INTRODUCTION
TO
THE SPECIAL THEORY
OF
RELATIVITY

Tai Cho (Thatcher) Lai

January 22, 2024

Abstract

This paper explores the basic principles of the special theory of relativity, formulated and developed mainly by physicists including but not only Albert Einstein, Hendrik Lorentz, Hermann Minkowski, and Henri Poincaré. Concepts such as Galilean transformations, Lorentz transformations, time dilation, length contraction, and tensors will be explored. This paper also discusses Maxwell's equations and their implications for special relativity.

Contents

1 Basics of Spacetime and Galilean Relativity **3**
1.1 Reference Frames and Coordinate Transformations 3
1.2 Spacetime Diagrams and Galilean Transformations 5

2 The Special Theory of Relativity **10**
2.1 The Invalidity of Galilean Relativity and Einstein's Postulates 10
2.2 Relativity of Simultaneity 12
2.3 Derivation of the Lorentz Transformation Equations 16
2.4 Time Dilation 20
2.5 Length Contraction 23
2.6 The Spacetime Interval and Light Cones 27
2.7 4-Vectors and the Minkowski Metric Tensor 32
2.8 The Lorentz Transformation Matrix 42

3 Electrodynamics **44**
3.1 Maxwell's Equations 44
3.2 Derivation of the Constancy of the Speed of Light 45
3.3 Lorentz Force and the Electromagnetic Tensor 47

A Matrix Multiplication **57**

B Vector Operations **59**
B.1 The Dot Product 59
B.2 The Cross Product 60

C Tensor Operations **60**
C.1 The Tensor-Vector Product 60
C.2 Tensor Contraction and Raising and Lowering Indices 63

D Vector Calculus Operators **64**
D.1 The Del Operator 64
D.2 The Gradient Operator 64
D.3 The Divergence Operator 64
D.4 The Curl Operator 65
D.5 The Laplace Operator 65

1 Basics of Spacetime and Galilean Relativity

1.1 Reference Frames and Coordinate Transformations

Reference frames form the foundation of special relativity and are crucial to understanding its associated principles and transformations. A *reference frame* is an abstract coordinate system defined or characterised by mathematically and physically defined geometric points with a non-changing orientation, scale, and origin. A reference frame is used to define properties such as velocity, position, and distance with respect to an observer within that frame. An inertial reference frame is a reference frame where there is no acceleration and Newton's first law holds. We will now abbreviate the terms "reference frame" and "inertial reference frame" as "RF" and "IRF", respectively

When we change from one RF to another, we are applying a *transformation*. There are different types of transformations, including rotation, translation, and shearing. Let us first consider a 2D Cartesian coordinate system representing an RF $\mathcal{S}$ on which a point P is plotted, with O denoting the origin with coordinates $(0, 0)$, like so:

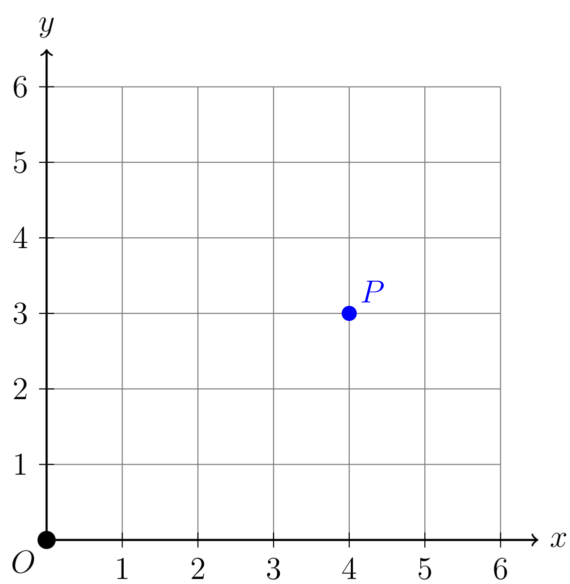

Figure 1: A 2D Cartesian coordinate system representing the RF $\mathcal{S}$ with a point P.

We may specify the point P with coordinates $(4, 3)$ — this tells us the *position* of P on a Cartesian plane with respect to the origin in $\mathcal{S}$. However, if we rotate the Cartesian plane clockwise by a certain number of degrees to form a new RF $\mathcal{S}'$, we will not get the same coordinates for P. Figure 2 illustrates this.

