Preface

General Relativity Lecture notes
Govind Menon

These are the lecture notes for Relativity I (PHY 4460/5560) and II (PHY 4478/5578) offered at Troy University. Chapters 1-5 are covered in full detail during the first semester. The assigned problems are integral to the lectures, and I would advise every student to work through them carefully. Should you find any errors or even a lack of clarity of explanation, please contact me via e-mail at gmenon(at)troy.edu.

There is hardly a need to say that the material presented is not original in content. But it is simply a version of how I envision the first reading of relativity should be. After going through my notes, you should be able to understand more advanced texts in general relativity by authors like Poisson and Straumann with relative ease.

A typical course sequence in relativity will not spend as much time as I have spent on special relativity. It is my strong personal belief that this trend is a shortcoming of how we teach general relativity. Inspired by the wonderful book on Minkowski spacetime by Gregory Naber [1], I have picked the bare essentials from Naber's text and included them in the first three chapters of the notes. Our notes start at the very beginning (it's a very good place to start), so, as such, there aren't any formal prerequisites for this course outside your freshman Calculus sequence. Indeed, sophomores have taken this course sequence successfully.

I would like to thank the Relativity class of 2020-21: Rachid Bowles, Soumitra Ganguly, Caroline Howell, Colin Jones, Nicholas Johnson, Ashik Kannan, and Natalie Larremore for helping me typeset a portion of these notes. Indeed, the students who have come before have helped me select the choice of topics, and I have valued their questions and input. I have primarily compiled these notes together since my students have requested a "textbook" for the class.

Live Long and Ponder,

Govind Menon
Troy, Alabama

Contents

Part I Lecture Notes for Relativity I (PHY 4460/5560)

1 The Lorentz Transformation . . . 3

2 Inertial Observers . . . 11
2.1 The Rotation Group . . . 11
2.2 Classes of Inertial Observers . . . 15
2.3 Minkowski Spacetime . . . 16
2.4 The Lorentz Group . . . 18
2.5 Causal Structure of Minkowski Spacetime . . . 23
2.6 Zeeman's Theorem . . . 27

3 Relativistic Mechanics and Electrodynamics . . . 29
3.1 Particle Dynamics . . . 29
3.1.1 Massless Particles . . . 40
3.2 A First Look at Tensors . . . 42
3.3 Relativistic Electrodynamics . . . 46
3.3.1 The Klein-Gordon Field . . . 46
3.3.2 Electrodynamics . . . 47
3.4 The Energy-Momentum Tensor . . . 49

4 Introduction to Gravity . . . 55
4.1 Gravitational Effects in Minkowski Spacetime . . . 55
4.1.1 Bending of Light . . . 55
4.1.2 Gravitational Redshift . . . 55
4.1.3 Gravitational Time Dilation . . . 56
4.2 Generalized Coordinates . . . 56
4.3 Tensors in Generalized Coordinates . . . 58
4.4 Free Particle in a Gravitational Field . . . 61
4.4.1 The Geodesic Equation of Motion . . . 62
4.4.2 Photon trajectories in General Relativity . . . 64

5 Schwarzschild Solution . . . 65
5.1 A Heuristic Argument for the Solution . . . 65
5.1.1 Timelike Geodesics . . . 66
5.1.2 The Perihelion Advance of Mercury . . . 67
5.1.3 Null Geodesics . . . 69
5.1.4 Shapiro Time Delay . . . 70
5.2 The Schwarzschild Black Hole . . . 71
5.2.1 The problem with time . . . 72
5.2.2 "Probing" the Event Horizon with Light . . . 72
5.2.3 Black Hole Coordinates . . . 73
5.3 The Schwarzschild Black Hole Event Horizon . . . 75
5.4 The Schwarzschild White Hole Spacetime . . . 76
5.5 Maximal Schwarzschild Geometry . . . 77
5.5.1 Transformations for the Exterior Black Hole Geometry . . . 77
5.5.2 Transformations for the Interior Black Hole Geometry . . . 79

5.5.3 Transformations for the White Hole Geometry . . . 80
5.5.4 Concluding Remarks . . . 80

Part II Lecture Notes for Relativity II (PHY 4478/5578)

6 Tensor Calculus . . . 85
6.1 Vector and Tensor Bases . . . 85
6.2 Lie Brackets of Vector Fields . . . 91
6.3 The Covariant Derivative . . . 92
6.4 Parallel Transport and Geodesics . . . 95
6.5 Torsion Tensor . . . 97
6.6 The Levi-Civita Connection . . . 98

7 Gravitation and Curvature . . . 101
7.1 The Riemann Curvature Tensor . . . 101
7.2 Symmetries of the Riemann Tensor . . . 103
7.3 Contractions of the Riemann Tensor . . . 105
7.4 The Einstein Tensor . . . 106
7.5 The Einstein Equation . . . 107
7.6 Fields in Curved Spacetime . . . 110

8 Cosmology . . . 113
8.1 Maximally Symmetric Space . . . 114
8.1.1 Flat Space . . . 115
8.1.2 Hyperbolic Space . . . 115
8.1.3 The HyperSphere . . . 116
8.2 Co-moving Coordinates . . . 117
8.2.1 FLRW Spacetime . . . 117
8.2.2 Proper Distance in the Galactic Spatial Grid . . . 118
8.2.3 Hubble's Law . . . 118
8.2.4 Light Rays in FLRW Spacetime . . . 119
8.2.5 Luminosity distance . . . 120
8.3 Dynamical Equations for FLRW Cosmology . . . 121
8.3.1 Friedmann Equations . . . 121
8.3.2 The Equation of State . . . 122
8.3.3 Heuristic Explanation of Friedmann Equations . . . 123
8.3.4 Critical Density and Spatial Geometry . . . 124
8.4 Dark Energy . . . 125
8.5 Exact Solutions is FLRW Cosmology . . . 126
8.5.1 The Radiation Filled Universe . . . 126
8.5.2 The Flat Universe . . . 126
8.5.3 The Matter Dominated Non-Flat Universe . . . 128
8.5.4 Familiarity with Dark Energy . . . 129
8.5.5 Asymptotic Behavior . . . 130
8.5.6 The general expression for a pure epoch . . . 132
8.6 A Brief Introduction to the Physical Universe . . . 132
8.6.1 The Early Universe . . . 133
8.6.2 The Cosmic Microwave Background . . . 134
8.6.3 Matter Domination . . . 134
8.7 Inflation . . . 135

9 Odds And Ends . . . 137
9.1 Relativistic Doppler Effect for Light . . . 137
9.2 Motivating Relativistic Momentum . . . 138
9.3 The Geodesic Equation . . . 139

References . . . 141

Part I

Lecture Notes for Relativity I (PHY 4460/5560)

1

The Lorentz Transformation

We will begin our study of (special) relativity by stating the postulates of the theory on which the entire subject is built.

Postulate 1: The speed of light, c, in vacuum has the same value in every inertial frame of reference.

Postulate 2: The laws of physics take the same form in all inertial frames of reference.

It is important to realize that a physical theory is only as true as the validity of its postulates. The remainder of the subject is a consequence of logic and analysis. Although relativity is a very robust and mature theory, continuous efforts are being made to test the limits of its postulates. Suffice it to say that we will consider the postulates of relativity to hold for the remainder of these notes.

Our job is to describe a set of preferred observers who carry a measuring stick and clocks so that they may record the spatial distance to any point of interest and also set up a clock at that location such that all clocks agree with the clock carried by the observer for all time. Consider such a potential "inertial observer" S. The coordinates used by the inertial observer S will be referred to as an **inertial frame** or an **inertial coordinate system**. From the first postulate of relativity, Newton's first law of motion is satisfied in this inertial frame, and so the observer S will see all potential free particles as moving with a constant velocity. This means that the observer S must not be accelerating. To locate the distance to any spatial point P from the spatial origin O of the observer, she/he simply sends a beam of light from O to P and have it reflect back to O and records the time taken as t_{OP}. From the first postulate of relativity, the unambiguous distance between O and P is simply given by

$$d_{OP} = c\,\frac{t_{OP}}{2}\ .$$

Also, if an event is recorded at the spatial location P, a signal is sent back to O via a light signal. Let t_P be the time when the signal was received at O by S. Then S records the time of the event as

$$t_P - \frac{d_{OP}}{c}\ .$$

In this manner, in principle, S is able to assign a set of four numbers (t, x, y, z) for all events. The point $x = y = z = 0$, locates the origin O mentioned above. Since space and time will be on equal footing in the development of the theory, we would instead have S use the coordinates (ct, x, y, z). This way, all of the 4-tuples have a dimension of distance.

We will soon have occasion to describe the complete catalog of all other possible inertial observers, and the coordinate transformations between them. But for now, we will focus on a very special type of observer. Let us suppose that an observer $\bar{S}$ is moving along the positive x-axis of S with a constant speed v. Since in Newtonian mechanics such an observer would also be deemed inertial (accelerations of objects as viewed by S and $\bar{S}$ are the same, and hence the expression for the net force is unchanged), we would want the speed of light in $\bar{S}$ to be exactly c. How can this be? For, after all, relative velocities must simply add vectorially giving that the speed of light $\bar{c}$ in $\bar{S}$ is given by $\bar{c} = c - v$, thus violating the second postulate of relativity!

The resolution to the above conundrum is given by the famous Lorentz transformation that students first encounter in a Modern Physics course. Let $\bar{S}$ use coordinates $(c\bar{t}, \bar{x}, \bar{y}, \bar{z})$. Notice that c is not "barred". Then the time and space coordinates of the two observers are related by the transformation

$$\boxed{\begin{array}{l} c\bar{t} = \gamma(ct - \beta x) \\ \bar{x} = \gamma(x - \beta ct) \\ \bar{y} = y \\ \bar{z} = z \end{array} \quad \text{where } \beta = v/c \text{ and } \gamma = \frac{1}{\sqrt{1-\beta^2}} .} \tag{1.1}$$

A few immediate remarks are in order. Note that the Lorentz transform is not valid when $\beta^2 \geq 1$ (we certainly don't want imaginary lengths and time). Therefore, we only allow for $\bar{S}$ to have a relative velocity below c. Further, when $\beta << 1$, we see that the Lorentz transformation reduces to its Galilean counterpart, i.e.,

$$\begin{array}{c} \bar{t} = t \\ \bar{x} = x - vt \\ \bar{y} = y \\ \bar{z} = z \end{array} . \tag{1.2}$$

Comparing the above equations, we see that, unlike Newtonian physics, here $\bar{t}$ is not equal to t; and this is the source of many of the interesting physics in special relativity.

Remark 1.1. Since the time coordinate undergoes a change under a Lorentz transformation, we can no longer treat it as an absolute quantity. I.e., we have to treat time and space on an equal footing. Thus, spacetime points are described by the 4-tuple (ct, x, y, z). We chose ct instead of t since ct has the dimension of length just like x, y, z.

Finally, it is important to note that there exists no meaningful derivation of the Lorentz transformation in the present context; however, we include a heuristic argument for the sake of completeness. It is only later, when we confront the powerful theorem by Zeeman [2] that we can conclude that we have the right result.

Observer $\bar{S}$ using coordinates $(\bar{t}, \bar{x}, \bar{y}, \bar{z})$ is moving with a speed v along the x-axis of another inertial observer S using coordinates (t, x, y, z). We want a transformation between the two coordinate systems such that both observers measure the same speed of propagation for a beam of light. Since relative motion is restricted to the x direction, we will set

$$\bar{y} = y, \quad \text{and} \quad \bar{z} = z \, .$$

As a first guess, we assume that

$$\bar{x} = \gamma(x - vt) \, . \tag{1.3}$$

Here γ is set to be a constant depended on the parameters of the transformation, namely c and v, that is yet to be determined. The above expression has the advantage that when $\gamma \to 1$, we get the familiar Galilean coordinate transformation of coordinates. The symmetry between the frames suggests that when $v \to -v$, the barred and unbarred coordinates interchange. This assumption tacitly implies that γ is an even function of v. Then, eq.(1.3) also implies that

$$x = \gamma(\bar{x} + v\bar{t}) \, . \tag{1.4}$$

Combining the above two equations yields

$$\bar{x} = \gamma(\gamma\bar{x} + \gamma v\bar{t}) \; - \gamma vt \, ,$$

or

$$t = \gamma\bar{t} \; + \frac{(\gamma^2 - 1)}{\gamma v}\bar{x} \, .$$

Once again, swapping the barred and unbarred coordinates and substituting $v \to -v$, we get that

$$\bar{t} = \gamma t \; + \frac{(1 - \gamma^2)}{\gamma v}x \, . \tag{1.5}$$

Now, imposing the requirement that, for the location of the tip of a beam of light

$$x = ct \implies \bar{x} = c\bar{t}$$

in eq.(1.3) we get that

$$\bar{t} = \gamma t(1 - \beta) \tag{1.6}$$

where,

$$\beta = \frac{v}{c} \, .$$

Unlike eqs.(1.3)-(1.5), eq.(1.6) is not generally valid. It is only true along the tip of the beam. Nonetheless, it enables us to fix γ. Inserting eq.(1.6) into eq.(1.5) yields,

$$\gamma t \ - \gamma\beta t = \gamma t + \frac{(1-\gamma^2)}{\gamma v} ct \ ,$$

which gives the final needed expression for γ as

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \ .$$

Note that $\beta < 1$ for real and finite transformations. Finally, inserting eq.(1.3) into eq.(1.5) gives us that

$$\bar{t} = \gamma \left(t - \frac{v}{c^2} x \right) \ .$$

To recap, the Lorentz transformation describing a boost along the x direction is given by

$$\bar{x} = \gamma(x - vt) \ , \bar{t} = \gamma \left(t - \frac{v}{c^2} x \right) \ ,$$

$$\bar{y} = y \ , \ \text{ and } \ \bar{z} = z \ ,$$

where

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \ , \ \text{ and } \ \beta = \frac{v}{c} \ .$$

Remark 1.2. It is only for convenience that we chose to derive the Lorentz transformation along the x axis. It is just as easily done along any axis, and we will have occasion to see this explicitly.

Problem 1.1. Show that the inverse Lorentz transformation is given by

$$\begin{array}{c} ct = \gamma(c\bar{t} + \beta\bar{x}) \\ x = \gamma(\bar{x} + \beta c\bar{t}) \\ \bar{y} = y \\ \bar{z} = z \end{array} \ . \tag{1.7}$$

Problem 1.2. Suppose a beam of light is sent from the spacetime point (ct_1, x_1, y_1, z_1) to (ct_2, x_2, y_2, z_2) in S, then we must have that

$$c^2(t_2 - t_1)^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 \ .$$

Show that in $\bar{S}$ the two spacetime coordinate components must also satisfy

$$c^2(\bar{t}_2 - \bar{t}_1)^2 = (\bar{x}_2 - \bar{x}_1)^2 + (\bar{y}_2 - \bar{y}_1)^2 + (\bar{z}_2 - \bar{z}_1)^2 \ .$$

Here, the barred coordinates are related to their unbarred coordinates by the Lorentz transformation in eq.(1.1).

The above problem does indeed verify that the speed of light does not change in a coordinate system that is moving with a constant velocity with respect to another inertial frame. For simplicity, we chose the direction of the moving frame to be in the x direction. Clearly, there is no need for such a restriction. Such frames (/observers) are in general referred to as **Lorentz Boosted** frames (/observers). One might even refer to the transformation in eq.(1.1) as a **Lorentz Boost**. The number γ in eq.(1.1) is often referred to as the **Lorentz factor**.

To understand the nature of space and time of a boosted observer, as compared to another inertial observer, we begin our analysis be constructing spacetime diagrams in the t, x plane. It is customary to graph x as the horizontal axis and ct as the vertical axis (the fact that these axes are chosen to be perpendicular has no physical significance; it is done only for ease). We will place both S and $\bar{S}$ coordinate grids in the same spacetime diagram to fully appreciate the consequences of a Lorentz boost. To this end, note that the $\bar{t}$ axis of $\bar{S}$ is the set of all points on which $\bar{x} = 0$, i.e., we set $\gamma(x - vt) = 0$. Here, as usual $\gamma = 1/\sqrt{1-\beta^2}$ and $\beta = v/c$. Therefore, the $c\bar{t}$ axis is the line

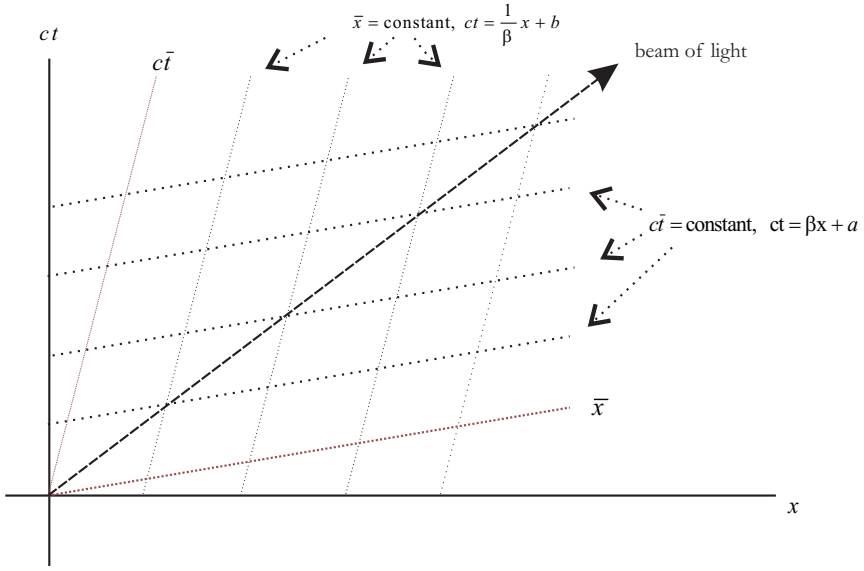


Fig. 1.1. Slanted dashed lines are lines of constant $c\bar{t}$ and $\bar{x}$. Notice how the beam of light has the same speed in both frames.

$$ct = \frac{1}{\beta}x \ .$$

Similarly, the $\bar{x}$ axis is obtained by setting $c\bar{t} = 0 = \gamma(ct - \beta x)$ and is given by the line

$$ct = \beta x \ ,$$

and the lines $c\bar{t}$ = constant (horizontal lines in the $\bar{S}$ frame) are given by

$$ct = \beta x + a$$

in the frame S. Here, different constants a generate different lines of constant $c\bar{t}$. These lines are parallel to the $\bar{x}$ axis.

The lines $\bar{x}$ = const (vertical lines in the $\bar{S}$ frame) are given by lines

$$ct = \frac{1}{\beta}x + b$$

Here, different constants b generate different lines of constant $\bar{x}$. These lines are parallel to the $c\bar{t}$ axis. The resulting diagram is represented in fig.(1.1). The observer in S lives on the line $x = 0$, i.e, the ct axis, while the observer in $\bar{S}$ lives on the line $\bar{x} = 0$, i.e., the $c\bar{t}$ axis. The fact that different observers live on different curves in spacetime leads to the famous time dilation and length contraction effects. Such special relativistic effects are meaningful only when β is significant and γ is greater than 1. In practice, this happens when $\beta \geq .1$ or equivalently when $v \geq .1c$.

Time Dilation

Example 1.1. With respect to a stationary inertial observer on earth (we will treat non-moving observers on earth as practically inertial), a rocket ship flies from a spacetime point $p_1 = (0, 0, 0, 0)$ on earth to the point $p_2 = (c\,T, d, 0, 0)$ to a distant galaxy. What is the lapsed time for an inertial observer that is fixed on the rocket ship? Provide the relevant spacetime diagram.

Solution: For ease of calculation, we will ignore the initial acceleration and the final deceleration of the rocket ship, as the time taken for launch and landing is minuscule compared to an intergalactic journey!

Let S and $\bar{S}$ be the inertial frames used by the observer on Earth and the rocket, respectively. As measured by S on Earth, the rocket ship starts at

$$t_1 = 0 \ \text{ and } \ x_1 = 0 \ ,$$

and at its destination has coordinate values

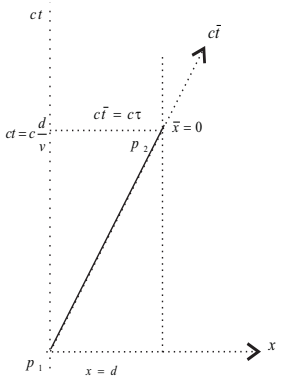


Fig. 1.2. The solid represents the path taken by the rocket ship.

$$t_2 = T \text{ and } x_2 = d = vT \ ,$$

where v is the speed of the rocket ship. On the rocket ship however, $\bar{S}$ locates the starting point $\bar{p}_1$ as

$$\bar{t}_1 = 0 \text{ and } \bar{x}_1 = 0 \ ,$$

and the time coordinate of the final destination as

$$\bar{t}_2 = \gamma \left(t_2 - \frac{v}{c^2} x_2 \right) = \gamma \left(t - \frac{v}{c^2} vT \right) = \sqrt{1-\beta^2}\, T \ .$$

Or

$$\bar{t}_2 = \frac{T}{\gamma} \ .$$

I.e., the observer on the rocket ship has aged less than the observer on Earth. This is an example of the well known time dilation effect in special relativity. As expected

$$\bar{x}_2 = \gamma(x_2 - vt_2) = \gamma(d - d) = 0 \ .$$

The relevant spacetime diagram for this example is given in fig.(1.2). □

Definition 1.1. *If a spacetime event occurs at the same spatial location* [1] *in a particular inertial frame, then the time lapsed for the event in this frame is called the* **proper time** *(denoted by τ) for this event. The frame itself is referred to as* **rest frame** *for the event.*

For the example considered above, the proper time lapsed for the trip $\tau = \Delta\bar{t}$. We will subsequently generalize the definition of proper time to include circumstances where there are no inertial rest frames during the entire span of the event. This will be the case, for example, in the initial and final stages of flight when the rocket ship is accelerating and or decelerating.

Length Contraction

Definition 1.2. *The length of an object measured in a frame where the object is at rest is referred to as its proper length.*

Example 1.2. A (really fast) jet airplane of proper length L_0 passes you by at a speed of βc. According to you, what is the length of the airplane? How long does it take for the airplane to pass you?

Solution On the airplane, in $\bar{S}$, the front end traces out a curve in spacetime denoted by $(c\bar{t}_2, \bar{x}_2)$, and so does that back end as described by $(c\bar{t}_1, \bar{x}_1)$. Note that we suppress the y and z coordinates for obvious reasons. Clearly, $\bar{S}$ measures the proper length to be

$$\Delta\bar{x} = \bar{x}_2 - \bar{x}_1 = L_0 \ .$$

When we transform the end point spacetime locations to your frame of reference S, the expressions become

$$\Delta x = \gamma(\Delta\bar{x} + v\Delta\bar{t}) \tag{1.8}$$

[1] By the same spatial location we mean x^i = constant for $i = 1, 2, 3$ in the particular inertial frame.

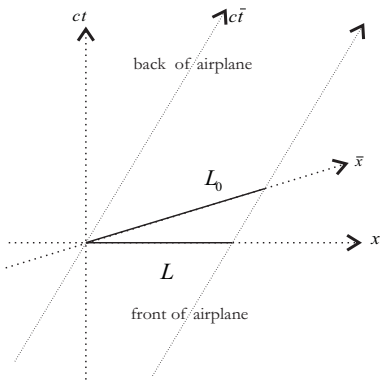


Fig. 1.3. The solid lines show the proper length L_0, and L is the length as measured in the frame S.

and

$$c\Delta t = \gamma(c\Delta\bar{t} + \beta\Delta\bar{x}) \ . \tag{1.9}$$

Note that Δ is the change in coordinate values from the front and the back end of the airplane. Also, we have used the inverse Lorentz transform (as we must). When you measure the length of the airplane in S, you locate the end points at the same instant of time in your frame (that's the simplest way to measure the length of a moving object). I.e., $c\Delta t = 0$ or from eq.(1.8)

$$c\Delta\bar{t} = -\beta\Delta\bar{x} \ .$$

Then, from eq.(1.9), the length L that you measure is given by

$$L = \Delta x = \gamma(\Delta\bar{x} + v\Delta\bar{t}) = \gamma\left(\Delta\bar{x} - \frac{v^2}{c^2}\Delta\bar{x}\right) = \frac{\Delta\bar{x}}{\gamma} \ ,$$

or

$$\boxed{L = \frac{L_0}{\gamma}} \ .$$

I.e., the length of the airplane appears contracted to you. The time taken for the airplane to pass you is given by

$$\Delta t = \frac{\text{distance}}{v} = \frac{L_0\sqrt{1-\beta^2}}{v} \ .$$

Fig.(1.3) shows the relevant spacetime diagram. □

The following example is from [3].

Example 1.3. A train of rest length $200m$ travels along a straight line of track past a train station of rest length $100m$ with $\beta = \sqrt{3}/2$. What does the engineer on the train measure the length of the train station to be? What does the station master measure the length of the train to be?

Solution Here $\gamma = 2$.

According to the station master:
Length of the station $= 100m$
Length of the train $= 200/2 = 100m$

According to the engineer on the train:
Length of the train $= 200m$
Length of the station $= 100/2 = 50m$

The relevant spacetime diagram is shown in fig.(1.4). Here we see that two views can co-exist harmoniously. □

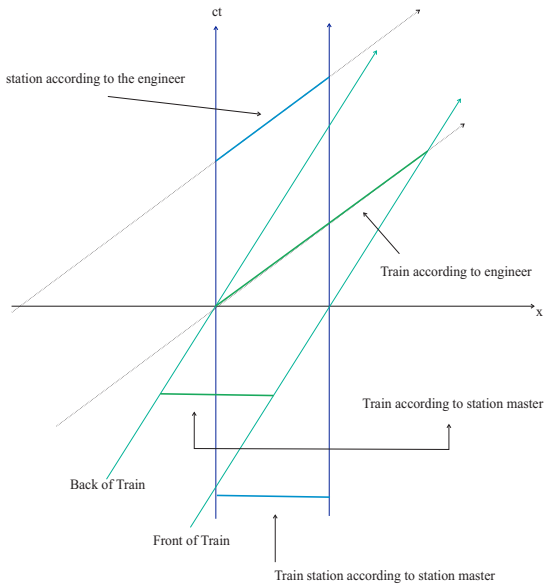


Fig. 1.4. The station is shown in blue, and the train is shown in green.

This portion of the homework is from a standard course in military science that every school child has to take in the Post Imperial Era [2]

Problem 1.3. You have been posted in the outer rim territory to monitor rebel activity in the sector. Near the end of a quiet shift, the Tantive IV streaks past, possibly carrying the stolen plans to the Death Star! The Imperial fleet must determine the accurate speed of the craft to track the Tantive IV. Your scans indicate that the spacecraft has a length of 135 m. Imperial records show the craft is 150 m long. In transmitting your report to the headquarters, what speed should you give for the spacecraft?

Problem 1.4. Anakin Skywalker dueled Count Dooku on board the Invisible Hand. During the duel, the Invisible Hand zoomed past the Integrity at a speed of .9c. Lieutenant Commander Needa looked through the window and witnessed part of the duel. If Anakin was holding his meter-long lightsaber at an angle of 30 degrees, how long does the saber appear to Needa? According to Needa, what angle does the lightsaber subtend (to the common horizontal in the direction of relative motion)?

Problem 1.5. Following the Rebel Alliance's smashing victory at the Battle of Yavin, Luke Skywalker (he was 19 years old at the time) flew his X-wing fighter to Hoth to scout its suitability for housing the new rebel base. Unfortunately, 4 light years into the trip (as measured by the rebel base) Luke discovered an Imperial blockade and that forced him to return immediately to Yavin, the round trip taking 6 years according to Luke, (a) how fast did he travel? (b) How old was Princess Leia, who was located at the base, when Luke returned?

Hint: Write 1 light year as c times 1 year. I.e., 8 lyrs = 8 c years.

Problem 1.6. Bonus 3 pts: Jacen Solo, the oldest son of Han and Leia, boards a spaceship and travel away from Coruscant toward Leritor, a mid rim world near the Bothan sector at a constant velocity of 0.85c. One year later on Coruscant clocks, Jacen's twin sister, Jaina, boards a second spaceship and follows Jacen at a constant velocity of 0.95c in the same direction. (a) When Jaina catches up with Jacen, what will be the differences in their (proper) ages? (b) Which twin will be older?

[2] Problems in special relativity were transformed into the world of "Star Wars" by Michael Cervera (Adjunct Lecturer, Department of History, Troy University). This storyline follows the expanded Star Wars universe and not the third theatrical trilogy.

2

Inertial Observers

In this chapter, we will classify the list of all possible inertial observers. This is important so that we can recognize a viable physical theory from an unphysical one. After all, according to the second postulate of relativity, any physical theory must have the same appearance in all inertial frames. So, we could select a candidate transformation for each member of an inertial class and subject the test theory to its transformation properties. If the new theory survives this test, then at the very least we have a plausible theory. In the next chapter, we will subject mechanics and electrodynamics to this test, and you will get a concrete sense of what "same appearance in all inertial frames" might mean. In practice, what we do is modify a theory if necessary so that the postulates do indeed hold.

It might seem odd that physicists consider the above-mentioned test as an important one. But, soon you will see that this has far-reaching consequences. In fact, just this test in special relativity determines all the possible types of particles that might exist in the universe [1]. It also determines the types of mathematical expressions that are allowed in a field theory. In this chapter, however, we will not consider the second postulate of relativity. Our sole attention will focus on the constancy of the speed of light. Turns out, this will modify most all previously existing equations of physics. Any correct physical theory must be recast in its relativistic form. We will take this issue up in the following chapter.

Before we enter into our classification scheme, it will be useful to have an understanding of the basic properties of the rotation group. After all, observers who are rotated with respect to an inertial observer must also be inertial (they would clearly measure the speed of light to be c).

2.1 The Rotation Group

Consider the simple case of rotation of an orthonormal bases frame about the z-axis. Here

$$\begin{pmatrix} \bar{\mathbf{e}}_x \\ \bar{\mathbf{e}}_y \\ \bar{\mathbf{e}}_z \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{e}_x \\ \mathbf{e}_y \\ \mathbf{e}_z \end{pmatrix} . \tag{2.1}$$

The rotation matrix about the z-axis about an angle φ denoted by $R_z(\varphi)$ is then given by

$$R_z(\varphi) = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

It is easy to verify some of the immediate properties of $R_z(\varphi)$:

$$[R_z(\varphi)]^{-1} = R_z(-\varphi) ,$$

$$\det R_z(\varphi) = 1 ,$$

and

$$[R_z(\varphi)]^T \cdot R_z(\varphi) = \mathbf{I} ,$$

where $\mathbf{I}$ is the identity matrix. Now, let $\mathbf{v}$ be any 3-dimensional vector given by

[1] This is beyond the scope of our purpose here. Most textbooks on Quantum Field Theory will explain the classification of particle types. For example, see [4].

$$\mathbf{v} = v_x\,\mathbf{e}_x + v_y\,\mathbf{e}_y + v_z\,\mathbf{e}_z\ .$$

The same physical vector in the rotated frame can be written as

$$\mathbf{v} = \bar{v}_x\,\bar{\mathbf{e}}_x + \bar{v}_y\,\bar{\mathbf{e}}_y + \bar{v}_z\,\bar{\mathbf{e}}_z\ .$$

Notice what we are doing is a *passive rotation*. The physical vector remains unchanged while the components and the bases change simultaneously. I.e.,

$$\bar{\mathbf{v}} = \mathbf{v}\ .$$

Under a rotation about the z-axis, we already know how the bases vectors in the two frames are related to each other (eq.(2.1)). It will be useful to work out the transformation between the components of the vector $\mathbf{v}$ in the two frames as well. To this end note that

$$\begin{aligned}\mathbf{v} &= \bar{v}_x\,\bar{\mathbf{e}}_x + \bar{v}_y\,\bar{\mathbf{e}}_y + \bar{v}_z\,\bar{\mathbf{e}}_z\\ &= \bar{v}_x\,(\cos\varphi\,\mathbf{e}_x\ + \sin\varphi\,\mathbf{e}_y)\ + \bar{v}_y\,(-\sin\varphi\,\mathbf{e}_x + \cos\varphi\,\mathbf{e}_y)\ + \bar{z}\,\mathbf{e}_z\\ &= (\cos\varphi\,\bar{v}_x\ - \sin\varphi\,\bar{v}_y)\,\mathbf{e}_x\ + (\sin\varphi\,\bar{v}_x\ + \cos\varphi\,\bar{v}_y)\,\mathbf{e}_y + \bar{z}\,\mathbf{e}_z\ .\end{aligned}$$

I.e.,

$$\begin{pmatrix} v_x\\ v_y\\ v_z\end{pmatrix} = \begin{pmatrix}\cos\varphi & -\sin\varphi & 0\\ \sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\bar{v}_x\\ \bar{v}_y\\ \bar{v}_z\end{pmatrix}\ .$$

Noting that the inverse of the above matrix is its transpose we get that

$$\begin{pmatrix} \bar{v}_x\\ \bar{v}_y\\ \bar{v}_z\end{pmatrix} = \begin{pmatrix}\cos\varphi & \sin\varphi & 0\\ -\sin\varphi & \cos\varphi & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}v_x\\ v_y\\ v_z\end{pmatrix}\ .$$

Curiously, here, under rotation the components of a vector change exactly the same way as the bases vector (eq.(2.1))! It is important to note that this is not true under any other types of bases transformation save rotation. We will see the reason for this shortly (problem 2.2).

Theorem 2.1. *Given any vector* $\mathbf{v}$ *in a particular frame, the bases can be rotated so that the new basis,* $\mathbf{v}$ *has only components in x-direction.*

Proof. The argument is very simple and hardly unique. Let θ_0 and φ_0 be the polar and azimuthal angle coordinates of the vector $\mathbf{v}$ [2]. Then, first rotate the bases about the z-axis through φ_0 and then rotate about the new y-axis by an angle of $\pi/2 - \theta$ when $\theta < \pi/2$ and by $-(\theta - \pi/2)$ otherwise. □

Example 2.1. Consider a vector $\mathbf{v}$ with magnitude 2, $\varphi = 90^0$ and $\theta = 30^0$. Find a bases where $\mathbf{v}$ only has an x component.

Solution: Note $\mathbf{v} = \mathbf{e}_y + \sqrt{3}\,\mathbf{e}_z$. As per the proof above we first rotate about z-axis through 90^0. The rotation matrix in this case is

$$R_z(\pi/2) = \begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\ .$$

Consequently the new basis is given by

$$\bar{\mathbf{e}}_x = e_y\ ,\ \ \bar{\mathbf{e}}_y = -e_x\ \text{ and }\ \bar{\mathbf{e}}_z = e_z\ .$$

In this basis, $\mathbf{v} = \bar{\mathbf{e}}_x + \sqrt{3}\,\mathbf{e}_z$. Now rotate about $\bar{\mathbf{e}}_y$ through 60^0. I.e.,

[2] Our choice of spherical coordinates are such that

$$x = r\cos\varphi\sin\theta\ ,$$

$$y = r\sin\varphi\sin\theta$$

and

$$z = r\cos\theta\ .$$

Here θ and φ are the polar and azimuthal angles respectively.

$$\begin{pmatrix}\tilde{\mathbf{e}}_x\\ \tilde{\mathbf{e}}_y\\ \tilde{\mathbf{e}}_z\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & 0 & \frac{\sqrt{3}}{2}\\ 0 & 1 & 0\\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\bar{\mathbf{e}}_x\\ \bar{\mathbf{e}}_y\\ \bar{\mathbf{e}}_z\end{pmatrix} .$$

Then

$$\begin{pmatrix}\tilde{v}_x\\ \tilde{v}_y\\ \tilde{v}_z\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & 0 & \frac{\sqrt{3}}{2}\\ 0 & 1 & 0\\ \frac{-\sqrt{3}}{2} & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}1\\ 0\\ \sqrt{3}\end{pmatrix} ,$$

and

$$\mathbf{v} = 2\,\tilde{\mathbf{e}}_x$$

where

$$\begin{pmatrix}\tilde{\mathbf{e}}_x\\ \tilde{\mathbf{e}}_y\\ \tilde{\mathbf{e}}_z\end{pmatrix} = \begin{pmatrix}\frac{1}{2} & 0 & \frac{\sqrt{3}}{2}\\ 0 & 1 & 0\\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\mathbf{e}_x\\ \mathbf{e}_y\\ \mathbf{e}_z\end{pmatrix} .$$

■

Problem 2.1. Let

$$R = \begin{pmatrix}\frac{1}{2} & 0 & \frac{\sqrt{3}}{2}\\ 0 & 1 & 0\\ -\frac{\sqrt{3}}{2} & 0 & \frac{1}{2}\end{pmatrix}\begin{pmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 1\end{pmatrix} .$$

Show that

$$\det R = 1 , \quad \text{and} \quad R^T R = \mathbf{I} .$$

So far we considered simple cases of very specific rotations. However, to understand rotations as abstract objects let us consider its defining properties.

Definition 2.1. *A transformation of vectors is a rotation R if and only if*

1. *R is linear. Clearly we would want that for any pair of vectors $\mathbf{v}$ and $\mathbf{w}$, $R(\mathbf{v}+\mathbf{w}) = R(\mathbf{v}) + R(\mathbf{w})$, and $R(a\mathbf{v}) = a\,R(\mathbf{v}))$ for any real number a.*
2. *R should preserve the lengths of vectors.*
3. *R should preserve the angle between any pair of vectors.*
4. *R should preserve the relative orientation of vectors; in particular for a orthonormal bases set, since*

$$\mathbf{e}_x \times \mathbf{e}_y = \mathbf{e}_z ,$$

we must have that

$$\bar{\mathbf{e}}_x \times \bar{\mathbf{e}}_y = \bar{\mathbf{e}}_z .$$

Here $\bar{\mathbf{e}}_x = R(\mathbf{e}_x)$, and similarly for the other basis vectors.

As a reminder, what we are doing here is a passive rotation as described previously. Property 1 implies that R can be written as a matrix. Since we can reconstruct how every vector transforms once we know how the basis vectors transform, we will specify R by its defining equations on an orthonormal basis:

$$\begin{pmatrix}\bar{\mathbf{e}}_x\\ \bar{\mathbf{e}}_y\\ \bar{\mathbf{e}}_z\end{pmatrix} = \begin{pmatrix}R_{11} & R_{12} & R_{13}\\ R_{21} & R_{22} & R_{23}\\ R_{31} & R_{32} & R_{33}\end{pmatrix}\begin{pmatrix}\mathbf{e}_x\\ \mathbf{e}_y\\ \mathbf{e}_z\end{pmatrix} . \tag{2.2}$$

Theorem 2.2. *If R is a rotation, then $R^T R = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix.*

Proof. Under rotation, the familiar dot product of vectors is unchanged (properties 2 and 3 above). I.e., for any pairs of vectors $\mathbf{v}$ and $\mathbf{w}$

$$\mathbf{v}\cdot\mathbf{w} = \sum_{i=1}^{3} v^i\, w^i = \sum_{i=1}^{3} \bar{v}^i\, \bar{w}^i ,$$

where

$$\bar{v}^i = \sum_{j=1}^{3} R_{ij}\ v^j \ ^{3}$$

for a rotation matrix R_{ij} . The vector $\mathbf{w}$ transforms in a similar manner. Then

$$\mathbf{v}\cdot\mathbf{w} = \sum_{i=1}^{3} \bar{v}^i\ \bar{w}^i = \sum_{ijk=1}^{3} R_{ij}\ v^j\ R_{ik}\ w^k = \sum_{jk}^{3} (R^T R)_{jk}\ v^j\ w^k = \sum_{i=1}^{3} v^i\ w^i\ .$$

This can happen for every pair of vectors $\mathbf{x}$ and $\mathbf{y}$ if and only if

$$(R^T R)_{kj} = \mathbf{I}\ .$$

□

Problem 2.2. Suppose under a rotation, the components of any vector $\mathbf{v}$ change according to the transformation

$$\bar{v}^i = \sum_{j=1}^{3} R_{ij}\ v^j\ ,$$

then show that the basis vectors change exactly the same way, i.e.,

$$\bar{\mathbf{e}}_i = \sum_{j=1}^{3} R_{ij}\ \mathbf{e}_j\ .$$

Theorem 2.3. *If R is a rotation, then* $\det R = 1$.

Proof. From property 4 of rotations and eq.(2.2),

$$1 = (\bar{\mathbf{e}}_x \times \bar{\mathbf{e}}_y)\cdot\bar{\mathbf{e}}_z = \det R\ .$$

This immediately gives that $\det R = 1$. □

Theorem 2.4. *R is a rotation* [4]*, if and only if*

$$\boxed{\det R = 1 \text{ and } \ R^T R = \mathbf{I}}\ .$$

Proof. The above two theorems satisfy all the requirements of rotation in definition 2.1. □

Problem 2.3. Show that the set of all rotations form a group. I.e., show that

1. $R = \mathbf{I}$, the identity matrix, is a rotation.
2. if R_1 and R_2 are rotations, then $R_1 R_2$ is a rotation.
3. if R is a rotation, then R^{-1} is a rotation.

Hint: Recall that for matrices A and B,

$$\det(AB) = (\det A)(\det B)\ ,$$

and

$$(AB)^T = B^T A^T\ .$$

Definition 2.2. *From the theorem above, the set of all matrices R such that* $\det R = 1$*, and $R^T R = 1$ forms an algebraic group* [5] *under matrix multiplication called the rotation group, and is denoted as $SO(3)$.*

The letters S stand for special, meaning determinant 1, or orientation preserving, and the letter O stands for orthogonal, meaning $R^T R = 1$ since they carry orthonormal bases to another set of orthonormal bases.

[3] Here we have assumed that the bases vectors $(\mathbf{e}_x, \mathbf{e}_y, \mathbf{e}_z)$ and transform exactly as the components of a vector (v^1, v^2, v^3). The next problem shows that this is indeed the case.

[4] As defined in eq.(2.2).

[5] If you are not familiar with Group Theory in Algebra, here we simply mean that conditions of problem 2.3 hold true.

2.2 Classes of Inertial Observers

There are four very distinct types of coordinate transformations that relate one inertial observer to another. For definiteness, consider one inertial frame/observer S that uses coordinates (ct, x, y, z). While the 3-tuple (x, y, z) is the Cartesian coordinates of a particle, we will never treat it as a vector (you will see soon that in differential geometry this is typical). However, when we have a Cartesian coordinate system (and only then), the difference in location between two spatial points (x_2, y_2, z_2) and (x_1, y_1, z_1) is denoted by

$$\Delta\mathbf{x} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$$

is considered to be a vector located at the point (x_1, y_1, z_1).

Time Translated Observer

Define a new set of coordinates, presumably carried by another observer, given by

$$\bar{t} = t + t_0 \ , \quad \text{and} \quad (\bar{x}, \bar{y}, \bar{z}) = (x, y, z)$$

for some constant t_0. Here, the two observers have not initially synchronised their otherwise identical clocks. Then the speed of a beam of light in S

$$\left| \frac{\Delta\mathbf{x}}{\Delta t} \right| = c$$

in turn implies that

$$\left| \frac{\Delta\bar{\mathbf{x}}}{\Delta\bar{t}} \right| = c \ .$$

Clearly this is so because $\Delta\bar{\mathbf{x}} = \Delta\mathbf{x}$ and $\Delta\bar{t} = \Delta t$. Here $|\cdot|$ calculates the usual magnitude of a vector $\mathbf{x}$.

Space Translated Observer

Now consider a new class of observers defined by

$$\bar{t} = t \ , \quad \text{and} \quad (\bar{x}, \bar{y}, \bar{z}) = (x + a_x, y + a_y, z + a_z)$$

Here a_x, a_y, and a_z are 3 fixed real numbers. Here too, for the exact same reason as above,

$$\left| \frac{\Delta\mathbf{x}}{\Delta t} \right| = c$$

implies that

$$\left| \frac{\Delta\bar{\mathbf{x}}}{\Delta\bar{t}} \right| = c \ .$$

Rotated observer

Now consider a class of observers defined by

$$\bar{t} = \ t \ , \quad \text{and} \quad \bar{x}^i = \sum_{i=1}^{3} R_{ij} \ x^j$$

for some rotation matrix R. Since the entries of the matrix R are constant

$$\frac{d\bar{x}^i}{dt} = \frac{d}{dt} \sum_{i=1}^{3} R_{ij} \ x^j = \sum_{i=1}^{3} R_{ij} \ \frac{dx^j}{dt} \ .$$

From properties 2 and 3 of definition 2.1, magnitudes of vectors do not change under rotation and so

$$\left| \frac{d\mathbf{x}}{dt} \right| = c \rightarrow \left| \frac{d\bar{\mathbf{x}}}{dt} \right| = c \ .$$

Therefore, a rotated observer measures the speed of light to be c just as S.

Lorentz Transformed Observer

It is important to mention the very significant new class of observers that are related by a Lorentz transformation here (eq. (1.1)). In the previous chapter, we ensured that the speed of light is indeed preserved under a Lorentz transformation. Naturally, we mean to include boosts in all directions.

2.3 Minkowski Spacetime

The coordinates (ct, x, y, z) constructed by clocks and light beams as described in the beginning of Chapter 1, where along a light beam, the change in coordinates is such that

$$c^2(\Delta t)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

is referred to as **Minkowski coordinates**. These coordinates are most "natural" to special relativity. Minkowski coordinates are usually labeled by Greek indices as $\{x^\mu\}$, where $\mu = 0, 1, 2, 3$ and $x^0 = ct, x^1 = x, x^2 = y$ and $x^3 = z$. Purely spatial objects are indicated by Latin indices $\{x^i\}$, where $i = 1, 2, 3$ and $x^1 = x, x^2 = y$ and $x^3 = z$.

In the previous section, we constructed four distinct classes of inertial observers. For compactness in notation, we can combine all of the transformations in these inertial frames into a single equation, which we write as

$$\bar{x}^\mu = \Lambda^\mu{}_\nu \, x^\nu + a^\mu \, , \tag{2.3}$$

where a^μ is a constant 4-tuple that could include a time and space translation, and the matrix $\Lambda^\mu{}_\nu$ includes both rotations and Lorentz transformations about arbitrary directions. [6] While the 4-tuples themselves are not vectors, the difference in coordinates of two points, x_2^μ and x_1^μ, defined by

$$\Delta x^\mu = x_2^\mu - x_1^\mu$$

is to treated as a vector located at the point x_1^μ. Then under a transformation given in eq.(2.3),

$$\Delta \bar{x}^\mu = \Lambda^\mu{}_\nu \, \Delta x^\nu \, . \tag{2.4}$$

In fact, the above equation defines what we mean by a vector; namely, *any object that transform according to eq.(2.4) under a coordinate transformation in eq.(2.3) is called a 4-vector*. Therefore, under inertial transformations, 4-vectors are only affected by boosts and rotations. Although Δx^μ is not invariant under $\Lambda^\mu{}_\nu$, it turns out that a "special" combination of components of Δx^μ are invariant under eq.(2.3).

---

[6] Always, the first index of a matrix, (μ in this case) denotes the row, and the second index (ν in this case) denotes the column of the matrix. Explicitly

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \cdot & \cdot \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \Lambda^3{}_3 \end{pmatrix} \, .$$

Additionally, we are using the Einstein summation convention, namely repeated upper and lower indices are being summed over. Explicitly

$$\Lambda^\mu{}_\nu \, x^\nu = \Lambda^\mu{}_0 \, x^0 + \Lambda^\mu{}_1 \, x^1 + \Lambda^\mu{}_2 \, x^2 + \Lambda^\mu{}_3 \, x^3$$

for each value of μ. The Einstein summation convention is to be understood throughout the remainder of the text. Any exceptions will be stated as such. When $\Lambda^\mu{}_\nu$ is a rotation, it will take the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & R_{ij} & \\ 0 & & & \end{pmatrix} \, ,$$

where R is a 3×3 rotation matrix, and for example, when $\Lambda^\mu{}_\nu$ is a boost along the x-direction, it will take the form

$$\Lambda^\mu{}_\nu = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \, .$$

Theorem 2.5. *Under eq.(2.3), the quantity*

$$-(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

is an invariant for rotations, time and space translations, and boosts along x axis. Here $(\Delta ct, \Delta x, \Delta y, \Delta z)$ are the components of a 4-vector.

Proof. Under space-time translations, the result is trivially true, since the components of a 4-vector remains unaffected. Under rotations

$$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (\Delta \bar{x})^2 + (\Delta \bar{y})^2 + (\Delta \bar{z})^2$$

and

$$\Delta c\bar{t} = \Delta ct \ .$$

Therefore, the above combination is again trivially true. Under a Lorentz transformation in the x-direction

$$\begin{aligned} &-(\Delta c\bar{t})^2 + (\Delta \bar{x})^2 + (\Delta \bar{y})^2 + (\Delta \bar{z})^2 \\ &= -\gamma^2[\Delta ct - \beta \Delta x]^2 + \gamma^2[\Delta x - \beta \Delta ct]^2 + (\Delta y)^2 + (\Delta z)^2 \\ &= \gamma^2[(\Delta x)^2(1-\beta^2) - (\Delta ct)^2(1-\beta^2)] + (\Delta y)^2 + (\Delta z)^2 \\ &= -(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \ . \end{aligned}$$

In a similar manner, it can be shown that

$$-(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

is invariant under arbitrary Lorentz transformations. □

Clearly, the notation Δx^μ is cumbersome. Moreover, as we shall see, there are 4-vectors other than coordinate differences. Therefore, we denote an arbitrary vector located at the point x in spacetime by A_x, B_x, X_x, Y_x etc. A 4-vector X_x at the point x is taken to mean

$$X_x \equiv (X^0, X^1, X^2, X^3) \equiv X^\mu_x \ .$$

Here X_x is the *abstract notation* for a vector, while in the *index notation* we write the same 4-vector as X^μ_x. Also, *Since spacetime is the basic object in special relativity, we will henceforth call 4-vectors simply as vectors.* If we are referring to the spatial components of a 4-vector in a particular frame, we will point it out by calling the object a 3-vector.

Remark 2.1. Tentatively, the set of points described by Minkowski coordinates $(ct, x, y, z) \in \mathbb{R}^4$ will be referred to as Minkowski spacetime. A precise definition with additional properties will follow shortly.

Problem 2.4. Consider any two vectors X^μ_x and Y^μ_x at the same spacetime point x. Show that the quantity

$$-X^0Y^0 + X^1Y^1 + X^2Y^2 + X^3Y^3$$

is an invariant under eq.(2.3), for rotations, time and space translations, and boosts along x axis.

We can use the invariant quantity to endow spacetime with an inner product η (in $\mathbb{R}^3$ we often call this the dot product). The problem above motivates the following definition.

Definition 2.3. *For any pair of vectors A^μ_x and B^μ_x at the spacetime point x, in Minkowski coordinates, define an inner product η by*

$$\eta(A_x, B_x) \equiv -A^0B^0 + A^1B^1 + A^2B^2 + A^3B^3 \ . \tag{2.5}$$

Notice that the inner product takes two vectors at the same spacetime point x to a real number. With a slight abuse of notation, it is also usual to write $\eta(A_x, B_x)$ as $A_x \cdot B_x$, and A^2_x is taken to mean $A_x \cdot A_x$. Define a matrix $\eta_{\mu\nu}$ by

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

Then, using the Einstein summation convention we write

$$\eta(A_x, B_x) = \eta_{\mu\nu} A_x^\mu B_x^\nu .$$

For any pair of vectors, the above quantity does not depend on the inertial frame. Note that $\eta^{-1} = \eta$, however we will denote η^{-1} as

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

so that

$$\eta^{\mu\alpha}\eta_{\alpha\nu} = \delta^\mu{}_\nu ,$$

where the Kronecker delta, $\delta^\mu{}_\nu$, is the identity matrix.

2.4 The Lorentz Group

We have shown that spacetime translations, rotations, and boosts preserve the inner product η. It is only natural to ask if we are missing any other linear transformations that will preserve η. To this end, consider any tentative linear transformations $\Lambda^\mu{}_\nu$. Then for any pair of vectors A and B, let $\bar{A}$, and $\bar{B}$ be defined by

$$\bar{A}^\mu = \Lambda^\mu{}_\nu A^\nu ,$$

and similarly for $\bar{B}$.

Theorem 2.6. *Let Λ be any linear transformation on vectors in Minkowski spacetime. Then Λ preserves the inner product η, i.e.,*

$$\eta(A, B) = \eta(\bar{A}, \bar{B})$$

if and only if

$$\eta = \Lambda^T \eta \Lambda .$$

Here $\bar{A}$ and $\bar{B}$ is as defined above.

Proof.

$$\eta_{\alpha\beta} A^\alpha B^\beta = \eta_{\mu\nu} \bar{A}^\mu \bar{B}^\nu = \eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta A^\alpha B^\beta .$$

Therefore

$$A^\alpha \left[\eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta - \eta_{\alpha\beta} \right] B^\beta = 0$$

for all A^α and B^β. This can happen only when

$$\eta_{\alpha\beta} = \eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta . \tag{2.6}$$

In matrix notation, the above equation reads

$$\eta = \Lambda^T \eta \Lambda .$$ [7]

□

Definition 2.4. *Any matrix that satisfies*

$$\eta = \Lambda^T \eta \Lambda \tag{2.7}$$

is called a general homogeneous Lorentz transformation. The set of all general homogeneous Lorentz transformations is denoted by $\mathcal{L}$.

[7]

$$\eta_{\mu\nu} \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta = \Lambda^\mu{}_\alpha \eta_{\mu\nu} \Lambda^\nu{}_\beta = \sum_\mu \left(\Lambda^T\right)^\alpha{}_\mu \eta_{\mu\nu} \Lambda^\nu{}_\beta .$$

The right hand side of the equation above is simply matrix multiplication.

Problem 2.5. Show that the boost in the x-direction given by eq.(1.1) is a general homogeneous Lorentz transformation.

Problem 2.6. Show that the matrix

$$\mathbb{T} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{2.8}$$

is a general homogeneous Lorentz transformation. $\mathbb{T}$ is the time-reversal transformation, and has the same entries as η, and $\mathbb{T}^2 = \mathbf{I}$

Problem 2.7. Show that the matrix

$$\mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} . \tag{2.9}$$

is a general homogeneous Lorentz transformation. $\mathbb{P}$ is called the parity transformation. $\mathbb{P}$ transforms a right-handed coordinate system into a left-handed one. Also, $\mathbb{P}^2 = \mathbf{I}$.

All general homogeneous Lorentz transformations will preserve the inner product η between any pair of vectors. Taking the determinant of the above equation, we also find that

$$\det \Lambda = \pm 1 .$$

Therefore, general homogeneous Lorentz transformations can be inverted as a matrix. Eq.(2.7) when multiplied by Λ^{-1} on both sides give

$$\Lambda^{-1} = \eta \, \Lambda^T \, \eta . \tag{2.10}$$

Theorem 2.7. *Define*

$$\Lambda_\nu{}^\mu \equiv \eta_{\nu\beta} \, \eta^{\mu\alpha} \, \Lambda^\beta{}_\alpha . \tag{2.11}$$

Then

$$(\Lambda^{-1})^\mu{}_\nu = \Lambda_\nu{}^\mu .$$

Proof.

$$(\Lambda^{-1})^\mu{}_\nu = (\eta \, \Lambda^T \, \eta)^\mu{}_\nu = \sum_{\alpha\beta} \eta^{\mu\alpha} \, (\Lambda^T)^\alpha{}_\beta \, \eta_{\beta\nu} = \eta^{\mu\alpha} \, \Lambda^\beta{}_\alpha \, \eta_{\beta\nu} = \eta_{\nu\beta} \, \eta^{\mu\alpha} \, \Lambda^\beta{}_\alpha .$$

□

Theorem 2.8. *Eq.(2.7) implies that*

$$\eta^{\mu\nu} = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \, \eta^{\alpha\beta} . \tag{2.12}$$

Proof.

$$\Lambda^a{}_m \, \Lambda_b{}^m = \delta^a{}_b .$$

I.e.,

$$\Lambda^a{}_m \, \eta_{\mu b} \, \eta^{\nu m} \, \Lambda^\mu{}_\nu = \delta^a{}_b .$$

Or,

$$\eta_{\mu b} \, (\Lambda^a{}_m \, \eta^{\nu m} \, \Lambda^\mu{}_\nu) = \delta^a{}_b .$$

Since $\eta_{\mu b}$ has a unique inverse, we have the necessary result. □

Problem 2.8. Show that $\mathcal{L}$ is a group. Hint: use eqs. (2.7) and (2.10).

