Preface

General Relativity Lecture notes
Govind Menon

These are the lecture notes for Relativity I (PHY 4460/5560) and II (PHY 4478/5578) that are 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 have explicitly shown all the fundamental constants, including c. If this bothers you, please pretend you do not "see" it.

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 15
- 2.1 The Rotation Group 16
- 2.2 Classes of Inertial Observers 21
- 2.3 Minkowski Spacetime 22
- 2.4 The Lorentz Group 26
- 2.5 Causal Structure of Minkowski Spacetime 33

References 41

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

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

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.,

$$\begin{aligned} \bar{t} &= t \\ \bar{x} &= x - vt \\ \bar{y} &= y \\ \bar{z} &= z \end{aligned} \qquad . \tag{1.2}$$

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

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

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

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

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

where

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

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

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

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

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

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

$$ct = \beta x \ ,$$

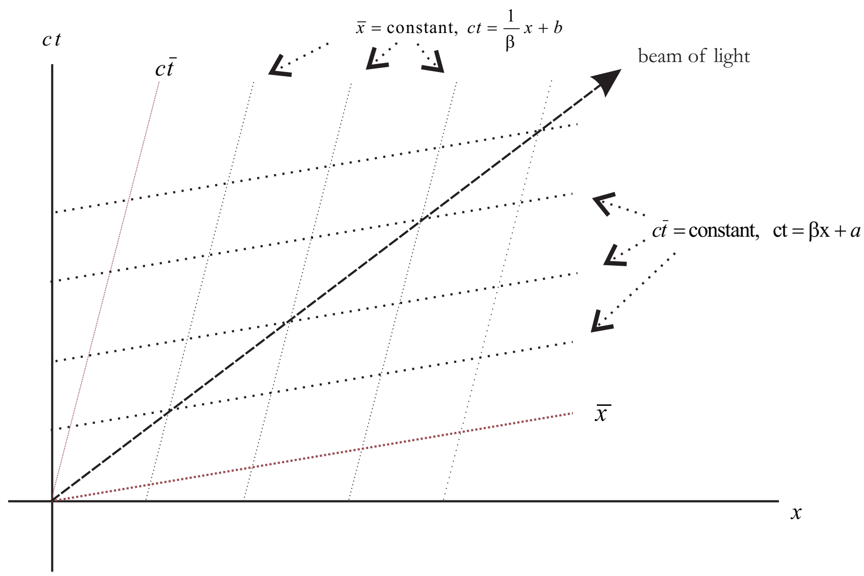

Fig. 1.1. Slanted dashed lines are lines of constant $c\bar{t}$ and $\bar{x}$. Notice how the beam of light has the same speed in both frames.

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

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

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 $\tau = \bar{t}$. In any other inertial frame, the duration for the event will be larger than τ by a factor of γ, i.e.,

$$\boxed{T = \gamma\,\tau} \,.$$

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.

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

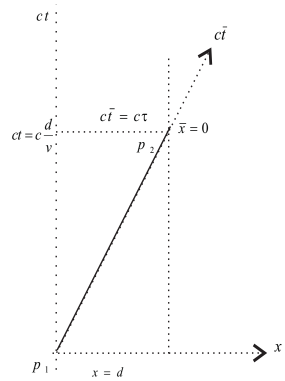

Fig. 1.2. The solid represents the path taken by the rocket ship.

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

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

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)

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

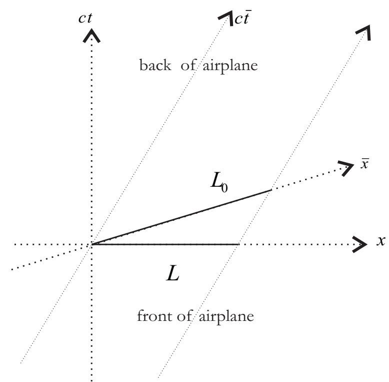

Fig. 1.3. The solid lines show the proper length L_0, and L is the length as measured in the frame S.