We can see from Figure 2 that P has the coordinates $(0, 5)$ in $\mathcal{S}'$, which is different from the coordinates of P in $\mathcal{S}$. Therefore, we can see that position is relative and not absolute under a rotation transformation. However, we can see that the distance between the origin and P is absolute. To calculate this, we can use the formula

$$\ell = \sqrt{x^2 + y^2}, \tag{1.1}$$

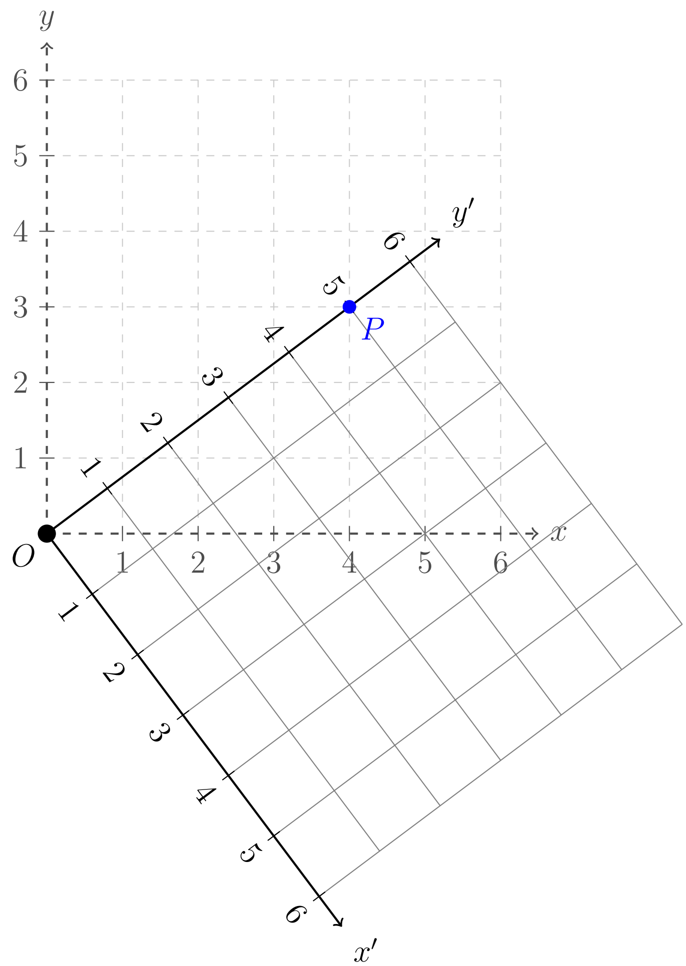

Figure 2: A 2D Cartesian coordinate system representing the primed $\mathcal{S}'$ (darker) and unprimed $\mathcal{S}$ (lighter) RFs with a point P.

where ℓ denotes the distance between O and P, x denotes the x-coordinate of P with respect to O in an RF, and y denotes the y-coordinate of P with respect to O in the same RF. Plugging in the coordinates for P in $\mathcal{S}$ and $\mathcal{S}'$, respectively, we get

$$\ell = \sqrt{4^2 + 3^2} = 5$$

in $\mathcal{S}$ and

$$\ell' = \sqrt{0^2 + 5^2} = 5$$

in $\mathcal{S}'$. Hence, we can see that the distance between O and P, as well as that between any two points we so choose on a Cartesian coordinate system, remains absolute under a rotation transformation. We say that distance, under a rotation transformation, is invariant. It is also invariant under a translation transformation (where we slide or move the coordinate system). An *invariant* is a quantity that remains absolute or unchanged after some coordinate transformation.

Note that properties such as the invariance of distance and relativity of position also apply to other types of coordinate systems, such as polar and cylindrical coordinate systems. However, we will mainly be dealing with Cartesian coordinate systems in this paper for convenience.

1.2 Spacetime Diagrams and Galilean Transformations

The study of relativity (Galilean and special) deals with what events look like and how objects appear to move or behave in different IRFs (or from the perspectives of different observers) as time progresses. Hence, we need a way to represent this. This is where a *spacetime diagram* comes in — this is an abstract coordinate system that allows us to plot the movement and appearance of different objects and events within an arbitrary RF. More specifically, a spacetime diagram is any graph that represents various objects' positions through time. We can compare it to a flipped position-time graph, where a horizontal axis represents position x (instead of time t), a vertical axis represents time t (instead of position x), and O represents the origin. This is shown in Figure 3.