Theorem 2.9. *If $\Lambda \in \mathcal{L}$, then*

$$(\Lambda^0{}_0)^2 = 1 + \sum_{k=1}^{3} (\Lambda^k{}_0)^2 = 1 + \sum_{k=1}^{3} (\Lambda^0{}_k)^2 . \tag{2.13}$$

Proof. By setting $\alpha, \beta = 0$ in eq.(2.6) we find that

$$\eta_{00} = -1 = -(\Lambda^0{}_0)^2 + \sum_{k=1}^{3} (\Lambda^k{}_0)^2 \ .$$

Similarly, eq.(2.12) implies that

$$\eta^{00} = -1 = -(\Lambda^0{}_0)^2 + \sum_{k=1}^{3} (\Lambda^0{}_k)^2 \ .$$

□

From eq.(2.13) we get that

$$\Lambda^0{}_0 \geq 1 \ , \quad \text{or} \quad \Lambda^0{}_0 \leq -1 \ .$$

Problem 2.9. Show that $\Lambda \in \mathcal{L}$ is a rotation if and only if $\Lambda^0{}_0 = 1$.

Definition 2.5. *If $\Lambda \in \mathcal{L}$ and $\Lambda^0{}_0 \geq 1$, we say that Λ is orthochronous, if $\Lambda^0{}_0 \leq -1$ it is non-orthochronous.*

Theorem 2.10. *The set of all orthochronous elements of $\mathcal{L}$ forms a subgroup of $\mathcal{L}$.*

Proof. Let Λ and $\bar{\Lambda}$ be two orthochronous elements of $\mathcal{L}$, then

$$(\Lambda\bar{\Lambda})^0{}_0 = \Lambda^0{}_0 \, \bar{\Lambda}^0{}_0 + \sum_i (\Lambda^0{}_i \bar{\Lambda}^i{}_0) \ .$$

From eq.(2.13), and the Schwartz inequality of 3-vectors, the above expression becomes

$$= + \sqrt{1 + \sum_i (\Lambda^0{}_k)^2} \sqrt{1 + \sum_i (\bar{\Lambda}^k{}_0)^2} + \sum_i (\Lambda^0{}_i \bar{\Lambda}^i{}_0)$$

$$\geq \sqrt{1 + \sum_i (\Lambda^0{}_k)^2} \sqrt{1 + \sum_i (\bar{\Lambda}^k{}_0)^2} - \sqrt{\sum_i (\Lambda^0{}_k)^2} \sqrt{\sum_i (\bar{\Lambda}^k{}_0)^2} \geq 0 \ .$$

I.e., $\Lambda \, \bar{\Lambda}$ is orthochronous.

Also,

$$\Lambda_0{}^0 = \eta_{0a} \, \eta^{0b} \, \Lambda^a{}_b = (-1)(-1)\Lambda^0{}_0 > 0 \ .$$

I.e., the inverse of an orthochronous matrix is orthochronous. □

Our inertial transformations thus far, described in eq.(2.3), all have the property that $\det \Lambda = +1$. But, there is no reason for us to ignore the case that includes a reflection of the spatial axis. Naturally, we excluded this case while classifying rotations, because they are not a rotation. However, an inertial observer could easily carry a spatial grid where $\mathbf{e}_x \times \mathbf{e}_y$ is what he/she calls the $-\mathbf{e}_z$ direction. This reassignment will not affect the speed of light.

Definition 2.6.

$$\mathcal{L}_+ = \Big\{ \Lambda | \ \eta = \Lambda^T \, \eta \, \Lambda, \text{ and } \det \Lambda = +1 \Big\} \ ,$$

$$\mathcal{L}_- = \Big\{ \Lambda | \ \eta = \Lambda^T \, \eta \, \Lambda, \text{ and } \det \Lambda = -1 \Big\} \ ,$$

$$\mathcal{L}_+^{\uparrow} = \Big\{ \Lambda | \Lambda \in \mathcal{L}_+, \ \Lambda^0{}_0 \geq 1 \Big\} \ ,$$

and

$$\mathcal{L}_+^{\downarrow} = \Big\{ \Lambda | \Lambda \in \mathcal{L}_+, \ \Lambda^0{}_0 \leq 1 \Big\} \ .$$

Problem 2.10. Verify that $\mathcal{L}_+^{\uparrow}$ forms a subgroup of $\mathcal{L}$.

Definition 2.7. *The proper* [8] *orthochronous homogeneous Lorentz group, denoted by* $\mathcal{L}_+^\uparrow$*, is the set of all orthochronous homogeneous Lorentz transformations with determinant* 1. $\mathcal{L}_+^\uparrow$ *is often called the Lorentz group for short.*

Problem 2.11. For any general homogeneous Lorentz transformation $\Lambda^\mu{}_\nu$,

$$T = (\Lambda^0{}_0, \Lambda^1{}_0, \Lambda^2{}_0, \Lambda^3{}_0) \ ,$$
$$X = (\Lambda^0{}_1, \Lambda^1{}_1, \Lambda^2{}_1, \Lambda^3{}_1) \ ,$$
$$Y = (\Lambda^0{}_2, \Lambda^1{}_2, \Lambda^2{}_2, \Lambda^3{}_2) \ ,$$

and

$$Z = (\Lambda^0{}_3, \Lambda^1{}_3, \Lambda^2{}_3, \Lambda^3{}_3) \ .$$

Show that

$$\eta(T,T) = -1 \ , \eta(X,X) = \eta(Y,Y) = \eta(Z,Z) = 1 \ ,$$

and

$$\eta(T,X) = \eta(T,Y) = \eta(T,Z) = \eta(X,Y) = \eta(X,Z) = \eta(Y,Z) = 0 \ .$$

Problem 2.12. Show that if a linear transformation of type

$$\Lambda = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

belongs to $\mathcal{L}_+^\uparrow$, then Λ is a boost in the + or − x-direction.

Theorem 2.11. *Let* Λ *be an arbitrary member of* $\mathcal{L}_+^\uparrow$*. Then there exists two rotations* $\mathcal{R}_1$ *and* $\mathcal{R}_2$ *such that*

$$\Lambda = \mathcal{R}_1 \ \ \Lambda_B(\beta_0) \ \ \mathcal{R}_2 \ .$$

Here, $\Lambda_B(\beta_0)$ *is a Lorentz boost in the* x*-direction with parameter* β_0*.*

Proof. If

$$\Lambda^1{}_0 = \Lambda^2{}_0 = \Lambda^3{}_0 = 0 \ ,$$

then the theorem is trivially true. Suppose this is not the case. Set

$$u_1 = \frac{(\Lambda^1{}_0, \ \Lambda^2{}_0, \ \Lambda^3{}_0)}{\sqrt{(\Lambda^1{}_0)^2 + (\Lambda^2{}_0)^2 + (\Lambda^3{}_0)^2}} \equiv (\alpha_1, \ \alpha_2, \ \alpha_3) \ .$$

Let

$$u_2 = (\beta_1, \beta_2, \beta_3)$$

and

$$u_3 = (\gamma_1, \gamma_2, \gamma_3)$$

be unit vectors such that $\{u_1, \ u_2, \ u_3\}$ is an orthonormal basis in $\mathbb{R}^3$. Note that $u_i^2 = 1$ for each i and $u_i \cdot u_j = 0$ for $i \neq j$. Therefore

$$(R_1)^{-1} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \\ \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix}$$

is an orthogonal matrix [9]. And by swapping the last pair of rows if necessary, we can ensure that the above matrix has a unit determinant. Therefore, we get the following rotation in $\mathbb{M}^4$:

[8] The term proper refers to the determinant of the transformation being +1.

[9] Since

$$(R_1)^{-1} \ (R_1^{-1})^T = \begin{bmatrix} (u_1 \cdot u_1) & (u_1 \cdot u_2) & (u_1 \cdot u_3) \\ (u_2 \cdot u_1) & (u_2 \cdot u_2) & (u_2 \cdot u_3) \\ (u_3 \cdot u_1) & (u_3 \cdot u_2) & (u_3 \cdot u_3) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \ .$$

$$(\mathcal{R}_1)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha_1 & \alpha_2 & \alpha_3 \\ 0 & \beta_1 & \beta_2 & \beta_3 \\ 0 & \gamma_1 & \gamma_2 & \gamma_3 \end{bmatrix} .$$

Then $(\mathcal{R}_1)^{-1}\ \Lambda \in \mathcal{L}_+^\uparrow$, and is of the form (as is explained below)

$$(\mathcal{R}_1)^{-1}\ \ \Lambda = \begin{bmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ a_{10} & a_{11} & a_{12} & a_{13} \\ (a_{20} = 0) & a_{21} & a_{22} & a_{23} \\ (a_{30} = 0) & a_{31} & a_{32} & a_{33} \end{bmatrix} .$$

Here $\{a_{ij}\}$ are real numbers one gets from the above matrix multiplication. Because of the explicit form of $(\mathcal{R}_1)^{-1}$, the top row is not affected. Here,

$$a_{10} = \alpha_1\ \Lambda^1{}_0 + \alpha_2\ \Lambda^2{}_0 + \alpha_3\ \Lambda^3{}_0 = \sqrt{(\Lambda^1{}_0)^2 + (\Lambda^2{}_0)^2 + (\Lambda^3{}_0)^2}\ >\ 0\ ,$$

and

$$\begin{aligned} a_{20} &= \beta_1\ \Lambda^1{}_0 + \beta_2\ \Lambda^2{}_0 + \beta_3\ \Lambda^3{}_0 \\ &= \sqrt{(\Lambda^1{}_0)^2 + (\Lambda^2{}_0)^2 + (\Lambda^3{}_0)^2}\ (u_1 \cdot u_2) = 0\ . \end{aligned}$$

Similarly, $a_{30} = 0$. Now let

$$v_2 = (a_{21}, a_{22}, a_{23})$$

and

$$v_3 = (a_{31}, a_{32}, a_{33})\ .$$

Note that v_2 and v_3 are orthonormal vectors since $a_{20} = 0 = a_{30}$. Let $v_1 = (c_1, c_2, c_3)$ be such that $\{v_1, v_2, v_3\}$ is an orthonormal basis for $\mathbb{R}^3$. If necessary, by allowing $v_1 \to -v_1$ we can arrange for

$$(\mathcal{R}_2)^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c_1 & a_{21} & a_{31} \\ 0 & c_2 & a_{22} & a_{32} \\ 0 & c_3 & a_{23} & a_{33} \end{bmatrix}$$

to be an orthogonal matrix in $\mathcal{L}_+^\uparrow$. Then $(\mathcal{R}_1)^{-1}\ \ \Lambda\ \ (\mathcal{R}_2)^{-1}$ must be in $\mathcal{L}_+^\uparrow$, and is of the form, which is easily verified using problem 2.11, that

$$(\mathcal{R}_1)^{-1}\ \ \Lambda\ \ (\mathcal{R}_2)^{-1} = \begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Here

$$\begin{aligned} a = \Lambda^0{}_0\ ,\ b = \Lambda^0{}_1 c_1 + \Lambda^0{}_2 c_2 + \Lambda^0{}_3 c_3\ , \\ c = a_{10}\ ,\ \text{and}\ \ d = a_{11} c_1 + a_{12} c_2 + a_{13} c_3\ . \end{aligned}$$

As we have already shown in problem 2.12, a proper orthochronous Lorentz transformation of the form

$$\begin{bmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

is a Lorentz transformation in the x-direction. □

Theorem 2.12. *All the elements of $\mathcal{L}$ are generated by $\mathbb{T}, \mathbb{P}$ and $\mathcal{L}_+^\uparrow$. Here, $\mathbb{T}$ and $\mathbb{P}$ are as in problem 2.6 and 2.7 respectively.*

Proof. Let $\Lambda \in \mathcal{L}$. Recall that $\mathbb{T}^{-1} = \mathbb{T}$, and $\mathbb{P}^{-1} = \mathbb{P}$. Then

if $\Lambda \in \mathcal{L}_+^\downarrow$, then $\mathbb{TP}\Lambda \in \mathcal{L}_+^\uparrow$. Therefore $\Lambda = \mathbb{PT}\Lambda_+^\uparrow$ for some $\Lambda_+^\uparrow \in \mathcal{L}_+^\uparrow$.

if $\Lambda \in \mathcal{L}_-^\downarrow$, then $\mathbb{T}\Lambda \in \mathcal{L}_+^\uparrow$. Therefore $\Lambda = \mathbb{T}\Lambda_+^\uparrow$ for some $\Lambda_+^\uparrow \in \mathcal{L}_+^\uparrow$.

if $\Lambda \in \mathcal{L}_-^\uparrow$, then $\mathbb{P}\Lambda \in \mathcal{L}_+^\uparrow$. Therefore $\Lambda = \mathbb{P}\Lambda_+^\uparrow$ for some $\Lambda_+^\uparrow \in \mathcal{L}_+^\uparrow$. □

In particle physics, there are interactions that do not respect time reversal symmetry and parity. I.e., the laws of physics in such cases are not invariant under $\mathbb{T}$ and $\mathbb{P}$, and are not considered as inertial transformations.

2.5 Causal Structure of Minkowski Spacetime

Definition 2.8. *The collection of Minkowski spacetime points denoted by the 4-tuple (ct, x, y, z), along with the inner product η acting on vectors is called Minkowski spacetime, and is denoted by $\mathbb{M}^4$.*

Definition 2.9. *The semi-direct product of the Lorentz group and all the translations forms the Poincaré group. This is the cartesian product (Λ, a), where $\Lambda \in \mathcal{L}_+^\uparrow$ and $a \in \mathbb{M}^4$. Under the action of an element of the Poincaré group on $\mathbb{M}^4$, we get the transformation $\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu$.*

Thus far, all known interactions respect the Poincaré group. I.e., all laws of physics are invariant under transformations of the type given above. The elements of the Poincaré transformations are the generators of new inertial frames.

Now that we have defined our background spacetime $\mathbb{M}^4$ and the inner product η, it is important to clarify that a vector X is located at some particular $x \in \mathbb{M}^4$. When this distinction is important, it is indicated by a subscript x. I.e., a vector at the spacetime point x will be written as X_x or more commonly as $X(x)$.

Definition 2.10. *The set of all vectors $X(x)$, at any $x \in \mathbb{M}^4$, clearly forms a vector space at x that is denoted by $T_x(\mathbb{M}^4)$. An element of $T_x(\mathbb{M}^4)$ is referred to as a tangent vector at x of the Minkowski spacetime.*

Definition 2.11. *A vector field Y in $\mathbb{M}^4$ is a differentiable map such that $Y(x) \in T_x(\mathbb{M}^4)$ for any $x \in \mathbb{M}^4$. The differentiability of Y implies that while Y takes on the form*

$$Y(x) = \left(Y^0(x), Y^1(x), Y^2(x), Y^3(x)\right) ,$$

each component function is differentiable with respect to the spacetime coordinates $\{x^\nu\}$. Specifically, this means that

$$\frac{\partial Y^\mu}{\partial x^\nu}$$

exists for all values of μ and ν. Typically, we require Y to be smooth. In this case, all orders of mixed partial derivatives exist. [10]

Since the inner product η is not dependent on the location of a spacetime point in Minkowski coordinates, in what follows we will often indicate a vector without reference to its location point.

Remark 2.2. However, since the inner product is a map between any two vectors at the same spacetime location, it is to be implicitly understood that, in what follows, $\eta(A, B)$ is short for $\eta(A_x, B_x)$ for some $x \in \mathbb{M}^4$.

Suppose $A^\mu = \Delta x^\mu$ is the difference in location between the starting point and the ending point of a particle moving with a constant speed, then

$$A^2 = -(\Delta ct)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 < 0$$

implies

$$\frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{(\Delta t)^2} < c^2 .$$

Therefore, the object is traveling below the speed of light, and hence it would take more time for the object to reach its destination than light would. When $A^2 = 0$, we have a vector that clearly describes the motion of a particle travelling at the speed of light. Finally, if $A^2 > 0$, the supposed particle the vector A^μ is describing travels a greater spatial distance than light would in the same amount of time. We are not claiming that such particles exists, however, the above discussion motivates the following definition.

[10] By a mixed partial derivative we mean objects like

$$\frac{\partial^2 Y^\mu}{\partial x^0 \partial x^1}$$

etc.

Definition 2.12. *Let A be any vector at any point in $\mathbb{M}^4$. Then A is called*

- *timelike if $A^2 < 0$.*
- *light-like (or null) if $A^2 = 0$.*
- *spacelike if $A^2 > 0$.*
- *A is causal if it is time-like or null.*

Definition 2.13. *Let A, B, X and Y be any set of vectors in $\mathbb{M}^4$. Then*

- *if $A^2 = -1$, we say that A is a unit timelike vector*
- *if $B^2 = 1$, we say that B is a unit spacelike vector*
- *if X and Y are vectors at any same point $x \in \mathbb{M}^4$, and if $\eta(X, Y) = 0$, we say that X and Y are orthogonal.*
- *if X and Y are unit orthogonal vectors, then they are called orthonormal.*

Problem 2.13. Let $T = (T^0, T^1, T^2, T^3)$ be a timelike vector. Then there exists a rotation $\Lambda \in \mathcal{L}$ such that $\bar{T}^\mu = \Lambda^\mu{}_\nu \, T^\nu$ is of the form $\bar{T}^\mu = (\bar{T}^0, \bar{T}^1, 0, 0)$. Hint: Use theorem 2.1.

Problem 2.14. Let T be a timelike vector of the form $T = (T^0, T^1, 0, 0)$. Then there exists a Lorentz boost $\Lambda \in \mathcal{L}$ such that $\bar{T}^\mu = \Lambda^\mu{}_\nu \, T^\nu$ is of the form $\bar{T}^\mu = (\bar{T}^0, 0, 0, 0)$.

Definition 2.14. *Let $A \in \mathbb{M}^4$. Then*

$$A^\perp = \{B \in \mathbb{M}^4 |\ \eta(A, B) = 0\}$$

is the set of all vectors orthogonal to A.

Problem 2.15. Let $A \in \mathbb{M}^4$. Show that $A^\perp$ is a vector space.

Theorem 2.13. *Let T be a timelike vector in $\mathbb{M}^4$. Then $T^\perp$ is a 3-dimensional subspace of $\mathbb{M}^4$ consisting of only spacelike vectors.*

Proof. Considering the results from the last two problems, let $\Lambda \in \mathcal{L}$ be such that

$$\bar{T}^\mu = \Lambda^\mu{}_\nu T^\nu$$

is of the form given by

$$\bar{T}^\mu = (\bar{T}^0, 0, 0, 0) \ .$$

In this frame set

$$\bar{S}_1^\mu = (0, 1, 0, 0), \quad \bar{S}_2^\mu = (0, 0, 1, 0) \ , \quad \text{and} \quad \bar{S}_3^\mu = (0, 0, 0, 1) \ .$$

Then

$$\bar{T}^\perp = \text{ span } \{\bar{S}_1, \bar{S}_2, \bar{S}_3\} \ .$$

Since η is invariant under $\mathcal{L}$, we get that

$$T^\perp = \text{ span } \{S_1, \ S_2, S_3\} \ ,$$

where

$$S_i^\mu = (\Lambda^{-1})^\mu{}_\nu \, \bar{S}_i^\nu$$

for $i = 1, 2, 3$ are orthonormal spacelike vectors. □

Problem 2.16. Let $N = (N^0, N^1, N^2, N^3)$ be a lightlike vector. Then there exists a rotation $\Lambda \in \mathcal{L}$ such that $\bar{N}^\mu = \Lambda^\mu{}_\nu \, N^\nu$ is of the form $\bar{N}^\mu = (a, \pm\, a, 0, 0)$, where a is a constant. Hint: Use theorem 2.1.

Theorem 2.14. *Let N be a lightlike vector in $\mathbb{M}^4$. Then $N^\perp$ is a 3-dimensional subspace of $\mathbb{M}^4$ of the type*

$$N^\perp = \text{ span } \{N, S_1, S_2\} \ ,$$

where S_1 and S_2 are spacelike vectors.

Proof. Let $\Lambda \in \mathcal{L}$ be such that

$$\bar{N}^\mu = \Lambda^\mu{}_\nu N^\nu$$

is of the form given by

$$\bar{N}^\mu = (a, \pm a, 0, 0) \, .$$

In this frame set

$$\bar{S}_1^\mu = (0, 0, 1, 0) \, , \text{ and } \bar{S}_2^\mu = (0, 0, 0, 1) \, .$$

Then

$$\bar{N}^\perp = \text{ span } \{\bar{N}, \bar{S}_1, \bar{S}_2\} \, .$$

Since η is invariant under $\mathcal{L}$, we get that

$$N^\perp = \text{ span } \{N, S_1, S_2\} \, ,$$

where

$$S_i^\mu = (\Lambda^{-1})^\mu{}_\nu \, \bar{S}_i^\nu$$

for $i = 1, 2$ are orthonormal spacelike vectors. □

Note that $N^\perp$ contains N. This is a peculiarity of special relativity stemming from the fact that η permits non-zero null vectors (since every null vector is orthogonal to itself). The 3-dimensional hyperplane $N^\perp$ is often referred to as a null hyperplane. In general relativity, where spacetime is pliable, we will have null hypersurfaces of interest. In particular, as we shall see, the event horizon of a black hole is in general a null hypersurface.

Time Orientation

We began our study of relativity by looking at how inertial coordinates are constructed. An inertial observer, free from any external forces, would use a light source, and a clock to setup a spacetime coordinate chart $\{x^\mu\}$ for all events of interest. Suppose in this coordinate system, a particle moves from point x_1^μ to x_2^μ, then the displacement vector from x_1^μ to x_2^μ is such that

$$\Delta x^0 = x_2^0 - x_1^0 > 0 \, . \tag{2.14}$$

Meaning, event x_2^μ occurs at a later time than x_1^μ. We will refer to such a vector that denotes particle displacement as a future-pointing timelike vector. Strictly speaking, while the above choice is convenient, it is not necessary. This, however, means that we have to keep up with our unnatural choice of coordinate system and compensate for this fact during every calculation. We will dispense with such tautology by defining a notion of "future" in precise terms.

Theorem 2.15. *Let T be timelike, and W any causal vector, then either*

$$T^0 W^0 > 0 \text{ in which case } \eta(T, W) < 0 \, ,$$

or

$$T^0 W^0 < 0 \text{ in which case } \eta(T, W) > 0 \, .$$

Proof. If $T^0 W^0 = 0$, then since T is timelike $T^0 \neq 0$, so $W^0 = 0$ i.e., W is spacelike and hence is not causal. So either $T^0 T^0 > 0$ or $T^0 W^0 < 0$.

We have that $(T^0)^2 > \sum_{i=1}^3 (T^i)^2$ and $(W^0)^2 \geq \sum_{i=1}^3 (W^i)^2$. Therefore,

$$(T^0 W^0)^2 > \left(\sum_{i=1}^3 (T^i)^2 \right) \left(\sum_{j=1}^3 (W^j)^2 \right) \geq \left(\sum_{i=1}^3 T^i \, W^i \right)^2 \, ,$$

i.e.,

$$|T^0 W^0| > \left| \sum_{i=1}^3 T^i \, W^i \right| \, .$$

If $T^0 W^0 > 0$, we have that

$$T^0 W^0 = |T^0 W^0| > \left| \sum_{i=1}^3 T^i \, W^i \right| \geq \sum_{i=1}^3 T^i \, W^i \, .$$

Therefore, $g(T, W) < 0$. If $T^0 W^0 < 0$, we have $g(T, -W) < 0$ and so $g(T, W) > 0$. □

Given a causal vector, it may not be clear if you are in an inertial chart such that for future pointing causal vectors, eq.(2.14) is true.

Definition 2.15. *Let $\mathcal{T}$ be any fixed timelike vector field in $\mathbb{M}^4$. Then $(\mathbb{M}^4, \mathcal{T})$ is referred to as a time-oriented Minkowski spacetime with orientation $\mathcal{T}$. In $(\mathbb{M}^4, \mathcal{T})$ a causal vector W is future pointing if and only if $\eta(\mathcal{T}, W) < 0$.*

Please note that,

a. $\mathcal{T}$ should evaluated at the same space-time location as W (see remark 2.2).
b. furthermore, by definition, $\mathcal{T}$ is future-pointing in $(\mathbb{M}^4, \mathcal{T})$.

Definition 2.16. *In $(\mathbb{M}^4, \mathcal{T})$, let T be a timelike vector in a given Minkowski coordinate system. The Minkowski coordinate system is positively time-oriented if $T^0 > 0$.*

Theorem 2.16. *Let X be a future-pointing causal vector in a positively time-oriented Minkowski coordinate system (ct, x, y, z). Then $X^0 > 0$.*

Proof. This follows immediately from definition 2.15 and theorem 2.15. □

Theorem 2.17. *Let (ct, x, y, z) be a positively time-oriented Minkowski coordinate system for $(\mathbb{M}^4, \mathcal{T})$, and let $\Lambda \in \mathcal{L}_+^\uparrow$. Then $(c\bar{t}, \bar{x}, \bar{y}, \bar{z})$ as generated by Λ in the usual manner is a positively time-oriented Minkowski coordinate system.*

Proof. Considering theorem 2.11, we can pick Λ to be a Lorentz boost in the x-direction. In this case

$$\bar{\mathcal{T}}^0 = \gamma \left(\mathcal{T}^0 - \beta \mathcal{T}^1 \right) .$$

But,

$$-(\mathcal{T}^0)^2 + (\mathcal{T}^1)^2 + (\mathcal{T}^2)^2 + (\mathcal{T}^3)^2 \leq 0 .$$

Since $\mathcal{T}^0 > 0$, we get that

$$\mathcal{T}^0 \geq \sqrt{(\mathcal{T}^1)^2 + (\mathcal{T}^2)^2 + (\mathcal{T}^3)^2} \geq \mathcal{T}^1 .$$

Hence, since $\beta < 1$, we get that $\bar{\mathcal{T}}^0 > 0$. □

Problem 2.17. Let N be a continuous future-pointing null vector field in $(\mathbb{M}^4, \mathcal{T})$. Then any causal vector W is future pointing if and only if

a. $\eta(W, N) < 0$, at any point in spacetime where W is not proportional to N.
b. $W = a^2 N$, for some non-zero number a, wherever W is proportional to N.

From the above problem, we see that a non-vanishing null vector field defines a time orientation on $\mathbb{M}^4$.

Problem 2.18. Find a spacelike vector S such that $S^0 > 0$. Then find a Lorentz boost in $\mathcal{L}_+^\uparrow$ such that $\bar{S}^0 < 0$.

From the above problem, it should be clear that *there is no temporal order to spacelike vectors.*

Problem 2.19. If V and W are future-pointing timelike and future-pointing causal vectors, respectively in $(\mathbb{M}^4, \mathcal{T})$, show that $V + W$ is future-pointing timelike.

Problem 2.20. Let $\mathcal{T}$ be any continuous timelike vector field in $\mathbb{M}^4$. Show that the only possible time orientations of $\mathbb{M}^4$ are given by $(\mathbb{M}^4, \mathcal{T})$ and $(\mathbb{M}^4, -\mathcal{T})$.

Problem 2.21. Let (ct, x, y, z) be an inertial coordinate system for $\mathbb{M}^4$. In this coordinate system, let $\mathcal{T} = (-3, 0, 1, 2)$ be a continuous timelike vector field in $\mathbb{M}^4$. Consider the time-oriented spacetime $(\mathbb{M}^4, \mathcal{T})$.

1. Confirm that $\mathcal{T}$ is a timelike vector.
2. Let $W = (1, 0, 0, 1)$. Confirm that W is a causal vector.
3. If so, is W future-pointing?

Problem 2.22. Let $(\mathbb{M}^4, \mathcal{T})$ be a time-oriented Minkowski spacetime. Let W be a future-pointing timelike vector in $(\mathbb{M}^4, \mathcal{T})$. Then show that

$$(\mathbb{M}^4, \mathcal{T}) \equiv (\mathbb{M}^4, W) \ .$$

I.e., show that X is future-pointing causal vector in $(\mathbb{M}^4, \mathcal{T})$ if and only if X is future-pointing causal in $(\mathbb{M}^4, W)$.

2.6 Zeeman's Theorem

Definition 2.17. *Let $f : \mathbb{M}^4 \to \mathbb{M}^4$ such that*

1. *f is 1-1 and onto. I.e., f^{-1} is well defined.*
2. *$(v - w)^2 = 0$ for $v, w \in \mathbb{M}^4 \iff (f(v) - f(w))^2 = 0$.*
3. *$v - w$ is future pointing lightlike $\iff$ $f(v) - f(w)$ is future pointing lightlike.*

Then f is called a causal automorphism.

A causal automorphism is the most general candidate for an inertial transformation: f takes the whole of $\mathbb{M}^4$ and maps it uniquely to the whole of $\mathbb{M}^4$. This means that the new inertial observer has complete access to spacetime. Further, from property 2, f maps lightlike vectors to lightlike vectors, i.e., f preserves the speed of light. Finally, f preserves the temporal order on light rays. I.e., it maps future-pointing null vectors to future-pointing null vectors. This is a mild and meaningful restriction. We certainly don't want light beams to be sent to the past by an inertial observer! However, we are not a priori restricting f to be a linear transformation.

Definition 2.18. *A dilation $K : \mathbb{M}^4 \to \mathbb{M}^4$ is the map*

$$K(v) = k\, v$$

for every $v \in \mathbb{M}^4$. Here k is a positive constant. It is usual for k to be referred to as a conformal factor.

We state the following theorem by Zeeman without proof [2]. The most intuitive proof of Zeeman's theorem is by S. Nanda and is definitely worth studying [5].

Theorem 2.18. *Let f be a causal automorphism of $\mathbb{M}^4$. Then there exists an orthochronous orthogonal transformation Λ, $a \in \mathbb{M}^4$, and a dilation K of $\mathbb{M}^4$ such that $f = K \circ \Lambda + a$. I.e.,*

$$\bar{x}^\mu = f^\mu(x) = k\, \Lambda^\mu{}_\nu\, x^\nu + a^\mu \ ,$$

for some $\Lambda \in \mathcal{L}^\uparrow_+$.

Let us see what this very important theorem tells us about special relativity. First we will tackle the conformal factor k. A transformation of type

$$(ct, x, y, z) \to k\ (ct, x, y, z)$$

has no physical content. It simply changes the length scale by a factor k. A dilation is similar to converting from metric units to English units, for example. Barring conformal factors, any transformations (that could include non-linear ones) that preserves the feature that the speed of light is constant and the fact that light flows into the future is necessarily an inertial transformation generated by the Poincaré group, and none other. Our quest for inertial transformations are, in fact, over. And, most importantly, since Poincaré transformations preserve the inner product η, it is truly a fundamental property of spacetime and is not preserved just for lightlike vectors. This inner product must be elevated to the status of a metric, in a differential geometric sense, when using non-inertial coordinates, or even non-Cartesian coordinates for that matter. [11] So, our hasty calculation

[11] This will be clarified along the way. For example, when using the familiar spherical coordinate system on the spatial coordinates (x, y, z), i.e., when we use coordinates (ct, r, θ, φ), η should be written as

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{bmatrix} .$$

We will learn how and why we do this when we study tensor calculus.

of the Lorentz transformation has proven to be the correct one under boosts. There is no other way the speed of light can be preserved, along with causality, without mixing space and time!

3

Relativistic Mechanics and Electrodynamics

In this chapter, we will assume that $\mathcal{T}$ has been fixed and that all positively time-oriented inertial charts have $\mathcal{T}^0 > 0$, and so $(\mathbb{M}^4, \mathcal{T})$ will simply be denoted by $\mathbb{M}^4$. Any future pointing causal vector W will have $W^0 > 1$.

3.1 Particle Dynamics

A curve α is a smooth (at least piecewise differentiable) map

$$\alpha : (a, b) \to \mathbb{M}^4 \ .$$

Here (a, b) is an open interval in $\mathbb{R}$. In a coordinate chart, α can be written in the form

$$\alpha(s) = \big(c\, t(s),\ x(s),\ y(s),\ z(s)\big) \ .$$

The curve α is timelike, spacelike or null if $g(\alpha', \alpha')$ is less than zero, greater than zero, or zero, respectively, for every $s \in (a, b)$, where

$$\alpha' = \left(c\, \frac{dt}{ds},\ \frac{dx}{ds},\ \frac{dy}{ds},\ \frac{dz}{ds} \right) \ .$$

For now, we will concentrate on future-pointing timelike curves since we are interested in the *worldline* (as they are called) of a massive particle. For this reason, we set

$$c\, \frac{dt}{ds} > 0 \ .$$

Example 3.1. Consider the curve

$$\alpha(s) = \big(.3c\ s^2,\ 0, 0,\ .3cs\big) \ .$$

Then

$$\alpha' = \frac{d}{ds}\, \alpha(s) = (.6c\ s,\ 0,\ .3c) \ .$$

To check if α is a timelike curve, we compute its Minkowski norm:

$$g(\alpha', \alpha') = -.36c^2 s^2 + .09c^2 \ .$$

I.e., α is timelike when $s > 1/2$, in which case it is also future pointing. ■

Example 3.2. In the previous example, consider any pair of $s_2 > s_1 > 1/2$. Then

$$\begin{aligned} \alpha(s_2) - \alpha(s_1) &= .3c\ \big(s_2^2 - s_1^2,\ 0,\ 0,\ s_2 - s_1\big) \\ &= .3c(s_2 - s_1)\ (s_2 + s_1,\ 0,\ 0,\ 1) \ . \end{aligned}$$

is clearly future-pointing timelike for all s. ■

In the previous example, α is a future-pointing timelike curve from p to q, and as expected $q - p$ is a future-pointing and timelike vector. This statement is true in general, as will be shown below. For ease of notation, for any vector X in $\mathbb{M}^4$, define the quadratic form

$$Q_M(X) = \eta(X, X) \ .$$

Lemma 3.1. *Let $\alpha : (a,b) \to \mathbb{M}^4$ be a future pointing timelike curve. Let $s_0 \in (a,b)$. Then there exists $\delta_1 > 0$ such that when $s \in (s_0,\ s_0+\delta_1) \subset (a,b)$ we have that $\alpha(s) - \alpha(s_0)$ is a future pointing timelike vector.*

Proof. Since $dt/ds > 0$ we have that $t(s_2) - t(s_1) > 0$.

Suppose the premise of the lemma is not true. Then there exists a monotonically decreasing sequence s_n that converges to s_0 from above such that

$$Q_M(\alpha(s_n) - \alpha(s_0)) \geq 0 \ .$$

The quadratic form cannot be less than zero, for then $\alpha(s_n) - \alpha(s_0)$ would be future pointing (since $t(s_n) - t(s_0) > 0$ as shown above) for every sequence, and then we will have nothing to prove. I.e.,

$$Q_M\left(\frac{\alpha(s_n) - \alpha(s_0)}{s_n - s_0}\right) \geq 0$$

for every n. Since Q_M is a continuous function,

$$0 \leq \lim_{n\to\infty} Q_M\left(\frac{\alpha(s_n) - \alpha(s_0)}{s_n - s_0}\right)$$

$$= Q_M\left(\lim_{n\to\infty} \frac{\alpha(s_n) - \alpha(s_0)}{s_n - s_0}\right) = Q_M(\alpha'(s_0)) \ .$$

I.e., $\alpha'(s_0)$ is either spacelike or null. And this is a contradiction of the premise. □

Theorem 3.1. *Let $\alpha : (a,b) \to \mathbb{M}^4$ be a future pointing timelike curve such that $\alpha(s_i) = p$ and $\alpha(s_f) = q$. Here $s_i < s_f \in (a,b)$. Then $q - p$ is future-pointing timelike.*

Proof. First, we note that the previous lemma can be extended to the past. I.e., for every $s_0 \in (a,b)$ there exists $\delta_2 > 0$ such that when $s \in (s_0 - \delta_2,\ s_0) \subset (a,b)$ we have that $\alpha(s_0) - \alpha(s)$ is future pointing timelike.

From lemma (3.1), there exists $\delta > 0$ such that when $s \in (s_i,\ s_i + \delta)$ we have that $\alpha(s) - p$ is future pointing timelike. Let $\epsilon = \sup\{\delta\}$ for all δ satisfying the condition above. The supremum exists because all δ are bounded by $(s_f - s_i)$. Then for $s_i < s < s_i + \epsilon$, we have that $\alpha(s) - \alpha(s_i)$ is future pointing timelike.

Suppose $\epsilon < s_f - s_i$. From lemma (3.1), there exists δ_2 such that when

$$s \in (s_i + \epsilon - \delta_2, s_i + \epsilon) \subset (s_i, s_f)$$

we have that $\alpha(s_i + \epsilon) - \alpha(s)$ is future pointing timelike. Then

$$\alpha(s_i + \epsilon) - \alpha(s_i) = (\alpha(s_i + \epsilon) - \alpha(s)) + (\alpha(s) - \alpha(s_i)) \ .$$

$$Q\Big(\alpha(s_i + \epsilon) - \alpha(s_i)\Big) = Q\Big(\alpha(s_i + \epsilon) - \alpha(s) + \alpha(s) - \alpha(s_i)\Big) = Q\Big(\alpha(s_i + \epsilon) - \alpha(s)\Big) + Q\Big(\alpha(s) - \alpha(s_i)\Big)$$

$$+2\eta\Big(\alpha(s_i + \epsilon) - \alpha(s), \alpha(s) - \alpha(s_i)\Big) \ .$$

All terms in the right-hand side above are negative (Why is the last term negative?). Therefore, $\alpha(s_i+\epsilon) - \alpha(s_i)$ is timelike and further its future pointing because $\alpha(s_i + \epsilon) - \alpha(s)$, and $\alpha(s) - \alpha(s_i)$ are. In exactly the same way, there exists δ_3 such that if $s \in (s_i, s_i + \epsilon + \delta_3)$, then

$$\alpha(s) - \alpha(s_i)$$

is future pointing timelike. This contradicts the definition of ϵ. □

There are two immediate concerns that we will address simultaneously. The parameter s for timelike curves seem arbitrary and devoid of physical meaning. Also, we have not yet introduced the notion of proper time for accelerating particles. We will now develop a notion of proper time for the accelerating particle and use it to parameterize its worldline.

Propertime For Accelerating Particles

Recall example 1.1. Here, the observer on Earth uses a coordinate system given by (ct, x, y, z), while the observer on the spaceship uses a coordinate system given by $(c\bar{t}, \bar{x}, \bar{y}, \bar{z})$. The change in position of the spaceship, which I call W for ease, as measured on Earth, is given by

$$W = (c\Delta t, \Delta x, 0, 0) = (c\Delta t, \beta c\Delta t, 0, 0).$$

Meanwhile, the observer on the spaceship measures the change in position to be

$$\bar{W} = (c\Delta\bar{t}, \Delta\bar{x}, 0, 0) = \Big(\gamma(c\Delta t - \beta\Delta x), \gamma(\Delta x - \beta c\Delta t), 0, 0\Big)$$

$$= \Big(\gamma(c\Delta t - \beta\Delta x), 0, 0, 0\Big) = (c\Delta\bar{t}, 0, 0, 0).$$

Clearly, here $\Delta\bar{t}$ measures the propertime lapsed. Why?

Then

$$W^2 = \eta(W, W) = \bar{W}^2 = \eta(\bar{W}, \bar{W}) = -c^2(\Delta\bar{t})^2 \ .$$

Or,

$$c\Delta\tau = c\Delta\bar{t} = \sqrt{-Q_M(W)} \ .$$

The expression above shows that proper time is independent of the inertial coordinate system.

Now, let us consider the case of an accelerating particle. Let $\alpha(s)$ for s in some open interval in $\mathbb{R}$ be the worldline of a particle traveling below the speed of light. Then, as mentioned earlier, α' is future-pointing timelike. Let $\alpha(s_i) = p$ and $\alpha(s_f) = q$ be any two events on the curve, also let $\{s_j\}_{j=1}^n$ be a monotonically increasing sequence such that

$$s_i < s_1 < s_2 < ... < s_n < s_f \ .$$

Then from the previous theorem, if $s_m > s_n$, $\alpha(s_m) - \alpha(s_n)$ is future-pointing timelike. Therefore,

$$c\Delta\tau_{total} \approx \sqrt{-Q_M(\alpha(s_1) - \alpha(s_i))} + \sqrt{-Q_M(\alpha(s_2) - \alpha(s_1))} + ...$$

$$+\sqrt{-Q_M(\alpha(s_n) - \alpha(s_{n-1}))} + \sqrt{-Q_M(\alpha(s_f) - \alpha(s_n))}$$

$$= \sqrt{\frac{-Q_M(\alpha(s_1) - \alpha(s_i))}{(s_1 - s_i)^2}}(s_1 - s_i) + ... + \sqrt{\frac{-Q_M(\alpha(s_f) - \alpha(s_n))}{(s_f - s_n)^2}}(s_f - s_n) \ .$$

Note that

$$\frac{Q_M(\alpha(s_b) - \alpha(s_a))}{(s_b - s_a)^2} = \eta\left(\frac{\alpha(s_b) - \alpha(s_a)}{s_b - s_a}, \frac{\alpha(s_b) - \alpha(s_a)}{s_b - s_a}\right) \ .$$

Now we let $n \to \infty$ such that all $\Delta s \to 0$ to obtain the exact propertime for the worldline.

Definition 3.1. *Let $\alpha(s)$ be a future pointing timelike curve, the propertime τ as a function of the parameter s is given by*

$$c\,\tau = \int^s \sqrt{-\eta(\alpha'(\bar{s}), \alpha'(\bar{s}))}\, d\bar{s} \ . \tag{3.1}$$

Here, $\alpha' = d\alpha/ds$, and $\bar{s}$ is a dummy integration variable.

Theorem 3.2. *The propertime along any future pointing timelike curve does not depend on the inertial coordinate system, nor does it depend on the choice of the smooth parameter s.*

Proof. Clearly, τ is independent of the inertial frame used (since $\eta(\alpha', \alpha')$ is invariant). It is not difficult to see that τ is independent of the parameter s. Since $\alpha(s)$ is future pointing, we have that $dx^0/ds > 0$. Let $u(s)$ be another parameter for α such that $\alpha(u)$ is also a future pointing timelike curve. Then $dx^0/du > 0$, and so $du/ds > 0$ since

$$\frac{dx^0}{ds} = \frac{dx^0}{du}\frac{du}{ds} > 0 \ .$$

Therefore,

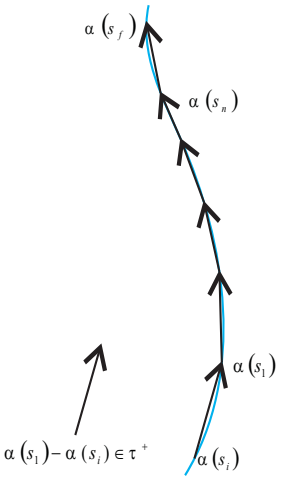


Fig. 3.1. The solid lines approximates the smooth curve α. Each arrow indicates that the displacement vector is future pointing.

$$c\,\tau = \int \sqrt{-\eta\left(\frac{d\alpha}{ds}, \frac{d\alpha}{ds}\right)}\, ds = \int \sqrt{-\left(\frac{du}{ds}\right)^2 \eta\left(\frac{d\alpha}{du}, \frac{d\alpha}{du}\right)}\, ds = \int \sqrt{-\eta\left(\frac{d\alpha}{du}, \frac{d\alpha}{du}\right)}\, \left|\frac{du}{ds}\right|\, ds$$
$$= \int \sqrt{-\eta\left(\frac{d\alpha}{du}, \frac{d\alpha}{du}\right)}\, \frac{du}{ds}\, ds = \int \sqrt{-\eta\left(\frac{d\alpha}{du}, \frac{d\alpha}{du}\right)}\, du\ .$$

□

Example 3.3. Find the proper time function for

$$\alpha(s) = \left(c\, s^3,\ 0,\ 0,\ 3\, c\, s^2\right)\ .$$
$$\alpha' = \frac{d}{ds}\,\alpha(s) = \left(3\, c\, s^2,\ 0,\ 0,\ 6\, c\, s\right)\ .$$

To check if α is a timelike curve, we compute its Minkowski norm:

$$\eta(\alpha', \alpha') = -9c^2\, s^4 + 36c^2\, s^2 = -9\, c^2 s^2 (s^2 - 4)\ .$$

Therefore, α is future pointing timelike when $s > 2$. Then

$$\tau = \int 3s\, \sqrt{s^2 - 4}\, ds + \text{ const}\ .$$

I.e.,

$$\tau = (s^2 - 4)^{3/2}$$

when the integration constant is set to zero. When α is proper time parameterized,

$$\alpha(\tau) = \left(c\, (\tau^{2/3} + 4)^{3/2},\ 0,\ 0,\ 3\, c\, (\tau^{2/3} + 4)\right) \tag{3.2}$$

for $\tau > 0$. ■

Relativistic Velocity And Acceleration

We use the propertime associated with the worldline of the particle to define the 4 velocity of the particle.

Definition 3.2. *Let $\alpha(\tau)$ be a propertime parameterized future pointing timelike curve. Then the proper 4 velocity of the particle associated to the curve is defined by*

$$U = \dot{\alpha} = \left(c\, \frac{dt}{d\tau}\ ,\ \frac{dx}{d\tau}\ ,\ \frac{dy}{d\tau}\ ,\ \frac{dz}{d\tau}\right)\ .$$

From now on $\cdot$ on timelike curves refers to derivative with respect to propertime. The 4-acceleration of the particle is defined in the usual way by

$$a = \dot{U} = \ddot{\alpha} = \left(c\, \frac{d^{\,2}t}{d\tau^2}\ ,\ \frac{d^{\,2}x}{d\tau^2}\ ,\ \frac{d^{\,2}y}{d\tau^2}\ ,\ \frac{d^{\,2}z}{d\tau^3}\right)\ .$$

From definition (3.1) we have that

$$\boxed{U^2 \equiv \eta(U,U) = -c^2 \ .} \tag{3.3}$$

Taking the propertime derivative of the equation above gives

$$\frac{d}{d\tau} U^{\,2} = 2\eta(U,a) = 0 \ ,$$

i.e.,

$$\boxed{\eta(U,a) = 0 \ ,} \tag{3.4}$$

and in particular a is spacelike (since U is timelike).

Transformation Of 4-velocity and acceleration

Consider a moving particle in an inertial frame S of $\mathbb{M}^4$. When parameterized by proper time, its world line takes the form

$$(x^0(\tau),\ x^1(\tau),\ x^2(\tau),\ x^3(\tau)).$$

Then the components of the 4-velocity are given by

$$U^\mu = \frac{dx^\mu}{d\tau} \ .$$

Since τ is invariant under translations, in a translated frame

$$\bar{U}^\mu = \frac{d\bar{x}^\mu}{d\tau} = \frac{d}{d\tau}(\bar{x}^\mu + a^\mu) = \frac{dx^\mu}{d\tau} = U^\mu \ .$$

I.e., U does not change under translations. Since τ is invariant under the action of $\Lambda \in \mathcal{L}$,

$$\bar{U}^\mu = \frac{d\bar{x}^\mu}{d\tau} \equiv \frac{d}{d\tau}(\Lambda^\mu{}_\nu\ x^\nu) = \Lambda^\mu{}_\nu\ \frac{dx^\nu}{d\tau} \ .$$

Or,

$$\boxed{\bar{U}^\mu = \Lambda^\mu{}_\nu\ U^\nu \ .} \tag{3.5}$$

I.e., 4-velocity, and similarly, 4-acceleration transforms like a vector in $\mathbb{M}^4$.

The decomposition of 4-velocity

It is possible to write the 4-velocity vector in terms of the familiar 3-velocity vector

$$\mathbf{u} = \left(\frac{dx}{dt}\ ,\ \frac{dy}{dt}\ ,\ \frac{dz}{dt}\right)$$

of an object.

$$\begin{aligned} -c^2 = U^2 &= -c^2\ \left(\frac{dt}{d\tau}\right)^2 + \left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2 \\ &= -\left(\frac{dt}{d\tau}\right)^2 \left(c^2 - \left(\frac{dx}{dt}\right)^2 - \left(\frac{dy}{dt}\right)^2 - \left(\frac{dz}{dt}\right)^2\right) \ . \end{aligned}$$

I.e.,

$$-c^2 = -\left(\frac{dt}{d\tau}\right)^2 (c^2 - \mathbf{u}\cdot\mathbf{u})$$

or

$$\gamma \equiv \left(\frac{dt}{d\tau}\right) = \frac{1}{\sqrt{1-\beta^2}}$$

where $\beta = \sqrt{\mathbf{u}\cdot\mathbf{u}}/c$.

Caution: Although the expressions for γ and β are the same here as they were in the case of the Lorentz boost, the physical meaning is different (related but different).

Thus

$$\boxed{U = (c\,\gamma, \gamma\,\mathbf{u}) = \frac{1}{\sqrt{1-\mathbf{u}^2/c^2}}\ (c,\ \mathbf{u})} \ . \tag{3.6}$$

Example 3.4. Find the proper velocity U for

$$\alpha(s) = \left(c\, s^3,\ 0,\ 0,\ 3\, c\, s^2\right) .$$

As we have seen, when α is proper time parameterized,

$$\alpha(\tau) = \left(c\, (\tau^{2/3}+4)^{3/2},\ 0,\ 0,\ 3\, c\, (\tau^{2/3}+4)\right)$$

for $\tau > 0$. Then

$$U = \frac{d}{d\tau}\,\alpha(\tau) = \left(\, c\, (\tau^{2/3}+4)^{1/2}\tau^{-1/3},\ 0,\ 0,\ 2\, c\, \tau^{-1/3}\right) .$$

■

Example 3.5. Find the 3-velocity of

$$\alpha(s) = \left(c\, s^3,\ 0,\ 0,\ 3\, c\, s^2\right) .$$

To compute the 3-velocity of the particle set $t = s^3$. Then

$$\alpha(t) = \left(c\, t,\ 0,\ 0,\ 3\, c\, t^{2/3}\right) .$$

Therefore,

$$\mathbf{u} = (0,\ 0,\ 2\, c\, t^{-1/3}) = (0,\ 0,\ 2\, c\, (\tau^{2/3}+4)^{-1/2})$$

■

Example 3.6. For α as given above, show explicitly that $U = \gamma(c,\ \mathbf{u})$.

$$U = \frac{d}{d\tau}\,\alpha(\tau) = \left(\, c\, (\tau^{2/3}+4)^{1/2}\tau^{-1/3},\ 0,\ 0,\ 2\, c\, \tau^{-1/3}\right) .$$

$$= (\tau^{2/3}+4)^{1/2}\tau^{-1/3}\left(\, c,\ 0,\ 0,\ 2\, c\, (\tau^{2/3}+4)^{-1/2}\right) .$$

Which is of the form

$$U = (\tau^{2/3}+4)^{1/2}\tau^{-1/3}\ (c,\ \mathbf{u}) .$$

It is easily verified from eq.(3.2) that

$$\frac{dt}{d\tau} = \frac{ds}{d\tau}\frac{dt}{ds} = 3s^2\frac{ds}{d\tau} = (\tau^{2/3}+4)^{1/2}\tau^{-1/3}$$

as required. ■

Transformation Of 3-velocity

Eq.(3.6) imply that

$$\mathbf{u}^i = \frac{U^i}{U^0/c} \tag{3.7}$$

Let $\bar{S}$ be the inertial frame generated by a Lorentz boost (eq.(1.1)) on a given frame S. Then,

$$\bar{U}^0 = \gamma\, U^0 - \beta\gamma\, U^1$$

$$\bar{U}^1 = \gamma\, U^1 - \beta\gamma\, U^0 .$$

Here, β and γ are obtained from the Lorentz boost (and not the speed of the particle). Also,

$$\bar{U}^2 = U^2$$

and

$$\bar{U}^3 = U^3 .$$

Eq.(3.7) gives that

$$\bar{\mathbf{u}}^1 = \frac{c\,\bar{U}^1}{\bar{U}^0} = \frac{c\,[U^1 - \beta U^0]}{U^0 - \beta U^1} = \frac{[c\, U^1/U^0 - \beta c]}{1 - (\beta/c)\ (c\, U^1/U^0)} .$$

I.e, setting $\mathbf{u}^1 = u_x$ we get,

$$\boxed{\bar{u}_x = \frac{(u_x - v)}{(1 - \frac{v}{c^2}\, u_x)}} .$$

Similarly,

$$\boxed{\bar{u}_y = \frac{u_y}{\gamma\,(1 - \frac{v}{c^2}\, u_x)} \quad \text{and} \quad \bar{u}_z = \frac{u_z}{\gamma\,(1 - \frac{v}{c^2}\, u_x)}} .$$

The two equations above give the transformation equation for 3-velocity under boost. Note that they do not transform like a vector.

Under rotations, 3-velocity transforms by a simple 3×3 rotation matrix like in Newtonian physics. So there is nothing new to consider since arbitrary Lorentz transformations are compositions of boosts and rotations.

Example 3.7. A spaceship traveling away from the earth at a speed of $.8c$ fires a torpedo at $.8c$ in the same direction. Find the speed of the torpedo as measured by Earth.

Let S be the frame fixed on the spaceship. All motion is fixed along the x-axis. Then

$$\mathbf{u} = (.8\ c,\ 0,\ 0) .$$

$\bar{S}$ is the frame fixed on Earth. Then the relative velocity of $\bar{S}$ with respect to S is $-.8c$.

$$\bar{u}_x = \frac{(.8c - (-.8c))}{(1 - \frac{(-.8c)}{c^2}\ .8c)} = \frac{1.6\ c}{1.64} = .9756\ c < c .$$

■

Problem 3.1. Find the components of $\mathbf{u}$ in terms of $\bar{\mathbf{u}}$. Hint: Deploy the usual strategy of applying $\beta \to -\beta$ while swapping the barred and the unbarred quantities.

Problem 3.2. The Starship Enterprise, retracting away from a Klingon vessel at half warp, fires a photon torpedo towards the Klingon vessel. Find the speed of the photon torpedo as measured by the Klingon vessel.

In the following problems, let

$$\alpha(s) = \big(2\ c\ (s^2 + 1),\ cs,\ \ 0,\ 0\big) .$$

Problem 3.3. Find α' and determine when α is a future-pointing timelike curve.

Problem 3.4. Find the proper time function τ for α given above.

Problem 3.5. Find the proper velocity $U = \dot{\alpha}$.

Problem 3.6. Compute the 3-velocity of the particle that has α as its world line.

Problem 3.7. A spaceship traveling away from the earth along the x-axis at a speed of $.8c$ fires a torpedo at $.8c$ at an angle of 60^o in the $x - y$ plane of the spaceship. Find the 3 velocity of the torpedo as measured by Earth. The inertial frames located on the Earth and the spaceship are related by the usual Lorentz boost in the x-direction.

