$. In reality this equation will not be deduced from any other more basic principle: The field equations were originally put forth by Einstein as an analogue in space-time of the potential equation $\nabla^2\Phi = 4\pi\mathrm{G}_N\mu$, providing a revolutionary new explanation of the mystery of gravity.

We start by formulating Newton's law in the form of Poisson's equation.

We recall that the gravitational potential in Euclidian space $\mathbb{R}^3$ defined by a body M located at the origin of coordinates is given the scalar function $\Phi(\underline{x}) = -\mathrm{G}_N\mathrm{M}/\left|\mathrm{r}(\underline{x})\right|$, where we have written $(\underline{x})$ instead of (x^i), and where $\mathrm{r}(\underline{x}) = (x^1, x^2, x^3)$ denotes the vector position, and $\left|\mathrm{r}(\underline{x})\right| = [(x^1)^2 + (x^2)^2 + (x^3)^2]^{1/2}$ its euclidean norm. By a potential function one means that $U(\underline{x}) = \operatorname{grad}\Phi(\underline{x})$ is a vector field with the property that the gravitational force exerted on a particle of mass m located as a point p will be given by $F(p) = -mU(p)$. In fact,

$$\begin{aligned} U(\underline{x}) = \operatorname{grad}\Phi(\underline{x}) &= \frac{\mathrm{G}_N\mathrm{M}}{\left|\mathrm{r}(\underline{x})\right|^3}(x^1, x^2, x^3) \\ &= \frac{\mathrm{G}_N\mathrm{M}}{\left|\mathrm{r}(\underline{x})\right|^2}\frac{r(\underline{x})}{\left|\mathrm{r}(\underline{x})\right|} = \frac{\mathrm{G}_N\mathrm{M}}{\left|\mathrm{r}(\underline{x})\right|^2}\mathrm{u}(\underline{x}), \end{aligned}$$

where $\mathrm{u}(\underline{x}) = r(\underline{x})/\left|\mathrm{r}(\underline{x})\right|$ is the unitary vector determined by p. Newton's law of gravitation is the assertion

$$F(p) = \frac{Mm\mathrm{G}_N}{\left|\mathrm{r}(p)\right|^2}(-\mathrm{u}(p)) = -m\,\mathrm{grad}\,\Phi(p) = -mU(p)$$

as claimed.

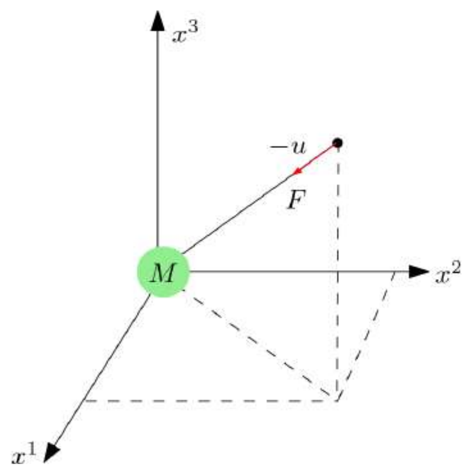


Figure 14.1: Inverse square law

14.2 Poisson's Equation

Poisson's equation is the analogue of Maxwell's first equation in electromagnetism, once the electrostatic potential is changed by the gravitational potential. In what follows, $U(\underline{x})$ will denote the gravitational vector field: $U(\underline{x}) = \mathrm{grad}\,\Phi(\underline{x})$. By $\mu(\underline{x})$ we will denote a function that gives the density of mass at the point with coordinates $(\underline{x})$. By a dot "$\cdot$" we will denote the standard inner product in $\mathbb{R}^3$.

Theorem 14.2.1. Let D be a domain in $\mathbb{R}^3$with smooth boundary.

1. The total flux of U along ∂D is equal to $4\pi\mathrm{G}_N$ times the total mass m contained inside D. This means that:

$$\int_{\partial D} U(\underline{x}) \cdot \mathrm{n}\; dA = 4\pi\mathrm{G}_N\mathrm{m},$$

where m $= \int_D \mu(\underline{x})dV$, and where n denotes the exterior normal vector to ∂D at each point with coordinates $x = (\underline{x})$.

2. Newton's law of universal gravitation can be encapsulated in the formula:

$$\nabla^2 \Phi(\underline{x}) = 4\pi G_N \mu(\underline{x}), \tag{14.1}$$

(Poisson's equation)

Proof. To verify (1), we imagine that the total mass distribution inside D can be thought of as a huge collection of small tiny balls inside D, each with mass m_δ. Hence, the total gravitational field U would be the sum of all the fields U_δ generated by all of these masses, with total mass m $= \int_D \mu(\underline{x})dx$. Example **??** shws that each m_δ accounts for a flux across $S = \partial D$ equal to $4\pi G_N \mathrm{m}_\delta$. Henceforth, all masses account for $\sum_\delta 4\pi G_N \mathrm{m}_\delta = 4\pi G_N \mathrm{m}$.

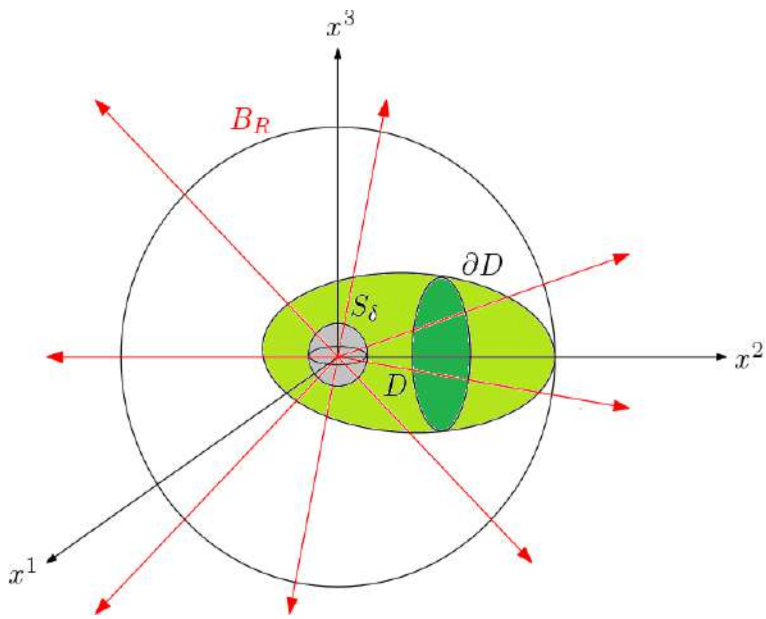


Figure 14.2: Gravitational flux

Assertion (2) is a consequence of the Divergence Theorem: The total flux over ∂D is $\int_{\partial D} U(\underline{x}) \cdot \mathrm{n}\ dA$. But this is equal to $\int_D \operatorname{div} U(\underline{x})dx$. Thus, $\int_D \operatorname{div} U(\underline{x})dx = 4\pi G_N \int_D \mu(\underline{x})d\underline{x}$, for any domain with smooth boundary D.

Then one must have $\operatorname{div} U(\underline{x}) = 4\pi G_N \mu(\underline{x})$. But

$$\operatorname{div} U(\underline{x}) = \operatorname{div}(\operatorname{grad} \Phi(\underline{x})) = \frac{\partial^2 U}{(\partial x^1)^2} + \frac{\partial^2 U}{(\partial x^2)^2} + \frac{\partial^2 U}{(\partial x^3)^2},$$

and from this one obtains $\nabla^2 \Phi(\underline{x}) = 4\pi G_N \mu(\underline{x})$. □

Example 14.2.2. In the interior of a homogeneous spherical shell the gravitational field cancels out and must be zero.

If fact, let us show that if p is any interior point in D the gravitational field F_p at p must be zero:Let's draw a sphere S_r in D containing p. If the

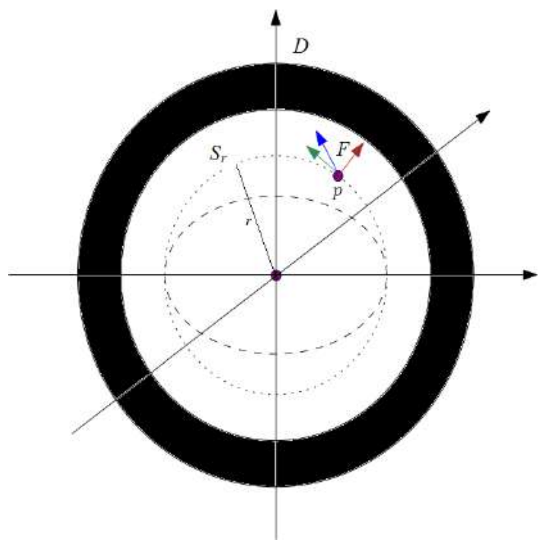


Figure 14.3: Gravitational field inside a shell

tangential component of F_p (in green) were not zero, by spherical symmetry, it would have to be nonzero at any other point over S_r. Thus, there would be a nonzero field on the sphere S_r, a well known topological impossibility ([25], page 120). Henceforth, the field at p must be normal, and, by symmetry, it has to be normal and of equal magnitude at any other point on S_r. This implies the total flux of F across S_r is $4\pi r^2 |F| \neq 0$. On the other hand, the divergence theorem of Gauss says this flux is equal to $\int_{\operatorname{int}(D)} \operatorname{div}(\operatorname{grad} \phi)\, dV$:

$$\int_{\operatorname{int}(D)} \nabla^2 \phi dV = 4\pi r^2 |F|, \tag{14.2}$$

where ϕ is the classical gravitational potential, and $\nabla^2\phi = \text{div}(\text{grad}\,\phi)$ denotes its Laplacian. But Poisson's equation (14.1) implies $\nabla^2\phi(x) = 4\pi G_N \mu(x)$, where μ is the mass density inside D, which we know is zero. Hence, $\nabla^2\phi$ must be zero and so would be the left hand side of (14.2), a contradiction.

Such a shell would behave like a gravitational *Faraday's Cage*. As we shall see in the next chapter, since we are assuming D is homogeneous and spherically symmetric, the geometry of the space-time it determines will be that of Schwarzschild (15.20). Inside D one should take $C = 0$, and outside one has $C = 2M$, as before. This says that in its interior the metric is Minkowski's, and outside the shell it is given by (15.21).

14.3 Tidal Forces

Tidal forces are characteristic of any gravitational field. In fact, an accelerated observer in flat space time, as in Section 10.10, would believe he is under a gravitational field apparently indistinguishable from that of Earth. However, the total absence of tidal forces would allow him to discover that he is not in the presence of a true gravitational field.

To see why this is true, suppose one releases two test particles that at time $t = 0$ are separated a short distance s_0, and that they fall freely towards the center of an object of mass M, let's say towards the center of the Earth. There we set a coordinate system with the x^3 axis pointing upwards, as shown below. Let $s(t)$ be the separation vector at time t between the test particles P_1 and P_2. Suppose particle P_1 is originally at position $(0, 0, d)$, where d is the distance between P_1 and the origin of coordinates: We assume that compared with the size of the massive body the original separation distance $s_0 = s(0)$ is very small (let's say 1 m compared with the radius of the Earth $\simeq 6.37 \times 10^6$ m). If $\alpha(t) = (a^i(t))$, and $\beta(t) = (b^i(t))$ are the trajectories of each particle, then according to Newton's Second Law, $d^2a^k/dt^2 = -\partial\Phi/\partial x^k(\alpha(t))$, and $d^2b^k/dt^2 = -\partial\Phi/\partial x^k(\beta(t))$. Let's write $s(t) = \beta(t) - \alpha(t)$ and define $h_k(\underline{x}) = (\partial\Phi/\partial x^k)(\underline{x})$. Approximating h_k by its differential one obtains $h_k(\underline{x} + \underline{s}) = h_k(\underline{x}) + \sum_i (\partial h_k/\partial x^i)(\underline{x}) s^i$. Thus,

$$\frac{\partial\Phi(\beta(t))}{\partial x^k} = \frac{\partial\Phi}{\partial x^k}(\alpha(t) + s(t)) = \frac{\partial\Phi(\alpha(t))}{\partial x^k} + \sum_i \frac{\partial^2\Phi(\alpha(t))}{\partial x^i \partial x^k} s^i(t).$$

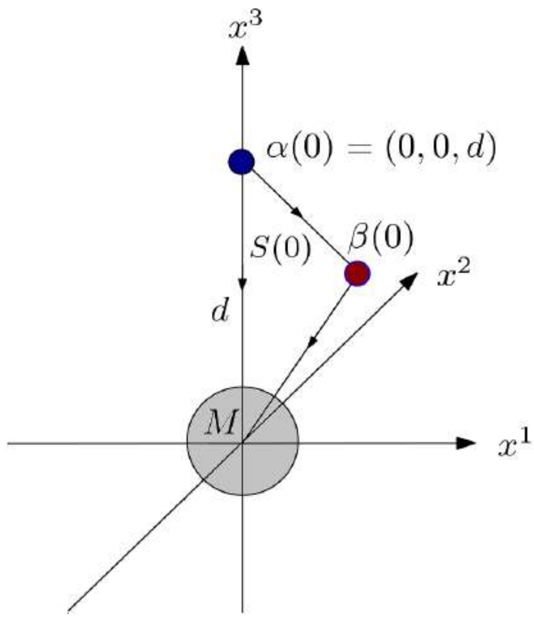


Figure 14.4: Tidal forces

From this one gets an approximate equation for the separation vector $s(t)$:

$$\begin{aligned} d^2s^k/dt^2 &= d^2b^k/dt^2 - d^2a^k/dt^2 = \\ &= \frac{-\partial\Phi(\alpha(t))}{\partial x^k} - \sum_i \frac{\partial^2\Phi(\alpha(t))}{\partial x^i \partial x^k} s^i(t) + \frac{\partial\Phi(\alpha(t))}{\partial x^k} \\ &= -\sum_i \frac{\partial^2\Phi(\alpha(t))}{\partial x^i \partial x^k} s^i(t). \end{aligned} \tag{14.3}$$

A computation shows

$\dfrac{\partial^2\Phi}{\partial x^i \partial x^k} = \dfrac{\mathrm{G}_N\mathrm{M}}{|\mathrm{r}(\underline{x})|^3}\left(\delta_{ik} - \dfrac{3x^i x^k}{|\mathrm{r}(\underline{x})|^2}\right)$, where δ_{ik} denotes Kronecker's function.

Thus, one gets from (14.3) the system of equations:

$$-\frac{\mathrm{G}_N\mathrm{M}}{r(t)^3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}\begin{bmatrix} s^1(t) \\ s^2(t) \\ s^3(t) \end{bmatrix} = \begin{bmatrix} d^2s^1/dt^2 \\ d^2s^2/dt^2 \\ d^2s^3/dt^2 \end{bmatrix}. \tag{14.4}$$

(We have used the fact that the test particle P_1 moves down the x^3 axis ($a^1(t) = a^2(t) = 0$, $a^3(t) = r(t) = d - 1/2gt^2$, for all t.) One then sees that the separation vector compresses in the horizontal direction due to attractive tidal forces, while it stretches in the vertical direction, due to tidal repulsion. The term "tidal" comes from the fact that these are the precisely the forces created by celestial bodies, like the Moon or the Sun responsible for the daily tides.

Let's analyze the acceleration of the separation vector $s(t) = \beta(t) - \alpha(t)$, at $t = 0$, of an element of ocean of mass m (this plays the role of particle t_2 in the example above) with respect to a particle t_1 located at the center of the Earth when they "fall towards the Moon". We set a coordinate system at the center of the Moon, as shown below. The coordinates of both particles at $t = 0$ are $\beta(0) = (R\sin\theta,\, 0,\, d - R\cos\theta)$, and $\alpha(0) = (0, 0, d)$, respectively, where d is the distance between the centers of the Earth and the Moon and R denotes the Earth's radius. Thus, $s(0) = (R\sin\theta, 0, -R\cos\theta)$, and the tidal forces at $t = 0$ are given by:

$$
\begin{aligned}
F_1 &= \frac{-\mathrm{G}_N m M_{\mathrm{Moon}}}{d^3} R\sin\theta \\
F_2 &= 0 \\
F_3 &= \frac{-2\mathrm{G}_N m M_{\mathrm{Moon}}}{d^3} R\cos\theta
\end{aligned}
$$

The force F_1 is responsible for compressing the oceans. On the other hand, when $\theta = 0$, the force $F_3 = -(2\mathrm{G}_N m M_{\mathrm{Moon}}/d^3)R$ is directed towards the Moon, while if $\theta = \pi$, $F_3 = (2\mathrm{G}_N m M_{\mathrm{Moon}}/d^3)R$ this same force is repulsive. Thus, F_3 is responsible for pulling the ocean away form the center of the Earth on both sides of the x^3-axis.

14.4 The Equivalence Principle

In 1920 Einstein commented that around 1907, while trying to make Newton´s theory fit within the framework of SR, all of a sudden, a thought came into his mind: The gravitational field has only a relative existence! For an observer freely falling there would be no gravitational field, at least in his immediate surroundings. A person in an elevator who accelerates at constant rate a, but does not know it, would believe he is in gravitational field (we know this is not globally true, since he wouldn't be able to detect any tidal forces). But identifying an accelerated observer with a gravitational field, the

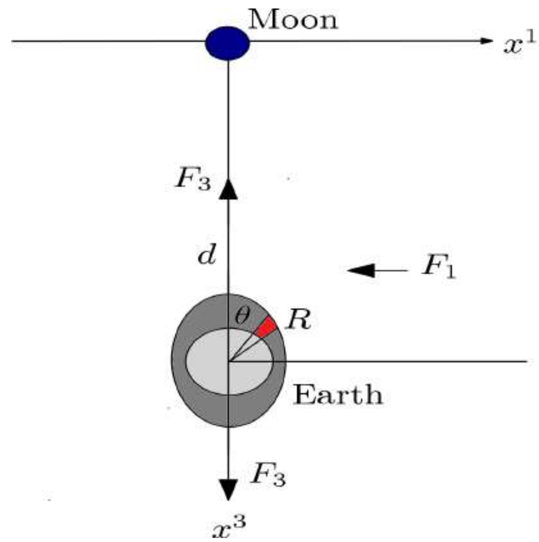


Figure 14.5: Tides

so called "equivalence principle", would indeed be true locally at a particular point.

The experiment Einstein imagines is more or less the following: An individual A is located at height h above another individual B in a uniform gravitational field where bodies fall with acceleration a. The equivalence principle would allow us to think as if A were at the nose of a rocket, h meters long, and B were at the tail, while the rocket accelerates at constant rate a in flat space-time. In Section 10.10 we deduce the equation of motion of such observers: If (t, x^i) are the standard coordinates of flat space-time, A's world-line (parametrized by proper time s) would be: $t_A(s) = \mathrm{a}^{-1}\sinh(\mathrm{a}s)$, and $x_A(s) = h + \mathrm{a}^{-1}\cosh(\mathrm{a}s)$, while B's would be $t_B(s) = \mathrm{a}^{-1}\sinh(\mathrm{a}s)$, and $x_B(s) = \mathrm{a}^{-1}\cosh(\mathrm{a}s)$.

Suppose A sends signals towards B at constant time intervals Δt_A (measured in his wristwatch). At what rate Δt_B would B receive them?We may assume these signals correspond to successive crests of an electromagnetic wave of light so that Δt_A would corresponds to its period (so that its frequency would be $\omega_A = 1/\Delta t_A$).

We know each light signal follows a straight line (a geodesic in flat space-time) with slope -1. At each point $(t_A(s),\ x_A(s))$ in the worldline of A, the equation $t = -x + x_A(s)) + t_A(s)$ gives the worldline of a photon that crosses A at that point. This line intersects the worldline of B at the point

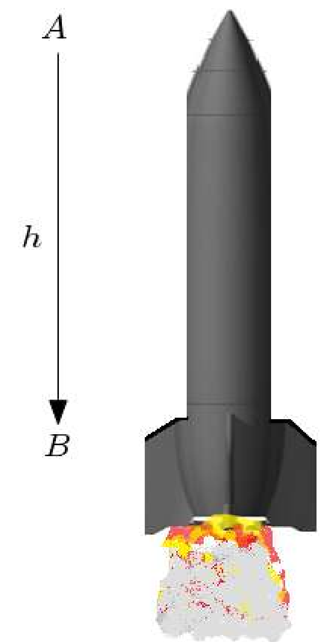


Figure 14.6: Equivalence Principle

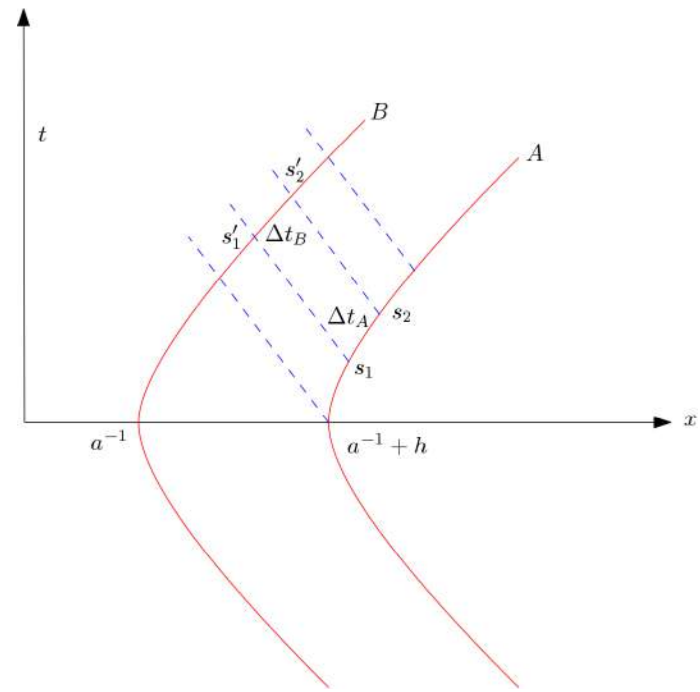
$t_B(s') = -x_B(s') + x_A(s)) + t_A(s)$.

That is, $\mathrm{a}^{-1}\sinh(\mathrm{a}s') + \mathrm{a}^{-1}\cosh(\mathrm{a}s') = h + \mathrm{a}^{-1}\cosh(\mathrm{a}s) + \mathrm{a}^{-1}\sinh(\mathrm{a}s)$. Thus $e^{\mathrm{a}s'} = \mathrm{a}h + e^{\mathrm{a}s}$. From this we obtain $\mathrm{a}s' = \ln(e^{\mathrm{a}s} + \mathrm{a}h)$. For very small h (let's say, of the order of a few meters) one can approximate $\ln(x+h) \simeq \ln x + h/x$. Hence, $\mathrm{a}s' = \mathrm{a}s + \mathrm{a}he^{-\mathrm{a}s}$ and one obtains: $s' = s + he^{-\mathrm{a}s}$. If two light signal are emitted by A at proper times $s_1$and s_2, they will be received by B at times $s'_1$and s'_2 (in his own watch), respectively. Henceforth, if $\Delta t_B = s'_2 - s'_1$, and $\Delta t_A = s_2 - s_1$ then $\Delta t_B = \Delta t_A + h(e^{-\mathrm{a}s_2} - e^{-\mathrm{a}s_1})$. Consequently:

$$\frac{\Delta t_B}{\Delta t_A} = 1 + h\left(\frac{e^{-\mathrm{a}s_2} - e^{-\mathrm{a}s_1}}{s_2 - s_1}\right) \simeq 1 + h\left.\frac{de^{-\mathrm{a}s}}{ds}\right|_{s=s_1} = 1 - \mathrm{a}he^{-\mathrm{a}s_1}.$$

The approximation is valid since $\Delta t_A = s_2 - s_1$ is very small (of the order of 10^{-6} ss, the period of a visible electromagnetic wave). By taking $s_1 = 0$ one obtains: $\Delta t_B = \Delta t_A(1 - \mathrm{a}h)$.

If, for instance, A sends 10^6 pulses in a short second, the observer B will receive these pulses in $1 - \mathrm{a}h$ ss.This says that the frequency of the wave, as



measure by A, which is $\omega_A = 10^6$ cycles/ss, will be, as measured by B equal to $\omega_B = 10^6/(1-\mathrm{a}h) = \omega_A/(1-\mathrm{a}h)$ cycles/ss. For instance, by taking $\mathrm{a} = \mathrm{g}$, the acceleration of gravity, we get: $\omega_A = \omega_B(1-\mathrm{g}h)$. Or in standard units: $\omega_A = \omega_B(1-\mathrm{g}h/\mathrm{c}^2)$, with $\mathrm{g} = 9.8$ m/s^2.

The Equivalence Principle would allow us to reverse the situation again, and think of A and B as the two original observers in a homogeneous gravitational field, where A stands at the top of a building of height h, while B remains on the surface of the corresponding body responsible for the gravitational field.

The difference of gravitational potential between A and B would be then equal to $\Phi(x_A) - \Phi(x_B)$, with $\Phi(x) = -\mathrm{G}_N\mathrm{M}/x$, where M is the mass of the body, $x_B = R$ denotes its radius, and $x_A = R+h$. But,

$$\Phi(x_A) - \Phi(x_B) = \frac{-\mathrm{G}_N\ \mathrm{M}}{R+h} + \frac{\mathrm{G}_N\ \mathrm{M}}{R} =$$
$$\mathrm{G}_N\mathrm{M}(\frac{1}{R} - \frac{1}{R+h}) = \frac{h\ \mathrm{G}_N\ \mathrm{M}}{R(R+h)} \simeq h\frac{\mathrm{G}_N\ \mathrm{M}}{R^2}.$$

A particular calculation taking $\mathrm{M} = 5.97 \times 10^{24}$ kg, the mass of the Earth, $\mathrm{G}_N = 6.674 \times 10^{-11}$ m^{3}kg^{-1}s^{-2}, and $R = 6.37 \times 10^6$, the radius of the Earth, gives $G_N M_E/R_E^2 = 9.8$ m/s^2, which is precisely the acceleration of gravity on the surface of the Earth.

Then one has $\Phi(x_A) - \Phi(x_B) \simeq h\mathrm{g}$, and therefore:

$$\omega_A = \omega_B(1 - \frac{h\mathrm{g}}{\mathrm{c}^2}) = \omega_B \left(1 + \frac{\Phi(x_B) - \Phi(x_A)}{\mathrm{c}^2}\right).$$

This equation suggests that *in the presence of a gravitational field the frequency of a light signal will be red shifted* by a factor $1 + (\Phi(x_B) - \Phi(x_A))/\mathrm{c}^2$ "as it climbs the gravitational potential". In general, one would expect

$$\omega_A = \omega_B \left(1 + \frac{\Phi(x_B) - \Phi(x_A)}{\mathrm{c}^2}\right) \tag{14.5}$$

for the redshift caused by a not too strong gravitational potential Φ.

On the other hand, we notice that the individual B at the bottom of the building would see his own watch running slower than that of A's: If we think of the pulses A is sending to B as the ticking of his watch, after n ticks in A's watch, B would registers $n(1 - \mathrm{g}h/\mathrm{c}^2) < n$ ticks of his own watch. If n ticks amount to, let's say, 80 years, then $(1 - \mathrm{g}h/\mathrm{c}^2)n$ ticks will amount to (taking $h = 30$ m) $\simeq 80(1 - 10^{-14})$ years. Observer A ages more quickly, but the age difference would only be less that 10^{-4} seconds in a life time!

14.5 Estimate of g_{00}

In this section we drop our standard assumption (c = 1) regarding units, and work in standard units.

Fix standard coordinates (t, x^i) for flat space-time. Imagine we suddenly introduce a body of mass M and radius R that we locate at the origin of the coordinate system. Suppose this body perturbs the Minkowski metric η creating a new *static* metric $g = \eta + \varepsilon$. By this we mean that the entries of ε would be functions of the spatial coordinates alone. Moreover, we would also expect $\varepsilon_{ij}(x^i) \to 0$ as the absolute value of the spatial coordinates $|x^i|$ approach ∞. This is a reasonable assumption, since the field gets closer and closer to zero far away from M. Summarizing, our basic assumptions are:

1. In standard coordinates (t, x^i) for $\mathbb{R}^4$ the metric takes the form $g = \eta + \varepsilon$, where $\varepsilon(x^i)$ is *static* (does not depend on t).

2. The entries $\varepsilon_{ij}(x^i)$ are very small far from M, i.e., $\varepsilon_{ij}(x^i) \to 0$ as the absolute value of each spatial coordinate $|x^i|$ approaches ∞. In particular, $g_{00}(x^i) \longrightarrow -1$, as $|x^i| \longrightarrow \infty$.

Under these assumptions we now want to estimate the metric coefficient g_{00}. We start by considering the case of a *steady* observer O at a fixed distance x in the direction of x^1 from the surface of the Body (where we denote its mass by M and its radius by R) responsible for the gravitational field. His worldline can be given the parametrization $O(s) = (s, x, 0, 0)$ in the standard coordinates of $\mathbb{R}^4$ (notice that O is not moving in a geodesic. It corresponds to a particle that stays at a *fixed* location outside M. In this case $\langle O'(s), O'(s)\rangle = g_{00}(O(s))$, and therefore

$$\langle O'(s), O'(s)\rangle = g_{00}(x, 0, 0), \tag{14.6}$$

(recall we are assuming g does not depend on t) so that

$$|O'(0)| = \sqrt{-g_{00}(x, 0, 0)},$$

and its 4-velocity at $s = 0$ would be $\mathbf{u} = (1/\sqrt{-g_{00}(x, 0, 0)})\partial_t$.

As in Section 14.4, we want to compare the frequencies ω_B and ω_A of a pulse of light, as measured by two steady observers B and A at fixed distances $x_B < x_A$, respectively. Let $\gamma(\tau)$ be a null curve corresponding to the worldline of the light signal emitted by A at $\tau_0 = 0$, and suppose it is received by B at $\tau = s_1$. As we saw in Section 10.8.2, the energy measured by both A and B is $E_A = -\langle\gamma'(0)), \mathbf{u}_A\rangle$, and $E_B = -\langle\gamma'(s_1), \mathbf{u}_B\rangle$, respectively. Since the energy of a circular wave is $\hbar$ times its frequency (as measured by each observer) we see then that

$$\begin{aligned}
\omega_B &= -\hbar^{-1}\langle\gamma'(\tau_1), \mathbf{u}_B\rangle = -\hbar^{-1}/\sqrt{-g_{00}(x_B, 0, 0)}\,\langle\gamma'(\tau_1), \partial_t\rangle \\
\omega_A &= -\hbar^{-1}\langle\gamma'(\tau_0), \mathbf{u}_A\rangle = -\hbar^{-1}/\sqrt{-g_{00}(x_A, 0, 0)}\,\langle\gamma'(\tau_0), \partial_t\rangle\,.
\end{aligned}$$

But since we are assuming g does not depend on t, one must have $\partial g_{ab}/\partial_t = 0$, and consequently $\langle\gamma'(s), \partial_t\rangle$ is constant (Proposition **??**). Thus, $\langle\gamma'(s_1), \partial_t\rangle = \langle\gamma'(0), \partial_t\rangle$, and henceforth:

$$\omega_A = \omega_B \frac{\sqrt{-g_{00}(x_B, 0, 0)}}{\sqrt{-g_{00}(x_A, 0, 0)}}. \tag{14.7}$$

Recall we are assuming $g_{00}(x^i) \to -1$, as the absolute value of each spatial coordinate $|x^i|$ increases. Thus, $g_{00}(x_A, 0, 0) \to -1$, as $x_A \to \infty$. If ω_∞ denotes $\lim_{x_A \to \infty} \omega_A$ we see that:

$$\omega_\infty = \omega_B \sqrt{-g_{00}(x_B, 0, 0)}. \tag{14.8}$$

On the other hand, for a not too strong potential we had calculated that $\omega_A = \omega_B(1 + (\Phi(x_B) - \Phi(x_A))/\mathrm{c}^2)$ (14.5). Again, by letting $x_A \to \infty$, and since the gravitational potential $\Phi(x_A)$ approaches zero, in the limit one obtains:

$$\omega_\infty = \omega_B \left(1 + \frac{\Phi(x_B)}{\mathrm{c}^2}\right). \tag{14.9}$$

Comparing (14.8) with (14.9) one gets the estimate:

$$g_{00}(x_B, 0, 0) = -\left(1 + \Phi(x_B)/\mathrm{c}^2\right)^2.$$

For a celestial body like Earth, or even the Sun, the term $\mathrm{G}_N\ \mathrm{M}/(\mathrm{c}^2 R)$ is very small (in the case of the Earth it is of order 10^{-10}). If $x_B > R$ the term $\Phi(x_B)/\mathrm{c}^2 = -\mathrm{G}_N\ \mathrm{M}/(\mathrm{c}^2 x_B)$ is even smaller. So one may approximate $(1 + \Phi(x_B)/\mathrm{c}^2)^2$ as $1 + 2\Phi(x_B)/\mathrm{c}^2$. Thus, $g_{00}(x, 0, 0) \simeq -(1 + 2\Phi(x, 0, 0)/\mathrm{c}^2)$, which gives the following estimate for g_{00} :

$$g_{00}(x, 0, 0) = -\left(1 - \frac{2\mathrm{G}_N\ \mathrm{M}}{\mathrm{c}^2 x}\right). \tag{14.10}$$

14.6 Geodesics in a Weak Gravitational Field

Now we look at the geodesics of free falling objects in a week gravitational field, where, as before, we assume the field determines a metric $g = \eta + \varepsilon$ very close to the metric of Minkowski. Consider, for instance, the geodesic described by a ball that moves through a parabola during two seconds before falling, as shown below:The ball follows a geodesic $\sigma_0(s)$ in space-time whose t-axis extends $2c = 6 \times 10^8$ meters that is very nearly a straight line. In fact, let $\sigma(s) = (t(s), c^i(s))$ be its parametric equation, where s denotes proper time.