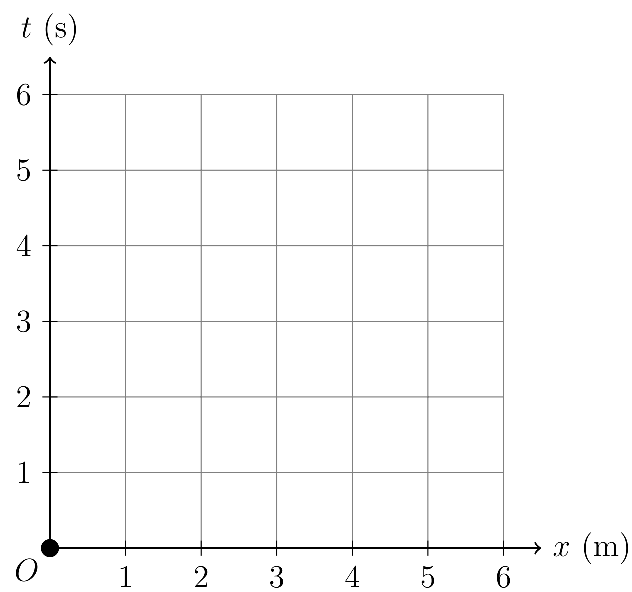

Figure 3: A 2D spacetime diagram.

Before special relativity, there was Newtonian or Galilean relativity. This theory of relativity was built upon the idea of *absolute space* and *universal time*, that is, Galilean relativity relies on the postulates that all objects are either moving or completely at rest with reference to absolute space (i.e. absolute motion) and that time ticks with equal pace for any observer in any IRF. Galilean relativity also states that *the laws of motion are identical in any IRF*. Finally, it states that *all objects in an IRF can move relative to each other in any form under uniform motion* (i.e. motion in a straight line with constant speed), implying that the speed of light can be surpassed — however, as we will discover in special relativity, this is not the case.

As mentioned in Section 1.1, changing between different RFs involves applying a transformation. It therefore follows that a *Galilean transformation* is applied to change between different IRFs in Galilean relativity. We often deal with relative motion between objects with constant velocities in Galilean relativity, hence, only spatial coordinates (in our 2D case, the x coordinate) vary under a Galilean transform.

To illustrate, consider a hypothetical scenario set in a park. Let us assume that a man is standing motionless relative to the ground; we will denote the man as A. Let us also assume that a cat, which we will denote as B, is moving away from A at a constant velocity v[1] of $0.5\,\mathrm{m/s}$ to the right relative to A. Using a spacetime diagram, we can plot the positions of A and B as time goes by. We will measure position x in metres and time t in seconds.

Let us first consider the spacetime diagram in A's IRF, which we will denote as $\mathcal{S}$. We can see that A is always at $x = 0$ for whatever value of t since A does not appear to be in motion with reference to himself. From A's perspective, at $t = 1$, B is at $x = 0.5$; similarly, at $t = 2$, B is at $x = 1$. By following this pattern, we see from Figure 4 that the points (represented by a series of dots) connect to form a line, representing B's motion at a velocity v of $0.5\,\mathrm{m/s}$ to the right relative to A. The line is called a *worldline*, which is a curve or path in spacetime that an object traces. We can hence see from above and Figure 4 that B's wordline in $\mathcal{S}$ is described by the equation

$$x = vt. \tag{1.2}$$

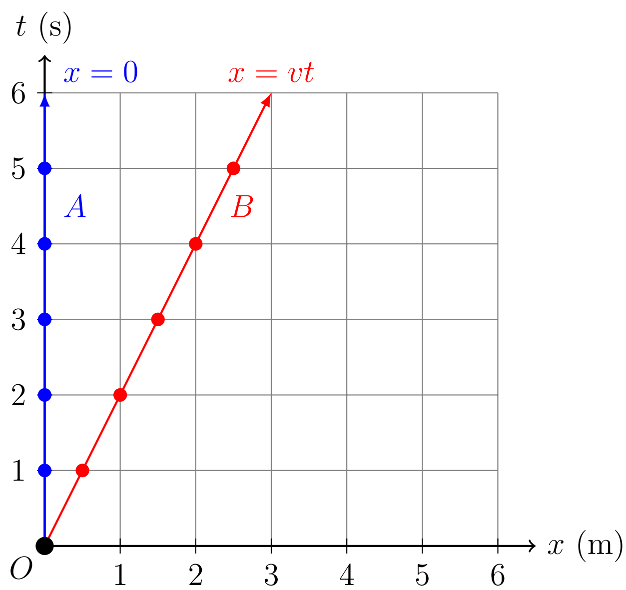

Figure 4: Worldlines of A and B in A's IRF $\mathcal{S}$.