Relativistic Force And Momentum

Definition 3.3. *The 4-momentum P of a particle of mass m is*

$$P(\tau) = mU(\tau) = (m\,c\,\gamma, m\,\gamma\,\mathbf{u}) \equiv (m\,c\,\gamma, \mathbf{p})\ .$$

Note, the 3-momentum $\mathbf{p} = m\gamma\mathbf{u}$. This is a redefinition of the classical momentum. It is important to note that conservation of the relativistic momentum has been experimentally verified in the elastic scattering of elementary particles thus lending credence to the modified definition. In the limit $\beta \ll 1$ we have that

$$\gamma = 1 + \frac{1}{2}\beta^2 + \frac{3}{8}\beta^4 + \ldots\ .$$

Therefore, in the small speed (as compared to c) limit, the relativistic momentum does indeed approach its Newtonian value. Since $g(U,U) = -c^2$, we have that

$$\eta(P,\ P) = -(mc)^2\ .$$

Definition 3.4. *(Modified Newton's 2nd Law) The net 4-force K acting on the particle is*

$$K(\tau) = \frac{dP}{d\tau}(\tau)\ .$$

From eq.(3.4) we have that

$$g(K,P) = 0\ .$$

The spatial component of the relativistic force is related to its Newtonian counterpart in the following way:

$$\mathbf{K} = \frac{d\mathbf{p}}{d\tau} = \frac{dt}{d\tau}\,\frac{d\mathbf{p}}{dt} = \frac{dt}{d\tau}\,\mathbf{F}\ .$$

Example 3.8. During a spacewalk, an astronaut, missing the Earth's gravitational field, decides to use his jet pack to accelerate at $1g$ along a fixed direction along the x axis. Find the propertime parameterized world line of the astronaut.

Since there is no motion along the $y,\ z$ axis, we will only concern ourselves with $U^0,\ U^1$. Clearly,

$$\frac{d}{dt}(m\gamma\ \mathbf{u}_x) = mg$$

implies

$$\frac{dU^1}{d\tau} = \frac{g}{c}\,U^0\ . \tag{3.8}$$

Similarly $g(U,\ a) = 0$ implies

$$U^0\,\frac{dU^0}{d\tau} = U^0\,a^0 = U^1\,a^1 = U^1\,\frac{dU^1}{d\tau} = U^1\,\frac{g}{c}\,U^0\ .$$

I.e.,

$$\frac{dU^0}{d\tau} = \frac{g}{c}\,U^1\ . \tag{3.9}$$

Equations (3.8) and (3.9) give that

$$\frac{d^2U^1}{d\tau^2} = \left(\frac{g}{c}\right)^2\,U^1\ .$$

This is easily solved to give

$$U^1(\tau) = a\sinh\left(\frac{g}{c}\,\tau\right) + b\cosh\left(\frac{g}{c}\,\tau\right)\ .$$

Now we set $\mathbf{u}_x = 0$ (starting from rest), and so

$$\gamma(0) = 1\ .$$

Therefore, $U^1(\tau = 0) = 0$, and

$$\frac{dU^1}{d\tau}(0) = \frac{g}{c}\,U^0(0) = g\ .$$

These initial conditions give

$$U^1(\tau) = c \sinh\left(\frac{g}{c}\,\tau\right) .$$

Eq.(3.8) along with $U^0(\tau = 0) = c$ now gives

$$U^0(\tau) = c \cosh\left(\frac{g}{c}\,\tau\right) .$$

Setting $t(0) = 0$ and $x(0) = c^2/g$ for convenience we get

$$ct = \frac{c^2}{g}\,\sinh\left(\frac{g}{c}\tau\right) , \quad \text{and} \quad x = \frac{c^2}{g}\,\cosh\left(\frac{g}{c}\tau\right) .$$

I.e., the particle moves along the hyperbolic curve

$$\left(\frac{g\,x}{c^2}\right)^2 - \left(\frac{g\,ct}{c^2}\right)^2 = 1 .$$

■

Problem 3.8. Show that the path of the astronaut in the problem above describes a future-pointing timelike curve.

Relativistic Energy

Recall that

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(m\gamma\mathbf{u}) .$$

Then the work done on a particle is given by

$$W = \int_{x_i}^{x_f} \mathbf{F}\cdot d\mathbf{x} = \int_{t_i}^{t_f} (\mathbf{F}\cdot\mathbf{u})\,dt = \int_{t_i}^{t_f} \frac{d}{dt}(m\gamma\mathbf{u})\cdot\mathbf{u}\,dt = m\int_{t_i}^{t_f} \mathbf{u}\cdot d(\gamma\mathbf{u}) .$$

On the other hand

$$d(\gamma\mathbf{u}) = \mathbf{u}d\gamma + \gamma d\mathbf{u}$$

and

$$d\gamma = \sum_{i=1}^{3} \frac{\partial\gamma}{\partial u_i}\,du_i = \frac{1}{c^2}\frac{\mathbf{u}\cdot d\mathbf{u}}{(1-\frac{u^2}{c^2})^{3/2}} .$$

Then,

$$\mathbf{u}\cdot d(\gamma\mathbf{u}) = u^2 d\gamma + \gamma\mathbf{u}\cdot d\mathbf{u} \tag{3.10}$$

Combining all the terms above, we get that

$$\mathbf{u}\cdot d(\gamma\mathbf{u}) = \frac{\beta^2\mathbf{u}\cdot d\mathbf{u}}{(1-\beta^2)^{3/2}} + \frac{\mathbf{u}\cdot d\mathbf{u}}{(1-\beta^2)^{1/2}} = \frac{\mathbf{u}\cdot d\mathbf{u}}{(1-\beta^2)^{3/2}} = c^2 d\gamma .$$

I.e.,

$$W = mc^2\int_i^f d\gamma .$$

From the work kinetic energy theorem, the change in kinetic energy of a particle is given by

$$\Delta\ \text{K.E}\ = W = mc^2\int_i^f d\gamma = mc^2\big(\gamma(f) - \gamma(i)\big) .$$

Therefore,

$$\boxed{\text{K.E} = m\gamma c^2 - mc^2} \tag{3.11}$$

so that K.E= 0 for a particle at rest. In this case, in the limit where $\beta << 1$, we get that

$$\text{K.E} = mc^2\left(1 - \frac{u^2}{c^2}\right)^{-1/2} - mc^2 \approx mc^2\left(1 + \frac{u^2}{2c^2} + ...\right) - mc^2 .$$

Therefore in the non-relativistic limit

$$\text{K.E} = \frac{1}{2}mu^2$$

as expected. Recall that,

$$\Delta E = \Delta(\text{K.E} + \text{P.E}) \ ,$$

where P.E is the potential energy of the particle. Then, for a free particle, since $\Delta\text{P.E} = 0$, we get that

$$\Delta\text{K.E} = \Delta E \ ,$$

where E is the total energy of the particle. I.e.,

$$\boxed{E = m\gamma c^2}$$

up to an arbitrary constant, which we set to 0. In this case, the rest energy of an object is given by the famous equation

$$\boxed{E_{rest} = mc^2} \ .$$

Here $\gamma = 1$ since an object at rest in not moving. The relation between E and $|\mathbf{p}| = \sqrt{\mathbf{p}\cdot\mathbf{p}}$ of a particle can now be obtained by noting that

$$|\mathbf{p}|^2c^2 = m\gamma^2u^2c^2 = m^2\gamma^2c^4\beta^2 = m^2\gamma^2c^4(1 - \frac{1}{\gamma^2}) = m^2\gamma^2c^4 - m^2c^4$$

or

$$\boxed{-E^2 + |\mathbf{p}|^2c^2 = -(mc^2)^2} \ [1] \ .$$

The above equation is sometimes referred to as the **mass-shell condition** for a free particle. Note that in the above equation, setting $m = 0$ implies that

$$\boxed{E = |\mathbf{p}|c} \ [2] \ .$$

All massless particles are expected to satisfy the above relationship. You should already know from your Modern Physics course that this is true for photons.

Recall that

$$1eV = 1.6 \times 10^{-19}J \ .$$

The electron volt $(eV)$is a useful unit of energy when considering elementary particles. For example, the rest energy of an electron is $m_ec^2 = .511MeV$. In this case, the unit for momentum becomes

$$[\mathbf{p}] = \frac{M_ev}{c} \ .$$

Example 3.9. An electron with energy $1MeV$ makes a head-on collision with a positron at rest. If the e^-e^+ pair annihilates to produce two gamma ray photons of equal energy moving at angle θ with respect to the initial path of the electron. Find the energy and momentum of the photons and the scattering angle θ.

Conservation of momentum implies that

$$\mathrm{p}_e = 2p\cos\theta \ .$$

Here p_e is the initial 3-momentum of the electron and p is the 3-momentum of the photon. Conservation of energy gives

$$m\gamma c^2 + mc^2 = 2pc \ .$$

[1] You might wonder why I placed the negative signs in such a curious order. In the next chapter, we will return to this point.

[2] We have taken liberties here and assumed that E and $|\mathbf{p}|$ does not depend on the mass of the object. As long as we are being careless, we can continue the argument and set

$$E = m\gamma c^2 = |\mathbf{p}|c = m\gamma uc \ ,$$

and thereby conclude that such particles must travel at the speed of light, i.e.,

$$u = c \ .$$

Also,

$$1MeV = m\gamma c^2 = \gamma(.511MeV)$$

implies that

$$\gamma = 1.96 \text{ and } \beta = 0.8 \ .$$

Therefore,

$$\mathrm{p}_e = m\gamma \frac{u}{c} c = m\gamma\beta c = \frac{m\gamma c^2 \beta}{c} = (.511)(1.96)(.8)\frac{M_e v}{c} = 0.8\frac{M_e v}{c} \ .$$

The energy conservation equation implies that

$$1MeV + .511MeV = 2pc$$

or

$$p = 0.76\frac{M_e v}{c} \ ,$$

and so for the photon

$$E = 0.76MeV \ .$$

Finally, momentum conservation implies that

$$0.8\frac{M_e v}{c} = 2(.76)\frac{M_e v}{c}\cos\theta$$

which gives

$$\theta = 64.4^o \ .$$

■

$g(K, P) = 0$ implies that

$$0 = -m\gamma c\frac{d}{d\tau}(m\gamma c) + \frac{dt}{d\tau}\ \mathbf{F}\cdot\mathbf{p}$$

or noting that $\dot{t} = \gamma$ we get

$$\frac{d}{dt}(m\gamma c^2) = \mathbf{F}\cdot\mathbf{u} \ .$$

I.e.,

$$d(m\gamma c^2) = \mathbf{F}\cdot d\mathbf{r} \ .$$

In the event that the net force acting on an object is conservative, i.e.,

$$\mathbf{F} = -\nabla\phi$$

for some scalar function ϕ, we have that

$$(m\gamma c^2)_{\mathbf{r}_f} - (m\gamma c^2)_{\mathbf{r}_i} = -\int_{\mathbf{r}_i}^{\mathbf{r}_f} \nabla\phi\cdot d\mathbf{r} = \phi(\mathbf{r}_i) - \phi(\mathbf{r}_f) \ .$$

Thus, we get the conserved quantity

$$E = m\gamma c^2 + \phi \ .$$

For small u

$$E = mc^2 + \left(\frac{1}{2}mu^2 + \phi\right) + ... \ .$$

The quantity in the parentheses is the sum total of the classical kinetic and potential energy of the object. The higher-order terms give relativistic corrections.

3.1.1 Massless Particles

The "photon" is an example of a massless particle. The speculated "graviton" is also a massless particle. The worldline of a photon is a curve $\alpha(s)$ such that

$$\frac{d\alpha}{ds} = P \ ,$$

the four momentum of a photon

$$P = \left(\frac{E}{c},\ \mathbf{p}\right) \ .$$

The important point is that for a particular photon, P is constant throughout its worldline (as long as it exists as a distinct entity, the photon is a free particle). Then

$$ct = \frac{E}{c}s + ct_0$$

and

$$\mathbf{x} = \mathbf{p}s + \mathbf{x}_0 \ .$$

There is no notion of proper time for a photon.

Transformation Of Momentum

$P = m\ dU/d\tau$ implies that, at least for particles with mass, P transforms as a vector. We postulate that the 4-momentum of massless particles also transform as vectors. For example, under a Lorentz boost in the x-direction, regardless of a particles mass

$$\frac{\bar{E}}{c} = \gamma\ \left(\frac{E}{c} - \frac{v}{c}\ \mathbf{p}_x\right)$$

and

$$\bar{\mathbf{p}}_x = \gamma\ \left(\mathbf{p}_x - v\ \frac{1}{c}(E/c)\right) \ ,$$

while $\bar{\mathbf{p}}_y = \mathbf{p}_y$ and $\bar{\mathbf{p}}_z = \mathbf{p}_z$.

Conservation Of 4-Momentum

Fundamental interactions of elementary particles respect Newton's 3rd law: *For every action, there is an equal and opposite reaction.* I.e., since forces come in equal and opposite pairs, the net force on a closed system is zero. Therefore,

$$P_{\text{net, initial}} = P_{\text{net, final}} \ .$$

Equivalently, the total 4-momentum before and after the event (collision, particle creation and/or annihilation, etc.) is the same. This is the law of conservation of momentum. Separating the space and time components we get the law of conservation of energy and 3-momentum $\mathbf{p}$.

There are numerous examples of particle interactions where the relativistic form of total energy and momentum of the interacting particles are conserved. These results essentially validate special relativity. Here is one such example:

Example 3.10. In a particle physics lab, a photon with energy E_{ph} collides with an electron initially at rest. Upon collision, the initial photon is annihilated, creating an electron-positron pair. The newly created electron position pair, along with the initial electron, moves with a speed u in the direction of motion of the initial photon. Find E_{ph} and u.

Solution In the frame S of the physics lab, the total initial energy

$$E_i = E_{\text{ph}} + m\ c^2 \ .$$

Here, m is the mass of electron (/positron). Total initial momentum

$$p_i = E_{\text{ph}}/c \ .$$

In the rest frame $\bar{S}$ of the electron-positron cluster, the total final energy

$$\bar{E}_f = 3mc^2 \ ,$$

and the total final momentum

$$\bar{p}_f = 0 \ .$$

Since

$$E = \gamma \ (\bar{E} + v \ \bar{p}_x) \quad \text{and} \quad p_x = \gamma \ (\bar{p}_x + \frac{v}{c^2}\bar{E}) \ ,$$

we get

$$E_f = 3m\gamma c^2 \quad \text{and} \quad p_f = 3 \ m \ \gamma \ u \ .$$

Setting $p_i = p_f$ we get $E_{\rm ph} = 3m \ \gamma \ u \ c$, and $E_i = E_f$ we get

$$3m\gamma uc + mc^2 = 3m\gamma c^2 \ .$$

Or

$$mc^2 = 3m\gamma \ (c^2 - uc)$$

or

$$\frac{1}{3} = \gamma \ (1-\beta) = \frac{1}{\sqrt{(1-\beta)(1+\beta)}}(1-\beta) = \sqrt{\frac{1-\beta}{1+\beta}} \ .$$

Therefore, $\beta = u/c = 4/5$ and $E_{\rm ph} = 4mc^2$.

In the previous example, $E_{\rm ph} = 4mc^2$ as opposed to $2mc^2$ (for the two new particles created) since the three particles carry kinetic energy of

$$T = 3m\gamma c^2 - 3mc^2 = 5mc^2 - 3mc^2 = 2mc^2 \ .$$

■

Binding Energy

The total energy of a bound system, like the nucleus of an atom, is less than the combined energy of the separated nucleons. This is in fact, why nucleons bind. For example, the nucleus of the deuterium atom, called the deuteron, consists of a proton and a neutron. The difference in mass

$$\Delta m = (m_p + m_n) - m_d = [\ (1.007825 + 1.008665) - 2.014102 \ ] \ u \ .$$

The binding energy

$$E_b = .002388 \ u \ \frac{931.5 MeV}{u} = 2.224 \ MeV \ .$$

Alpha Decay

Unstable nuclei have a larger mass than its constituent nucleons. A frequently seen decay process is by the emission of the nucleus of a Helium atom (the α particle). A typical decay of Radium to Radon looks like

$$^{226}_{88} Ra \ \rightarrow \ ^{222}_{86} Rn \ + \ ^{4}_{2}He \ .$$

The mass difference in this mechanism is given by

$$\begin{aligned} \Delta m = m_{Ra} - (m_{Rn} + m_{He}) &= [\ 226.025402 - (222.017571 + 4.002602) \ ] \ u \ . \\ &= \ .005229 \ u \ . \end{aligned}$$

The decay energy

$$E_d = .005229 \ u \ \frac{931.5 MeV}{u} = 4.871 \ MeV \ .$$

Here, the lighter object (namely the alpha particle) carries most of this energy in the form of kinetic energy. This is can easily be seen by noting that, in the non-relativistic case

$$KE = \frac{p^2}{2m}$$

and the momentum p is the same for both Radon and the alpha particle due to momentum conservation.

Problem 3.9. Show that

$$1u = 931.494 \ \frac{MeV}{c^2} \ .$$

Problem 3.10. Show that when $\beta \ll 1$ we have that

$$\gamma = 1 + \frac{1}{2}\beta^2 + \frac{3}{8}\beta^4 + ... \ .$$

Problem 3.11. In example 3.8, assume that the astronaut launched from Earth.

1. When 30 years has passed on earth, how much older is the astronaut? How far from Earth has the astronaut traveled?
2. When the astronaut is 30 years older, how much time has lapsed on Earth? How far from the Earth is the astronaut?

Problem 3.12. A photon traveling in the y-direction has momentum $3.2 \times 10^{-9} Kg \ m/s$. Find the energy of the photon.

Problem 3.13. An electron traveling in the x-direction has energy $2MeV$. Find the magnitude of the momentum of the electron. Find the speed of the electron.

Problem 3.14. Calculate the binding energy of ${}^4He$ (atomic mass $= 4.002603u$). ${}^4He$ has two protons and two neutrons.

Problem 3.15. Calculate the energy released in the alpha decay of ${}^{238}_{92}U$ (atomic mass $= 238.050783 \ u$) into ${}^{234}_{90}Th$ (atomic mass $= 234.043583 \ u$).

Relativistic Covariance of Mechanics

The laws of mechanics now take the form

$$ma^\mu = K^\mu \ .$$

As long as under the action of $\Lambda \in \mathcal{L}_+^\uparrow$, the external force K is such that

$$\bar{K}^\mu = \Lambda^\mu{}_\nu \ K^\nu$$

we would have that

$$m\bar{a}^\mu = \bar{K}^\mu \ .$$

This is what we mean when we say that the laws of mechanics take the same form in every inertial frame. As you will see below, this is indeed the case for K in electrodynamics.

3.2 A First Look at Tensors

Recall that under a coordinate transformation of type

$$\bar{x}^\mu = \Lambda^\mu{}_\nu \ x^\nu + a^\mu$$

vectors transform as

$$\bar{A}^\mu = \Lambda^\mu{}_\nu \ A^\nu \ . \tag{3.12}$$

Recall that the tangent space at any $x \in \mathbb{M}^4$ is denoted as $T_x(\mathbb{M}^4)$. In the past we referred to elements of $T_x(\mathbb{M}^4)$ as just vectors. But since we are getting ready to define a new vector space, henceforth elements of $T_x(\mathbb{M}^4)$ will be referred to as Minkowski *tangent vectors* or *contravariant Vectors*. In particular, Minkowski tangent vectors will transform as given by eq. (3.12).

Dual Vectors

As we have seen before, for vectors A and B, the quantity $\eta(A,B)$ is invariant under the action of a general homogeneous Lorentz transformation and spacetime translations. I.e.,

$$\eta(A,B) = \eta(\bar{A},\bar{B}) \ .$$

This allows us to define a new kind of vector with lower indices, namely, objects of type A_μ defined by

$$A_\mu = \eta_{\mu\nu} A^\nu \ .$$

Theorem 3.3. *$A_\mu B^\mu \equiv A_0 B^0 + A^1 B_1 + A^2 B_2 + A^3 B_3$ is invariant under a general homogeneous transformation and spacetime translations, provided a dual vector A_μ transforms as*

$$\bar{A}_\mu = \Lambda_\mu{}^\nu A_\nu \ .$$

Proof.

$$\bar{A}_\mu \bar{B}^\mu = \Lambda_\mu{}^\theta A_\theta \ \Lambda^\mu{}_\nu B^\nu = (\Lambda^{-1})^\theta{}_\mu \ \Lambda^\mu{}_\nu A_\theta B^\nu = A_\nu B^\nu \ .$$

□

Thus, we see that A_μ has well-defined transformation properties. Objects that transform according to a rule

$$\bar{A}_\mu = \Lambda_\mu{}^\nu A_\nu \tag{3.13}$$

shall henceforth be referred to as a *covariant vector*, or a *dual vector*. Note that the above transformation property is equivalent to

$$\bar{A}_\mu = (\Lambda^{-1})^\nu{}_\mu \ A_\nu \ .$$

Definition 3.5. *The set of all vectors $A_\mu(x)$ at any $x \in \mathbb{M}^4$, clearly forms a vector space at x that is denoted by $T_x^\star(\mathbb{M}^4)$. An element of $T_x^\star(\mathbb{M}^4)$ is referred to as a dual vector at x.*

Definition 3.6. *A dual vector field X in $\mathbb{M}^4$ is a differentiable map such that $X(x) \in T_x^\star(\mathbb{M}^4)$ for any $x \in \mathbb{M}^4$.*

Consider the map from

$$T_x(\mathbb{M}^4) \to T_x^*(\mathbb{M}^4)$$

given by

$$A^\mu \to A_\mu = \eta_{\mu\nu} A^\nu \ .$$

In coordinates,

$$(A^0, A^1, A^2, A^3) \to (A_0, A_1, A_2, A_3) = (-A^0, A^1, A^2, A^3) \ .$$

Its inverse

$$T_x^*(\mathbb{M}^4) \to T_x(\mathbb{M}^4)$$

is clearly given by

$$A^\mu = \eta^{\mu\nu} A_\nu \ .$$

Note:

1. From above, we see that $\eta_{\mu\nu}$ and $\eta^{\mu\nu}$ provide a $1-1$, onto and invertible map between $T_x(\mathbb{M}^4)$ and $T_x^*(\mathbb{M}^4)$.
2. Theorem 3.3 articulates the fundamental role of a dual vector: it operates on a tangent vector in a linear fashion to yield a real number, in the sense that

$$A_\mu \left(B^\mu + C^\mu\right) = A_\mu B^\mu + A_\mu C^\mu \ .$$

Problem 3.16. Show that, under a coordinate transformation of type

$$\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \,,$$

we get that

$$\Lambda^\mu{}_\nu = \frac{\partial \bar{x}^\mu}{\partial x^\nu} \,,$$

and that

$$\Lambda_\nu{}^\mu = \frac{\partial x^\mu}{\partial \bar{x}^\nu} \,.$$

Consequently, we have that

$$\frac{\partial \bar{x}^\mu}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial \bar{x}^\nu} = \frac{\partial x^\mu}{\partial \bar{x}^\alpha}\frac{\partial \bar{x}^\alpha}{\partial x^\nu} = \delta^\mu_\nu \,.^{3}$$

From the problem above we have that, under a coordinate transformation of type

$$\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \,,$$

tangent vectors/dual vectors transform according to the rule

$$\bar{A}^\mu = \frac{\partial \bar{x}^\mu}{\partial x^\nu}\, A^\nu \,, \quad \bar{A}_\mu = \frac{\partial x^\nu}{\partial \bar{x}^\mu}\, A_\nu \,. \tag{3.14}$$

Problem 3.17. Verify

$$\frac{\partial \bar{x}^\mu}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial \bar{x}^\nu} = \frac{\partial x^\mu}{\partial \bar{x}^\alpha}\frac{\partial \bar{x}^\alpha}{\partial x^\nu} = \delta^\mu_\nu$$

for the case of transformation of coordinates (x, y, z) to (r, θ, φ) in $\mathbb{R}^3$.

Definition 3.7. *Under a coordinate transformation of type*

$$\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \,,$$

a tensor of rank (r, s) at a point $p \in \mathbb{M}^4$ is an object that transforms according to the rule

$$\bar{T}^{\mu_1\ldots\mu_r}{}_{\nu_1\ldots\nu_s}(\bar{p}) = \Lambda^{\mu_1}{}_{\alpha_1}\ldots\Lambda^{\mu_r}{}_{\alpha_r}\,\Lambda_{\nu_1}{}^{\beta_1}\ldots\Lambda_{\nu_s}{}^{\beta_s}\,T^{\alpha_1\ldots\alpha_r}{}_{\beta_1\ldots\beta_s}(p) \,.$$

Definition 3.8. *Under a coordinate transformation of type*

$$\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \,,$$

a tensor field of rank (r, s) is an object that depends on the points $x \in \mathbb{M}^4$ that transforms according to the rule

$$\bar{T}^{\mu_1\ldots\mu_r}{}_{\nu_1\ldots\nu_s}(\bar{x}) = \Lambda^{\mu_1}{}_{\alpha_1}\ldots\Lambda^{\mu_r}{}_{\alpha_r}\,\Lambda_{\nu_1}{}^{\beta_1}\ldots\Lambda_{\nu_s}{}^{\beta_s}\,T^{\alpha_1\ldots\alpha_r}{}_{\beta_1\ldots\beta_s}(x) \,.$$

Here $\bar{x}$ locates the point x in the transformed coordinate system.

From eq.(3.14), we see that the tensor transformation rule can also be written in the form

$$\bar{T}^{\mu_1\ldots\mu_r}{}_{\nu_1\ldots\nu_s}(\bar{x}) = \frac{\partial \bar{x}^{\mu_1}}{\partial x^{\alpha_1}}\ldots\frac{\partial \bar{x}^{\mu_r}}{\partial x^{\alpha_r}}\;\frac{\partial x^{\beta_1}}{\partial \bar{x}^{\nu_1}}\;\ldots\frac{\partial x^{\beta_s}}{\partial \bar{x}^{\nu_s}}\;\;T^{\alpha_1\ldots\alpha_r}{}_{\beta_1\ldots\beta_s}(x) \,.$$

Problem 3.18. Show that under

$$\bar{x}^\mu = \Lambda^\mu{}_\nu\, x^\nu + a^\mu \,,$$

for $\Lambda \in \mathcal{L}$, the metric is invariant as a rank $(0, 2)$ tensor. I.e., verify that $\bar{\eta}_{\mu\nu} = \eta_{\mu\nu}$, where

$$\bar{\eta}_{\mu\nu} = \Lambda_\mu{}^\alpha\,\Lambda_\nu{}^\beta\,\eta_{\alpha\beta} \,.$$

[3] In our case this expression follows immediately. But this is true in general. See problem 3.17 for a non-trivial case.

Mixed Tensors and Tensor Products

There is no real need to place all the contravariant indices in the beginning, followed by all the covariant indices. Nothing precludes the existence of tensors of type

$$S_{\alpha}{}^{\beta}{}_{\gamma} \, .$$

The transformation properties of such an object should be obvious. Such objects are sometimes called *mixed* tensors. It is also possible to define a product of tensors. For example the product of tensors $S^{\mu}{}_{\nu}$ and $T^{\alpha\beta}$ denoted by

$$S^{\mu}{}_{\nu} \cdot T^{\alpha}{}_{\beta}$$

can be viewed as a new tensor that transforms according to the rule

$$\bar{S}^{\mu}{}_{\nu} \cdot \bar{T}^{\alpha}{}_{\beta} = \Lambda^{\mu}{}_{a} \Lambda_{\nu}{}^{b} \Lambda^{\alpha}{}_{c} \Lambda_{\beta}{}^{d} \, S^{a}{}_{b} \cdot T^{c}{}_{d} \, .$$

The above transformation property can be extended to the product of arbitrary mixed tensors.

Contraction of Tensor Indices

Similarly, it is possible to contract any two indices of a tensor. For example, consider a tensor of type $S^{\mu\nu}{}_{\alpha}$. Then the quantity

$$S^{\mu\nu}{}_{\nu}$$

is a well-defined tangent vector. To see this, note that

$$\bar{S}^{\mu\nu}{}_{\alpha} = \Lambda^{\mu}{}_{a} \, \Lambda^{\nu}{}_{b} \, \Lambda_{\alpha}{}^{c} \, S^{ab}{}_{c} \, .$$

Therefore

$$\bar{S}^{\mu\nu}{}_{\nu} = \Lambda^{\mu}{}_{a} \, \Lambda^{\nu}{}_{b} \, \Lambda_{\nu}{}^{c} \, S^{ab}{}_{c} = \Lambda^{\mu}{}_{a} \, \; S^{ab}{}_{b} \, .^{4}$$

Or

$$\bar{S}^{\mu\nu}{}_{\nu} = \Lambda^{\mu}{}_{\lambda} \; S^{\lambda\nu}{}_{\nu} \, ,$$

which means that you can contract the indices and then transform the remaining indices of a tensor, or transform the tensor (without contracting the indices) and then contract the indices. Regardless of the order of operations, we get the same result. In exactly the same way, we can contract indices of products of tensors. As an example, this could take on the form

$$S^{\mu}{}_{\alpha} \cdot T^{\alpha}{}_{\beta} \, ,$$

or even the familiar operation

$$\eta_{\mu\nu} \cdot A^{\nu} = \eta_{\mu\nu} \, A^{\nu} \, .$$

When indices of tensors are contracted, the resulting tensor is clearly of one less covariant and contravariant rank.

Raising and Lowering of Indices

Problem 3.19. Show that

$$\eta_{\mu\nu} \bar{A}^{\nu} = \overline{\eta_{\mu\nu} A^{\nu}} \, ,$$

and

$$\eta^{\mu\nu} \bar{A}_{\nu} = \overline{\eta^{\mu\nu} A_{\nu}} \, .$$

In this manner, all tensor indices can be lowered and raised by η and η^{-1}. For example, given a tensor $S^{\mu\nu}{}_{\alpha}$, we can obtain a new tensor $S^{\mu}{}_{\nu\alpha}$ by

$$S^{\mu}{}_{\nu\alpha} = \eta_{\nu\lambda} \, S^{\mu\lambda}{}_{\alpha}$$

Problem 3.20. How are the components of $S^{\mu}{}_{\nu}{}^{\alpha}$ related to $S^{\mu\nu}{}_{\alpha}$?

[4] What is crucial here is the fact that

$$\Lambda^{\nu}{}_{b} \, \Lambda_{\nu}{}^{c} = \delta^{c}_{b} \, .$$

3.3 Relativistic Electrodynamics

3.3.1 The Klein-Gordon Field

As a warm-up to relativistic field theory, let us first consider the simplest case of a scalar field. Here, it will be easy to see how Lorentz invariance applies to fields. Recall that the mass-shell equation is given by

$$-\frac{E^2}{c^2} + p^2 + m^2c^2 = 0 \, .$$

From Quantum Mechanics, the Hamiltonian Operator E is given by,

$$E = \hat{H} \equiv i\hbar \frac{\partial}{\partial t} \, ,$$

and the momentum operator

$$\hat{\mathbf{p}} \equiv -i\hbar\vec{\nabla} \, .$$

Inserting the operator expressions for E^2 and p^2 into the mass shell equation, we get that

$$\hbar^2 \frac{\partial^2}{\partial(ct)^2} - \hbar^2\nabla^2 + m^2c^2 = 0$$

which can be viewed as an operator equation. Consider a particle described by a scalar field (a single real-valued function) ϕ on $\mathbb{M}^4$. The operator equation above can then be thought of as acting on ϕ, resulting in

$$-\frac{\partial^2\phi}{\partial(ct)^2} + \nabla^2\phi - \frac{m^2c^2}{\hbar^2}\phi = 0 \, .$$

The above equation can be written more compactly as

$$\eta^{\mu\nu}\frac{\partial^2\phi}{\partial x^\mu \partial x^\nu} - \frac{m^2c^2}{\hbar^2}\phi = 0 \, . \tag{3.15}$$

This is the Klein-Gordon equation. The Higgs field in particle physics and the inflaton field in cosmology is indeed attributed to the above Klein-Gordon equation.

Now, we must determine whether the Klein-Gordon equation is relativistically correct or not, i.e., whether or not the equation is Lorentz invariant. To this end, consider two frames with coordinate charts $\{x^\mu\}$ and $\{\bar{x}^\mu\}$ related by

$$\bar{x}^\mu = \Lambda^\mu{}_\nu \, x^\nu + a^\mu \, .$$

In these two frames, the scalar field is such that

$$\bar{\phi}(\bar{x}) = \phi(x) \, .^{5}$$

Suppose in $\{x^\mu\}$, eq. (3.15) holds true. The derivative operator transforms according to the chain rule

$$\frac{\partial}{\partial x^\mu}\phi(x) = \frac{\partial \bar{x}^\alpha}{\partial x^\mu}\frac{\partial}{\partial \bar{x}^\alpha}\bar{\phi}(\bar{x}) = \Lambda^\alpha{}_\mu \frac{\partial}{\partial \bar{x}^\alpha}\bar{\phi}(\bar{x}) \, .$$

For ease of notation, we set

$$\partial_\mu \equiv \frac{\partial}{\partial x^\mu} \, , \text{and} \;\; \bar{\partial}_\mu \equiv \frac{\partial}{\partial \bar{x}^\mu} \, .$$

Then

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\phi(x) = \eta^{\mu\nu}\,\Lambda^\alpha{}_\mu\,\Lambda^\beta{}_\nu\,\bar{\partial}_\alpha\bar{\partial}_\beta\,\bar{\phi}(\bar{x}) = \eta^{\alpha\beta}\bar{\partial}_\alpha\bar{\partial}_\beta\,\bar{\phi}(\bar{x}) \, .$$

The last equality was obtained using eq.(2.12). Thus, it follows that, eq.(3.15) in the chart $\{x^\mu\}$ implies that

$$\eta^{\alpha\beta}\bar{\partial}_\alpha\bar{\partial}_\beta\,\bar{\phi}(\bar{x}) - \frac{m^2c^2}{\hbar^2}\bar{\phi}(\bar{x}) = 0 \, .$$

I.e., the Klein-Gordon equation has the same appearance in the chart $\{\bar{x}^\mu\}$ This is what we mean when we say that the Klein-Gordon equation is Lorentz invariant.

Problem 3.21. Let $K^\mu = (K^0, \mathbf{K})$ be a constant vector. How are the components of K related so that

$$\phi(x) = A\, e^{i\eta_{\mu\nu}K^\mu x^\nu}$$

is a solution to the Klein-Gordon equation. Here A is a constant. Is ϕ an eigenfunction of E and $\mathbf{p}$? If so, what are the corresponding eigenvalues? What can you conclude about the causal character of K? Note that the solution is clearly Lorentz invariant.

[5] It is precisely this requirement that makes ϕ a scalar field.

3.3.2 Electrodynamics

Maxwell's equations for the electric and magnetic field in **S.I** [6] units are given by

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} \,, \quad \nabla \cdot \mathbf{B} = 0 \,,$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \,, \text{ and } \nabla \times \mathbf{B} = \frac{1}{c^2}\frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{J} \,.$$

Here, ρ is the charge density, $\mathbf{J}$ is the current density, $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field and $c^2 = (\mu_0 \epsilon_0)^{-1}$. The Lorentz force on a charged particle q is given by

$$\mathbf{F} = q(\mathbf{E} + \mathbf{u} \times \mathbf{B}) \,. \tag{3.16}$$

In order to formulate a relativistic theory of electrodynamics, we have experimentally determine how these quantities transform under a Lorentz transformation. Suffice it to say that upon extensive verification, it has been deemed that indeed the 4-current

$$J^\mu = (c\rho, \mathbf{J})$$

transforms a 4-vector, and the quantities

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}$$

transforms as a rank two contravariant tensor. In any frame, given $F^{\mu\nu}$, we can determine $\mathbf{E}$ and $\mathbf{B}$ using the mnemonic

$$E_i = F^{ti} \,, \text{ and } B_i = \varepsilon_{ijk} F^{jk} \,.^{7}$$

$F^{\mu\nu}$ is referred to as the Maxwell tensor, and under the action of the Poincaré group element (Λ, a)

$$\bar{F}^{\mu\nu}(\bar{x}) = \Lambda^\mu{}_\alpha \Lambda^\nu{}_\beta \, F^{\alpha\beta}(x) \,,$$

where, as usual

$$\bar{x}^\mu = \Lambda^\mu{}_\alpha x^\alpha + a^\mu \,.$$

Problem 3.22. Show that

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix} \,.$$

Define $\partial_{[\mu} F_{\nu\lambda]}$ by the expression

$$\partial_{[\mu} F_{\nu\lambda]} \equiv \frac{1}{3!}[\partial_\mu F_{\nu\lambda} - \partial_\mu F_{\lambda\nu} + \partial_\nu F_{\lambda\mu} - \partial_\nu F_{\mu\lambda} + \partial_\lambda F_{\mu\nu} - \partial_\lambda F_{\nu\mu}] \,.$$

In the above expression, we are taking an anti-symmetric sum of all the indices within the square brackets. The factor of 1/3! is a convention and a useful normalization. [8] Since $F_{\mu\nu}$ is anti-symmetric, we get that

$$\partial_{[\mu} F_{\nu\lambda]} = \frac{1}{3}[\partial_\mu F_{\nu\lambda} + \partial_\nu F_{\lambda\mu} + \partial_\lambda F_{\mu\nu}] \,.$$

Theorem 3.4. *The homogeneous Maxwell's equations are given by the equation*

$$\partial_{[\mu} F_{\nu\lambda]} = 0 \,. \tag{3.17}$$

The above equation is referred to as the Bianchi identity for electromagnetism.

[6] $c^2 = (\mu_0 \epsilon_0)^{-1}$.
[7] These are clearly not tensor equations. Here ε_{ijk} is the completely anti-symmetric object such that $\varepsilon_{123} = 1$.
[8] In the language of differential calculus, this combination of derivatives is called the exterior derivative of F.

Proof. Set $\mu = \nu = 0, \lambda = 1$. Then, trivially

$$\frac{\partial F_{01}}{\partial ct} + \frac{\partial F_{10}}{\partial ct} + \frac{\partial F_{00}}{\partial x} = 0 \,,$$

and we do not get any meaningful information in this case. In the same way, if any two of the three indices are the same, we get a trivial result that do not correspond to Maxwell's equations. Now set $\mu = 0, \nu = 1,$ and $\lambda = 2$. In this case, we get that

$$\frac{1}{c}\frac{\partial B_z}{\partial t} + \frac{\partial E_x}{\partial y} + \frac{\partial E_y}{\partial x} = 0 \,,$$

or

$$(\nabla \times \mathbf{E})_z + \frac{1}{c}\frac{\partial B_z}{\partial t} = 0 \,.$$

Similarly, $\mu = 0, \nu = 1, \lambda = 3$ and $\mu = 0, \nu = 2, \lambda = 3$ give the remaining two components of Faraday's law. The only remaining choice of indices is $\mu = 1, \nu = 2, \lambda = 3$, wherein we get

$$\frac{\partial F_{23}}{\partial x} + \frac{\partial F_{31}}{\partial y} + \frac{\partial F_{12}}{\partial z} = 0 \,,$$

or $\nabla \cdot \mathbf{B} = 0$ □

Theorem 3.5. *Under the action of the Poincaré group, eq.(3.17) preserves its form. In other words, the Bianchi identity is covariant, i.e.,*

$$\bar{\partial}_{[\mu}\bar{F}_{\nu\lambda]} = 0 \iff \partial_{[\mu}F_{\nu\lambda]} = 0 \,.$$

Proof.

$$\bar{\partial}_{[\mu}\bar{F}_{\nu\lambda]} = (\Lambda^{-1})^{\alpha}{}_{\mu}(\Lambda^{-1})^{\beta}{}_{\nu}(\Lambda^{-1})^{\gamma}{}_{\lambda}\,\partial_{\alpha}F_{\beta\gamma}$$
$$+(\Lambda^{-1})^{\alpha}{}_{\nu}(\Lambda^{-1})^{\beta}{}_{\lambda}(\Lambda^{-1})^{\gamma}{}_{\mu}\,\partial_{\alpha}F_{\beta\gamma} + (\Lambda^{-1})^{\alpha}{}_{\lambda}(\Lambda^{-1})^{\beta}{}_{\mu}(\Lambda^{-1})^{\gamma}{}_{\nu}\,\partial_{\alpha}F_{\beta\gamma} \,.$$

A simple rearrangement of dummy indices shows that

$$(\Lambda^{-1})^{\alpha}{}_{\nu}(\Lambda^{-1})^{\beta}{}_{\lambda}(\Lambda^{-1})^{\gamma}{}_{\mu}\,\partial_{\alpha}F_{\beta\gamma} = (\Lambda^{-1})^{\alpha}{}_{\mu}(\Lambda^{-1})^{\beta}{}_{\nu}(\Lambda^{-1})^{\gamma}{}_{\lambda}\partial_{\beta}F_{\gamma\alpha} \,,$$

and

$$(\Lambda^{-1})^{\alpha}{}_{\lambda}(\Lambda^{-1})^{\beta}{}_{\mu}(\Lambda^{-1})^{\gamma}{}_{\nu}\,\partial_{\alpha}F_{\beta\gamma} = (\Lambda^{-1})^{\alpha}{}_{\mu}(\Lambda^{-1})^{\beta}{}_{\nu}(\Lambda^{-1})^{\gamma}{}_{\lambda}\partial_{\gamma}F_{\alpha\beta} \,.$$

Substituting the above two expressions into the original expression for $\bar{\partial}_{[\mu}\bar{F}_{\nu\lambda]}$ gives

$$\bar{\partial}_{[\mu}\bar{F}_{\nu\lambda]} = (\Lambda^{-1})^{\alpha}{}_{\mu}(\Lambda^{-1})^{\beta}{}_{\nu}(\Lambda^{-1})^{\gamma}{}_{\lambda}[\partial_{\alpha}F_{\beta\gamma} + \partial_{\beta}F_{\gamma\alpha} + \partial_{\gamma}F_{\alpha\beta}]$$
$$= (\Lambda^{-1})^{\alpha}{}_{\mu}(\Lambda^{-1})^{\beta}{}_{\nu}(\Lambda^{-1})^{\gamma}{}_{\lambda}\partial_{[\alpha}F_{\beta\gamma]} \,.$$

The claim is now obvious. □

Problem 3.23. Show that in covariant form, the Lorentz force on a charge q is given by

$$K^{\mu} = qF^{\mu\nu}U_{\nu} \,. \tag{3.18}$$

Here U is the 4-velocity of the charge. I.e., show that the spatial components of the Lorentz force in eq.(3.18) give eq.(3.16), and when $\mu = 0$, this is simply the equation for power.

Problem 3.24. Show that the inhomogeneous Maxwell's Equation is given by

$$\partial_{\nu}F^{\mu\nu} = \mu_0 J^{\mu} \,.$$

Problem 3.25. Show that the above equation is covariant. I.e. show that

$$\frac{\partial}{\partial \bar{x}^{\nu}}\bar{F}^{\mu\nu} = \mu_0 \bar{J}^{\mu}$$

where $\bar{J}^{\mu} = \Lambda^{\mu}{}_{\nu}\,J^{\nu}$ and $\bar{F}^{\mu\nu}$ is as given above.

Problem 3.26. Show that

$$F_{\mu\nu}F^{\mu\nu} = 2\left(B^2 - \frac{E^2}{c^2}\right) ,$$

where $E^2 = \mathbf{E}\cdot\mathbf{E}$ and similarly for B^2.

Problem 3.27. Let

$$A_\mu = \left(-\frac{V}{c}, \mathbf{A}\right) ,$$

where V is the scalar potential, and $\mathbf{A}$ is the vector potential of electrodynamics. Then show that

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu .$$

Also, show that if A_μ transforms as a dual vector, then $F_{\mu\nu}$ transforms as a tensor.

3.4 The Energy-Momentum Tensor

The Energy-Momentum tensor for the electromagnetic field is given by the symmetric tensor [9]

$$T^{\mu\nu} = \frac{1}{\mu_0}\left[\eta_{\alpha\beta}F^{\mu\alpha}F^{\nu\beta} - \frac{1}{4}\eta^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right] .$$

Theorem 3.6. *In vacuum,*

$$\partial_\mu T^{\mu\nu} = 0 .$$

Proof. Since $\partial_\mu F^{\mu\nu} = 0$ in vacuum,

$$\partial_\mu T^{\mu}{}_{\nu} = \frac{F^{\alpha\beta}}{\mu_0}\left[\partial_\alpha F_{\nu\beta} - \frac{1}{2}\partial_\nu F_{\alpha\beta}\right] \tag{3.19}$$

Swapping the dummy variables α and β, from above we get that

$$\partial_\mu T^{\mu}{}_{\nu} = \frac{F^{\beta\alpha}}{\mu_0}\left[\partial_\beta F_{\nu\alpha} - \frac{1}{2}\partial_\nu F_{\beta\alpha}\right] .$$

Or

$$\partial_\mu T^{\mu}{}_{\nu} = \frac{-F^{\alpha\beta}}{\mu_0}\left[\partial_\beta F_{\nu\alpha} + \frac{1}{2}\partial_\nu F_{\alpha\beta}\right] . \tag{3.20}$$

Eqs. (3.19) + (3.20) gives

$$2\partial_\mu T^{\mu}{}_{\nu} = \frac{F^{\alpha\beta}}{\mu_0}\left[\partial_\alpha F_{\nu\beta} - \partial_\beta F_{\nu\alpha} - \partial_\nu F_{\alpha\beta}\right] .$$

Substituting

$$\partial_\alpha F_{\nu\beta} = -\partial_\nu F_{\beta\alpha} - \partial_\beta F_{\alpha\nu} ,$$

we get that

$$2\partial_\mu T^{\mu}{}_{\nu} = \frac{F^{\alpha\beta}}{\mu_0}\left[-\partial_\nu F_{\beta\alpha} - \partial_\beta F_{\alpha\nu} - \partial_\beta F_{\nu\alpha} - \partial_\nu F_{\alpha\beta}\right] = 0 .$$

□

Problem 3.28. For the Maxwell field, show that T^{00} describes the energy density of the electromagnetic field. I.e.,

$$T^{00} = \frac{1}{2}\epsilon_0 E^2 + \frac{1}{2}\frac{B^2}{\mu_0} .$$

[9] $T^{\mu\nu} = T^{\nu\mu}$.

Problem 3.29. For electromagnetic fields, show that,

$$T^{0i} = \frac{1}{\mu_0 c}(\mathbf{E} \times \mathbf{B})^i \ .$$

In electrodynamics, the Poynting vector is given by $\mathbf{S} = (\mathbf{E} \times \mathbf{B})/\mu_0$, and the momentum density vector $\mathbf{p}$ is $\mathbf{p} = \mathbf{S}/\mathbf{c^2}$.

In the relativistic formalism, the energy-momentum density 4-tuple p^μ is defined by

$$p^\mu = \frac{1}{c} T^{0\mu} \ . \tag{3.21}$$

Thus,

$$p^\mu = \left(\frac{\text{energy density}}{c}, \text{ momentum density vector} \right) .$$

It should be clear to you that p^μ as defined by eq.(3.21) is not a 4-vector. Since the energy-momentum tensor is divergence-free

$$\frac{\partial}{\partial t} p^\mu = \frac{\partial}{\partial ct} T^{0\mu} = -\nabla_i T^{i\mu} \ .$$

The rate of change of total energy-momentum, defined by

$$\frac{dP^\mu}{dt} = \frac{d}{dt} \int_{\mathbb{R}^3} p^\mu \ d^3x = \int_{\mathbb{R}^3} \frac{\partial}{\partial ct} T^{0\mu} \ d^3x \ .$$

Then, using the 3-dimensional divergence theorem, we get that

$$\frac{dP^\mu}{dt} = -\int_{\mathbb{R}^3} \nabla_i T^{i\mu} d^3x = -\lim_{r\to\infty} \int_{S^2(r)} T^{i\mu} \ \mathbf{n}_i \ da \ .$$

In the above equation, $S^2(r)$ is the 2-dimensional sphere of radius r, da is the area element on the sphere, and $\mathbf{n}$ is the outward normal to the sphere. In spherical coordinates,

$$da = r^2 \sin\theta \ d\theta d\varphi$$

and $\mathbf{n} = \hat{r}$. When fields are restricted to compact regions of space, we will generally have that

$$\lim_{r\to\infty} r^2 \ T^{i\mu} \to 0 \ ,$$

in which case

$$\frac{dP^\mu}{dt} = 0 \ .$$

Equivalently, the total energy and momentum described by P^μ is a conserved quantity. Although p^μ is not a 4-vector, it can be shown that P^μ is indeed a 4-vector under Lorentz transformation ([6]).

Problem 3.30. For non-vacuum fields $J^\mu \neq 0$. In this case, show that for the electromagnetic energy-momentum tensor

$$\partial_\mu T^{\mu\nu} = -F^\nu{}_\alpha \ J^\alpha \ .$$

Problem 3.31. The energy-momentum tensor for the Klein-Gordon field is given by

$$T^{\mu\nu}_{K.G} = \eta^{\mu\alpha}\eta^{\nu\beta}\partial_\alpha\phi \ \partial_\beta\phi - \frac{1}{2}\eta^{\mu\nu}\left[\eta^{\alpha\beta}\partial_\alpha\phi\partial_\beta\phi + \frac{m^2c^2}{\hbar^2}\phi^2\right] \ .$$

Show that

$$\partial_\mu T^{\mu\nu}_{K.G} = 0 \ .$$

Problem 3.32. For electromagnetic fields, show that,

$$\partial_\mu J^\mu = 0 \ .$$

Problem 3.33. Find the electric and magnetic field created by an electron moving along the x-axis with a speed v.

The Ingredients of the Energy-Momentum Tensor

Since T^{0i}/c is the momentum density, the dimensions of T^{0i} denoted by $[T^{0i}]$ is given by

$$[T^{0i}] = \frac{M}{LT^2} \, .$$

For reasons that will become evident when we consider the Einstein equation, we would like $T^{\mu\nu}$ to be a symmetric tensor field. With this in mind, let us try to understand the meaning and dimensions of T^{i0}. Since $T^{\mu\nu}$ is divergence free

$$\partial_\mu T^{\mu 0} = \frac{\partial T^{00}}{\partial ct} + \partial_i T^{i0} = 0 \, ,$$

or, integrating over a closed spatial cuboid V with boundary B, we get that

$$\frac{1}{c}\frac{d}{dt}\int_V T^{00} dxdydz = -\int_B T^{i0} \cdot \hat{n}_i \, da$$

$$= -\int_{B_x} T^{10} dydz - \int_{B_y} T^{20} dxdz - \int_{B_z} T^{30} dxdy \, .^{10}$$

Here $\hat{n}_i$, is, as usual, the outward normal, and the three area integrals are taken over the six faces of the cuboid. Then, ignoring the limits of integration for ease of presentation

$$\int T^{00} dxdydz = -\int cT^{10} dydzdt - \int cT^{20} dxdzdt - \int cT^{30} dxdydt \, .$$

Since the left-hand side above represents total energy, we have that

$$[cT^{i0}] = \frac{\text{energy flowing out in the i-th direction}}{\text{area} \cdot \text{time}} \, .$$

Therefore, as desired

$$[T^{i0}] = \frac{M}{LT^2} = [T^{0i}] \, .$$

Furthermore cT^{i0} can be interpreted as the flux of energy flux per unit area per unit time, which is exactly the Poynting flux $\mathbf{S}^i$. Thus, we see that $T^{i0} = cp^i$ as we expected. *It is a curious fact that the energy flux per unit area per unit time is exactly c times the momentum density that in part allows for the energy momentum tensor to be symmetric.*

The conservation equation for the i^{th}-component of momentum is given by (once again, for ease, we ignore the limits/regions of integration)

$$\frac{1}{c}\frac{d}{dt}\int T^{0i} dxdydz = -\int T^{1i} dydz - \int T^{2i} dxdz - \int T^{3i} dxdy$$

or,

$$\int \frac{T^{0i}}{c} dxdydz = -\int T^{1i} dydzdt - \int T^{2i} dxdzdt - \int T^{3i} dxdydt \, .$$

As before, since the left-hand side above represents total momentum, we can conclude that

$$[T^{ii}] = \frac{\text{i-momentum flowing out in i-direction}}{\text{area} \cdot \text{time}}$$

$$= \frac{\text{force in the i-direction}}{\text{area}} = \text{pressure in the i-direction} \equiv \mathrm{P}_i \, .$$

Additionally,

$$[T^{ji}] = \frac{\text{i-momentum flowing out in j-direction}}{\text{area} \cdot \text{time}} \equiv \sigma_{ij} \, .$$

It should immediately be clear that $[\sigma_{ij}] = [\sigma_{ji}]$. We are now able to visualize the energy-momentum tensor in the following way:

$$T^{\mu\nu} = \begin{pmatrix} \rho c^2 & cp_x & cp_y & cp_z \\ cp_x & \mathrm{P}_x & \sigma_{12} & \sigma_{13} \\ cp_y & \sigma_{21} & \mathrm{P}_y & \sigma_{23} \\ cp_z & \sigma_{31} & \sigma_{32} & \mathrm{P}_z \end{pmatrix} \, . \tag{3.22}$$

σ_{ij} describe the shear in the field.

[10] Here $\int_{B_x} T^{10} dydz$ would mean integral over $\int T^{10} dydz$ when $x = x_f$ minus the same integral when $x = x_i$.

Non-interacting Dust

Let us consider a non-interacting dust evolving in spacetime simply due to their initial velocities in such a way the at any point we can define a 4-vector u that describes the average fluid 4-velocity. Further, from point to point, we assume that u varies smoothly. Naturally u is a timelike future pointing vector field. Consider one such u at any point in this "fluid". In its local rest from $u = (c, \vec{0})$ at any instant of motion. Then the mass contained in a small cubical region of the dust is given by

$$\Delta m = \rho_0 \Delta x \Delta y \Delta z$$

where ρ_0 is the rest mass density. Now consider the same region in the lab frame. For ease let us consider the case where this cubical region is moving in the x-direction. There will be length contraction in the direction of motion. The cubical section will now have a cuboidal shape, and the mass can be expressed as

$$\Delta m = \rho \Delta \bar{x} \Delta y \Delta z$$

where $\Delta \bar{x} = \Delta x/\gamma$ due to length contraction. Comparing the two expressions for mass, we see that

$$\rho = \gamma \rho_0 \ .$$

It appears that mass density transforms as the 0th component of a 4-vector [11]. Thus, we define the mass density vector by

$$J^\mu = \frac{1}{c} \rho_0 u^\mu$$

so that $J^0 = \rho_0 \gamma = \rho$. The continuity equation given by

$$\partial_\mu J^\mu = 0 \tag{3.23}$$

is imposed on this current so that the total mass of the system is conserved. This can be seen as follows.

$$\frac{1}{c}\frac{d}{dt}\int J^0 \, dxdydz = -\int_{\mathbb{R}^3} \nabla \cdot \mathbf{J} \, dxdydz = -\lim_{r\to\infty} \int_{S^2(r)} \mathbf{J} \cdot \hat{n} \, da \ .$$

The last term above vanishes since J vanishes on the boundary, and so we get that the total mass

$$M = \int J^0 \, dxdydz$$

is conserved. Thus, the continuity equation implies conservation of mass and is a needed ingredient for fluid dynamics.

To obtain the energy-momentum tensor for a non-interacting dust, we simply note that in the rest frame $T^{00} = \rho_0 c^2$, or equivalently $T^{00} = \rho_0 u^0 u^0$. Thus, we define the symmetric energy-momentum tensor for dust by the expression

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu \ .$$

It is easily checked that this tensor is divergence-free, for

$$\partial_\mu T^{\mu\nu} = \partial_\mu(\rho_0 u^\mu u^\nu) = \partial_\mu(\rho_0 u^\mu) u^\nu + \rho_0 u^\mu \partial_\mu u^\nu$$

$$= c(\partial_\mu J^\mu) u^\nu + \rho_0 \frac{du^\nu}{d\tau} \ .$$

The first term in the right-hand side above goes to zero due to the continuity equation, and the second term goes to zero because there are no forces acting on the dust. Thus, the energy-momentum tensor for a non-interacting dust is divergence-free as required.