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:

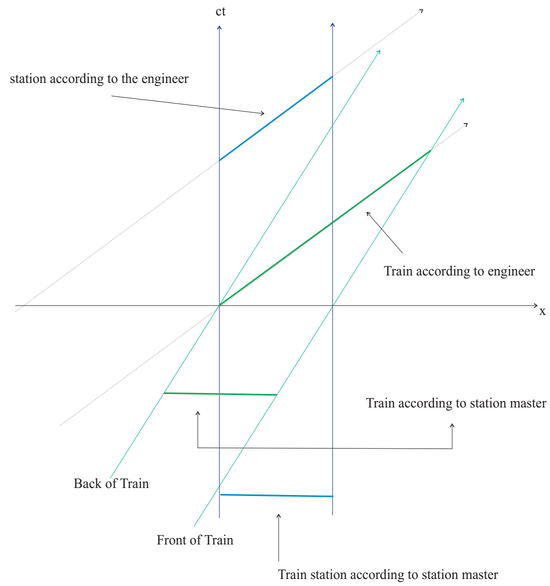

Fig. 1.4. The station is shown in blue, and the train is shown in green.

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. □

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).

[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].

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

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

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

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

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.,

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

---

[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.

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 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 bases 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 bases vectors transform, we will specify R by its defining equations on an orthonormal bases:

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

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 are 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 bases vectors change exactly the same way, i.e.,

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

---

[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.

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 satisfies 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

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

---

[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) are 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) 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 have ensured that the speed of light is indeed preserved under 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, wherein 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 labelled 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 object 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 have constructed four distinct classes of inertial observers. For compactness in notation, we can combine all of transformations in these inertial 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).

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

---

[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

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

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

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

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

Here X_x is the *abstract notation* for a vector, while in the *index notation* we write the same 4-vector as X_x^μ. 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_x^2 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

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

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*.*

---

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

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 $\mathbb{M}^4$:

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

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,

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

[9] Since

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

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

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

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

$$a = \Lambda^0{}_0 \; , \; b = \Lambda^0{}_1 c_1 + \Lambda^0{}_2 c_2 + \Lambda^0{}_3 c_3 \; ,$$

$$c = a_{10} \; , \text{ and } \; d = a_{11} c_1 + a_{12} c_2 + a_{13} c_3 \; .$$

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 $\mathbb{T}\mathbb{P}\Lambda \in \mathcal{L}_+^\uparrow$. Therefore $\Lambda = \mathbb{P}\mathbb{T}\Lambda_+^\uparrow$ 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.

Definition 2.10. *The set of all vector X_x, for any $x \in \mathbb{M}^4$, clearly forms a vector space at x and is denoted by $T_x(\mathbb{M}^4)$ and is referred to as the tangent space at x.*

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

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

implies

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

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.

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 $A^2 > 0$.*

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*

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

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

etc.

- *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.*

Definition 2.14. *A vector in $\mathbb{M}^4$ is causal if it is time-like or null.*

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 2.15. *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

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

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*

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

or

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

Proof. If $T^0W^0 = 0$, 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 $T^0T^0 > 0$ or $T^0W^0 < 0$.

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,

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

i.e.,

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

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

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

Therefore, $g(T,W) < 0$. If $T^0W^0 < 0$, 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 2.16. *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.*

Definition 2.17. *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 $(\mathbb{M}^4, \mathcal{T})$.
c. $\mathcal{T}$ fixes the direction of future in $(\mathbb{M}^4, \mathcal{T})$.

Definition 2.18. *In* $(\mathbb{M}^4, \mathcal{T})$*, a 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 From definition 2.17 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.

Remark 2.3. 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, show that $V + W$ is future-pointing timelike.

Zeeman's Theorem

Definition 2.19. *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 2.20. *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 speed of light is constant and the fact that light flows into the future is necessarily an inertial transformation generated by the Poincare group, and none other. Our quest for inertial transformations are in fact over. And, most importantly, since Poincare 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 of the Lorentz transformation has proven to be the correct one under boosts. There is no other way speed of light can be preserved, along with causality, without mixing space and time!

[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

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

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

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.