[1]Throughout this section on Galilean relativity, we will denote velocities without making them bold (which is contrary to the convention for denoting vectors in physics). While velocity in general is a vector, since we are only dealing with one dimension of space and uniform motion here, we may treat velocity as a scalar. As we move onto special relativity and introduce more dimensions, in some cases, we will use proper denotations for vectors and scalars.

Let us now consider the spacetime diagram from B's perspective, in which a new pair of axes describe the motion of A and B through spacetime in B's IRF, which we will denote as $\mathcal{S}'$. It is evident that B always remains motionless relative to itself. Thus, using a new pair of axes (the x'- and t'-axis, representing position and time in $\mathcal{S}'$, respectively), we see from Figure 5 that B is always at $x' = 0$ for any value of t'. Here, we trace out B's worldline in $\mathcal{S}'$. Since from A's perspective, B moves away from him at 0.5 m/s to the right, it follows that from B's perspective, A moves away from it at 0.5 m/s to the left. Therefore, we can see that at $t' = 1$, A is at $x' = -0.5$; at $t' = 2$, A is at $x' = -1$. Following this pattern, we trace out A's worldline in $\mathcal{S}'$. This is also shown in Figure 5. We can hence see from above and Figure 5 that A's wordline in $\mathcal{S}'$ is described by the equation

$$x' = -vt'. \tag{1.3}$$

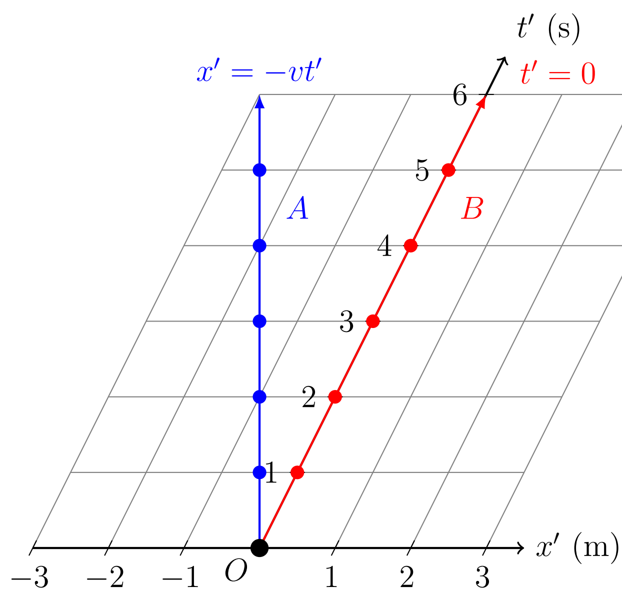

Figure 5: Worldlines of A and B in B's IRF $\mathcal{S}'$.

Note that a dot or point on a spacetime diagram that represents its present position and time away from the origin is called an *event*. Also, note that from Figures 4, 5, and 6, we can see that when we choose whose RF and worldline we are considering, the worldline always lines up with the time axis (i.e. in their own IRF, their position is always at $x = 0$). This must be true because any moving object is always motionless with respect to itself.

From Figures 4 and 5, we see that the change in position between each corresponding event at the same time from A's and B's worldlines, which we will denote as Δx, increases with time according to their relative and constant velocity v. For instance, in both IRFs $\mathcal{S}$ and $\mathcal{S}'$, at $t' = t = 1$, $\Delta x = 0.5$; at $t' = t = 2$, $\Delta x = 1$, etc. Hence, another way we can transform $\mathcal{S}$ into $\mathcal{S}'$ is by shifting both A's and B's worldlines to the left, allowing B's worldline to lie on the vertical t'-axis, as shown in Figure 6 — that is, with a positive relative velocity v associated with a left shift. Both ways of transforming between IRFs given a constant velocity are valid; it does not matter as long as the worldline under question lies on the time axis and the other worldlines are shifted accordingly. However, the convention in special relativity is to shift the coordinate system (representing its corresponding IRF) to the right with a positive v, leaving the worldlines fixed.