Equivalently, we could have required that for a non-interacting dust, the energy-momentum tensor is divergence-free to obtain the continuity equation given by eq.(3.23).

[11] While $u = (c, \vec{0})$ in the rest frame, in the lab frame $u = (c\gamma, \gamma\vec{u})$.

Perfect Fluids

A perfect fluid has a well defined fluid velocity as before. In addition, the fluid, unlike dust, can exert pressure on itself. In the local rest frame of the fluid, the momentum density vanishes, and we can take $\mathrm{P}_x = \mathrm{P}_y = \mathrm{P}_z = \mathrm{P}$ in eq.(3.22). Further, a perfect fluid does not suffer from shear terms described by σ_{ij}. Then, in a local rest frame,

$$T^{\mu\nu} = \begin{pmatrix} \rho_0 c^2 & 0 & 0 & 0 \\ 0 & \mathrm{P} & 0 & 0 \\ 0 & 0 & \mathrm{P} & 0 \\ 0 & 0 & 0 & \mathrm{P} \end{pmatrix} = \begin{pmatrix} \rho_0 c^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \mathrm{P} & 0 & 0 \\ 0 & 0 & \mathrm{P} & 0 \\ 0 & 0 & 0 & \mathrm{P} \end{pmatrix}$$

or,

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu + \mathrm{P}\left(\eta^{\mu\nu} + \frac{u^\mu u^\nu}{c^2}\right) ,$$

since in the rest frame $u = (c, \vec{0})$. But this is a tensorial equation and therefore we set the energy-momentum tensor for a perfect fluid as

$$T^{\mu\nu} = \left(\rho_0 + \frac{\mathrm{P}}{c^2}\right) u^\mu u^\nu + \mathrm{P}\, \eta^{\mu\nu} . \tag{3.24}$$

The equations of motion for a perfect fluid can be obtained by setting $\partial_\mu T^{\mu\nu} = 0$.

Problem 3.34. Verify that the energy-momentum tensors of the electromagnetic field, the Klein Gordon field and the perfect fluid is a tensor. In particular, for the perfect fluid how does ρ_0 and P transform? Hint: Don't overthink the last question. Read their definitions carefully.

Energy Conditions

For the inertial observer whose 4-velocity is $u = (c, 0, 0, 0)$, from problem 3.28, we see that $T^{00} \geq 0$ for the electromagnetic field. I.e., the energy density is non-negative. Clearly,

$$T^{00} = \frac{1}{c^2} T^{\mu\nu} u_\mu u_\nu = \frac{1}{c^2} T_{\mu\nu} u^\mu u^\nu \geq 0.$$

Therefore, for any inertial observer, described by a future pointing timelike vector W, we must have that $T_{\mu\nu} W^\mu W^\nu \geq 0$ for the electromagnetic field. We will indeed capture this feature as a requirement for any physically relevant energy-momentum tensor.

Definition 3.9. *The energy-momentum tensor satisfies the weak energy condition if $T_{\mu\nu} W^\mu W^\nu \geq 0$ for every future pointing timelike vector field W.*

Problem 3.35. Let W be any future-pointing timelike vector. Show that $T_{\mu\nu} W^\mu W^\nu \geq 0$ for the Klein-Gordon field. Pick an inertial frame where $W = (1, 0, 0, 0)$.

Problem 3.36. Let W be any future-pointing timelike vector. Show that $T_{\mu\nu} W^\mu W^\nu \geq 0$ for a non-interacting dust.

Energy-momentum tensors do not automatically satisfy the weak energy condition. These are, in general, extra conditions imposed on the constituents of the tensor.

Problem 3.37. Show that the energy-momentum tensor for a perfect fluid satisfies the weak energy condition if and only if

$$\rho_0 > 0 , \quad \textit{and} \quad \rho_0 c^2 + P > 0 .$$

Hint: Pick $u = (c, 0, 0, 0)$. Let W be any future-pointing timelike vector. Since W is future pointing timelike, note that

$$0 < \left(\frac{\vec{W}}{W^0}\right)^2 < 1,$$

however the above quantity can get arbitrarily close to 1.

Since $cp^\mu = T^{0\mu}$ describes energy and momentum density, we would like the flow or transport of these quantities to be below or at the speed of light. For an observer at rest in the inertial frame when $u^\mu = (c, 0, 0, 0)$,

$$T^{0\mu} = -\frac{1}{c} T^{\nu\mu} u_\nu \ .$$

As before we would like this feature to remain true for any observer.

Definition 3.10. *The energy-momentum tensor satisfies the dominant energy condition if for any future-pointing causal vector field W, $-T_\nu{}^\mu W^\nu$ is also a future-pointing causal vector field.*

4

Introduction to Gravity

Just like in special relativity, Einstein began his work in general relativity with a premise he elevated to the status of a principle. Stated simply, the Equivalence Principle (EP) takes the following form: *Any freely falling observer cannot detect local effects due to gravity.*

Before we consider the implications of the equivalence principle in detail, let us begin by considering some simple examples of how gravitation effects the Minkowski spacetime. The basic idea here is that all objects are affected by gravity in the same way, and hence by falling along with the objects you do not encounter any gravitational effects locally (very close to you). In Einstein's time this was a giant leap for humankind. However, watching a video of the goings on in the International Space Station, one is quickly convinced of its validity. But the usual caution applies, and the quest for EP-violation is ongoing. Any concrete case of EP-violation is admittedly a quick path to a Nobel prize!

4.1 Gravitational Effects in Minkowski Spacetime

4.1.1 Bending of Light

Consider a free-falling spaceship with an observer at one end of the ship. The freely falling observer emits a beam of light towards the other end of the craft from a height h from the "bottom" of the ship. The observer, as per the equivalence principle, cannot feel the effects of gravity and records that the light beam reached the other end of the ship exactly a height h from the bottom. However, to an inertial observer outside the ship, the beam of light appear to bend as the ship falls downward in order to reach the other end of the craft at the same level h from the bottom of the ship. This is an important and immediate consequence of EP. As we shall see in detail, this effect has indeed been verified experimentally. [1]

4.1.2 Gravitational Redshift

As a thought experiment, consider a free falling room [2] and two observers - one inside the room (FF for Free Falling) and one at rest outside (IO for Inertial Observer). In the beginning of the experiment, i.e., $t = 0$, the room is at rest with respect to IO, and then it falls freely under a uniform gravitational field $\mathbf{g}$. At this instant, a beam of light of frequency ν_{em} is emitted vertically to the top of the room which is a height h above. There is a light detector placed on the ceiling of the room. According to FF, the light detector records a light signal at time $\Delta t = h/c$ with frequency $\nu_{rec} = \nu_{em}$ as required by EP. Thus for FF [3]

$$\frac{\nu_{rec}}{\nu_{em}} - 1 = 0 \, .$$

However, accounting for special relativity, for IO, the corresponding equation takes the form

$$\frac{\nu_{rec}}{\nu_{em}} - 1 \approx \frac{u}{c} \, ,$$

[1] We will discuss the bending of light in the context of a Schwarzschild black hole.

[2] Einstein was famous for his thought experiments. In 1915, mankind was not yet prepared to conduct free-fall experiments in space.

[3] Recall that the relativistic doppler effect, for small β, is given by:

$$\frac{\nu}{\nu_0} - 1 = \sqrt{\frac{1+\beta}{1-\beta}} \approx \beta \, .$$

where $u = g\Delta t$ is the downward speed of the ceiling/detector at the time of signal detection. EP grants the information presented by FF as the preferred one when gravitation is included. Thus, to reconcile between the above two expressions for the frequency shift, IO "conjures" a gravitational frequency shift of

$$\frac{\nu_{rec}}{\nu_{em}} - 1 \approx \frac{u}{c} + \Delta_{g\nu} \ ,$$

where, clearly, we must have that $u/c + \Delta_{g\nu} = 0$. To explain the gravitational origin of $\Delta_{g,\nu}$, we will re-write it as following:

$$\Delta_{g\nu} = -\frac{u}{c} = -\frac{g\Delta t}{c} = -\frac{gh}{c^2} = -\frac{\Delta\Phi}{c^2} \ ,$$

where Φ is the Newtonian gravitational potential. More generally,

$$\frac{\nu_{rec}}{\nu_{em}} - 1 \approx \frac{u}{c} - \frac{\Delta\Phi}{c^2} \ .$$

Thus, when light signals are sent to and from stationary points in a gravitational field, we get that

$$\boxed{\frac{\nu_{rec} - \nu_{em}}{\nu_{em}} \approx -\frac{1}{c^2}[\Phi_{rec} - \Phi_{em}] \ .} \tag{4.1}$$

The above equation describes the gravitational correction to redshift at the lowest order. As light signals are emitted from a star, as it travels to towards Earth, there is a gravitational redshift

Note: $\nu_{em} > \nu_{rec}$ so this is a gravitational redshift.

4.1.3 Gravitational Time Dilation

For a special relativistic inertial observer, the frequency of light received will depend inversely on their proper time, i.e.,

$$\nu \sim \frac{1}{\Delta\tau} \ .$$

I.e., from eq.(4.1) we get that when a stationary observer with propertime τ_1 emits light in a gravitational field and is received by another stationary observer with propertime τ_2,

$$\frac{\nu_{rec} - \nu_{em}}{\nu_{em}} = \left(\frac{1}{\Delta\tau_2} - \frac{1}{\Delta\tau_1}\right)\Delta\tau_1 \approx -\frac{\Delta\Phi}{c^2} \ ,$$

or,

$$\boxed{\frac{\Delta\tau_1}{\Delta\tau_2} \approx 1 - \frac{\Delta\Phi}{c^2} \ .} \tag{4.2}$$

This is a case of gravitational time dilation. Physically, this signifies that two stationary observers at two locations have clocks that *do not* tick the same way in a gravitational field.

4.2 Generalized Coordinates

Gravitational time dilation suggests the spacetime coordinate dependence of the metric. A stationary observer at position 1 in a gravitation field notes that

$$-c^2\Delta\tau_1^2 = (\eta_{00})_1 \ c^2\Delta t^2 \ .$$

But another stationary observer in the same inertial frame at location 2 notes that

$$-c^2\Delta\tau_2^2 = (\eta_{00})_2 \ c^2\Delta t^2 \ .$$

Eq. (4.2) then implies that

$$(\eta_{00})_2 \neq (\eta_{00})_1; \ .$$

Thus, if we want to keep the metric as a tensorial object, we can no longer assume that

$$(\eta_{\mu\nu})_2 \neq (\eta_{\mu\nu})_1 \ .$$

Therefore, when gravitation is included, $\eta_{\mu\nu}$, and consequently, $\eta^{\mu\nu}$ are dependent on coordinates (ct, x, y, z). It is typical in this case to refer to the metric by $g_{\mu\nu}$. In this case, to include gravitation, we let

$$\eta_{\mu\nu} \to g_{\mu\nu}(ct, x, y, z) \ ,$$

such that when gravitation is "turned off", i.e., in the limit that the Universal Gravitational constant $G \to 0$, we must have that

$$g_{\mu\nu}(ct, x, y, z) \to \eta_{\mu\nu} \ .$$

Also, just like $\eta_{\mu\nu}$, we require that $g_{\mu\nu}$ is symmetric [4]. If $g_{\mu\nu}$ is position dependent, we should be prepared to handle more transformations than those given by the Poincaré group. Hence, in the following, we will describe the most general class of coordinates allowable to describe any region of spacetime. In order to understand the nature of a generalized coordinate system, consider first the simple case of the cylindrical coordinate system. The coordinate transformation between the Cartesian and cylindrical coordinate system

$$(x, y, z) \to (s, \phi, z)$$

is given by

$$\begin{aligned} x &= s\cos\varphi \\ y &= s\sin\varphi \ . \\ z &= z \end{aligned}$$

The matrix of partial derivatives in this case is given by

$$\begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial z} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -s\sin\varphi & 0 \\ \sin\varphi & s\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \ .$$

The determinant of the above matrix is called the Jacobian of the transformation, namely J, and in this case is given by s. The inverse transform between the cylindrical coordinate system and the Cartesian coordinate system is given by

$$\begin{aligned} s &= \sqrt{x^2 + y^2} \\ \tan\varphi &= y/x \ . \\ z &= z \end{aligned}$$

In this case,

$$\begin{pmatrix} \frac{\partial s}{\partial x} & \frac{\partial s}{\partial y} & \frac{\partial s}{\partial z} \\ \frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial y} & \frac{\partial \varphi}{\partial z} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} & \frac{\partial z}{\partial z} \end{pmatrix} = \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi/s & \cos\varphi/s & 0 \\ 0 & 0 & 1 \end{pmatrix} \ ,$$

and $J = \frac{1}{s}$. It is important to observe that the above two transformations are inverses of each other, i.e.,

$$\begin{pmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial \varphi} & \frac{\partial x}{\partial z} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial \varphi} & \frac{\partial y}{\partial z} \\ \frac{\partial z}{\partial s} & \frac{\partial z}{\partial \varphi} & \frac{\partial z}{\partial z} \end{pmatrix} \begin{pmatrix} \frac{\partial s}{\partial x} & \frac{\partial s}{\partial y} & \frac{\partial s}{\partial z} \\ \frac{\partial \varphi}{\partial x} & \frac{\partial \varphi}{\partial y} & \frac{\partial \varphi}{\partial z} \\ \frac{\partial z}{\partial x} & \frac{\partial z}{\partial y} & \frac{\partial z}{\partial z} \end{pmatrix}$$

$$= \begin{pmatrix} \cos\varphi & -s\sin\varphi & 0 \\ \sin\varphi & s\cos\varphi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi/s & \cos\varphi/s & 0 \\ 0 & 0 & 1 \end{pmatrix} = \delta^{\mu}_{\nu} \ .$$

The previous example illustrates all the essential properties and requirements of a valid coordinate system. By definition, $\mathbb{R}^3$ consists of all points given by the 3-tuple (x, y, z). Every point has a distinct value for the 3-tuple. Coordinate labels are like addresses. They should uniquely specify the location of a point. This is, in fact, not the case for the cylindrical coordinate system. Along the z, there is no unique specification of the coordinate φ. I.e., the 1-1 map between the Cartesian and the cylindrical coordinate system breaks down. As you can see, the points were cylindrical coordinate system is ill-defined defined is precisely given by the location $s = 0$. This is exactly when the Jacobian of the original transformation vanishes, or equivalently, where the Jacobian of the inverse transformation is ill-defined. This result is true in general and forms the benchmark test for coordinate systems. The general result is called the inverse function theorem, and we state it below without proof.

Suppose $(x^1, ..., x^n)$ is a coordinate system. [5] Then, another n-tuple $(y^1, \ldots, y^n)$ is a valid coordinate system if every point denoted by $(x^1, ..., x^n)$ is uniquely denoted by $(y^1, \ldots, y^n)$ and vice versa. I.e., we can write the coordinate functions in the form.

[4] There is no reason to consider the possibility that $g(A, B) \neq g(B, A)$.

[5] We are not just considering spacetimes here. The coordinate system could describe a sphere or a torus or a complicated surface, even in higher dimensions. Such objects are in general, referred to as a differential manifold.

$$y^1(x^1,\ldots,x^n),\ldots,y^n(x^1,\ldots,x^n)$$

and

$$x^1(y^1,\ldots,y^n),\ldots,x^n(y^1,\ldots,y^n)\ .$$

By this, we mean that one set of coordinate system is a differentiable function of the other set and vice versa. The *inverse function theorem* [7] states that this happens precisely when the matrix given by

$$\begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} & \cdots & \frac{\partial y^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y^m}{\partial x^1} & \frac{\partial y^m}{\partial x^2} & \cdots & \frac{\partial y^m}{\partial x^n} \end{pmatrix}$$

is well defined, and further, the Jacobian of the transformation given by the determinant of the above matrix is not equal to zero. Moreover, when this is the case

$$\begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} & \cdots & \frac{\partial y^2}{\partial x^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y^m}{\partial x^1} & \frac{\partial y^m}{\partial x^2} & \cdots & \frac{\partial y^m}{\partial x^n} \end{pmatrix} \begin{pmatrix} \frac{\partial x^1}{\partial y^1} & \frac{\partial x^1}{\partial y^2} & \cdots & \frac{\partial x^1}{\partial y^n} \\ \frac{\partial x^2}{\partial y^1} & \frac{\partial x^2}{\partial y^2} & \cdots & \frac{\partial x^2}{\partial y^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x^m}{\partial y^1} & \frac{\partial x^m}{\partial y^2} & \cdots & \frac{\partial x^m}{\partial y^n} \end{pmatrix} = \delta^\mu_\nu\ .$$

The above equation can be written using the Einstein summation convention as

$$\frac{\partial y^\mu}{\partial x^\alpha}\frac{\partial x^\alpha}{\partial y^\nu} = \frac{\partial x^\mu}{\partial y^\alpha}\frac{\partial y^\alpha}{\partial x^\nu} = \delta^\mu_\nu\ . \tag{4.3}$$

Problem 4.1. Write the transformation and the inverse transformation between the Cartesian coordinate system and the spherical coordinate system. Locate the points where the Jacobian of the transformation is either 0 or undefined. Argue as to why these locations are not describable by the spherical coordinate system. Further, verify eq.(4.3) in this case.

4.3 Tensors in Generalized Coordinates

The transformation rules for covariant and contravariant vectors in special relativity are given by problem 3.16 and eq.(3.14).

$$\bar{A}^\mu = \Lambda^\mu{}_\nu A^\nu = \frac{\partial \bar{x}^\mu}{\partial x^\nu}A^\nu$$

and

$$\bar{A}_\mu = \Lambda_\mu{}^\nu A_\nu = \frac{\partial x^\nu}{\partial \bar{x}^\mu}A_\nu\ .$$

If we have vector fields instead, the rules generalize to

$$\bar{A}^\mu(\bar{x}) = \frac{\partial \bar{x}^\mu}{\partial x^\nu}A^\nu(x)\ ,$$

and

$$\bar{A}_\mu(\bar{x}) = \frac{\partial x^\nu}{\partial \bar{x}^\mu}A_\nu(x)\ .$$

Here x and $\bar{x}$ denote the same spacetime point in two different frames. The notation $\bar{A}_\mu(\bar{x})$ is redundant and can be simplified to $A_\mu(\bar{x})$. For our generalized coordinates, we use x^μ and y^μ instead of x^μ and $\bar{x}^\mu$. Therefore, as an obvious generalization [6] we take the transformation rules for tangent and dual vector fields to be

$$A^\mu(y) = \frac{\partial y^\mu}{\partial x^\nu}A^\nu(x)\ ,$$

and

$$A_\mu(y) = \frac{\partial x^\nu}{\partial y^\mu}A_\nu(x)\ .$$

[6] Vectors can only transform by the action of a matrix, and we only have a choice of the transformation matrix and its inverse.

As in the special relativistic case, y and x will denote the same physical event in the two coordinate systems. Also note the usual chain rules:

$$\frac{\partial}{\partial y^\mu} = \frac{\partial x^\nu}{\partial y^\mu}\frac{\partial}{\partial x^\nu} \ , \tag{4.4}$$

and [7]

$$dy^\mu = \frac{\partial y^\mu}{\partial x^\nu} dx^\nu \ . \tag{4.5}$$

The above rules imply the following transformation rule for tensors:

$$T^{\mu_1,\ldots,\mu_r}{}_{\nu_1,\ldots,\nu_s}(y) = \frac{\partial y^{\mu_1}}{\partial x^{\alpha_1}}\ldots\frac{\partial y^{\mu_r}}{\partial x^{\alpha_r}}\frac{\partial x^{\beta_1}}{\partial y^{\nu_1}}\ldots\frac{\partial x^{\beta_s}}{\partial y^{\nu_s}} T^{\alpha_1\ldots\alpha_r}{}_{\beta_1,\ldots,\beta_s}(x) \ . \tag{4.6}$$

Contraction of tensors continue to work the same way, since all we really needed was the property given by eq.(4.3) (see theorem 3.3 along with the discussion below the theorem, and problem 3.16).

In special relativity

$$(\Delta s)^2 = \eta_{\mu\nu}\Delta x^\mu \Delta x^\nu$$

is an invariant quantity. In general relativity, we can use the metric tensor $g_{\mu\nu}$ to create an invariant quantity as well.

Definition 4.1.

$$(ds)^2 \equiv g_{\mu\nu} dx^\mu dx^\nu \ . \tag{4.7}$$

Theorem 4.1. *$(ds)^2$ is independent of the coordinate system. I.e., it's an invariant quantity.*

Proof. Since $g_{\mu\nu}$ is a tensor, we get that

$$ds^2 = g_{\mu\nu}(y)dy^\mu dy^\nu$$

$$= \frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial x^\beta}{\partial y^\nu} g_{\alpha\beta}\ \frac{\partial y^\mu}{\partial x^\lambda}dx^\lambda\ \frac{\partial y^\nu}{\partial x^\sigma}dx^\sigma = \frac{\partial x^\alpha}{\partial y^\mu}\frac{\partial y^\mu}{\partial x^\lambda}\frac{\partial x^\beta}{\partial y^\nu}\frac{\partial y^\nu}{\partial x^\sigma}dx^\lambda dx^\sigma g_{\alpha\beta}(x)$$

$$= \delta^\alpha{}_\lambda \delta^\beta{}_\sigma g_{\alpha\beta}(x)dx^\lambda dx^\sigma = g_{\lambda\sigma}(x)dx^\lambda dx^\sigma \ .$$

□

[7] What the heck is dy^μ? Let

$$\alpha(s) = (x^1(s), ..., x^n(s)) \ \ \forall s \in (a, b)$$

be a curve given in the x^μ coordinate system. Then in the y^μ coordinate system, α can be written as

$$\alpha(s) = (y^1(s), ..., y^n(s)) \ .$$

The velocity of this curve in the x^i coordinate system is given by

$$\dot{\alpha} = \left(\frac{dx^1}{ds}, ..., \frac{dx^n}{ds}\right) \ ,$$

while the velocity of this curve in the y^i coordinate system is given by

$$\dot{\alpha} = \left(\frac{dy^1}{ds}, ..., \frac{dy^n}{ds}\right) \ .$$

The components of the velocity in the two coordinate frames are related by the expression

$$\frac{dy^i}{ds} = \frac{\partial y^i}{\partial x^1}\frac{dx^1}{ds} + ... + \frac{\partial y^i}{\partial x^n}\frac{dx^n}{ds} = \frac{\partial y^i}{\partial x^i}\frac{dx^i}{ds} \ .$$

Since s is an arbitrary parameter, and with the understanding that we can take derivatives along any arbitrary curve, we hide its curve independence and write

$$dy^i = \frac{\partial y^i}{\partial x^j}dx^j$$

for short.

The Generality of Tensors

Although all discussions have focused on spacetime, the formalism we have developed is more general. It applies to a large class of geometric structures called Differentiable Manifolds. For example, the Klein bottle, the surface of a sphere, and that of a torus are 2-dimensional manifolds. The Minkowski spacetime is a 4-dimensional manifold. A more complicated 3-dimensional manifold is the rotation group $SO(3)$. We will not discuss how $SO(3)$ forms a differentiable manifold. Such discussions belong in a course on Lie Groups.

If we can further define a metric tensor $g_{\mu\nu}$ on a manifold, given tangent vectors X and Y, we can compute $g(X,Y)$ in the usual manner by the recipe

$$g(X,Y) \equiv g_{\mu\nu} X^\mu Y^\nu \ .$$

If the metric tensor is positive definite, i.e., $g(X,X) \geq 0$ for every tangent vector X, we say the manifold is **Riemannian**. In relativity because of the presence of a timelike vector, i.e., vectors such that $g(X,X) < 0$, we say the manifold is **Semi-Riemannian**. In particular, in orthonormal bases, since we have only one timelike vector it is a **Lorentzian Manifold**. [8]

The simplest 3-dimensional manifold is $\mathbb{R}^3$. Consider for example $\mathbb{R}^3$, where the Euclidean metric in the natural coordinates (x,y,z) is given by

$$ds^2|_E = (dx)^2 + (dy)^2 + (dz)^2 \ .^{9}$$

Comparing the above expresion with eq.(4.7) we see that, by definition, $g_{xx} = g_{yy} = g_{zz} = 1$ while $g_{xy} = g_{xz} = g_{yz} = 0$.

Problem 4.2. In the spherical coordinate system (r,θ,φ) of $\mathbb{R}^3$, where

$$x = r\cos\varphi\ \sin\theta\ , \quad y = r\sin\varphi\ \sin\theta\ , \quad \text{and} \quad z = r\cos\theta\ ,$$

show that

$$ds^2|_E = dr^2 + r^2(d\theta^2 + \sin^2\theta\ d\varphi^2)\ .$$

Here $g_{rr} = 1$, $g_{\theta\theta} = r^2$, and $g_{\varphi\varphi} = r^2\sin^2\theta$, while, $g_{r\theta} = g_{r\varphi} = g_{\theta\varphi} = 0$.

Problem 4.3. In spherical coordinates, let $W = (r^2, 3\sin^2\theta, -1)$ and $X = (r^2, 0, 2)$ be vector fields. Find $g(W,W)$, $g(W,X)$, $g(X,X)$, and $g(X-Y,2X)$ at the point (1,0,0).

From the above problem, we see that the induced metric on a sphere of radius r is given by

$$ds^2|_{sphere} = r^2(d\theta^2 + \sin^2\theta\ d\varphi^2)$$

since $dr = 0$ for a sphere. The presence of $\sin^2\theta$ in the metric for a sphere may be puzzling! After all, there aren't any special points on a sphere. The reason for this is our choice of coordinate system, where there is a preferred pole (z axis that defines the polar angle θ).

Problem 4.4. Suppose we rotate our Cartesian coordinate system by

$$\bar{x}^i = R^i{}_j x^j \ ,$$

where $R^i{}_j$ is any element of the rotation group. Show that

$$ds^2|_E = d\bar{x}^2 + d\bar{y}^2 + d\bar{z}^2 \ ,$$

where $(\bar{x},\bar{y},\bar{z})$ is given by the rotation matrix above.

Now consider the transformation

$$\bar{x} = \bar{r}\cos\bar{\varphi}\ \sin\bar{\theta}\ , \quad \bar{y} = \bar{r}\sin\bar{\varphi}\ \sin\bar{\theta}\ , \quad \text{and} \quad \bar{z} = \bar{r}\cos\bar{\theta}\ .$$

Then, as before

$$ds^2|_E = d\bar{\theta}^2 + \sin^2\bar{\theta}\ d\bar{\varphi}^2\ .$$

Therefore, by picking a rotated coordinate system, we have a new pole ($\bar{z}$ axis that defines a new $\bar{\theta}$ axis) and the form metric is unchanged. Hence the metric of the sphere is said to be **spherically symmetric**.

[8] These general metrics are required to be symmetric and non-degenerate, i.e., $\det g \neq 0$.
[9] When writing quantities like $(ds)^2$, we often ignore the parentheses and write ds^2 instead.

4.4 Free Particle in a Gravitational Field

In a gravitational field at any one fixed point p, the EP guarantees the existence of a local inertial coordinate system ξ^μ such that

$$g_{\mu\nu}(\xi)|_p = \eta_{\mu\nu} \quad \text{and} \quad \frac{\partial}{\partial\xi^\alpha} g_{\mu\nu}(\xi)\Big|_p = 0 \,. \tag{4.8}$$

Since we expect physics to have governing equations described by the second derivatives (an old habit instilled by Newton), the above equation can be considered as initial conditions implied by the EP. Further, the equation of motion of an object in free fall (under the influence of gravitation alone) in the ξ^μ coordinate system is given by

$$\frac{d^2\xi^\alpha}{d\tau^2}\Big|_p = 0 \,.$$

Here τ is the proper time of the falling object. The observer using the coordinate system $\{\xi^\alpha\}$ is said to be locally inertial at p . Such observers do not see effects of gravitation at p. Inertial observers exists because the rate at which the observer accelerates matches the rate at which any particle accelerates at that point. This happens because, irrespective of their masses, objects fall exactly the same way in a gravitational field. The equivalence of the inertial mass of an object to the gravitational mass of the object leads to the independence of the mass of the particle in its equations of motion.

Let x^μ be any other coordinate system at p. Clearly, the product rule implies that

$$\frac{\partial}{\partial x^\alpha}\, g_{\mu\nu}(\xi)\Big|_p = \frac{\partial\xi^\beta}{\partial x^\alpha}\, \frac{\partial}{\partial\xi^\beta}\, \eta_{\mu\nu}(\xi)\Big|_p = 0 \,. \tag{4.9}$$

Also, in the x^μ coordinate system

$$g_{\mu\nu}(x)\Big|_p = \frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial\xi^\beta}{\partial x^\nu}\, g_{\alpha\beta}(\xi)\Big|_p \equiv \frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial\xi^\beta}{\partial x^\nu}\Big|_p\, \eta_{\alpha\beta} \,. \tag{4.10}$$

It is important to note a subtle point. Eq.(4.9) does not mean that

$$\frac{\partial}{\partial x^\alpha}\, g_{\mu\nu}(x)\Big|_p = 0$$

since $g_{\mu\nu}(x)\Big|_p$ is given by the right hand side of eq.(4.10). Eq.(4.10) gives the immediate following corollary:

Corollary 4.1. $\det g_{\mu\nu} \neq 0$.

Definition 4.2. *Since* $\det g_{\mu\nu} \neq 0$ *we get that* $g^{-1}_{\mu\nu}$ *exists. As usual, denote* $g^{-1}_{\mu\nu}$ *by* $g^{\mu\nu}$*. Then*

$$g_{\mu\nu}g^{\nu\alpha} = \delta^\alpha_\mu \,.$$

Then we can raise and lower tensors just like in special relativity by replacing $\eta_{\mu\nu}$ with $g_{\mu\nu}$. For example

$$A_\mu = g_{\mu\nu}A^\nu \,,$$

$$F^\mu{}_\nu = F^{\mu\alpha}g_{\alpha\nu} \,,$$

etc.

Theorem 4.2. *The tangent space at any point p in spacetime has the same structure as in special relativity, i.e., we can construct a basis with 4 orthonormal vectors such that one vector is timelike, and the remaining 3 is spacelike.*

Proof. In ξ at p, there exist four 4-vectors $e_{(\mu)}$ such that

$$\eta_{\alpha\beta}\, e^\alpha_{(\mu)} e^\beta_{(\nu)} = \eta_{\mu\nu} \,.$$

This is literally what we mean by the first equation in eq.(4.8). For example, we can set

$$e_{(0)} = (1,0,0,0) \,, \quad e_{(1)} = (0,1,0,0) \,,$$

etc. Then, at p, in an arbitrary coordinate system x^μ, define

$$E^\mu_{(\alpha)} = \frac{\partial x^\mu}{\partial \xi^\nu} e^\nu_{(\alpha)} \, .$$

Clearly, then

$$g_{\alpha\beta} \, E^\alpha_{(\mu)} E^\beta_{(\nu)} = \eta_{\mu\nu} \, .$$

□

Definition 4.3. *An arbitrary spacetime with an assigned Lorentz metric $g_{\mu\nu}$ is denoted by $(\mathcal{M}, g)$* [10]*. We will also require $\mathcal{M}$ to be time oriented which we may choose to specify as $(\mathcal{M}, g, \mathcal{T})$.* [11]

4.4.1 The Geodesic Equation of Motion

As we have seen, in a local inertial frame centered at the point p, the equation of motion for an object in free fall is given by

$$\left. \frac{d^2 \xi^\alpha}{d\tau^2} \right|_p = 0 \, .$$

In the following calculation, we will rewrite the above equation in an arbitrary coordinate system x^μ with no reference to ξ^μ. This is important since while EP guarantees the existence of ξ^μ it does not offer a method to construct it (but construct it we will, see corollary 4.3). Hence, without the rewrite, the above equation does not yield any meaningful results.

Analogous to the special relativistic case (definition 3.1), in general relativity, propertime is given by the expression

$$c\tau = \int \sqrt{-g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}} ds \, , \tag{4.11}$$

where

$$U^\mu = \frac{dx^\mu}{d\tau}$$

is the proper velocity of the particle. Then, just as before, we will have that

$$g(U, U) = g_{\mu\nu} \, U^\mu U^\nu = -c^2 \, .$$

Theorem 4.3. *For a free particle, EP implies the geodesic equation of motion given by*

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta} \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} = 0 \, , \tag{4.12}$$

where

$$\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu} [\partial_\alpha g_{\beta\nu} + \partial_\beta g_{\alpha\nu} - \partial_\nu g_{\alpha\beta}] \, . \tag{4.13}$$

Proof. Let p be any point in the worldline of the particle. Then at p,

$$\left. \frac{d^2 \xi^\alpha}{d\tau^2} \right|_p = 0$$

implies that

$$\frac{d}{d\tau} \left(\frac{\partial \xi^\alpha}{\partial x^\mu} \frac{dx^\mu}{d\tau} \right) = 0 \, .$$

Here, $\{x^\mu\}$ is any coordinate system at p. Therefore,

$$\frac{\partial \xi^\alpha}{\partial x^\mu} \frac{d^2 x^\mu}{d\tau^2} + \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0 \, .$$

Since

[10] Later on we will add a few more requirements for $g_{\mu\nu}$, chief of them being that $g_{\mu\nu}$ has to satisfy the Einstein equation.

[11] Considering definition 2.15 and problem 2.17, this implies an assignment of a continuous global timelike or lightlike vector field as future pointing.

$$\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial x^\lambda}{\partial \xi^\alpha} = \delta^\lambda_\mu,$$

we have that

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\lambda\nu}\frac{dx^\lambda}{d\tau}\frac{dx^\nu}{d\tau} = 0$$

where

$$\Gamma^\mu_{\lambda\nu} \equiv \frac{\partial x^\mu}{\partial \xi^\alpha}\frac{\partial^2 \xi^\alpha}{\partial x^\lambda \partial x^\nu} \ . \tag{4.14}$$

Note that $\Gamma^\lambda_{\mu\nu}$ as defined above is symmetric in the μ and ν indices. All of the above equations are only valid at p . All that remains to be shown is that the above expression for $\Gamma^\lambda_{\mu\nu}$ agrees with eq.(4.13). To see this, note that at p, the metric tensor in the x^μ coordinate system is given by eq.(4.10), and hence has derivatives

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \frac{\partial^2 \xi^\alpha}{\partial x^\lambda \partial x^\mu}\frac{\partial \xi^\beta}{\partial x^\nu}\,\eta_{\alpha\beta} + \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial^2 \xi^\beta}{\partial x^\lambda \partial x^\nu}\,\eta_{\alpha\beta} \ .$$

From eq.(4.14), the above equation becomes

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\rho_{\lambda\mu}\frac{\partial \xi^\alpha}{\partial x^\rho}\frac{\partial \xi^\beta}{\partial x^\nu}\,\eta_{\alpha\beta} + \Gamma^\rho_{\lambda\nu}\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial \xi^\beta}{\partial x^\rho}\,\eta_{\alpha\beta} \ ,$$

or

$$\left.\frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right|_p = \Gamma^\rho_{\lambda\mu}\ g_{\rho\nu} + \Gamma^\rho_{\lambda\nu}\ g_{\rho\mu}\Big|_p \ . \tag{4.15}$$

The usual procedure of adding and subtracting clever combinations of metric derivatives give (from the above equation)

$$\left.\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu}\right|_p = g_{\kappa\nu}\Gamma^\kappa_{\lambda\mu} + g_{\kappa\mu}\Gamma^\kappa_{\lambda\nu}$$
$$+g_{\kappa\nu}\Gamma^\kappa_{\mu\lambda} + g_{\kappa\lambda}\Gamma^\kappa_{\mu\nu} - g_{\kappa\lambda}\Gamma^\kappa_{\nu\mu} - g_{\kappa\mu}\Gamma^\kappa_{\nu\lambda} \ .$$

After a rearrangement of indices, multiplying the above equation with $g^{\nu\sigma}$ gives the needed result . □

Corollary 4.2. *If we know the metric $g_{\mu\nu}$ and the Christoffel symbols $\Gamma^\kappa_{\lambda\mu}$ at one point on a semi-Riemannian manifold, we can obtain the coordinate derivatives of the metric at that point.*

Proof. This is obvious from eq.(4.15). At any point in the manifold, in any coordinate system

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \Gamma^\rho_{\lambda\mu}\ g_{\rho\nu} + \Gamma^\rho_{\lambda\nu}\ g_{\rho\mu} \ .$$

□

Corollary 4.3. *(Working backwards,) it is easy to see that given $g_{\mu\nu}$ and $\Gamma^\kappa_{\lambda\mu}$ at one point p^α on the manifold, we can construct $\{\xi^\alpha\}$ that is inertial at p.*

Proof. This is guaranteed by the EP, and we will show this explicitly. It is sufficient to show that we can pick $\{\xi^\alpha\}$ such that at p, they satisfy eq.(4.14). Pick

$$\xi^\alpha(x) = \xi^\alpha(p) \ + \frac{\partial \xi^\alpha}{\partial x^\mu}(p)\ (x^\mu - p^\mu) + \frac{1}{2}\frac{\partial \xi^\alpha}{\partial x^\lambda}(p)\ \Gamma^\lambda_{\mu\nu}(p)\ (x^\mu - p^\mu)(x^\nu - p^\nu) + ...$$

The third order terms will not effect eq.(4.14) at p. Here, $\xi^\alpha(p)$ are arbitrary constants, and

$$\frac{\partial \xi^\beta}{\partial x^\nu}(p)$$

are to be picked so that

$$\frac{\partial \xi^\alpha}{\partial x^\mu}(p)\ \ \frac{\partial \xi^\beta}{\partial x^\nu}(p)\ \ \eta_{\alpha\beta} = g_{\mu\nu}(p).$$

□

Remark 4.1. In the above corollary, note that if we pick $\frac{\partial \xi^\beta}{\partial x^\nu}(p)$ so that

$$\frac{\partial \xi^\alpha}{\partial x^\mu}(p) \ \frac{\partial \xi^\beta}{\partial x^\nu}(p) \ \eta_{\alpha\beta} = g_{\mu\nu}(p) \ ,$$

then the matrices

$$\Lambda^\nu{}_\beta \ \frac{\partial \xi^\beta}{\partial x^\nu}(p)$$

satisfies the above equation as well, provided $\Lambda^\nu{}_\beta$ is an arbitrary Lorentz transformation. Therefore, if $\{\xi^\alpha\}$ is inertial at p^α, then so is $\{\Lambda^\nu{}_\alpha \xi^\alpha + c^\nu\}$, where c^ν are arbitrary constants. Clearly, this is not surprising.

From the above equation of motion, it is clear that the motion of an object in a gravitational field is determined by the metric $g_{\mu\nu}$. But what determines $g_{\mu\nu}$? This is a long story which leads to the "other" famous Einstein equation, and is the topic of study in the chapters to follow.

4.4.2 Photon trajectories in General Relativity

From subsection 3.1.1, we get that in special relativity, for a photon

$$\frac{d^2 x^\mu}{ds^2} = 0 \ ,$$

and

$$\eta(P,P) = \eta_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0 \ .$$

Here, since $g = \eta$, we also have that $\Gamma^\mu_{\alpha\beta} = 0$. Then in General Relativity, we can take a photon trajectory to be given by a null geodesic satisfying

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{\alpha\beta} \frac{dx^\alpha}{ds} \frac{dx^\beta}{ds} = 0 \ ,$$

and

$$g(P,P) = g_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} = 0 \ .$$

It is important to note that propertime is not defined for a null geodesic. So here, s is just any parameter that describes the null geodesic $x^\mu(s)$. In general, be it propertime or otherwise, a parameter for which the curve satisfies the geodesic equation is referred to as an **affine** parameter.

5

Schwarzschild Solution

5.1 A Heuristic Argument for the Solution

In this section, we will learn to work with a known solution to the Einstein equation. The Schwarzschild metric describes a vacuum, spherically symmetric, and static spacetime. For now, although we motivate the Schwarzschild metric, we are not in a position to derive its precise form. Nonetheless, we will still learn to extract physical meaning from this geometry. Once we formulate the Einstein equation, the Schwarzschild solution is easily derived (as we will).

Spherically symmetric, static spacetimes take the form,

$$ds^2 = g_{tt}\, d(ct)^2 + g_{rr}\, dr^2 + r^2 d\Omega^2 \ ,^{1}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2$ is the metric of the unit sphere, and g_{tt} and g_{rr} depend only on r. From eq.(4.2), we have that, for locations labeled by 1 and 2 in a weak gravitational field, the proper times are related by

$$d\tau_1 \approx [1 + \Phi_1 - \Phi_2]\, d\tau_2$$

Let the location of 2 be very far away, i.e., let $r_2 \to \infty$. Then $\Phi_2 = 0$ and $d\tau_2 = dt$, and let location 1 be another arbitrary point (i.e., we drop the subscript). Then,

$$d\tau \approx \left(1 + \frac{\Phi}{c^2}\right) dt \ ,$$

or, for a stationary object (fixed r, θ, φ),

$$ds^2 = -c^2 d\tau^2 \approx -\left(1 + \frac{2\Phi}{c^2}\right) d(ct)^2 \ .$$

In the general case, we get that

$$ds^2 \approx -\left(1 + \frac{2\Phi}{c^2}\right) d(ct)^2 + g_{rr} dr^2 + r^2 d\Omega^2$$

Outside a spherically symmetric compact object, Newtonian gravity gives that

$$\Phi = \frac{-GM}{r} \ .$$

The Schwarzschild metric describing a vacuum, spherically symmetric, static spacetime is given by

$$ds^2_{Sch} = -\left(1 - \frac{2GM}{c^2 r}\right) d(ct)^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 + r^2 d\Omega^2 \ . \tag{5.1}$$

Although we have motivated the expression for g_{00}, it is a mere coincidence that we guessed it correctly. Further, g_{rr} is stated without any explanation. I cannot offer any additional justification at this point. Suffice it to say that we will derive this exact expression in the next chapter. The Schwarzschild metric suffers from two singularities (points in spacetime where the metric is not defined). They are located by

[1] If we are being pedantic since $x^0 = ct$, we should write g_{tt} as g_{ctct}.

$$r = 0 \text{ and } r = \frac{2GM}{c^2} .$$

We will analyze the implications of these singularities shortly. Set

$$r^* = \frac{2GM}{c^2} .$$

$r > r^*$ is referred to as the Schwarzschild exterior spacetime, and $r < r^*$ is the interior spacetime. For Sun,

$$r^*_\odot = \frac{2GM_\odot}{c^2} \approx 3\text{km}, \text{ while } \frac{r^*_\odot}{R_\odot} \approx 10^{-6} .$$

Clearly, $T^\mu = (1, 0, 0, 0)$ is timelike in the exterior spacetime, spacelike in the interior spacetime, and lightlike at $r = r^*$, while $R^\mu = (0, 1, 0, 0)$ is spacelike in the exterior spacetime, timelike in the interior spacetim,e and undefined when $r = r^*$.

5.1.1 Timelike Geodesics

Setting $ds^2 = -c^2 d\tau^2$, the invariant interval becomes

$$-c^2 = -\left(1 - \frac{r^*}{r}\right) c^2 \dot{t}^2 + \left(1 - \frac{r^*}{r}\right)^{-1} \dot{r}^2 + +r^2\dot{\theta}^2 + r^2 \sin^2\theta\dot{\varphi}^2 , \tag{5.2}$$

and the Lagrangian for the geodesic equation of motion for a free particle in Schwarzschild spacetime is given by

$$L = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = -\left(1 - \frac{r^*}{r}\right) c^2 \dot{t}^2 + \left(1 - \frac{r^*}{r}\right)^{-1} \dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2\theta\dot{\varphi}^2 .$$

Here $\cdot$ denotes derivative with respect to proper time. The Lagrangian is minimized when

$$\frac{d}{d\tau}\left(\frac{\partial L}{\partial \dot{x}^\mu}\right) - \frac{\partial L}{\partial x^\mu} = 0 .$$

For the Schwarzschild metric

$$\frac{\partial L}{\partial t} = 0 .$$

Therefore,

$$\frac{d}{d\tau}\left(\frac{\partial L}{\partial c\dot{t}}\right) = 0 \Rightarrow \frac{\partial L}{\partial c\dot{t}} = \text{const} .$$

I.e.,

$$\frac{1}{2}\frac{\partial L}{\partial c\dot{t}} = g_{tt} c\dot{t} = \text{const} = -\kappa$$

is a constant of motion. In the limit as r approaches infinity, we have

$$-g_{tt} mc^2 \dot{t} \Rightarrow mc^2\gamma = \epsilon$$

is the energy of the particle, or

$$\kappa = \frac{\epsilon}{mc} .$$

and similarly,

$$\frac{\partial L}{\partial \varphi} = 0 ,$$

gives rise to the conservation of angular momentum l given by

$$l = m g_{\varphi\varphi}\dot{\varphi} = m r^2 sin^2\theta\dot{\varphi} . \tag{5.3}$$

Since geodesic motion is always planar in Schwarzschild spacetime (just as in planetary motion in classical mechanics, while this is clear on physical grounds, see problem 5.2 below) , we may restrict the orbit to the equatorial plane given by $\theta = \pi/2$. Then, eq.(5.2) reduces to

$$\frac{1}{2} m\dot{r}^2 + \left(1 - \frac{r^*}{r}\right)\frac{\ell^2}{2mr^2} - \frac{GMm}{r} = \frac{m}{2}(\kappa^2 - c^2) . \tag{5.4}$$

Problem 5.1. Verify the above equation.

Problem 5.2. Find the geodesic equation of motion for the θ coordinate. Show that a constant value of θ satisfies this equation.

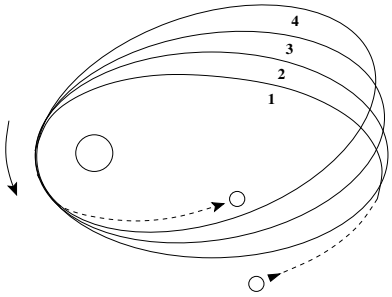


Fig. 5.1. Successive orbits of Mercury about the sun are shown. The perihelion advance is greatly exaggerated for visual clarity.

5.1.2 The Perihelion Advance of Mercury

Eq.(5.4) is the governing equation for timelike geodesics in Schwarzschild spacetime. The solutions to this equation will give corrections to the classical planetary orbits described by the Newtonian theory. For most of the outer planets, where the effects of general relativity can be ignored, Newtonian orbits are sufficient. However, the orbit of Mercury, being closest to the Sun, will deviate from the Newtonian expectation. One such effect is the advancement of the perihelion (the closest point in the orbit) of Mercury. For the Sun

$$\frac{r^*_\odot}{R_\odot} \approx 10^{-6} \, .$$

Then, for planetary motion, $r > r^*$, and so we won't have to concern ourselves with the metric singularity at r^*. Also, if needed, clearly, r^*/r will serve as a good parameter for a Taylor expansion of difficult expressions. From eq.(5.3) we get that

$$\dot{r} = \frac{dr}{d\tau} = \frac{dr}{d\varphi}\frac{d\varphi}{d\tau} = \frac{dr}{d\varphi}\frac{l}{mr^2} \, . \tag{5.5}$$

Problem 5.3. Setting

$$A = \left(1 - \frac{r^*}{r}\right)^{-1} \quad , \quad J = l/m\kappa \, , \quad \text{and} \quad E = c^2/\kappa^2 \, ,$$

show that equation eq.(5.4) reduces to

$$d\varphi = \int \frac{\sqrt{A}}{r^2}\left(\frac{A}{J^2} - \frac{E}{J^2} - \frac{1}{r^2}\right)^{-1/2} dr \, .$$

We are looking for bounded motion. In this case, the turning points $r_\mp$ are given by

$$\frac{dr}{d\varphi} = 0 \implies \frac{1}{r_\mp^2} - \frac{A(r_\mp)}{J^2} = -\frac{E}{J^2} \, .$$

This allows us to write E and J in terms of $r_\mp$.

$$\frac{1}{r_+^2} - \frac{A(r_+)}{J^2} = -\frac{E}{J^2} = \frac{1}{r_-^2} - \frac{A(r_-)}{J^2}$$

or,

$$J^2 = \Big[A(r_+) - A(r_-)\Big]\left(\frac{1}{r_+^2} - \frac{1}{r_-^2}\right)^{-1} \, .$$

In a similar way:

Problem 5.4. Show that

$$E = \frac{r_+^2 A(r_+) - r_-^2 A(r_-)}{r_+^2 - r_-^2} \, .$$

Note that our specific choice of E and J allows for turning points.

Problem 5.5. Show that

$$\frac{A(r)}{J^2} - \frac{E}{J^2} = \frac{r_-^2\left[A(r) - A(r_-)\right] - r_+^2\left[A(r) - A(r_+)\right]}{r_+^2 r_-^2\left[A(r_+) - A(r_-)\right]} \ .$$

From the above problem

$$\left(1 - \frac{r^*}{r}\right)\left[\frac{A}{J^2} - \frac{E}{J^2} - \frac{1}{r^2}\right] = C\left(\frac{1}{r_-} - \frac{1}{r}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right)\left(1 - \frac{r^*}{r}\right) \ ,$$

for some constant C, since the L.H.S. is a cubic polynomial in $1/r$. Therefore,

$$\frac{A}{J^2} - \frac{E}{J^2} - \frac{1}{r^2} = C\left(\frac{1}{r_-} - \frac{1}{r}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right) \ . \tag{5.6}$$

Problem 5.6. Evaluate the expression above as $r \to \infty$ and show that

$$C = 1 - r^*\left(\frac{1}{r_-} + \frac{1}{r_+}\right) \ .$$

Therefore,

$$\int_{r_-}^{r} d\varphi = C^{-\frac{1}{2}} \int_{r_-}^{r} \frac{\sqrt{A}}{r^2}\left[\left(\frac{1}{r_-} - \frac{1}{r}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right)\right]^{-\frac{1}{2}} dr \ .$$

To solve the above integral, use the following substitution

$$\frac{1}{r} = \frac{1}{2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right) + \frac{1}{2}\left(\frac{1}{r_+} - \frac{1}{r_-}\right)\sin\psi \ .$$

In this case, when $r = r_\pm$, we have that $\psi = \pm\pi/2$.

Problem 5.7. Using the above substitution, show that

$$\left[\left(\frac{1}{r_-} - \frac{1}{r}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right)\right]^{1/2} = \frac{1}{2}\left(\frac{1}{r_-} - \frac{1}{r_+}\right)\cos\psi \ .$$

Clearly,

$$-\frac{dr}{r^2} = \frac{1}{2}\left(\frac{1}{r_+} - \frac{1}{r_-}\right)\cos\psi \, d\psi \ .$$

Therefore,

$$\int_{r_-}^{r} \frac{\sqrt{A}}{r^2}\left[\left(\frac{1}{r_-} - \frac{1}{r}\right)\left(\frac{1}{r} - \frac{1}{r_+}\right)\right]^{-\frac{1}{2}} dr = \int_{-\pi/2}^{\psi} \sqrt{A}\, d\psi \ .$$

Expanding $C^{-\frac{1}{2}}$ and $\sqrt{A}$ (using eq.(5.6)) to first order in r^* we get that

$$\varphi(r) - \varphi(r_-) \approx \left[1 + \frac{r^*}{2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right]\int_{-\pi/2}^{\psi}\left(1 + \frac{1}{2}\frac{r^*}{r}\right)d\psi$$

$$= \left[1 + \tfrac{r^*}{2}\left(\tfrac{1}{r_+} + \tfrac{1}{r_-}\right)\right] \times$$

$$\int_{-\pi/2}^{\psi}\left[1 + \frac{r^*}{4}\left\{\left(\frac{1}{r_+} + \frac{1}{r_-}\right) + \left(\frac{1}{r_+} - \frac{1}{r_-}\right)\sin\psi\right\}\right]d\psi$$

$$= \left[1 + \tfrac{r^*}{2}\left(\tfrac{1}{r_+} + \tfrac{1}{r_-}\right)\right] \times$$

$$\left\{\left[1 + \frac{r^*}{4}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right]\left(\psi + \frac{\pi}{2}\right) - \frac{r^*}{4}\left(\frac{1}{r_+} - \frac{1}{r_-}\right)\cos\psi\right\} \ .$$

Problem 5.8. To order r^*, show that

$$\varphi(r) - \varphi(r_-) = \left[1 + \frac{3r^*}{4}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right]\left(\psi + \frac{\pi}{2}\right) - \frac{r^*}{4}\left(\frac{1}{r_+} - \frac{1}{r_-}\right)\cos\psi \ .$$

I.e.,

$$\varphi(r_+) - \varphi(r_-) \approx \left[1 + \frac{3GM}{2c^2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\right]\left(\frac{\pi}{2} + \frac{\pi}{2}\right) \ .$$

Then the perihelion advance of Mercury is given by

$$\Delta\varphi = 2\left[\varphi(r_+) - \varphi(r_-)\right] - 2\pi = 3\frac{GM}{c^2}\left(\frac{1}{r_+} + \frac{1}{r_-}\right)\pi \text{ rad/rev} \ .$$

For Mercury

$$\left(\frac{1}{r_+} + \frac{1}{r_-}\right) = 3.6166 \times 10^{-8} \text{ km}^{-1} \ ,$$

and

$$\frac{GM_\odot}{c^2} = 1.475 \text{ km} \ .$$

Therefore,

$$\Delta\varphi = 3 \times 1.475 \times \pi \times 3.6166 \times 10^{-8} \ \frac{\text{rad}}{\text{rev}} = 5 \times 10^{-7}\frac{\text{rad}}{\text{rev}} \ .$$
$$= 0.103 \ \frac{\text{arcsec}}{\text{rev}} \ .^{2}$$

Mercury makes 415 revolutions in a century, so over the course of a century, the perihelion shift in Mercury's orbit is given by

$$\Delta\varphi = 415 \times 0.103 = 43 \ \frac{\text{arcsec}}{\text{century}} \ .$$

5.1.3 Null Geodesics

For timelike geodesics we set $ds^2 = -c^2 d\tau^2$ to obtain eq.(5.2). For null geodesics, $ds^2 = 0$ and so

$$0 = -\left(1 - \frac{r^*}{r}\right)c^2\left(\frac{dt}{d\lambda}\right)^2 + \left(1 - \frac{r^*}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\theta}{d\lambda}\right)^2 + r^2\sin^2(\theta)\left(\frac{d\varphi}{d\lambda}\right)^2 \ ,$$

for some affine parameter λ. Once again,

$$L = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} \ ,$$

and

$$\frac{\partial\mathcal{L}}{\partial c\dot{t}} = \text{const} \equiv -\kappa$$

implies

$$\kappa = \left(1 - \frac{r^*}{r}\right)c\frac{dt}{d\lambda} \ . \tag{5.7}$$

Similarly, due to the φ-independence of the Lagrangian, we get that the angular momentum of the geodesic is conserved about the symmetry axis. I.e.,

$$j = r^2\sin^2\theta\frac{d\varphi}{d\lambda} \ .$$

Once again, without loss of generality, we restrict motion to the equatorial plane ($\theta = \frac{\pi}{2}$)), and the invariant interval becomes

$$0 = -\left(1 - \frac{r^*}{r}\right)c^2\dot{t}^2 + \left(1 - \frac{r^*}{r}\right)^{-1}\dot{r}^2 + r^2\dot{\varphi}^2 \ ,$$

or,

$$\dot{r}^2 = \kappa^2 - \frac{j^2}{r^2 A} \ .$$

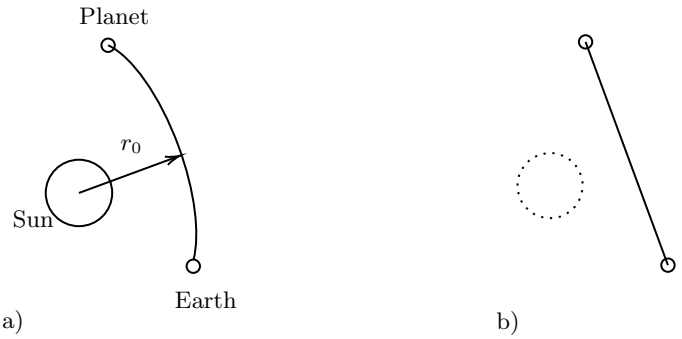


Fig. 5.2. a) shows a bent beam of light due to the presence of the Sun. b) shows a shorter path taken by the beam of light when the Sun is elsewhere.