We claim that if the 3-velocity of the corresponding object is small compared with the speed of light, then $t(s) \simeq s$, and its 3-velocity is approximately $\mathrm{v} = \sum_i dc^i/ds\ \partial_{x^i}$.

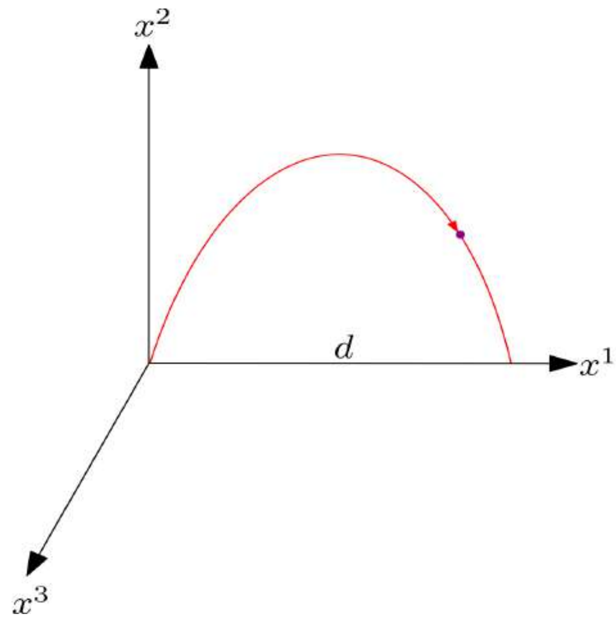


Figure 14.7: Parabolic Trajectory

By computing $\sigma'(s)$ we see that:

$$\begin{aligned}\sigma'(s) &= (dt/ds)\partial_t + \textstyle\sum_i (dc^i/ds)\partial_{x^i} \\ &= (dt/ds)\partial_t + (dt/ds)\textstyle\sum_i (dc^i/dt)\partial_{x^i} \\ &= (dt/ds)(\partial_t + \mathrm{v}),\end{aligned}$$

and consequently

$$-1 = \langle \sigma'(s), \sigma'(s)\rangle = (dt/ds)^2 g_{00}(\sigma(s)) + 2(dt/ds)\,\langle \partial_t, \mathrm{v}\rangle + \langle \mathrm{v}, \mathrm{v}\rangle^2 .$$

Now, the term $\langle \partial_t, \mathrm{v}\rangle$ is very small, since its is equal to $\sum_i v^i \varepsilon_{i0}$ and $|\sum_i v^i \varepsilon_{i0}| < |\mathrm{v}| \left|\sqrt{\sum_i \varepsilon_{i0}}\right|$ (this is the ordinary Schwartz's inequality). Since we are assuming $|\mathrm{v}| \simeq 0$, one obtains

$$-1 = \langle \sigma'(s), \sigma'(s)\rangle \simeq (dt/ds)^2 g_{00}(\sigma(s)).$$

Thus, $t(s) \simeq s$, and from this we immediately see that the 3-velocity v would approximately be equal to $\mathrm{v} = \sum_i dc^i/ds\ \partial_{x^i}$.

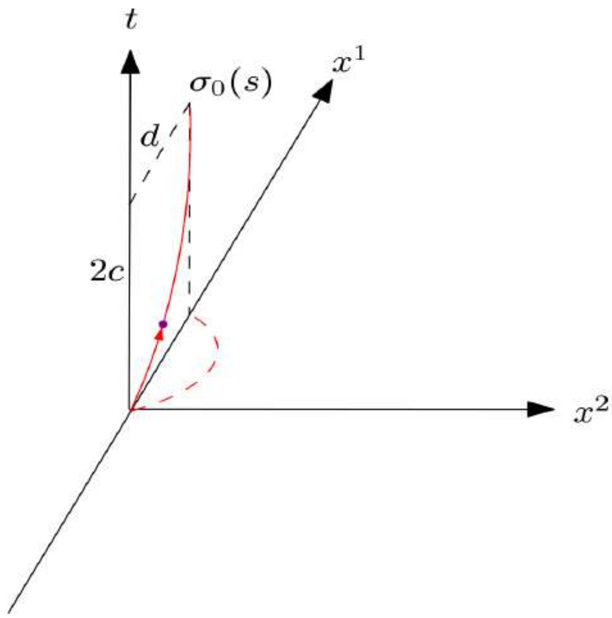


Figure 14.8: Geodesic in a weak field

14.7 The Astonishing Analogy

In this section we work again in units (ss) where c = 1.

In Section 6.4 we deduced the equation $A(s) = R(\mathbf{u}(s), Y(s), \mathbf{u}(s)) = \nabla_{\mathbf{u}(s)}\nabla_{\mathbf{u}(s)}Y(s)$ for the geodesic deviation, where $A(s)$ was the acceleration of the separation vector $Y(s)$ between $\sigma_0(s)$ and an "infinitesimally closed" timelike geodesic $\sigma_1(s)$ ($\mathbf{u}(s)$ denotes the unitary tangent vector at the point $p = \sigma_0(s)$).This equation strongly resembles Equation 14.3. In fact, let (t, x^i) be the standard coordinates of $\mathbb{R}^4$, and, as before, we assume $\sigma_0(s) = (s, a^i(s))$, and $\sigma_1(s) = (s, b^i(s))$ are arc length parametrizations of the geodesics of two falling test particles that move towards the center of the body of mass M (let's say the Earth). If the particles are very closed together, the separation vector

$$D(s) = \sigma_1(s) - \sigma_0(s) = (0, b^i(s) - a^i(s))$$

can be approximated as the vector $Y(s)$. Computing in the standard coordinates we observe that at p

$$A(s) = \textstyle\sum_d R^d{}_{cab}(\sigma_0(s))\ \mathbf{u}^a(s)Y^b(s)\mathbf{u}^c(s). \tag{14.11}$$

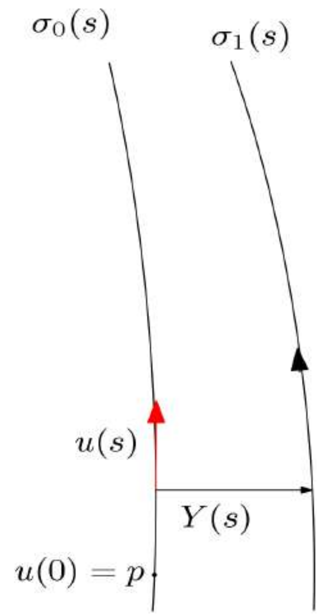


Figure 14.9: Geodesic deviation

But since $\mathbf{u}(s) = \partial t + \mathrm{v}$, and we are assuming $|\mathrm{v}| \simeq 0$, then in (14.11) all terms are negligible, except when $a = c = 0$. Since $Y^0(s) \simeq D(s) = 0$, Equation 14.11 becomes

$$A(s) = \sum_k R^k{}_{00i}(\sigma_0(s))\ Y^i(s) = -\sum_k R^k{}_{0i0}(\sigma_0(s))\ Y^i(s).$$

Thus, Equation 14.11 becomes (to a good approximation)

$$A(s) = -\sum_i R^k{}_{0i0}(\sigma_0(t))\ D^i(s). \tag{14.12}$$

But Equation 14.3 says that:

$$\frac{d^2 D^k}{ds^2} = -\sum_i \frac{\partial^2 \Phi}{\partial x^i \partial x^k}(\sigma_0(s)) D^i(s). \tag{14.13}$$

Since $A(s) \simeq d^2 D^i/ds^2$ one sees that $R^k{}_{0i0} \simeq \partial^2\Phi/\partial x^i \partial x^k$ at p. If we think of this equation as a tensor equation, we could then "contract" indices on both sides to obtain: $\mathsf{Ric}_{00} = \sum_k R^k_{0k0} = \sum_i \partial^2\Phi/\partial x^i \partial x^i = \nabla^2\Phi$. Consequently:

$$\mathsf{Ric}_{00} = \nabla^2\Phi = 4\pi \mathrm{G}_N \mu. \tag{14.14}$$

In empty space-time (devoid of masses), the latter equation just becomes $\mathsf{Ric}_{00} = 0$. As we shall see in the next section, if this is true in any Lorentz system of coordinates, at any point p, then it is true that $\mathsf{Ric} = 0$. This was the first of the field equations discovered by Einstein. With that in hand, he was able to explain the "anomalous" precession of the perihelion of Mercury, and predicted the bending of a ray of light when it passes near a celestial body. After trial and error he was able to discover the general field equation. His reasoning could be summarized as follows.

It is reasonable to expect that the tensor of energy-momentum T should be the mathematical object replacing μ. So one would expect that the analogue of Newton's law would be given by an equation of the form $Ric_{ab} = kT_{ab}$ for some suitable constant k. But we know that $\sum_a (\nabla_{\partial_a} \overline{T})^{ab} = 0$ must be zero, but is not the case that $\sum_a (\nabla_{\partial_a} \overline{\mathsf{Ric}})^{ab}$ is equal to zero. However, we know that for $G^{ab} = Ric^{ab} - \frac{1}{2}\mathrm{R}g^{ab}$ one has $\sum_a (\nabla_a \overline{G})^{ab} = 0$ (Proposition 6.2.7). So, as Einstein himself suggested, it would be reasonable to seek for an equation of the form $G^{ab} = kT^{ab}$, or equivalently, of the form $G_{ab} = kT_{ab}$, i.e..,

$$G_{ab} = \mathsf{Ric}_{ab} - \frac{1}{2}\mathrm{R}\ g_{ab} = kT_{ab}. \tag{14.15}$$

Let's see how to determine k by passing to the "Newtonian limit".

First, we notice that form this equation one gets:

$$\sum_{r,s} g^{rs}\mathsf{Ric}_{rs} - \frac{1}{2}\sum_{r,s}\mathrm{R}\ g^{rs}g_{rs} = k\sum_{r,s} g^{rs}T_{rs},$$

and henceforth

$$\mathrm{R} - \frac{1}{2}4\mathrm{R} = k\sum_{r,s} g^{rs}T_{rs} = k\mathrm{T},$$

(we have used the fact that $\sum_{r,s} g^{rs}g_{rs} =$ Trace of the 4×4 identity matrix). Thus, $\mathrm{R} = -k\mathrm{T}$, with $\mathrm{T} = \sum_{r,s} g^{rs}T_{rs}$ and Equation (14.15) can be written as

$$\mathsf{Ric}_{ab} = kT_{ab} + \frac{1}{2}\mathrm{R}\ g_{ab} = k(T_{ab} - \frac{1}{2}g_{ab}\mathrm{T}). \tag{14.16}$$

Our hypothesis on M implies that the components of the tensor of energy momentum tensor in the standard coordinates (t, x^i), $T^{ab} = \mu v^a v^b$ reduce to $T^{00} = \mu$ and to $T^{ab} = 0$, for $(a, b) \neq (0, 0)$, since $v^0 \simeq 1$, and $v^i \simeq 0$. Also, under our hypotheses we already know that $g_{00} \simeq -1$, as we proved in the previous section. Since we are assuming the entries of ε are very small, neglecting terms of quadratic order one sees that $(\eta + \varepsilon)(\eta - \varepsilon) = \eta^2 = Id_{4\times 4}$

(the 4×4 identity matrix) so that for our metric $g = \eta + \varepsilon$ we have $g^{ab} = \eta_{ab} - \varepsilon_{ab}$. In particular, $g_{00}g^{00} + \sum_i g_{0i}g^{i0} = 1$. Hence, $g_{00}g^{00} - \sum_i \varepsilon_{0i}\varepsilon_{i0} = 1$, and since $\sum_i \varepsilon_{0i}\varepsilon_{i0} \simeq 0$, we get $g_{00}g^{00} \simeq 1$.

But $T_{00} = g_{00}g_{00}T^{00} = \mu$. Henceforth, $\mathrm{T} = g^{00}T_{00} \simeq -\mu$.

Taking $a = b = 0$ in Equation (14.16) one obtains:

$$\mathsf{Ric}_{00} = k(T_{00} - \frac{1}{2}g_{00}\mathrm{T}) = k(\mu - \frac{1}{2}(-1)(-\mu)) = \frac{1}{2}k\mu,$$

But Equation (14.14) implies that $4\pi\mu = 1/2k\mu$, from which we get $k = 8\pi\mathrm{G}_N$. Summarizing, Einstein's field equations must be: $G = 8\pi\mathrm{G}_N T$. In nonstandard units of time (ss), the constant G_N has units of $\mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$. Hence, in standard units G_N must be divided by c^2. Similarly, $T_{ab} = \rho v^a v^b$ (13.1) hence in standard units one must also divide by c^2, and henceforth one can write Einstein´s equation in the usual form:

$$G = \frac{8\pi\mathrm{G}_N}{\mathrm{c}^4}T \qquad \text{(Einstein's Field Equations)}$$

14.8 Einstein's Equations in Plain English

Einstein's field equation is, in fact, a compact way to write a set of 10×10 nonlinear partial differential equations in the metric coefficients, once a particular system of coordinates is fixed. This is because G and T are symmetric tensors, so that if we regard them as given by symmetric 4×4 matrices, each tensor would then entail 4 entries in the main diagonal plus 6 additional entries repeated two times symmetrically with respect to this diagonal. In spite of its mathematical complexity, Einstein's equations have a very simple geometric interpretation ([2]). So simple in fact, that they could be explained to the layman in plain English. To understand this, we first notice that the validity of Einstein's equations $Ric_{ab} = 8\pi\mathrm{G}_N(T_{ab} - \frac{1}{2}g_{ab}\mathrm{T})$ in a particular system of coordinates $x = (x^a)$ for space-time (M, g) is equivalent to the validity of this equation for $a = b = 0$ in *all Lorentz frames at each point of* M. To state this precisely, we first make the following elementary remark:

Remark 14.8.1. Let L be a symmetric bilinear form in $\mathbb{R}^4$. If $L(v, v) = 0$ for all $v \in U$, a non empty open set of $\mathbb{R}^4$, then L must be the identically zero form.

Proof. Since L is symmetric there is a basis B for $\mathbb{R}^n$ where L can be represented by a diagonal matrix diag$[\lambda_a]$, with $\lambda_a \in \mathbb{R}$. Hence, $L(v, v) = 0$ for all $v \in U$ implies that $\lambda_1 u_1^2 + \cdots + \lambda_n u_n^2 = 0$, for $v = (u_1, \ldots, u_n) \in U$, where we have written the vector $v = [u_i]$ in the basis B. for a fixed $v_0 = (b_0, \ldots, b_n) \in U$, we know that there is $\varepsilon > 0$ such that $\lambda_1 b_1^2 + \cdots + \lambda_n (b_n + y)^2 = 0$, for all y with $|y| < \varepsilon$. But this forces $\lambda_n = 0$. By induction one gets in the same way that $\lambda_{n-1} = 0, \ldots, \lambda_1 = 0$. □

Proposition 14.8.2. Let (M, g) be space-time. Then $G = 8\pi \mathrm{G}_N T$ holds in an open region $W \subset M$ if and only if for each point $p \in W$ one has $G_{00} = 8\pi \mathrm{G}_N T_{00}$ in all Lorentz coordinate frames $x = (x^a)$ at p.

Proof. The implication $\Longrightarrow$ is obvious. For the reciprocal, we know that for a fixed $p \in W$, all timelike vectors v in $T_p(M)$ form a non empty open cone U. If we let L to be the bilinear form $G - 8\pi \mathrm{G}_N T$ in $T_p(M)$, then given any vector $v \in U$ there is always a Lorentz frame whose "time coordinate" is v (Proposition 11.2.7). By hypothesis we know that $L_{00} = 0$ in this frame, which is equivalent to saying that $L(v, v) = 0$. Thus L vanishes for all pairs (v, v), with $v \in U$, and by the previous remark this implies that $L = 0$ at p, and consequently Einstein's equation $L = G - 8\pi \mathrm{G}_N T = 0$ holds in W. □

We already know that Einstein's equation can be written as

$$\mathsf{Ric} = 8\pi \mathrm{G}_N (T - \frac{1}{2} g \mathrm{T}),$$

and by the previous proposition this amounts to saying that

$$\mathsf{Ric}_{00} = 8\pi \mathrm{G}_N (T_{00} - \frac{1}{2} g_{00} \mathrm{T})$$

in all Lorentz coordinate frames $x = (x^a)$, at each point p. But in a Lorentz frame $g_{00} = -1$, and $\mathrm{T} = \sum_{r,s} g^{rs} T_{rs} = -T_{00} + T_{11} + T_{22} + T_{33}$, and consequently Einstein's equation becomes

$$\begin{aligned} \mathsf{Ric}_{00} &= 8\pi \mathrm{G}_N \left(T_{00} + \frac{1}{2}(-T_{00} + T_{11} + T_{22} + T_{33}) \right) \\ &= 4\pi \mathrm{G}_N (T_{00} + T_{11} + T_{22} + T_{33}). \end{aligned} \tag{14.17}$$

The right hand side of this equation has a clear physical meaning: The first term is the energy density at the point p while T_{ii} represents pressure on each spatial direction (13.1.3).

Imagine now a test particle P_0 at p that is initially surrounded by three nearby test particles P_k. We denote by $\sigma_u(s)$ the (geodesic) world-line of each particle P_u, $u = 0, 1, 2, 3$.

Now, let us fix a Lorentz frame $x = (x^a)$ for P_0 at $p = \sigma_0(0)$, and let's denote by $\sigma_u^a(s) = x^a(\sigma_u(s))$ the components of each σ_u in this coordinate system. Let $Y_{P_u}(s) = \sigma_u(s) - \sigma_0(s)$ be the separation vector between P_u and P_0 in space-time. We assume each P_k, $k = 1, 2, 3$, is initially *very close to* P_0, and located on the x^k-axis at a distance ε_k. That is, we assume that we can make the approximation $Y_{P_k}(0) = \varepsilon_k \partial_k$, where ε_k denotes a small positive number. Let's denote by $\mathbf{u}(s)$ the four velocity of P_0. Since we are assuming $x = (x^a)$ is a Lorentz frame for P_0 we have that $\mathbf{u}(\mathbf{0}) = \partial x^0$. Since $\sigma_0(s)$ is a geodesic, one has $\nabla_{\mathbf{u}(s)}\mathbf{u}(s) = 0$, and then the curvature tensor $R(\mathbf{u}(s), \mathbf{u}(s), \mathbf{u}(s))$ is zero. Therefore,

$$R(\partial_{x^0}, \partial_{x^0}, \partial_{x^0}) = \sum_d R^d{}_{000}\, \partial_d = 0,$$

and in particular $R^0{}_{000} = 0$.

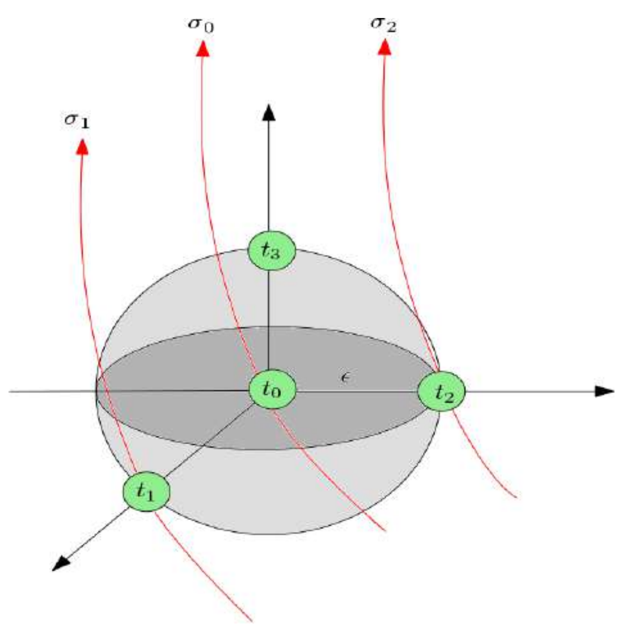


Figure 14.10: Einstein's equations

Additionally, we will suppose that the particles P_u are initially "at rest". Strictly speaking, except for P_0, this is a meaningless assumption, because the frame $x = (x^a)$ has physical meaning only at the point p, and not necessarily at any other point. However, for particles very close to p it is reasonable to define "being initially at rest" by the condition $d\sigma_u^j/d\sigma_u^0\big|_{s=0} = 0$, $j = 1, 2, 3$, i.e., each spatial component of the 4-velocity is null. Since $d\sigma_u^0/ds\big|_{s=0} \neq 0$ (each P_u moves in a timelike geodesic) being initially at rest implies that $d\sigma_k^j/ds\big|_{s=0} = 0$. In particular, if $r_u(s) = Y^k_{\mathrm{P}_u}(s)$ denotes the k-component of $Y^k_{\mathrm{P}_u}$, then $r'_k(0) = 0$.

As we notice in Equation 14.11, in a Lorentz frame at p for $\sigma_0(s)$ each vector $Y_{\mathrm{P}_k}(s) = \sum_b Y^b_{\mathrm{P}_k}(s)\partial_b$ must satisfy the equation:

$$\frac{d^2 Y^c_{\mathrm{P}_k}(s)}{ds^2} = -\sum_b R^c{}_{0b0}(\sigma_0(s))Y^b_{\mathrm{P}_k}(s) = -\sum_j R^c{}_{0j0}(\sigma_{\mathrm{P}_0}(s))Y^j_{\mathrm{P}_k}(s),$$

(this last equality, since $Y^0_{\mathrm{P}_k}(s) = 0$). For $c = k$ one obtains $r''_k(0) = -R^k{}_{0k0}(\sigma_0(s))\varepsilon_k$. Thus,

$$-\sum_k \frac{r''_k(0)}{r_k(0)} = \sum_k R^k{}_{0k0}(p) = \sum_u R^u{}_{0u0}(p) = \mathsf{Ric}_{00}(p),$$

where we have used the fact that $R^0{}_{000} = 0$.

Denote by $V(s) = 4/3\pi r_1(s)r_2(s)r_3(s)$ the volume of the ellipsoid with axes $r_k(s)$. If we use the fact that $r'_k(0) = 0$, a simple computation shows that $V''(0)/V(0) = \sum_k r''_k(0)/r_k(0)$. From this one obtains $V''(0)/V(0) = -Ric_{00}(p)$. By Einstein's equation (14.17) we then must have

$$\frac{V''(0)}{V(0)} = -4\pi \mathrm{G}_N(T_{00}(p) + T_{11}(p) + T_{22}(p) + T_{33}(p)).$$

Thus, Einstein's equation would say in plain English that *the rate at which the relative acceleration rate of the volume of a ball of test particles is equal to* $-4\pi\mathrm{G}_N$ *times the sum of the energy density at the center of the ball plus the sum of the energy pressure in each spatial direction.*

14.8.1 Geometric Units

It is convenient to choose units not only for time but also for mass, so that in these units the constant of gravitation $\mathrm{G}_N = 6.674 \times 10^{-11}\ \mathrm{m}^3\ \mathrm{kg}^{-1}\ \mathrm{s}^{-2}$ is equal to one. These are often called *geometrized units or geometric units.*

We define a *big kilogram*, kgg as $c^2/(G_N)$ kg. Since $c = 3 \times 10^8$ m$\times$s^{-1}, then:

$$\begin{aligned} 1 \text{ kgg} &= \frac{9 \times 10^{16}}{6.674} \times 10^{11} \simeq 1.348 \times 10^{27} \text{ kg.} \\ 1 \text{ kg} &= \frac{6.674}{9} \times 10^{-27} \simeq 7.41 \times 10^{-26} \text{ kgg.} \end{aligned}$$

Obviously, in units of m, kgg, ss (meter, big kilogram, short second) the constant of gravitation becomes equal to one. The mass of the Earth (5.972×10^{24} kg) in geometric units becomes 0.0044 kgg. The mass of our Sun (1.9×10^{30} kg) would be approximately 1408.9 kgg.

Our terminology is by no means standard. We have converted units of mass into units of length, so the natural name for "big kilograms" would be "meters". In geometric units, mass and time are measured in meters, or in centimeters, as it customary in cosmology. For instance, the mass of the Earth would be 0.4 cm. Terms like "big kilogram" or "short second" used in these notes are capricious. The reader may disregard them if he or she prefers.

From now on we will use geometric units, unless we specify otherwise.

14.9 EINSTEIN'S EQUATIONS FROM A VARIATIONAL VIEWPOINT XXXXX

Part IV

Solutions to Einstein's Equation

Chapter 15

The Schwarzschild Metric

The first exact solution of Einstein's field equations were discovered by Karl Schwarzschild, in 1915, only a few months after Einstein introduced his general theory of relativity. Surprisingly, Schwarzschild discovered his celebrated solution while serving in the German army during World War I. He died the following year from a rare autoimmune disease, at the early age of forty two.

In this chapter we want to deduce Schwarzschild's solutions for a static and spherically symmetric space-time manifolds. We start by considering the vacuum case, that is when the manifold corresponds to a region of space-time devoid of masses. More precisely, we will consider an open set in space-time where the *energy-momentum tensor vanishes.* There, Einstein's equation (Equation 14.16)

$$\mathsf{Ric}_{ab} = 8\pi \mathrm{G}_N (T_{ab} - \frac{1}{2} g_{ab} \mathrm{T})$$

becomes $\mathsf{Ric}_{ab} = 0$. We want to solve this equation under the assumption that the corresponding space-time manifold (M, g) is *static and spherically symmetric*, concepts we define below. These hypothesis are satisfied when the source of curvature comes from a *spherically symmetric object*, like a star. In these case M can be taken as the manifold $U \times \mathbb{R}$, where U is the region in $\mathbb{R}^3$ outside the celestial body.

For such a manifold, we want to show that we can choose appropriated coordinates (t, r, θ, ϕ) for (M, g) so that g can be written as:

$$g = -\left(1 - \frac{\mathrm{C}}{r}\right) dt \otimes dt + \left(1 - \frac{\mathrm{C}}{r}\right)^{-1} dr \otimes dr + r^2 dA, \tag{15.1}$$

where C is an arbitrary constant and $dA = d\theta \otimes d\theta + \sin^2\theta \; d\phi \otimes d\phi$ is the standard metric of a *round sphere*.

15.1 Static Space-Time Manifolds

Definition 15.1.1. The 4-manifold (M, g) is *static* if it satisfies:

ST1 There is a timelike Killing vector field V defined in M. The existence of such vector field is equivalent to the existence of a parametric complete group of isometries $\lambda_t : M \to M$ (by complete we mean $t \in \mathbb{R}$) such that for each $p \in M$, if $O_p(t) = \lambda_t(p)$ is the orbit of p, with $O_p(0) = p$, then $O_p'(0)$ is a timelike vector (**??**).

ST2 There is a 3-dimensional hypersurface H diffeomorphic to an open set W of $\mathbb{R}^3$ of the form $W = \mathbb{R}^3$ or $W = \mathbb{R}^3 - \{(0)\}$, such that for each p in H the tangent space $T_p(M)$ is spacelike (each nonzero vector is spacelike), and such that V_p is *orthogonal* to $T_p(H)$.

ST3 Each point $q \in M$ lies in the orbit of a unique point p in H, i.e., $q = O_p(t)$, for a unique t. We notice that if H_t denotes the surface $\lambda_t(H)$ then M must be equal to the disjoint union of the 3-dimensional slices H_t, which are all isometric. Hence, by giving q local coordinates (t, x), where $x = (x^i)$ are local coordinates for $H = H_0$ around p, and t is the only value for which $O_p(t) = q$, one may decompose M as a product space $\mathbb{R} \times H$.

Remark 15.1.2.

1 Let (M, g) be static. By (15.1) M decomposes as $\mathbb{R} \times H$, hence, if $x = (\underline{x})$ is any set of local coordinates for H, by choosing coordinates (t, x^i) we may write g as:

$$g = a(t,x)dt \otimes dt + \textstyle\sum_i b_i(t,x)dt \otimes dx^i + \sum_{j,k} c_{j,k}(t,x)dx^j \otimes dx^k.$$

Let q be any point in W, and let p and t be the unique values for which $q = O_p(t)$. Since λ_t are isometries, the product

$$\left\langle \partial_t|_q , \partial_t|_q \right\rangle = \left\langle \partial_t|_p , \partial_t|_p \right\rangle \tag{15.2}$$

must be a negative constant in each orbit $O_p(t)$. Thus, the function $a(t,x)$ must be *independent of the coordinate* t. Using again the fact that λ_t are isometries we have

$$\left\langle \partial_{x^i}\big|_q , \partial_t\big|_q \right\rangle = \left\langle \partial_{x^i}\big|_p , \partial_t\big|_p \right\rangle = 0,$$

where the last equality holds due to (15.1). This says that all *the coefficients* $b_i(t, \underline{x})$ *must vanish*. For the same reason as in (15.2), the product $\left\langle \partial_{x^i}\big|_q , \partial_{x^j}\big|_q \right\rangle$ is also constant along any orbit $O_p(t)$, and henceforth *the functions* c_{jk} *neither depend on* t. Hence, for a static space-time the metric g can be written as:

$$g = -a(\underline{x})dt \otimes dt + \sum c_{j,k}(\underline{x})dx^j \otimes dx^k.$$

2 In a static space-time (M, g), the Gaussian coordinates provide a frame of reference with physical meaning for each observer $O_p(s) = (s/\mathrm{a}, p)$, where $\mathrm{a} = a(p)^{1/2}$. Proper time provides a natural clock, while $x = (x^i)$ give spatial coordinates for each 3-dimensional slice. These slices with induced metric $h = \sum c_{j,k}(\underline{x})dx^j \otimes dx^k$ look all the same to O_p, since they are all isometric. Hence, as his clock moves forward, *the collection of all spatial slices are regarded by* O_p *as a single unit* W: The set of all points $q' = (x^i(q))$ in 3-dimensional space corresponding to all possible events $(t(q), x^i(q))$ in M.

15.1.1 Fermat's Principle

We want to see that O_p sees any beam of light moving in W following a geodesic path. Classically, this is known as *Fermat's Principle*. By this we mean the following: if $\gamma(s) = (t(s), x^i(s))$ is a null geodesic, with $t'(s) > 0$ (hence, $s = s(t)$ is a function of t) then the path followed by the ray is the curve $\alpha(t) = (x^i(s(t)))$ in (W, h). We are claiming *that* $\alpha(t)$ *is a geodesic in* (W, h).

In fact, we first notice that the Christoffel symbols $\Gamma^0_{jk}(q)$ must all vanish, since (**??**) implies that:

$$\begin{aligned}\Gamma^0_{jk} &= \frac{1}{2}\sum_c g^{c0}\left(\frac{\partial g_{kc}}{\partial x^j} + \frac{\partial g_{cj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^c}\right). \\ &= \frac{1}{2} g^{00}\left(\frac{\partial g_{k0}}{\partial x^j} + \frac{\partial g_{0j}}{\partial x^k} - \frac{\partial c_{jk}}{\partial t}\right) = 0,\end{aligned}$$

since $g_{k0} = g_{0j} = 0$, and $\partial c_{jk}(x)/\partial t = 0$.