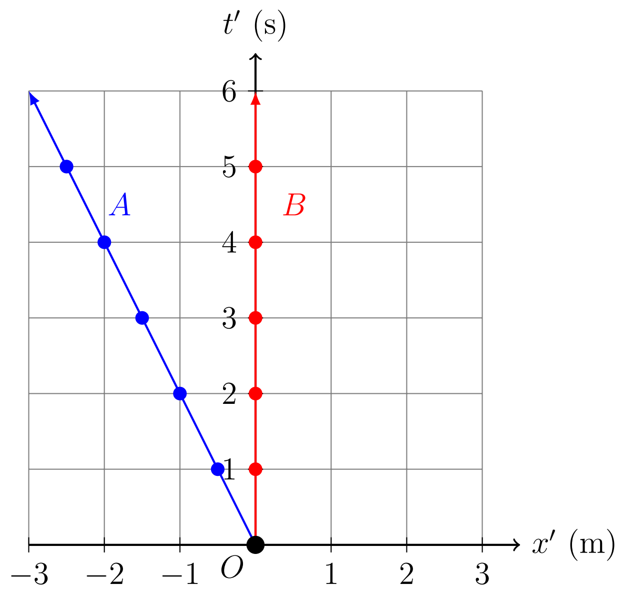

Figure 6: Worldlines of A and B in B's IRF $\mathcal{S}'$.

We now know what a Galilean transformation looks like given a constant relative velocity v and two moving objects each in an IRF. With all the information above, we can formulate eight equations that describe a Galilean transformation from an unprimed IRF to a primed IRF and back.[2] Taking the hypothetical park scenario described above, it is evident that

$$t' = t, \tag{1.4}$$

where t' and t are the time coordinates for a given event in $\mathcal{S}'$ and $\mathcal{S}$, respectively, since for any two objects moving relative to each other at some constant velocity each in an IRF, time runs at the same pace — this relates to the concept of universal time in Galilean relativity. We can also see that

$$x' = x - vt, \tag{1.5}$$

where x' is the x-coordinate in $\mathcal{S}'$, x is the x-coordinate in $\mathcal{S}$, and v is the relative velocity between two objects each in an IRF. Equation 1.5 is true because an object with its motion described by coordinates in $\mathcal{S}'$ is displaced from another object with its motion described by coordinates in $\mathcal{S}$ by their relative velocity multiplied by the time elapsed. We may also think of Equation 1.5 as the "old" coordinates in the unprimed frame "becoming" a new coordinate under a Galilean transformation, in which the set of all such coordinates under

[2]The terms "primed" and "unprimed" used throughout this paper refer to coordinates, RFs, or other mathematical objects alike that have and have not undergone a coordinate transformation, respectively.

the same transformation make up a new primed coordinate system or IRF. Regarding the two other axes of space, that is, the y- and z-axis, we also consider them to be equal to their primed counterparts, as in

$$y' = y \tag{1.6}$$

and

$$z' = z. \tag{1.7}$$

Together, Equations 1.4, 1.5, 1.6, and 1.7 describe a Galilean transformation from an unprimed IRF to a primed IRF.

We may also represent a Galilean transformation in 2D (that is, only with the x and t coordinates) from an unprimed IRF to a primed IRF using matrix notation, which is[3]

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} 1 & -v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} x - vt \\ t \end{pmatrix}. \tag{1.8}$$

If we wish to transform from a primed IRF back to an unprimed IRF, we simply add vt to both sides of Equation 1.5, giving

$$x = x' + vt, \tag{1.9}$$

and leave all other components the same, that is,

$$y = y', \tag{1.10}$$

$$z = z', \tag{1.11}$$

and

$$t = t'. \tag{1.12}$$

Together, Equations 1.9, 1.10, 1.11, and 1.12 describe an *inverse Galilean transformation*, which is a Galilean transformation from a primed IRF to an unprimed IRF.