5.1.4 Shapiro Time Delay

In 1964, Irwin Shapiro calculated the time delay in a radar echo from Earth to Venus and also led the effort to measure this fully general relativistic effect. Consider a null geodesic traveling from Earth located at $r = r_A$ to Venus located at $r = r_B$ and back. Let r_0 be the smallest value of r attained by this null geodesic (see figure 5.2). One can already guess that the time delay will be inversely related to r_0. Set $b = j/\kappa$. Then, from the equation above, along the null geodesic (light signal)

$$\frac{dr}{d\lambda} = \pm j \left[\frac{1}{b^2} - \frac{1}{r^2}\left(1 - \frac{r*}{r}\right)\right]^{\frac{1}{2}}$$

The time taken,

$$c\Delta T = c\int dt = c\int \frac{dt}{dr} dr \ ,$$

and

$$\frac{dt}{dr} = \frac{dt}{d\lambda}\frac{d\lambda}{dr}$$

$$= \frac{\kappa A}{cj}\left[\frac{1}{b^2} - \frac{1}{r^2}\left(1 - \frac{r*}{r}\right)\right]^{-\frac{1}{2}}$$

$$= \frac{A}{c}\left[1 - \frac{b^2}{r^2}\left(1 - \frac{r*}{r}\right)\right]^{-\frac{1}{2}} \ .$$

Note:

$$\frac{dr}{d\lambda}\Big|_{r=r_0} = 0 \rightarrow \frac{1}{b^2} = \frac{1}{r_0^2}\left(1 - \frac{r*}{r_0}\right) \ ,$$

or,

$$b^2 = r_0^2\left(1 - \frac{r*}{r_0}\right) \ .$$

Therefore,

$$\frac{dt}{dr} = \frac{1}{c}\left(1 - \frac{r*}{r}\right)^{-1}\left[1 - \frac{r_0^2}{r^2}\left(1 - \frac{r*}{r_0}\right)\left(1 - \frac{r*}{r}\right)\right]^{-\frac{1}{2}} \ .$$

Set

$$f(r^*) = \left(1 - \frac{r*}{r}\right)^{-1}\left[1 - \frac{r_0^2}{r^2}\left(1 - \frac{r*}{r_0}\right)\left(1 - \frac{r*}{r}\right)\right]^{-\frac{1}{2}} \ .$$

Then, to first order in r^*,

$$f(r^*) \approx f(r^* = 0) + \frac{df}{dr*}\Big|_{r^*=0} r^* \ .$$

Clearly,

$$f(r^* = 0) = \left(1 - \frac{r_0^2}{r^2}\right)^{-\frac{1}{2}} \ .$$

---

[2] $$1 \text{ rad} = \frac{360 \text{ deg} \times 3600 \text{ arcsec}}{2\pi \text{ deg}} = 2.06 \times 10^5 \text{ arcsec} \ .$$

The time delay $c\Delta T$ can be written as

$$c\Delta T = c\Delta T_0 + c\,\delta T\ ,$$

where $c\Delta T_0$ is the 0^{th} order term which leads to the Newtonian answer, while $c\,\delta T$ is the relativistic correction we are looking for. The 0^{th} order term,

$$\begin{aligned} c\Delta T_0 &= 2\left[\int_{r_0}^{r_A}\left(1-\frac{r_0^2}{r^2}\right)^{-\frac{1}{2}}dr + \int_{r_0}^{r^B}\left(1-\frac{r_0^2}{r^2}\right)dr\right] \\ &= 2\left[\sqrt{r_A^2-r_0^2}+\sqrt{r_B^2-r_0^2}\right]\ , \end{aligned}$$

is the expected Euclidean answer.

Problem 5.9. Show that the first-order term

$$\frac{df}{dr*}\Big|_{r^*=0} r^* = r^*\frac{1}{\sqrt{r^2-r_0^2}}\,\frac{r+\frac{3}{2}r_0}{r+r_0}\ .$$

The first order correction, and hence the approximate time delay, is then given by

$$c\delta T = 2r^*\left[\int_{r_0}^{r_A}\frac{1}{\sqrt{r^2-r_0^2}}\,\frac{r+\frac{3}{2}r_0}{r+r_0}\,dr + \int_{r_0}^{r_B}\frac{1}{\sqrt{r^2-r_0^2}}\,\frac{r+\frac{3}{2}r_0}{r+r_0}\,dr\right]\ .$$

Problem 5.10. Show that

$$\int\frac{1}{\sqrt{r^2-r_0^2}}\,\frac{r+\frac{3}{2}r_0}{r+r_0}\,dr = r^*\ln\left[\frac{r+\sqrt{r^2-r_0^2}}{r_0}\right]+\frac{r^*}{2}\left[\frac{\sqrt{r^2-r_0^2}}{r_0+r}\right]\ .$$

Hint: Use the following substitution $r = r_0\sec\theta$, and the fact that

$$\tan(\theta/2) = \frac{\sin(\theta/2)}{\cos(\theta/2)} = \frac{\sin\theta}{2\cos^2(\theta/2)} = \frac{\sin\theta}{1+\cos\theta}\ .$$

Then

$$\begin{aligned} c\delta T &= r^*\ln\left[\frac{r_B+\sqrt{r_B^2-r_0^2}}{r_0}\right]+\frac{r^*}{2}\left[\frac{\sqrt{r_B^2-r_0^2}}{r_0+r_B}\right] \\ &+r^*\ln\left[\frac{r_A+\sqrt{r_A^2-r_0^2}}{r_0}\right]+\frac{r^*}{2}\left[\frac{\sqrt{r_A^2-r_0^2}}{r_0+r_A}\right]\ . \end{aligned}$$

The above expression can be simplified for $r >> r_0$. In this case

$$c\,\delta T = 2\left[\frac{r^*}{2}+r^*\ln\frac{2r_A}{r_0}+\frac{r^*}{2}+r^*\ln\frac{2r_B}{r_0}\right] = \frac{4GM}{c^2}\left[1+\ln\left(\frac{4\,r_Ar_B}{r_0^2}\right)\right]\ .$$

Shapiro established this correction for a radar sent between Earth and Venus to be approximately $200\mu s$ [8].

5.2 The Schwarzschild Black Hole

Consider the vector field $\mathcal{T}^\mu_{Sch} = (1,0,0,0)$ in the Schwarzschild coordinate system. Then

$$\mathcal{T}_{Sch}\cdot\mathcal{T}_{Sch} = g_{\mu\nu}\mathcal{T}^\mu_{Sch}\mathcal{T}^\nu_{Sch} = -\left(1-\frac{r^*}{r}\right) < 0$$

when $r > r^*$. Thus, we set $\mathcal{T}_{Sch}$ as being future-pointing timelike in the exterior Schwarzschild spacetime. This choice is consistent with $\mathbb{M}^4$ as $r\to\infty$ as $ds^2_{sch}\to ds^2_{flat}$.

5.2.1 The problem with time

Consider an infalling timelike radial geodesic ($l = 0$) in the equatorial plane, such that $\dot{r} \to 0$ when $r = \infty$. Then, in eq.(5.4), $\epsilon = mc^2$, and

$$\frac{1}{2}\dot{r}^2 - \frac{GM}{r} = 0$$

i.e.,

$$cd\tau = \pm\sqrt{\frac{r}{r^*}}dr \ ,$$

or

$$\tau(r) - \tau(r_0) = -\frac{2r^*}{3c}\left[\left(\frac{r}{r^*}\right)^{3/2} - \left(\frac{r_0}{r^*}\right)^{3/2}\right] \ .$$

I.e, in finite amount of proper time, the object reaches r^*. Let us compare this to the time taken for the same event as measured in the Schwarzschild coordinate system. Noting that

$$\frac{dt}{dr} = \frac{dt}{d\tau}\cdot\frac{d\tau}{dr}$$

we get that

$$c\frac{dt}{dr} = -\sqrt{\frac{r}{r^*}}\left(1 - \frac{r^*}{r}\right)^{-1} \ .$$

Problem 5.11. Show that, in the equation above

$$c\Delta t = c\int dt = -\frac{1}{\sqrt{r^*}}\int \frac{r^{3/2}dr}{r - r^*}$$

$$= -\frac{2}{\sqrt{r^*}}\left[\frac{r^{3/2}}{3} + r^*\sqrt{r} + \frac{1}{2}(r^*)^{3/2}\left[\ln(\sqrt{r} - \sqrt{r^*}) - \ln(\sqrt{r} + \sqrt{r^*})\right]\right] \ .$$

From the above problem, we see that unlike the propertime interval, $c\Delta t \to +\infty$ as $r \to r^*$. This suggests that the Schwarzschild time coordinate may not be a judicious clock. Since light is the wild card in relativity, to understand how we might remedy the time coordinate, let us try a similar tactic with null geodesics.

5.2.2 "Probing" the Event Horizon with Light

Radial null geodesics satisfy

$$-\left(1 - \frac{r^*}{r}\right)c^2dt^2 + \left(1 - \frac{r^*}{r}\right)^{-1}dr^2 = 0$$

or,

$$cdt = \pm\left(1 - \frac{r^*}{r}\right)^{-1}dr \ . \tag{5.8}$$

What the above equation means is that an outgoing null geodesic (one where the radial coordinate r is increasing) is given by

$$\alpha(\lambda) = (ct(\lambda), r(\lambda), 0, 0) \ ,$$

where the 4-velocity is defined by

$$\frac{d\alpha}{d\lambda}(\lambda) = \left(\frac{dct}{d\lambda}(\lambda), \frac{dr}{d\lambda}(\lambda), 0, 0\right) = \left(\left(1 - \frac{r^*}{r}\right)^{-1}\frac{dr}{d\lambda}(\lambda), \frac{dr}{d\lambda}(\lambda), 0, 0\right) \ .$$

Therefore, the vector field l given by

$$l = \left(\left(1 - \frac{r^*}{r}\right)^{-1}, 1, 0, 0\right)$$

is proportional to the outgoing null geodesics. Similarly, in falling geodesics have 4-velocity

$$n = \left(\left(1 - \frac{r^*}{r}\right)^{-1}, -1, 0, 0\right) \ .$$

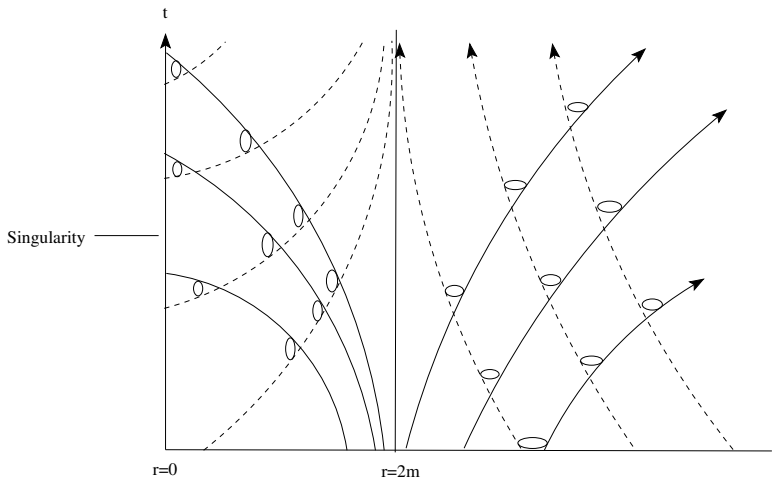


Fig. 5.3. The infalling radial geodesics are indicated by the dashed curves, while the outgoing geodesics are drawn in solid lines. In the exterior geometry, the arrows indicate that the time coordinate increases along the geodesics. At this stage, we are not yet able to make a similar assertion in the interior geometry.

Eq.(5.8) can be integrated to give

$$ct = \pm\big(r + r^* \ln|r - r^*|\big) + \text{const.}$$

In order to remove terms with units of natural log of length, we rewrite the above equation using a new integration constant as

$$ct = \pm\left(r + r^* \ln\left|\frac{r}{r^*} - 1\right|\right) + \text{const.}^{3}$$

This curve describes both radially infalling and outgoing null geodesics (see figure 5.3).

5.2.3 Black Hole Coordinates

We are going to use the infalling geodesic equation

$$ct + \left(r + r^* \ln\left|\frac{r}{r^*} - 1\right|\right) = \text{const.}$$

to define a new set of black hole coordinates given by the transformation

$$c\bar{t} = ct + r + r^* \ln\left|\frac{r}{r^*} - 1\right| \ , \ \bar{r} = r, \ \bar{\theta} = \theta, \ \text{ and } \ \bar{\varphi} = \varphi \ . \tag{5.9}$$

All quantities in the black hole coordinates will be indicated with a "bar" on the top, unlike their Schwarzschild coordinate counterparts. When transforming tensors, it is important to separate the old coordinates from the new ones.

Problem 5.12. Write the transformation matrix between the black hole coordinates and the Schwarzschild coordinates. When $r \neq r^*$ verify that the transformation matrix is invertible.

Problem 5.13. Show that in the black hole coordinate system $(c\bar{t}, \bar{r}, \bar{\theta}, \bar{\varphi})$, the Schwarzschild metric takes the form

$$ds^2_{BH} = -\left(1 - \frac{r^*}{\bar{r}}\right) d(\bar{ct})^2 + 2d(\bar{ct})d\bar{r} + \bar{r}^2\left(d\bar{\theta}^2 + \sin^2\bar{\theta} d\bar{\varphi}^2\right)^{4} \ . \tag{5.10}$$

3

$$ct = \pm\left(r + r^* \ln\left|r^*\left(\frac{r}{r^*} - 1\right)\right|\right) + \text{const.}$$

$$= \pm\left(r + r^* \ln\left|\frac{r}{r^*} - 1\right|\right) + r^* \ln r^* + \text{const.}$$

4 $2d(\bar{ct})\, d\bar{r}$ is really short for $d(\bar{ct})\, d\bar{r} + d\bar{r}\, d(\bar{ct})$, in other words $g_{ct\, r} = g_{r\, ct} = 1$.

In the above expression for the metric, since $\bar{r} = r$, and $\bar{\theta} = \theta$, it would not be erroneous to also write

$$ds^2_{BH} = -\left(1 - \frac{r^*}{r}\right) d(\bar{ct})^2 + 2d(\bar{ct})dr + r^2\left(d\theta^2 + \sin^2\theta d\varphi^2\right) \ .$$

In the black hole coordinate system, $r = r^*$ is no longer singular in eq.(5.10). The price to pay was the singular transformation (at $r = r^*$) in eq.(5.9).

Problem 5.14. Find $(g_{BH})^{\mu\nu}$.

In order to understand the time orientation in the interior geometry, we will write $\mathcal{T}_{Sch}$ in the Eddington-Finkelstein coordinate system. In this case

$$\bar{\mathcal{T}}^{\mu}_{Sch} = \frac{\partial \bar{x}^{\mu}}{\partial x^{\nu}} \mathcal{T}^{\nu}_{Sch}$$

implies that $\bar{\mathcal{T}}^{\mu}_{Sch} = (1, 0, 0, 0)$. Consequently, to be consistent with the exterior geometry, $\bar{\mathcal{T}}^{\mu}_{Sch}$ is set to be future-pointing timelike when $r > r^*$. Note, however, that we cannot use $\bar{\mathcal{T}}^{\mu}_{Sch}$ to define time orientation when $r < r^*$. Why?

Problem 5.15. In the black hole coordinate system, show that

$$\bar{l} = \left(\frac{2}{(1 - r^*/r)}, 1, 0, 0\right) \ ,$$

and that

$$\bar{n} = (0, -1, 0, 0) \ .$$

Theorem 5.1. *$\bar{n}$ is future pointing null where $r > r^*$.*

Proof.

$$\bar{\mathcal{T}}_{Sch} \cdot \bar{n} = -1 \ .$$

■

But $\bar{n}$ is nowhere vanishing and well defined for all values of $r > 0$. Therefore, we set $\bar{n}$ as future-pointing everywhere in the black hole spacetime.

Let

$$R^{\mu} = (0, -1, 0, 0) \tag{5.11}$$

in the Schwarzschild coordinates when $0 < r < r^*$. Clearly, R is timelike when $0 < r < r^*$.

Problem 5.16. Show that

$$\bar{R}^{\mu} = \left(\left(\frac{r^*}{r} - 1\right)^{-1}, -1, 0, 0\right) \ .$$

Theorem 5.2. *$\bar{R}^{\mu}$ is future pointing timelike when $0 < r < r^*$.*

Proof.

$$\bar{R} \cdot \bar{n} = -\frac{1}{(r^*/r - 1)} < 0$$

when $r < r^*$. ■

Consider a causal particle in the interior geometry (i.e., when $0 < r < r^*$) described by a curve

$$\alpha(\lambda) = (ct(\lambda), r(\lambda), \theta(\lambda), \varphi(\lambda)) \ .$$

Here, if α is a timelike, λ could be the propertime τ, or just an affine parameter if α is lightlike.

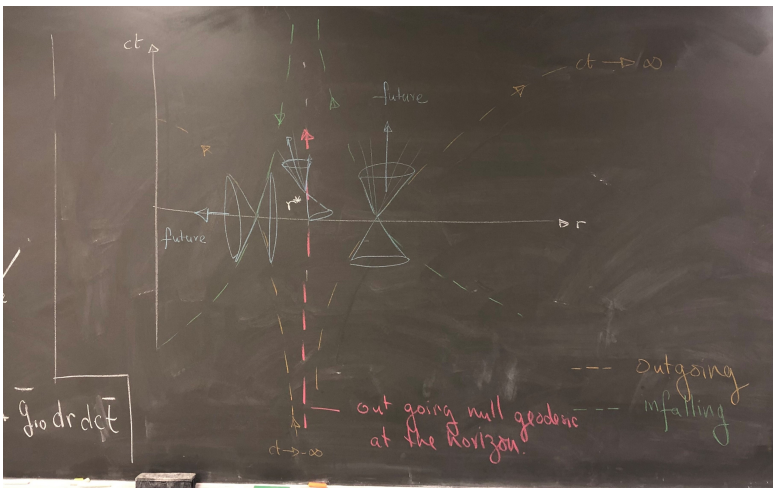


Fig. 5.4. Light cones tip over inside the event horizon.

Corollary 5.1. *In Schwarzschild coordinates, $dr/d\lambda < 0$ for a causal particle when $0 < r < r^*$.*

Proof. Let α be as above. Then

$$\dot{\alpha} = \left(c\frac{dt}{d\lambda}, \frac{dr}{d\lambda}, \frac{d\theta}{d\lambda}, \frac{d\varphi}{d\lambda} \right) .$$

Since $\dot{\alpha}$ and R (as given by eq.(5.11)) are future-pointing causal vector when $0 < r < r^*$, we must have that $g(\dot{\alpha}, R) < 0$. I.e.,

$$\left(1 - \frac{r^*}{r}\right)^{-1} \frac{dr}{d\lambda}(-1) < 0 .$$

Since here $(1 - r^*/r)^{-1} < 0$ we must have that $dr/d\lambda < 0$. ■

From the corollary above, we see that the metric given by eq.(5.10) does indeed describe a black hole. A causal particle in the interior geometry cannot escape to the region given by $r > r^*$.

Definition 5.1. *The metric in eq.(5.10), along with the time orientation given by $\bar{n}$ will be referred to as the Schwarzschild* **black hole** *spacetime.*

Problem 5.17. Show that $-l$ is future-pointing when $r < r^*$.

5.3 The Schwarzschild Black Hole Event Horizon

The surface described by $\bar{r} = r^*$ is 3-dimensional. In the black hole coordinates $(c\bar{t}, r, \theta, \varphi)$ (since Schwarzschild coordinates are not valid here), the tangent space of this surface is clearly spanned by

$$\tilde{N} = (1, 0, 0, 0) \, , \tilde{S}_1 = (0, 0, 1, 0), \text{ and } \tilde{S}_2 = (0, 0, 0, 1) \, .$$

The set $(\tilde{N}, \tilde{S}_1, \tilde{S}_2)$ hearkens back to theorem 2.14.

Problem 5.18. Show that $\tilde{N}$ is future-pointing null at $\bar{r} = r^*$.

Theorem 5.3. *$\tilde{N}$ is tangent to a null geodesic at $\bar{r} = r^*$.*

Proof.

$$\left(1 - \frac{r^*}{r}\right) \bar{l} = \left(2, \left(1 - \frac{r^*}{r}\right), 0, 0\right)$$

is tangent to the null geodesic tangent vector l. Now set $\bar{r} = r^*$, and the right hand side above becomes $(2, 0, 0, 0) = 2(1, 0, 0, 0)$. □

Problem 5.19. (Trivially) show that $\tilde{S}_1$, and $\tilde{S}_2$ are spacelike at $\bar{r} = r^*$.

Now consider a causal particle described by a curve

$$\alpha(\lambda) = (c\bar{t}(\lambda), \bar{r}(\lambda), \bar{\theta}(\lambda), \bar{\varphi}(\lambda))$$

that could potentially cross the surface $\bar{r} = r^*$. The 4-velocity of such a particle is future-pointing causal and is given by

$$\dot{\alpha} = \left(c\frac{d\bar{t}}{d\lambda}, \frac{d\bar{r}}{d\lambda}, \frac{d\bar{\theta}}{d\lambda}, \frac{d\bar{\varphi}}{d\lambda} \right) .$$

Theorem 5.4. *For a causal particle as described above,*

$$\frac{d\bar{t}}{d\lambda} > 0$$

at $\bar{r} = r^*$.

Proof.

$$0 > \dot{\alpha} \cdot \bar{n} = \frac{d\bar{t}}{d\lambda}(-1).$$

□

Problem 5.20. For a causal particle as described above,

$$\frac{d\bar{r}}{d\lambda} < 0$$

at $\bar{r} = r^*$.

From the results above, we conclude that the only causal vector tangent to $\bar{r} = r^*$ is $\tilde{N}$, which describes a lightlike geodesic. Any other causal vector at $\bar{r} = r^*$ points into the black hole. For this reason, the surface $\bar{r} = r^*$ is called the **event horizon**. Moreover, a beam of light can potentially live on the event horizon. See fig.(5.5) for the causal structure of the event horizon.

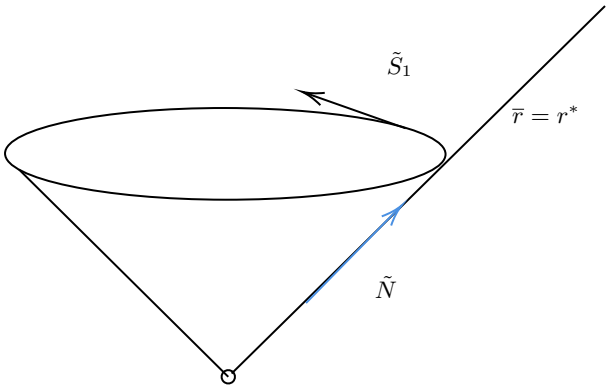


Fig. 5.5. The causal structure of the Schwarzschild event horizon. The circle represents a 2-dimensional sphere. Hence, S_2 (not shown) is also tangent to this sphere.

Problem 5.21. How will you place time orientation arrows on the null geodesics in fig.(5.3) when $r < r^*$? Convince yourself that it is consistent with fig.(5.4).

5.4 The Schwarzschild White Hole Spacetime

Just as in the black hole case, we will use the outgoing geodesic equation

$$ct - \left(r + r^* \ln \left| \frac{r}{r^*} - 1 \right| \right) = \text{const.}$$

to define a new set of white hole coordinates given by the transformation

$$c\tilde{t} = ct - r - r^* \ln \left| \frac{r}{r^*} - 1 \right| \ , \ \tilde{r} = r, \ \tilde{\theta} = \theta, \ \text{ and } \ \tilde{\varphi} = \varphi \ . \tag{5.12}$$

All quantities in the white hole coordinates $(c\tilde{t}, \tilde{r}, \tilde{\theta}, \tilde{\varphi})$ will be indicated with a "tilde" on top.

Problem 5.22. Rewrite the Schwarzschild metric using the white hole coordinates.

Problem 5.23. In white hole coordinates, show that the vector field $\tilde{l} = (0, 1, 0, 0)$ provides a time orientation for this spacetime that is consistent with the time orientation of the Schwarzschild spacetime in the exterior geometry.

Problem 5.24. Find the expression for $\tilde{n}$ in the white hole coordinates.

Problem 5.25. Write a complete set of arguments showing that in the white hole spacetime, all particles are ejected out from the region $\tilde{r} < r^*$.

Definition 5.2. *The Schwarzschild metric in the $(c\tilde{t}, r, \theta, \varphi)$ coordinates, along with the time orientation given by $\tilde{l}$ will be referred to as the Schwarzschild* **white hole** *spacetime.*

Definition 5.3. *The surface defined by $\tilde{r} = r^*$ is the horizon of the Schwarzschild white hole spacetime.*

Problem 5.26. Show that $-n$ is future-pointing when $r < r^*$ in the Schwarzschild white hole spacetime.

Problem 5.27. Compute n_μ in the Schwarzschild black hole and white hole spacetimes.

5.5 Maximal Schwarzschild Geometry

By now, you should be convinced that the Schwarzschild metric alone does not define whether it describes a white hole or a black hole. This happens because we can extend the region the metric describes by using clever coordinates and choosing judicious time orientation. Also, recall that constant values of $c\bar{t}$ describe infalling null geodesics, while constant values of $c\tilde{t}$ describe outgoing null geodesics. Here will stitch the black hole and white hole spacetime using an extended coordinate patch.

5.5.1 Transformations for the Exterior Black Hole Geometry

When $r > r^*$ note that

$$\ln \left| \frac{r}{r^*} - 1 \right| = \ln \left(\frac{r}{r^*} - 1 \right) \ .$$

Then

$$d\bar{ct} = cdt + r^* . \left(\frac{r}{r^*} - 1 \right)^{-1} \frac{1}{r^*} dr + dr = cdt + \left(1 - \frac{r^*}{r} \right)^{-1} dr \ ,$$

and similarly

$$d\tilde{ct} = cdt - \left(1 - \frac{r^*}{r} \right)^{-1} dr \ .$$

Therefore,

$$d\bar{ct} d\tilde{ct} = c^2 dt^2 - \left(1 - \frac{r^*}{r} \right)^{-2} dr^2 = - \left(1 - \frac{r^*}{r} \right)^{-1} \left[- \left(1 - \frac{r^*}{r} \right) c^2 dt^2 + \left(1 - \frac{r^*}{r} \right)^{-1} dr^2 \right] \ .$$

The Schwarzschild metric takes a simple form in the $(c\bar{t}, c\tilde{t}, \theta, \varphi)$ coordinate system given by

$$ds^2 = -\left(1 - \frac{r^*}{r}\right) dc\bar{t}dc\tilde{t} + r^2 d\Omega^2 \ .$$

In order to cross over to the interior region of the Schwarzschild spacetime, it is possible to get rid of the problematic term given by $(1 - r^*/r)$ in the above expression by introducing a scaled coordinate system

$$p = \exp\left(\frac{c\bar{t}}{2r^*}\right) \ , \ \text{ and } \ q = -\exp\left(\frac{-c\tilde{t}}{2r^*}\right) \ .$$

Then,

$$dc\bar{t}dc\tilde{t} = 4(r^*)^2 \exp\left(-\frac{(c\bar{t} - c\tilde{t})}{2r^*}\right) dpdq \ .$$

Since

$$\begin{aligned} c\bar{t} - c\tilde{t} &= \left[ct + r + r^* \ln\left(\frac{r}{r^*} - 1\right)\right] - \left[ct - r - r^* \ln\left(\frac{r}{r^*} - 1\right)\right] \\ &= 2r + 2r^* \ln\left(\frac{r}{r^*} - 1\right) \ , \end{aligned}$$

we get that

$$\exp\left(\frac{c\bar{t} - c\tilde{t}}{2r^*}\right) = \exp\left(\frac{r}{r^*}\right)\left(\frac{r}{r^*} - 1\right) \ .$$

Hence

$$ds^2 = -\left(1 - \frac{r^*}{r}\right) dc\bar{t}dc\tilde{t} + r^2 d\Omega^2 = -\left(1 - \frac{r^*}{r}\right) 4(r^*)^2 \exp\left(\frac{r}{r^*}\right)\left(\frac{r}{r^*} - 1\right)^{-1} dpdq + r^2 d\Omega^2 \ ,$$

or,

$$ds^2 = -\frac{4(r^*)^3}{r} \exp\left(\frac{r}{r^*}\right) dpdq + r^2 d\Omega^2 \ .$$

To obtain a Minkowski like coordinate system in the (ct, r) plane, define

$$cT = \frac{p + q}{2}, \ \text{ and } \ X = \frac{p - q}{2} \ .$$

Then,

$$ds^2 = \frac{4(r^*)^3}{r} \exp\left(\frac{r}{r^*}\right)(-c^2 dT^2 + dX^2) + r^2 d\Omega^2 \ [5] \ .$$

In order to obtain the transformation between (ct, r) and (cT, X), note that,

$$\begin{aligned} cT = \frac{p - q}{2} &= \frac{1}{2}\left[\exp\left(\frac{c\bar{t}}{2r^*}\right) - \exp\left(\frac{-c\tilde{t}}{2r^*}\right)\right] \\ &= \frac{1}{2}\left[\exp\left(\frac{ct}{2r^*}\right) - \exp\left(\frac{-ct}{2r*}\right)\right] \exp\left(\frac{r}{2r^*}\right) \exp\left(\frac{1}{2}\ln\left(\frac{r}{r^*} - 1\right)\right) \ . \end{aligned}$$

I.e.,

$$cT = \sqrt{\frac{r}{r^*} - 1}\ \exp\left(\frac{r}{2r^*}\right) \sinh\frac{ct}{2r^*} \ .$$

In exactly the same way

$$X = \sqrt{\frac{r}{r^*} - 1}\ \exp\left(\frac{r}{2r^*}\right) \cosh\frac{ct}{2r^*} \ .$$

The following observations will illuminate the advantages of the (cT, X) coordinate system when $r \geq r^*$ (we are going to allow the limit as $r \to r^*$ since the metric in no longer at this point in the (cT, X, θ, φ) coordinate system. See fig.(5.6).

[5] In the (cT, X) plane, the metric is almost flat except for the factor $(4(r^*)^3/r)\exp(r/r^*)$. Such terms are referred to as a conformal factor in the literature.

a. Clearly $X \geq 0$.

b.
$$\frac{cT}{X} = \tanh \frac{ct}{2r^*}$$
implies that $-1 \leq cT/X \leq 1$. Therefore, $-X \leq cT \leq X$. Additionally, constant values of ct are straight lines, and $cT = \pm X$ as $ct \to \pm\infty$ respectively.

c.
$$c^2T^2 - X^2 = pq = \exp\left(-\frac{(c\bar{t} - c\tilde{t})}{2r^*}\right) = -\left(\frac{r}{r^*} - 1\right)\exp\left(\frac{r}{r^*}\right)$$
implies curves of constant r are hyperbolas.

d. As $r \to r^*$, one is tempted to say that $cT = X = 0$. However, this is not necessarily the case as the limit is not well defined when $ct \to \pm\infty$. But, regardless of the value of ct, we have that $c^2T^2 - X^2 \to 0$ from above. Therefore, in general, all we can conclude is that $cT = \pm X$. Therefore, the hyperbolas of constant r degenerate into straight lines $cT = \pm X$ at r^*.

5.5.2 Transformations for the Interior Black Hole Geometry

When $r < r*$,
$$\ln\left|\frac{r}{r^*} - 1\right| = \ln\left(1 - \frac{r}{r^*}\right) .$$
In this case setting
$$cT = \frac{p-q}{2} \quad \text{and} \quad X = \frac{p+q}{2} ,$$
once again, we get that
$$ds^2 = \frac{4(r^*)^3}{r}\exp\left(\frac{-r}{r^*}\right)\left[-c^2dT^2 + dX^2\right] + r^2d\Omega^2 \ ^{6} .$$
In this case, it is easily shown that when $r < r^*$
$$cT = \sqrt{1 - \frac{r}{r^*}}\ \exp\left(\frac{r}{2r^*}\right)\cosh\frac{ct}{2r^*}$$
and
$$X = \sqrt{1 - \frac{r}{r^*}}\ \exp\left(\frac{r}{2r^*}\right)\sinh\frac{ct}{2r^*} .$$

Here are the various limits of coordinate values of the interior geometry in the (T, X, θ, φ) coordinate system. See fig.(5.6).

a. Clearly $cT \geq 0$.

b.
$$\frac{X}{cT} = \tanh\frac{ct}{2r^*}$$
implies that $-1 \leq X/cT \leq 1$. Therefore, $-cT \leq X \leq cT$. Additionally, constant values of ct are straight lines, and $X = \pm cT$ as $ct \to \pm\infty$ respectively.

c.
$$c^2T^2 - X^2 = \left(1 - \frac{r}{r^*}\right)\exp\left(\frac{r}{r^*}\right)$$
implies curves of constant r are hyperbolas.

[6] The old definition for cT and X would give
$$ds^2 = \frac{4(r^*)^3}{r}\exp\left(\frac{-r}{r^*}\right)\left[c^2dT^2 - dX^2\right] + r^2d\Omega^2 .$$
This is not quite consistent with the expression we had before.

d. Just as before, when $r \to r^*$, we get that $X = \pm cT$.

e. When $r < r^*$, r is bounded below and spacetime ends with the hyperbola at $r = 0$. This is given by $c^2T^2 - X^2 = 1$ and $cT = \sqrt{X^2 + 1}$.

5.5.3 Transformations for the White Hole Geometry

To extend spacetime to all 4 quadrants in the (cT, X) plane, all we need to do is the substitution $cT \to -cT$ and $X \to -X$.

In the White Hole region where $0 < r < r^*$,

$$cT = -\sqrt{1 - \frac{r}{r^*}}\ \exp\left(\frac{r}{2r^*}\right) \cosh\frac{ct}{2r^*}$$

and,

$$X = -\sqrt{1 - \frac{r}{r^*}}\ \exp\left(\frac{r}{2r^*}\right) \sinh\frac{ct}{2r^*}\ .$$

In a region describing another asymptotically flat region where $r > r^*$,

$$cT = -\sqrt{\frac{r}{r^*} - 1}\ \exp\left(\frac{r}{2r^*}\right) \sinh\frac{ct}{2r^*}$$

and,

$$X = -\sqrt{\frac{r}{r^*} - 1}\ \exp\left(\frac{r}{2r^*}\right) \cosh\frac{ct}{2r^*}\ .$$

5.5.4 Concluding Remarks

Then in all four regions, the maximal Schwarzschild spacetime is given by the metric

$$ds^2_{MaxSch} = \frac{4(r^*)^3}{r} \exp\left(\frac{-r}{r^*}\right) \left[-c^2dT^2 + dX^2\right] + r^2 d\Omega^2\ .$$

- Here, (cT, X) takes on all values bounded within the hyperbolas $c^2T^2 - X^2 = 1$.
- We set $\mathcal{T}^\mu_{MaxSch} = (1, 0, 0, 0)$ in the (cT, X, θ, φ) coordinate system as future-pointing and timelike everywhere.
- In both the black hole and white hole region $r = 0$ is manifestly problematic.
- As suggested by the term, in the white hole, the outgoing radial null vector is future pointing and crosses the horizon smoothly. I.e., all matter and energy in this region will pour out to either asymptotically flat regions or the black hole in its immediate future as shown in the diagram below.

Problem 5.28. Show that the notion of future is correctly captured by $\mathcal{T}^\mu_{MaxSch} = (1, 0, 0, 0)$ in the (cT, X, θ, φ) coordinate system. I.e., show that $\bar{n} \cdot \mathcal{T}_{MaxSch} < 0$ in the black hole spacetime and $\tilde{l} \cdot \mathcal{T}_{MaxSch} < 0$ in the white hole spacetime.

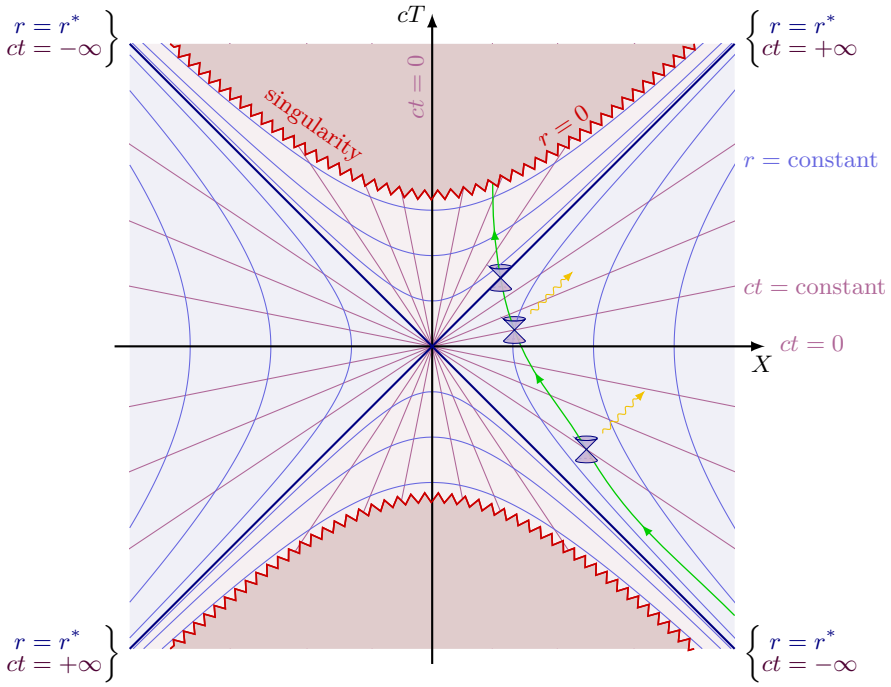


Fig. 5.6. The Kruskal diagram for the maximal Schwarzschild geometry (courtesy of https://tikz.net/relativity_kruskal_diagram/).

Part II

Lecture Notes for Relativity II (PHY 4478/5578)

6

Tensor Calculus

6.1 Vector and Tensor Bases

So far, we have written vectors as A^μ. This is clearly half the story. From previous experience, we know that vectors are written as

$$\mathbf{A} = A^\mu e_\mu \ ,$$

where $\{A^\mu\}$ are the components and $\{e_\mu\}$ are the bases vectors/vector fields. Now, suppose that we did insert the basis vectors back in. For the moment, we will write the bases set in a coordinate system $\{x^\mu\}$ for $T_x(\mathcal{M})$ as $\{e_\mu(x)\}$ as a placeholder. In short order, we will make it more precise. Note that, in general, it is not necessary for $\{e_\mu\}$ to be an orthonormal set. The vector $\mathbf{A}$ has to be the same physical vector in any two coordinate systems, say $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$. By this, we mean that

$$\mathbf{A} = A^\nu(x)e_\nu(x) = A^\mu(y)e_\mu(y) \ .$$

Then, it is easy to figure out how $\{e_\mu(x)\}$ transforms since

$$A^\nu(x)e_\nu(x) = A^\mu(y)e_\mu(y) = \left(\frac{\partial y^\mu}{\partial x^\nu}A^\nu(x)\right)e_\mu(y) = A^\nu(x)\left(\frac{\partial y^\mu}{\partial x^\nu}e_\mu(y)\right) \ .$$

Thus, we require that

$$e_\nu(x) = \frac{\partial y^\mu}{\partial x^\nu}e_\mu(y)$$

or

$$e_\mu(y) = \frac{\partial x^\nu}{\partial y^\mu}e_\nu(x) \ .$$

From eq.(4.4), we see that there is a direct correspondence between

$$e_\mu(x) \quad \text{and} \quad \frac{\partial}{\partial x^\mu}$$

in terms of their transformation properties. We formalize the above observation in the following definition.

Definition 6.1. *In a coordinate system $(x^1, \ldots, x^n)$ containing a point p in a spacetime $\mathcal{M}$, the tangent vector space at p is given by*

$$\text{span} \left\{ \frac{\partial}{\partial x^\mu}\Big|_p \ \Big| \mu = 1, \ldots, n \right\} \ .^{[1]}$$

Vector fields in this chart can be written as

$$A = A^\mu(x) \ \frac{\partial}{\partial x^\mu} \ ,$$

for any component functions $A^\mu(x)$. As a generalisation to the definition of definition 2.10, the tangent space at $p \in \mathcal{M}$ is denoted as $T_p(\mathcal{M})$, while the space of all tangent vector fields is denoted as $T(\mathcal{M})$:

$$T(\mathcal{M}) = \bigcup_p T_p(\mathcal{M})$$

and is referred to as the tangent bundle.

[1] In general relativity, we typically label coordinate indices from $0-4$ (there are exceptions). Although our sole attention will be devoted to spacetime manifolds, all of the results in the remainder of this chapter will be valid for arbitrary manifolds.

Notice that a tangent vector v_p at p takes a function f and maps it to a real number given by

$$v_p[f] = v^\mu \left.\frac{\partial f(x)}{\partial x^\mu}\right|_p ,$$

while tangent vector fields map functions to functions given by

$$A[f] = A^\mu(x) \frac{\partial f(x)}{\partial x^\mu} .$$

Remark 6.1. This means that, if we know how vector fields act on functions, we can recreate the vector field. In particular,

$$A[x^\nu] = A^\mu(x) \frac{\partial x^\nu}{\partial x^\mu} = A^\nu(x) .$$

As an example, the time orientation in the maximal geometry of the Schwarzschild spacetime given by $\mathcal{T}^\mu_{MaxSch} = (1,0,0,0)$ in the (cT, X, θ, φ) coordinate system can now be written as

$$\mathcal{T}^\mu_{MaxSch} = \frac{\partial}{\partial cT} .^{2}$$

Using the same arguments as above and eq.(4.5) show that:

Problem 6.1. Show that in a coordinate system $(x^1, \ldots, x^n)$ containing a point p in a spacetime $\mathcal{M}$, the space of dual vectors at p is given by

$$\text{span} \left\{ dx^\mu|_p \middle| \mu = 1, \ldots, n \right\} .$$

I.e., knowing how dx^μ transforms, show that

$$A_\mu(x(p))\, dx^\mu = A_\nu(y(p))\, dy^\nu .$$

Here $(y^1, \ldots, y^n)$ is another coordinate chart.

Definition 6.2. *The dual vector fields in this chart can be written as $A_\mu(x)\, dx^\mu$. The space of dual vectors at $p \in \mathcal{M}$ is denoted as $T^*_p(\mathcal{M})$, while the space of all the dual vector fields is denoted as $T^*(\mathcal{M})$ and is referred to as the cotangent bundle.*

As we have shown, tangent vectors map functions to numbers. But, what do dual vectors do? Dual vectors act on vectors and produce a number. You have already seen this, Suppose χ_p is a dual vector and A_p is a tangent vector at the same point p then χ_p acting on A_p is given by

$$\chi_p(A_p) = (\chi_p)_\mu A^\mu_p .$$

We previously referred to this as a contraction and demonstrated that this action is coordinate-independent.

Then

$$\chi(A) = \chi_\mu A^\mu = \chi_\mu(x)\, dx^\mu \left(A^\nu(x) \frac{\partial}{\partial x^\nu} \right) .$$

This can happen only if we define the action of the dual basis on a vector basis as follows:

$$dx^\mu \left(A^\nu(x) \frac{\partial}{\partial x^\nu} \right) = A^\nu(x)\, dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) ,^{3}$$

and

$$dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta^\mu_\nu .$$

The above equation holds when dx^μ acts on ∂_ν at the same fixed point.

[2] We have already gained quite a bit in specificity. In the past, we had to write $\mathcal{T}^\mu_{MaxSch} = (1,0,0,0)$ in the (cT, X, θ, φ) coordinate system. On the other hand

$$\mathcal{T}^\mu_{MaxSch} = \frac{\partial}{\partial cT}$$

need no further reference to a coordinate system.

[3] This makes dx^μ is function linear, meaning that its not just constants that come out of the parenthesis, but even functions like $A^\nu(x)$ do.

Problem 6.2. Show that the above equation is true regardless of the choice of coordinates. I.e., for are any two coordinate charts $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$ show that

$$dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta^\mu_\nu$$

implies that

$$dy^\mu \left(\frac{\partial}{\partial y^\nu} \right) = \delta^\mu_\nu \ .$$

Okay, so we have been able to create a basis for covariant and contravariant vector fields. But what about tensors? Consider first the metric tensor $g_{\mu\nu}$ written in the form

$$g = g_{\mu\nu}(x)\ dx^\mu \otimes dx^\nu \ [4] \ .$$

Notation: When the coordinate system being used is clear, say it is $(x^1, \ldots, x^n)$ we often write the tangent basis vectors $\partial/\partial x^\mu$ as ∂_μ.

The metric tensor is designed to take any two vectors and return a number. Let A_p and B_p be any two tangent vectors at the point $p \in \mathcal{M}$. Define the action $g(A_p, B_p)$ by

$$g(A_p, B_p) = g_{\mu\nu}(x)\ dx^\mu \otimes dx^\nu(A_p, B_p) \equiv g_{\mu\nu}(x)\ dx^\mu(A_p)\ dx^\nu(B_p) = g_{\mu\nu}(x)\ \ dx^\mu(A^\alpha_p\ \partial_\alpha)\ \ dx^\nu(B^\beta_p\ \partial_\beta)$$

$$= g_{\mu\nu}(x) A^\alpha_p\ B^\beta_p\ \ dx^\mu(\ \partial_\alpha)\ \ dx^\nu(\ \partial_\beta) = g_{\mu\nu}(p) A^\alpha_p\ B^\beta_p\ \ \delta^\mu_\alpha\ \ \delta^\nu_\beta \ [5] = g_{\mu\nu}(p) A^\mu_p\ B^\nu_p \ .$$

This is exactly what we wanted. In the event that A and B are vector fields, we would then have

$$g(A, B)(x) = g_{\mu\nu}(x) A^\mu(x)\ B^\nu(x) \ .$$

g belongs to the space of covariant tensors of rank 2, denoted by $T_2(\mathcal{M})$. Clearly,

$$T_2(\mathcal{M}) = \text{span}\ \left\{ dx^\mu \otimes dx^\nu \middle| \mu, \nu = 1, \ldots, n \right\} \ .$$

The components of the metric tensor g are easily seen to be given by

$$g_{\mu\nu} = g(\partial_\mu, \partial_\nu) \ .$$

To see how the tensor transforms, consider two coordinate systems given by $(x^1, \ldots, x^n)$ and $(y^1, \ldots, y^n)$. Then

$$g(y) = g_{\mu\nu}(y) dy^\mu \otimes dy^\nu = g_{\mu\nu}(y) \frac{\partial y^\mu}{\partial x^\alpha} dx^\alpha \otimes \frac{\partial y^\nu}{\partial x^\beta} dx^\beta = g_{\mu\nu}(y) \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial y^\nu}{\partial x^\beta}\ dx^\alpha \otimes dx^\beta \ ,$$

which we write as

$$g_{\alpha\beta}(x)\ dx^\alpha \otimes dx^\beta \ .$$

I.e.,

$$g_{\alpha\beta}(x) = g_{\mu\nu}(y) \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial y^\nu}{\partial x^\beta} \ ,$$

or equivalently

$$g_{\mu\nu}(y) = g_{\alpha\beta}(x) \frac{\partial x^\alpha}{\partial y^\mu} \frac{\partial x^\beta}{\partial y^\nu} \ .$$

This is precisely what eq.(4.6) gives. The contravariant form of the metric tensor is given by (also denoted by g)

$$g = g^{\mu\nu} \frac{\partial}{\partial x^\mu} \otimes \frac{\partial}{\partial x^\nu}$$

where $g^{\mu\nu}$ is the metric inverse of $g_{\mu\nu}$. In this case, the components are given by

[4] $\otimes$ is referred to as a tensor product, and $dx^\mu \otimes dx^\nu$ is read as dx^μ tensor product dx^ν.

[5] Here we tacitly implying that

$$dx^\mu|_p \left(\frac{\partial}{\partial x^\nu}|_q \right) = \delta^\mu_\nu$$

if $p = q$, and 0 otherwise.

$$g^{\mu\nu} = g(dx^\mu, dx^\nu) \ .$$

What about tensors with covariant and contravariant components? For example, a tensor T with covariant rank 2, and contravariant rank 1 can be written as

$$T = T^\mu{}_{\nu\lambda}\ \partial_\mu \otimes dx^\nu \otimes dx^\lambda \ .$$

T would then take 1 covariant vector, say χ, and two contravariant vectors A and B, and return a scalar (function or a number depending on whether it is a tensor field or just a tensor at a point). Here,

$$T(\chi, A, B) = T^\mu{}_{\nu\lambda}\ \partial_\mu \otimes dx^\nu \otimes dx^\lambda(\chi, A, B) = T^\mu{}_{\nu\lambda}\ \chi(\partial_\mu)\ dx^\nu(A)\ dx^\lambda(B)^6$$
$$= T^\mu{}_{\nu\lambda}\ \chi_a dx^a(\partial_\mu)\ dx^\nu(A^b\partial_b)\ dx^\lambda(B^c\partial_c) = T^\mu{}_{\nu\lambda}\ \chi_a\ A^b\ B^c\ \delta^a_\mu\ \delta^\nu_b\ \delta^\lambda_c = T^\mu{}_{\nu\lambda}\ \chi_\mu\ A^\nu\ B^\lambda \ .$$

In this case $T \in T^1_2(\mathcal{M})$, and

$$T^\mu{}_{\nu\lambda} = T\left(dx^\mu, \frac{\partial}{\partial x^\nu}, \frac{\partial}{\partial x^\lambda}\right) \ .$$

In a similar manner we can construct tensors fields $T^s_r(\mathcal{M})$ as a multilinear functional on r many vectors and s many dual vectors:

$$T : T(\mathcal{M}) \times \cdots \times T(\mathcal{M}) \times T^*(\mathcal{M}) \times \cdots \times T^*(\mathcal{M}) \times \to \mathbb{R} \ . \tag{6.1}$$

In the above expression, there are r many $T(\mathcal{M})$ and s many $T^*(\mathcal{M})$. It is important to note that the defining property of tensors is that they are **multi-function-linear objects**, i.e.,

$$T(A, B, \alpha\ C + \beta D, ..., E) = \alpha\ T(A, B, C, ..., E) + \beta\ T(A, B, D, ..., E) \ . \tag{6.2}$$

Here α and β are functions on the manifold and A, B, C, D, E are either tangent vector fields (*contravariant vectors*) or one forms (*covariant vectors*) acting on a tensor field of an appropriate type. This property would hold true at any entry of tensors of arbitrary type.The components of such a tensor in an arbitrary coordinate basis can be obtained by the formula

$$T_{i_1 i_2 \ldots i_r}{}^{j_1 j_2 \ldots j_s} = T\left(\partial_{i_1}, \ldots, \partial_{i_r}, dx^{j_1}, \ldots, dx^{j_s}\right) \ , \tag{6.3}$$

and having obtained the components, the tensor can be written as

$$T = T_{i_1 i_2 \ldots i_r}{}^{j_1 j_2 \ldots j_s}\ dx^{i_1} \otimes \cdots \otimes dx^{i_r} \otimes \partial_{j_1} \otimes \cdots \otimes \partial_{j_s} \ . \tag{6.4}$$

Tensor products can also be used to construct other tensors. For example, if T is a tensor field of type $T^s_r(\mathcal{M})$ and S is a tensor field of type $T^m_l(\mathcal{M})$, then $T \otimes S$ is a tensor field of type $T^{s+m}_{r+l}(\mathcal{M})$ whose components are given by

$$(T \otimes S)_{i_1 \ldots i_r\ p_1 \ldots p_l}{}^{j_1 \ldots j_s\ q_1 \ldots q_m} = T_{i_1 \ldots i_r}{}^{j_1 \ldots j_s}\ S_{p_1 \ldots p_l}{}^{q_1 \ldots q_m} \ . \tag{6.5}$$

The components of tensors transform exactly as given in eq.(4.6), all the previous properties like raising and lowering of indices work the same way. Also, if f is a function and T is a tensor, then their tensor product $f \otimes T$ is simply taken to be fT. I.e, functions can be thought of as rank 0 tensors. Tensor indices can be raised, lowered, and contracted just as before.

The following examples will help clarify the new machinery that we have developed. The Schwarzschild metric tensor can now be written in the form

$$g_{Sch} = -\left(1 - \frac{r^*}{r}\right) dct \ \otimes \ dct + \left(1 - \frac{r^*}{r}\right)^{-1} dr \ \otimes \ dr + r^2 d\theta \ \otimes \ d\theta + r^2 \sin^2\theta d\varphi \ \otimes \ d\varphi \ ,$$

and the Schwarzschild black hole metric is given by

$$g_{BH} = -\left(1 - \frac{r^*}{\bar{r}}\right) d\bar{ct} \ \otimes \ d\bar{ct} + d\bar{ct} \ \otimes \ d\bar{r} + d\bar{r} \ \otimes \ d\bar{ct} + \bar{r}^2 \left(d\bar{\theta} \ \otimes \ d\bar{\theta} + \sin^2\bar{\theta}\ d\bar{\varphi} \ \otimes \ d\bar{\varphi}\right)$$

[6] Notice how ∂_μ acting on χ is written as $\chi(\partial_\mu)$.

Example 6.1. Find the contravariant form (both indices raised) of the Schwarzschild metric.

Solution: Since the Schwarzschild metric is diagonal,

$$g^{\mu\nu} = \frac{1}{g_{\mu\nu}} \, .$$

Then

$$g_{Sch} = -\left(1 - \frac{r^*}{r}\right)^{-1} \frac{\partial}{\partial ct} \otimes \frac{\partial}{\partial ct} + \left(1 - \frac{r^*}{r}\right) \frac{\partial}{\partial r} \otimes \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial}{\partial \theta} \otimes \frac{\partial}{\partial \theta} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial}{\partial \varphi} \otimes \frac{\partial}{\partial \varphi} \, .$$

■

Example 6.2. Find an orthonormal bases vectors for the Schwarzschild metric.

Solution: Since

$$g_{Sch}\left(\frac{\partial}{\partial ct}, \frac{\partial}{\partial ct}\right) = -\left(1 - \frac{r^*}{r}\right) \, ,$$

set

$$e_0 = \left(1 - \frac{r^*}{r}\right)^{-1/2} \frac{\partial}{\partial ct} \, .$$

Then

$$g_{Sch}(e_0, e_0) = g_{Sch}\left(\left(1 - \frac{r^*}{r}\right)^{-1/2} \frac{\partial}{\partial ct}, \left(1 - \frac{r^*}{r}\right)^{-1/2} \frac{\partial}{\partial ct}\right)$$

$$= \left(1 - \frac{r^*}{r}\right)^{-1} g_{Sch}\left(\frac{\partial}{\partial ct}, \frac{\partial}{\partial ct}\right) = \left(1 - \frac{r^*}{r}\right)^{-1} (-1) \left(1 - \frac{r^*}{r}\right) = -1 \, .$$

Similarly

$$e_1 = \left(1 - \frac{r^*}{r}\right)^{1/2} \frac{\partial}{\partial r}$$

is a unit vector, i.e. $g(e_1, e_1) = 1$. Moreover $g(e_0, e_1) = 0$, i.e., e_0 and e_1 are orthonormal. We can now easily complete the basis by setting

$$e_2 = \frac{1}{r} \frac{\partial}{\partial \theta} \, ,$$

and

$$e_3 = \frac{1}{r \sin \theta} \frac{\partial}{\partial \varphi} \, .$$

■

Example 6.3. Write the Schwarzschild orthonormal basis vectors in the black hole coordinate system, and show that their properties hold.