Now, The Levi-Civita connection ∇ in M induces a natural connection in W compatible with h, with Christoffel symbols $\widetilde{\Gamma}^i_{jk}(q')$ given by $\widetilde{\nabla}_{\partial_j}\partial_k = \sum_i \widetilde{\Gamma}^i_{jk}(q')\partial_i$. But, by definition of $\widetilde{\nabla}$

$$\widetilde{\nabla}_{\partial_j}\partial_k = \nabla_{\partial_j}\partial_k = \textstyle\sum_c \Gamma^c_{jk}(q)\partial_c = \sum_i \widetilde{\Gamma}^i_{jk}(q')\partial_i,$$

since $\Gamma^0_{jk}(q) = 0$. Thus, $\widetilde{\Gamma}^i_{jk}(q') = \Gamma^i_{jk}(q)$.

We know $\langle \gamma'(s), \gamma'(s)\rangle = 0$, which implies that

$$\left(\frac{dt}{ds}\right)^2 = \frac{1}{a(x)}\left\langle \textstyle\sum_i \frac{dx^i(s)}{ds}\partial_{x^i}, \sum_j \frac{dx^j(s)}{ds}\partial_{x^j}\right\rangle,$$

and in particular we must have: $\frac{d}{dt}(dt/ds) = 0$. Thus,

$$\begin{aligned}\frac{d^2x^i}{dt^2} &= \frac{d}{dt}\left(dx^i/dt\right) = \frac{d}{dt}\left(\frac{dx^i/ds}{dt/ds}\right)\\ &= \frac{dt/ds\ \frac{d}{dt}\left(dx^i/ds\right) - dx^i/ds\ \frac{d}{dt}(dt/ds)}{(dt/ds)^2}\\ &= \frac{1}{(dt/ds)^2}\frac{d^2x^i}{ds^2}.\end{aligned}$$

On the other hand, the geodesic equations for $\gamma(s)$ implies, in particular, that:

$$\frac{d^2x^i}{ds^2} + \textstyle\sum_{a,b}\Gamma^i_{ab}(\gamma(s))\frac{dx^a}{ds}\frac{dx^b}{ds} \tag{15.3}$$

$$= \frac{d^2x^i}{ds^2} + \textstyle\sum_{j,k}\Gamma^i_{jk}(\gamma(s))\frac{dx^j}{ds}\frac{dx^k}{ds} = 0, \tag{15.4}$$

(since $\Gamma^i_{0k}(\gamma(s)) = 0$). Since $dt/ds(\gamma(s)) > 0$, equation (15.3) is equivalent to

$$\begin{aligned}&\frac{1}{(dt/ds)^2}\left(\frac{d^2x^i}{ds^2} + \textstyle\sum_{j,k}\Gamma^i_{jk}(\gamma(s))\frac{dx^j}{ds}\frac{dx^k}{ds}\right)\\ &= \frac{d^2x^i}{dt^2} + \textstyle\sum_{j,k}\Gamma^i_{jk}(\gamma(s))\frac{dx^j}{dt}\frac{dx^k}{dt}\\ &= \frac{d^2x^i}{dt^2} + \textstyle\sum_{j,k}\Gamma^i_{jk}(\gamma(s))\frac{dx^j}{dt}\frac{dx^k}{dt} = 0,\end{aligned} \tag{15.5}$$

(using that $dx^j/ds = dx^j/dt\ dt/ds$). But $\Gamma^i_{jk}(\gamma(s)) = \widetilde{\Gamma}^i_{jk}(\alpha(s))$, and henceforth (15.5) is just the geodesic equation for $\alpha(t)$ in W.

15.2 Spherical symmetry

Definition 15.2.1. The 4-manifold (M, g) is *spherically symmetric* if it satisfies the following three conditions:

SS1 There is a subgroup Rot of the group of isometries of M isomorphic to the group of rotations SO(3) which preserves the Killing field V. This last condition means that for $p \in M$, and $\rho \in \mathrm{Rot}$, one has $V_{\rho(p)} = d\rho(p)V_p$.

SS2 If $q \in M$, and if S_q denotes the orbit of q (the set of all points of the form $\rho(q)$, with $\rho \in \mathrm{Rot}$) then S_q is topologically a 2-dimensional sphere.

Conditions SS1 and SS2 imply that:

i If $S = S_q$ is the orbit of q, then at any point $c \in S$ one must have that V_c is orthogonal to $T_c(S)$, the tangent space to the sphere at c.

ii Each 2-sphere S_q must be entirely contained in some spatial slice H_{t_0}.

Hence, the action of Rot restricts to H. We also want this action to be *equivariant* with the standard action of SO(3) on euclidean space. More precisely, if we denote the action of $\mathrm{Rot} \simeq \mathrm{SO}(3)$ by a *dot*, $\cdot$, one has:

SS3 There is a diffeomorphism $f : H \to W \subseteq \mathbb{R}^3$ (not necessarily an isometry), where W is equal to $\mathbb{R}^3$ or to $\mathbb{R}^3$ minus the origin, such that $\rho \cdot f(x) = f(\rho \cdot x)$, for $x \in H$ and every rotation ρ.

We now prove the above remark.

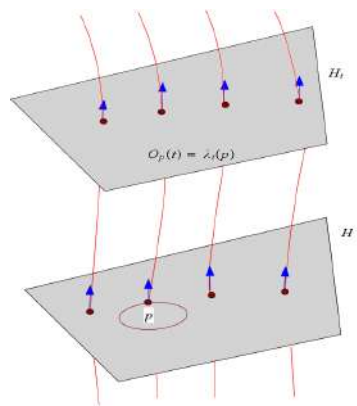
Proof. i) We first notice that there must be at least one point in S where the field V is orthogonal to the tangent space of S at that point. For if not, the projection of V onto the tangent space at each point of S would provide a continuos (smooth) nonzero vector field on a 2-sphere, a topological impossibility ([25], page 120). Hence, let $c \in S_q$ be a point where V_c is orthogonal to $T_c(S_q) = 0$. For another point $c' = \rho(c)$ on the same orbit as c, (i.e., on S) condition (15.2.1) would imply that

$$\begin{aligned}\langle V_{c'}, T_{c'}(S)\rangle &= \langle d\rho(c)V_c, d\rho(c)T_c(S)\rangle \\ &= \langle V_c, T_c(S)\rangle = 0,\end{aligned}$$

since each vector in $T_{\rho(c)}(S)$ is the image under $d\rho(c)$ a vector in $T_c(S)$. Thus, V is orthogonal to the tangent space at every point on S. ii) We know that $q \in H_{t_0}$, for some t_0. Let (t, x) be local coordinates in a open neighborhood U_q for which $\partial_t|_q = V_q$, and so that $x = (x^i)$ are coordinates for H_{t_0}. Let us show that any other point q' in S_q that lies in U_q must also be in H_{t_0}: Let $\beta : [0, 1] \to M$ be any curve in S_q from q to q' with $\beta(0) = q$ and $\beta(1) = q'$. In these local coordinates we may write $\beta(s) = (b^0(s), b^i(s))$, with $b^0(0) = t_0$. We claim db^0/ds must be zero for all values s of the parameter, and therefore $b^0(s) = t_0$, for all the values of the parameter s. In fact, by (15.1) V is *orthogonal* to each H_t, and consequently

$$\left\langle \beta'(s), V_{\beta(s)} \right\rangle = \frac{db^0}{ds}(s) \ \left\langle V_{\beta(s)}, V_{\beta(s)} \right\rangle = 0,$$

which forces $db^0/ds(s) = 0$. Henceforth, $b^0(1) = t_0$ which means that $q' \in H_{t_0}$. Since any two points on S can be connected by a curve contained in S we see that $S_q \subset H_{t_0}$. □



Proposition 15.2.2. Suppose (H, g) is a Riemannian 3-manifold diffeomorphic to either $\mathbb{R}^3$ or to $\mathbb{R}^3 - \{(0)\}$ whose group of isometries contains a subgroup Rot satisfying (15.2.1) above. Then one can choose coordinates (r, θ, ϕ), $0 < r$, $0 < \theta < \pi$, $0 < \phi < 2\pi$ that put the metric in the form:

$$\mathrm{g} = l(r)dr \otimes dr + \lambda(r)^2 dA,$$

where $l(r) > 0$, $\lambda(r)$ are smooth functions defined for $r > 0$, and where $dA = d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi$ is the standard form of the metric of a round unitary sphere. These coordinates are defined, except, as usual for the coordinate singularities $\theta = 0, \pi$. In case $H \simeq \mathbb{R}^3$, we must also exclude the singular point $r = 0$.

Proof. Via the diffeomorphism f one can transport the metric g to W so that H becomes trivially isometric to W. Hence, it suffices to prove the proposition for the 3-manifold (W, g), and where Rot = SO(3) is the usual group of rotations in $\mathbb{R}^3$. Let (r, θ, ϕ) be the standard spherical coordinates defined on W after removing the origin. We know the orbits of SO(3) in W are the standard spheres in euclidean space (though the metric, of course, is not necessarily the same). Fix S_1 the unitary sphere in W (its radius is one in the standard euclidean metric). Condition SS2 tells us that S_1 is a topological sphere. But the curvature at any two points p, p' on S_1 must be the same, since a rotation taking p into p' is *a priori* an isometry. A well know theorem of Riemannian Geometry implies that S_1 must be isometric to the standard sphere ([5], Page 317). Hence, after rotating axis if necessary one can choose spherical coordinates (r, θ, ϕ) for $\mathbb{R}^3$ such that the restriction of g to S looks like:

$$\mathrm{g} = a_1^2 \ (d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi),$$

for a suitable nonzero constant a_1. Now we want to *geodesically propagate* the coordinates (θ, ϕ) defined on S_1 to any other of the concentric spheres S_r. For this, fix a point $p \in S_1$, and let n_p be the unitary normal vector at p (i.e., unitary and normal in the geometry induced by g). Let $\gamma(s)$ be the unique geodesic that starts at p, with $\gamma'(0) = n_p$. We know $\gamma(s)$ must stay invariant under any isometry, in particular, it stays invariant under the rotation $\rho_p$whose axis is the (euclidean) line L_p that passes through p. Since L_p are precisely the points that remain fixed under this rotation, we see that $\gamma(s)$ must be entirely contained in L_p. This shows that geodesics in (W, g) are straight lines, though not of the same length as in euclidean geometry. Clearly, the line L_p pierces S_r at a unique point $q \in S_r$. Define $\lambda(r) = \sqrt{4\pi / r^2}$. We claim that in these coordinates, g restricted to S_r can be written as

$$\mathrm{g} = \lambda(r)^2 (d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi).$$

Define $R > 0$ as the ratio

$$R = \left\langle \partial_\theta|_q \,, \partial_\theta|_q \right\rangle / \left\langle \partial_\theta|_p \,, \partial_\theta|_p \right\rangle . \tag{15.6}$$

Let ρ_{pq} be the rotation about the line trough p and q, that takes $\partial_\theta|_p$ into $\partial_\phi|_p$, and $\partial_\theta|_q$ into $\partial_\phi|_q$. Since ρ is an isometry one has:

$$\begin{aligned}\left\langle \partial_\phi|_q, \partial_\phi|_q \right\rangle &= \left\langle d\rho(q)\, \partial_\theta|_q,\ d\rho(q)\, \partial_\theta|_q \right\rangle = \\ \left\langle \partial_\theta|_q, \partial_\theta|_q \right\rangle &= R\left\langle \partial_\theta|_p, \partial_\theta|_p \right\rangle = R\left\langle d\rho^{-1}(p)\, \partial_\phi|_p, d\rho^{-1}(p)\, \partial_\phi|_p \right\rangle \\ &= R\left\langle \partial_\phi|_p, \partial_\phi|_p \right\rangle.\end{aligned} \tag{15.7}$$

Similarly, by taking the rotation that takes $\partial_\theta|_p$ into $\partial_\theta|_p + \partial_\phi|_p$, and $\partial_\theta|_q$ into $\partial_\theta|_q + \partial_\phi|_q$, respectively, one see in the same fashion that

$$R = \left\langle \partial_\theta|_q + \partial_\phi|_q, \partial_\theta|_q + \partial_\phi|_q \right\rangle / \left\langle \partial_\theta|_p + \partial_\phi|_p, \partial_\theta|_p + \partial_\phi|_p \right\rangle. \tag{15.8}$$

Henceforth, (15.6), (15.7), and (15.8) imply that $[\mathrm{g}(q)]$, the matrix representing g at q in the coordinates (θ, ϕ), is equal to R times the matrix representing the metric at p: $[\mathrm{g}(q)] = R[\mathrm{g}(p)]$. On the other hand, if p' is any other point in S_1, there is a rotation taking p to p'. Again, since this is an isometry $[\mathrm{g}(p)] = [\mathrm{g}(p')]$, and consequently we must have $[\mathrm{g}(q)] = R[\mathrm{g}(p)]$, for any pair of points $p \in S_1$ and $q \in S_r$ lying in the line pq. This shows the metric g restricted to S_r must have the form

$$\mathrm{g} = a_r^2(d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi), \tag{15.9}$$

for a suitable nonzero constant a_r. But this constant must then be equal to $\lambda(r)$, since the area of S_r (integrating the two form determined by 15.9) must be $4\pi b^2$ which we know is equal to $4\pi\lambda(r)^2$. Finally, for θ_0 and ϕ_0 fixed, if we let $l(r)$ = length of the geodesic determined by the line from the origin to the point with coordinates (r, θ_0, ϕ_0), we clearly have that $l(r) > 0$ and g is written in W as:

$$\mathrm{g} = l(r) dr \otimes dr + \lambda(r)^2(d\theta \otimes d\theta + \sin^2(\theta) d\phi \otimes d\phi).$$

□

We want to show that if the space-time manifold (M, g) is *spherically symmetric* and *static,* the metric g can be written as (15.1) in suitable coordinates.

15.3 Birkhoff's Theorem

Theorem 15.3.1 (Birkhoff)**.** Let (M, g) be a *spherically symmetric* and *static* space-time manifold (M, g). Suppose $\mathsf{Ric} = 0$. Then, there are coordinates (r, θ, ϕ) for M, (r, θ, ϕ), where $0 < r$, $0 < \theta < \pi$, $0 < \phi < 2\pi$, called *Schwarzschild coordinates,* for which the metric can be written as:

$$g = -\left(1 - \frac{\mathrm{R}_S}{r}\right) dt \otimes dt + \left(1 - \frac{\mathrm{R}_S}{r}\right)^{-1} dr \otimes dr + r^2 dA, \tag{15.10}$$

for a suitable constant R_S.

Proof. Since (M, g) is static, in appropriated coordinates $(t, \underline{x})$ the metric g can be written as (**??**). By Proposition 15.2.2 coordinates $(\underline{x})$ in H can be chosen as (r, θ, ϕ) so that the second term in this sum can be replaced by $l(r) dr \otimes dr + \lambda^2(r) dA$, for $r > 0$. Finally, let us see that in these coordinates, the function $a(r, \theta, \phi)$ does not depend on θ, ϕ. But this is the case, since by definition of the time coordinate one has $V_q = \partial t|_q$, and the group of rotations Rot preserves the Killing field, which then implies that for a fixed value of r:

$$\begin{aligned} a(r, \theta, \phi) &= \left\langle \partial t|_{\rho(p)}, \partial t|_{\rho(p)} \right\rangle \\ &= \left\langle d\rho(p)\, \partial t|_p,\ d\rho(p)\, \partial t|_p \right\rangle = \left\langle \partial t|_p, \partial t|_p \right\rangle, \end{aligned}$$

for each $p \in H$. But varying θ or ϕ amounts to apply a suitable ρ in Rot, hence $a(r)$ is a smooth function of r, with $a(r) < 0$. Summarizing, the metric g, in the coordinates (t, r, θ, ϕ) assumes the form:

$$g = a(r) dt \otimes dt + l(r) dr \otimes dr + \lambda^2(r) dA,$$

where $a(r) < 0$, and $l(r) > 0$. By writing $a(r) = -e^{2\alpha(r)}$, $l(r) = e^{2\beta(r)}$, and $\lambda^2(r) = r^2 h(r)$, for suitable functions $\alpha(r), \beta(r)$, and $h(r) > 0$, we may write:

$$g = - - e^{2\alpha(r)} dt \otimes dt + e^{2\beta(r)} dr \otimes dr + h(r) r^2 dA. \tag{15.11}$$

Moreover, after scaling r one may assume h is the constant function 1. In fact, let $\rho(r) = r e^{\gamma(r)}$, with $\gamma(r)$ an increasing function. Then $d\rho = (1 + r\gamma'(r)) e^{\gamma(r)} dr$, and therefore g can be written as:

$$g = -e^{2\alpha(r)} dt \otimes dt + ((1 + r\gamma'(r))^{-2} e^{2\beta(r) - 2\gamma(r)}\ d\rho \otimes d\rho + \rho^2 dA.$$

One can express r as a function of ρ, since $\rho(r)$ is an increasing function. Hence, one can choose $\widetilde{\beta}(\rho)$ so that

$$e^{2\widetilde{\beta}(\rho)} = (1 + r\gamma'(r))^{-2} e^{2\beta(r)-2\gamma(r)}.$$

In terms of the coordinates $(x^0, x^1, x^2, x^3) = (t, \rho, \theta, \phi)$ the metric looks like

$$g = -e^{2\widetilde{\alpha}(\rho)} dt \otimes dt + e^{2\widetilde{\beta}(\rho)} d\rho \otimes d\rho + \rho^2 dA, \tag{15.12}$$

for a suitable exponent $\widetilde{\alpha}(\rho)$. We may refresh the notation and write α, β instead of $\widetilde{\alpha}$, $\widetilde{\beta}$, and r instead of ρ, as it is customary. A computation of the Christoffel symbols using the program *Maple* (see box 1.1 below) gives:

$$\begin{aligned}
&\Gamma^0_{0,1} = \alpha'(r),\\
&\Gamma^1_{0,0} = \alpha'(r)e^{2\alpha(r)-2\beta(r)},\\
&\Gamma^1_{1,1} = \beta'(r), && (15.13)\\
&\Gamma^1_{2,2} = -re^{-2\beta(r)}, && (15.14)\\
&\Gamma^1_{3,3} = -r\sin^2(\theta)e^{-2\beta(r)}, && (15.15)\\
&\Gamma^2_{1,2} = \frac{1}{r},\ \Gamma^2_{3,3} = -\sin\theta\cos\theta,\\
&\Gamma^3_{1,3} = \frac{1}{r},\quad \Gamma^3_{2,3} = \frac{\cos\theta}{\sin\theta}, && (15.16)\\
&\Gamma^c_{a,b} = 0 \text{ in any other case.}
\end{aligned}$$

Now, the components of the Ricci tensor in terms of the Christoffel symbols are given by (**??**):

$$\mathsf{Ric}_{ab} = \textstyle\sum_u \frac{\partial \Gamma^u_{a,b}}{\partial x^u} - \frac{\partial \Gamma^u_{a,u}}{\partial x^b} + \sum_{u,r} \Gamma^u_{a,b}\Gamma^r_{u,r} - \Gamma^u_{a,r}\Gamma^r_{b,u}.$$

For this particular metric $\mathsf{Ric}_{ab} = 0$, except in the following cases (see boxes 1.2-1.3 below):

$$\begin{aligned}
\mathsf{Ric}_{00} &= e^{2(\alpha(r)-\beta(r))}[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r)] && (15.17)\\
\mathsf{Ric}_{11} &= -\alpha''(r) - \alpha'(r)^2 + \alpha'(r)\beta'(r) + \frac{2}{r}\beta'(r)\\
\mathsf{Ric}_{22} &= e^{-2\beta(r)}[r(\beta'(r) - \alpha'(r)) - 1] + 1\\
\mathsf{Ric}_{33} &= \sin^2(\theta)\mathsf{Ric}_{22}.
\end{aligned}$$

The tensor Ric vanishes identically. In particular, $\mathsf{Ric}_{00} = \mathsf{Ric}_{11} = 0$. This implies

$$e^{-2(\alpha(r)-\beta(r))}\mathsf{Ric}_{00} + \mathsf{Ric}_{11} = \frac{2}{r}(\alpha'(r) + \beta'(r)) = 0, \tag{15.18}$$

from which one gets $\beta(r) = -\alpha(r) + c$, for a constant c, that without lose of generality one can choose equal to zero after a scaling of coordinate time by a factor of $e^{-c}t$. On the other hand, $Ric_{22} = 0$ implies that

$$e^{-2\beta(r)}[r(\beta'(r) - \alpha'(r)) - 1] + 1 = 0. \tag{15.19}$$

By substituting $\beta(r) = -\alpha(r)$ in (15.19) one gets: $e^{2\alpha(r)}(2r\alpha'(r) + 1) = 1$. That, in turn, may be written as $(re^{2\alpha(r)})' = 1$. Henceforth, $re^{2\alpha(r)} = r - \mathrm{R}_S$, or equivalently, $e^{2\alpha(r)} = 1 - \mathrm{R}_S/r$ for a suitable constant R_S. This constant is known as the *Schwarzschild radius*. Summarizing, $f(r) = -(1 - \mathrm{R}_S/r)$, and from this we readily obtain:

$$l(r) = e^{2\beta(r)} = e^{-2\alpha(r)} = \left(1 - \frac{\mathrm{R}_S}{r}\right)^{-1},$$

Thus, the metric should looks like:

$$g = -\left(1 - \frac{\mathrm{R}_S}{r}\right) dt \otimes dt + \left(1 - \frac{\mathrm{R}_S}{r}\right)^{-1} dr \otimes dr + r^2 dA, \tag{15.20}$$

where $dA = d\theta \otimes d\theta + \sin^2\theta \; d\phi \otimes d\phi$, and where $0 < r$, $0 < \theta < \pi$, $0 < \phi < 2\pi$. □

15.4 The Exterior of a Non Rotating Star

The exterior of a non rotating spherical star provides an example of a *spherically symmetric* and *static* space-time. More precisely, if we assume the body of the star is a perfect ball of radius $a > 0$, B_a, the 4-manifold $M = \mathbb{R} \times (\mathbb{R}^3 - B_a)$ can be modeled as a spherically symmetric, static, *empty* space-time. However, under rotation, the star ceases to be a spherical body (it bulges at the equator) and therefore the vacuum space surrounding the material body is no longer a spherically symmetric 3-manifold (although the space-time around it can still be assumed to be static).

The Ricci flat pseudo-riemannian 4-manifold (M, g) must then have a metric that in Schwarzschild's coordinates has the form (15.10). In order to determine the value of the constant R_S, we use the fact that a star must create a gravitational field around it. For *a weak gravitational field,* we had already obtained the estimate (14.10) for the $(0,0)$ coefficient of the metric: $g_{00}(r) = -(1 - 2\mathrm{M}/r)$. Hence, for $r >> 0$, the field created by the star behaves like a weak field, and one must have $\mathrm{R}_S/r = 2\mathrm{M}/r$, from which one obtains $\mathrm{R}_S = 2\mathrm{M}$ (or $\mathrm{R}_S = 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2$ in standard unit). The constant $\mathrm{R} = 2\mathrm{M}$ is called the *Schwarzschild radius of the star.* Therefore, the metric g, known as the *Schwarzschild metric,* can be written in spherical coordinates as:

$$g = -\left(1 - \frac{2\mathrm{M}}{r}\right) dt \otimes dt + \left(1 - \frac{2\mathrm{M}}{r}\right)^{-1} dr \otimes dr + r^2 dA. \qquad (15.21)$$

Or, in standard units, as:

$$g = -\left(1 - \frac{2\mathrm{G}_N\mathrm{M}}{\mathrm{c}^2 r}\right) \mathrm{c}dt \otimes \mathrm{c}dt + \left(1 - \frac{2\mathrm{G}_N\mathrm{M}}{\mathrm{c}^2 r}\right)^{-1} dr \otimes dr + r^2 dA.$$

For a star to be *stable* it is necessary that $\mathrm{R}_S < a$, otherwise, as we shall see, it is destined to disappear under *gravitational collapse.* In its place, only a "singularity" remains, and around it, a *black hole* is created, a bizarre empty region of space-time that we will analyze in the next chapter.

We must point out that the metric (15.21) is *only valid for the exterior region outside the body of the star.* Inside, it is different, since it is no longer true that the energy-momentum tensor vanishes. Hence, although (15.21) is undefined for $r = 2\mathrm{M}$, this metric is perfectly well defined in M, since $\mathrm{R}_S < a$. As we shall see in the next chapter, one can isometrically embed M as an open set of $M' = \mathbb{R} \times (\mathbb{R}^3 - \{(0)\})$, where (15.21) is defined for all $r > 0$.

Remark 15.4.1.

1. From (15.21) we see that if $r >> 0$ then g is approximately the Minkowski metric η, written in spherical coordinates. For a far away steady observer: For an observer whose worldline is $O(s) = (s, r_0, \phi_0, \theta_0)$, with $r_0 >> 2\mathrm{M}$, one has $O'(s) = \partial_t$, and $|O'(s)| = (1 - 2\mathrm{M}/r_0)^{1/2} \simeq 1$. Hence, t *would approximately represent its proper time, while* r, ϕ, θ *are the usual spherical coordinates for each of its spatial slices. An infinite far away observer is usually called the observer at infinity.*

2. For any r greater that $R_S = 2M$, O's proper time between two points in $O(s)$ with coordinates t_0, t_1 would be given by $(1 - 2Mr_0)^{1/2}\,(t_1 - t_0)$. As we noticed in Remark 15.1.2 (2), $t^* = (1 - 2Mr_0)^{-1/2}\,t$, together with $x = (r, \theta, \phi)$, provide coordinates with physical meaning for O. The collection of all spatial slices t =constant, being all isometric, conform his 3-dimensional universe, while t^* provides his clock.

3. For values of $r < 2M$, the curve $O(s)$ is no longer timelike, but spacelike, so it does not correspond to the worldline on any observer. "Inside these region", the coordinate r (but not t) should be interpreted as a "time coordinate", so the singularity $r = 0$ would actually be" a moment in time".

15.5 Redshift in Schwarzschild Space-Time

In Section 14.5 we compared the frequencies ω_B and ω_A of a pulse of light as measured by two steady observers B and A at fixed distances $x_B < x_A$ from the surface of an object of mass M. We obtained the relationship

$$\omega_A = \omega_B \frac{\sqrt{-g_{00}(x_B, 0, 0)}}{\sqrt{-g_{00}(x_A, 0, 0)}}. \tag{15.22}$$

For a not too massive object, we estimated that as x_A moves further and further away, and if ω_∞ denotes the limit $\lim_{x_A \to \infty} \omega_A$, then:

$$\omega_\infty = \omega_B\,(1 + \Phi(x_B)) = \omega_B(1 - \frac{M}{x}), \tag{15.23}$$

(in geometric units). This estimate is accurate for weak gravitational fields. To obtain the general relationship between ω_∞ and ω_B one uses the Schwarzschild metric: Again, we let $O_A(s) = (s, r_A, \theta_A, \phi_A)$ and $O_B(s) = (s, r_B, \theta_B, \phi_B)$ be two steady observers at fixed places in euclidean space hovering above a spherically symmetric body of mass M , where we assume $r_A > r_B > R_S = 2M$. For the Schwarzschild metric we know that $g_{00} = -(1 - 2M/r)$. Hence, by letting $r_A \to \infty$ one obtains:

$$\omega_\infty = \omega_B \sqrt{-g_{00}(r_B)} = \omega_B(1 - \frac{2M}{r_B})^{1/2} \tag{15.24}$$

In standard units, the term inside the radical is equal to $1 - 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2 r$. Hence, we see again that if M is not too big (so that $2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2 r$ is small) one can use the approximation $(1-\delta)^{1/2} = 1 - \frac{1}{2}\delta + \cdots$ from which one obtains back formula (15.23).

Formula (15.24) shows that light gets red shifted (looses frequency) as "it climbs the gravitational potential" when it moves away from M. The constant R_S is known as the *Schwarzschild radius*. We notice that the closer the observer B gets to the *horizon*, the set of spatial points at radius R_S, the more red shifted any light signal coming from B would be, as perceived by an observer very distant from M; and its frequency would be zero if it comes from a point right on the horizon!