We can also represent an inverse Galilean transformation in 2D from a primed IRF to an unprimed IRF in matrix notation, which is

$$\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} x' + vt' \\ t' \end{pmatrix}. \tag{1.13}$$

In Galilean relativity, velocity is also additive. To illustrate, consider three objects, A, B, and C, each in an IRF. Let us say that A is a man standing motionless relative to the

[3] See Appendix A for matrix multiplication.

ground on a road. On the same road is B, a driver in a car moving away from A at a constant velocity v of $10\,\mathrm{m/s}$ to the right relative to A. B then throws out a ball C moving away from B at a constant velocity u of $5\,\mathrm{m/s}$ to the right relative to B. Since velocity is additive, we can see that the velocity u' at which the ball travels relative to A is given by

$$u' = v + u = 10\,\mathrm{m/s} + 5\,\mathrm{m/s} = 15\,\mathrm{m/s}.$$

2 The Special Theory of Relativity

2.1 The Invalidity of Galilean Relativity and Einstein's Postulates

The core of special relativity involves two postulates proposed by Einstein. *Einstein's first postulate* is that *the laws of physics are identical in any IRF*. For instance, if you were to throw a ball up and then catch it, the laws of physics that describe this event will always be the same whether you are standing motionless relative to the ground, sitting on a bus moving at some constant velocity, or walking in a park at some fixed velocity. This is because all such events take place in an IRF.[4]

We have seen that in Galilean relativity, velocity is additive. So let us now consider another hypothetical scenario where we have three objects, A, B, and C, each in an IRF. Let us say that A is a man in a train station standing motionless relative to the platform. Let us also suppose that a passenger B holding a torch is sitting in a train and moving away from A at a constant velocity v of $1 \times 10^8\,\mathrm{m/s}$ to the right relative to A.[5] Finally, a beam of light C is shone from B's torch which moves away from B at a constant velocity u of $3 \times 10^8\,\mathrm{m/s}$ to the right relative to B.[6] In Galilean relativity, the velocity u' at which the beam of light travels relative to A is given by

$$u' = v + u = 1 \times 10^8\,\mathrm{m/s} + 3 \times 10^8\,\mathrm{m/s} = 4 \times 10^8\,\mathrm{m/s}. \tag{2.1}$$

We can plot a spacetime diagram in A's and B's respective IRFs to see how the motion of the beam of light appears to each of them, in which position will be measured in $x \times 10^8$ metres. Figures 7 and 8 illustrates this.

One might think that Equation 2.1 along with Figures 7 and 8 correctly describe the relationship between the relative motion of C to A and B, but they would be wrong. They would be wrong because the speed of light (in free space or a vacuum) is always equal to $c = 3 \times 10^8\,\mathrm{m/s}$ to any inertial observer. This is known as *Einstein's second postulate*, which states that *the speed of light in free space remains invariant under transformations between IRFs.*

[4]We will assume that gravitational acceleration is equal in all hypothetical scenarios given here.

[5]Since this is a hypothetical scenario, yes, we will assume that the passenger and train can reach such a velocity.

[6]To be pedantic, it should be $299\,762\,458\,\mathrm{m/s}$, but we will approximate it and write $3 \times 10^8\,\mathrm{m/s}$.

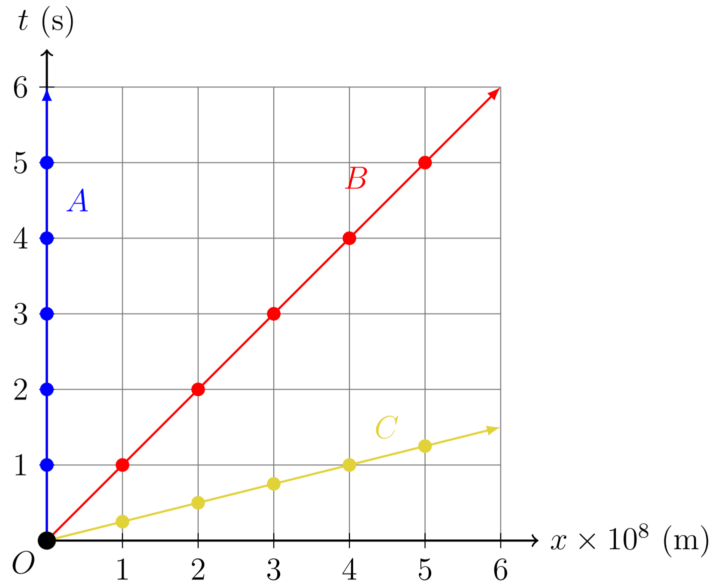

Figure 7: Worldlines of A, B, and C in A's IRF $\mathcal{S}$.