Solution: While we can independently search for an orthonormal basis, all we have to do is transform the Schwarzschild orthonormal basis to the black hole coordinate system. Since

$$\frac{\partial}{\partial ct} = \frac{\partial c\bar{t}}{\partial ct} \frac{\partial}{\partial c\bar{t}} \, ,$$

set

$$e_0 = \left(1 - \frac{r^*}{r}\right)^{-1/2} \frac{\partial}{\partial c\bar{t}} \, .$$

Then clearly $g_{Sch}(e_0, e_0) = -1$. On the other hand

$$\frac{\partial}{\partial r} = \frac{\partial c\bar{t}}{\partial r} \frac{\partial}{\partial c\bar{t}} + \frac{\partial \bar{r}}{\partial r} \frac{\partial}{\partial \bar{r}} = \left(1 - \frac{r^*}{r}\right)^{-1} \frac{\partial}{\partial c\bar{t}} + \frac{\partial}{\partial \bar{r}} \, .$$

Then

$$e_1 = \left(1 - \frac{r^*}{r}\right)^{1/2} \frac{\partial}{\partial r} = \left(1 - \frac{r^*}{r}\right)^{1/2} \left[\left(1 - \frac{r^*}{r}\right)^{-1} \frac{\partial}{\partial c\bar{t}} + \frac{\partial}{\partial \bar{r}}\right] \, .$$

Then

$$g_{BH}(e_1, e_1) = 1 \ ,$$

and similarly

$$g_{BH}(e_0, e_1) = 0$$

as required. Also, e_2 and e_3 do not change their form or properties in the Eddington-Finkelstein coordinate system.

■

The orthonormal basis vectors for the black hole metric derived above is not valid across the horizon since we simply transformed the orthonormal basis vectors of the Schwarzschild metric. We remedy this deficiency in the following example.

Problem 6.3. Find an orthonormal bases vectors for the black hole metric that is valid across the horizon.

Problem 6.4. Find the contravariant form of the black hole metric.

Problem 6.5. Show that the contravariant form of the black hole metric in the basis e_0, e_1, e_2, e_3 (as derived in the example above) is given by

$$g = -e_0 \otimes e_0 + e_1 \otimes e_1 + e_2 \otimes e_2 + e_3 \otimes e_3 \ [7] \ .$$

Hint: If you expand the above metric by putting in the explicit form of the basis vectors, you should get the contravariant form of the Eddington-Finkelstein metric.

Problem 6.6. Find the dual frame for the orthonormal frame in the black hole metric in example 6.3. I.e., raise the index of all four of the vector fields, and label them (e^0, e^1, e^2, e^3) respectively.

The above result is surely expected. A spacetime metric will always have the appearance of the Minkowski metric when written in an orthonormal basis.

Problem 6.7. Using the dual frame (e^0, e^1, e^2, e^3) found above, show that the black hole metric can be written in the form

$$g_{EF} = -e^0 \otimes e^0 + e^1 \otimes e^1 + e^2 \otimes e^2 + e^3 \otimes e^3 \ .$$

Example 6.4. In the Schwarzschild spacetime, let

$$F = -q \sin\theta \big(d\theta \ \otimes \ d\varphi - d\varphi \ \otimes \ d\theta \big) \ .$$

Here q is a constant. Raise the first index of this tensor.

Solution: The expression for F tells us that

$$F_{\theta\varphi} = -q \sin\theta = -F_{\varphi\theta} \ .$$

We need to first obtain $F^{\mu}{}_{\nu}$.

$$F^{\mu}{}_{\nu} = g^{\mu\lambda}_{Sch} F_{\lambda\nu} \ .$$

Since the Schwarzschild metric is diagonal

$$F^{\theta}{}_{\varphi} = g^{\theta\lambda}_{Sch} \ F_{\lambda\varphi} = g^{\theta\theta}_{Sch} \ F_{\theta\varphi} = -\frac{1}{r^2} \ q \sin\theta \ ,$$

and

$$F^{\varphi}{}_{\theta} = g^{\varphi\lambda}_{Sch} \ F_{\lambda\theta} = g^{\varphi\varphi}_{Sch} \ F_{\varphi\theta} = \frac{1}{r^2 \sin^2\theta} q \sin\theta = \frac{q}{r^2 \sin\theta} \ .$$

These are all the non-trivial components. Then, when we raise the first index of F we get

$$F = -\frac{q \sin\theta}{r^2} \ \frac{\partial}{\partial\theta} \otimes d\varphi + \frac{q}{r^2 \sin\theta} \ \frac{\partial}{\partial\varphi} \otimes d\theta \ .$$

■

[7] In an orthonormal basis, all spacetime metrics will take the form of the Minkowski metric.

Problem 6.8. In the black hole spacetime, let

$$F = u(\varphi)\big(d\varphi \otimes dc\bar{t} - dc\bar{t} \otimes d\varphi\big) .$$

Here u is an arbitrary function of φ. Write F in a completely contravariant form. Also, find the tensor

$$F^{\mu\nu}F_{\nu\lambda}\,\frac{\partial}{\partial x^{\mu}} \otimes dx^{\lambda} .$$

6.2 Lie Brackets of Vector Fields

Definition 6.3. *Following remark 6.1, the lie bracket $[X, Y]$ of two tangent vector fields X and Y is a vector field determined by the expression*

$$[X, Y]f = X(Yf) - Y(Xf) \tag{6.6}$$

for any real valued function f.

Recall that

$$\frac{\partial}{\partial x^{\mu}} \equiv \partial_{\mu} .$$

Now we further set

$$\partial_{\mu} f = f_{,\mu} , \quad \text{and} \quad \partial_{\mu}\partial_{\nu} f = f_{,\nu,\mu} \text{ etc.}$$

Theorem 6.1. *In components*

$$[X, Y] = \big(Y^{i}_{,j}X^{j} - X^{i}_{,j}Y^{j}\big)\ \partial_{i} . \tag{6.7}$$

Here, $X = X^{i}\partial_{i}$ and $Y = Y^{j}\partial_{j}$.

Proof.

$$\begin{aligned}[X, Y]f &= X(Y^{j}f_{,j}) - Y(X^{i}f_{,i}) \\ &= X^{i}\ \partial_{i}(Y^{j}f_{,j}) - Y^{j}\ \partial_{j}(X^{i}f_{,i}) .\end{aligned}$$

Since $f_{,i,j} = f_{,j,i}$, we have, after a rearrangement of dummy labels, that

$$[X, Y]f = \big(Y^{i}_{,j}X^{j} - X^{i}_{,j}Y^{j}\big)\ \partial_{i}f$$

for arbitrary f. □

Example 6.5. In Schwarzschild coordinates, let

$$X = \sin^{2}\theta\ \frac{\partial}{\partial ct} - r^{2}\cos^{2}\theta\frac{\partial}{\partial\theta} ,$$

and let

$$Y = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta} .$$

Find $[X, Y]$.

Solution:

$$[X, Y] = \left(\frac{r^{2}\cos^{3}\theta}{\sin^{2}\theta}\right)\frac{\partial}{\partial\theta} - 2\cos\theta\frac{\partial}{\partial ct} - 2r^{2}\cos\theta\frac{\partial}{\partial\theta} .$$

■

Problem 6.9. In $\mathbb{R}^{2}$ with coordinates (x, y) define vector fields $Y = 3y\partial_{x} + 2x\partial_{y}$ and $Z = x^{2}\partial_{y}$. Find $L = [Y, Z]$. Now consider a change of coordinates from (x, y) to (u, v) defined by $u = x + y$ and $v = x - y$. Find Y Z and L in the (u, v) coordinate system and denote it as $\bar{Y}$, $\bar{Z}$ and $\bar{L}$. Find $[\bar{Y}, \bar{Z}]$. Confirm that $\bar{L} = [\bar{Y}, \bar{Z}]$.

The above result is true in general. I.e., the Lie Bracket defined by eq. (6.7) is independent of the coordinate system, as is shown below.

Theorem 6.2. *Lie brackets of two vector fields are a coordinate-independent definition, i.e.,*

$$\left[\bar{X},\bar{Y}\right]^l = [X,Y]^i \, \frac{\partial \bar{x}^l}{\partial x^i} \, .$$

Here, $\{x^\alpha\}$ *and* $\{\bar{x}^\alpha\}$ *are two overlapping coordinate systems.*

Proof.

$$\begin{aligned}\left[\bar{X},\bar{Y}\right]^l &= \bar{X}^j \frac{\partial \bar{Y}^l}{\partial \bar{x}^j} - \bar{Y}^j \frac{\partial \bar{X}^l}{\partial \bar{x}^j} = X^m \frac{\partial \bar{x}^j}{\partial x^m}\frac{\partial x^n}{\partial \bar{x}^j}\frac{\partial \bar{Y}^l}{\partial x^n} - Y^m \frac{\partial \bar{x}^j}{\partial x^m}\frac{\partial x^n}{\partial \bar{x}^j}\frac{\partial \bar{X}^l}{\partial x^n} \\ &= X^n \frac{\partial \bar{Y}^l}{\partial x^n} - Y^n \frac{\partial \bar{X}^l}{\partial x^n} = X^n \frac{\partial}{\partial x^n}\left(Y^i \frac{\partial \bar{x}^l}{\partial x^i}\right) - Y^n \frac{\partial}{\partial x^n}\left(X^i \frac{\partial \bar{x}^l}{\partial x^i}\right) \\ &= \left(X^n \frac{\partial Y^i}{\partial x^n} - Y^n \frac{\partial X^i}{\partial x^n}\right)\frac{\partial \bar{x}^l}{\partial x^i} \, .\end{aligned}$$

□

Clearly $[X,Y] = -[Y,X]$, and in any coordinate system $\{x^i\}$, $[\partial_i, \partial_j] = 0$.

6.3 The Covariant Derivative

In order to take derivatives of tensor fields that live on an arbitrary manifold in a coordinate-independent way, we axiomatically introduce the following set of properties we take for granted in the Calculus of $\mathbb{R}^n$. Derivatives of tensors are taken with respect to a tangent vector. This derivative should not affect the tensor type, i.e., the derivative of an $T^s_r(\mathcal{M})$ tensor is another $T^s_r(\mathcal{M})$ tensor. By $\nabla_X T$ we mean the covariant derivative (as it is called) with respect to the tangent vector X of the tensor field T [8]. Following is a list of properties that we would like the derivative operator to have:

1. $$\nabla_X(aT + bS) = a\nabla_X T + b\nabla_X S \tag{6.8}$$
   Here $a, b \in \mathbb{R}$ and T and S are tensors of arbitrary type. Additionally,
   $$\nabla_{fX+gY} T = f\nabla_X T + g\nabla_Y T \tag{6.9}$$
   Here f and g are functions, X and Y are any tangent vectors, and T is any tensor.

2. For scalar functions f
   $$\nabla_X f = X(f) \, . \tag{6.10}$$

3. For arbitrary tensors T and S
   $$\nabla_X(T \otimes S) = (\nabla_X T) \otimes S + T \otimes \nabla_X S \, . \tag{6.11}$$
   This is nothing more than the usual product rule for derivatives. A scalar function f can be thought of a tensor of type $T^0_0(\mathcal{M})$ where $f \otimes T \equiv f\, T$.

4. $$\nabla_X \, CT = C \, \nabla_X T \tag{6.12}$$
   Here C is a contraction of any pair/pairs of covariant and contravariant indices. The need for this axiom will be made clear in theorem 6.4.

In order to pin down the requirements of coordinate- independence, let us fix a coordinate system and set

$$\nabla_{\partial_\mu}\partial_\nu \equiv \Gamma^\alpha_{\mu\nu} \, \partial_\alpha \, . \tag{6.13}$$

The functions $\Gamma^\alpha_{\mu\nu}$'s are called *connection coefficients* or *Christoffel symbols*. As we shall see, the $\Gamma^\alpha_{\mu\nu}$'s do not form the components of a tensor. Using the properties satisfied by the covariant derivative we see that

$$\nabla_X Y \equiv \nabla_X(Y^\nu \partial_\nu) = (X^\mu \, Y^\alpha_{\ ,\mu} + X^\mu \, Y^\nu \, \Gamma^\alpha_{\mu\nu}) \, \partial_\alpha \, . \tag{6.14}$$

Here Y is a vector field and X is a tangent vector at the point we wish to take the derivative.

Notation

[8] While in general T is a tensor field, X can simply be a tangent vector at a particular point.

- $\nabla_X Y = \nabla_{(X^\mu \partial_\mu)} Y \equiv X^\mu \nabla_{\partial_\mu} Y \equiv X^\mu \nabla_\mu Y$.

- Let T be an arbitrary tensor field of type $T^s_r(\mathcal{M})$. Since

$$\nabla_{f\ X} T = f\ X^\mu\ \nabla_\mu T\ ,$$

for any tangent vector X and smooth function f. I.e., the covariant derivative is automatically function-linear in its subscripted entry. We can then define an absolute covariant derivative

$$\nabla : T^s_r(\mathcal{M}) \to T^s_{r+1}(\mathcal{M})$$

defined by

$$\nabla T(Y_1, \cdots, Y_r, X, W^1, \cdots, W^s) = (X^\mu \nabla_\mu T)(Y_1, \cdots, Y_r, W^1, \cdots, W^s)\ .$$

Here, as usual $Y_1, \ldots, Y_r$ and X are vectors, and $W^1, \ldots, W^s$ are dual vectors.

Abuse of Notation

$$\nabla_\nu Y^\mu \equiv (\nabla_\nu Y)^\mu\ ,$$

where $(\nabla_\nu Y)^\mu$ is the μth component of the vector $\nabla_\nu Y = (\nabla_\nu Y)^\mu\ \partial_\mu$. Mathematicians generally avoid this abused notation. We may not! And hence,

$$\nabla_X Y^\alpha = X^\mu\ Y^\alpha_{\ ,\mu} + X^\mu\ Y^\nu\ \Gamma^\alpha_{\mu\nu}\ ,$$

or even

$$\nabla_\mu Y^\alpha = Y^\alpha_{\ ,\mu} + Y^\nu\ \Gamma^\alpha_{\mu\nu}\ .$$

Theorem 6.3. *Eq.(6.14) is coordinate independent provided*

$$\Gamma^\alpha_{\mu\nu}(p) = \frac{\partial x^\alpha}{\partial \bar{x}^\beta}\ \frac{\partial^2 \bar{x}^\beta}{\partial x^\mu \partial x^\nu} + \bar{\Gamma}^\beta_{\sigma\lambda}(\bar{p}) \frac{\partial \bar{x}^\sigma}{\partial x^\mu} \frac{\partial \bar{x}^\lambda}{\partial x^\nu} \frac{\partial x^\alpha}{\partial \bar{x}^\beta}\ . \tag{6.15}$$

Here, $\{x^\alpha\}$ and $\{\bar{x}^\alpha\}$ are two overlapping coordinate systems such that the point p in the $\{x^\alpha\}$ coordinate system is mapped to the point $\bar{p}$ in the $\{\bar{x}^\alpha\}$ coordinate system. Clearly, $\Gamma^\alpha_{\mu\nu}$ and $\bar{\Gamma}^\alpha_{\mu\nu}$ are the connection coefficients in the two frames.

Proof. This is nothing more than a straightforward computation, which we include here for completeness. For the covariant derivative to be a truly coordinate independent operation, we must have that

$$(\nabla_{\bar{X}} \bar{Y})^\beta = \frac{\partial \bar{x}^\beta}{\partial x^\alpha}\ (\nabla_X Y)^\alpha\ , \tag{6.16}$$

where

$$\bar{X}^\mu = X^\nu\ \frac{\partial \bar{x}^\mu}{\partial x^\nu} \quad \text{and} \quad \bar{Y}^\mu = Y^\nu\ \frac{\partial \bar{x}^\mu}{\partial x^\nu}$$

which follows from the transformation property of tangent vectors. In component form, the R.H.S of eq.(6.16) gives

$$\frac{\partial \bar{x}^\beta}{\partial x^\alpha}\ (\nabla_X Y)^\alpha = (Y^\alpha_{\ ,\mu} + Y^\nu \Gamma^\alpha_{\mu\nu})\ X^\mu\ \frac{\partial \bar{x}^\beta}{\partial x^\alpha}\ .$$

Whereas, the L.H.S gives

$$(\nabla_{\bar{X}} \bar{Y})^\beta = (\bar{Y}^\beta_{\ ,\mu} + \bar{Y}^\nu \bar{\Gamma}^\beta_{\mu\nu})\ \bar{X}^\mu = X^\alpha \frac{\partial \bar{x}^\mu}{\partial x^\alpha} \left[\frac{\partial}{\partial \bar{x}^\mu} \left(Y^\theta \frac{\partial \bar{x}^\beta}{\partial x^\theta} \right) + \bar{\Gamma}^\beta_{\mu\nu}\ Y^\theta \frac{\partial \bar{x}^\nu}{\partial x^\theta} \right]$$

$$= X^\alpha\ Y^\theta_{\ ,\alpha}\ \frac{\partial \bar{x}^\beta}{\partial x^\theta} + X^\alpha\ Y^\theta\ \frac{\partial^2 \bar{x}^\beta}{\partial x^\alpha \partial x^\theta} + \bar{\Gamma}^\beta_{\mu\nu}\ X^\alpha\ Y^\theta\ \frac{\partial \bar{x}^\mu}{\partial x^\alpha}\ \frac{\partial \bar{x}^\nu}{\partial x^\theta}\ .$$

Setting the R.H.S and L.H.S equal to each other we get that

$$\Gamma^\alpha_{\mu\nu}\ Y^\nu\ X^\mu\ \frac{\partial \bar{x}^\beta}{\partial x^\alpha} = X^\alpha\ Y^\theta\ \frac{\partial^2 \bar{x}^\beta}{\partial x^\alpha \partial x^\theta} + \bar{\Gamma}^\beta_{\mu\nu}\ X^\alpha\ Y^\theta\ \frac{\partial \bar{x}^\mu}{\partial x^\alpha}\ \frac{\partial \bar{x}^\nu}{\partial x^\theta}\ .$$

Relabeling indices so that all terms have $X^\mu\ Y^\nu$ factors, and comparing coefficients of arbitrary $X^\mu\ Y^\nu$ terms we get the necessary result. □

Problem 6.10. Using the equivalence principle, we found in eq.(4.13) that, at least in GR,

$$\Gamma^{\alpha}_{\mu\lambda} = \frac{1}{2} g^{\nu\alpha}[\partial_\lambda g_{\mu\nu} + \partial_\mu g_{\nu\lambda} - \partial_\nu g_{\lambda\mu}] \, .$$

Show that the above expression for $\Gamma^{\alpha}_{\mu\lambda}$ is consistent with eq.(6.15). I.e., show that

$$\frac{1}{2} g^{a\alpha}[\partial_\nu g_{\mu a} + \partial_\mu g_{a\nu} - \partial_a g_{\nu\mu}] =$$

$$\frac{\partial x^\alpha}{\partial \bar{x}^\beta} \frac{\partial^2 \bar{x}^\beta}{\partial x^\mu \partial x^\nu} + \frac{1}{2} \bar{g}^{m\beta}[\bar{\partial}_b \bar{g}_{am} + \bar{\partial}_a \bar{g}_{mb} - \bar{\partial}_m \bar{g}_{ba}] \frac{\partial \bar{x}^a}{\partial x^\mu} \frac{\partial \bar{x}^b}{\partial x^\nu} \frac{\partial x^\alpha}{\partial \bar{x}^\beta} \, . \tag{6.17}$$

Here, $\{x^\alpha\}$ and $\{\bar{x}^\alpha\}$ are overlapping coordinate systems. Here, both Latin and Greek indices run through the same set, also

$$\bar{\partial}_b \equiv \frac{\partial}{\partial \bar{x}^b} \, .$$

In the above problem, we have given one (very important in GR) candidate for $\Gamma^{\alpha}_{\mu\lambda}$. That should not be taken to mean that, in general, the existence of a covariant derivative requires a metric, or is even dependent on it.

If it weren't for the first term in the R.H.S of eq.(6.15), the Christoffel symbols would transform like a tensor. Indeed, it is the existence of this term that allows for the Christoffel symbols to vanish at any given point in a locally inertial frame, while in an arbitrary coordinate system that may not be the case (which cannot happen if the Christoffel symbols were a true tensor), thus allowing for the equivalence principle to exist. Once we know how to take the derivative of vector fields, property 4. fixes the covariant derivative of dual vector fields.

Theorem 6.4. *The properties of the covariant derivative imply that*

$$\nabla_\mu \chi_\nu = \chi_{\nu,\mu} - \Gamma^{\alpha}_{\mu\nu} \chi_\alpha \, , \tag{6.18}$$

where χ is any dual vector field, and $\nabla_\mu \chi_\nu = (\nabla_\mu \chi)_\nu$.

Proof. Let Y be any vector field. From property 2. of covariant derivatives

$$\nabla_\mu \ (\chi_\alpha Y^\alpha) = \chi_{\alpha,\mu} \, Y^\alpha + \chi_\alpha \, Y^\alpha_{,\mu} \, .$$

On the other hand from property 4. of covariant derivatives

$$\nabla_\mu \ (\chi_\alpha Y^\alpha) = (\nabla_\mu \, \chi_\alpha) Y^\alpha + \chi_\alpha \, \nabla_\mu Y^\alpha$$

$$= (\nabla_\mu \, \chi_\alpha) Y^\alpha + \chi_\alpha \left[Y^\alpha_{,\mu} + \Gamma^{\alpha}_{\mu\nu} Y^\nu \right] \, .$$

Therefore, comparing the two expressions on the right-hand side, we get that

$$\chi_{\alpha,\mu} \, Y^\alpha + \chi_\alpha \, Y^\alpha_{,\mu} = (\nabla_\mu \, \chi_\alpha) Y^\alpha + \chi_\alpha \left[Y^\alpha_{,\mu} + \Gamma^{\alpha}_{\mu\nu} Y^\nu \right] \, ,$$

or

$$(\nabla_\mu \, \chi_\alpha) Y^\alpha = \chi_{\alpha,\mu} \, Y^\alpha - \chi_\nu \Gamma^{\nu}_{\mu\alpha} Y^\alpha \, .$$

Since this is true for any vector field Y, we get the necessary result. □

The theorem above implies that

$$(\nabla_\mu dx^\nu) = -\Gamma^{\nu}_{\mu\alpha} dx^\alpha \, ,$$

so that

$$\nabla_\mu \chi = (\chi_{\nu,\mu} - \Gamma^{\alpha}_{\mu\nu} \chi_\alpha) dx^\nu \, .$$

A similar calculation as before shows that when the Christoffel symbols transforms as in eq.(6.15), the resulting covariant derivative of 1-forms are coordinate independent. Now that we can take derivatives of tangent vector fields and 1-forms, we can easily extend the covariant derivative to arbitrary tensors, for example,

$$\nabla_\mu \left[S^\alpha{}_\beta \, \partial_\alpha \otimes \, dx^\beta \right]$$

$$= (\partial_\mu S^\alpha{}_\beta) \, \partial_\alpha \otimes dx^\beta + S^\alpha{}_\beta \, (\nabla_\mu \partial_\alpha) \otimes \, dx^\beta + S^\alpha{}_\beta \, \partial_\alpha \otimes \, (\nabla_\mu dx^\beta)$$

$$= (\partial_\mu S^\alpha{}_\beta)\,\partial_\alpha \otimes dx^\beta + S^\alpha{}_\beta\,\Gamma^\nu_{\mu\alpha}\,\partial_\nu \otimes dx^\beta - S^\alpha{}_\beta\,\Gamma^\beta_{\mu\nu}\,\partial_\alpha \otimes dx^\nu$$

$$= \Big(\partial_\mu S^\alpha{}_\beta + S^\nu{}_\beta\,\Gamma^\alpha_{\mu\nu} - S^\alpha{}_\nu\Gamma^\nu_{\mu\beta}\Big)\partial_\alpha \otimes dx^\beta\ .$$

Or equivalently

$$\nabla_\mu S^\alpha{}_\beta = \partial_\mu S^\alpha{}_\beta + S^\nu{}_\beta\,\Gamma^\alpha_{\mu\nu} - S^\alpha{}_\nu\Gamma^\nu_{\mu\beta}\ .$$

Considering problem 6.10, let us set the Christoffel symbols to be given by eq.(4.13) in the following set of problems. Furthermore, all tensors are given in Schwarzschild spacetime in Schwarzschild coordinates.

Problem 6.11. Let

$$X = 2\sin\theta\,\partial_{ct} - \partial_\theta \quad \text{and} \quad Y = e^{-(ct)^2}\partial_\phi$$

in Schwarzschild spacetime, in Schwarzschild coordinates. Find

1. $\nabla_X Y$
2. $\nabla_Y X$
3. $[X, Y]$
4. $\nabla_X Y - \nabla_Y X - [X, Y]$.

Problem 6.12. Let $\chi = -dct + 3M\sin^2\theta\partial_\theta + \left(1 - \frac{r^*}{r}\right)dr$. Find $\nabla_{\partial_r}\chi$.

Problem 6.13. Let

$$F = \sin^4\theta\big(l^\flat \otimes d\theta - d\theta \otimes l^\flat\big)\ .$$

where

$$l^\flat = cdt + \left(1 - \frac{2M}{c^2 r}\right)^{-1} dr\ .$$

Find

1. $j^\nu = \nabla_\mu F^{\mu\nu}$
2. $g(j, j)$
3. $j_\mu F^{\mu\nu}$.

6.4 Parallel Transport and Geodesics

Definition 6.4. *Let V be a vector field about a point $p \in \mathcal{M}$. Suppose $\alpha(s) = (x^1(s), \ldots, x^n(s))$ is a curve in $\mathcal{M}$ such that $\alpha(0) = p$, and*

$$\dot\alpha(s) = \left(\frac{dx^1}{ds}(s), \ldots, \frac{dx^n}{ds}(s)\right) = V(\alpha(s))\ ,$$

we say that α is an integral curve of V with initial value p.

Theorem 6.5. *Let $V \in T_p(\mathcal{M})$ for some $p \in \mathcal{M}$. Then for any vector field W, $\nabla_V W|_p$ depends only on $V(p)$ and the values of W along any curve α such that $\dot\alpha(0) = V(p)$.*

Proof. In a coordinate chart $\{x^a\}$, let $\alpha(s) = (x^1(s), \ldots, x^n(s))$ be any curve such that $\alpha(0) = p$ and $V = \dot\alpha|_{s=0}$ Then, at p,

$$\nabla_V W^\beta\Big|_p = V(W^\beta) + \Gamma^\beta_{\mu\nu}V^\mu W^\nu\Big|_p = \frac{dx^\sigma}{ds}\Big|_{s=0} W^\beta_{,\sigma}(p) + \Gamma^\beta_{\mu\nu}(p)V^\mu(p)W^\nu(p)$$

$$= \frac{d}{ds}W^\beta(\alpha(s))\Big|_{s=0} + \Gamma^\beta_{\mu\nu}(p)V^\mu(p)W^\nu(p)\ .$$

It should also be clear that the answer does not depend on the choice of α since

$$\frac{dx^\sigma}{ds}\Big|_{s=0} W^\beta_{,\sigma}(p) = V^\sigma(p)W^\beta_{,\sigma}(p)\ .$$

□

Definition 6.5. *Suppose $\alpha(s) = (x^1(s), \ldots, x^n(s))$ is a curve in $\mathcal{M}$. Let V be a vector field defined along α. I.e., $V = V(\alpha(s)) \equiv V(s)$. Then the covariant derivative of V along α is defined by*

$$\frac{D}{\partial s}V^\beta = \frac{dV^\beta}{ds} + \Gamma^\beta_{\mu\nu}\,\frac{dx^\mu}{ds}\,V^\nu \tag{6.19}$$

Remark 6.2. Clearly, in the event that V is defined in a region around α, we have that

$$\frac{D}{\partial s}V = \nabla_{\dot\alpha}V\ .$$

Definition 6.6. *Consider a manifold with a given covariant derivative. Let $\alpha(s) = (x^1(s), ..., x^n(s))$ be a curve, and let v_0 be a tangent vector at some point $\alpha(s_0)$ on the curve. The vector field $v(s)$ along α is said to be parallel translated along α with initial condition v_0 if*

$$v^\nu(s) = v_0^\nu - \int_{s_0}^{s} \Gamma^\nu_{\rho\sigma}(s')v^\sigma(s')\frac{dx^\rho}{ds'}ds'\ , \tag{6.20}$$

for all $\nu = 1, \ldots, n$.

Theorem 6.6. *$v(s)$ above is parallely translated along α if and only if $\frac{D}{\partial s}v^\nu = 0$ for all s.*

Proof. From (6.20),

$$\frac{D}{\partial s}v^\nu = \frac{dv^\nu}{ds} + \Gamma^\nu_{\mu\sigma}(s)v^\sigma(s)\frac{dx^\mu}{ds} = \Gamma^\nu_{\mu\sigma}(s)v^\sigma(s)\frac{dx^\mu}{ds} - \Gamma^\nu_{\mu\sigma}(s)v^\sigma(s)\frac{dx^\mu}{ds} = 0\ .$$

□

Definition 6.7. *Let $\alpha(s) = (x^1(s), ..., x^n(s))$ be a curve with tangent u, i.e., $u = \dot\alpha$. Then, α is a geodesic if*

$$\frac{D}{\partial s}u = 0. \tag{6.21}$$

We say that u is a geodesic tangent vector field. From eq.(6.19), in components, the geodesic equation takes the form

$$\ddot x^\mu + \Gamma^\mu_{\alpha\beta}\dot x^\alpha \dot x^\beta = 0\ . \tag{6.22}$$

In the above equation, the dot refers to the derivative with respect to s. The parameter s is referred to as an **affine parameter**.

Definition 6.8. *A curve $\alpha(s)$ is a pre-geodesic, if α can be re-parameterized by a parameter τ such that $\alpha(\tau)$ is a geodesic.*

Theorem 6.7. *Let $\alpha(s)$ be an intergral curve of a vector field u, i.e., $\dot\alpha(s) = u(\alpha(s))$. Then α is a pre-geodesic (and hence u is a pre-geodesic tangent vector field), if and only if there exists a function f on $\alpha(s)$ such that*

$$\frac{D}{\partial s}u\,\Big|_s = f(s)\ u(s)\ . \tag{6.23}$$

Proof. Along $\alpha(s) = (x^1(s), ..., x^n(s))$, we have

$$\frac{d^2}{ds^2}x^\lambda + \Gamma^\lambda_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = f(s)\ \frac{dx^\lambda}{ds}$$

For any re-parametrization τ of the curve, we have

$$\frac{d}{ds} = \frac{d\tau}{ds}\ \frac{d}{d\tau}$$

and

$$\frac{d^2}{ds^2} = \frac{d}{ds}\left[\frac{d\tau}{ds}\,\frac{d}{d\tau}\right] = \frac{d^2\tau}{ds^2}\,\frac{d}{d\tau} + \left[\frac{d\tau}{ds}\right]^2\,\frac{d^2}{d\tau^2}\,.$$

Then, along $\alpha(\tau)$

$$\frac{d^2\tau}{ds^2}\,\frac{dx^\lambda}{d\tau} + \left(\frac{d\tau}{ds}\right)^2\,\left[\frac{d^2}{d\tau^2}x^\lambda + \Gamma^\lambda_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}\right] = f(s)\,\frac{d\tau}{ds}\,\frac{dx^\lambda}{d\tau}\,.$$

or

$$\left(\frac{d\tau}{ds}\right)^2\,\left[\frac{d^2}{d\tau^2}x^\lambda + \Gamma^\lambda_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}\right] = \left(f(s)\,\frac{d\tau}{ds} - \frac{d^2\tau}{ds^2}\right)\,\frac{dx^\lambda}{d\tau}\,. \tag{6.24}$$

This is an affinely parameterized geodesic equation if

$$f(s)\,\frac{d\tau}{ds} = \frac{d^2\tau}{ds^2}\,. \tag{6.25}$$

You will show in the problems section that the unique solution to the above equation is

$$\tau = \int^s ds''\,\left[\exp\int^{s''} f(s')ds'\right]\,. \tag{6.26}$$

Incidentally, notice that

$$\frac{d\tau}{ds} = \exp\int^s f(s')\,ds' > 0\,, \tag{6.27}$$

and this ensures that s and τ are invertible parameters. Therefore, the affine parameter $\tau(s)$ is well defined. Now, $\alpha(\tau)$ satisfies the geodesics equation. □

Problem 6.14. Show that the unique solution to eq.(6.25) is given by eq.(6.27).

Problem 6.15. Show that if s and τ are two affine parameters for the same geodesic, then $\tau = as + b$. Here, a and b are constants. Hint: In eq.(6.25), set $f(s) = 0$.

Problem 6.16. Verify that in Schwarzschild coordinates (ct, r, θ, ϕ),

$$T = \left(1 - \frac{r^*}{r}\right)^{-1}\partial_{ct} + \frac{\sqrt{r^* r}}{r}\partial_r$$

is a geodesic vector field. I.e., show that $\nabla_T T = 0$ (see remark 6.2). Also find $g(T,T)$. These are the rain-geodesics of the Schwarzschild geometry.

6.5 Torsion Tensor

From eq.(6.15) we see that

$$\Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\nu\mu} = \left[\bar\Gamma^\beta_{ab} - \bar\Gamma^\beta_{ba}\right]\frac{\partial\bar x^a}{\partial x^\mu}\frac{\partial\bar x^b}{\partial x^\nu}\frac{\partial x^\alpha}{\partial\bar x^\beta}\,.$$

I.e., the quantity $\Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\nu\mu}$ transforms like a type $T^1_2(\mathcal{M})$ tensor. Clearly,

$$\Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\nu\mu} = (\nabla_\mu\partial_\nu - \nabla_\nu\partial_\mu)^\alpha\,.$$

Indeed, in a coordinate chart, the components of the torsion tensor are given by

$$T^\alpha{}_{\mu\nu} = (\nabla_\mu\partial_\nu - \nabla_\nu\partial_\mu)^\alpha\,. \tag{6.28}$$

Or, equivalently

$$T(\chi, X, Y) = \chi_\alpha\,X^\mu Y^\nu(\Gamma^\alpha_{\mu\nu} - \Gamma^\alpha_{\nu\mu}) = \chi_\alpha\,\Gamma^\alpha_{\mu\nu}\,(X^\mu Y^\nu - Y^\mu X^\nu)\,.$$

So, we may tempted to set the the action of the torsion tensor T on any $X, Y \in T(\mathcal{M})$ and $\chi \in T^*(\mathcal{M})$ as

$$T(\chi, X, Y) \stackrel{?}{=} \chi(\nabla_X Y - \nabla_Y X)\,.$$

But, it is easily verified that the above expression is not function linear in all its entries [9]. This is easily remedied by the following definition.

Definition 6.9. *The torsion tensor T is of type $T^1_2(\mathcal{M})$ and is given by*

$$T(\chi, X, Y) = \chi(\nabla_X Y - \nabla_Y X - [X, Y]) \; . \tag{6.29}$$

Here X and Y are tangent vector fields, and χ is a dual vector field.

Note that this definition is consistent with all the requirement listed above. In particular, the Lie bracket term does not show up in the expression for the components of the torsion tensor in a coordinate chart (eq.(6.28)) since $[\partial_\mu, \partial_\nu] = 0$. T is a true tensor, i.e., its a function linear object in all its entries. For example,

$$T(\chi, fX, Y) = \chi \left(\nabla_{fX} Y - \nabla_Y fX - [fX, Y] \right)$$
$$= \chi \left(f\nabla_X Y - Y(f)X - f\nabla_Y X - f[X, Y] + Y(f)X \right) = f \; T(\chi, X, Y) \; .$$

6.6 The Levi-Civita Connection

A Manifold may admit many covariant derivatives. The properties of a covariant derivative listed above do not, by any means fix the connection coefficients. However, there is a preferred connection (in General Relativity) if the manifold is endowed with a metric.

Theorem 6.8. *Let g be a metric tensor on a manifold $\mathcal{M}$. Then there exists a unique covariant derivative on $\mathcal{M}$ such that*

1. $\nabla_X g = 0$, for every $X \in T(\mathcal{M})$,
2. *the torsion tensor $T = 0$.*

This is the Levi-Civita Connection.

Proof. Since $T = 0$, we have that [10]

$$\nabla_{\partial_\mu} \partial_\nu = \nabla_{\partial_\nu} \partial_\mu \; ,$$

or,

$$\Gamma^\alpha_{\mu\nu} = \Gamma^\alpha_{\nu\mu} \; .$$

This is the feature we need from our torsion-free requirement. Now, let us consider the implications of the requirement that the covariant derivative of the metric be trivial. Using the product rule and eq.(6.18)

$$\nabla_X g_{\mu\nu} \; dx^\mu \otimes dx^\nu$$
$$= X(g_{\mu\nu}) \; dx^\mu \otimes dx^\nu - g_{\mu\nu} X^\beta \Gamma^\mu_{\alpha\beta} \; dx^\alpha \otimes dx^\nu - g_{\mu\nu} X^\beta \Gamma^\nu_{\alpha\beta} \; dx^\mu \otimes dx^\alpha = 0 \; . \tag{6.30}$$

After changing dummy variables so that all tensor products in the above equations are of the form $dx^\mu \otimes dx^\nu$, and setting $X = \partial_\lambda$ we get

$$\partial_\lambda g_{\mu\nu} = g_{\alpha\nu} \Gamma^\alpha_{\lambda\mu} + g_{\alpha\mu} \Gamma^\alpha_{\lambda\nu} \; . \tag{6.31}$$

Let us rewrite the above equation by permutations of subscripts:

$$\partial_\mu g_{\nu\lambda} = g_{\alpha\lambda} \Gamma^\alpha_{\mu\nu} + g_{\alpha\nu} \Gamma^\alpha_{\mu\lambda} \; , \tag{6.32}$$

and

$$\partial_\nu g_{\lambda\mu} = g_{\alpha\mu} \Gamma^\alpha_{\nu\lambda} + g_{\alpha\lambda} \Gamma^\alpha_{\nu\mu} \; . \tag{6.33}$$

Then eq.(6.31)+ eq.(6.32) − eq.(6.33) gives

$$\Gamma^\alpha_{\mu\lambda} = \frac{1}{2} g^{\nu\alpha} [\partial_\lambda g_{\mu\nu} + \partial_\mu g_{\nu\lambda} - \partial_\nu g_{\lambda\mu}] \; . \tag{6.34}$$

We have already shown (problem (6.10)) that eq.(6.34) generates a well-defined connection. We have the result we want. The connection coefficients are completely specified, and hence, it is unique. □

[9] For $f, g \in C^\infty(\mathcal{M})$, if T is defined as above, then

$$T(\chi, fX, gY) \neq fg \; T(\chi, X, Y) \; .$$

[10] $[\partial_\mu, \partial_\nu] = 0$ for coordinate vector fields.

Theorem 6.9. *Let $\mathcal{M}$ be a manifold with a given metric g, and let ∇ be the associated Levi-Civita connection. Let $\alpha(\tau)$ be a geodesic with velocity $\dot{\alpha} = u$, then $g(\dot{\alpha}, \dot{\alpha})$ is a constant along α.*

Proof.

$$\frac{d}{d\tau}\, g(\dot{\alpha}, \dot{\alpha}) = \frac{dx^i}{d\tau}\, \frac{\partial}{\partial x^i}\, (g(u,u)) = u(g(u,u)) \ .$$

On the other hand

$$u(g(u,u)) = \nabla_u g(u,u) = (\nabla_u g)(u,u) + g(\nabla_u u, u) + g(u, \nabla_u u) \ .$$

The first equality above is true, since for a function f, $\nabla_u f \equiv u[f]$. Also, since we are using the Levi-Civita connection, $\nabla_u g = 0$, and since u is a geodesic tangent vector field, we have $\nabla_u u = 0$. □

The above theorem implies that the speed of a geodesic is a constant. This is consistent with special relativity, in which the 4-velocity U has a constant magnitude given by $U^2 = g(U,U) = -c^2$.

Problem 6.17. Let $\mathcal{M}$ be a manifold with a given metric g, and let ∇ be the associated Levi-Civita connection. Let $\alpha(\tau)$ be a geodesic with velocity $\dot{\alpha} = u$. Let v be a vector field that is parallel translated along α, then show that $g(v, \dot{\alpha})$ is a constant along α.

Problem 6.18. Show that the torsion tensor T is function-linear in all the entries.

In the following, let $M = \mathbb{R}^3$ with the Euclidean metric

$$g = dx \otimes dx + dy \otimes dy + dz \otimes dz$$

in the standard cartesian coordinates (x, y, z).

Problem 6.19. Show that the Levi-Civita connection coefficients are trivial (i.e., $\Gamma^{\mu}_{\alpha\beta} = 0$).

Problem 6.20. Write down the general equation for a geodesic and solve it (should be done with relative ease). Note that in $\mathbb{R}^3$, geodesics are just straight lines.

The following problems refer to $S^2(r)$ (the sphere of radius r):

Problem 6.21. In the previous chapter, we found the Euclidean metric g of $\mathbb{R}^3$ in the spherical coordinate system. By setting $dr = 0$, write down the metric of $S^2(r)$ in the coordinate system (θ, φ). Clearly, here the poles are excluded.

Problem 6.22. Find the connection coefficients $\Gamma^{\mu}_{\alpha\beta}$ of $S^2(r)$.

Problem 6.23. Great circles on a sphere:

1. Write down the geodesic equation on $S^2(r)$. I.e., let $\alpha(s) = (\theta(s), \varphi(s))$ be a curve in $S^2(r)$. Find the equations that $\theta(s)$ and $\varphi(s)$ must satisfy, if $\alpha(s)$ is to be a geodesic.
2. Show that $c \equiv g(\dot{\alpha}(s), \partial_\varphi)$ is a constant along α.
3. Show that $c \leq 1$.
4. Using the above equations and fact that $g(\dot{\alpha}(s), \dot{\alpha}(s)) = \text{const} = 1$, show that along a geodesic

$$\cos\theta = \frac{\sqrt{1-c^2}}{c}\, (\cos\varphi \sin\theta \cos a + \sin\varphi \sin\theta \sin a)$$

   for some constant a.
5. Multiplying the above equation by r, the radius of the circle, argue that the geodesics on $S^2(r)$ are great circles.
6. Show that $\alpha(s) = (\pi/2, s/r)$, $0 \leq s \leq 2/\pi$ is a geodesic.

Problem 6.24. Consider the Meridians on $S^2(r)$, i.e., let $\alpha(s) = (\theta_0, s)$ for any fixed value of θ be a curve in $S^2(r)$.

1. Show that for $\theta_0 \neq \pi/2$, α is not a pre-geodesic (hence not a geodesic).
2. Let $p = (0, 0) \in S^2(r)$. Let $v_p \in T_p(S^2(r))$ be given by

$$v_p = \partial_\theta - \partial_\varphi \, .$$

Find $v(s) \in T_{\alpha(s)}(S^2(r))$ such that $v(s)$ parallely translates v_p along $\alpha(s)$. I.e., solve

$$\frac{dv^\alpha}{d\tau} + \Gamma^\alpha_{\mu\sigma}(\tau) v^\sigma(\tau) \frac{dx^\mu}{d\tau} = 0$$

along with the initial condition $v(0) = v_p$.

7

Gravitation and Curvature

7.1 The Riemann Curvature Tensor

Like the Torsion tensor, the Riemann Curvature Tensor will also be extracted from the transformation properties of the Christoffel symbols (eq.(6.15)). But this time, we will have an additional derivative involved in our calculations. The Riemann Curvature Tensor will be defined for an arbitrary covariant derivative. I.e., for the moment, we will not insist on the Levi-Civita connection. However, the full symmetry of the curvature tensor will only be revealed for Levi-Civita connections.

Definition 7.1. *The Riemann Curvature Tensor is of type $T^1_3(\mathcal{M})$ and is defined by*

$$R(X,Y,Z,w) = [\nabla_X \nabla_Y w - \nabla_Y \nabla_X w - \nabla_{[X,Y]} w](Z) \tag{7.1}$$

Here, X, Y and Z are tangent vector fields and w is a 1 form.

Problem 7.1. Show that eq.(7.1) defines a true tensor. I.e., it is multi-function linear at all entries. For example, show that

$$R(f\ W + g\ X, Y, Z, w) = f\ R(W,Y,Z,w) + g\ R(X,Y,Z,w)$$

Here, W, X, Y and Z are tangent vector fields and w is a 1 form. Now show the analogous result for all entries.

Theorem 7.1. *In component form, the Riemann tensor is given by*

$$R_{\beta\alpha\mu}{}^{\nu} = \left[\partial_\alpha \Gamma^\nu_{\beta\mu} - \partial_\beta \Gamma^\nu_{\alpha\mu} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\mu} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\mu}\right] . \tag{7.2}$$

Proof. First note that

$$\nabla_Y \nabla_X w_\theta = Y^\alpha \left[X^\beta\ w_{\theta,\beta} - w_\nu\ X^\beta\ \Gamma^\nu_{\beta\theta}\right]_{,\alpha} - Y^\alpha \left[X^\beta\ w_{c,\beta} - w_\nu\ X^\beta\ \Gamma^\nu_{\beta c}\right]\ \Gamma^c_{\alpha\theta}\ .$$

Therefore,

$$\nabla_X \nabla_Y w_\theta - \nabla_Y \nabla_X w_\theta = Y^\alpha w_\nu X^\beta \left[\partial_\alpha \Gamma^\nu_{\beta\theta} - \partial_\beta \Gamma^\nu_{\alpha\theta} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\theta} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\theta}\right]$$

$$+ w_{\theta,\beta}\left[X^\alpha Y^\beta_{,\alpha} - Y^\alpha X^\beta_{,\alpha}\right] + w_\nu \Gamma^\nu_{\beta\theta}\left[Y^\alpha X^\beta_{,\alpha} - X^\alpha Y^\beta_{,\alpha}\right] .$$

$$= Y^\alpha w_\nu X^\beta \left[\partial_\alpha \Gamma^\nu_{\beta\theta} - \partial_\beta \Gamma^\nu_{\alpha\theta} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\theta} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\theta}\right] + \nabla_{[X,Y]} w_\theta\ .$$

Then

$$\nabla_X \nabla_Y w_\theta - \nabla_Y \nabla_X w_\theta - \nabla_{[X,Y]} w_\theta = Y^\alpha w_\nu X^\beta \left[\partial_\alpha \Gamma^\nu_{\beta\theta} - \partial_\beta \Gamma^\nu_{\alpha\theta} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\theta} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\theta}\right] ,$$

or

$$R_{\beta\alpha\mu}{}^{\nu}\ X^\beta\ Y^\alpha\ Z^\mu\ w_\nu \equiv R(X,Y,Z,w) = \left[\nabla_X \nabla_Y w_\theta - \nabla_Y \nabla_X w_\theta - \nabla_{[X,Y]} w_\theta\right] Z^\theta$$

$$= Y^\alpha w_\nu X^\beta Z^\theta \left[\partial_\alpha \Gamma^\nu_{\beta\theta} - \partial_\beta \Gamma^\nu_{\alpha\theta} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\theta} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\theta}\right] = Y^\alpha w_\nu X^\beta Z^\mu \left[\partial_\alpha \Gamma^\nu_{\beta\mu} - \partial_\beta \Gamma^\nu_{\alpha\mu} + \Gamma^\nu_{\alpha c}\Gamma^c_{\beta\mu} - \Gamma^\nu_{\beta c}\Gamma^c_{\alpha\mu}\right] .$$

Since this is true for every $Y^\alpha, w_\nu, X^\beta, Z^\mu$ we have the necessary result. □

Theorem 7.2. *In component form eq.(7.1) becomes*

$$\nabla_\beta \nabla_\alpha w_\theta - \nabla_\alpha \nabla_\beta w_\theta = R_{\beta\alpha\theta}{}^{a} w_a \ . \tag{7.3}$$

Proof. From the above theorem

$$\nabla_X \nabla_Y w_\theta - \nabla_Y \nabla_X w_\theta = Y^\alpha w_a X^\beta R_{\beta\alpha\theta}{}^{a} + \nabla_{[X,Y]} w_\theta \ .$$

When X and Y are some coordinate bases vector fields ∂_β and ∂_α, clearly the last Lie bracket term disappears, and

$$\nabla_\beta \nabla_\alpha w_\theta - \nabla_\alpha \nabla_\beta w_\theta = w_a R_{\beta\alpha\theta}{}^{a} \ .$$

Here, it's important to remember that

$$\nabla_\beta \nabla_\alpha w_\theta = (\nabla_\beta \nabla_\alpha w)_\theta \ .$$

□

In component form eq.(7.1) becomes

$$R_{\beta\alpha\mu}{}^{\nu} \, X^\beta \, Y^\alpha \, Z^\mu \, w_\nu = X^\beta Y^\alpha [\nabla_\beta \nabla_\alpha w_\nu - \nabla_\alpha \nabla_\beta w_\nu] Z^\nu \ . \tag{7.4}$$

The result follows immediately from the previous theorem. However, there is a hidden subtlety. The point here is that

$$\nabla_\beta \nabla_\alpha w_\nu - \nabla_\alpha \nabla_\beta w_\nu$$

acts like a tensor of type $T^0_3(\mathcal{M})$. It is a linear map! Problem 7.1 conclusively proves that eq.(7.1) defines a tensor. But if you are still not convinced by the abstract argument, the following calculation should clinch the deal.

Theorem 7.3. *The following object transforms like a $T^1_3(\mathcal{M})$ tensor:*

$$\frac{\partial}{\partial x^\nu}\Gamma^\sigma_{\mu\rho} - \frac{\partial}{\partial x^\mu}\Gamma^\sigma_{\nu\rho} + \Gamma^\alpha_{\mu\rho}\Gamma^\sigma_{\alpha\nu} - \Gamma^\alpha_{\nu\rho}\Gamma^\sigma_{\alpha\mu} \ . \tag{7.5}$$

Proof. I have used both Latin and Greek indices in the calculations below. All indices run through all the coordinates of the spacetime (/manifold). Equation (6.15) can be rewritten as

$$\frac{\partial^2 \bar{x}^\beta}{\partial x^\mu \partial x^\rho} = \frac{\partial \bar{x}^\beta}{\partial x^a}\Gamma^a_{\mu\rho} - \bar{\Gamma}^\beta_{cd}\frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial \bar{x}^d}{\partial x^\rho} \ . \tag{7.6}$$

Therefore,

$$\frac{\partial}{\partial x^\nu}\frac{\partial^2 \bar{x}^\beta}{\partial x^\mu \partial x^\rho} = \Gamma^a_{\mu\rho}\frac{\partial}{\partial x^\nu}\frac{\partial \bar{x}^\beta}{\partial x^a} + \frac{\partial \bar{x}^\beta}{\partial x^a}\frac{\partial}{\partial x^\nu}\Gamma^a_{\mu\rho} - \frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial \bar{x}^d}{\partial x^\rho}\frac{\partial}{\partial x^\nu}\bar{\Gamma}^\beta_{cd} - \bar{\Gamma}^\beta_{cd}\frac{\partial \bar{x}^d}{\partial x^\rho}\frac{\partial}{\partial x^\nu}\frac{\partial \bar{x}^c}{\partial x^\mu} - \bar{\Gamma}^\beta_{cd}\frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial}{\partial x^\nu}\frac{\partial \bar{x}^d}{\partial x^\rho} \ .$$

Now for every term of type

$$\frac{\partial}{\partial x^\nu}\frac{\partial \bar{x}^\beta}{\partial x^a}$$

in the above equation, we can use eq.(7.6) to compute the second derivatives. Then

$$0 = \frac{\partial}{\partial x^\nu}\frac{\partial^2 \bar{x}^\beta}{\partial x^\mu \partial x^\rho} - \frac{\partial}{\partial x^\mu}\frac{\partial^2 \bar{x}^\beta}{\partial x^\nu \partial x^\rho} = \frac{\partial \bar{x}^\beta}{\partial x^a}\left[\partial_\nu \Gamma^a_{\mu\rho} - \partial_\mu \Gamma^a_{\nu\rho}\right] + \frac{\partial \bar{x}^\beta}{\partial x^d}\left[\Gamma^a_{\mu\rho}\Gamma^d_{\nu a} - \Gamma^a_{\nu\rho}\Gamma^d_{\mu a}\right] + \frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^n}{\partial x^\rho}\partial_\mu \bar{\Gamma}^\beta_{mn}$$

$$-\frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial \bar{x}^d}{\partial x^\rho}\partial_\nu \bar{\Gamma}^\beta_{cd} - \frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^\theta}{\partial x^\mu}\frac{\partial \bar{x}^\pi}{\partial x^\rho}\bar{\Gamma}^\beta_{mn}\bar{\Gamma}^n_{\theta\pi} + \frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^n}{\partial x^\rho}\bar{\Gamma}^\beta_{cd}\bar{\Gamma}^d_{mn}$$

$$-\Gamma^a_{\mu\rho}\bar{\Gamma}^\beta_{mn}\frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^n}{\partial x^a} + \Gamma^l_{\mu\rho}\bar{\Gamma}^\beta_{mn}\frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^n}{\partial x^l} + \Gamma^a_{\nu\rho}\bar{\Gamma}^\beta_{mn}\frac{\partial \bar{x}^m}{\partial x^\mu}\frac{\partial \bar{x}^n}{\partial x^a} - \Gamma^a_{\nu\rho}\bar{\Gamma}^\beta_{cd}\frac{\partial \bar{x}^c}{\partial x^\mu}\frac{\partial \bar{x}^d}{\partial x^a} \ .$$

The last two lines in the R.H.S of the above equation cancels itself out. Finally, we get

$$0 = \left[\frac{\partial}{\partial x^\nu}\Gamma^l_{\mu\rho} - \frac{\partial}{\partial x^\mu}\Gamma^l_{\nu\rho} + \Gamma^a_{\mu\rho}\Gamma^l_{\nu a} - \Gamma^a_{\nu\rho}\Gamma^l_{\mu a}\right] - \frac{\partial x^l}{\partial \bar{x}^\beta}\frac{\partial \bar{x}^m}{\partial x^\nu}\frac{\partial \bar{x}^n}{\partial x^\rho}\frac{\partial \bar{x}^c}{\partial x^\mu}\left[\frac{\partial}{\partial \bar{x}^c}\bar{\Gamma}^\beta_{mn} - \frac{\partial}{\partial \bar{x}^m}\bar{\Gamma}^\beta_{cn} + \bar{\Gamma}^\beta_{cd}\bar{\Gamma}^d_{mn} - \bar{\Gamma}^\beta_{m\alpha}\bar{\Gamma}^\alpha_{cn}\right] \ .$$

This is the desired result. □

Theorem 7.4.

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)(T\otimes S) = [(\nabla_a\nabla_b - \nabla_b\nabla_a)T]\otimes S + T\otimes(\nabla_a\nabla_b - \nabla_b\nabla_a)S\ .$$

Here T and S are any tensor fields.

Proof.

$$\nabla_b(T\otimes S) = (\nabla_b T)\otimes S + T\otimes\nabla_b S\ .$$

Therefore,

$$\nabla_a\nabla_b(T\otimes S) = (\nabla_a\nabla_b T)\otimes S + \nabla_b T\otimes\nabla_a S + \nabla_a T\otimes\nabla_b S + T\otimes\nabla_a\nabla_b S\ .$$

Clearly, the middle two terms in the R.H.S of the above equation will not contribute in an anti-symmetric combination. □

Problem 7.2. For a torsion-free connection, for any smooth function f, show that

$$\nabla_{[a}\nabla_{b]}f = 0\ .$$

Theorem 7.5. *For a torsion-free connection,*

$$(\nabla_a\nabla_b V - \nabla_b\nabla_a V)^d = -R_{abc}{}^d\, V^c\ . \tag{7.7}$$

Here, $V\in T(\mathcal{M})$.