When M represents a non-rotating star (this last conditions is required to ensure spherical symmetry), the value of R_S is smaller than the radius of the star, so that *the geometry of space-time for point with $r < \mathrm{R}_S$ is not given by (15.21), since these points would lie in a region where the energy momentum tensor is not zero!*

15.6 Spherical Shells

In this section we want to analyze the geometry of space-time inside of a non rotating star, or more generally, inside a spherical shell

$$\mathrm{S} = \{(r, \theta, \phi) : 0 \leq r_0 < r \leq r_1,\ 0 < \theta < \pi,\ 0 < \phi < 2\pi\}.$$

For this, we suppose S consists of a perfect fluid of density and pressure given by smooth functions $\rho(r)$ and $\mathrm{P}(r)$. We assume this fluid moves in space-time in the direction of the *unitary* vector field $V = 1/\left|\partial_t\right|\, \partial_t$. Let M denote the 4-manifold $M = \mathbb{R} \times \mathrm{S}$, with coordinates $x = (t, r, \theta, \phi)$. Since we are assuming M is static and spherically symmetric, we already know the metric g can be written as (15.12)

$$g = -e^{2\alpha(r)} dt \otimes dt + e^{2\beta(r)} dr \otimes dr + r^2 dA.$$

Hence, the components of V in the coordinates x are given by $v^0 = 1/\left|\partial_t\right| = e^{-\alpha(r)}$, and $v^i = 0$. Hence, associated to $\rho(r)$ and $\mathrm{P}(r)$ there is an energy-momentum tensor $\overline{T}$ whose components in these same coordinates are given by (Proposition 13.2.1):

$$T^{ab} = (\rho(r) + \mathrm{P}(r)) v^a v^b + g^{ab} \mathrm{P}(r),$$

and therefore

$$T^{00} = (\rho(r) + \mathrm{P}(r))e^{-2\alpha(r)} + (-e^{-2\alpha(r)})\mathrm{P}(r) = \rho(r)e^{-2\alpha(r)}$$
$$T^{11} = e^{-2\beta(r)}\mathrm{P}(r),\ T^{22} = r^{-2}\mathrm{P}(r),\ T^{33} = (r^2\sin^2\theta)^{-1}\mathrm{P}(r),$$
$$T^{ab} = 0, \text{ if } a \neq b.$$

Lowering the indices one gets:

$$T_{00} = \textstyle\sum_{r,s} g_{0,s}g_{0,r}T^{00} = g_{00}^2T^{00} = e^{4\alpha(r)}e^{-2\alpha(r)}\rho(r) = e^{2\alpha(r)}\rho(r)$$
$$T_{11} = g_{11}^2T^{11} = e^{4\beta(r)}e^{-2\beta(r)}\mathrm{P}(r) = e^{2\beta(r)}\mathrm{P}(r)$$
$$T_{22} = g_{22}^2T^{11} = r^2r^{-2}\mathrm{P}(r) = \mathrm{P}(r)$$
$$T_{33} = g_{33}^2T^{33} = (r^2\sin^2\theta)(r^2\sin^2\theta)^{-1}\mathrm{P}(r) = \mathrm{P}(r).$$

In (15.17) we had computed

$$\mathsf{Ric}_{00} = e^{2(\alpha(r)-\beta(r))}[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r)]$$
$$\mathsf{Ric}_{11} = -\alpha''(r) - \alpha'(r)^2 + \alpha'(r)\beta'(r) + \frac{2}{r}\beta'(r)$$
$$\mathsf{Ric}_{22} = e^{-2\beta(r)}[r(\beta'(r) - \alpha'(r)) - 1] + 1$$
$$\mathsf{Ric}_{33} = \sin^2(\theta)\mathsf{Ric}_{22}.$$

Thus, the scalar curvature $\mathrm{R} = \sum_{r,s} g^{rs}\mathsf{Ric}_{rs}$ would be equal to

$$\begin{aligned}\mathrm{R} &= -e^{-2\alpha(r)}\mathsf{Ric}_{00} + e^{-2\beta(r)}\mathsf{Ric}_{11} + r^{-2}\mathsf{Ric}_{22} + (r^2\sin^2\theta)^{-1}\mathsf{Ric}_{33} \\ &= -e^{-2\alpha(r)}\mathsf{Ric}_{00} + e^{-2\beta(r)}\mathsf{Ric}_{11} + \frac{2}{r^2}\mathsf{Ric}_{22} \\ &= -2e^{-2\beta(r)}\left[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r) - \frac{2}{r}\beta'(r) + \frac{1}{r^2} - \frac{1}{r^2}e^{2\beta(r)}\right].\end{aligned}$$

We know the field equations are

$$\mathsf{Ric}_{ab} - \frac{1}{2}\mathrm{R}g_{ab} = 8\pi T_{ab}.$$

For $a = b = 0$, we obtain:

$$\begin{aligned}&e^{2(\alpha(r)-\beta(r))}[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r)] + \\ &- e^{2\alpha(r)-2\beta(r)}\left[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r) - \frac{2}{r}\beta'(r) + \frac{1}{r^2} - \frac{1}{r^2}e^{2\beta(r)}\right] \\ &= 8\pi e^{2\alpha(r)}\rho(r).\end{aligned}$$

Simplifying one obtains:

$$1 - e^{-2\beta(r)} + 2\beta' r e^{-2\beta(r)} = 8\pi r^2 \rho(r). \tag{15.25}$$

For $a = b = 1$, one gets

$$\begin{aligned}
&-\alpha''(r) - \alpha'(r)^2 + \alpha'(r)\beta'(r) + \frac{2}{r}\beta'(r) + \\
&+ e^{-2\beta(r)} e^{2\beta(r)} \left[\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{2}{r}\alpha'(r) - \frac{2}{r}\beta'(r) + \frac{1}{r^2} - \frac{1}{r^2} e^{2\beta(r)}\right] \\
&= 8\pi e^{2\beta(r)} \mathrm{P}(r).
\end{aligned}$$

Simplifying one obtains:

$$\frac{e^{-2\beta(r)}}{r^2}\left(2r\alpha'(r) + 1 - e^{2\beta(r)}\right) = 8\pi \mathrm{P}(r). \tag{15.26}$$

In a similar fashion one gets for $a = b = 2$:

$$e^{-2\beta(r)}\left(\alpha''(r) + \alpha'(r)^2 - \alpha'(r)\beta'(r) + \frac{\alpha'(r)}{r} - \frac{\beta'(r)}{r}\right) = 8\pi \mathrm{P}(r). \tag{15.27}$$

Let us now use the equation of local conservation of energy $\sum_a (\nabla_{\partial_a} \overline{T})^{ab} = 0$ (13.3). Taking $b = 1$ this equation becomes (for simplicity, we will not write the variable r in the equations)

$$\nabla_{\partial_a} \overline{T} = \sum_c \frac{\partial T^{cc}}{\partial x^a} \partial x^c \otimes \partial x^c + \sum_{c,r} T^{cc} \Gamma^r_{ac} \partial x^r \otimes \partial x^c + \sum_{c,s} T^{cc} \Gamma^s_{ac} \partial x^c \otimes \partial x^s.$$

Hence the component $(\nabla_{\partial_a} \overline{T})^{a1}$ is given by

$$(\nabla_{\partial_a} \overline{T})^{a1} = \begin{cases} \frac{\partial T^{11}}{\partial x^1} + 2T^{11}\Gamma^1_{11}, & \text{if } a = 1 \\ \\ T^{11}(\Gamma^0_{01} + \Gamma^2_{21} + \Gamma^3_{31}) + T^{00}\Gamma^1_{00} + T^{22}\Gamma^1_{22} + T^{33}\Gamma^1_{33}, & \text{if } a \neq 1 \end{cases}.$$

Thus,

$$(\nabla_{\partial_a} \overline{T})^{a1} = e^{-2\beta}(\mathrm{P}' - 2\beta'\mathrm{P} + 2\beta'\mathrm{P} + \alpha'\mathrm{P} + \frac{2}{r} + \rho\alpha' - \frac{2}{r}) = 0,$$

From this we obtain:

$$(\rho(r) + \mathrm{P}(r))\alpha'(r) + \mathrm{P}'(r) = 0. \tag{15.28}$$

Let us define

$$m(r) = \frac{1}{2}(r - re^{-2\beta(r)}), \text{ so that}$$
$$e^{2\beta(r)} = \left(1 - \frac{2m(r)}{r}\right)^{-1}.$$

Taking the derivative with respect to r we obtain:

$$m'(r) = \frac{1}{2}(1 - e^{-2\beta(r)} + 2\beta' re^{-2\beta(r)})$$
$$= \frac{1}{2} \times 8\pi r^2 \rho(r) = 4\pi r^2 \rho(r),$$

where in the second line we have used equation (15.25) to substitute the expression inside the parenthesis for $8\pi r^2 \rho(r)$. This immediately gives us

$$m(r) = 4\pi \int_{r_0}^{r} r^2 \rho(r) dr, \text{ with } 0 \leq r_0 < r \leq r_1.$$

Now, in terms of $m(r)$, equation (15.27) can be written as:

$$\alpha'(r) = \frac{m(r) + 4\pi r^3 \mathrm{P}(r)}{r(r - 2m(r))}. \tag{15.29}$$

Combining (15.28) and (15.29) one gets the *Tolman-Oppenheimer-Volkoff* equation

$$\mathrm{P}'(r) = -\frac{(\rho(r) + \mathrm{P}(r))(m(r) + 4\pi r^3 \mathrm{P}(r))}{r(r - 2m(r))}. \tag{15.30}$$

We impose the natural boundary condition $\mathrm{P}(r_1) = 0$ since one does not expect any pressure at the very surface of S. Assuming $\rho(r) = \rho$ is *constant*, then, in terms of $\mathrm{P}(r)$, the function $\alpha(r)$ can be expressed as:

$$\alpha(r) = \int_{r_0}^{r} \frac{-\mathrm{P}'(s)}{\rho + \mathrm{P}(s)} ds = \ln\left(\frac{\rho + \mathrm{P}(r_0)}{\rho + \mathrm{P}(r)}\right) + \alpha(r_0),$$

and therefore:

$$e^{2\alpha(r)} = e^{2\alpha(r_0)} \left(\frac{\rho + \mathrm{P}(r_0)}{\rho + \mathrm{P}(r)}\right)^2. \tag{15.31}$$

For $r = r_1$ the boundary condition $\mathrm{P}(r_1) = 0$ implies:

$$e^{2\alpha(r_1)} = e^{2\alpha(r_0)} \left(1 + \frac{\mathrm{P}(r_0)}{\rho}\right)^2.$$

In order to match the outer Shwarzschild solution at the boundary $r = r_1$ one needs $e^{2\alpha(r_1)}$ to be equal to $(1 - 2\mathrm{M}/r_1)$, where $\mathrm{M} = 4/3\pi(r_1^3 - r_0^3)$ is the total Newtonian mass of S. Thus,

$$e^{2\alpha(r_0)} = \frac{e^{2\alpha(r_1)}}{\left(1 + \frac{\mathrm{P}(r_0)}{\rho}\right)^2} = \frac{1 - 2\mathrm{M}/r_1}{\left(1 + \frac{\mathrm{P}(r_0)}{\rho}\right)^2}.$$

From (15.31) one obtains

$$e^{2\alpha(r)} = \frac{(1 - 2\mathrm{M}/r_1)\left(\frac{\rho + \mathrm{P}(r_0)}{\rho + \mathrm{P}(r)}\right)^2}{\left(1 + \frac{\mathrm{P}(r_0)}{\rho}\right)^2} = \frac{(1 - 2\mathrm{M}/r_1)}{(1 + \mathrm{P}(r)/\rho)^2},$$

and the metric inside S would be given by:

$$g_{\mathrm{S}} = \frac{(1 - 2\mathrm{M}/r_1)}{(1 + \mathrm{P}(r)/\rho)^2} dt \otimes dt + \left(1 - \frac{2m(r)}{r}\right)^{-1} dr \otimes dr + r^2 dA. \qquad (15.32)$$

where $\mathrm{P}(r)$ is the unique solution of the *Tolman-Oppenheimer-Volkoff* with initial condition $\mathrm{P}(r_1) = 0$.

15.6.1 Interior of a Uniformly Dense Star

If our shell S is represented by the interior of a *uniformly dense* star of radius R, i.e., $\rho(r) = \rho$, $r_0 = 0$, $r_1 = \mathrm{R}$ there is an exact solution to (15.30). In this case $m(r) = 4/3\pi r^3$, for $0 < r \leq \mathrm{R}$, and equation (15.30) can be integrated ([4], [34]) to give:

$$\mathrm{P}(r) = \rho \frac{(1 - 2\mathrm{M}/\mathrm{R})^{1/2} - (1 - 2\mathrm{M}r^2/\mathrm{R}^3)^{1/2}}{(1 - 2\mathrm{M}r^2/\mathrm{R}^3)^{1/2} - 3(1 - 2\mathrm{M}/\mathrm{R})^{1/2}},$$

where $\mathrm{M} = m(r_1)$. We notice that the denominator vanishes if

$$r = \mathrm{R}/\mathrm{M}\sqrt{\mathrm{M}(9\mathrm{M} - 4\mathrm{R})},$$

which is a real number if and only if $9\mathrm{M} - 4\mathrm{R} \geq 0$. Hence, if we assume the pressure inside the star is finite (a reasonable physical assumption) then one must have $\mathrm{M} < 4/9\mathrm{R}$ (in standard units, $\mathrm{M} < 4\mathrm{R}/(9\mathrm{G}_N)$). Even under the more general assumption that $\rho(r)$ is not constant but monotonically decreasing, $d\rho/dr \leq 0$ (as a matter of fact, our original assumption that $\rho(r)$ is a constant function is rather unrealistic), one can show ([38], page 130) that for a star to be physically stable it is required that $M < 4\mathrm{R}/(9\mathrm{G}_N)$ (*Buchdahl's theorem*). If the mass of a star is not too big, once its fuel is exhausted, and cools down, it will attain a final state of equilibrium and becomes a white dwarf or a neutron star. However, if the mass of the star is greater than the *Tolman–Oppenheimer–Volkoff limit,* (three to four times the mass of the Sun) then equilibrium will never be achieved: Inner pressure will not support its own weight and the star will undergo a complete gravitational collapse, shrinking until its radius becomes smaller than the corresponding Schwarzschild radius. Once this threshold is surpassed, it will continue shrinking until it finally disappears, becoming a black hole, as we see in the next chapter. This final process occurs extremely fast: A stars a little larger than our Sun would disappear in about 10^{-5} seconds.

In terms of r, it is easy to see that the function $e^{2\alpha(r)}$ is given by

$$e^{2\alpha(r)} = \left[\frac{3}{2}\left(1 - \frac{2\mathrm{M}}{\mathrm{R}}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{2\mathrm{M}r^2}{\mathrm{R}^3}\right)^{1/2}\right]^2, \quad 0 < r < \mathrm{R}.$$

And we already know that

$$e^{2\beta(r)} = \left(1 - \frac{2m(r)}{r}\right)^{-1} = \left(1 - \frac{8\pi\rho}{3}r^2\right)^{-1}.$$

This completely determines the metric inside the star.

15.7 Geometry Inside a Spherical Empty Cavity

We want to show that the geometry inside an empty spherical cavity must be flat. It is reasonable to expect it, since such a shell would behave like a gravitational *Faraday's Cage,* where the gravitational forces cancels out and must be zero at any interior point, as we showed in Example 14.2.2.

By Birkhoff's theorem, one can choose coordinates $(\bar{t}, r, \theta, \phi)$ so the geometry inside any spherically symmetric static cavity must look like

$$g = -\left(1 - \frac{\mathrm{C}}{r}\right) d\bar{t} \otimes d\bar{t} + \left(1 - \frac{\mathrm{C}}{r}\right)^{-1} dr \otimes dr + r^2 dA, \ \text{with } 0 < r \leq r_1, \tag{15.33}$$

for a suitable constant C. But for all values of $0 < r < r_1$, one has $\mathrm{P}(r) = m(r) = 0$ in equation (15.29) , and therefore $\alpha'(r) = 0$. Consequently, $\alpha(r)$ must be constant. This forces $\mathrm{C} = 0$ in (15.33), and henceforth the metric is flat in the empty cavity inside the shell. *However, this does not imply that in the Schwarzschild global coordinates the metric is given by* (15.21), since Birkohff's theorem does not guarantee that the coordinates $(\bar{t}, r, \theta, \phi)$ must be the same as those of an observer at infinity. In fact we will show that for the time coordinate one has the relation $\bar{t} = \kappa t$, with

$$\kappa = \frac{(1 - 2\mathrm{M}/r_1)^{1/2}}{(1 + \mathrm{P}(r_0)/\rho)}.$$

Let see why: In Section 15.6 we showed that inside the material body of the shell the metric is given by

$$g = -e^{2\alpha(r)} dt \otimes dt + \left(1 - \frac{2m(r)}{r}\right)^{-1} dr \otimes dr + r^2 dA.$$

Since the metric inside the empty cavity determined by S is constant flat, by continuity it should coincide with the metric at the inner boundary $r = r_0$. Thus,

$$g_{\text{inside}} = -\kappa dt \otimes dt - dr \otimes dr + dA.$$

We notice that *after the change of coordinates* $\bar{t} = \kappa t$ *the metric becomes the standard Minkowski metric of flat space time.*

15.8 Time Machines

Formula (15.22) can also be interpreted as time dilation. We already know that the compared frequencies ω_B and ω_A of a pulse of light as measured by B and A is given by

$$\omega_A = \omega_B \frac{\sqrt{-g_{00}(r_B)}}{\sqrt{-g_{00}(r_A)}} = \omega_B Q(\mathrm{M}, r_B, r_A). \tag{15.34}$$

where the factor $Q(\mathrm{M}, r_B, r_A)$ is less than one. In standard units

$$Q(\mathrm{M}, r_B, r_A) = \left(\frac{1 - 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2 r_B}{1 - 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2 r_A}\right)^{1/2}. \tag{15.35}$$

We may imagine that each pulse of light emitted by B corresponds to the ticking of a clock he uses to measure his proper time. The frequency of the light signal emitted is measured by him to be $2\pi/\Delta s_B$, where Δs_B is the corresponding period of the light wave. Now, suppose A receives these signals at intervals Δs_A (measured in A's proper time), so that $\omega_A = 2\pi/\Delta s_A$. From (15.34) one obtains $\Delta s_B = Q(\mathrm{M}, r_B, r_A)\Delta s_A < \Delta s_A$.

To see how time dilates, suppose that B hovers very close to M, let's say just a meter away form the horizon of *V616 Monocerotis*, the closest known black hole, believe to be located about three thousand light years away. Its mass is estimated to be eleven times that of our Sun: $\mathrm{M} = 11 \times 1.989 \times 10^{30}$ kg. Its Schwarzschild radius would then be $R = 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2 \simeq 32.4$ km. On the other hand, we assume observer A hovers 10 km away from B. With these parameters one can calculate the factor $Q(\mathrm{M}, r_B, r_A)$ as approximately equal to 0.01. Now, suppose that in A's proper time a hole century has elapsed. Let's divide A's worldline into a collection events, one per minute, as measured in A's proper time, for a total of about 5.2×10^7 equidistant points on this line, equally separated as if they were frames in a movie. But then, the factor Q tells us that one hundred frames in A's worldline are compressed into one frame of B's worldline. Thus, the entire "movie", a whole century in A's world, could be watched by B in fast-motion in just one year! But, how much force would B need to hover just one meter above the horizon of this black hole? Suppose, for instance, B manages to hover steadily by using the thrust of a rocket. Let's compute how much force this would demand from the engines. B's worldline, parametrized by arc length, would be

$$O_B(s) = \left((1 - \frac{2\mathrm{M}}{R_B})^{-1/2}s, R_B, \theta_B, \phi_B\right)$$

(R_B, θ_B, ϕ_B are constants).

As we discussed in Section 10.4, at any point $p = O_B(s_0)$ the 4-acceleration A_p of O_B at p coincides with O_B's 3-acceleration a_p, as measured in his own frame of reference. An orthonormal base for O_B at p is given by the vectors $e_a = \partial_a/|\partial_a|$. Then the total force he experiences would then be his rest mass times $|\mathrm{a}_p|$. Let us compute A_p in the coordinate frame of O_B at p : If

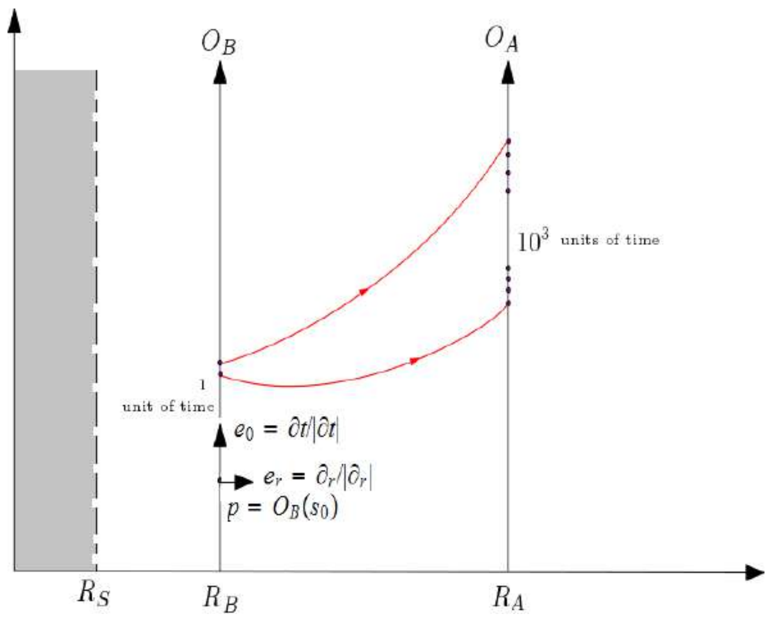


Figure 15.1: Time dilation

we denote $(1 - 2\mathrm{M}/R_B)^{-1/2}$ by δ we see that

$$\begin{aligned} A_p = \nabla_{\mathbf{u}(s)}\mathbf{u}(s) &= \delta^2\nabla_{\partial^t}\partial^t = \delta^2{\textstyle\sum_c}\Gamma^c_{00}\,\partial_r|_p \\ &= \frac{\mathrm{M}}{R_B^2}(1-\frac{2\mathrm{M}}{R_B})^{-1}(1-\frac{2\mathrm{M}}{R_B})\,\partial_r|_p \\ &= \frac{\mathrm{M}}{R_B^2}\partial_r = \frac{\mathrm{M}}{R_B^2}\,|\partial_r|\,e_r = \frac{\mathrm{M}}{R_B^2}(1-\frac{2\mathrm{M}}{R_B})^{-1/2}e_r, \end{aligned}$$

since $|\partial_r| = (1 - 2\mathrm{M}/R_B)^{-1/2}$ at the point p. In standard units of mass and time:

$$|\mathrm{a}_p| = \frac{\mathrm{G}_N\mathrm{M}}{R_B^2}(1-\frac{2\mathrm{G}_N\mathrm{M}}{\mathrm{c}^2R_B})^{-1/2}\ \mathrm{m}\times\mathrm{s}^{-2}.$$

This differs from Newtonian acceleration by the factor $(1 - 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2R_B)$, which is very small when $R_B >> 2\mathrm{G}_N\mathrm{M}/\mathrm{c}^2$. Substituting the values for G_N, M, $R_B = 2\mathrm{G}_N\mathrm{M} + 1 = 32449.98$ m, $R_A = R_B + 10000$ m, we obtain $|\mathrm{a}_p| \simeq 2.5\times10^{14}$ m$\times$s^{-2}. That is, at one meter from the horizon, B would experience a force exerted by his rocket engines similar to that he would feel on Earth being under the weight of a mass the size of mount Everest!

The mental experiment we just discussed tells us that using a black hole as a time machine does not seem to be feasible. There is, however, one way of canceling the overwhelming gravitational forces surrounding a big mass: one could stay inside a homogeneous spherical shell where the total gravitational force must be zero.

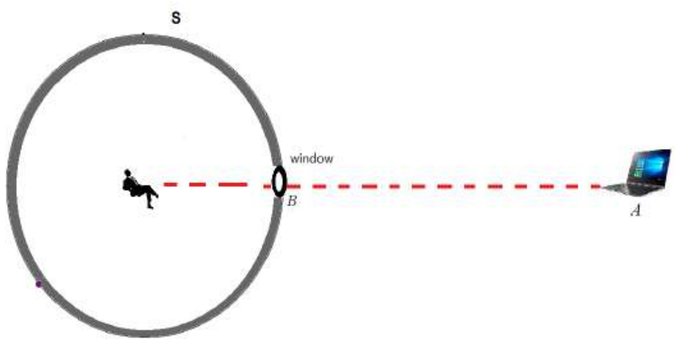


Figure 15.2: Time machine

We already know that inside the cavity determined by S space-time is flat, where the metric is given in global Schwarzschild coordinates by

$$g_{\text{inside}} = -\frac{(1 - 2\mathrm{M}/r_1)}{(1 + \mathrm{P}(r_0)/\rho)^2} dt \otimes dt - dr \otimes dr + dA.$$

As an example, consider a "thin" shell, let's say with dimensions $r_0 = 100$, $r_1 = 101$ m. We assume it has a constant density equal to $\rho = 0.000393$ kgg/m^3. Then, its total mass in *big kilograms* would be M = 49.8 kgg. In standard kilograms, M would be equal to 6.4×10^{28} kg, approximately 60 times the mass of Jupiter. By solving numerically equations (15.29) and (15.30) (see program below) one can show that

$$e^{2\alpha(r_0)} \simeq k(1 - 2\mathrm{M}/r_1),$$

since $k \simeq 1 + 10^{-5}$. Hence, the proper time of an observer inside the shell would be approximately equal to one right outside, next to it, that we know is equal to $(1 - 2\mathrm{M}/r_1)^{1/2} \simeq 0.1$, which is also the time dilation factor with respect to an observer very far away (at "infinity") (15.35). If one used the shell as a time machine, it would be possible to observe an entire century of events in the exterior world in just ten years. The only problem, of course, would be that to construct such a shell one would need a material with a density equal to 5.1×10^{23} kg/m^3, approximately a million times more dense than the densest object know in the universe, a neutron star!

Metric in a spherical shell

We fix coordinates (t,r,θ,φ), and consider a perfectly symmetric spherical shell

$$S = \{(r,\theta,\varphi): A \le r \le B,\ 0 < \theta < \pi/2,\ 0 < \phi < 2\pi\}$$

We assume the metric is given by

$$g = -e^{2a(r)}dt^2 + e^{-2b(r)}dr^2 + r^2 d\theta^2 + r^2\sin(\theta)^2 d\phi^2$$

The program *shell* takes as input the dimensions of the shell, *A, B,* and its density, ρ(r) = *rho* (we assume it is a constant), and a value $A \le u \le B$, and returns a list with the values of

1) $-g_{tt}(u) = e^{2a(u)}$,

2) *The pressure function* $P(u)$. The pressure is given in standard units of $\frac{kF}{m^2}$, *where kF* $= 1$ *kilogram force*, or *equivalently* $1\ kF = 9.8$ *Newtons*.

3) The total mass of the shell, m(u).

Computations are carried out numerically.

```
> shell:=proc(A,B,rho,u) local m, e1, e2, f, g, P, t, C, dsol,cc;
> m:=r-> piecewise(r<=A,0,r<=B,(4/3)*Pi*(r^3-A^3),B<r,(4/3)*Pi*(B^3
  -A^3))*rho;
> e2:=diff(P(r),r)+(m(r)+4*Pi*P(r))/(r^2-2*r*m(r))*(rho+P(r)); C:=
  (1-2*m(B)/B);cc:=1.34/0.9*10^42;
> dsol:=dsolve({e2,P(B)=0}, range = A..B, numeric, output =
  listprocedure);
> g:= eval(P(r),dsol); t:=C*((1+g(A)/rho)^(-2))*(rho+g(A))^2/(rho+g
  (u))^2;
> return([t(u),cc*g(u),m(u)]);
> end proc:
>
```

Figure 15.3: Metric inside a shell

15.9 Planetary Motion

Before discussing how particle and photons move in Schwarzschild space-time we want to review the fundamentals of Newton´s theory of planetary motion.

In Newtonian gravity a particle P of mass m moves under the action of a *central force* F. If the motion of P is given by a curve $\alpha(t) = (r(t), \theta(t), \phi(t))$ in standard spherical coordinates (r, θ, ϕ) then $F = -\frac{\mathrm{G}_N \mathrm{M}m}{|\alpha(t)|^2}\alpha(t)$, where $\alpha(t) = \alpha(t)/\left|\alpha(t)\right|$ is the unitary vector determined by $\alpha(t)$:

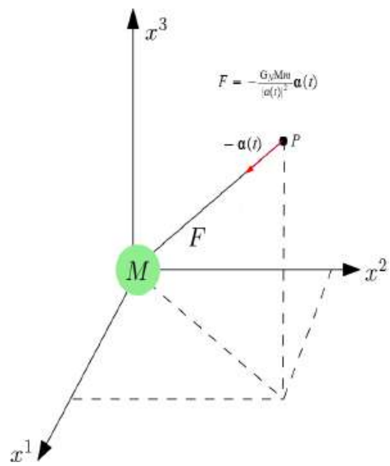


Figure 15.4: Newtonian force of gravity

We recall that the *angular momentum* of P is defined as the vector product of its position vector times its linear momentum: $L(t) = m\alpha(t) \times \alpha'(t)$. Since $\alpha''(t) = \frac{-\mathrm{MG}_N}{|\alpha(t)|^3}\ \alpha(t)$ one has $\alpha(t) \times \alpha''(t) = 0$. This implies:

$$L'(r) = m\ \alpha'(t) \times \alpha'(t) + m\ \alpha(t) \times \alpha''(t) = 0,$$

and therefore $L = L(t)$ *must be constant.* Since $\alpha(t)$ is at all times perpendicular to L it must be lie on the plane perpendicular to L. By rotating coordinates, if necessary, we may assume this is the equatorial plane $\theta = \pi/2$. Using formulas (**??**) to express $\alpha'(t)$ (the variable t is suppressed) one gets:

$$\begin{aligned} \alpha' &= r'e_r + r\theta'\ e_\theta + r\phi' \sin\theta\ e_\phi \\ \alpha'' &= [(r'' - r(\phi')^2 \sin^2\theta - r(\theta')^2]e_r \\ &\quad + [r\theta'' + 2r'\phi' - r(\phi')^2 \sin\theta\cos\theta]e_\theta \\ &\quad + [r\phi'' \sin\theta + 2r'\phi' \sin\theta + 2r'\phi' \cos\theta]e_\phi. \end{aligned} \tag{15.36}$$

Thus, the magnitude of L is equal to $\mathrm{L} = mr^2(t)\phi'(t)$.

On the other hand, let us recall that the gravitational potential is given by $U(t) = -\mathrm{G}_N\mathrm{M}m/\left|\alpha(t)\right|$, so that the total work W done by F to move a particle from $p = \alpha(t_1)$ to $q = \alpha(t_2)$ is equal to

$$W = \int_{r(t)} F \bullet \alpha'(t) ds = -\int_{t_1}^{t_2} \operatorname{grad} U(\alpha(t)) \cdot \alpha'(t) dt$$
$$= U(\alpha(t_1)) - U(\alpha(t_2)) = U(p) - U(q).$$

But the work done by F is equal to the increment in kinetic energy of the

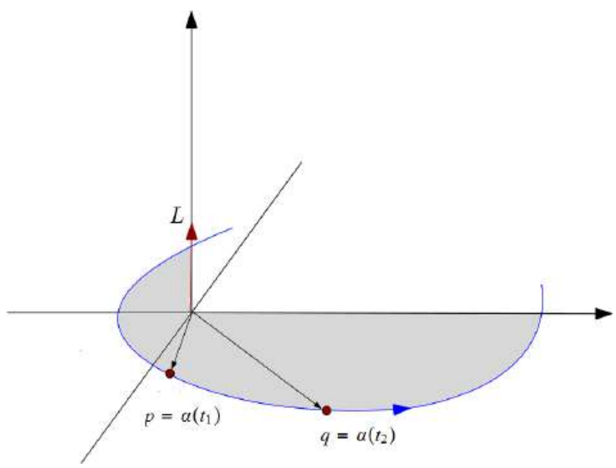


Figure 15.5: Orbits under a central force

particle: $\mathrm{K}(t) = \frac{1}{2} m |\alpha'(t)|^2$. That is: $U(p) - U(q) = \mathrm{K}(q) - \mathrm{K}(p)$. If the total energy of the particle is defined to be $\mathrm{E} = U(\alpha(t)) + \mathrm{K}(\alpha(t))$, then we have showed that the total energy of P *stays constant along its path*. Using (15.36) we compute $\mathrm{K}(\alpha(t))$ as

$$\mathrm{K}(\alpha(t)) = \frac{1}{2} m \left(r'(t)^2 + r^2(t) \phi'(t)^2 \right). \tag{15.37}$$

Substituting $\phi'(t) = \mathrm{L}/mr^2(t)$, and $\mathrm{K} = \mathrm{E} - U$ in (15.37) one obtains the equation of motion for P:

$$\mathrm{E} = \frac{1}{2} m \left(\frac{dr}{dt} \right)^2 + \left(\frac{\mathrm{L}^2}{2mr^2(t)} - \frac{\mathrm{G}_N \mathrm{M} m}{r(t)} \right) \tag{15.38}$$

(we used that $|\alpha(t)| = r(t)$). The term inside the parenthesis is called the *effective potential*, and will be denoted by $V(t)$.

Equation (15.38) can be solved for r in terms of ϕ, as follows: First, we write

$$\frac{dr}{dt} = \frac{dr}{d\phi}\frac{d\phi}{dt} = \frac{dr}{d\phi}\frac{\mathrm{L}}{mr^2(t)}.$$

Define a new function $u(t)$ such that $r(t)u(t) = \mathrm{L}^2/(\mathrm{M}m)$. In terms of ϕ one has:

$$\frac{dr}{d\phi} = \frac{dr}{du}\frac{du}{d\phi} = \frac{-\mathrm{L}^2}{\mathrm{M}m}\frac{1}{u^2},$$

and therefore:

$$\frac{dr}{dt} = \frac{\mathrm{L}}{mr^2}\frac{(-\mathrm{L}^2)}{m\mathrm{M}}\frac{1}{u^2}\frac{du}{d\phi} = \frac{-\mathrm{M}}{\mathrm{L}}\frac{du}{d\phi}.$$

We can write (15.38) in terms of u as follows:

$$\mathrm{E} = \frac{1}{2}\frac{m\mathrm{M}^2}{\mathrm{L}^2}\left(\frac{du}{d\phi}\right)^2 + \left(\frac{m\mathrm{M}^2}{2L^2}u^2 - \frac{m\mathrm{M}^2}{\mathrm{L}^2}u\right).$$

Multiplying both sides by $2\mathrm{L}^2/(m\mathrm{M}^2)$ one obtains:

$$\frac{2L^2\mathrm{E}}{m\mathrm{M}^2} = \left(\frac{du}{d\phi}\right)^2 + u^2 - 2u.$$

Taking derivatives with respect to ϕ on both sides gives the equation:

$$2\frac{du}{d\phi}(\frac{d^2u}{d\phi^2} + u - 1) = 0.$$

Hence, in order to obtain the motion for P one just has to solve the differential equation $d^2u/d\phi^2 + u - 1 = 0$. It is easy to see that its solutions are given by $u = 1 + \varepsilon\cos\phi$, which represents the equation of an ellipse in polar coordinates, with eccentricity $\varepsilon = 1 - b^2/a^2$, where $a > b$ are the lengths of the corresponding axes. *Kepler's law* is just the fact that $\mathrm{L} = r^2(t)\phi'(t)$ is constant: This means that the area swept by a planet during an interval of time $\Delta t = t_2 - t_1$ is equal to $\int_{t_1}^{t_2} r^2(t)\phi'(t)dt = \Delta t\ \mathrm{L}$.

15.10 Motion of Nonzero Mass Particles

In this section we will derive an equation similar to (15.38) for a particle that moves in Schwarzschild space-time $(\mathbb{R}^4, g)$, where g denotes the metric (15.21).

We start discussing the motion of a particle P with nonzero mass that moves along a geodesic $\sigma : I \to \mathbb{R}^4$, that we assume parametrized by arc length Since the Schwarzschild metric is independent of the coordinates t and ϕ, the fields ∂_t and ∂_ϕ are Killing vector fields. Hence, by Proposition **??** $\langle \sigma'(s), \partial_t \rangle$ and $\langle \sigma'(s), \partial_\phi \rangle$ stay constant along $\sigma(s)$. Define $\mathrm{e} = -\langle \sigma'(s), \partial_t \rangle$, and $l= \langle \sigma'(s), \partial_\phi \rangle$. If we write $\sigma(s) = (t(s), r(s), \theta(s), \phi(s))$, then we see that $\mathrm{e} = t'(s)(1 - 2\mathrm{M}/r(s))$ and $l= r^2(s) \sin^2 \theta(s) \phi'(s)$.