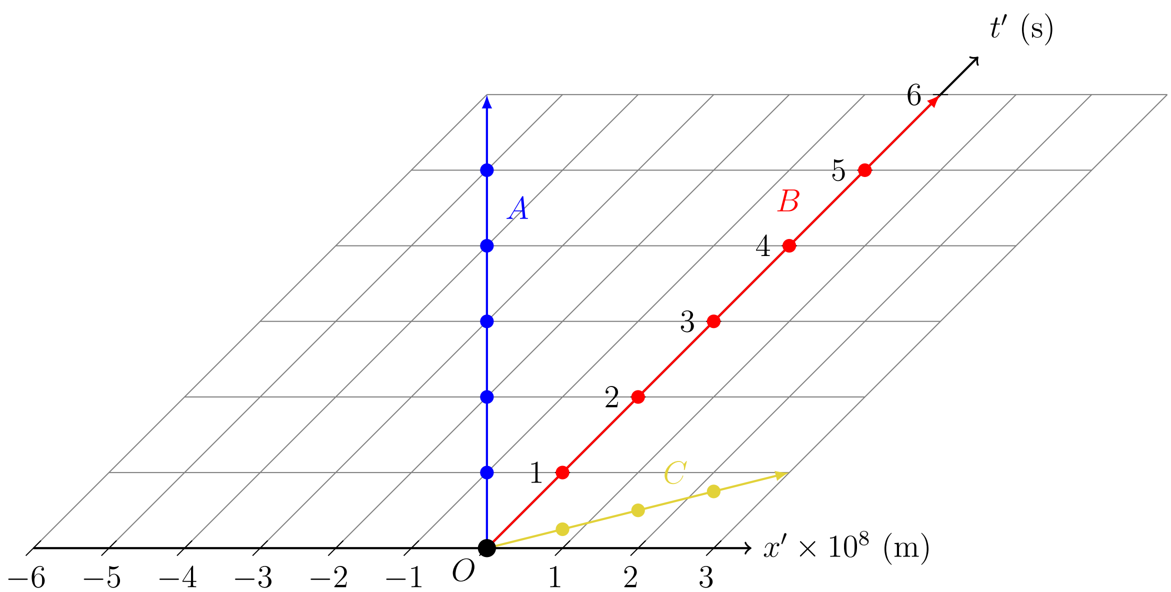

Figure 8: Worldlines of A, B, and C in B's IRF $\mathcal{S}'$.

There are quite a few experimental and theoretical pieces of evidence for the constancy of the speed of light in any IRF. One of the most famous is the Michelson-Morley experiment. It was an experiment performed in 1887 by physicists Albert A. Michelson and Edward W. Morley in an attempt to measure the relative motion of the Earth to the hypothetical "luminiferous ether". The physicists found no difference in the speed of light, whether measured in the direction of Earth's motion or other directions. This disproved the existence of a luminiferous ether and contradicted Newtonian physics.

From this, we can see that Galilean transformations do not accurately describe the motion of objects in IRFs at velocities or speeds close to the speed of light. Hence, we require a better and more accurate description of the relationship between transformations of IRFs.

2.2 Relativity of Simultaneity

We have seen that the speed of light does not appear to be identical in different IRFs when we apply Galilean transformations. Therefore, we require another kind of transformation that allows the speed of light to remain the same no matter the IRF.

Before we continue, let us reconsider the units we will use for our spacetime diagrams. Assuming time t to be measured in seconds, if position x were to be measured in metres, then the wordline of a rightward-travelling light beam would appear to be almost entirely horizontal — this is not what we want. In special relativity, the convention is for the worldline of a light beam to travel 1 unit of distance in 1 unit of time. Hence, what we want — and what would be more convenient and illustrative — is to have a light beam trace out a worldline at a 45° angle to both axes on a spacetime diagram. To achieve this, we will let each unit of distance be the distance that light travels in 1 second, which is 3×10^8 m,[7] also known as a *light-second* (ls). Therefore, the slope m of the worldline of a rightward-travelling light beam would be[8]

$$m = \frac{\Delta x}{\Delta t} = \frac{3 \times 10^8\,\mathrm{m}}{1\,\mathrm{s}} = c. \tag{2.2}$$

We see from Equation 2.2 that despite the worldline of a rightward-travelling light beam being at a 45° angle to both axes (in the first quadrant) on a spacetime diagram, the slope is not 1 — it is c. Hence, to make its slope 1 — and to make the units of position and time consistent[9] — we multiply t by c, giving us a ct-axis with units of length (measured in light-seconds) instead of just a t-axis. We will also say that $c = 1$ when using the relativistic unit of light-seconds. This is illustrated in Figure 9.[10] Now, the velocity or speed of any object in an IRF will be a dimensionless quantity since we have a quantity in light-seconds divided by another quantity in light-seconds.