Proof. For a torsion-free connection, $\nabla_{[a}\nabla_{b]}(V^c w_c) = 0$ for any $w\in T^*(M)$. Then

$$0 = V^c(\nabla_a\nabla_b w_c - \nabla_b\nabla_a w_c) + w_c(\nabla_a\nabla_b V - \nabla_b\nabla_a V)^c = V^c R_{abc}{}^d w_d + w_c(\nabla_a\nabla_b V - \nabla_b\nabla_a V)^c\ .$$

Or

$$V^d R_{abd}{}^c w_c = -w_c(\nabla_a\nabla_b V - \nabla_b\nabla_a V)^c\ .$$

□

Corollary 7.1. *For a torsion-free connection,*

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\, T^{m_1 m_2 \ldots m_k}{}_{n_1 n_2 \ldots n_l} = -\sum_{i=1}^{k} R_{abp}{}^{m_i}\, T^{m_1\ldots p\ldots m_k}{}_{n_1\ldots n_l} + \sum_{j=1}^{l} R_{abn_j}{}^{p}\, T^{m_1\ldots m_k}{}_{n_1\ldots p\ldots n_l}\ . \tag{7.8}$$

Here p would be placed in the i^{th} and j^{th} entry of the first and second term of the R.H.S respectively.

Proof. Theorem 7.4 along with eqs.(7.3) and (7.7) gives an immediate proof by induction. □

7.2 Symmetries of the Riemann Tensor

Theorem 7.6. Property 1 *Clearly $R(X,Y,Z,w) = -R(Y,X,Z,w)$.*

Theorem 7.7. Property 2 *For the Riemann tensor stemming from a Torsion free connection*

$$R_{[abc]}{}^d = 0\ . \tag{7.9}$$

Here,

$$R_{[abc]}{}^d = \frac{1}{6}\left(R_{abc}{}^d - R_{acb}{}^d - R_{bac}{}^d + R_{bca}{}^d + R_{cab}{}^d - R_{cba}{}^d\right)\ .$$

Proof. For $w \in T^*(\mathcal{M})$,

$$\nabla_a \nabla_b w_c = (w_{c,b} - w_d \Gamma^d_{bc})_{,a} - (w_{p,b} - w_d \Gamma^d_{bp})\Gamma^p_{ac} \ .$$

Then

$$\nabla_{[a} \nabla_b w_{c]} = w_{[c,b,a]} - w_{d,[a}\Gamma^d_{bc]} - w_d \Gamma^d_{[bc,a]} - w_{p,[b}\Gamma^p_{ac]} + w_d \Gamma^d_{p\,[b}\Gamma^p_{ac]} \ .$$

In the last term Γ^d_{bp} has been switched with Γ^d_{pb} . Clearly, the first term on the R.H.S vanishes. The second term cancels the forth term (note the change in the sequence of the anti-symmetrized indices). The third and fifth terms vanishes because the Christoffel symbols are symmetric for torsion-free connections, and therefore we get an anti-symmetrized sum of a symmetric object. Therefore, $\nabla_{[a} \nabla_b w_{c]} = 0$, and so, from eq.(7.3)

$$0 = \nabla_{[a} \nabla_b w_{c]} - \nabla_{[b} \nabla_a w_{c]} = R_{[abc]}{}^d w_d$$

for arbitrary w. □

Note that the above symmetry of the Riemann tensor does not require the Levi-Civita connection. It is sufficient that the torsion tensor vanishes. But, the next result requires the Levi-Civita connection.

Theorem 7.8. Property 3 *The Riemann curvature tensor stemming from the Levi-Civita connection is anti-symmetric in the last two indices. I.e.,*

$$R_{abcd} = -R_{abdc} \ . \tag{7.10}$$

Proof.

$$0 = (\nabla_a \nabla_b - \nabla_b \nabla_a) g_{cd} = -R_{abc}{}^e g_{ed} - R_{abd}{}^e g_{ce} = -R_{abcd} - R_{abdc} \ . \tag{7.11}$$

□

Theorem 7.9. Property 4 *For the Riemann curvature tensor stemming from the Levi-Civita connection, properties* 1*,* 2*, and* 3 *imply that*

$$R_{abcd} = R_{cdab} \ . \tag{7.12}$$

Proof. In this proof, we constantly rely on properties 1, 2, and 3 of the Riemann curvature tensor.

$$R_{[abc]d} = 0$$

implies

$$R_{abcd} - R_{acbd} - R_{bacd} + R_{bcad} + R_{cabd} - R_{cbad} = 0 \ .$$

I.e.,

$$R_{abcd} = R_{acbd} + R_{cbad} \ . \tag{7.13}$$

Similarly,

$$R_{[cda]b} = 0$$

implies

$$R_{cdab} = R_{cadb} + R_{adcb} \tag{7.14}$$

and

$$R_{[cbd]a} = 0$$

implies

$$R_{cbda} = R_{cdba} + R_{dbca} \ . \tag{7.15}$$

Finally,

$$R_{[abd]c} = 0$$

implies

$$R_{dabc} = R_{badc} + R_{dbac} \ . \tag{7.16}$$

Substituting eq.(7.15) in eq.(7.13), we get

$$R_{abcd} = R_{acbd} - R_{cdba} - R_{dbca} \ .$$

Similarly, substituting eq.(7.16) in eq.(7.14), we get

$$R_{cdab} = R_{cadb} + R_{badc} + R_{dbac} \ .$$

Subtracting the last two equations, we have the necessary result. □

Problem 7.3. Effortlessly show that the Riemann tensor vanishes for $M = \mathbb{R}^3$ with the Euclidean metric

$$g = dx \otimes dx + dy \otimes dy + dz \otimes dz \ .$$

Therefore conclude that the Riemann tensor (stemming from the Levi-Civita connection) vanishes in the spherical (or for that matter any) coordinate system.

Problem 7.4. In contrast to the last problem, for the Euclidean space $\mathbb{R}^3$, painfully show that

$$R_{123}{}^{1} = 0$$

in the spherical coordinate system. If you are still not convinced that the Riemann tensor (stemming from the Levi-Civita connection) vanishes for the Euclidean space $\mathbb{R}^3$ regardless of the coordinate system, you are free to compute all its remaining components in the spherical coordinate system (to begin with).

Problem 7.5. Clearly a sphere ain't flat: Find all the non-trivial components of the Riemann tensor for $S^2(r)$.

7.3 Contractions of the Riemann Tensor

Theorem 7.10.

$$R_a{}^a{}_{cd} = 0 \ .$$

Proof. From theorem 7.6, $g^{ab}R_{abcd} = 0$, since it is the symmetric sum of an anti-symmetric object. □

In exactly the same way:

Theorem 7.11.

$$R_{abc}{}^{c} = 0 \ .$$

Not all contractions of the Riemann tensor is trivial.

Theorem 7.12.

$$R_{abc}{}^{a} = -R_{ab}{}^{a}{}_{c} = R_{ba}{}^{a}{}_{c} = -R_{bac}{}^{a} \ .$$

Proof.

$$R_{abc}{}^{a} = g^{ad}R_{abcd} = -g^{ad}R_{abdc} = -R_{ab}{}^{a}{}_{c} \ .$$

The other cases are similar. □

There is no symmetry that requires $R_{abc}{}^{a}$ to be trivial. Therefore, there exists only one independent tensor contraction of the Riemann tensor.

Definition 7.2. *The Ricci tensor R_{ab} is given by*

$$R_{ab} = R_{acb}{}^{c} \ . \tag{7.17}$$

From the theorems above, all contractions of the Riemann tensor are proportional to the Ricci tensor.

Theorem 7.13. *The Ricci tensor is symmetric, i.e.,*

$$R_{ab} = R_{ba} \ .$$

Proof.

$$R_{ab} = g^{cd}R_{acbd} = g^{cd}R_{bdac} = R_{ba} \ .$$

□

Definition 7.3. *The scalar curvature R is given by*

$$R = R_a{}^{a} \ . \tag{7.18}$$

Bianchi Identity for the Riemann Curvature Tensor

Theorem 7.14. *The Riemann curvature tensor satisfies the Bianchi identity when the torsion tensor* $T = 0$, *i.e.,*

$$\nabla_{[a} R_{bc]d}{}^{e} = 0 \tag{7.19}$$

Proof. From eq.(7.8)

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)\nabla_c w_d = R_{abc}{}^{e} \nabla_e w_d + R_{abd}{}^{e} \nabla_c w_e \ .$$

While,

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

If we anti-symmetrize over indices a, b and c in both the equations above, we find the L.H.S's are equal. Therefore, by setting the R.H.S equal to each other, we get

$$R_{[abc]}{}^{e} \nabla_e w_d + R_{[ab|d|}{}^{e} \nabla_{c]} w_e = w_e \nabla_{[a} R_{bc]d}{}^{e} + R_{[bc|d|}{}^{e} \nabla_{a]} w_e \ [1] \ .$$

But, from theorem 7.7, $R_{[abc]}{}^{e} = 0$, and clearly,

$$R_{[ab|d|}{}^{e} \nabla_{c]} w_e = R_{[bc|d|}{}^{e} \nabla_{a]} w_e \ .$$

Therefore, $w_e \nabla_{[a} R_{bc]d}{}^{e} = 0$ for arbitrary w . □

7.4 The Einstein Tensor

Definition 7.4. *The Einstein tensor* G *is of type* $T^0_2(\mathcal{M})$ *and is defined by*

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab} \ . \tag{7.20}$$

Clearly, the Einstein tensor is symmetric.

Theorem 7.15. *For a Levi-Civita connection, the Einstein tensor is divergence free. I.e.,*

$$\nabla^a G_{ab} \equiv g^{ac} \nabla_c G_{ab} = 0 \ . \tag{7.21}$$

Proof. This follows from contracting the Bianchi equation. Due to the symmetries of the Riemann curvature tensor, the Bianchi identity takes the form

$$\nabla_a R_{bcd}{}^{e} + \nabla_b R_{cad}{}^{e} + \nabla_c R_{abd}{}^{e} = 0 \ .$$

Contracting c and e we get

$$\nabla_a R_{bd} - \nabla_b R_{ad} + \nabla_c R_{abd}{}^{c} = 0 \ .$$

Now we further contract a and d noting that $\nabla_c R_{abd}{}^{c} = \nabla^c R_{abdc} = \nabla^c R_{bacd}$ to obtain

$$\nabla^a R_{ba} - \nabla_b R + \nabla^c R_{bc} = 0 \ ,$$

i.e.,

$$2\nabla^a R_{ba} - \nabla_b R = 0 \ .$$

Therefore,

$$\nabla^a G_{ab} = \nabla^a R_{ab} - \frac{1}{2} g_{ab} \nabla^a R = \nabla^a R_{ab} - \frac{1}{2} \nabla_b R = 0$$

for Levi-Civita connections. □

Problem 7.6. Compute the Ricci tensor of $S^2(r)$.

Problem 7.7. Compute the scalar curvature of $S^2(r)$.

Problem 7.8. Compute the Einstein tensor of $S^2(r)$.

[1] Here, we simply ignore the index d in the anti-symmetrization process. More precisely:

$$R_{[bc|d|}{}^{e} \nabla_{a]} \ w_e = \frac{1}{6}(R_{bcd}{}^{e} \nabla_a w_e - R_{bad}{}^{e} \nabla_c w_e + R_{abd}{}^{e} \nabla_c w_e$$

$$- R_{acd}{}^{e} \nabla_b w_e - R_{cbd}{}^{e} \nabla_a w_e + R_{cad}{}^{e} \nabla_b w_e \ ) \ .$$

7.5 The Einstein Equation

So, what about gravity? We have seen that gravitation is described by the metric and the Levi-Civita connection (the Christoffel symbols determined by the equivalence principle govern the equations of motion for a test particle). Therefore, the field equations for gravity must be consistent with the statement that the Einstein tensor is divergence free. We have see in special relativity that the energy-momentum tensor is divergence free in the Minkowski coordinates, i.e.,

$$\frac{\partial}{\partial x^\mu} T^{\mu\nu} = 0 \, .$$

It is then reasonable to expect that in general relativity, when gravitational effects are included, we must have that

$$\nabla_\mu T^{\mu\nu} = 0 \tag{7.22}$$

with respect to the Levi-Civita connection. Also, both the energy-momentum tensor of matter fields and the Einstein tensor are symmetric. It is then reasonable to set

$$G_{ab} = R_{ab} - \frac{1}{2} R g_{ab} = \kappa T_{ab}$$

as a tentative equation for the gravitational field. Here κ is some coupling constant that is yet to be determined. Then we would have that

$$\nabla_a (R^{ab} - \frac{1}{2} R g^{ab}) = 0 = \kappa \, \nabla_a T^{ab}$$

provided we use the Levi-Civita connection. This is the essential story, but we will take a slightly longer way home! Let us however decide once and for all that *as implied by the equivalence principle, all covariant derivatives in general relativity will be taken using the Levi-Civita connection.*

Heuristic "Derivation" of the Einstein Equation

In Electrostatics, Coulombs law gives us that the electric field $\vec{E}$ and it's potential ϕ is prescribed by

$$\vec{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \hat{r} \text{ and } \nabla^2 \phi = -\frac{\rho}{\epsilon_0} \, .$$

Analogously, the Newtonian gravitational counterparts are given by

$$\vec{g} = -\frac{GM}{r^2} \hat{r} \text{ and } \nabla^2 \phi = 4\pi G \rho \, .$$

Here ϕ represents the gravitational potential. Also, recall for weak gravity,

$$g_{00} \approx -\left(1 + \frac{2\phi}{c^2}\right) \text{ and } T^{00} = \rho c^2 \, .$$

In the following calculation, since we want general relativity to be consistent with Newtonian gravity we will make some assumptions regarding the underlying conditions for applicability:

1. Gravity is time independent. I.e., we will require that $g_{\mu\nu,t} = 0$.

2. Gravity is weak. I.e., we can write the metric as a small perturbation of the Minkowski metric:

   $$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \, .$$

   Since $h_{\mu\nu}$ is small, we ignore all second or higher order terms in $h_{\mu\nu}$.

3. $T_{\mu\nu}$ is of order $h_{\mu\nu}$.

Claim. To first order in $h_{\mu\nu}$,

$$R_{00} \approx -\frac{1}{2} \nabla^2 g_{00} \approx \frac{1}{c^2} \nabla^2 \phi \, .$$

Proof. Recall that (note that we have raised and lowered indices differently here; verify this using eq.(7.2))

$$R^{\alpha}{}_{\beta\mu\nu} = \partial_\mu \Gamma^{\alpha}_{\nu\beta} - \partial_\nu \Gamma^{\alpha}_{\mu\beta} + \Gamma^{\alpha}_{\mu\sigma}\Gamma^{\sigma}_{\nu\beta} + \Gamma^{\alpha}_{\nu\sigma}\Gamma^{\sigma}_{\mu\beta} \, .$$

To first order in $h_{\mu\nu}$, Christoffel symbols become

$$\Gamma^{\alpha}_{\nu\beta} = \frac{1}{2}\eta^{\alpha\sigma}\{\partial_\beta g_{\sigma\nu} + \partial_\nu g_{\sigma\beta} - \partial_\sigma g_{\nu\beta}\} \, .$$

I.e.,

$$\partial_\mu \Gamma^{\alpha}_{\nu\beta} = \frac{1}{2}\eta^{\alpha\sigma}\{\partial_\mu\partial_\beta g_{\sigma\nu} + \partial_\mu\partial_\nu g_{\sigma\beta} - \partial_\mu\partial_\sigma g_{\nu\beta}\}$$

and

$$\partial_\nu \Gamma^{\alpha}_{\mu\beta} = \frac{1}{2}\eta^{\alpha\sigma}\{\partial_\nu\partial_\beta g_{\sigma\mu} + \partial_\nu\partial_\mu g_{\sigma\beta} - \partial_\nu\partial_\sigma g_{\mu\beta}\} \, .$$

Therefore, ignoring $\Gamma\Gamma$ terms we get that

$$R_{\alpha\beta\mu\nu} = \eta_{\alpha\theta}R^{\theta}{}_{\beta\mu\nu} = \frac{1}{2}\{\partial_\mu\partial_\beta g_{\alpha\nu} + \partial_\nu\partial_\beta g_{\alpha\mu} - \partial_\mu\partial_\alpha g_{\nu\beta} - \partial_\nu\partial_\alpha g_{\mu\beta}\} \, .$$

The expression for R_{00} becomes

$$R_{00} = g^{\alpha\beta}R_{\alpha 0\beta 0} = \frac{-g^{ij}}{2}\{\partial_i\partial_j g_{00} + \partial_0\partial_0 g_{ij} - \partial_0\partial_j g_{i0} - \partial_i\partial_0 g_{0j}\}$$

Then, due to the time independence of the metric tensor

$$R_{00} \approx -\frac{1}{2}g^{ij}\partial_i\partial_j g_{00} \approx \frac{1}{c^2}\nabla^2\phi$$

as needed (note $i, j = 1, 2, 3$). □

From the above claim, we see that

$$R_{00} \propto \nabla^2\phi \propto \rho \propto T_{00} \, .$$

Indeed, one is tempted to consider a field equation of type

$$R_{\mu\nu} \propto T_{\mu\nu}.$$

However, unlike $T_{\mu\nu}$, $R_{\mu\nu}$ is not divergence free (see eqs. (7.21) and (7.22)). Einstein easily resolved this problem by modifying the above equation to

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \kappa T_{\mu\nu} \, .$$

Contracting the free indices of the equation above we get that [2]

$$R = -\kappa \, T \, .$$

Here $T = T^{\mu}{}_{\mu}$. Thus

$$R_{\mu\nu} = \kappa\left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right) \, .$$

In particular

$$R_{00} = \kappa\left(T_{00} - \frac{1}{2}Tg_{00}\right) \, .$$

As we approximate the above equation, since T is considered as a small quantity, $Tg_{00} \approx T\eta_{00}$, and further in the rest frame of the dust, $T \approx \eta_{\mu\nu}T^{\mu\nu} \approx -T^{00} = -T_{00}$ (since $u^i = 0$). Thus

$$\frac{4\pi G}{c^2}\rho = \frac{1}{c^2}\nabla^2\phi \approx R_{00} \approx \kappa\left(T_{00} - \frac{1}{2}T_{00}\right) = \frac{1}{2}\kappa T_{00} \approx \frac{1}{2}\kappa\rho c^2 \, ,$$

thereby giving

$$\kappa = \frac{8\pi G}{c^4} \, .$$

The Einstein Field equation now takes the form

[2] Note $g^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} = 4$.

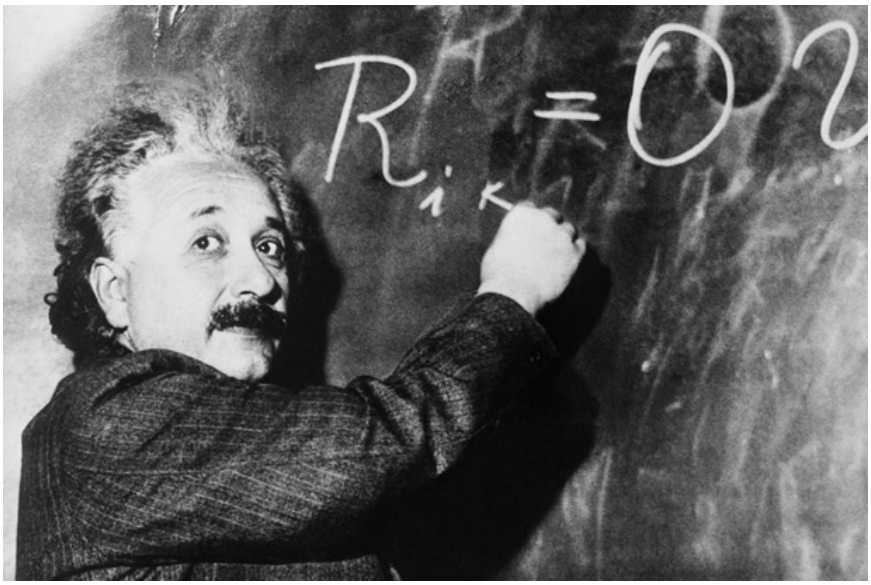


Fig. 7.1. A famous picture of Einstein displaying the vacuum equations of general relativity.

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

It is indeed possible to generalize the above equation to the form

$$\boxed{R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu}} = \frac{8\pi G}{c^4} T_{\mu\nu} \, . \tag{7.23}$$

Here Λ is a constant, famously known as the cosmological constant. The modification to include the cosmological constant was possible only due to the fact that $g_{\mu\nu}$ is divergence free. Typically, outside of a cosmological setting, we will choose the constant Λ to be vanishing. For vacuum spacetimes, $T_{\mu\nu} = 0$. I.e., when $\Lambda = 0$, $R = -\kappa T = 0$ as well. In this case, the Einstein equation reduces to

$$R_{\mu\nu} = 0 \, .$$

For this reason, vacuum spacetimes are often referred to as Ricci flat (see figure 7.1). While energy and momentum are a source for gravitation via the energy-momentum tensor, the gravitational field tells how matter and fields should evolve via the covariant derivative. This means that we must solve for gravitation and all energy and matter content at the same time. This is a daunting task as the Einstein equation is not linear in $g_{\mu\nu}$. However, when gravity is very strong compared to the other fields (for example, near a black hole), we can fix a background spacetime and study fields that evolve in it. Here the connection is already fixed since the metric determines the Christoffel symbols. For example, we could study how electromagnetic fields evolve in Schwarzschild spacetime (see problem 7.13). In this case, Maxwell's equations in a fixed background continues to be linear.

All equations of general relativity can be written in arbitrary coordinate systems. This is often referred to as the principle of *general covariance*. One often refers to as equations being *covariant*.

In addition to the weak and dominant energy condition for the energy-momentum tensor, for completeness we also include the following requirement that can be useful for certain applications.

Definition 7.5. *The energy-momentum tensor satisfies the strong energy condition if for any future pointing timelike vector field X, we have that*

$$\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) X^\mu X^\nu \geq 0 \, .$$

Here $T = g^{\mu\nu} T_{\mu\nu}$.

Problem 7.9. Show that the Schwarzschild metric satisfies the Einstein vacuum equation.

7.6 Fields in Curved Spacetime

Dust

Consider a non-interacting dust evolving in a purely gravitational field with the usual energy-momentum tensor

$$T^{\mu\nu} = \rho_0 u^\mu u^\nu \ .$$

Then

$$\nabla_\mu T^{\mu\nu} = \nabla_\mu(\rho_0 u^\mu)u^\nu + \rho_0 u^\mu \nabla_\mu u^\nu = (c\nabla_\mu J^\mu)u^\nu + \rho_0 \nabla_u u^\nu \ .$$

The first term in the right hand side above vanishes due to the continuity equation (3.23), and the second term vanishes due to the geodesic equation of motion. Once again we have that the energy-momentum tensor of dust is divergence free.

Perfect Fluid

In the case of a perfect fluid, as a natural generalization of eq.(3.24), we set

$$T^{\mu\nu} = \left(\rho_0 + \frac{\mathrm{P}}{c^2}\right) u^\mu u^\nu + \mathrm{P}\ g^{\mu\nu} \ .$$

The equations of motion for a perfect fluid can be obtained by setting $\nabla_\mu T^{\mu\nu} = 0$.

Electromagnetism

The following equations of electrodynamics will clearly reduce to the special relativistic case if we set

$$\nabla_\nu F^{\mu\nu} = \mu_0 J^\mu \ , \quad \nabla_{[\mu} F_{\nu\lambda]} = 0 \ ,$$

where the Lorentz force on a charged particle is given by

$$K^\mu = qF^{\mu\nu}U_\nu \ ,$$

along with the modified Newton's formula for force

$$K^\mu = ma^\mu \equiv m\nabla_U U^\mu \ .^{3}$$

The energy-momentum tensor for the electromagnetic field in a spacetime is then given by

$$T^{\mu\nu}_{EM} = \frac{1}{\mu_0}\left[g_{\alpha\beta}F^{\mu\alpha}F^{\nu\beta} - \frac{1}{4}g^{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right] \ .$$

Problem 7.10. When $J^\mu = 0$, show that $\nabla_\mu T^{\mu\nu}_{EM} = 0$.

Problem 7.11. Show that

$$\partial_{[\mu} F_{\nu\lambda]} = 0$$

is a covariant equation.

Problem 7.12. Show that

$$\nabla_{[\mu} F_{\nu\lambda]} = 0 \iff \partial_{[\mu} F_{\nu\lambda]} = 0 \ .$$

Problem 7.13. Let

$$F = q\sin\theta(d\theta \otimes d\varphi - d\varphi \otimes d\theta) \ .$$

[3] In special relativity

$$a^\mu = \frac{d}{d\tau}U^\mu = \frac{\partial x^\nu}{d\tau}\partial_\nu U^\mu = U^\nu \partial_\nu U^\mu$$

This naturally generalizes to

$$a^\mu = U^\nu \nabla_\nu U^\mu = \nabla_U U^\mu \ .$$

1. Show that $F_{\mu\nu}$ is a solution to Maxwell's equation in the Schwarzschild spacetime.

2. Find the current density vector J^μ described by the field.

3. Just for fun, find two linearly independent vector fields X and Y such that

$$F_{\mu\nu}X^\mu = 0 = F_{\mu\nu}Y^\mu \ .$$

   Don't over complicate this problem, a mere inspection will do the trick.

4. Calculate

$$\lim_{r\to\infty} \int_{r=\text{const}} \mathbf{B} \cdot d\mathbf{A} \ .$$

5. What are your thoughts regarding the answer to the previous question?

8

Cosmology

We begin by attempting to write a metric for the entire Universe, or even possibly a set of reasonable choices for a spacetime metric. This is not an easy task, and we will need simplifying assumptions. Ultimately, the validity of the expression will rely on the strength of our assumptions. The study of Cosmology then begins with the description of the very large-scale structure of the universe. At this scale, the universe appears to homogeneous, which is to say that space looks similar at any location, and isotropic, which is to say that it is spherically symmetric at any location [1]. This is the *cosmological principle* we will rely on. Images from the Sloan Digital Sky Survey (SDSS) shows that the distribution of galaxies and the voids in between forms a pattern that is consistent with our principle (see figure 8.1). Then, at this very large scale a single spatial coordinate will locate a cluster of galaxies and a void in a coordinate cube of volume $dxdydx$ at some location (see figure 8.2). For example, you and I, and the Milky Way galaxy in the larger Virgo super cluster of galaxies will be denoted by a single coordinate value.

Homogeneity and isotropy is implied only at any fixed value of *cosmic time*. We certainly do not expect the universe to have the same temperature and matter content a few billion years ago or even in the distant future. The notion of a cosmic time, at least theoretically, implies that there are potential *cosmic observers* at every spatial location with synchronized clocks. The set of all points with a fixed value of cosmic time t will generate all of *cosmic space* at time t. Further these cosmic observers will not have a spatial velocity in any direction, for then there will surely be a preferred direction at that location violating the notion of isotropy. Then the clock used by the cosmic observers will generate a timelike vector that is orthogonal to cosmic space that is generated by a fixed value of cosmic time. Thus the cosmological principle implies that the spacetime metric takes the form

$$g = -f(t)c^2\ dt \otimes dt + R^2(t)\left(h_{ij}\ dx^i \otimes dx^j\right)\ .^{2}$$

Here, $i, j = 1, 2, 3$, and $h_{ij} = h_{ij}(x^1, x^2, x^3)$ for some yet to be determined functions $f(t)$, and $R(t)$. The quantity

$$h(t) = R^2(t)\left(h_{ij}\ dx^i \otimes dx^j\right)$$

is the spatial metric at cosmic time t. Before we further analyze the spatial metric, it is easy to see that by a simple re-scaling of the time coordinate given by

$$\int^t \sqrt{f(t')}\ dt' \to t\ ^{3}$$

The spacetime metric now appears as

$$g = -c^2 dt \otimes dt + R^2(t)\left(h_{ij}\ dx^i \otimes dx^j\right)\ . \tag{8.1}$$

[1] Clearly then, for example, Schwarzschild spacetime is not isotropic.

[2] Since the cosmic observer is not moving with respect to cosmic space, her/his 4-velocity is of the form $u = (\dot{t}, 0, 0, 0)$. Here $\cdot$ represents the derivative with respect to proper time. Since this velocity is orthogonal to cosmic space, we have the $g_{0i} = 0$ for $i = 1, 2, 3$.

[3] For then

$$\sqrt{f(t)}\ dt \to dt\ ,$$

or

$$f(t)\ dt \otimes dt \to dt \otimes dt\ .$$

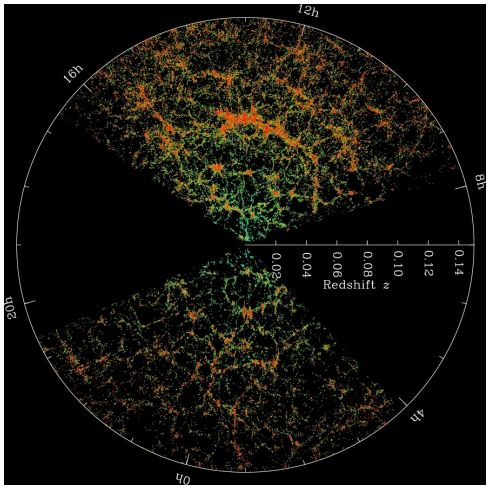


Fig. 8.1. The SDSS map of the Universe. Each dot is a galaxy; the color bar shows the local density.

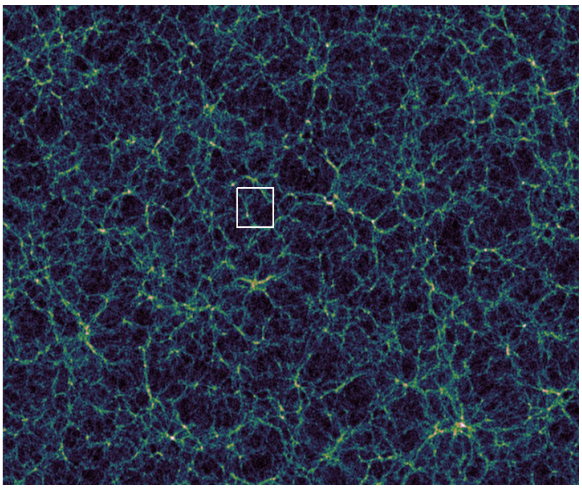


Fig. 8.2. This simulated view of the large-scale structure of the universe shows the vast cosmic web of galaxies, as well as the dark, empty expanses of the cosmic voids in between. A typical coordinate value encompasses a structure as large as the rectangle shown in the figure. Credit: Nico Hamaus, Universitäts-Sternwarte München, courtesy of The Ohio State University

$R^2(t)$ will differ in form in the re-scaled version. It is then the spatial metric $h(t)$ that is expected to be homogeneous and isotropic. Naturally, we will assign the global timelike vector ∂_t as future pointing.

8.1 Maximally Symmetric Space

The factor of $R^2(t)$ does not play an active role in the notion of homogeneity and isotropy. So for now, will ignore its role. Homogeneous and isotropic spaces posses a high degree of symmetry. In fact they are considered to be maximally symmetric. We do not delve into a study of maximally symmetric space. Suffice it to say that in general, for maximally symmetric spaces, the Riemann curvature tensor takes the form given by

$$R_{\mu\nu\alpha\beta} = k(g_{\alpha\nu}g_{\mu\beta} - g_{\beta\nu}g_{\mu\alpha}) \ . \tag{8.2}$$

Here k is a constant, and with a suitable choice of coordinates, can be restricted to take on values $0, -1$, and 1. Moreover for each value of k, there is a unique spatial metric associated to it. It then remains to find a(/the only) choice for the three cases.

Problem 8.1. Show that the Riemann curvature tensor given in eq.(8.2) satisfies the properties given in theorems 7.6 through 7.9.

8.1.1 Flat Space

To understand homogeneity and isotropy of space (not spacetime), consider the simplest case of the Euclidean space $\mathbb{R}^3$ with coordinates (x, y, z) and metric

$$h = dx \otimes dx + dy \otimes dy + dz \otimes dz \ .$$

Under a global coordinate transformation of type $\mathbf{x}' = \mathbf{x} + \mathbf{a}$, for any constant 3-vector $\mathbf{a}$, the spatial metric takes the form

$$h = dx' \otimes dx' + dy' \otimes dy' + dz' \otimes dz' \ .$$

The above two expressions for h has the same form; this is what we mean by homogeneity. In spherical coordinates (r, θ, φ), h is given by

$$h = dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2\theta d\varphi \otimes d\varphi \ .$$

The isotropy of the metric is then revealed since in the translated spherical coordinates (r', θ', φ')

$$h = dr' \otimes dr' + r'^2 d\theta' \otimes d\theta' + r'^2 \sin^2\theta' d\varphi' \otimes d\varphi' \ ,$$

where

$$x' = r' \sin\theta' \cos\varphi', y' = r' \sin\theta' \sin\varphi' \ , \ \text{ and } \ z' = r' \cos\theta' \ .$$

So while $\mathbb{R}^3$ possesses a translational symmetry, it is also spherically symmetric about any translated point. This is isotropy. The simplest choice for space in cosmology is a flat Euclidean space. For $\mathbb{R}^3$, the curvature tensor vanishes and so $k = 0$ in eq.(8.2).

8.1.2 Hyperbolic Space

Consider the 3-dimensional space given by the set of points (X, Y, Z, W) satisfying the equation for a hyperbolic space

$$X^2 + Y^2 + Z^2 - W^2 = -R^2 \ .$$

Here R is any fixed non-zero constant. This space can be parameterized by 3 coordinates (θ, χ, φ) by the equations [4]

$$X = R \cos\theta \sinh\chi \tag{8.3a}$$
$$Y = R \sin\theta \sinh\chi \cos\varphi \tag{8.3b}$$
$$Z = R \sin\theta \sinh\chi \sin\varphi \tag{8.3c}$$
$$W = R \cosh\chi \ . \tag{8.3d}$$

Here, $0 < \chi < \infty$ [5]. Now X, Y and Z have to be independently determined. Therefore we set $0 \leq \varphi < 2\pi$, while $0 \leq \theta < \pi$. On the hyperbolic space we impose the metric

$$dX \otimes dX + dY \otimes dY + dZ \otimes dZ - dW \otimes dW$$

subject to the constraints in eq.(8.3). The resulting metric is denoted as h_{H^3}.

Problem 8.2. Show that h_{H^3} in the (χ, θ, φ) coordinate system is given by

$$h_{H^3} = R^2 \left[d\chi \otimes d\chi + \sinh^2\chi \left(d\theta \otimes d\theta + \sin^2\theta d\varphi \otimes d\varphi \right) \right] \ . \tag{8.4}$$

We will now show that the hyperbolic space is homogeneous. What we are looking for is an invertible transformation that will take the point $(0, 0, 0, \pm R)$ to any point of the type $R(a_1, a_2, a_3, \pm\sqrt{1 + a^2})$. Here $\mathbf{a} = (a_1, a_2, a_3)$, and $a^2 = \mathbf{a} \cdot \mathbf{a}$. Moreover, under this transformation, as we shall show, the form of the metric will be preserved. Let $\mathbf{X} = (X, Y, Z)$. Then consider the transformation given by

$$\overline{\mathbf{X}} = \mathbf{X} + b(\mathbf{a} \cdot \mathbf{X})\mathbf{a} + W\mathbf{a} \tag{8.5a}$$
$$\overline{W} = (\mathbf{a} \cdot \mathbf{X}) + \sqrt{1 + a^2}\, W \ . \tag{8.5b}$$

[4] In this chart $W \geq R$. Therefore, we must construct a similar chart to describe the space when $W \leq R$. However, these two space are disconnected, and so we generally ignore this case.

[5] Clearly, the transformation is singular when $\chi = 0$ (look at the form of the metric). It can be shown that this is just a coordinate singularity by a simple translation of coordinates. Such translations are possible since the space is homogeneous and is explicitly constructed shortly.

All dot products above are Euclidean, and $b = (\sqrt{1+a^2}-1)/a^2$. Then

$$\overline{X}^2 + \overline{Y}^2 + \overline{Z}^2 - \overline{W}^2 = \mathbf{X}\cdot\mathbf{X} + b^2a^2(\mathbf{X}\cdot\mathbf{a})^2 + a^2W^2 + 2b(\mathbf{X}\cdot\mathbf{a})^2 +$$

$$2(\mathbf{a}\cdot\mathbf{x})W + 2bW(\mathbf{X}\cdot\mathbf{a})a^2 - (\mathbf{a}\cdot\mathbf{X})^2 - (1+a^2)W^2 - 2(\mathbf{a}\cdot\mathbf{X})\sqrt{1+a^2}W$$

$$= -R^2 + (\mathbf{X}\cdot\mathbf{a})^2[b^2a^2 + 2b - 1] + 2(\mathbf{a}\cdot\mathbf{X})W[1 + ba^2 - \sqrt{1+a^2}] = -R^2 \,.^{6}$$

Therefore, the transformation maps the hyperbolic space into itself, and further,

$$d\bar{\mathbf{X}} = d\mathbf{X} + b(\mathbf{a}\cdot d\mathbf{X})\,\mathbf{a} + \mathbf{a}\,dW, \quad \text{and} \quad d\bar{W} = (\mathbf{a}\cdot\ d\mathbf{X}) + \sqrt{1+a^2}dW$$

in exactly the same way as before implies that

$$dX^2 + dY^2 + dZ^2 - dW^2 = d\bar{X}^2 + d\bar{Y}^2 + d\bar{Z}^2 - d\bar{W}^2 \,.$$

Now define

$$\overline{X} = R\sinh\overline{\chi}\cos\overline{\theta} \tag{8.6a}$$
$$\overline{Y} = R\sinh\overline{\chi}\sin\overline{\theta}\cos\overline{\varphi} \tag{8.6b}$$
$$\overline{Z} = R\sinh\overline{\chi}\sin\overline{\theta}\sin\overline{\varphi} \tag{8.6c}$$
$$\overline{W} = R\cosh\overline{\chi}\,. \tag{8.6d}$$

Clearly in the $(\overline{\chi},\overline{\theta},\overline{\varphi})$ coordinate system, h_{H^3} is given by

$$h_{H^3} = R^2\left[d\overline{\chi}\otimes d\overline{\chi} + \sinh^2\overline{\chi}\left(d\overline{\theta}\otimes d\overline{\theta} + \sin^2\overline{\theta}\,d\overline{\varphi}\otimes d\overline{\varphi}\right)\right],$$

which is exactly in the form given in eq.(8.4). The transformations in eqs.(8.5) do not change the form of the metric, and can be applied successively so that any point is mapped to any other point in h_{H^3}, showing that it is a homogeneous space. Also, the presence of

$$d\theta\otimes d\theta + \sin^2\theta d\varphi\otimes d\varphi \quad \text{and} \quad d\overline{\theta}\otimes d\overline{\theta} + \sin^2\overline{\theta}\,d\overline{\varphi}\otimes d\overline{\varphi}$$

in h_{H^3} shows that isotropy is preserved as well.

8.1.3 The HyperSphere

Just as we proceeded in the case of the hyperbolic space, consider the 3-dimensional space given by the set of points (X,Y,Z,W) satisfying the equation for a hypersphere S^3

$$X^2 + Y^2 + Z^2 + W^2 = R^2 \,.$$

Here R is any fixed non-zero constant. This space can be parameterized by 3 coordinates (θ,χ,φ) by the equations

$$X = R\cos\theta\sin\chi \tag{8.7a}$$
$$Y = R\sin\theta\sin\chi\cos\varphi \tag{8.7b}$$
$$Z = R\sin\theta\sin\chi\sin\varphi \tag{8.7c}$$
$$W = R\cos\chi\,. \tag{8.7d}$$

Here, $0\le\chi\le\pi$, $0\le\theta\le\pi$ and $0\le\varphi<2\pi$ characterizes the whole space. On the hyper-sphere we impose the metric

$$h_{S^3} = dX\otimes dX + dY\otimes dY + dZ\otimes dZ + dW\otimes dW$$

subject to the constraints in eq.(8.7).

Problem 8.3. Show that in the (χ,θ,φ) coordinate system, h as given above, subject to the constraints in eq.(8.7) becomes

$$h_{S^3} = R^2\left[d\chi\otimes d\chi + \sin^2\chi\left(d\theta\otimes d\theta + \sin^2\theta d\varphi\otimes d\varphi\right)\right].$$

[6] Note

$$b^2a^2 + 2b - 1 = 0 = 1 + ba^2 - \sqrt{1+a^2}\,.$$

Problem 8.4. Exactly as in the hyperbolic space case, using the transformations

$$\overline{\mathbf{X}} = \mathbf{X} - b(\mathbf{a}\cdot\mathbf{X})\mathbf{a} + W\mathbf{a} \tag{8.8a}$$

$$\overline{W} = -(\mathbf{a}\cdot\mathbf{X}) + \sqrt{1-a^2}\, W\ . \tag{8.8b}$$

show that S^3 is homogeneous and isotropic.

Problem 8.5. Show that $\mathbb{R}^3$ with the Euclidean metric, h_{H^3} and h_{S^3} satisfies eq.(8.2).

8.2 Co-moving Coordinates

A metric of the form given in eq.(8.1) is said to be in *comoving coordinates*, and we will find it particularly useful in cosmology. A spatial coordinate label of fixed $x^i, i = 1, 2, 3$ locates a galaxy (or cluster of galaxies as in figure 8.2) for all time t. Then, distances between galaxies may change in time due to the presence of $R(t)$, while the coordinates themselves are fixed. Hence, the coordinates are comoving along with the galaxies, and the worldlines of galaxies are given by the curve $x^\mu = (ct, 0, 0, 0)$.

8.2.1 FLRW Spacetime

Problem 8.6. Show that the metric for the hyperbolic space, R^3 and S^3 can be written as

$$g = R^2\left[d\chi\otimes d\chi + \frac{\sin^2(\sqrt{k}\chi)}{k}(d\theta\otimes d\theta + \sin^2\theta d\varphi\otimes d\varphi)\right]\ ,$$

where in the case of

- R^3, the limit $k\to 0$ is implied.
- S^3, $k = 1$.

  and
- the hyperbolic space, $k = -1$.

Problem 8.7. Show that the metric for the hyperbolic space, R^3 and S^3 can be written as

$$g = R^2\left[\frac{1}{1-k\xi^2}\, d\xi\otimes d\xi + \xi^2(d\theta\otimes d\theta + \sin^2\theta d\varphi\otimes d\varphi)\right]\ ,$$

where in the case of

- R^3, $k = 0$, and $\xi = \chi$.
- S^3, $k = 1$, and $\xi = \sin\chi$.

  and
- the hyperbolic space, $k = -1$, and $\xi = \sinh\chi$.

Problem 8.8. Show that the value of k in the previous two problems are consistent with the value of k in eq.(8.2).

The present cosmological epoch will be denoted as t_0, and all quantities with a subscript of 0 will indicate their value at the present epoch. For example, $R_0 = R(t = t_0)$. Define a dimensionless scale factor $a(t)$ by

$$a(t) = \frac{R(t)}{R_0}\ .$$

Clearly, $a_0 = 1$.

The Friedmann Lemaître Robertson Walker (FLRW) metric for a homogeneous and isotropic spacetime then takes the form given by

$$g = -c^2 dt \otimes dt + a^2(t) R_0^2 \left[d\chi \otimes d\chi + \frac{\sin^2(\sqrt{k}\chi)}{k} (d\theta \otimes d\theta + \sin^2\theta d\varphi \otimes d\varphi) \right] ,$$

in the $(ct, \chi, \theta, \varphi)$ coordinate system or

$$g = -c^2 dt \otimes dt + a^2(t) R_0^2 \left[\frac{1}{1 - k\xi^2} \, d\xi \otimes d\xi + \xi^2 (d\theta \otimes d\theta + \sin^2\theta d\varphi \otimes d\varphi) \right] , \tag{8.9}$$

in the $(ct, \xi, \theta, \varphi)$ coordinate system for $k = 0, 1, -1$.

Claim. For a metric of the form given in eq.(8.1), the curve $x^\mu = (ct, 0, 0, 0)$ is an affinely parametrized future pointing timelike geodesic.

Proof. For $i = 1, 2, 3$,

$$\frac{d^2 x^i}{d\tau^2} = 0, \quad \text{and} \quad \Gamma^i_{00} = 0 .$$

Therefore, the geodesic equations of motion for the spatial coordinates are satisfied. Since $\Gamma^0_{00} = 0$, the only remaining requirement is given by

$$\frac{d^2 x^0}{d\tau^2} = 0 .$$

Therefore,

$$c\dot{t} \equiv c \frac{dt}{d\tau} = \text{constant} .$$

But the 4-velocity $u = (c\dot{t}, 0, 0, 0)$ is such that $u^2 = -c^2$, i.e., $-c^2 \dot{t}^2 = -c^2$. Therefore, $dt = d\tau$ as needed. Also, $u = (c, 0, 0, 0)$ is clearly future pointing. □

The previous claim tells us that cosmic space uses t as a propertime function, and that there is a notion of proper galactic distances that is reminiscent of special relativity.

8.2.2 Proper Distance in the Galactic Spatial Grid

One advantage of homogeneity is that we can chose a preferred special location without any loss of generality. Henceforth, we locate ourselves at $\xi = 0$. Then the *proper distance* [7] at any time t to another location ξ is clearly given by

$$d_p(\xi, t) = a(t) R_0 \int_0^\xi \frac{d\xi'}{\sqrt{1 - k\xi'^2}} .$$

Then

$$d_p(\xi, t) = a(t) \; d_p(\xi, t_0) ,$$

where

$$d_p(\xi, t_0) = R_0 \int_0^\xi \frac{d\xi'}{\sqrt{1 - k\xi'^2}} = \frac{R_0}{\sqrt{k}} \sin^{-1}(\sqrt{k}\xi) .$$

8.2.3 Hubble's Law

In 1929, Edwin Hubble noticed that there was a linear relationship between the redshift factor z and the distance d to several galaxies. This observation can be expressed in the form $z = (H_0/c)d$. Here, H_0 is a constant of proportionality conveniently scaled by a factor of c. Using the non-relativistic limit of Doppler's formula, $z \approx \beta$, he then concluded that all galaxies are receding away from us at a rate given by

$$v = H_0 d .$$

The quantity H_0 came to be known as Hubble's constant. Modern estimates, using measurements of the cosmic microwave background radiation left over from the Big Bang, place the value of the Hubble constant at

$$H_0 \approx 70 \, \frac{\text{Km/s}}{\text{Mpc}} .$$

To understand the above result within the context of FLRW cosmology, note that the *proper velocity* of a galaxy is given by

[7] Obtained by measuring length in cosmic space given by constant slices of t.

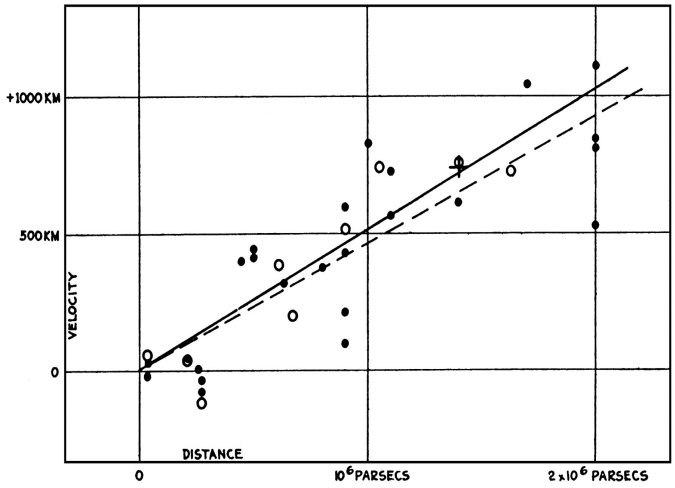


Fig. 8.3. As Robert P. Kirshner explains in [9], the graph shows the velocity–distance relation among extragalactic nebulae. "Radial velocities corrected for solar motion (but labeled in the wrong units) are plotted against distances estimated from involved stars and mean luminosities of nebulae in a cluster. The black discs and full line represent the solution for solar motion by using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually." (Copyright 1929, The Huntington Library, Art Collections and Botanical Gardens)

$$V_p(t) = \frac{d}{dt} d_p(\xi, t) = \dot{a}(t)\; d_p(\xi, t_0) = \frac{\dot{a}}{a}\; d_p(\xi, t)\;.$$

Define the Hubble function by

$$H = \frac{\dot{a}}{a}\;.$$

Then

$$V_p(t) = H(t)\; d_p(\xi, t)\;.$$

Hubble discerned this fact empirically as a result of his meticulous measurement of distance and velocities. Of course, Hubble was not aware of any time dependence in Hubble's Law, instead, he merely observed that

$$V_p(t_0) = H_0\; d_p(\xi, t_0)\;.$$

Since space is homogeneous, we can just as easily conclude that Hubble's Law is valid at any spatial location in the universe, and since $H_0 > 0$ we have that the universe is currently expanding.

8.2.4 Light Rays in FLRW Spacetime

Consider a ray of light (null geodesic) emitted at time t_{em} at some ξ_{em} along constant values of θ and φ that reaches us at the present epoch. Since along this path

$$-c^2 dt^2 + a^2(t) R_0^2 \frac{d\xi^2}{1 - k\xi^2} = 0\;,$$

we get that

$$\int_{t_{em}}^{t_0} \frac{c dt}{a(t)} = -R_0 \int_{\xi_{em}}^{0} \frac{d\xi}{\sqrt{1 - k\xi^2}} = R_0 \int_0^{\xi_{em}} \frac{d\xi}{\sqrt{1 - k\xi^2}} = d_p(\xi_{em}, t_0)\;.$$

The minus was inserted after the first equality above to make all quantities positive. Since

$$c\frac{dt}{a(t)} = c\frac{dt}{da}\frac{da}{a(t)} = c\frac{da}{a(t)\dot{a}(t)}$$

we get that

$$d_p(\xi_{em}, t_0) = c \int_{a_{em}}^{a_0 = 1} \frac{da}{a^2 H(a)}\;.$$

The equation above relates distance to the scale factor at emission.

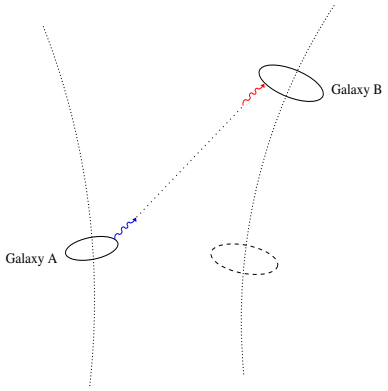


Fig. 8.4. A photon emitted from a distant galaxy becomes redshifted in an expanding universe as it reaches us at the present epoch.

Problem 8.9. Let $\alpha(s) = \big(ct(s), \xi(s), \theta(s), \varphi(s)\big)$ be a null geodesic in FLRW spacetime. Then show that

$$\frac{d}{ds}\left(a\frac{dt}{ds}\right) = 0\ .$$

Hint: Use the Euler-Lagrange equations of motion for a null geodesic.

A photon traversing the FLRW spacetime will experience a change in wavelength, λ, that depends directly only on the scale factor. For a null geodesic as above, the energy of the corresponding photon is given by

$$\frac{E}{c} = c\frac{dt}{ds}\ .$$

Since $E = hc/\lambda$, the above equation along with the previous claim implies that a/λ is a constant along the geodesic. I.e., in an expanding universe, the photon would appear to lose energy and vice versa! In general relativity, when the metric does not have a time translational symmetry, one must not expect energy conservation. Since λ is directly proportional to a, we have that the redshift factor

$$z = \frac{\lambda_{rec}}{\lambda_{em}} - 1 = \frac{a_0}{a(t_{em})} - 1 = \frac{1}{a(t_{em})} - 1\ .$$

Therefore, since

$$dz = \frac{-1}{a^2}da\ ,$$

from above, we get that

$$d_p(\xi_{em}, t_0) = c\int_{a_{em}}^{1}\frac{da}{a^2H(a)} = -c\int_{z}^{0}\frac{dz}{H(z)} = c\int_{0}^{z}\frac{dz}{H(z)}\ .$$

8.2.5 Luminosity distance

Let $\mathcal{L}$ be the intrinsic luminosity (total power radiated) of an object. Since the frequencies of the emitted and received photons are related by

$$\frac{\omega_{em}}{\omega_{rec}} = \frac{\lambda_{rec}}{\lambda_{em}} = 1 + z\ ,$$

the received energy of the photon is diminished from the transmitted energy by a factor given by

$$\omega_{rec} = \frac{\omega_{em}}{1+z}\ .$$

In a similar way, since frequency is inversely proportional to time, the lapse of time at emission and reception are related by

$$\frac{\Delta t_{rec}}{\Delta t_{em}} = 1 + z\ .$$

As we observe an object, the perceived luminosity $\mathcal{L}_0$ appears as

$$\mathcal{L}_0 \propto \frac{\omega_{rec}}{\Delta t_{rec}} = \frac{1}{(1+z)^2}\frac{\omega_{em}}{\Delta t_{em}} = \frac{1}{(1+z)^2}\,\mathcal{L}\,.$$

Then the flux of energy $\mathfrak{F}$ received is given by

$$\mathfrak{F} = \frac{(\mathcal{L}/(1+z)^2)}{4\pi\; d_p(\xi_{em}, t_0)^2}\,.$$

Here, flux is the received energy per unit area per unit time. The luminosity distance to an object is defined by

$$d_L = (1+z)d_p(\xi_{em}, t_0)$$

so that

$$\mathfrak{F} = \frac{\mathcal{L}}{4\pi\; d_L^2}\,.$$

8.3 Dynamical Equations for FLRW Cosmology

As it stands, we have a general idea about the spacetime metric and even about the matter content which is given by a perfect fluid energy momentum tensor, at least under the assumption of homogeneity and isotropy. The Einstein equation will play a major role in determining the governing equations for the scale factor $a(t)$, the density $\rho(t)$, and pressure $P(t)$ of the fluid. On the way, we hope to define experimental parameters that will help discern the value of k, which in turn will pin down the precise geometry of spacetime.

8.3.1 Friedmann Equations

Problem 8.10. For the form of the metric in eq.(8.9), show that

$$G_{00} = \frac{3\left(kc^2 + {R_0}^2\dot{a}^2\right)}{{R_0}^2a^2c^2}\,,\quad \text{and}\quad G_{\xi\xi} = \frac{-1}{1-k\xi^2}\frac{\left(kc^2 + 2{R_0}^2\ddot{a}a + {R_0}^2\dot{a}^2\right)}{c^2}\,.$$

Here $G_{\mu\nu}$ are components of the Einstein tensor defined by eq.(7.20).

In our comoving coordinates, the proper velocity of the fluid is given by

$$u^\mu = \left(c, \vec{0}\right)\,,\quad \text{and}\quad u_\mu = \left(-c, \vec{0}\right)\,.$$

Problem 8.11. Show that for a perfect fluid with u as given above

$$T_{00} = \left(\rho + \frac{P}{c^2}\right)c^2 - P = \rho c^2$$

and

$$T_{\xi\xi} = Pg_{\xi\xi} = \frac{P{R_0}^2a^2}{1-k\xi^2}\,.$$

The 00 component of the Einstein equation then implies that

$$\frac{3\left(kc^2 + {R_0}^2\dot{a}^2\right)}{{R_0}^2a^2c^2} = \frac{8\pi G}{c^4}\rho c^2$$

or equivalently

$$\frac{kc^2}{{R_0}^2a^2} + \frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho\,, \tag{8.10}$$

and the $\xi\xi$ component gives

$$\frac{-1}{1-k\xi^2}\frac{\left(kc^2 + 2{R_0}^2\ddot{a}a + {R_0}^2\dot{a}^2\right)}{c^2} = \frac{8\pi G}{c^4}\frac{P{R_0}^2a^2}{1-k\xi^2}$$

or equivalently

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{c^2}\left[\frac{1}{3}\rho c^2 + P\right]\,. \tag{8.11}$$

Equations (8.10) and (8.11) will be referred to as the first and second Friedmann equations, respectively.