Remark 15.10.1. The constant e *can be interpreted as the total energy of a particle P of total rest mass one $m_0 = 1$, as measured by an observer at infinity.* In fact, as we showed (15.4.1), the 4-velocity $\mathbf{u}_0$ of a very far away observer is given approximately by ∂t. By (10.25). The 4-velocity of P is given by $\mathbf{v}_0 = t'(s)\partial_t +$ v, where $\langle \partial_t, \mathrm{v} \rangle = 0$. Its total energy, as measured by O, would then be

$$\begin{aligned} E &= -m_0 \langle \mathbf{u}_0, \mathbf{v}_0 \rangle = -m_0 \langle \partial_t,\ t'(s)\partial_t + \mathrm{v} \rangle \\ &= m_0 t'(s)(1 - \frac{2\mathrm{M}}{r(s)}) = m_0 \mathrm{e}. \end{aligned}$$

As for the planetary motion, we claim that *the spatial trajectory of P is contained in a plane in $\mathbb{R}^3$.* After performing an appropriated translation, and then suitable a rotation of the coordinates, one can always assume $r(s) \neq 0$, for all $s \in I$, and that $\phi'(0) = 0$ (for this last condition to be satisfied one just have to choose the axes so that $\partial_\theta = \sigma'(0)$). Hence, in this new coordinate system $l = 0$. We claim in this frame $\phi(s)$ must be a constant, and therefore P moves in a plane perpendicular to the equatorial plane $\theta = \pi/2$. In fact, $r^2(s) \sin^2 \theta(s) \phi'(s) = 0$ forces $\phi'(s) = 0$, since $\sin \theta(s) = 0$ only for $\theta(s) = 0, \pi$, which are excluded in spherical coordinates.

Again, rotating coordinates, if necessary, we may assume without loss of generality that the plane of motion is the equatorial plane $\theta = \pi/2$. In these coordinates one obtains the same formula as for the conservation of angular momentum in Newtonian mechanics of a particle of unit mass, i.e., $r^2(s)\phi'(s)$ must be constant: $l = r^2(s)\phi'(s)$. Hence, *l should be interpreted as angular momentum.*

On the other hand, P moves in a timelike geodesic $\langle \sigma'(s), \sigma'(s) \rangle = -1$. Thus,

$$-(1 - \frac{2\mathrm{M}}{r(s)})t'(s)^2 + (1 - \frac{2\mathrm{M}}{r(s)})^{-1} r'(s)^2 + r^2 \phi'(s)^2 = -1 \tag{15.39}$$

Solving for $t'(s)$ and $\phi'(s)$ in $\mathrm{e} = t'(s)(1 - 2\mathrm{M}/r(s))$ and $l = r^2(s)\phi'(s)$, respectively, and substituting in (15.39) one obtains:

$$-(1 - \frac{2\mathrm{M}}{r(s)})^{-1}\mathrm{e}^2 + (1 - \frac{2\mathrm{M}}{r(s)})^{-1}\left(\frac{dr}{ds}\right)^2 + \frac{l^2}{r(s)^2} = -1,$$

which can be rewritten as:

$$\frac{(\mathrm{e}^2 - 1)}{2} = \frac{1}{2}\left(\frac{dr}{ds}\right)^2 + \left(\frac{l^2}{2r(s)^2} - \frac{\mathrm{M}}{r(s)} - \frac{\mathrm{M}l^2}{r(s)^3}\right).$$

In standard units of mass and time this equation becomes:

$$\frac{(\mathrm{e}^2 - 1)}{2} = \frac{1}{2}\frac{1}{\mathrm{c}^2}\left(\frac{dr}{ds}\right)^2 + \frac{1}{\mathrm{c}^2}\left(\frac{l^2}{2r(s)^2} - \frac{\mathrm{G}_N\mathrm{M}}{r(s)} - \frac{\mathrm{G}_N\mathrm{M}l^2}{\mathrm{c}^2r(s)^3}\right), \tag{15.40}$$

Suppose now that P has rest mass m. If we multiply both sides of (15.40) by $m\mathrm{c}^2$, and define $\mathrm{L} = ml$, then (15.40) becomes:

$$m\mathrm{c}^2\frac{(\mathrm{e}^2 - 1)}{2} = \frac{1}{2}m\left(\frac{dr}{ds}\right)^2 + \left(\frac{\mathrm{L}^2}{2mr(s)^2} - \frac{\mathrm{G}_N\mathrm{M}m}{r(s)} - \frac{\mathrm{G}_N\mathrm{L}^2}{m\mathrm{c}^2r(s)^3}\right). \tag{15.41}$$

Let's denote by E the term on the left. Then

$$\mathrm{e} = \left(1 + \frac{2\mathrm{E}}{m\mathrm{c}^2}\right)^{1/2} \simeq 1 + \frac{\mathrm{E}}{m\mathrm{c}^2},$$

when $\mathrm{E} << m\mathrm{c}^2$. Hence, $m\mathrm{c}^2\mathrm{e} = m\mathrm{c}^2 + \mathrm{E}$, and if $r >> 0$ we know that $m\mathrm{c}^2\mathrm{e}$ would correspond to the total energy of P, as measured by an observer at infinity which is the sum of its rest energy + E. Therefore E *can be interpreted as the Newtonian or kinetic energy of* P. In this way, (15.41) would be the analogue of (15.38), but with an extra term $\mathrm{G}_N\mathrm{L}^2/m\mathrm{c}^2r(s)^3$, which is very small if r is large. Hence, Newtonian orbits are just approximations of orbits in Einstein's theory. However, their qualitative behavior could be quite different. As we shall see, they are not necessarily conics, or even closed orbits.

15.11 Orbits in Schwarzschild's Space-Time

In the previous section we obtained the fundamental equation of motion of a particle. In geometric units:

$$\mathrm{E} = \frac{1}{2}\left(\frac{dr}{ds}\right)^2 + \left(\frac{l^2}{2r(s)^2} - \frac{\mathrm{M}}{r(s)} - \frac{\mathrm{M}l^2}{r(s)^3}\right). \tag{15.42}$$

The term

$$V(r) = \frac{l^2}{2r^2} - \frac{\mathrm{M}}{r} - \frac{\mathrm{M}l^2}{r^3}$$

is called the *effective potential*, and E the Newtonian energy, or simply the *energy of* P. By analyzing the graph of $\mathrm{E} - V(r)$ one can deduce the qualitative behavior of P. For this we write:

$$\frac{dr}{ds} = \pm\sqrt{2}[\mathrm{E} - V(r(s))]^{1/2}, \tag{15.43}$$

where the sign of the radical is the same as the sign of $dr/ds|_{s=0}$: It is positive, if P starts moving with radial velocity pointing away from the central mass M, and negative if the radial velocity vector points towards M. Clearly, motion can only occur if $\mathrm{E} \geq V(r(s))$. Now, the function V can be written in terms of l/M and r/M as:

$$V(r) = \frac{(l/\mathrm{M})^2}{2(r/\mathrm{M})^2} - \frac{1}{(r/\mathrm{M})} - \frac{(l/\mathrm{M})^2}{(r/\mathrm{M})^3}.$$

For different values of l/M the graph of V looks like:

The behavior of orbits then depends on $V(r)$ and its relationship to E. By equating $V'(r)$ to zero one easily sees that $V(r)$ has critical points at

$$r = \frac{l^2}{2\mathrm{M}}\left[1 \pm \sqrt{1 - 12\left(\frac{\mathrm{M}}{l}\right)^2}\right].$$

If $l/\mathrm{M} \geq \sqrt{12}$, the function $V(r)$ has a maximum and a minimum, and it has only one maximum if $l/\mathrm{M} = \sqrt{12}$. On the other hand, the function $V(r)$ is strictly increasing if $l/\mathrm{M} < \sqrt{12}$. Let us analyze each case separately.

1. Case: $l/\mathrm{M} < \sqrt{12}$ the particle P either plunges into M or escapes to infinity. Let's take, for instance, $l/\mathrm{M} = \sqrt{12} - 1/2$, and $\mathrm{E} = -0.15$. i) Assume $dr/ds|_{s=s_1} \geq 0$. Let's say P starts at $r_0 = r(0) = 2.5$. Then r must increase until it gets to the *turning point* $r_1 = r(s_1) \simeq 3.22$, where $V(r_1) = \mathrm{E}$. At this point $dr/ds|_{r=r_1} = 0$. After it reaches this point r starts decreasing to zero (the sign would be negative in front of the square root in 15.43). In fact, it cannot continue increasing, since for $r > r_1$ (15.43) has no solutions. On the other hand, it cannot stay

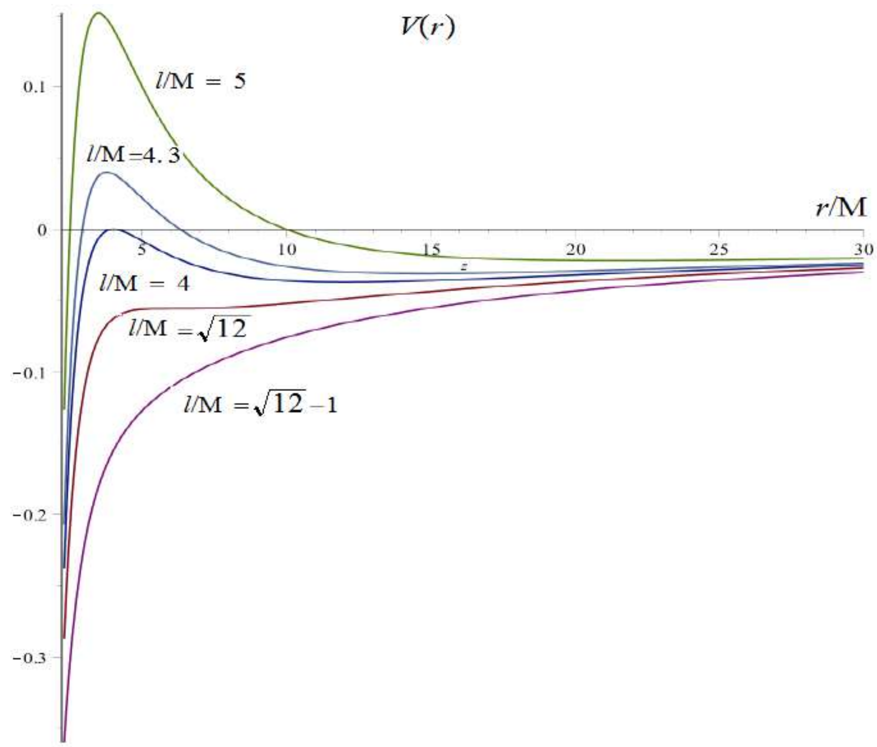


Figure 15.6: Effective potential for nonzero mass particles

constant equal to zero, either. To see this, one can differentiate both sides of (15.42) to obtain:

$$\frac{dr}{ds}\frac{d^2r}{ds^2} = \frac{-dV(r)}{dr}. \tag{15.44}$$

Hence, $dr/ds|_{s=s_1} = 0$ forces $d^2r/ds^2|_{s=s_1} = +\infty$, so that

$$\lim_{s\to s_1} \frac{dr}{ds}\frac{d^2r}{ds^2} = \left.\frac{-dV(r)}{dr}\right|_{r=r_1}.$$

But $r(s)$ cannot stay constant with value $r_1 = r(s_1)$, for if $dr/ds = 0$ for $s \geq s_1$ then r would not be a smooth function (not even of class C^2), since d^2r/ds^2 would be undefined at $s = s_1$: ii) If $dr/ds|_{r=r_0} < 0$, then clearly r must decrease forever.

2. Using that same example, if we take $E \geq 0$ instead, then if $dr/ds|_{r=r_0} \geq 0$ the particle will never reach a turning point and escapes to infinity.

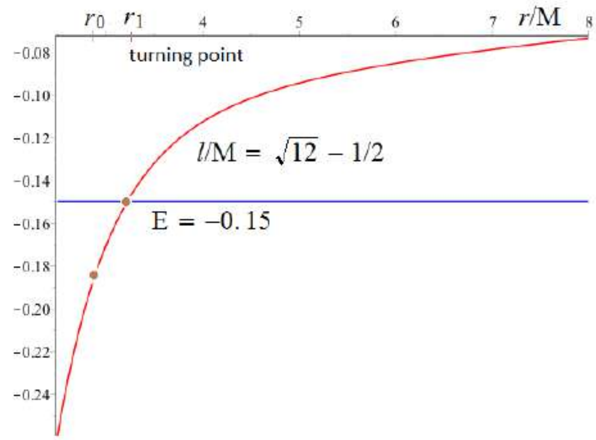


Figure 15.7:

On the other hand, if $dr/ds|_{r=r_0} < 0$ then r must decrease forever, and then P plunges into M.

3. The same analysis as before shows that whatever value of l/M, if $\mathrm{E} > V(r)$ then the particle plunges into M, if $dr/ds|_{r=r_0} < 0$, and escapes to infinity, if $dr/ds|_{r=r_0} > 0$.

4. Bound orbits can appear when $l/\mathrm{M} > \sqrt{12}$ and there are two turning points. For instance, if $l/\mathrm{M} = 4.5$, and $\mathrm{E} = -0.025$, the points $r_1 \simeq 12.1$ and $r_2 \simeq 25.2$ satisfy the equation $V(r) = \mathrm{E}$:

If P starts at r_0, with $r_1 < r_0 < r_2$, then, if $dr/ds|_{r=r_0} > 0$ the value of r increases until it reaches r_2, and then decreases to r_1. After reaching this point it starts increasing again, and the cycle repeats. In a similar way, if $dr/ds|_{r=r_0} < 0$, then r decreases until it reaches r_1, then it starts increasing up to r_2, and the it decreases again down to r_1, and the cycle repeats. One *orbit is, by definition, the motion of P between two successive inner (or outer) turning points.* The following picture shows nineteen successive points of a bound orbit where the red dots (points 8 and 17) mark the passage of P at two successive inner turning points. The angle swept is equal to $2\pi + \delta\phi$, where $\delta\phi \simeq \pi/2$. The orbit is called a *closed orbit* if $\delta\phi = 0$. Otherwise, the turning point is said to *precess*.

5. Among the closed orbits, circular orbits are the most simple. They occur if $r = r_1$ is constant, and $V(r_1) = \mathrm{E}$.

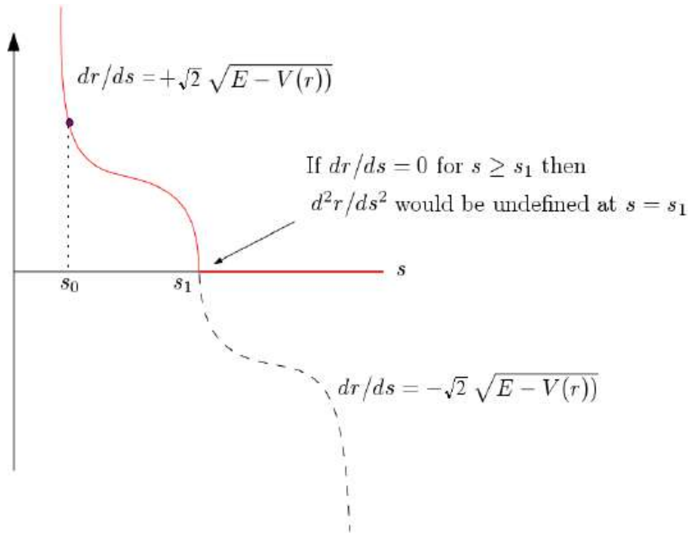


Figure 15.8:

For $\mathrm{E} = \mathrm{E}_1$ the orbit is unstable (any small perturbation of r will get P out of its orbit) while it is stable for $\mathrm{E} = \mathrm{E}_2$, since for all r near r_1 one has $\mathrm{E}_2 < V(r)$, and therefore there are no solutions to (15.43). It is easy to see that the most inner stable orbit occurs when $l/\mathrm{M} = \sqrt{12}$, for $r = 6\mathrm{M}$.

15.12 Radial Plunge Orbits

One specially simple case of an orbit is the radial free fall of a particle P originally at rest, coming from far away $r_0 = r(0) >> 0$. By this we mean that the particle starts with no angular momentum, i.e., $l = 0$, and no kinetic energy $\mathrm{E} = 0$. Hence, equation 15.42 becomes:

$$\frac{dr}{ds} = \pm\sqrt{\frac{2\mathrm{M}}{r(s)}}.$$

Since we are assuming $dr/ds < 0$, (P is coming inward) we must choose the negative root, and the equation of motion becomes

$$r(s)^{1/2} r'(s) = -\sqrt{2\mathrm{M}}. \tag{15.45}$$

By integrating both sides one obtains:

$$r(s) = (3/2)^{2/3}(2\mathrm{M})^{1/3}(s_0 - s)^{2/3}, \tag{15.46}$$

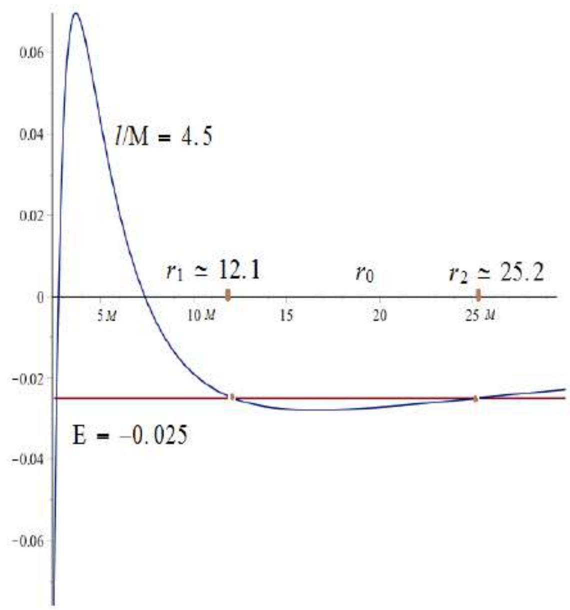


Figure 15.9:

where s_0 is the constant of integration satisfying $r_0 = (3/2)^{2/3}(2\mathrm{M})^{1/3}(s_0)^{2/3}$. Clearly, s_0 *represents the proper time measured by* P *from* $r = r_0$ to $r = 0$.

Now, $\mathrm{E} = 0$ implies $\mathrm{e} = 1$, and since $\mathrm{e} = t'(s)(1-2\mathrm{M}/r(s))$, using (15.45) we obtain:

$$\frac{dt}{dr} = \frac{dt/ds}{dr/ds} = \frac{t'(s)}{r'(s)} = \frac{-\ r^{1/2}}{\sqrt{2\mathrm{M}}(1-2\mathrm{M}/r)} = \frac{-(r/2\mathrm{M})^{1/2}}{1-2\mathrm{M}/r}.$$

If we let $z = r/2\mathrm{M}$, then the equation above becomes

$$\frac{dt}{dz} = \frac{dr}{dz}\frac{dt}{dr} = \frac{-z^{1/2}}{1-z^{-1}} = -2\mathrm{M}\frac{z^{3/2}}{z-1}.$$

Integrating both sides gives:

$$\begin{aligned} t(r) &= 2\mathrm{M}\left(-(2/3)z^{3/2} - 2z^{1/2} + \ln\left(\frac{z^{1/2}-1}{z^{1/2}+1}\right)\right) + t_0 \qquad (15.47) \\ &= 2\mathrm{M}\left[-\frac{2}{3}\left(\frac{r}{2\mathrm{M}}\right)^{3/2} - 2\left(\frac{r}{2\mathrm{M}}\right)^{1/2} + \ln\left|\frac{(r/2\mathrm{M})^{1/2}+1}{(r/2\mathrm{M})^{1/2}-1}\right|\right] + t_0, \end{aligned}$$

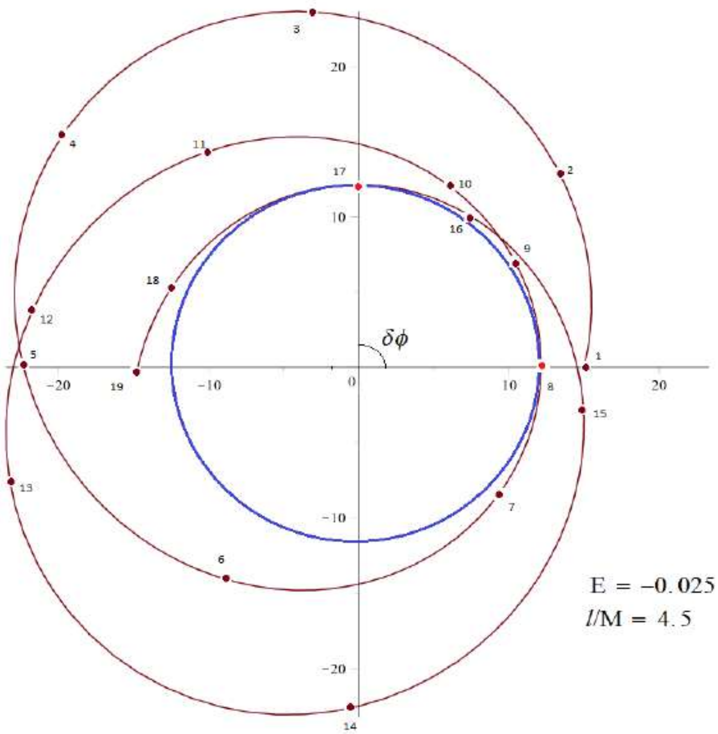


Figure 15.10: Precession of the orbit

where t_0 is an arbitrary constant of integration which can be fixed by choosing $t(r_0) = 0$.

We notice that even though the particle P takes only a finite amount of proper time to reach $r = 0$, it takes infinite time, as measured by an observer at infinity (infinite coordinate time). This, since $t(r) \to \infty$ as $r \to 2\text{M}$. As we shall see in the next chapter, $r = 2\text{M}$ may be interpreted as the horizon of a black hole. Thus, an observer plunging into it is never seen by an outside observer to cross this threshold: As it approaches the horizon, light coming from P becomes dimer and dimer, as it gets more and more red shifted. But P indeed crosses the horizon, and, as we will see, once that occurs it is almost instantly destroyed by the enormous tidal forces inside before it disappears at the central singularity.

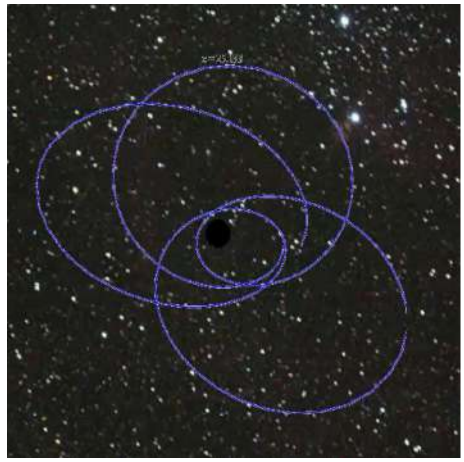


Figure 15.11:

15.13 Precession of Mercury's Perihelion

An "anomalous" precession of the perihelion of Mercury had been noticed since 1859. By analyzing observations of transits of Mercury over the Sun's disk from 1697 to 1848, the famous french astronomer Urbain Le Verrier showed that the actual rate of precession of Mercury's perihelion disagreed from that predicted from Newton's theory by 38" (arc seconds) per century (later reestimated at 43") [41]. Many ad-hoc explanations were devised: The existence of another planet, Vulcan, was postulated. Later it was suggested that dark dust between the Sun and Mercury was responsible for this anomaly, but none of these hypothesis were consistent with observations. The phenomenon was fully explained for the first time by Einstein, and it was the first empirical evidence of his new theory of gravitation (we follow [27]).

In the previous section we defined an *orbit* as the motion between two successive turning points r_1 and r_2, and the precession of the orbit as $\delta\phi$, where the angle swept between these two points is equal to $\Delta\phi = 2\pi + \delta\phi$. Hence, we want to estimate $\Delta\phi$ between two two successive inner turning points p_1 and p_3 : We assume P moves with effective potential and energy given by the graph: and that it starts at r_0 such that $dr/ds|_{r=r_0} < 0$, and with $r_1 < r_0 < r_2$, where r_1, r_2 are the respective radius of the inner turning point p_1 and the outer turning point p_2. Using that $d\phi/ds = l/r^2$ and equation

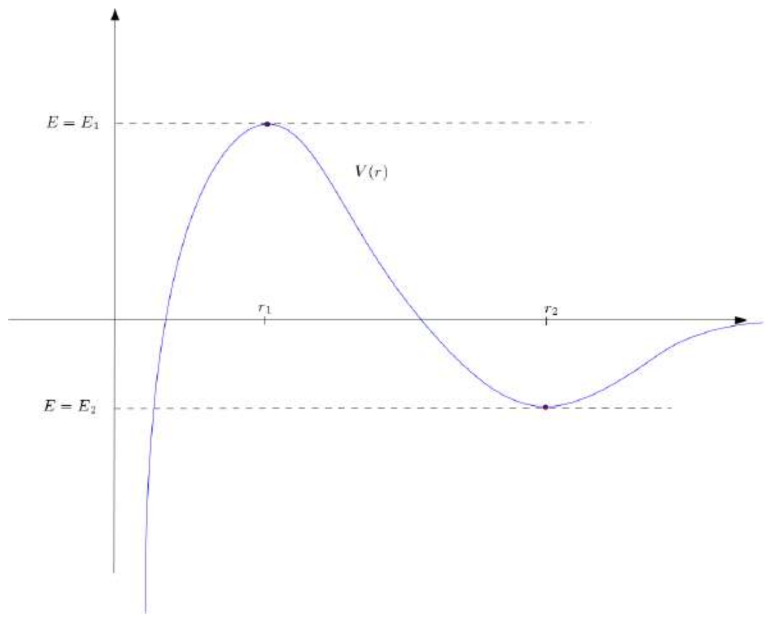


Figure 15.12:

(15.43) we see that

$$\frac{d\phi}{dr} = \frac{d\phi/ds}{dr/ds} = \frac{l/r^2}{\pm\sqrt{2}[\mathrm{E} - V(r(s))]^{1/2}} = \pm\frac{1}{\sqrt{\frac{r^4}{l^2}(2\mathrm{E} - 2V(r))}}, \qquad (15.48)$$

where

$$V(r) = \frac{l^2}{2r^2} - \frac{\mathrm{M}}{r} - \frac{\mathrm{M}l^2}{r^3}.$$

The sign of the radical is chosen as the sign of dr/ds. The expression inside the radical is a polynomial of degree four:

$$\begin{aligned} p(r) &= \frac{2\mathrm{E}}{l^2}r^4 + \frac{2\mathrm{M}}{l^2}r^3 - r^2 + 2\mathrm{M}r \\ &= \frac{2\mathrm{E}}{l^2}r^4 + \frac{\mathrm{R}_s}{l^2}r^3 - r^2 + \mathrm{R}_s r. \end{aligned}$$

Since r_1, r_2 are turning points, then they are roots of $p(r)$. We denote by ε the third nonzero root of $p(r)$. We may write $p(t) = \frac{2\mathrm{E}}{l^2}r(r-r_1)(r-r_2)(r-\varepsilon)$, where $\varepsilon+r_1+r_2 = -\mathrm{R}_s/2\mathrm{E}$. We denote by ϕ_1, ϕ_2, ϕ_3 the corresponding angles

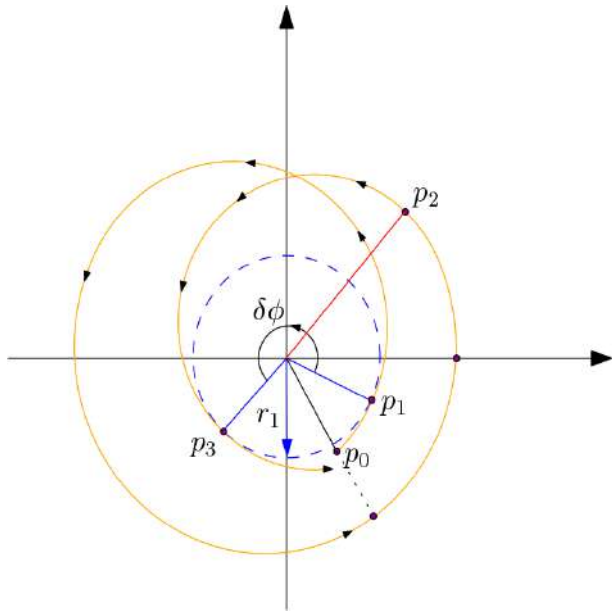


Figure 15.13: Precession of the orbit

of the points p_1, p_2 and p_3. We want to calculate $\Delta\phi = \phi_3 - \phi_1$. For this, we write $\Delta\phi = (\phi_2 - \phi_1) + (\phi_3 - \phi_2)$ so that

$$\Delta\phi = \int_{p_1}^{p_2} \frac{dr}{d\phi} d\phi + \int_{p_2}^{p_3} \frac{dr}{d\phi} d\phi$$

$$= \int_{r_1}^{r_2} \frac{+1}{\sqrt{p(r)}} dr + \int_{r_2}^{r1} \frac{-1}{\sqrt{p(r)}} dr = 2 \int_{r_1}^{r_2} \frac{1}{\sqrt{p(r)}} dr.$$

To estimate the later integral we write the radical expression inside as follows:

$$\frac{1}{\sqrt{p(r)}} = \frac{1}{\sqrt{\frac{-2\mathrm{E}}{l^2} r^2 (r_1 - r)(r - r_2)(1 - \frac{\varepsilon}{r})}} = \frac{1}{\sqrt{1 - \mathrm{e}^2}} \frac{1}{r\sqrt{(r_1 - r)(r - r_2)(1 - \frac{\varepsilon}{r})}}$$

$$= \frac{1}{\sqrt{1 - \mathrm{e}^2}} \frac{1}{r\sqrt{(r_1 - r)(r - r_2)}} (1 - \frac{\varepsilon}{r})^{-1/2},$$

where we have used that $\mathrm{E} = (\mathrm{e}^2 - 1)/2$ (notice that for P one must have e <1). Using the binomial theorem we may write:

$$(1 - \varepsilon/r)^{-1/2} \simeq 1 + \varepsilon/2r + h,$$

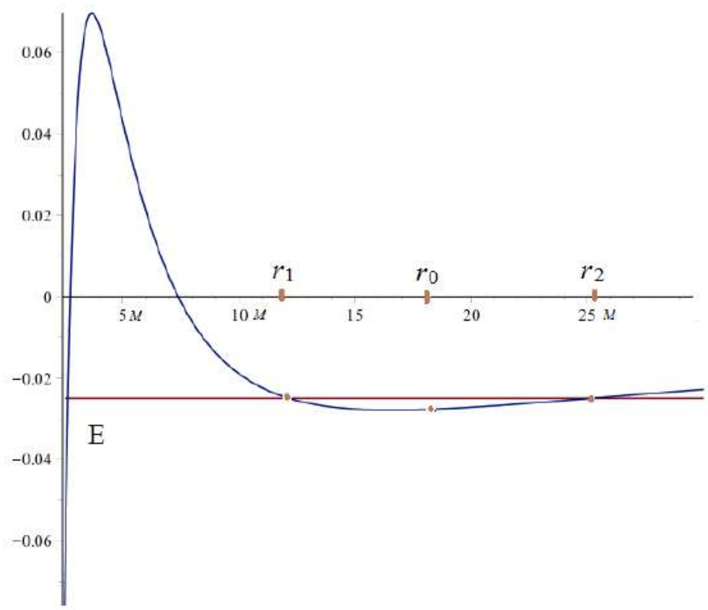


Figure 15.14:

where the error h can be estimated as:

$$|h| < \frac{3}{8}(1-\varepsilon/r)^{-5/2}(\frac{\varepsilon}{r})^2 \leq \frac{3}{8}(1-\varepsilon/r_2)^{-5/2}(\frac{\varepsilon}{r_1})^2. \qquad (15.49)$$

Henceforth,

$$\frac{1}{\sqrt{p(r)}} = \frac{1}{\sqrt{1-\mathrm{e}^2}}\left(\frac{1+h}{r\sqrt{(r_1-r)(r-r_2)}} + \frac{\varepsilon/2}{r^2\sqrt{(r_1-r)(r-r_2)}}\right).$$

But now the integrals of both terms inside the parenthesis can be evaluated in closed form (One can see this using *Maple*, for instance) as:

$$\int_{r_1}^{r_2} \frac{1+h}{r\sqrt{(r_1-r)(r-r_2)}}dr = \frac{\pi(1+h)}{\sqrt{r_1 r_2}}$$

$$\int_{r_1}^{r_2} \frac{\varepsilon/2}{r^2\sqrt{(r_1-r)(r-r_2)}}dr = \frac{\pi\varepsilon}{4\sqrt{r_1 r_2}}\frac{(r_1+r_2)}{r_1 r_2}.$$

From this one obtains:

$$\Delta\phi = 2\int_{r_1}^{r_2} \frac{1}{\sqrt{p(r)}}dr = \frac{2\pi}{\sqrt{1-\mathrm{e}^2}}\left(\frac{(1+h)}{\sqrt{r_1 r_2}} + \frac{\varepsilon}{4\sqrt{r_1 r_2}}\frac{(r_1+r_2)}{r_1 r_2}\right). \qquad (15.50)$$

Finally, we want to write e in terms of r_1 and r_2. We use the fact that $p(r_1) = p(r_2) = 0$:

$$2\mathrm{E}r_1^4 - l^2(r_1^2 + \mathrm{R}_s r_1) - \mathrm{R}_s r_1^3 = 0.$$
$$2\mathrm{E}r_2^4 - l^2(r_2^2 + \mathrm{R}_s r_2) - \mathrm{R}_s r_2^3 = 0.$$

Solving for E and l in terms of r_1 and r_2 one gets:

$$2\mathrm{E} = \frac{-r_1 r_2 \mathrm{R}_S + \mathrm{R}_S^2(r_1+r_2)}{r_1 r_2(r_1+r_2+\mathrm{R}_S) - (r_1+r_2)^2\mathrm{R}_S}$$

(we do not write the value of l since will not need it). To simplify matters let's write $d = r_1 r_2/(r_1+r_2)$. In terms of r_1, r_2 and d, equation (15.50) becomes:

$$\Delta\phi = 2\pi\left[\frac{1}{\sqrt{1-\mathrm{R}_S/d}}\left(1 + \frac{1}{4}\frac{R_S/d}{(1-R_S/d)}\right) + \frac{h}{\sqrt{1-\mathrm{R}_S/d}}\right]. \qquad (15.51)$$

For the planet Mercury, $r_1 = 46\times 10^6$ km, $r_2 = 69.8\times 10^6$ km, so that $d = 27.7\times 10^6$ km. On the other hand, the Schwarzschild radius of the Sun is approximately 2.95 km. Combining this with (15.49) one obtains the estimate $\left|\frac{h}{\sqrt{1-\mathrm{R}_S/d}}\right| < 10^{-15}$. Substituting r_1, r_2, d in (15.51) one obtain that $\Delta\phi \simeq 2\pi + 5.03\times 10^{-7}$ so that the precession of its perihelion is approximately $\delta\phi \simeq 5.03\times 10^{-7}$. In one century Mercury orbits the Sun 415.2 times that account for a total displacement of its perihelion of $415.2\times\delta\phi = 2088.456\times 10^{-7}$ radians, or equivalently, for a total displacement of

$$\frac{360\times 3600}{2\pi}\times 2088.456\times 10^{-7} = 43.084 \text{ seconds of arc},$$

as has been observed!