Problem 8.12. Show that the Friedmann equations imply that

$$\frac{d}{dt}\left(\rho c^2 a^3\right) = -P\frac{d}{dt}a^3 \ .$$

The remaining non-trivial components of the Einstein equation give no further information. In fact they are identical to eq.(8.11). So we face the conundrum of having three variables, namely a, ρ, and P, and having only two equations. The only way to determine a unique time evolution require a third equation. In this case, this is the equation of state.

8.3.2 The Equation of State

The first law of thermodynamics tells us that $\Delta E = \Delta q + \Delta w$. In this context, $\Delta q = 0$ since no one is extracting heat from the universe, and so $dE = -PdV$. I.e.,

$$P \propto \frac{dE}{dV} = \rho c^2.$$

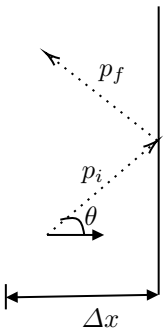
With this is mind, we will consider an equation of state of the type

$$P = w\rho c^2 \ , \tag{8.12}$$

where w is constant. In a matter-dominated universe, $P = 0$, or equivalently, $w = 0$. However, this is not the case when the universe is dominated by radiation. Consider a box of photons bouncing off the walls. For now, assume the photons to be monochromatic. Let n be the number of photons, ϵ the energy of photons, and A the area of each face of the box. The x-displacement Δx traversed by a photon on a given time interval Δt is given by

$$\Delta x = c\cos\theta\Delta t \ .$$

Here θ is the angle the momentum vector makes with the x-axis as shown in the figure below.



The x-momentum imparted when the photon reflects off the wall is

$$2p_x = 2p\cos\theta \ .$$

Here p is the momentum of the photon so that $\epsilon = pc$. Therefore, the force imparted by a single photon is given by

$$F = \frac{2p\cos\theta}{\Delta t} \ ,$$

and hence the pressure P exerted by the single photon is given by $P = F/A$. Of course, the photons are free to move in any direction inside the box, so only half will be moving to the right. Therefore, the total average pressure exerted on the wall is found by integrating across all directions, and is given by

$$P = \frac{1}{2}\frac{1}{\int \sin\theta \ d\theta d\phi}\int \frac{(2p\cos\theta)}{A\Delta t}\,(nA\Delta x)\sin\theta \ d\theta d\phi$$

$$= \frac{1}{8\pi}\int \frac{(2p\cos\theta)}{\Delta t}\,(nc\cos\theta\Delta t)\sin\theta \ d\theta d\phi$$

$$= \frac{1}{2}pnc \int \cos^2\theta \sin\theta d\theta = \frac{1}{2}pnc \frac{\cos^3\theta}{3}\Big|_0^\pi = \frac{1}{3}(pc)n = \frac{1}{3}\epsilon n \ .$$

Thus, with the understanding that energy density ϵn can be written as ρc^2, we get that the equation of state for radiation is given by

$$P = \frac{1}{3}\rho c^2 \ . \tag{8.13}$$

Now if there are photons of different wavelength and frequency, we include their individual energy and wind up with the same expression for pressure. Clearly, for radiation, $w = 1/3$.

Claim. When eq.(8.12) is satisfied, the mass density is given by

$$\rho = \rho_0 \left[a(t)\right]^{-3(1+w)} \ . \tag{8.14}$$

Proof. From problem 8.12, we get that

$$\dot{\rho}c^2 = -\frac{1}{a^3}\frac{da^3}{dt}\left(\rho c^2 + P\right) = -\frac{3\dot{a}}{a}\left(1+w\right)\rho c^2 \ .$$

I.e.,

$$\frac{\dot{\rho}}{\rho} = -\frac{3\dot{a}}{a}\left(1+w\right) \ .$$

Therefore,

$$\frac{d\rho}{\rho} = -3\left(1+w\right)\frac{da}{a} \ ,$$

or

$$\rho = \rho_0 \left[a(t)\right]^{-3(1+w)} \ .$$

□

For a matter dominated universe, by definition $P = 0$ and so $w = 0$, hence

$$\rho_M(t) = \rho_{M,0}\left[a(t)\right]^{-3} \ .$$

For radiation dominated universe $w = 1/3$, and hence

$$\rho_R(t) = \rho_{R,0}\left[a(t)\right]^{-4} \ .$$

Finally, consider a negative pressure equation of state given by the choice $w = -1$. Here

$$P = -\rho c^2 \ ,$$

and the energy density is a constant in time, even in an expanding universe. I.e.,

$$\rho(t) = \rho\left(t_0\right) \ .$$

As we shall see, this has useful interpretations in the case of dark energy.

8.3.3 Heuristic Explanation of Friedmann Equations

It is a remarkable fact that the Friedmann equations can be understood in simple Newtonian terms. Consider a universe filled with galaxies in a spherically symmetric pattern about our location here on Earth. Further, in comoving coordinates, let the coordinates of a particular galaxy, denoted by A, have fixed spatial coordinates $\vec{x}_A$. We will place ourselves at $\vec{x} = \vec{0}$. Since the coordinates are fixed for all time, the distance D between us and galaxy A is given by

$$D(t) = a(t)\Delta x$$

where a(t) is a time dependent scale factor as before, and $\Delta x = |\vec{x}_A|$. The velocity and acceleration of A are given by

$$V(t) = \dot{a}(t)\Delta x \ ,$$

and

$$A(t) = \ddot{a}(t)\Delta x \ .$$

If the total mass of all galaxies contained in a sphere of radius D is M, then by Gauss' Law for Newtonian gravity, we have that the force equation gives

$$mA = -\frac{GmM}{D^2} ,$$

where m is the mass of galaxy A. If ρ is the universal mass density, we have that

$$M = \frac{4}{3}\pi(a^3\Delta x^3)\rho .$$

Therefore,

$$m\ddot{a}(t)\Delta x = \frac{-Gm\rho\frac{4}{3}\pi a^3\Delta x^3}{a^2\Delta x^2} ,$$

or equivalently

$$\frac{\ddot{a}}{a} = -\frac{4\pi G\rho}{3} .$$

This is precisely the second Friedmann Equation (eq.(8.11)) for a matter-dominated universe when $P = 0$. As expected, due to the presence of $\ddot{a}$, eq.(8.11) has the interpretation of a simple force equation. Meanwhile, the conservation of energy of m is given by

$$\frac{1}{2}mV^2 - \frac{Gm}{D}M = E ,$$

where E is the total energy of the galaxy A. I.e.,

$$\frac{1}{2}m\dot{a}^2\Delta x^2 - \frac{Gm}{a\Delta x}\rho\frac{4}{3}\pi a^3\Delta x^3 = E$$

or equivalently,

$$\left(\frac{\dot{a}}{a}\right)^2 - \frac{8\pi G}{3}\rho = \frac{2E}{\Delta x^2 m a^2} .$$

This is the first Friedmann Equation (eq.(8.10)) if we set

$$\frac{2E}{\Delta x^2 m} = -kc^2 .$$

To understand the qualitative features of FLRW cosmology, note that the Friedmann equation above can be written as

$$\dot{a}^2 - \frac{8\pi G}{3}a^2\rho = -kc^2 .$$

Drawing an analogy from classical mechanics, regardless of whether the universe is matter or radiation dominated, the potential energy term is of the form

$$U = -\frac{8\pi G}{3}\frac{\rho_0}{a} , \quad \text{or} \quad -\frac{8\pi G}{3}\frac{\rho_0}{a^2} .$$

In either case, from the behavior of the potential, we can conclude that

- When $k \leq 0$, and a is unbounded.
  and

- When $k = +1$, i.e., the spatial geometry is that of the hypersphere, and a is bounded. I.e, the universe does not expand forever.

8.3.4 Critical Density and Spatial Geometry

Eq.(8.10) can be rewritten in the form

$$-k = \frac{{R_0}^2}{c^2}\dot{a}^2\left[1 - \frac{8\pi G}{3H^2}\rho\right] .$$

Define critical density $\rho_c(t)$ by

$$\rho_c(t) = \frac{3H^2(t)}{8\pi G} . \tag{8.15}$$

Then

$$-k = \left(\frac{R_0\dot{a}}{c}\right)^2\left[1 - \frac{\rho}{\rho_c}\right] . \tag{8.16}$$

Therefore, if

- $\rho(t) < \rho_c(t)$, we have a hyperbolic universe ($k = -1$).
- $\rho(t) = \rho_c(t)$, then we have a flat universe ($k = 0$).
  and
- $\rho(t) > \rho_c(t)$, then $k = +1$, i.e., the spatial geometry is that of the hypersphere.

A simple comparison of ρ_c, which is easily obtained from the Hubble constant H_0, with our observed ρ determines the value of k. The current estimation of energy density is given by

$$\rho = \rho_c(1 \pm 0.01) \ .$$

Sadly, this is exactly where we do not want to be, and the value of k is yet unfixed.

8.4 Dark Energy

The Einstein Equation with a cosmological constant given by eq.(7.23) when rewritten as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} - \Lambda g_{\mu\nu}$$

has a simple interpretation of an additional source of energy and momentum represented by the term involving the cosmological constant. In the literature, this term often appears as a leading candidate for the postulated *dark energy*. A deeper dive into the need for dark energy, and other various models for it are beyond the scope of our discussion [8]. The Friedmann equations are modified by the presence of the cosmological constant, and are given by

$$\frac{kc^2}{R_o^2 a^2} + \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{1}{3}\Lambda c^2 \ ,$$

and

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{c^2}\left[\frac{1}{3}\rho c^2 + P\right] + \frac{1}{3}\Lambda c^2 \ .$$

Problem 8.13. Show that the presence of the cosmological constant in the Einstein equation leads to the modified Friedmann equations given above.

The Equation of State for Dark Energy

Let ρ_{other} and P_{other} denote energy and pressure density due to everything in the universe excluding dark energy. Then

$$\frac{kc^2}{R_o^2 a^2} + \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left[\rho_{\text{other}} + \rho_\Lambda\right] = \frac{8\pi G}{3}\ \rho_{\text{Total}} \ ,$$

where $\rho_\Lambda = \Lambda c^2/8\pi G$, and $\rho_{\text{tot}} = \rho_{\text{other}} + \rho_\Lambda$. The second Friedmann equation can be written as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{c^2}\left[\frac{1}{3}\rho_{\text{other}}c^2 + P_{other}\right] + \frac{1}{3}\Lambda c^2$$

$$= -\frac{4\pi G}{c^2}\left[\frac{1}{3}\rho_{\text{tot}}c^2 + P_{\text{tot}}\right] \ ,$$

where $P_{\text{tot}} = P_{\text{other}} + P_\Lambda$ and $P_\Lambda = -\Lambda c^4/8\pi G$. Therefore, we can interpret

$$P_\Lambda = -\rho_\Lambda c^2 \ ,$$

as the pressure due to dark energy, and so, here $w = -1$.

[8] The interested is referred to [10]. Another leading candidate for dark energy is Quintessence which is described by a scalar field. Here is a readable article describing Quintessence.

8.5 Exact Solutions is FLRW Cosmology

8.5.1 The Radiation Filled Universe

For a radiation-dominated universe, when there is no "matter",

$$\frac{\dot{a}^2}{a^2} + \frac{kc^2}{{R_0}^2 a^2} = \frac{8\pi G}{3}\frac{\rho_{R,0}}{a^4} \ .$$

In this case, set

$$A = \sqrt{\frac{8\pi G}{3}\rho_{R,0}} \ .$$

Then

$$\dot{a^2} = \frac{A^2}{a^2} - \frac{kc^2}{{R_0}^2} \ .$$

It is then easily shown that

$$a(t) = \sqrt{t}\sqrt{2A - \frac{kc^2}{{R_0}^2}t} \ . \tag{8.17}$$

The above expression gives the evolution equation for $a(t)$ in a radiation filled universe regardless of the value of k.

Problem 8.14. Verify that eq.(8.17) is the solution in a radiation-dominated universe.

8.5.2 The Flat Universe

When $k = 0$, from eqs.(8.10) and (8.14) we get that

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho_0 a^{-3(1+w)} \ .$$

Using the ansatz

$$a(t) = \left(\frac{t}{t_0}\right)^s \ , \tag{8.18}$$

and noting that $\dot{a}/a = s/t$, we get that

$$\left(\frac{s}{t}\right)^2 = \frac{8\pi G}{3}\rho\left(\frac{t}{t_0}\right)^{-3s(1+w)} \ .$$

When

$$s = \frac{2}{3(1+w)} \ ,$$

the powers of t match to obtain

$$\left(\frac{s}{t}\right)^2 = \frac{8\pi G}{3}\rho_0\frac{t_0^2}{t^2} \ .$$

In this case

$$s = \sqrt{\frac{8\pi G}{3}\rho_0}\ t_0 \ .$$

Defining a Hubble time t_H by

$$t_H = \sqrt{\frac{3}{8\pi G\rho_0}} \ ,$$

we get that the present epoch is given by

$$t_0 = s t_H \ .$$

Since

$$H_0 = \left(\frac{\dot{a}}{a}\right)_{t=0} = \frac{s}{t_0} \ ,$$

the Hubble time is simply the inverse of the Hubble constant, i.e., $t_H = H_0^{-1}$. For matter dominated universe, $w = 0$, and therefore $s = \frac{2}{3}$ and

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3}$$

and

$$t_0 = \frac{2}{3} t_H \ .$$

where t_0 represents the age of the universe. At $t = 0$, $a(0) = 0$ making the metric singular. All spatial distances, in this limit, becomes 0. This is the *Big Bang* singularity in FLRW cosmology when $k = 0$. [9] For a radiation dominated universe, $w = \frac{1}{3}$, and therefore we get $s = \frac{1}{2}$ which gives that

$$a(t) = \left(\frac{t}{t_0}\right)^{1/2}$$

and

$$t_0 = \frac{1}{2} t_H = \frac{1}{2H_0} \ .$$

Regardless of the details of the equation of state, i.e., the value for w, from the ansatz for a, we see that the metric exhibits a big bang singularity at $t = 0$.

Problem 8.15. Show that the above expression for $a(t)$ in a radiation-dominated universe agrees with eq.(8.17).

Horizons in FLRW Cosmology

For simplicity, we consider the $k = 0$ case. For a null geodesic along constant values of θ and φ, we have that

$$-c^2 dt^2 + a^2 {R_0}^2 d\xi^2 = 0 \ .$$

where

$$a = \left(\frac{t}{t_0}\right)^s \ .$$

This in turn implies that

$$\int_{\xi_0}^{\xi} d\xi = \pm \frac{c}{R_0} {t_0}^s \int_{t_0}^{t} t^{-s} dt \ ,$$

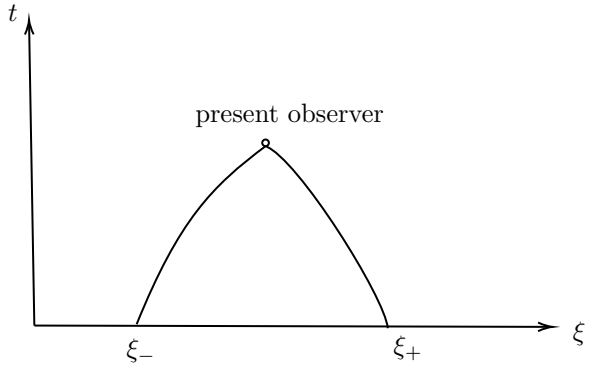
or

$$\xi = \xi_0 \pm \frac{c}{R_0} \frac{t_0^s}{1-s} \left(t^{1-s} - {t_0}^{1-s}\right) \ .$$

Going backwards in time, as $t \to 0$,

$$\xi \to \xi_0 \mp \frac{c}{R_0} \frac{t_0}{1-s} \equiv \xi_\mp \ .$$

I.e., an observer located at ξ_0 at the present epoch will not have any communication from values of ξ in the complement of the set (ξ_-, ξ_+). Thus $\xi_\mp$ locates the causal horizon for the present observer (see below).



[9] Recall that $H_0 \approx 70$ Km/s/Mpc, and

$$1 \text{ Mpc} = 3.0857 \times 10^{19} \text{ km} \ .$$

Therefore,

$$\frac{1}{H_0} = \frac{3.0857 \times 10^{19} \text{ km}}{70 \text{ km/s}} = 4.4081 \times 10^{17} \text{ s} = 1.397 \times 10^{10} \text{ yr} = 13.97 \text{ billion yrs.}$$

For our physical universe, the proper distance to the horizon is given by

$$d_p(\xi_{em}, t_0) = \int_{t_{em}=0}^{t_0} \frac{cdt}{a(t)} \sim 40 \text{ billion light years} .$$

We can see much further back than thirteen billion light-years, as we might otherwise expect, since the universe has been expanding.

8.5.3 The Matter Dominated Non-Flat Universe

In preparation for solving for $a(t)$ when $k = \pm 1$ in a matter-dominated universe, it is advantageous to define a deceleration parameter $q(t)$ by

$$q(t) = -a\frac{\ddot{a}}{\dot{a}^2} .$$

Note that if $\ddot{a} > 0$ we have $q < 0$. I.e., q is proportional to "negative acceleration". Evaluating eq.(8.11) at the present epoch gives

$$-q_0 {H_0}^2 = -\frac{4\pi G}{c^2}\left[P_0 + \frac{1}{3}\rho_0 c^2\right]$$

or,

$$P_0 = -\frac{1}{3}\rho_0 c^2 + \frac{c^2}{4\pi G} q_0 {H_0}^2 .$$

Evaluating eq.(8.10) at the present epoch gives

$$ {H_0}^2 + \frac{kc^2}{{R_0}^2} = \frac{8\pi G}{3}\rho_0 .$$

The above two equations can be used to solve for P_0.

$$P_0 = -\frac{1}{3}c^2\frac{3}{8\pi G}\left[{H_0}^2 + \frac{kc^2}{{R_0}^2}\right] + \frac{c^2}{4\pi G} q_0 H_0^2$$

$$= -\frac{c^2}{8\pi G}\left[\frac{kc^2}{{R_0}^2} + H_0^2 - 2q_0 {H_0}^2\right] .$$

I.e.,

$$P_0 = -\frac{c^2}{8\pi G}\left[\frac{kc^2}{{R_0}^2} + (1 - 2q_0){H_0}^2\right] .$$

For a matter-dominated universe,

$$P_0 = 0 \implies \frac{kc^2}{{R_0}^2} = {H_0}^2 (2q_0 - 1) . \tag{8.19}$$

Therefore,

$$\rho_0 = \frac{3}{8\pi G}\left[{H_0}^2 + {H_0}^2 (2q_0 - 1)\right] = \frac{3}{8\pi G} 2q_0 {H_0}^2 .$$

The $k = +1$ Matter Dominated Universe

Here

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\frac{\rho_0}{a^3} - \frac{c^2}{{R_0}^2 a^2} ,$$

i.e.,

$$\dot{a}^2 = \frac{2q_0 {H_0}^2}{a} - \frac{c^2}{{R_0}^2} = \frac{c^2}{{R_0}^2}\left[\frac{2q_0 {H_0}^2 {R_0}^2}{c^2 a} - 1\right] .$$

When $k + 1$, eq.(8.19) gives that

$$\frac{c^2}{{R_0}^2} = {H_0}^2 (2q_0 - 1) .$$

Therefore

$$\dot{a^2} = \frac{c^2}{{R_0}^2}\left[\frac{2q_0}{(2q_0 - 1)}\frac{1}{a} - 1\right] .$$

Let

$$\alpha = \frac{2q_0}{(2q_0 - 1)} \, .$$

I.e.,

$$\dot{a^2} = \frac{c^2}{{R_0}^2} \left[\frac{\alpha}{a} - 1\right] \, ,$$

or,

$$c \int dt = R_0 \int \frac{\sqrt{a}}{\sqrt{\alpha - a}} da \, .$$

To perform a trigonometric substitution, note that

$$\dot{a} \geq 0 \;\rightarrow\; \frac{\alpha}{a} \geq 1 \;\rightarrow\; \alpha \geq a \geq 0 \, .$$

Now, let

$$a = \alpha \sin^2 \frac{\Theta}{2} \, ,$$

for a coordinate range for Θ that is yet to be determined. The integral now becomes

$$c \int dt = \alpha R_0 \int \sin^2 \frac{\Theta}{2} d\Theta \, .$$

or,

$$ct - ct_i = \frac{\alpha R_0}{2} \left(\Theta - \sin\Theta\right) \, ,$$

for some integration constant ct_i. Clearly, when $t = t_i$, $\Theta = 0$. Now pick t_i to be 0, and set the range for Θ to be $0 \leq \Theta \leq 2\pi$. Setting $a(t_0) = 1$, we have that

$$1 = \alpha \sin^2 \frac{\Theta_0}{2} = \frac{\alpha}{2} \left[1 - \cos\Theta_0\right] = \frac{q_0}{2q_0 - 1} \left[1 - \cos\Theta_0\right]$$

implying that

$$\cos\Theta_0 = \frac{1 - q_0}{q_0} \, .$$

Thus

$$\sin\Theta_0 = \frac{\sqrt{2q_0 - 1}}{q_0} \, .$$

Therefore,

$$t_0 = \frac{\alpha R_0}{2c} \left(\Theta_0 - \sin\Theta_0\right) \, .$$

Using eq.(8.19), when $k + 1$, the present epoch is given by

$$t_0 = \frac{q_0}{(2q_0 - 1)^{3/2}} \frac{1}{H_0} \left[\cos^{-1} \frac{(1 - q_0)}{q_0} - \frac{(2q_0 - 1)^{1/2}}{q_0}\right] \, .$$

The lifetime of the universe (when $\Theta = 2\pi$)

$$t_L = \frac{\alpha R_0^2}{2c} 2\pi = \frac{2\pi q_0}{(2q_0 - 1)^{3/2} H_0} \, .$$

8.5.4 Familiarity with Dark Energy

In this subsection, we will set $\rho = 0 = P$. Only contribution to dynamics will be k and ρ_Λ, and consequently P_Λ. In this case, the first and second Friedmann equations will take the form

$$\frac{kc^2}{{R_0}^2 a^2} + \frac{\dot{a}^2}{a^2} = (H_\Lambda)^2$$

and

$$\frac{\ddot{a}}{a} = (H_\Lambda)^2$$

respectively, where

$$H_\Lambda = \sqrt{\frac{1}{3}\Lambda c^2}$$

is the Hubble constant of the De-Sitter geometry (as we shall see). Consider first the case when $\Lambda > 0$. Then the second Friedmann equation is satisfied by the general solution

$$a = C_1 e^{H_\Lambda t} + C_2 e^{-H_\Lambda t} .$$

The first equation then imposes the condition

$$4C_1 C_2 {R_0}^2 (H_\Lambda)^2 = kc^2 .$$

- When $k = 1$, $\Lambda > 0$ and dark energy is the only source, then we set

$$C_1 = C_2 = \frac{c}{2R_0 H_\Lambda} ,$$

and then

$$a = \frac{c}{R_0 H_\Lambda} \cosh H_\Lambda t . \tag{8.20}$$

I.e., a contracting universe reaches a minimum, and then expands again forever. The resulting spacetime is referred to as de Sitter geometry.

- When $k = 0$, $\Lambda > 0$, and dark energy is the only source. Then set $C_2 = 0$ for an expanding universe (i.e., $\dot{a} > 0$) and

$$a = e^{-H_\Lambda t_0} \, e^{H_\Lambda t} .$$

where $e^{-H_\Lambda t_0}$ is the integration constant C_1 so that $a(t_0) = 1$. This is clearly an expanding universe. If instead we wanted a contracting universe, we could have just as easily set $C_1 = 0$.

- When $k = -1$, $\Lambda > 0$ and dark energy is the only source, then we set

$$C_1 = -C_2 = \frac{c}{2R_0 H_\Lambda} ,$$

and then

$$a = \frac{c}{R_0 H_\Lambda} \sinh H_\Lambda t . \tag{8.21}$$

It can be shown that in the case where $\Lambda > 0$, and $k = \pm 1$, the FLRW spacetime is a portion of the maximal de Sitter spacetime of the $k = 0$ case. Consequently, the regions described in the $k = \pm 1$ g are geodesically incomplete, meaning that causal geodesics run out of spacetime in a finite amount of affine parameter. This is certainly artificial. For a readable account of the various submanifolds in these cases see the book by [11]. [10] For a full four-dimensional account of coordinate transformations, also see [12].

8.5.5 Asymptotic Behavior

Early Universe

In the early universe, ρ can fall off as a^{-3} or a^{-4} for a matter or radiation dominated universe, respectively. This implies that

$$\left(\frac{\dot{a}}{a}\right)^2 \approx \frac{8\pi G}{3}\rho ,$$

in the early universe since from the first Friedmann equation

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G}{3}\rho - \frac{kc^2}{{R_0}^2 a^2} ,$$

and a is very small. The term with k only has a^{-2} dependence as opposed to ρ. Therefore, the early universe agrees in all three cases regardless of the value of k.

[10] There are a few eccentricities in Moukhanov's book. In our notation, equations (1.90) and (1.91) of Moukhanov's should read as

$$-T^2 + X^2 + Y^2 = (H_\Lambda)^{-2}$$

and

$$ds^2 = -dT^2 + dX^2 + dY^2$$

respectively. Here T is a timelike coordinate, and X and Y are two spacelike coordinates. Mukhanov suppresses two spatial coordinates. This makes the diagrams easier to draw. Please note that, as a result, $0 \le \chi \le 2\pi$, unlike the full four dimensional spacetime geometry when $k = 1$ we have that $0 \le \chi \le \pi$. The basic reason is that a circle has 2π radians, while the polar angle θ in S^2, for example, has domain $[0, \pi]$. Nonetheless, barring details, the basic ideas remain the same.

The Late Flat Universe

When $k = 0$ in a matter-dominated universe,

$$a(t) = \left(\frac{t}{t_0}\right)^{2/3}$$

implies

$$\dot{a} = \frac{2}{3}\left(\frac{1}{t_0}\right)^{2/3}\left(\frac{1}{t}\right)^{1/3} .$$

Therefore,

$$\dot{a} \to 0 \ \text{ as } \ t \to \infty .$$

For a radiation-dominated universe, when $k = 0$,

$$a(t) = \left(\frac{t}{t_0}\right)^{1/2} .$$

I.e.,

$$\dot{a} = \left(\frac{1}{2\sqrt{t_0}}\right)\frac{1}{\sqrt{t}} .$$

Again, we get that

$$\dot{a} \to 0 \ \text{ as } \ t \to \infty .$$

I.e., the rate of expansion in a flat universe slows down to a halt at the very end.

The Late Hyperbolic Universe

When $k = -1$,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{{R_0}^2 a^2} .$$

and regardless of whether it is a matter-dominated universe where $\rho = \rho_0\, a^{-3}$, or a radiation dominated universe where $\rho = \rho_0\, a^{-4}$, for late times,

$$\left(\frac{\dot{a}}{a}\right)^2 \approx \frac{c^2}{{R_0}^2 a^2} .$$

Therefore,

$$\dot{a} \approx \frac{c}{R_0} \to a(t) \approx \frac{c}{R_0} t .$$

I.e., at late times, when $k = -1$, the expansion rate reaches a non-zero constant value, unlike the $k = 0$ case.

The Late Universe in the case of a Hypersphere

Here $k = 1$. Since

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{c^2}{{R_0}^2 a^2} ,$$

somewhere along the way, regardless of whether its a matter dominated universe where $\rho = \rho_0\, a^{-3}$, or a radiation dominated universe where $\rho = \rho_0\, a^{-4}$, the expression on the right hand side will eventually become 0, implying that $\dot{a} \to 0$. Further,

$$\frac{\ddot{a}}{a} = \frac{-4\pi G}{c^2}\left[\frac{1}{3}\rho c^2 + P\right] .$$

The above expression is always negative, and therefore $a(t)$ starts decreasing again and could (as we shall see) go to zero again. This will be the case for a big crunch.

8.5.6 The general expression for a pure epoch

By definition, a pure epoch is an idealized era where there is only one type of source of gravity and the equation of state takes the form $P = w\rho c^2$ for a particular constant w. In this case, when $w \neq -1/3$, it is easy to obtain a closed for expression for a as a function of conformal time η defined by

$$\eta = \frac{c}{R_0}\int a(t)dt \ .$$

In conformal time coordinates, the FLRW metric takes the form

$$g = a^2(t)R_0^2\left[-c^2 d\eta \otimes d\eta + \left(d\chi \otimes d\chi + \frac{\sin^2(\sqrt{k}\chi)}{k}d\Omega^2\right)\right] \ .$$

Then set

$$\alpha = \frac{2}{1+3w}$$

and define A by

$$A^{1+3w} = \frac{8\pi G\rho_0}{3} \ .$$

Finally define b by

$$a = Ab^{\alpha} \ .$$

Then

$$H = \frac{1}{a}\frac{da}{d\eta}\frac{d\eta}{dt} = \frac{c}{R_0 a^2}\frac{da}{d\eta} = \frac{\alpha}{A}\frac{c}{R_0}\frac{b'}{b^{1+\alpha}} \ ,$$

where $b' = db/d\eta$. Simple algebra reduces the first Friedmann equation to the form

$$1 = \frac{c^2}{R_0^2}\left(\alpha b'^2 + kb^2\right) \ .$$

The general solution to the above equation is given by

$$b(\eta) = \frac{R_0^2}{c^2}\left[S_k(\eta/\alpha)\right]^{\alpha} \ .$$

where

$$S_k(\chi) = \frac{\sin(\sqrt{k}\chi)}{k} \ .$$

The expression for $a(\eta)$ is now easily obtained from the definition of b.

8.6 A Brief Introduction to the Physical Universe

Ordinary matter radiates when particles collide with it, and so we tend to refer to such objects as luminous matter. Astronomers observed that, away from the centers of galaxies, stars were orbiting much faster than they should if the galaxy only had the luminous matter distributed within it. For simplicity, consider a spherically symmetric distribution of stars about a galactic center. Let $M(r)$ denote the visible mass of all the stars contained in a spherical volume of radius r about the center. Then, classically, Gauss's law gives that the velocity of a star of mass m in a circular orbit at radius r is given by

$$\frac{mGM(r)}{r^2} = \frac{mv^2}{r} \ ,$$

or

$$v \sim \sqrt{\frac{M(r)}{r}} \ .$$

However, the observed data suggested that v was a constant at large r, giving a mass distribution that behaved like $M(r) \sim r$. Such a distribution of matter was contrary to observation, and this led to postulating the existence of *dark matter*. There are several other reasons that strongly suggests the existence of dark matter. But, we will not enter into such a discussion here. Suffice it to say that just like ordinary matter, w for dark matter is taken to be 0.

In order to categorize the separate ingredients of the energy and momentum content of the universe, for matter and radiation define

$$\Omega = \frac{\rho_0}{\rho_{c,_0}} \ , \ \text{ where } \ \rho_{c,0} = \frac{3H_0^2}{8\pi G} \ .$$

Also, define

$$\Omega_k = \frac{-kc^2}{R_0^2 H_0^2} \ , \ \text{ and } \ \Omega_\Lambda = \frac{1}{3}\frac{\Lambda c^2}{H_0^2} \ .$$

Then,

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{R_0^2 a^2} + \frac{1}{3}\Lambda c^2$$

and

$$H^2 = H_0^2\left[\frac{\Omega_M + \Omega_{DM}}{a^3} + \frac{\Omega_R}{a^4} + \frac{\Omega_k}{a^2} + \Omega_\Lambda\right] \ .$$

In the above equation, the subscripts M, DM, R, k, and Λ denotes ordinary matter mostly made up of baryons, dark matter, radiation, curvature constant, and cosmological constant, respectively. Evaluation the above equation at $t = t_0$, we get that

$$\Omega_M + \Omega_{DM} + \Omega_R + \Omega_k + \Omega_\Lambda = 1 \ .$$

Current estimates give that $\Omega_{DM} \approx 0.27$, and $\Omega_M \approx 0.05$. The dark energy contribution is given by $\Omega_\Lambda \approx 0.68$, and the remaining error belongs to the curvature contribution such that $|\Omega_k| \leq 0.01$. The current matter energy density is roughly 10 hydrogen atoms per meter cubed.

Looking at the inverse powers of a, we can conclude that, the early universe, as $a \to 0$, was radiation dominated. As the universe expands, it becomes more matter-dominated, and as the universe expands even further, in an ever-expanding model, dark energy or the cosmological constant term dominates. The cosmic inventory for the different epochs then reads as given in the following table.

Matter	Galaxies, Dust, Dark Matter	$w = 0$	$\rho \propto a^{-3}$
Radiation	Photons, Gravitons, massless Neutrinos	$w = \frac{1}{3}$	$\rho \propto a^{-4}$
Cosmological Constant	Λ	$w = -1$	$\rho = \frac{\Lambda c^2}{8\pi G}$
Curvature	k	$w = -\frac{1}{3}$	$\rho \propto a^{-2}$

8.6.1 The Early Universe

The energy density of radiation is given by

$$u = \rho_R c^2 = \frac{\rho_{0,R} c^2}{a^4} \ .$$

Assuming that this radiation has a blackbody spectrum, from Stefan-Boltzmann's law, we get that

$$u = \sigma T^4 \ .$$

Here σ is the Stefan-Boltzmann constant given by

$$\sigma = \frac{2\pi^5 k^4}{15c^2 h^3} \ ,$$

and k is the Boltzmann's constant. Finally, recall that, for radiation

$$a = \sqrt{\frac{t}{t_0}} \ ,$$

where

$$t_0 = \frac{t_H}{2} = \sqrt{\frac{3}{32\pi G\rho_{0,R}}} \, .$$

Combining all these equations above, we get that,

$$\frac{\rho_{0,R}c^2}{t^2}t_0^2 = \sigma T^4 \, .$$

Then

$$T = \left(\frac{3c^2}{32\pi G\sigma}\right)^{\frac{1}{4}} t^{-1/2} \, .$$

I.e., in a radiation-dominated universe

$$T \propto \frac{1}{a} \, ,$$

and

$$T_{Kelvin} = 1.52 \times 10^{10} t^{-1/2} \, .$$

8.6.2 The Cosmic Microwave Background

When photons, electrons, and protons interacted with each other they were in thermal equilibrium. But as the universe expanded, the energy of the photons redshifted sufficiently so that they could no longer ionize the molecules. As a result, stable atoms were formed, and the photons freely propagated through space. The cosmic microwave background (CMB) that we see today is a relic of this event.

The ratio of the number of photons to protons in the universe is roughly

$$\frac{N_\gamma}{N_P} \sim 10^8 - 10^{10} \, .$$

Photon decoupling would happen when there is no longer one photon per proton with the ionization energy of 13.6 eV (the ionization energy of the Hydrogen atom). The probability that a given photon has energy E at temperature T is given by $e^{-E/kT}$. A crude estimate gives that the number of photons per proton with energy greater than or equal to $13.6eV$ is

$$10^8 e^{\frac{-13.6}{kT}} \, .$$

When this quantity is less than 1, the photons would be set free. This happens when $T \sim 8000K$. A more careful analysis shows that the photon decoupling temperature is $\sim 3000K$. The observed temperature of CMB today determined from Wein's displacement law is $\sim 2.7K$ or equivalently 10^{-4} eV. The cosmological redshift of this decoupling event is then given by

$$z + 1 = \frac{\lambda_{rec}}{\lambda_{em}} = \frac{1}{a(t_{em})} = \frac{T_{decoupling}}{T_{today}} = \frac{3000}{2.7} \, ,$$

I.e., $z \sim 1100$ and $a \sim 10^{-3}$. The universe was 1000 times smaller during the decoupling epoch.

8.6.3 Matter Domination

Since the energy of a proton is $\sim 10^9$ eV, the ratio of the density of matter to the density of radiation at the present epoch is given by

$$\frac{\rho_{M,0}}{\rho_{R,0}} = \frac{10^9 \times 10}{10^8 \times 10^{-4}} \approx 10^6 \, .$$

The additional factor of 10 in the numerator accounts for the fact that dark matter is responsible for nearly 90% of the matter content. Clearly, we live in a matter-dominated universe. Then, as a function of the scale factor, the ratio of the density of matter to the density of radiation at any epoch is given by

$$\frac{\rho_M}{\rho_R} = \frac{\rho_{M_0}}{a^3}\frac{a^4}{\rho_{R_0}} \approx 10^6 a \, .$$

In the past when $a \sim 10^{-6}$, we had that $\rho_M = \rho_R$. Even further past, the universe, as expected was radiation-dominant.

8.7 Inflation

From eq.(8.11), we see that a FLRW spacetime is decelerating if

$$\left(\frac{1}{3}+w\right)\rho c^2 > 0 \ ,$$

or equivalently,

$$w > -1/3 \ .$$

Conversely, when $w < -1/3$, we have an accelerating universe.

Definition 8.1. *An FLRW epoch is said to be inflationary if $\ddot{a} > 0$, or equivalently when $w < -1/3$.*

Further, eq.(8.16) can be rewritten as

$$\left[\Omega(t)-1\right] \equiv \left[\frac{\rho}{\rho_c}-1\right] = k\left(\frac{c}{R_0\dot{a}}\right)^2 \ .$$

Current estimates suggest that, $\Omega(t_0)-1 \sim 0$, or that we live in a flat $k=0$ universe. Naturally, the question arises as to how the energy density could have been fine-tuned in this way. To consider a possible explanation, imagine a universe where $k \neq 0$, such that $\Omega(t_0)-1$ is small but not vanishing. The argument that follows will not depend on whether $k = \pm 1$. If such a universe is decelerating, $\dot{a}$ must have been much greater in the past [11]. Which mean that if $\Omega(t_0)-1$ is small, $\Omega(t)-1$ must have been much smaller (~ 0) in the past, and we run into the same problem of fine-tuning. In conclusion the condition that $\Omega(t)-1$ is small is not a very stable condition in a decelerating universe when $k \neq 0$.

One mechanism to finely tune $\Omega(t_0)-1$ to almost 0 is an inflationary epoch where $\dot{a}$ increases rapidly to make the right-hand side of the equation above vanishingly small. *An early inflationary period was thus postulated to explain the so-called "flatness" problem.* From an analysis of CMB, it is estimated that the ratios of the scale factor before and after inflation are given by

$$\frac{a(t_{\text{after}})}{a(t_{\text{before}})} = \frac{a(10^{-32}s)}{a(10^{-43}s)} \sim 10^{60} \ .$$

One imagines that after the very brief inflationary period, after which $\Omega(t)-1$ became 0, the universe underwent a phase transition to a $k=0$ state, resulting in our very special universe.

In some models, the inflation is caused by a scalar field ϕ subject to a potential of the form $V(\phi)$. Here

$$w = \frac{P}{\rho c^2} = \frac{\frac{1}{2}\dot{\phi}^2 + V(\phi)}{\frac{1}{2}\dot{\phi}^2 - V(\phi)} \ .$$

Figure 8.5 shows a possible plot for the potential energy function. In the very early universe $V(\phi)$ was very large and fluctuations in ϕ was rather minimal, resulting in

$$w = \frac{\frac{1}{2}\dot{\phi}^2 + V(\phi)}{\frac{1}{2}\dot{\phi}^2 - V(\phi)} \approx -1 \ .$$

Since w is less than $-1/3$, this would certainly result in an inflationary epoch.

In addition to solving the flatness problem, inflation also addresses several other problem in cosmology in a natural way. The existence of horizons for an observer in big bang cosmology suggests that the homogeneity of CMB radiation in all directions can be particularly puzzling, considering that patches of the universe affecting us at the current time where not in causal contact! Inflation fixes the *horizon problem* by causally connecting all relevant space at $10^{-36}s$, allowing for radiation to reach a uniform temperature.

Alan Guth and Andrei Linde originally postulated inflation to explain why we don't find magnetic monopoles and other exotic particles suggested by grand unified theories in nature. Such particles created during the pre-inflationary epoch is thought to have been scattered over an inflated space with infinitesimally small densities, and as such, we are not able to detect them. Inflation thus find a cure for the *magnetic monopole problem.*

[11] For example, this can be seen by considering a radiation-dominated early universe where $a \sim \sqrt{t}$, wherein $\dot{a} \sim t^{-1/2}$.

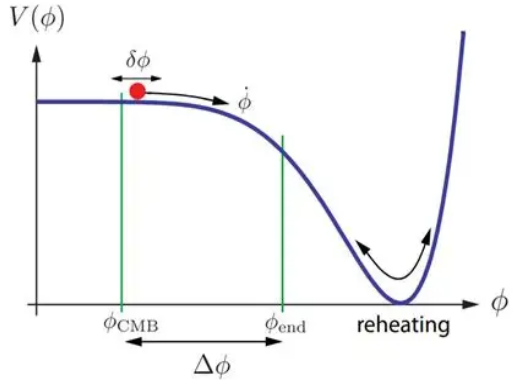


Fig. 8.5. A possible plot of $V(\phi)$ versus ϕ. [13]

The inflationary epoch can also explain the tiny anisotropies in the CMB spectrum and even the creation of all matter. Quantum fluctuations of the ϕ can result in the anisotropies. Further, as the large potential energy function $V(\phi)$ decays, the resulting universe is populated by particles to conserve energy.

Cosmological Horizons in an Inflationary Universe

Consider a $k = 0$ universe like ours. One could ask if there is a cosmological horizon in this case, meaning, are there regions of space that an immortal observer cannot observe? The answer, as we shall see, depends on the late time behaviour of the scale factor. The proper distance to the horizon, d_H is given by

$$d_H = \int_0^\infty \frac{cdt}{a(t)} \,,$$

where

$$s = \frac{2}{3(1+w)} \,.$$

Now, through different epochs, the values of w maybe different. But for a cosmological horizon to exists, all we need is that

$$\int_{t'}^\infty \frac{cdt}{a(t)} < \infty \,,$$

for some fixed late time t'. Equivalently, we need that

$$\int_{t'}^\infty t^{-s} dt < \infty \,.$$

This happens if and only if $s > 1$, or equivalently $w < -1/3$. Thus, we see that if the late flat universe is inflationary (like we think it is due to dark energy), then the universe comes with cosmological horizons.

Problem 8.16. Show that in an FLRW spacetime, an inflationary epoch has an energy-momentum tensor that does not satisfy the strong energy condition.

9

Odds And Ends

9.1 Relativistic Doppler Effect for Light

Suppose we have a distant star emitting light while moving towards the Earth with speed v. Let T_0 be the time in the frame of the star such that n complete waves are emitted. i.e.

$$\nu_0 = \frac{n}{T_0} \, .$$

Due to time dilation, an observer on Earth measures a time $T = \gamma T_0$ for the same event. If we so choose, we can think of the star as being at rest and the earth as moving a distance vT in time T. Then the frequency measured by the observer (ν) on Earth is given by

$$\nu = \frac{n + (vT/\lambda)}{T} \, .$$

Since $c = \nu\lambda$,

$$\nu = \frac{n}{T} + \frac{v}{c}\nu$$

or

$$(1-\beta)\nu = \frac{u}{\gamma T_0} = \frac{\nu_0}{\gamma} \, ,$$

and so

$$\nu = \frac{\sqrt{1-\beta^2}}{1-\beta}\nu_0 \, .$$

This finally gives us the expression we need.

$$\boxed{\nu = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}}\,\nu_0} \, . \tag{9.1}$$

Example 9.1. A red bumper sticker on a car reads, "If this bumper sticker looks green, you are driving too fast". How fast do you have to drive so that it appears green to you?

Solution For red light, $\lambda = 450$ Hz, and for green light, $\lambda = 530$ Hz. Eq. 9.1 then gives

$$\frac{530}{450} = \frac{\sqrt{1-\beta}}{\sqrt{1-\beta}} \implies \beta = 0.16 \, .$$

■

Doppler Redshift occurs when the source of light is receding. In this case, we set $\beta \to -\beta$ and obtain

$$\frac{\nu}{\nu_0} = \frac{\sqrt{1-\beta}}{\sqrt{1+\beta}}$$

or

$$\frac{\lambda}{\lambda_0} = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} \, .$$

Define the **redshift factor** z by

$$z = \frac{\lambda - \lambda_0}{\lambda_0} = \frac{\lambda}{\lambda_0} - 1 = \frac{\sqrt{1+\beta}}{\sqrt{1-\beta}} - 1 \, . \tag{9.2}$$

Problem 9.1. In the non-relativistic limit, show that

$$z \approx \beta \, .$$

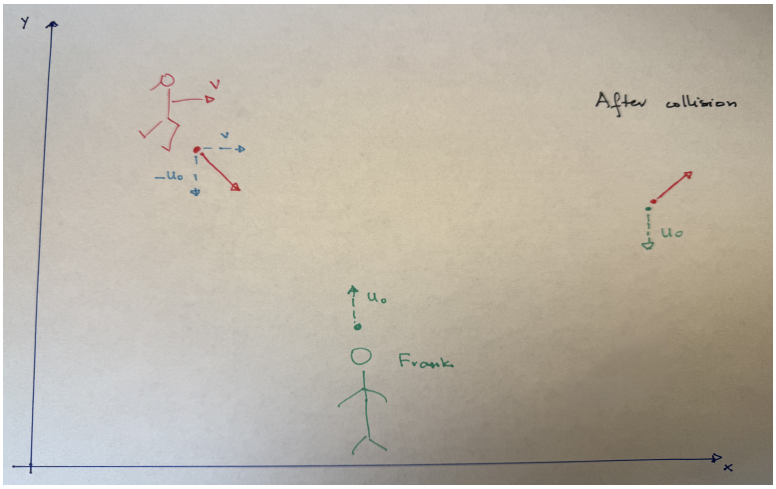


Fig. 9.1. The red and green balls eventually collides and scatters back as shown.

9.2 Motivating Relativistic Momentum

This problem is directly taken from [14], and it provides physical insight into why momentum should contain the γ factor.

Frank and Mary are in the voids of space, i.e., far away from any gravitational influence, and Mary is traveling at speed v with respect to Frank along their mutual x direction. Frank throws a mass m in the $+y$ with speed U_0. Mary throws a similar mass m in the $-y$ direction at speed U_0. The arrangement is such that the two masses collide elastically and return back at the same speed (from our experience in classical mechanics). See fig. (9.1).

In Frank's frame, say S, **before bollision**

$$U_{Fx} = 0 \text{ , and } U_{Fy} = U_0 \text{ ,}$$

and **after collision**

$$U_{Fx} = 0 \text{ , and } U_{Fy} = -U_0 \text{ .}$$

In Mary's frame, say $\bar{S}$: **before collision**

$$\bar{U}_{Mx} = 0 \text{ , and } \bar{U}_{My} = -U_0 \text{ ,}$$

and **after collision**

$$\bar{U}_{Mx} = 0 \text{ , and } \bar{U}_{My} = U_0 \text{ .}$$

Now let's transform Mary's quantities into Frank's frame in order to compare and check for the expected law of conservation of momentum. Using the inverse transform as per problem 3.1, we get that in

S **before collision**

$$U_{Mx} = v, \text{and } U_{My} = -\frac{U_0}{\gamma} = -U_0\sqrt{1-\beta^2}$$

and **after collision**

$$U_{Mx} = v, \text{and } U_{My} = \frac{U_0}{\gamma} = U_0\sqrt{1-\beta^2} \text{ .}$$

Here $\beta = v/c$ as usual. If we assume the usual expression for momentum, namely

$$\mathbf{p} = m\mathbf{u} \text{ ,}$$

then

$$\Delta\mathbf{p}_F = 0\ \mathbf{e}_x + -2mU_0\ \mathbf{e}_y \text{ ,}$$

and

$$\Delta\mathbf{p}_M = 0\ \mathbf{e}_x + 2mU_0\sqrt{1-\beta^2}\ \mathbf{e}_y\ .$$

Therefore, the change in total momentum

$$\Delta\mathbf{p}_{Total} = \Delta\mathbf{p}_F + \Delta\mathbf{p}_M = \left(2mU_0\sqrt{1-\beta^2} - 2mU_0\right)\mathbf{e}_y \neq 0\ .$$

This means that we have to either give up the law of conservation of momentum or modify the expression for momentum for the case of relativistic particles. Consequently, and with hindsight, define the relativistic 3-momentum of a particle with mass m to be

$$\mathbf{p} = \frac{m\mathbf{u}}{\sqrt{1-\frac{u^2}{c^2}}}\ ,$$

where $u^2 \equiv \mathbf{u}\cdot\mathbf{u}$. For ease of notation, define γ in this context [1] to be

$$\gamma = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}\ ,$$

and so the relativistic 3-momentum of a particle with mass m is given by

$$\boxed{\mathbf{p} = m\gamma\mathbf{u}}\ . \tag{9.3}$$

Using the new expression for momentum, for Frank we get that

$$\Delta\mathbf{p}_F = \frac{-2mU_0}{\sqrt{1-\frac{U_0^2}{c^2}}}\ \mathbf{e}_y\ .$$

For Mary, noting that

$$\mathbf{u} = v\ \mathbf{e}_x \pm U_0\sqrt{1-\beta^2}\ \mathbf{e}_y\ ,$$

the momentum before collision is given by

$$\frac{mv}{\sqrt{1-\frac{v^2+U_0(1-\beta^2)}{c^2}}}\ \mathbf{e}_x - \frac{mU_0\sqrt{1-\beta^2}}{\sqrt{1-\frac{v^2+U_0(1-\beta^2)}{c^2}}}\ \mathbf{e}_y\ ,$$

and after collision the momentum becomes

$$\frac{mv}{\sqrt{1-\frac{v^2+U_0(1-\beta^2)}{c^2}}}\ \mathbf{e}_x + \frac{mU_0\sqrt{1-\beta^2}}{\sqrt{1-\frac{v^2+U_0(1-\beta^2)}{c^2}}}\ \mathbf{e}_y\ .$$

Therefore,

$$\Delta\mathbf{p}_M = \frac{2mU_0\sqrt{1-\beta^2}}{\sqrt{1-\frac{v^2+U_0(1-\beta^2)}{c^2}}}\ \mathbf{e}_y = \frac{2mU_0}{\sqrt{\frac{1}{1-\beta^2}-\frac{v^2}{(1-\beta^2)c^2}-\frac{U_0^2}{c^2}}}\ \mathbf{e}_y = \frac{2mU_0}{\sqrt{1-\frac{U_0^2}{c^2}}}\ \mathbf{e}_y\ .$$

Finally, as required, we have that

$$\Delta\mathbf{p}_F + \Delta\mathbf{p}_M = 0\ ,$$

and so the law of conservation of momentum is restored in this case all is well with the universe once again! It is important to note that one example may not give a unique correct expression for momentum. It only paves the way for a good guess which must be justified later by more sophisticated means.

9.3 The Geodesic Equation

In the case of a Riemannian metric, where the metric is positive definite, the length of a curve $x^\mu(s)$ for $s_i \leq s \leq s_f$ is defined by

$$\int_{s_i}^{s_f} \sqrt{g_{\mu\nu}\frac{dx^\mu}{ds'}\frac{dx^\nu}{ds'}}ds'\ .$$

[1] Despite appearances, γ as used here and as a factor in Lorentz transformations have completely different meaning. Here $\mathbf{u}$ is the speed of the particle under consideration, while in the case of a Lorentz transformation, the v contained in the expression for γ is the relative speed of the observers in the two inertial frames S and $\bar{S}$.

A curve minimizes length when the integral above is minimized. This is equivalent to minimizing

$$\int_{s_i}^{s_f} g_{\mu\nu} \frac{dx^\mu}{ds'} \frac{dx^\nu}{ds'} ds' \ .$$

The standard technique from classical mechanics sets the Lagrangian as

$$L = g_{\mu\nu}(x(s)) \frac{dx^\mu}{ds} \frac{dx^\nu}{ds} \equiv g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu \ .$$

The Euler-Lagrange equations are given by

$$\frac{d}{ds} \left(\frac{\partial L}{\partial \dot{x}^\lambda} \right) - \frac{\partial L}{\partial x^\lambda} = 0 \ . \tag{9.4}$$

Then, since

$$\frac{\partial L}{\partial x^\lambda} = \frac{\partial}{\partial x^\lambda} \left(g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu \right) = g_{\mu\nu,\lambda} \ \dot{x}^\mu \dot{x}^\nu \ ,$$

and

$$\frac{\partial L}{\partial \dot{x}^\lambda} = 2 g_{\mu\nu} \frac{\partial \dot{x}^\mu}{\partial \dot{x}^\lambda} \dot{x}^\nu = 2 g_{\lambda\nu} \dot{x}^\nu \ ,$$

we get that

$$\frac{d}{ds} \left(g_{\lambda\nu} \dot{x}^\nu \right) - \frac{1}{2} g_{\mu\nu,\lambda} \dot{x}^\mu \dot{x}^\nu = 0 \ .$$

This gives

$$g_{\lambda\nu} \ddot{x}^\nu + g_{\lambda\nu,\mu} \dot{x}^\mu \dot{x}^\nu - \frac{1}{2} g_{\mu\nu,\lambda} \dot{x}^\mu \dot{x}^\nu = 0 \ ,$$

or

$$\ddot{x}^\sigma + \frac{g^{\sigma\lambda}}{2} \left[g_{\lambda\nu,\mu} + g_{\lambda\mu,\nu} - g_{\mu\nu,\lambda} \right] \dot{x}^\mu \dot{x}^\nu = 0 \ .$$

This above equation is called the geodesic equation, since it minimizes length in the variational sense.

References

1. G. Naber, *The Geometry of Minkowski Spacetime*, 2nd ed., ser. Applied Mathematical Sciences (Book 92). Springer-Verlag New York, Inc., 2011.
2. E. C. Zeeman, "Causality implies the lorentz group," *Journal of Mathematical Physics*, vol. 5, no. 4, pp. 490–493, 1964. [Online]. Available: https://doi.org/10.1063/1.1704140
3. B. O'Neill, *Semi-Riemannian geometry. With applications to relativity*, 1st ed., ser. Pure and Applied Mathematics. Academic Press, 1983.
4. L. Ryder, *Quantum Field Theory*, 2nd ed. Cambridge University Press, England, 1996.
5. S. Nanda, "A geometrical proof that causality implies the lorentz group," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 79, no. 3, p. 533–536, 1976.
6. S. Weinberg, *Graitation and Cosmology*. John Wiley & Sons, Inc., USA, 1972.
7. W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. Mc Graw Hill, India, 2013.
8. I. I. Shapiro, "Fourth test of general relativity," *Phys. Rev. Lett.*, vol. 13, pp. 789–791, Dec 1964. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.13.789
9. R. P. Kirshner, "Hubble's diagram and cosmic expansion," *Proceedings of the National Academy of Sciences*, vol. 101, no. 1, pp. 8–13, 2004. [Online]. Available: https://www.pnas.org/content/101/1/8
10. P. J. E. Peebles and B. Ratra, "The cosmological constant and dark energy," *Rev. Mod. Phys.*, vol. 75, pp. 559–606, Apr 2003. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.75.559
11. V. Mukhanov, *Physical Foundations of Cosmology*, 1st ed. Cambridge University Press, 2005.
12. A. V. Marcus Spradlin, Andrew Strominger, "Les houches lectures on de sitter space," *hep-th/0110007*, 2001. [Online]. Available: https://arxiv.org/abs/hep-th/0110007
13. T. Mukherjee, "Generation of primary and secondary gravitational waves in the early universe," Master's thesis, Indian Institute of Technology, Madras, 05 2022.
14. J. Marion and S. Thornton, *Classical Dynamics of Particles and Systems*, 5th ed. Cengage Learning India, 2012.