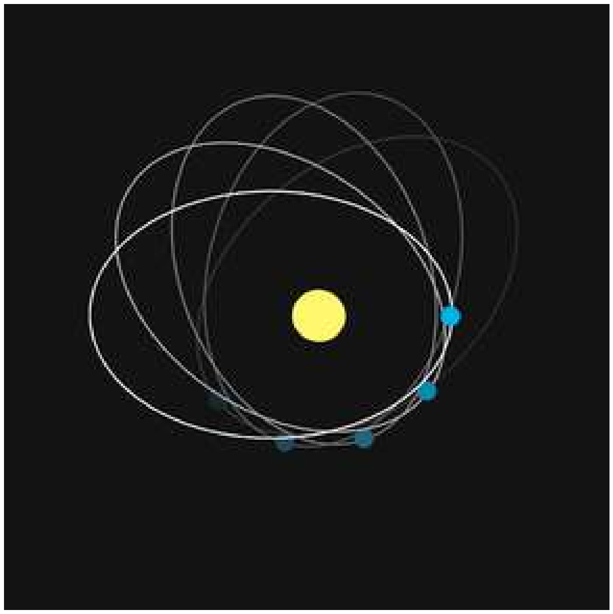


Figure 15.15: Precession of Mercury's perihelion

15.14 Motion of Massless Particles

In this section we will derive the equation of motion for particles that move in Schwarzschild space-time $(\mathbb{R}^4, g_s)$, where g_s is the metric (15.21). We proceed the same way as in Section 15.10: If $\sigma(s)$ is a null geodesic, we define as before $\mathrm{e} = -\langle\sigma'(s), \partial_t\rangle$, and $l = \langle\sigma'(s), \partial_\phi\rangle$. Writing $\sigma(s) = (t(s), r(s), \theta(s), \phi(s))$, we see that $\mathrm{e} = t'(s)(1 - 2\mathrm{M}/r(s))$ and $l = r^2(s)\sin^2\theta(s)\phi'(s)$ are constant functions. The same argument as in Section 15.10 shows that *the spatial trajectory of a zero mass particle P that moves along σ is contained in a plane in $\mathbb{R}^3$*. Without loss of generality one can assume this is the equatorial plane $\theta = \pi/2$. Hence, as, before, $l = r^2(s)\phi'(s)$. however, Equation 15.39 changes. It becomes:

$$-(1 - \frac{2\mathrm{M}}{r(s)})t'(s)^2 + (1 - \frac{2\mathrm{M}}{r(s)})^{-1}r'(s)^2 + r^2\phi'(s)^2 = 0.$$

Substituting $t'(s) = \mathrm{e}/(1 - 2\mathrm{M}/r(s))$ and $\phi'(s) = l/r^2(s)$ one obtains:

$$\frac{\mathrm{e}^2}{l^2} = \frac{1}{l^2}\left(\frac{dr}{ds}\right)^2 + \frac{1}{r(s)^2}\left(1 - \frac{2\mathrm{M}}{r(s)}\right),$$

or equivalently

$$\frac{1}{\mathrm{b}^2} = \frac{1}{l^2}\left(\frac{dr}{ds}\right)^2 + W(r), \tag{15.52}$$

where $\mathrm{b} = l/\mathrm{e}$, and $W(r) = (1 - 2\mathrm{M}/r)/r^2$ is the *effective potential* for particles with zero mass.

Let's see what b means in physical terms. We imagine a photon P that is heading towards M, coming from very from far away, so that its initial values of r satisfy that $r_0 >> 0$, and $dr/ds|_{s=0} < 0$. We assume P moves in the equatorial plane, and that its starts its trajectory at a distance d form the x-axis

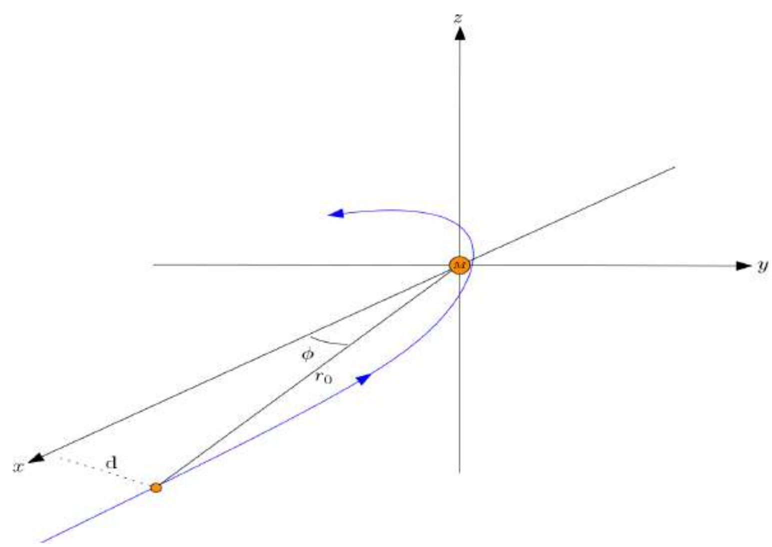


Figure 15.16: A photon coming from infinity

We claim $\mathrm{d} \simeq \mathrm{b} = l/\mathrm{e}$. Now,

$$\frac{l}{\mathrm{e}} = \frac{r^2(s)d\phi/ds}{dt/ds(1 - 2\mathrm{M}/r(s))} = \frac{r^2(s)d\phi/dt}{(1 - 2\mathrm{M}/r(s))}.$$

But when $r(s) >> 0$, and $(1 - 2\mathrm{M}/r) \simeq 1$. Hence:

$$\frac{l}{\mathrm{e}} \simeq r^2(s)\frac{d\phi}{dt}. \tag{15.53}$$

On the other hand, $r(s) >> 0$ implies that the corresponding angle $\phi(s)$ is small. Hence, $\phi(s) \simeq \sin\phi(s)$, and therefore $\mathrm{d} \simeq r(s)\phi(s)$. Thus, $\phi(s) \simeq$

$\mathrm{d}/r(s)$, and $d\phi/dr \simeq -\mathrm{d}/r^2(s)dr/ds$. Moreover, (15.52) implies that for $r >> 0$, $W(r) \simeq 0$ and therefore $(dr/ds) \simeq l^2/\mathrm{b}^2 = \pm\mathrm{e}$. Since P moves inward we have $dr/ds \simeq -\mathrm{e}$, and

$$dr/dt = \frac{dr/ds}{dt/ds} = -\mathrm{e}\frac{(1 - 2\mathrm{M}/R)}{\mathrm{e}} \simeq -1.$$

From this one gets:

$$\frac{d\phi}{dt} = \frac{d\phi}{dr}\frac{dr}{dt} \simeq \frac{-\mathrm{d}}{r^2(s)}\frac{dr}{dt} \simeq \frac{-\mathrm{d}}{r^2(s)}(-1) = \frac{\mathrm{d}}{r^2(s)}.$$

Substituting this value of $d\phi/dt$ in (15.53) one obtains: $l/\mathrm{e} \simeq r^2(s)(\mathrm{d}/r^2(s)) = \mathrm{d}$. The constant b is called the *impact parameter* of P.

We observed in the previous section that $V(r)$ can have two critical points, a maximum and a minimum. The effective potential $W(r)$, on the other hand, has only one critical point: a maximum at $r = 3\mathrm{M}$, where $W(\mathrm{M}) = 1/(27\mathrm{M}^2)$.

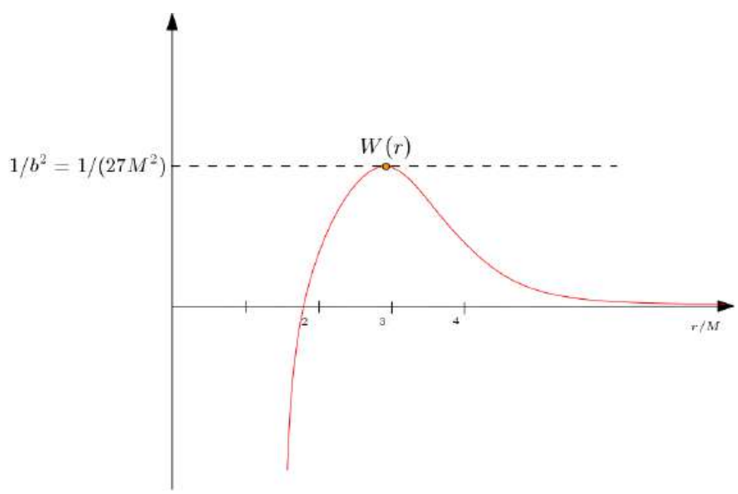


Figure 15.17: Effective potential for a zero mass particle

The analysis of possible trajectories is very similar to the one we did in the last section. The following graphs show three possible scenarios:

1. Circular orbit: $\mathrm{b}^2 = 27\mathrm{M}^2$:

   A scattering orbit: $1/\mathrm{b}^2 < 1/(27\mathrm{M}^2)$, and $dr/ds|_{s=0} < 0$:

   Absorption of the photon at M : $1/\mathrm{b}^2 > 1/(27\mathrm{M}^2)$, and $dr/ds|_{s=0} < 0$:

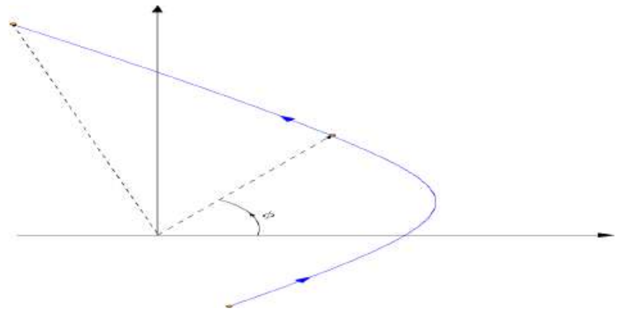


Figure 15.18: scattering orbit

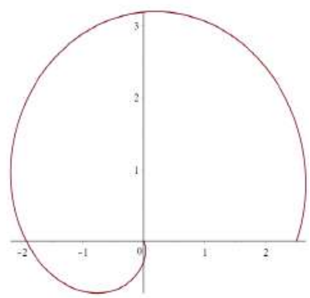


Figure 15.19: Absorption of the photon at M :

15.15 Gravitational Bending of Light

One of the first tests of the theory of General Relativity was the deflection of light rays. Positions of star images within the field near the Sun were used to test Albert Einstein's prediction of the bending of light around the Sun. The first observation of light deflection was performed by Arthur Eddington and his collaborators during the total solar eclipse of May 29, 1919, when stars near the Sun could be observed. The result made the front page of most major newspapers, and made Einstein instantly world-famous. However, many argued that the results had been plagued by systematic errors, although reanalysis of the data suggests that Eddington's analysis was accurate [42].

To see how light can be deflected, we start by analyzing the trajectory of a photon coming from very far away, which has a scattering orbit as in example (2), above. That is, we assume $1/\mathrm{b}^2 < 1/(27\mathrm{M}^2)$, and $dr/ds|_{s=0} < 0$. We rewrite equation (15.52) as follows:

$$\frac{dr}{ds} = \pm l \left(\frac{1}{\mathrm{b}^2} - W(r) \right)^{1/2}. \tag{15.54}$$

Since P starts with $dr/ds|_{s=0} < 0$ we must choose the negative sign for the radical in (15.54), so that r decreases. When r reaches the turning point $r = r_2$, r start increasing, hence, from here on, one must choose the positive sign in (15.54).

On the other hand, since $l = r^2 d\phi/ds$ one has:

$$\frac{d\phi}{dr} = \frac{d\phi/ds}{dr/ds} = \frac{l/r^2}{\pm l\,(1/\mathrm{b}^2 - W(r))^{1/2}} = \pm\frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2}, \tag{15.55}$$

where the sign of the radical is chosen as explained above. The deflection angle is defined to be $\delta_\phi = \lim \Delta\phi - \pi$, where $\Delta\phi = \phi_2 - \phi_1$ is the change in angle when the photon travels from position 1 to 2, and $\lim \Delta\phi$ is the limit of the differences $\Delta\phi$ when $r_0, r_3 \to \infty$:

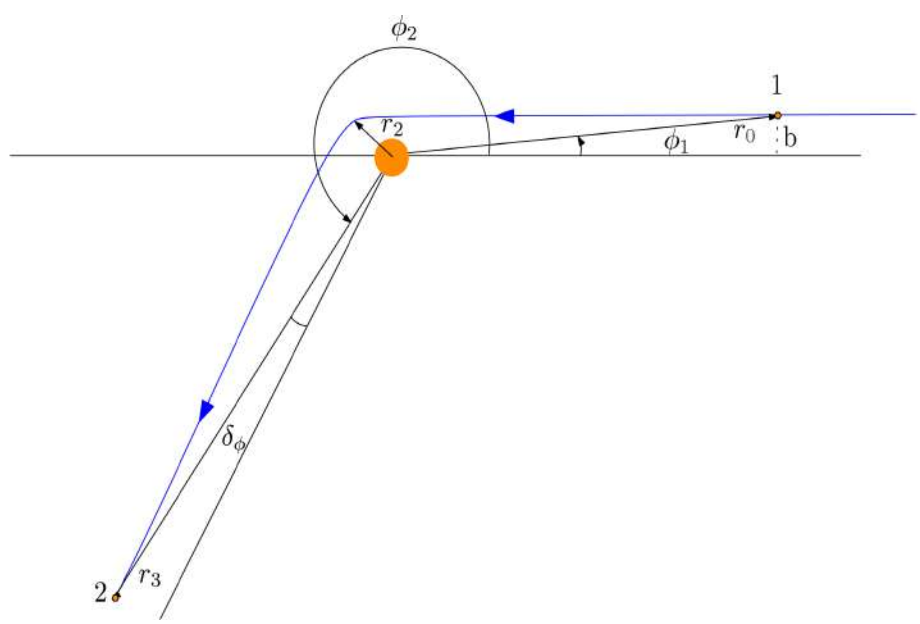


Figure 15.20: Bending of light

On he other hand,

$$\begin{aligned}\Delta\phi = \int_{r_0}^{r_3} \frac{d\phi}{dr} &= \int_{r_0}^{r_2} \frac{-1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr + \int_{r_2}^{r_3} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr \\ &= \int_{r_2}^{r_0} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr + \int_{r_2}^{r_3} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr.\end{aligned}$$

Taking limits when $r_0, r_3 \to \infty$ we obtain:

$$\begin{aligned}\lim \Delta\phi &= \int_{r_2}^{\infty} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr + \int_{r_2}^{\infty} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr \\ &= 2\int_{r_2}^{\infty} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr.\end{aligned}$$

We change the variable of integration by letting $r = \mathrm{b}/u$, so that $dr = -\mathrm{b}/u^2 du$, and

$$\begin{aligned}\int_{r_2}^{\infty} \frac{1}{r^2}\left(\frac{1}{\mathrm{b}^2} - W(r)\right)^{-1/2} dr &= \int_{u_2}^{0} \frac{u^2}{\mathrm{b}^2}\left[\frac{1}{\mathrm{b}^2} - \frac{u^2}{\mathrm{b}^2}(1 - \frac{2\mathrm{M}u}{\mathrm{b}})\right]^{-1/2} \frac{-\mathrm{b}}{u^2} du \\ &= \int_{0}^{u_2} \left[1 - u^2(1 - \frac{2\mathrm{M}u}{\mathrm{b}})\right]^{-1/2} du,\end{aligned}$$

where $u_2 = \mathrm{b}/r_2$ is a root of the expression inside the brackets, since r_2 is by definition a root of $1/\mathrm{b}^2 - W(r)$.

By factoring $(1 - \frac{2\mathrm{M}u}{\mathrm{b}})$, the expression in brackets can be written as:

$$\left[1 - u^2(1 - \frac{2\mathrm{M}u}{\mathrm{b}})\right]^{-1/2} = \left(1 - \frac{2\mathrm{M}u}{\mathrm{b}}\right)^{-1/2}\left[\left(1 - \frac{2\mathrm{M}u}{\mathrm{b}}\right)^{-1} - u^2\right]^{-1/2}.$$

If we take the impact parameter b about the size (slightly bigger) of R, the radius of the celestial body responsible for the gravitational bending, then quantity 2M/b is very small if M is not too big. For instance, for the Sun this number is of the order of 10^{-6}. Henceforth, one can approximate

$$\left(1 - \frac{2\mathrm{M}u}{\mathrm{b}}\right)^{-1/2} \simeq 1 + \frac{\mathrm{M}u}{\mathrm{b}},$$

and

$$\left(1 - \frac{2\mathrm{M}u}{\mathrm{b}}\right)^{-1} \simeq 1 + \frac{2\mathrm{M}u}{\mathrm{b}},$$

and therefore

$$\lim \Delta\phi \simeq 2 \int_0^{u_2} \frac{1 + \frac{\mathrm{M}u}{\mathrm{b}}}{\left[1 + \frac{2\mathrm{M}u}{\mathrm{b}} - u^2\right]^{1/2}} du.$$

This can be integrated as an exact function. In *Maple* one can verify that

$$2 \int_0^{u_2} \frac{1 + \frac{\mathrm{M}u}{\mathrm{b}}}{\left[1 + \frac{2\mathrm{M}u}{\mathrm{b}} - u^2\right]^{1/2}} du = \pi + \frac{4\mathrm{M}}{\mathrm{b}}$$

from which we obtain $\delta_\phi = \lim \Delta\phi - \pi \simeq 4\mathrm{M/b}$. Since $\mathrm{b} \simeq R$, the radius of the Sun, this gives a value of $\delta_\phi \simeq 1.7''$ seconds of arc, a prediction that corroborates Einstein's theory.

SIMULATION OF ORBITS IN SCHWARZSCHILD GEOMETRY

The routine *Orbit* takes parameters:

1. r0 = the radial initial position of a particle of unitary mass.
2. By λ we denote L/M, where L is the angular momentum and M the mass of the central body or the black hole.
3. By E we denote the Newtonian Energy of motion.
4. s = +1 or -1 is taken if for the initial radial velocity $\frac{dr}{ds}(0) > 0$ or $\frac{dr}{ds}(0) < 0$, respectively.
5. a = the top angle of variation for ϕ, i.e., $0 \le \varphi \le a$.

The routine *Orbit*(r_0, λ, E, s, a) gives the graph of $r = r(\varphi)$, where $r(s)$, $\varphi(s)$ are solution of the system of equations:

$$E = \frac{1}{2}\left(\frac{dr}{ds}\right)^2 + \left(\frac{L^2}{2\,r(s)^2} - \frac{M}{r(s)} - \frac{ML^2}{r(s)^3}\right)$$

$$r(s)^2\,\phi'(s) = L,$$

and animates the orbit of a unitary mass particle moving under such conditions.

The function $W(r) = \frac{L^2}{2\,r(s)^2} - \frac{M}{r(s)} - \frac{ML^2}{r(s)^3}$ is called the *effective potential*.

```
> with(plots):
> orbit:=proc(r0,lambda,E,s,a)
> local r,V,v0,z,sol,T,L; L:=lambda;
> V:=(L,z)->L^2/(2*z^2)-1/z-L^2/z^3;
> v0:=-s*sqrt(2)*L*sqrt(E-V(L,r0));
> sol:=dsolve({diff(u(phi),phi,phi)=-u(phi)+1+3*u(phi)^2/L^2,u(0)=
  L^2/r0, D(u)(0)=v0},numeric,output=listprocedure);
> r:=L^2/eval(u(phi),sol);
> T:=[plot(r(phi),phi=0..a),animate(plot, [[r(phi)*cos(phi), r(phi)
  *sin(phi), phi = 0 .. x]], x=0..a, frames=100)];
> print(T[1]);print(T[2]);
> end proc:
```

A Program for computing the Chistoffel symbols and the Ricci tensor corresponding to a metric g.

1) The routine "christoffel" receives a matrix G whose entries are functions in the coordinates x[a], corresponding to certain metric g, and three integers c, a, b in the range 0,...,3, and returns the Chistoffel symbol $\Gamma^{c}_{a,b}$ corresponding to g.

```
> with(linalg):
```

```
> christoffel:=proc(G,c,a,b) local T,GG,r;
> GG:=evalm(inverse(G));
> T:=0;
> for r from 0 to 3 do
> T:=T+1/2*GG[c+1,r+1]*(diff(G[r+1,b+1],x[a])+diff(G[r+1,a+1],x[b])
  -diff(G[a+1,b+1],x[r]));
> end do;
> return(T);
> end proc:
>
```

Example: Let us compute $\Gamma^{2}_{1,2}$ for the metric given by

the matrix $A = \begin{bmatrix} -e^{2\alpha(x_1)} & 0 & 0 & 0 \\ 0 & e^{2\beta(x_1)} & 0 & 0 \\ 0 & 0 & x_1^2 & 0 \\ 0 & 0 & 0 & x_1^2 \sin(x_2)^2 \end{bmatrix}$, *one obtains* :

```
> A := matrix( [ [-exp(2*alpha(x[1])), 0, 0, 0], [0, exp(2*beta(x[1])), 0, 0], [0, 0, x[1]^2,
   0], [0, 0, 0, x[1]^2*sin(x[2])^2] ]) :
```

```
> c=2, a=1, b=2
```

$$c = 2, a = 1, b = 2 \qquad (1)$$

```
> christoffel(A,2,1,2);
```

$$\frac{1}{x_1} \qquad (2)$$

Figure 15.21: Box 1

> *That is*, $\Gamma^{2}_{1,2} = \frac{1}{x_1}$

$$\textit{That is, } \Gamma^{2}_{1,2} = \frac{1}{x_1} \quad (3)$$

>

2) The routine "LChristoffel" receives a matrix G whose entries are functions in the coordinates x[a], corresponding to certain metric g, and returns all the nonzero Chistoffel symbol

$\Gamma^{c}_{a,b}$ corresponding to g $\left(\text{recall } \Gamma^{c}_{b,a} = \Gamma^{c}_{a,b}\right)$. Each symbol $\Gamma^{c}_{a,b}$ is written as $\Gamma_{c,a,b}$.

```
> LChristoffel:=proc(G) local a,b,c;
> for c from 0 to 3 do
> for a from 0 to 3 do
> for b from a to 3 do
> if christoffel(G,a,b,c)<>0 then print(GAMMA [c,a,b]= christoffel
  (G,c,a,b)); end if;
> end do; end do; end do;
> end proc:
```

>

For the example of the metric given above by *A* one gets:

```
> LChristoffel(A);
```

$$\Gamma_{0,0,1} = \frac{d}{dx_1}\alpha(x_1)$$

$$\Gamma_{1,0,0} = \frac{\left(\frac{d}{dx_1}\alpha(x_1)\right)e^{2\alpha(x_1)}}{e^{2\beta(x_1)}}$$

$$\Gamma_{1,1,1} = \frac{d}{dx_1}\beta(x_1)$$

$$\Gamma_{1,2,2} = -\frac{x_1}{e^{2\beta(x_1)}}$$

$$\Gamma_{1,3,3} = -\frac{x_1 \sin(x_2)^2}{e^{2\beta(x_1)}}$$

$$\Gamma_{2,1,2} = \frac{1}{x_1}$$

$$\Gamma_{2,3,3} = -\sin(x_2)\cos(x_2)$$

$$\Gamma_{3,1,3} = \frac{1}{x_1} \quad (4)$$

Figure 15.22: Box 2

The routine "ric" receives a matrix A, corresponding to a certain metric g, and two integers in the range 0,..3, and returns the component $Ric_{a,b}$ of the Ricci tensor associated to g.

The program "Ricci" produces a list of all the nonzero entries of this tensor.

```
> ric:=proc(G,a,b) local u,r,R,T;
> T:=0;
> for u from 0 to 3 do
> T:=T+diff(christoffel(G,u,a,b),x[u])-diff(christoffel(G,u,a,u),x
  [b]);
> end do;
> R:=T;
> for u from 0 to 3 do
> for r from 0 to 3 do
> R:=R+christoffel(G,u,a,b)*christoffel(G,r,u,r)-christoffel(G,u,a,
  r)*christoffel(G,r,b,u);
> end do; end do;
> return(R); end proc:
```

```
> Ricci:=proc(G)
> local a,b;

> for a from 0 to 3 do
> for b from a to 3 do
> if ric(G,a,b)<>0 then print(Ricc [a,b]= ric(G,a,b)); end if;
> end do; end do; end proc:
```

For instance, if g is given by A as above one gets:

```
>
> Ricci(A);
```

$$Ricc_{0,0} = -\frac{\left(\frac{d}{dx_1}\alpha(x_1)\right)e^{2\alpha(x_1)}\left(\frac{d}{dx_1}\beta(x_1)\right)}{e^{2\beta(x_1)}} + \frac{\left(\frac{d^2}{dx_1^2}\alpha(x_1)\right)e^{2\alpha(x_1)}}{e^{2\beta(x_1)}} + \frac{\left(\frac{d}{dx_1}\alpha(x_1)\right)^2 e^{2\alpha(x_1)}}{e^{2\beta(x_1)}} + \frac{2\left(\frac{d}{dx_1}\alpha(x_1)\right)e^{2\alpha(x_1)}}{e^{2\beta(x_1)}x_1}$$

$$Ricc_{1,1} = -\left(\frac{d^2}{dx_1^2}\alpha(x_1)\right) - \left(\frac{d}{dx_1}\alpha(x_1)\right)^2 + \left(\frac{d}{dx_1}\beta(x_1)\right)\left(\frac{d}{dx_1}\alpha(x_1)\right) + \frac{2\left(\frac{d}{dx_1}\beta(x_1)\right)}{x_1}$$

$$Ricc_{2,2} = \frac{x_1\left(\frac{d}{dx_1}\beta(x_1)\right)}{e^{2\beta(x_1)}} - \frac{1}{e^{2\beta(x_1)}} + 1 - \frac{x_1\left(\frac{d}{dx_1}\alpha(x_1)\right)}{e^{2\beta(x_1)}}$$

$$Ricc_{3,3} = \frac{x_1\sin(x_2)^2\left(\frac{d}{dx_1}\beta(x_1)\right)}{e^{2\beta(x_1)}} - \frac{\sin(x_2)^2}{e^{2\beta(x_1)}} + \sin(x_2)^2 - \frac{x_1\sin(x_2)^2\left(\frac{d}{dx_1}\alpha(x_1)\right)}{e^{2\beta(x_1)}} \qquad (5)$$

```
>
```

Figure 15.23: Box 3

Chapter 16

Black Holes

16.1 Introduction

In science fiction movies black holes are represented by massive dark stars surrounded by a whirling bright reddish halo, sucking in anything that comes near them. According to a popular myth, nothing can escape their gravitational field, not even light! But this is a misconception: In the first place, a black hole is not a celestial body, but really an empty region of space-time! And strictly speaking, since photons do not have any mass, they are not subjected to the action of forces in the Newtonian sense, regardless of their intensity. As we shall see, a black hole is not more capable of attracting anything into it than any other celestial body of the same mass. Objects lying beyond the *event horizon* of a black hole are not necessarily destined to be swallowed by the monster. However, it is true that any particle, even without mass, that goes inside its event horizon is destined to disappear at the central *singularity*, a point where space and time cease to exist!

The interpretation of the whole Schwarzschild's solution as corresponding, not to the exterior region outside a star, but as the geometry of an empty region of space-time appeared in the work of Finkelstein, in 1958. For a long time, this solution was considered just a mathematical curiosity until some physicist decided to take the mathematical formalism more "seriously". Eminent physicists like Robert Oppenheimer and Hartland Snyder had already envisioned the possibility that some massive stars could undergo a total gravitational collapse. The discovery of neutron stars in the early 1960's contributed to spark interest in this phenomena, and some started to suspect that black holes may be born when some heavy stars collapse, at the end of their life cycle. Then, by absorbing other stars, they could grow, and black holes of millions of solar masses may form.

Nowadays there is a general consensus that such bizarre objects exist everywhere in the universe. Super massive black holes very likely inhabit the centers of most galaxies. It is believed, for instance, that the radio wave source at the core of our Milky Way galaxy, known as *Sagittarius* A^*, contains a giant black hole of about 4.3 million solar masses.

Despite its invisible interior, the presence of a black hole can be inferred indirectly: Matter falling around it forms an external *accretion disk,* heated by friction, creating some of the brightest objects known in the universe [43].

16.2 Eddigton-Finkelstein Coordinates

In Section 15.4 we defined the metric

$$g = -\left(1 - \frac{2\mathrm{M}}{r}\right) dt \otimes dt + \left(1 - \frac{2\mathrm{M}}{r}\right)^{-1} dr \otimes dr + r^2 (d\theta \otimes d\theta + \sin^2\theta\ d\phi \otimes d\phi), \tag{16.1}$$

in spherical coordinates (t, r, θ, ϕ) for the 4-manifold $\mathbb{R} \times (\mathbb{R}^3 - B_a)$, as modeling the region outside a perfectly symmetric round star. But if we imagine that we remove the star then one could think of g as the geometry of a larger manifold:

$$M = (\mathbb{R} \times W_1) \cup (\mathbb{R} \times W_2), \tag{16.2}$$

where W_1 is the *inner* region $0 < r < 2\mathrm{M}$, (usually inside the star) and W_2 is the *outer region,* defined by $2\mathrm{M} < r$. We are forced to exclude $r = 2\mathrm{M}$ since (16.1) is not defined there.

In W_1 the vector field ∂_t is timelike, where t represents time measured by an observer at infinity, and r the usual radial coordinate. However, the field ∂_t is no longer timelike inside W_2. In this region the roles of the coordinates t and r are reversed, and ∂_t becomes spacelike while ∂_r becomes timelike. This can be readily seen fixing the coordinates θ and ϕ, and drawing the light cones corresponding to M. For this, one needs to compute the null geodesics, that would be given by $\gamma(s) = (t(s), r(s), \theta_0, \phi_0)$, where

$$-\left(1 - \frac{2\mathrm{M}}{r(s)}\right) t'(s)^2 + \left(1 - \frac{2\mathrm{M}}{r(s)}\right)^{-1} r'(s)^2 = 0,$$

or equivalently, $t'(s) = \pm r'(s)/(1 - \frac{2M}{r(s)})$. Integrating both sides one obtains:

$$\begin{aligned} t(s) &= r(s) + 2M \log |r(s) - 2M| + \text{constant} \\ &= r(s) + 2M \log \left| \frac{r(s)}{2M} - 1 \right| + k, \end{aligned}$$

after absorbing the term $2M \ln(2M)$ into the constant. It is also clear from

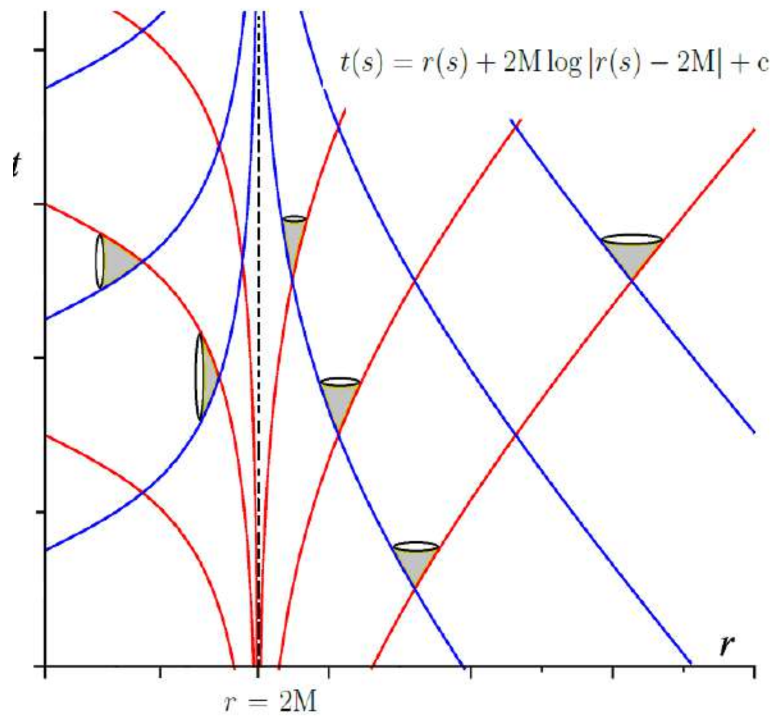


Figure 16.1: Schwarzschild's coordinates

(16.1) that in W_1 the metric is not static. In fact, interchanging t and r the Schwarzschild metric looks like

$$g = \left(1 - \frac{2M}{t}\right)^{-1} dt \otimes dt - \left(1 - \frac{2M}{t}\right) dr \otimes dr + t^2 dA,$$

To better understand the geometry of M, one can "change coordinates" in such a way that g becomes defined at $r = 2M$, showing that this is not an *essential singularity*, but a just a *coordinate singularity*, as it is usually called. In more rigorous terminology, one shows how to construct a larger pseudo-Riemannian 4-manifold $M' = \mathbb{R} \times \mathbb{R}^3 - \{0\}$ in which (M, g) embeds isometrically. In fact, we define a new variable

$$v = t + r \ln \left| \frac{r}{2\mathrm{M}} - 1 \right|, \text{ for } r \neq 2\mathrm{M}. \tag{16.3}$$

A straightforward computation shows that

$$dt = dv + \left(-1 - \frac{2\mathrm{M}}{r - 2\mathrm{M}} \right) dr = dv + \frac{r}{2\mathrm{M} - r} dr. \tag{16.4}$$

Replacing this equation in (16.1) one obtains:

$$g' = -\left(1 - \frac{2\mathrm{M}}{r}\right) dv \otimes dv + dv \otimes dr + dr \otimes dv + r^2 (d\theta \otimes d\theta + \sin^2 \theta \ d\phi \otimes d\phi).$$

Now, we define coordinates (v, r, θ, ϕ) in M', where, $-\infty < v < \infty$, $r > 0$, and, as usual, $0 < \theta < \pi$, $0 < \phi < 2\pi$. Clearly the map given in coordinates by

$$h(t, r, \theta, \phi) = (v(t, r), r, \theta, \phi),$$

embeds (M, g) isometrically as an open subset of (M', g'), where the image of h is precisely the region in M' of all points with $r > 2\mathrm{M}$.

In order to understand M' one can consider the radial null curves $\gamma(s) = (v(s), r(s), \theta_0, \phi_0)$. Since $\gamma'(s) = v'(s)\partial_v + r'(s)\partial_r$, being null means that

$$-\left(1 - \frac{2\mathrm{M}}{r}\right) v'(s)^2 + 2v'(s)r'(s) = 0. \tag{16.5}$$

There are two possible solutions to this equation: $v(s) = k_1$, with k_1 an arbitrary constant, and

$$v'(s) = \frac{2r'(s)}{1 - 2\mathrm{M}/r} = 2r'(s) + \frac{4\mathrm{M}r'(s)}{r(s) - 2\mathrm{M}}.$$

Integrating both sides one obtains:

$$v(s) = 2r(s) + 4\mathrm{M} \ln |r(s) - 2\mathrm{M}| + \text{constant}.$$

By absorbing the term $4\mathrm{M} \ln 2\mathrm{M}$ into the constant we may write:

$$v(s) = 2r(s) + 4\mathrm{M} \ln \left| \frac{r(s)}{2\mathrm{M}} - 1 \right| + k_2. \tag{16.6}$$

If we let $\bar{t} = v - r$, then $v = k_1$ is equivalent to $\bar{t} + r = k_1$ which represents the equation of straight line with slope -1. On the other hand, in terms of $\bar{t}$ (16.6) becomes

$$\bar{t}(s) = r(s) + 4\mathrm{M}\ln\left|\frac{r(s)}{2\mathrm{M}} - 1\right| + k_2.$$

From the figure above (where we have fixed θ and ϕ) one can see that the

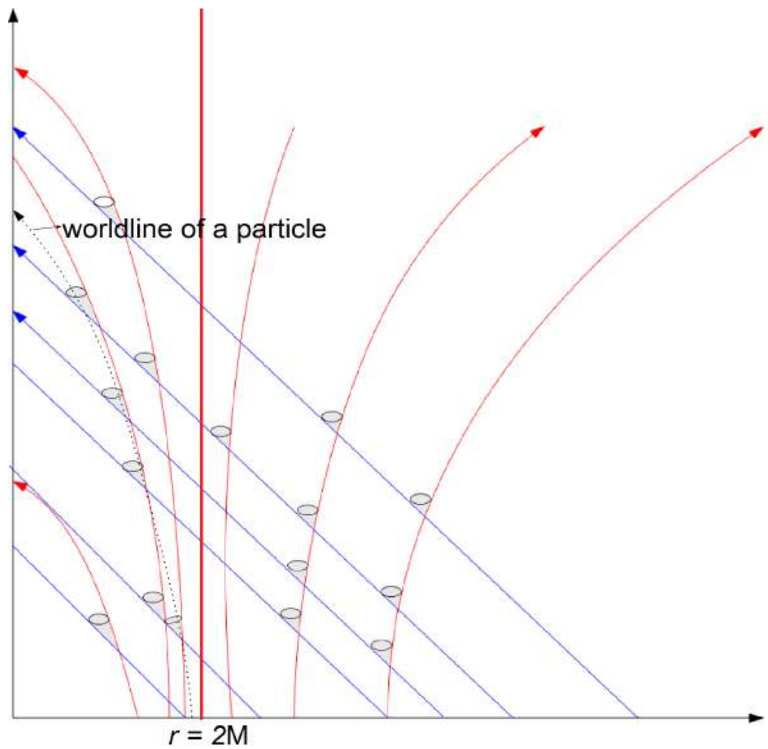


Figure 16.2: Finkelstein's coordinates

worldline of any particle inside the region $r < 2\mathrm{M}$ is destined to end up at the singularity $r = 0$: This is clear for massless particles, since they move along null geodesics, and these end at $r = 0$. And it is also true for massive particles, since any point in their worldline must lie inside a future cone (dotted line in the figure above). The region $r < 2\mathrm{M}$ is called a *Schwarzschild's black hole*, and the hypersurface $r = 2\mathrm{M}$ is called its *event horizon*. We notice from (16.5) that in M' the equation $r(s) = 2\mathrm{M}$ defines a null geodesic.

In (15.47) we showed that a particle P radially approaching the horizon is never seen by an outside observer to ever cross it. Light coming from P becomes dimer and dimer as it gets more and more red shifted. But P indeed crosses the horizon! The proper time of a particle P falling radially from $r = r_0$ to $r = 0$ was calculated in (15.46) as:

$$s_0 = \frac{2r_0^{3/2}}{3\sqrt{2}\mathrm{M}^{1/2}}.$$

If $r_0 < 2\mathrm{M}$, we see that

$$s_0 = \frac{2r_0^{3/2}}{3\sqrt{2}\mathrm{M}^{1/2}} < \frac{2(2\mathrm{M})^{3/2}}{3\sqrt{2}\mathrm{M}^{1/2}} = \frac{4}{3}\mathrm{M},$$

or, measuring s_0 in seconds, $s_0 < 4\mathrm{M}/(3c)$. For a solar mass, $\mathrm{M}_{\mathrm{sun}} \simeq 1480$ kgg, one has $s_0 < 10^{-5}$ seconds. Hence, in such a case the particle P would disappear almost instantly once it crosses the event horizon. For the biggest black holes known in the universe, whose total mass could be something like $10^{10}\mathrm{M}_{\mathrm{sun}}$, the proper time for an astronaut who wanted to make a free-fall into the central singularity would be less than 28 hours. However, there is not much time to enjoy the trip before he is torn apart, and utterly destroyed, by the enormous tidal forces inside! Finkelstein's coordinates can also be

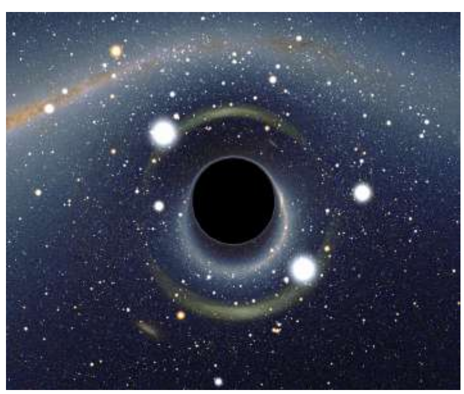


Figure 16.3: Black Hole

used to see how the gravitational collapse of a star occurs. For a star whose mass is at least three to four times $\mathrm{M}_{\mathrm{sun}}$ (the Tolman–Oppenheimer–Volkoff limit) no known mechanism is sufficiently powerful to stop the implosion, and the star will inevitably shrink until its radius becomes less than 2M. Once that happens, the worldline of any of its particles must then inevitably end up at the singularity, and the star will collapse to form a black hole. This corroborates what we proved in Section (15.6), that if the radius of the star is less than 9M/4, a state of hydrostatic equilibrium was not possible.

16.3 Kruskal-Szekeres Coordinates

Now we want to maximally extend the Schwarzschild solution by introducing the Kruskal-Szekeres coordinate system, named after Martin Kruskal and George Szekeres. More precisely, we want to introduce a manifold (M'', g'') together with a maximal embedding $(M', g') \hookrightarrow (M'', g'')$. This means that M'' cannot be isometrically embedded as a proper subset of a larger manifold. For this to happen, M'' has to be *geodesically complete*, which means that any geodesic path in M'' can be extended to arbitrarily large positive or negative values of its affine parameter, unless it runs into a singularity (we will not prove this last property).

In order to construct (M'', g'') we start by choosing coordinates (u, x, θ, ϕ) for $\mathbb{R}^2 \times S^2$, where $-\infty < u < \infty$, $-\infty < x < \infty$, and θ, ϕ are the usual spherical coordinates for the unitary sphere S^2. We define M'' as the open subset $M'' = U \times S^2 \subset \mathbb{R}^2 \times S^2$, where $U = \{(u, x) : u^2 - x^2 < 1\}$: In M'' one defines the metric

$$g'' = \frac{32\mathrm{M}^3}{r(x,u)} e^{-r(x,u)/2\mathrm{M}}(-du \otimes du + dx \otimes dx) + r^2(x,u)(d\theta \otimes d\theta + \sin^2\theta \; d\phi \otimes d\phi),$$

where $r(u, x)$ is a function implicitly defined by the equation

$$\left(\frac{r(x,u)}{2\mathrm{M}} - 1\right) e^{r(x,u)/2\mathrm{M}} = x^2 - u^2.$$

Now, the original Schwarzschild manifold (16.2) is isometrically embedded in (M'', g'') by the function

$$h(t, r, \theta, \phi) = (u(t, r), x(t, r), \theta, \phi)$$

where

$$u(t,r) = \begin{cases} \left(\frac{r}{2\mathrm{M}} - 1\right)^{1/2} e^{r/4\mathrm{M}} \sinh\left(\frac{t}{4\mathrm{M}}\right), \text{ if } r > 2\mathrm{M} \\ \left(\frac{r}{2\mathrm{M}} - 1\right)^{1/2} e^{r/4\mathrm{M}} \cosh\left(\frac{t}{4\mathrm{M}}\right), \text{ if } r < 2\mathrm{M} \end{cases}$$

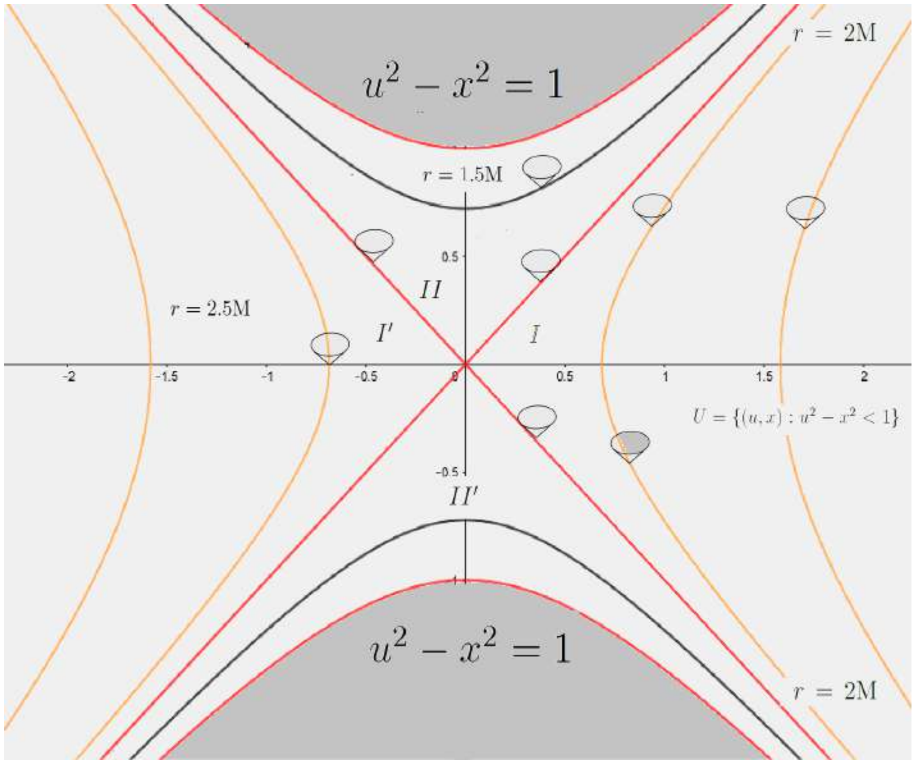


Figure 16.4: Kruskal-Szekeres coordinates

$$x(t,r) = \begin{cases} \left(\frac{r}{2\mathrm{M}} - 1\right)^{1/2} e^{r/4\mathrm{M}} \cosh\left(\frac{t}{4\mathrm{M}}\right), \text{ if } r > 2\mathrm{M} \\ \left(\frac{r}{2\mathrm{M}} - 1\right)^{1/2} e^{r/4\mathrm{M}} \sinh\left(\frac{t}{4\mathrm{M}}\right), \text{ if } r < 2\mathrm{M} \end{cases}.$$

One can easily see that

$$\tanh\left(\frac{t}{4\mathrm{M}}\right) = \frac{u}{x}, \text{ if } r > 2\mathrm{M} \tag{16.7}$$
$$\tanh\left(\frac{t}{4\mathrm{M}}\right) = \frac{x}{u}, \text{ if } r < 2\mathrm{M}.$$

Remark 16.3.1. Fix the coordinates θ and ϕ, as above.

1 The vector ∂_u is a timelike vector in M''; the region $u^2-x^2 \geq 1$ corresponds to the original singularity $r = 0$ in M.

2 The open set $h(W_1)$ is region I, where $-x < u < x$, corresponding to $r > 2\mathrm{M}$, while $h(W_2)$ is region II, defined by $|x| < u < \sqrt{1+x^2}$, corresponding to the black hole, $0 < r < 2\text{M}$ in M. Hence, the image of h is the whole region above the line $u = -x$. This line (as well as $u = x$) corresponds to the event horizon $r = 2\text{M}$. It also corresponds to the image of the lines $t = +\infty$, and $t = -\infty$, respectively, in M.

3 According to (16.7), straight lines crossing the origin are the images of the lines $t =$constant in M.

4 If $\theta = \theta_0$ and $\phi = \phi_0$, null geodesics in M'' are given by

$$\gamma(s) = (u(s), x(s), \theta_0, \phi_0)$$

with $u'(s)^2 = \pm x(s)^2$. These are the equations of lines with slope $= \pm 1$. Hence, light cones in M'' look just the same as a light cone in a Minkowski diagram in special relativity.

5 In Schwarzschild coordinates, $\alpha(s) = (s, r_0, \theta_0, \phi_0)$, with $r_0 > 2\text{M}$, is the world line of a steady observer. Via h this is mapped into the hyperbolas

$$x^2 - u^2 = \left(\frac{r_0}{2\text{M}} - 1\right) e^{r_0/2\text{M}}. \tag{16.8}$$

For $r_0 < 2\text{M}$, however, equation (16.8) does not correspond to the worldline of any steady observer, since it is not timelike. If fact, this corroborates what we had observed before: Inside the black hole (region II) any particle is destined to end up in finite proper time at the central singularity. Vertical lines, on the other hand, always correspond to timelike observers.

6 The event horizons bounding the black hole and white hole interior regions are the lines $r = \pm 2\text{M}$ at 45 degrees. Any light ray emitted at the horizon in a radial direction would remain on the horizon forever.

There is a new and surprising feature in this construction: In M'', regions I' and II' are not in the image of the Schwarzschild manifold M. A closer look at the figure above shows that region II', where

$$-\sqrt{1 + x^2} < u < -|x|\, t, \quad 0 < r < 2\text{M},$$

is the opposite of region II, in the sense that any particle whose worldline starts there must leave this region in finite proper time. Therefore, it is natural to call region II a *white hole*. As we already noticed, in the Schwarzschild manifold, a far away observer sees infalling particles taking infinite time to reach the event horizon. For white holes, on the other hand, outgoing particles are regarded as being traveling outward for an infinite time, as if they had originated at boundary $t = -\infty$ since time immemorial.

Nobody knows if white holes exist in the real universe, and there isn't any known physical processes through which a white hole could be formed. Some astronomers, however, have suggested that GRB 060614, a remarkable gamma-ray burst detected by the Swift satellite on June 14, 2006 is indeed the first documented occurrence of a white hole.

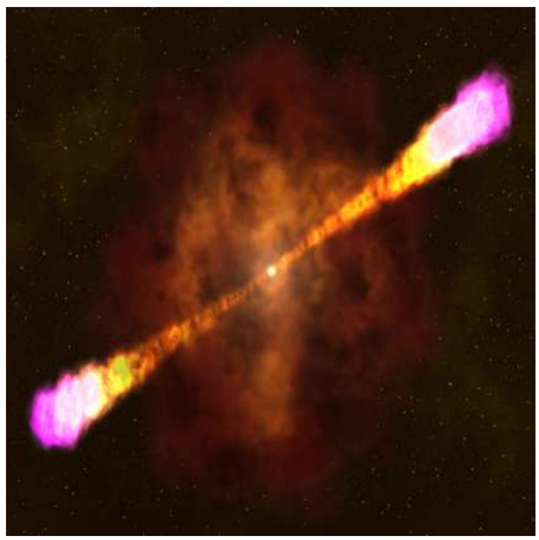


Figure 16.5: White Hole

Region I', defined by $x < u < -x$, and $r > 2\mathrm{M}$, is also novel. This is a exact replica of region I, a *parallel exterior region,* which is also asymptotically flat. There is no way, however to send a signal from one region to the other, as one can see if one follows the direction of the future light cones. Thus, in a sense, regions I and I' are parallel universes only connected by a singularity.

A different perspective of the geometry of M'' can be see if we fix $u = u_0$, $\theta = \theta_0$, while varying r and ϕ. The corresponding 2-dimensional surface would have the geometry of the embedded surface in $\mathbb{R}^3$: The throat of this surface, $r = 2\mathrm{M}$, corresponds to the event horizon. The upper part lies in region I, and the lower, in region I'. Following any of the vertical curves (keeping ϕ also constant while moving from $x = -\infty$ to $x = +\infty$) does

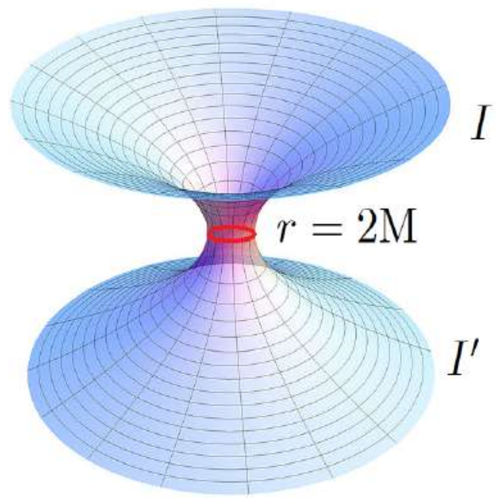


Figure 16.6: $u = u_0$, $\theta = \theta_0$

not provide a bridge between the two parallel universes, since it would not correspond to any timelike curve in M''.

16.4 Hawking Radiation

Particles and antiparticles annihilate to produce radiation. In the opposite direction, with sufficient amounts of energy, particles and antiparticles can be created in particle accelerators. But according to quantum mechanics, even in empty space, with zero energy, pairs of virtual particles and antiparticles, a tiny distance apart, can be created due to *quantum fluctuations*, which are temporary changes in the amount of energy at a point in space, as permitted by Heisenberg's uncertainty principle. But since pair of particles created this way must have positive mass, they have to disappear almost instantly, otherwise the principle of local energy conservation would be violated.

There is one special case, however, when a particle is created just outside the event horizon of a black hole, and the antiparticle appears just inside it. Let's denote by p and $\overline{p}$ the corresponding 4-momenta of the pair. In a small region U containing both particles the geometry is approximately that of Minkowski's space-time, since, for instance, the Schwazschild metric factor $1 - 2\mathrm{M}/r$ varies negligibly around certain value r_0. Hence, conservation of momentum in U (10.9) implies that $p+\overline{p}=0$, and consequently

$$\langle p, \partial_t\rangle + \langle \overline{p}, \partial_t\rangle = 0. \tag{16.9}$$

As noticed above, ∂_t is timelike for $r > 2\mathrm{M}$, but spacelike inside the black hole, since $\langle \partial_t, \partial_t\rangle = -(1-2\mathrm{M}/r)$ (in the figure below, moving along the hyperbolas in Kruskal-Szekeres coordinates amounts to keeping r θ, ϕ constant, and therefore these hyperbolas can be parametrized in Schwarzschild coordinates as $t = t(s), r = r_0, \theta = \theta_0, \phi = \phi_0$ so that the tangent vector at any point would have the same direction as ∂_t). If O is an observer whose 4-velocity lies in the direction of ∂_t, the energy of the particle measured by O would be $E_p = -1/(1-2\mathrm{M}/r_0)^{1/2}\ \langle p, \partial_t\rangle$. But E_p is always a positive quantity, hence $\langle p, \partial_t\rangle$ must be a negative number. Were the antiparticle also outside the horizon we would also have $\langle \overline{p}, \partial_t\rangle < 0$, and equation (16.9) could not hold. *But inside the black hole ∂_t being spacelike does not correspond to energy measured by any observer* and $\langle \overline{p}, \partial_t\rangle$ could be positive. In fact, if we choose an observer O' together with a Lorentz frame where one of the spatial vectors is ∂_t, the quantity $\langle \overline{p}, \partial_t\rangle$ would correspond to 3-momentum, which could be perfectly be negative. Thus, equation (16.9) can be satisfied, as the particle recedes from the horizon, and the antiparticle moves toward the singularity, and it is finally absorbed. In the total balance of energy, the black hole loses a tiny fraction of its mass (the particle absorbed is an antiparticle) and radiates that same amount of energy. It has been estimated

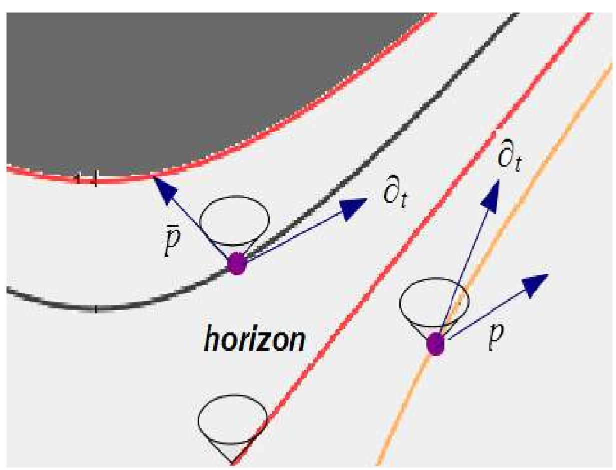


Figure 16.7:

that if $\mathrm{M_{kg}}$ is the black hole's mass (in kilograms) then its evaporation time would be $T = \frac{5129\pi G_N^2}{\hbar c^4}\mathrm{M}_{\mathrm{kg}}^3$ seconds ([44]). Then the lifetime of a black hole whose mass is three times that of our Sun would be greater than 10^{67} years, vastly longer than the age of the universe! For primordial tiny black holes, a hypothetical type formed during the high-density inhomogeneous phase of the Big Bang, with masses of $\mathrm{M} = 10^{11}$ kg (their Schwarzschild's radii would be less than 10^{-14} m) the evaporation time would be about 2.6 billion years. This is why some astronomers search for signs of exploding primordial black holes.

Chapter 17

ROTATING BLACK HOLES XXXXX

Chapter 18

Cosmological Models

18.1 The FRLW Metric

In Section 12.4 we assumed without any further discussion that a spatially flat, homogeneous, isotropic universe could be modeled by the manifold $M = \mathbb{R}_+ \times \mathbb{R}^3$ with coordinates $x = (t, x^i)$, $t > 0$, and metric

$$g = -dt \otimes dt + A(t)^2 (\textstyle\sum_i dx^i \otimes dx^i),$$

where $a(t)$ is a monotonically increasing smooth function. In this chapter we want to give a rigorous definitions of homogeneity and isotropy, and then we want to deduce all possible metrics for a universe that satisfies these hypothesis. Einstein's field equations can then be used to determine the form of the coefficient functions $A(t)$, depending on a set of particular physical hypothesis. The models are sometimes called the *Standard Models* of modern cosmology, or the Friedmann–Robertson–Walker–Lemaître models of the universe, developed independently by the these authors in the 1920s and 1930s.

Let's start with the fundamental definitions.

Definition 18.1.1. A space-time manifold (M, g) is called *homogeneous* if:

1.i There exist a family of commoving observers $O_\alpha(t)$ moving in time-like geodesics such that each point in M belongs to a unique $O_\alpha(t)$. We demand this family to be parametrized by a space-like 3-hypersurface H: That is, we require each observer to be a commoving observer of the form $O_p(t) = (t, p)$, for a unique $p \in H$, where $t \in I$ a (possibly infinite) open interval in $\mathbb{R}$. Therefore, M decomposes as a product $I \times H$, where each point in M belongs to the worldline of a unique geodesic observer O_p.

1.ii If we denote by H_t the spatial slice $\{t\} \times H$ then for any pair of points $p, q \in H_0 = H$ there is an isometry $\phi : M \to M$ such that $\phi(p) = q$.

2 The space-time manifold (M, g) is called *isotropic* if given $p \in H_0$, and two unitary vectors $u, v \in T_p(H_0)$, there exists an isometry $\rho : M \to M$ of the form $\rho = \rho_H \times Id$, such that $d\rho(p)u = v$. Clearly ρ fixes O_p, i.e., $\rho(t, p) = (t, p)$, for all $t \in I$, since $\rho(t, q) = (t, \rho_H(q))$, for all $q \in H_0$.

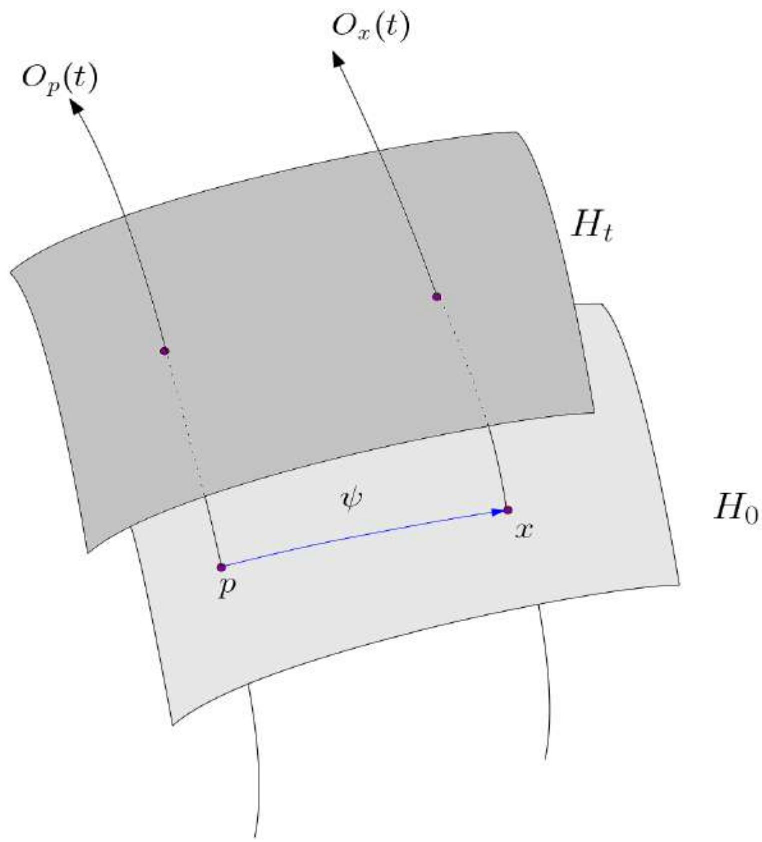


Figure 18.1: Homogeneous space

Remark 18.1.2. The observers $O_p(t)$ provide a global system of Gaussian coordinates for $M = I \times H$.

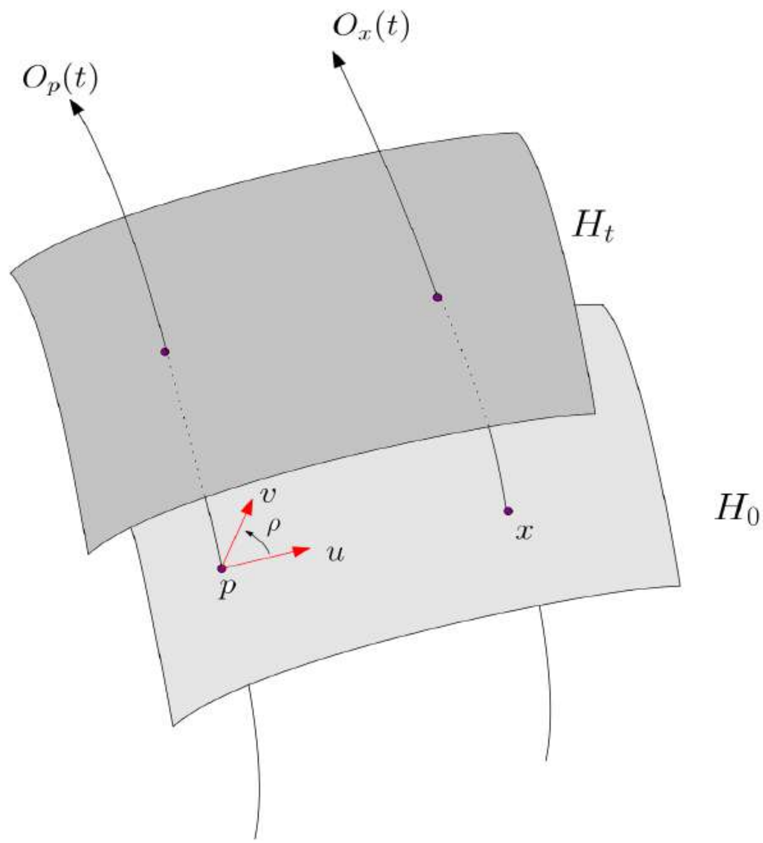


Figure 18.2: Isotropic space

Proof. As we showed in (11.2.9), for the 4-velocity $\mathbf{u}_{t,p}$ of $O_p(t)$ at each point (t, p) to be a unitary normal vector to $T_p(H_t)$ at p, it suffices that this is true for $H_0 = H$. If w denotes the projection of the 4-velocity $\mathbf{u}_p$ in $T_p(H)$, we must show that $w \neq 0$. If not, we could choose a different vector w' of the same length in $T_p(H)$, and an isometry ρ that fixes $\mathbf{u}_p$ and sends w into w'. But since an isometry must preserve projections one would have $w' = w$, a contradiction. □

The family of observers $O_p(t)$ represent all the galaxies, regarded as tiny specks of dust moving in space-time. Conditions 1.ii and 2 just mean that "at present time", $t = 0$, *the spatial universe looks the same for any two observers O_p and O_q, regardless in which direction they look.* In modern cosmology, the notion that the spatial distribution of matter in the universe is homogeneous and isotropic when viewed on a large enough scale is called the *cosmological principle.* In simple words this means that, viewed on a sufficiently large scale, the universe looks the same for all observers, and in every direction they look. Although the universe is not homogeneous at smaller scales, it is statistically homogeneous on scales larger than 250 million light years. And it is isotropic to a large degree: The intensity of the cosmic microwave background, for instance, is about the same in every direction, even though some recent data collected by the Planck Mission has concluded that there are anisotropies statistically significant that can no longer be ignored [29].

Let us denote by $f_t : H_0 \to H_t$ the map that sends each point $z \in H_0$ into the point $\widetilde{z} = (t, z) \in H_t$. This point is, by definition, the unique point of intersection of $O_z(t)$ with H_t. This map is not necessarily an isometry, even though, as we shall see below, it must be a conformal map. Before we state this precisely, we want to first observe the following:

Remark 18.1.3.

1. Let $p \in H_0$, and ϕ any isometry of the homogeneous space-time M. Let $q = \phi(p)$. Then ϕ must also send $\widetilde{p}$ into $\widetilde{q}$, and therefore $\phi \circ f_t = f_t \circ \phi$.
2. Suppose (M, g) is homogeneous and isotropic. Then given $p, q \in H_0$, and $u \in T_p(H_0)$, $v \in T_q(H_0)$ unitary vectors, there is an isometry $\psi : M \to M$ that takes p into q, and such that $d\psi(p)u = v$.

Proof. In fact, since an isometry must send geodesics into geodesics, the curve $\phi(O_p(t))$ must be a geodesic that passes through q, and that is normal to H at this point. But this curve is unique, since M is homogeneneous, thus this curve coincide with $O_q(t)$. From this one immediately sees that $\phi(\widetilde{p}) = \widetilde{q}$. That is, $\phi(f_t(p)) = f_t(\phi(p))$, for any p. To prove the second assertion one just needs to take the composite of any isometry ϕ taking p into q (M is homogeneous) followed by an isometry that fixes q and takes u into v (M is isotropic). □

Let us denote by $g_t = \langle -, - \rangle_t$ the corresponding Riemannian metric induced by g in each spatial slice H_t. We want to see next that for each t fixed then there is $a_t > 0$ such that $\langle -, - \rangle_t = a_t \langle -, - \rangle_0$, that is all the induced metrics g_t are *conformal.*

Theorem 18.1.4. For each fixed value of t there is positive constant a_t such that $\langle -, - \rangle_t = a_t \langle -, - \rangle_0$.

Proof. We use the notation $f_t(z) = \widetilde{z}$ for a fixed t. Let $p \in H_0$ be an arbitrary point, and let U be a suffciently small neighborhood of p such that for any $x \in U \subset H_0$ there exists a unique geodesic $\alpha(s)$ that passes through x and p, let's say with $\alpha(0) = p$, and $\alpha(s_0) = x$. denote by u the (unitary) tangent vector $u = \alpha'(0)$. We know there is ψ, an isometry of M, that sends p into x, and u into $v = \alpha'(s_0)$ (18.1.3, 2). In a smaller open interval $I_0 \subset I$ containing zero, the curve $\beta : I_0 \to M$ given by $\beta(s) = \psi(\alpha(s))$ is also a geodesic, since an isometry sends geodesics into geodesics. By construction $\beta(0) = x$, and $\beta'(0) = v$, thus the uniqueness of a geodesic passing through x with fixed tangent vector implies that $\beta(s) = \alpha(s+s_0)$, $s \in I_0$. By (18.1.3, 1) above $f_t\psi(\alpha(s)) = \psi f_t(\alpha(s))$. Hence, $f_t(\alpha(s+s_0)) = \psi f_t(\alpha(s))$, or equivalently: $\widetilde{\alpha}(s+s_0) = \psi(\widetilde{\alpha}(s))$. taking derivates at $s = 0$ one obtains $\widetilde{\alpha}'(s_0) = d\psi(\widetilde{p})\widetilde{\alpha}'(0)$. If $\widetilde{v} = \widetilde{\alpha}'(s_0) = df_t(x)v$, we imediately obtain $|\widetilde{v}| = |\widetilde{\alpha}'(s_0)| = |\widetilde{\alpha}'(0)|$. Define $a_t > 0$ as

$$a_t = |\widetilde{\alpha}'(0)| \,/\, |\alpha'(0)| = |\widetilde{\alpha}'(0)| = |\widetilde{v}|$$

(notice $|\alpha'(0)| = |u| = 1$). Thus we have proved that for any point $\widetilde{x}$ in $\widetilde{\alpha}(s)$ the tangent vector $\widetilde{v}$ of $\widetilde{\alpha}(s)$ at $\widetilde{x}$ has constant norm a_t. Now, fix any unitary vector $w \in T_x(H_0)$, and denote by $\widetilde{w}$ the corresponding vector $\widetilde{w} = df_t(x)w$ in $T_{\widetilde{x}}(H_t)$. We know there is an isometry ρ that takes $v = \alpha'(s_0)$ into w, and fixes x. By (18.1.3, 1)

$$df_t(x)d\rho(x)v = d\rho(x)df_t(x)v = d\rho(x)\widetilde{v}.$$

But $df_t(x)d\rho(x)v = df_t(x)w = \widetilde{w}$, henceforth $\widetilde{w} = d\rho(x)\widetilde{v}$. Since ρ is an isometry one gets

$$\begin{aligned}\langle \widetilde{w}, \widetilde{w} \rangle_t &= \langle \widetilde{v}, \widetilde{v} \rangle_t = |\widetilde{v}|^2 = a_t^2 \\ &= a_t^2 \langle w, w \rangle_0 ,\end{aligned}$$

this last equality, since w is unitary. But any vector in $T_{\widetilde{x}}(H_t)$ can be obtained as the image of some $w \in T_x(H_0)$. Thus, we have proved that for an arbitrary point $\widetilde{x}$ in a small neighborhood $\widetilde{U} \subset H_t$, and any vector $\widetilde{w} \in T_{\widetilde{x}}(H_t)$, one has:

$$\langle df_t(x)w, df_t(x)w \rangle_t = a_t^2 \langle w, w \rangle_0 .$$

This shows that $h_t = 1/a_t^2 \ f_t$ is such that

$$\langle dh_t(x)w, dh_t(x)w \rangle_t = \langle w, w \rangle_0 , \tag{18.1}$$

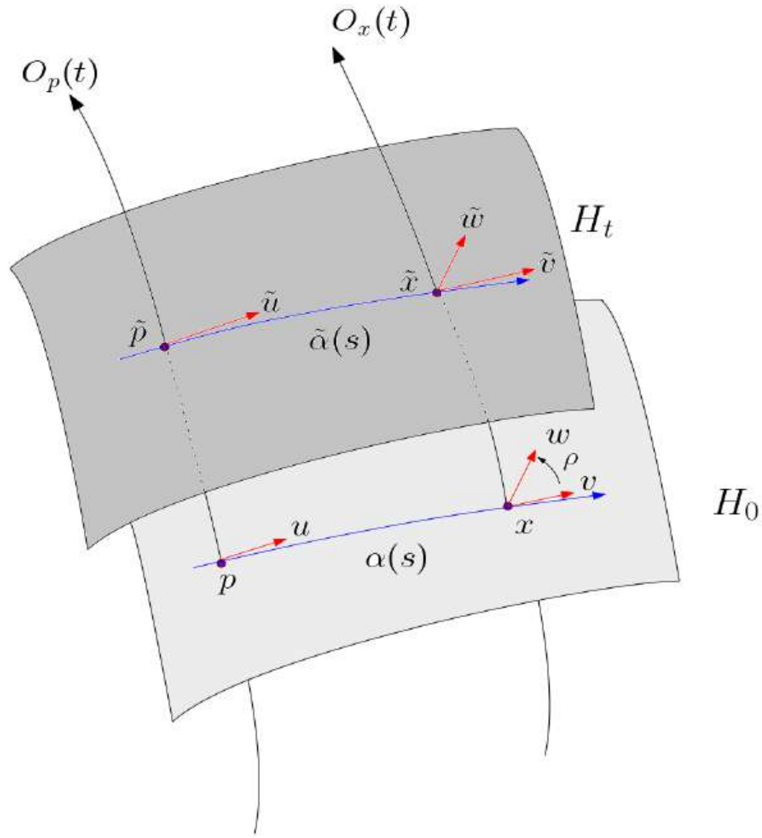


Figure 18.3

for all w. By applying (18.1) to any vector of the form $w = w_1 + w_2$ one readily sees that

$$\langle dh_t(x)w_1, dh_t(x)w_2\rangle_t = \langle w_1, w_2\rangle_0 \,,$$

and therefore $h_t(x)$ is an isometry. Thus, $\langle -, -\rangle_t = a_t \langle -, -\rangle_0$ in a sufficiently small neighborhood U of p. Since H_0 is connected one must have that $\langle -, -\rangle_t = a_t \langle -, -\rangle_0$, in all H_0. □

Since $a_t = \langle -, -\rangle_t / \langle -, -\rangle_0$, the function $a(t) = a_t$ is smooth. We want to show next that all the hyperspaces H_t have constant sectional curvature.

Theorem 18.1.5. All the hyperspaces H_t have constant sectional curvature κ/a_t, with $\kappa = \kappa(H_0)$. If we further assume these are geodesically complete, connected and simply connected, then they must be isometric to $(L_k, \frac{a_t}{\kappa} g_k)$ where (L_k, g_k) is one of the three standard models (Example **??**) $\mathbb{R}^3$, S^3 or $\mathbb{H}^3$, depending whether $\kappa(H_t)$ is zero, positive or negative, respectively. Then in each case the metric of M can be written in coordinates (t, ψ, θ, ϕ) $(\psi, \theta, \phi$ as in **??**):

1. $g = -dt \otimes dt + A^2(t) g_0,\ g_0 = d\psi \otimes d\psi + \psi^2 (d\theta \otimes d\theta + \sin^2(\theta) d\phi^2)$
2. $g = -dt \otimes dt + A^2(t) g_1,\ g_1 = d\psi \otimes d\psi + \sin^2(\psi)(d\theta \otimes d\theta + \sin^2(\theta) d\phi^2)$
3. $g = -dt \otimes dt + A^2(t) g_{-1},\ g_{-1} = d\psi \otimes d\psi + \sinh^2(\psi)(d\theta \otimes d\theta + \sin^2(\theta) d\phi^2)$

By replacing ψ in 1,2,3 by the radial coordinate $r = \sin\psi$, $r = \psi$, $r = \sinh\psi$, respectively, the metrics g_0,g_1, g_{-1} can be written in a unified form as $(k = 0, 1, -1)$:

$$g = -dt \otimes dt + A^2(t) \left[\frac{1}{1 - kr^2} dr \otimes dr + r^2 (d\theta \otimes d\theta + \sin^2\theta \ d\phi \otimes d\phi) \right]. \tag{18.2}$$

Proof. First, let us prove that $H_0 = H$ has constant sectional curvature. For any $p \in H$ we know there is an isotropy ρ of the form $\rho_H \times Id$ and therefore $\rho_H : H \to H$ is an isotropy at p. By Proposition 6.3.7, $\kappa_p(H)$ must be constant for all tangent planes at p. On the other hand, if $p, q \in H$ we know there exists an isometry ϕ taking p into q. Thus, Proposition **??** shows that $\kappa_p(H) = \kappa_q(H)$ and therefore H has constant sectional curvature $\kappa(H) = \kappa$ (for this last assertion, isotropy would be enough by Theorem 6.3.8). Since we know that $f_t : H \to H_t$ is a homothety with constant a_t, one must have at every point $p \in H$ (Proposition **??**) that $\kappa_p(H_t) = \kappa / a_t$ and this shows that the section curvature of every spatial slice H_p is constant, equal to κ / a_t. By Remark **??**, under the hypothesis of H_t being geodesically complete, connected and simply connected, one concludes that H_t is isometric to $(L_k, a_t/\kappa \ g_k)$ for one of the three models (L_k, g_k), depending whether $\kappa = 0, \kappa > 0$ or $\kappa < 0$. Finally, we choose the coefficient function $A(t)$ so that $\langle -, - \rangle_{A^2(t) g_k} = \langle -, - \rangle_{H_t}$. This can be done as follows: If $\kappa = 0$, we choose $A(t) = a_t^{1/2}$. If $\kappa > 0$, we choose $A(t) = a_t / \kappa^{1/2}$. And if $\kappa < 0$, we choose $A(t) = a_t / (-\kappa)^{1/2}$. This concludes the proof. □

Our next objective is to determine the form of $A(t)$, by using Einstein's Field Equation

$$\mathsf{Ric}_{ab} = 8\pi(T_{ab} - \frac{1}{2}g_{ab}\mathrm{T})$$

We use global coordinates (t, r, θ, ϕ) as in (18.2). The main assumption is that matter and energy in the universe move as a perfect fluid in the direction of time with 4-velocity $\mathbf{u} = \partial_t$, with density and pressure given by two fixed functions $\rho(t)$ and $\mathrm{P}(t)$, so that the Tensor of energy-momentum is given by $\overline{T}^{ab} = (\rho(t) + \mathrm{P}(t))u^a u^b + g^{ab}\mathrm{P}(t)$. Since $\mathbf{u}^0 = 1$ and $\mathbf{u}^i = 0$, the only non zero components of $\overline{T}$ are

$$\overline{T}^{00} = (\rho(t) + \mathrm{P}(t)) + (-1)\mathrm{P}(t) = \rho(t)$$
$$\overline{T}^{ii} = g_{ii}^{-1}\mathrm{P}(t).$$

The tensor T_{ab} is obtained by lowering indices: $T_{ab} = \sum_{r,s} g_{ar} g_{bs} \overline{T}^{rs}$. Hence, the only non zero components of this tensor are:

$$T_{00} = g_{00}^2 \overline{T}^{00} = \rho(t)$$
$$T_{ii} = g_{ii}^2 g_{ii}^{-1}\mathrm{P}(t) = g_{ii}\mathrm{P}(t).$$

Hence, the trace of T_{ab}, $\mathrm{T} = \sum_{r,s} g^{rs} T_{rs}$ is, equal to

$$\mathrm{T} = g^{00}\rho(t) + \textstyle\sum_i g_{ii}^{-1} g_{ii}\mathrm{P}(t) = -\rho(t) + 3\mathrm{P}(t).$$

A computation using Maple gives the following Christoffel symbols:

$\Gamma^0_{11} = \dfrac{A(t)A'(t)}{1 - kr^2},\ \Gamma^0_{22} = r^2 A(t)A'(t),\ \Gamma^0_{33} = r^2 \sin^2\theta A(t)A'(t),$

$\Gamma^1_{11} = \dfrac{kr}{1 - kr^2},\ \Gamma^1_{22} = -r(1 - kr^2),\ \Gamma^1_{33} = -r(1 - kr^2)\sin^2\theta,\ \Gamma^1_{01} = \dfrac{A'(t)}{A(t)},$

$\Gamma^2_{12} = \dfrac{1}{r},\ \Gamma^2_{33} = -\sin\theta\cos\theta,\ \Gamma^2_{02} = \dfrac{A'(t)}{A(t)}$

$\Gamma^3_{13} = \dfrac{1}{r},\ \Gamma^3_{23} = \cot\theta,\ \Gamma^3_{03} = \dfrac{A'(t)}{A(t)}$

$\Gamma^c_{ab} = 0$, in any other case.

On the other hand, the Ricci tensor can be computed (**??**) as

$$\mathsf{Ric}_{ab} = \sum_u \left(\frac{\partial\Gamma^u_{ab}}{\partial x^u} - \frac{\partial\Gamma^u_{au}}{\partial x^b}\right) + \sum_{u,r}\left(\Gamma^u_{ab}\,\Gamma^r_{ur} - \Gamma^u_{ar}\,\Gamma^r_{bu}\right).$$

This gives:

$$\begin{aligned}
\mathsf{Ric}_{00} &= -3\frac{A''(t)}{A(t)} \\
\mathsf{Ric}_{11} &= \frac{A(t)A''(t) + 2(A')^2 + 2k}{1 - kr^2} \\
\mathsf{Ric}_{22} &= r^2\left[A(t)A''(t) + 2(A')^2 + 2k\right] \\
\mathsf{Ric}_{33} &= r^2 \sin^2\theta \left[A(t)A''(t) + 2(A'(t))^2 + 2k\right] \\
\mathsf{Ric}_{ab} &= 0, \text{ in any other case.}
\end{aligned}$$

Einstein's equation for $a = b = 0$ gives:

$$-3\frac{A''(t)}{A(t)} = 8\pi(\rho(t) + \frac{1}{2}\left[-\rho(t) + 3\mathrm{P}(t))\right] = 4\pi(\rho(t) + 3\mathrm{P}(t)), \tag{18.3}$$

or equivalently

$$\frac{A''(t)}{A(t)} = \frac{-4\pi}{3}(\rho(t) + 3\mathrm{P}(t)) \tag{18.4}$$

For $a = i$, $b = i$, Einstein's equations coincide, and they are all the equation:

$$\frac{A''(t)}{A(t)} + 2\left(\frac{A'(t)}{A(t)}\right)^2 + 2\frac{k}{A(t)^2} = 4\pi(\rho(t) - \mathrm{P}(t)). \tag{18.5}$$

Substituting $A''(t)/A(t)$ in (18.5) one obtains:

$$\left(\frac{A'(t)}{A(t)}\right)^2 = \frac{8\pi}{3}\rho(t) - \frac{k}{A(t)^2}, \tag{18.6}$$

or equivalently

$$A'(t)^2 - \frac{8\pi\rho(t)}{3}A(t)^2 = -k. \tag{18.7}$$

Equations 18.4 and 18.7 are called *Friedman equations.* We notice that in Equation 18.6 the term $A'(t)/A(t)$ is the Hubble function $\mathrm{H}(t)$ (Section 12.4.3). If we evaluate (18.6) at the *present time* $t = t_0$, one obtains

$$\mathrm{H}_0^2 - \frac{8\pi\rho_0}{3} = -\frac{k}{A(t_0)^2} = -k,$$

where H_0 is the Hubble constant, $\rho_0 = \rho(t_0)$ is the density of the universe at present time, and where we have normalized $A(t)$ in such a way that so that $A(t_0) = 1$, as in Section 12.4. Clearly, if $k = 0$, corresponding to a flat universe, the density of the universe at present time must be equal to $\rho_{\text{crit}} = 3\mathrm{H}_0^2/8\pi$, called the *critical density.* For if $\rho_0 > \rho_{\text{crit}}$ one must have $k > 0$, hence $k = +1$, and the universe is a *closed universe* of finite volume. On the other hand, if $\rho_0 < \rho_{\text{crit}}$ one must have $k < 0$, hence $k = -1$, and the universe is *open*, and infinite.

From the Friedmann equations one can obtain the first law of thermodynamics for cosmology: If we differentiate 18.7 with respect to t we get (for simplicity we omit t from the notation):

$$2A'A'' - \frac{8\pi}{3}(\rho' A^2 + 2\rho A A') = 0. \tag{18.8}$$

From (18.4) we see that $A'' = \frac{-4\pi}{3}(\rho + 3\mathrm{P})A$. If we substitute A'' in (18.8) we obtain:

$$\frac{-8\pi}{3}A'(\rho + 3\mathrm{P})A - \frac{8\pi}{3}(\rho' A^2 + 2\rho A A') = 0,$$

or equivalently:

$$3\rho(t)A(t)A'(t) + 3\mathrm{P}(t)A(t)A'(t) + \rho'(t)A(t)^2 = 0. \tag{18.9}$$

After multiplying by $A(t)$, this last equation can be written as:

$$\frac{d}{dt}(\rho(t)A^3(t)) = -\mathrm{P}(t)\frac{d}{dt}A^3(t), \tag{18.10}$$

which is the *first law of thermodynamics for cosmology.* This equation has a clear physical meaning: if at a particular time t we consider a cube C of unit coordinate length, hence of volume $V(t) = A^3(t)$ in the metric, the amount of energy inside C would be $E(t) = \rho(t)V(t)$. As the universe expands, the energy density changes by $\frac{d}{dt}E(t) = \frac{d}{dt}(\rho(t)A^3(t))$. Equation (18.10) then says that it diminishes by $\mathrm{P}(t)\frac{d}{dt}V(t)$, which can be interpreted as the differential of work done against the external pressure in order to expand. It is an easy exercise to show that (18.10) can also be obtained from the local conservation of energy law: $\sum_b(\nabla_{\partial_b}\overline{T})^{0b} = 0$.

There are two important instances in which (18.10) can be solved exactly. These are when $\mathrm{P}(t) = 0$, that corresponds to a universe dominated by matter, where galaxies move as dust in space-time. The other case is when the universe is dominated by radiation. In this case it can be shown that the vanishing of the energy momentum tensor for electromagnetism suggests that

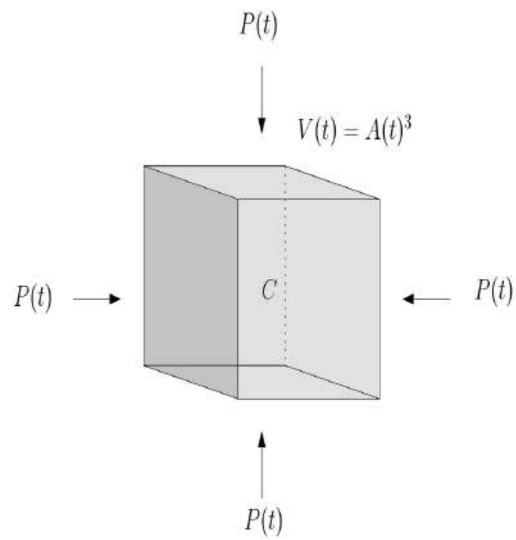


Figure 18.4:

the pressure and the density are related by a simple equation: $\mathrm{P}(t) = \rho(t)/3$ (see [11], Page 373, 374), ([18]). In the first case one gets from (18.10) that $\rho(t)A^3(t) =$ constant. To determine this constant we evaluate this equation at $t = t_0$, (present time). This constant must then be $\rho(t_0)A^3(t_0) = \rho(t_0)$ (we recall $A(t)$ has been normalized so that $A(t_0) = 1$). Hence, the dilation factors must be equal to $\rho(t) = \rho_0/A^3(t)$.

In the second case, after substituting $\mathrm{P}(t) = \rho(t)/3$ in (18.9) one gets:

$$3\rho(t)A(t)A'(t) + \rho(t)A(t)A'(t) + \rho'(t)A(t)^2 = 0,$$

which can be written as

$$4\rho(t)A(t)A'(t) + \rho'(t)A(t)^2 = 0,$$

or more simply (after multiplying by $A^2(t)$) as:

$$\frac{d}{dt}(\rho(t)A^4(t)) = 0.$$

Again, evaluating at $t = t_0$ one gets $\rho(t) = \rho_0/A^4(t)$.

In a universe dominated by matter, Friedman equation (18.7) reads:

$$A'(t)^2 - \frac{8\pi\rho_0}{3A(t)} = -k. \tag{18.11}$$

While in a universe dominated by radiation one gets

$$A'(t)^2 - \frac{8\pi\rho_0}{3A(t)^2} = -k. \tag{18.12}$$

For a flat universe $k = 0$, using that $\rho_{\text{crit}} = 3\text{H}_0^2/8\pi$, these differential equations can be written as

$$A'(t)^2 - \frac{\text{H}_0^2}{A(t)} = 0$$
$$A'(t)^2 - \frac{\text{H}_0^2}{A^2(t)} = 0.$$

Since $\text{H}_0 = A'(t_0)/A(t_0)t$, one can readily see that these differential equations can be solved as:

$$A(t) = (t/t_0)^{2/3}, \text{ and } A(t) = (t/t_0)^{1/2}, \text{ respectively,}$$

a fact that we had assumed in Section 12.4 for a flat universe.

In general, the quantitative behavior of (18.11) and (18.12) is easy to deduce. By (18.4) $A(t)$ must have negative concavity (assuming the energy condition $\rho(t) + 3\text{P}(t) > 0$):

In Particular, one can observed that a closed universe expands from the Big-Bang, and then contracts to a Big-Crunch, while the flat an open universes expand forever.

18.2 A FURTHER SECTION ON COSMOLOGY XXXXX

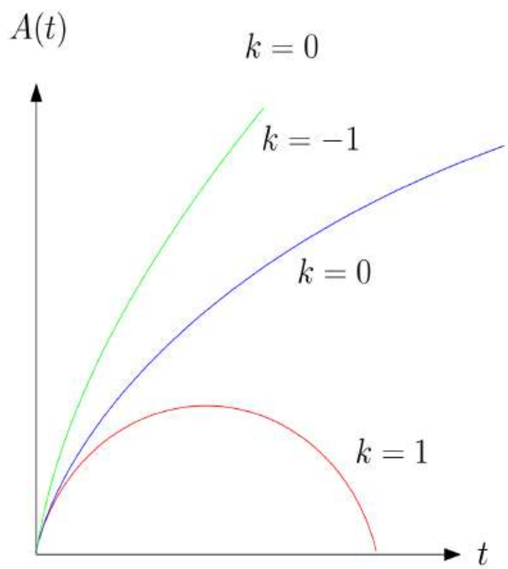


Figure 18.5: Dust-Filled and radiation-filled universe

Chapter 19

HAWKING AND PENROSE SINGULARITIES THEOREM XXXXX

Bibliography

[1] Baez, J. C., & Muniain, J. P. (1994). Gauge fields, knots and gravity (Vol. 6). Singapore: World Scientific.

[2] Baez, J. C., & Bunn, E. F. (2005). The meaning of Einstein's equation. American journal of physics, 73(7), 644-652.

[3] Boothby, W. M. (1986). An introduction to differentiable manifolds and Riemannian geometry (Vol. 120). Academic press.

[4] Carroll, S. M. (2004). Spacetime and geometry. An introduction to general relativity (Vol. 1).

[5] Perdigão do Carmo, M. (1976). Differential geometry of curves and surfaces.

[6] Feynman, R. P., Leighton, R. B., & Sands, M. (1969). [Lectures on physics]; The Feynman lectures on physics. 2 (1969). Mainly electromagnetism and matter. Addison-Wesley.

[7] Franklin, J. (2005). Classical electromagnetism. Pearson.

[8] Alonso, M., & Finn, E. J. (1967). Fundamental university physics (Vol. 2). Reading, MA: Addison-Wesley.

[9] Griffiths, D. J., & Reed College. (1999). Introduction to electrodynamics (Vol. 3). Upper Saddle River, NJ: prentice Hall.

[10] Guillemin, V., & Pollack, A. (2010). Differential topology (Vol. 370). American Mathematical Soc.

[11] Hartle, J. B. (2002). Gravity: an introduction to Einstein's general relativity. Pearson Education India.

[12] Hawking, S. W., & Ellis, G. F. R. (1973). The large scale structure of space-time (Vol. 1). Cambridge university press.

[13] Hehl, F. W., & Obukhov, Y. N. (2012). Foundations of classical electrodynamics: Charge, flux, and metric (Vol. 33). Springer Science & Business Media.

[14] Hicks, N. J. (1965). Notes on differential geometry.

[15] Hirsch, M. W. (2012). Differential topology (Vol. 33). Springer Science & Business Media.

[16] Jackson, J. D., & Jackson, J. D. (1962). Classical electrodynamics (Vol. 3). New York etc.: Wiley.

[17] Kobayashi, S., & Nomizu, K. (1963). Foundations of differential geometry (Vol. 1, No. 2). New York.

[18] Kolb, E., & Turner, M. (1994). The early universe (Vol. 69). Westview press.

[19] Lang, S., Algebra, Addison-Wesley 1984.

[20] Levinson, N., Theory of Ordinary Differential Equations, N. Y. Mc Graw-Hill, 1955.

[21] https://en.wikisource.org/wiki/A_Treatise_on_Electricity_and_Magnetism

[22] Chorin, A. J., Marsden, J. E., & Marsden, J. E. (1990). A mathematical introduction to fluid mechanics (Vol. 3). New York: Springer.

[23] Madsen, I. H., & Tornehave, J. (1997). From calculus to cohomology: de Rham cohomology and characteristic classes. Cambridge University Press.

[24] Morita, S. (2001). Geometry of differential forms (Vol. 201). American Mathematical Soc.

[25] Munkres J., Elements of Algebraic Topology, Benjamin Pub., 1984.

[26] Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973). Gravitation. Macmillan.

[27] Biesel, O. (2008). The Precession of Mercury's Perihelion.

[28] Papachristou, C. J., & Magoulas, A. N. (2012). Electromotive force: A guide for the perplexed. arXiv preprint arXiv:1211.6463.

[29] http://sci.esa.int/planck/51551-simple-but-challenging-the-universe-according-to-planck/

[30] Pollack, G. L., Stump, D. R., & Inglefield, C. (2005). Electromagnetism. American Journal of Physics, 73(6), 574-574.

[31] Rindler, W. (2001). Relativity: special, general, and cosmological (Vol. 1).

[32] Sachs, R. K., & Wu, H. H. (2012). General relativity for mathematicians (Vol. 48). Springer Science & Business Media.

[33] Schutz, B. (2009). A first course in general relativity. Cambridge university press.

[34] Schwarzschild, K. (1916, March). Über das Gravitationsfeld einer Kugel aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie. In Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, Phys.-Math. Klasse, 424-434 (1916).

[35] Taubes, C. H. (2011). Differential geometry: bundles, connections, metrics and curvature (Vol. 23). Oxford University Press.

[36] Tu, L. (2011). An introduction to manifolds, Universitext, Springer Verlag, second edition.

[37] http://www.aeromental.com/2009/05/01/un-iman-cae-lentamente-por-un-tubo-de-cobre/9

[38] Wald, R. M. (2010). General relativity. University of Chicago press.

[39] Warner, F. W. (2013). Foundations of differentiable manifolds and Lie groups (Vol. 94). Springer Science & Business Media.

[40] Weinberg, S., & Dicke, R. H. (1973). Gravitation and cosmology: principles and applications of the general theory of relativity. American Journal of Physics, 41(4), 598-599.

[41] https://en.wikipedia.org/wiki/Relativistic_aberration

[42] Ball, Philip (2007). " Arthur Eddington was innocent!". News@nature. doi:10.1038/news070903-20

[43] https://en.wikipedia.org/wiki/Black_hole

[44] https://en.wikipedia.org/wiki/Hawking_radiation

[45] Ref: https://en.wikipedia.org/wiki/RD1