From Figure 9, we can also see that the worldline of a rightward-travelling light beam is given by the equation

$$x = ct, \tag{2.3}$$

and from this, we can see that the slope is indeed 1 since

[7]Note that this is also equivalent to $c \times 1\,\mathrm{s}$, where $c = 3 \times 10^8\,\mathrm{m/s}$ (the speed of light in free space).

[8]Since time is plotted on the vertical axis and position is plotted on the horizontal axis, to obtain the slope of any worldline on a spacetime diagram, we take the "run over rise", which is equivalent to the change in position divided by the change in time, as shown in Equation 2.2.

[9]For reasons we will discuss later in this section, having a ct-axis instead of a t-axis is much more convenient and simplifies important calculations.

[10]From now on, we will omit the numbers and units labelling each axis on a spacetime diagram and assume each grid spacing (representing units of position and time on their respective axes) is separated by an interval of 1 ls. We will also omit the majority of events plotted on worldlines except for those of significance. Lastly, we will omit the labelling of the origin and only represent it with a black dot.

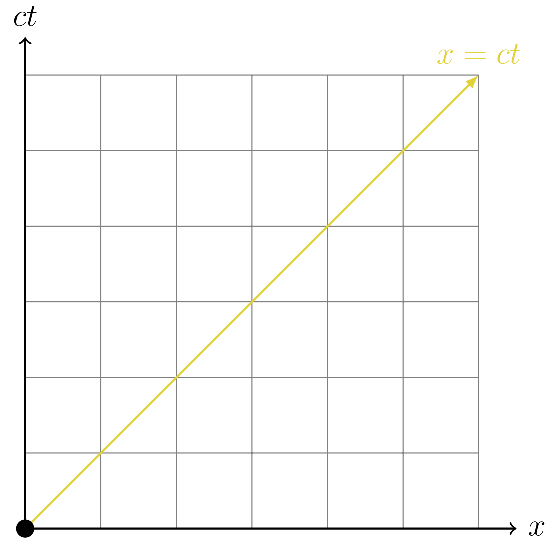

Figure 9: A revised 2D spacetime diagram on which the wordline of a light beam travelling to the right with respect to the origin is plotted.

$$m = \frac{\Delta x}{c\Delta t} = \frac{1\,\text{ls}}{1\,\text{ls}} = 1. \tag{2.4}$$

You now might wonder how we are supposed to measure time in units of length — this is entirely possible and valid. For instance, a friend might ask you how much longer it will take you to arrive at their home. You might say you are 500 m away. Assuming that you travel at a constant speed of 5 m/s on a bike, you and your friend would then deduce that it would take you another $500/5 = 100$ s until you arrive. Hence, 500 m can be thought of as the distance you travel on a bike in 100 s assuming a constant speed of 5 m/s.

Similarly, assuming time t runs at intervals of 1 s, each unit along the ct-axis equals 3×10^8 m (or 1 ls). We can therefore say that 3×10^8 m (or 1 ls) of time is the amount of time it takes a light beam to travel 3×10^8 m (or 1 ls) of distance.

With this in mind, let us discuss the geometric interpretation of the Lorentz transformation. Before anything, we can already have a fairly good guess at what the Lorentz transformation might look like; the key is to use Einstein's second postulate, which states that the speed of light remains invariant under transformation between different IRFs. To achieve this, when we transform between two IRFs, on their respective spacetime diagrams, we must make the angle between the wordline of a light beam and each axis equal, as we will see later on in this section.

We must also clarify what we mean by two events being *simultaneous* and how we can measure it. When two events are simultaneous, they happen *at the same time* according to the observer's RF. Since the speed of light is indifferent to any observer in an IRF, we may use it to define simultaneity.

Suppose we have three people A, B, and C. A is standing motionless relative to the ground, B is sitting in a train that travels away from A to the right at 1/3 the speed of light, and C is also sitting in a train that travels away from A to the right at 1/3 the speed of light, but he starts at 2 ls to the right of B. Let us now suppose B sends a light beam to C, and when C receives the light beam, he reflects it directly back to B. Figure 10 shows the resultant spacetime diagram.

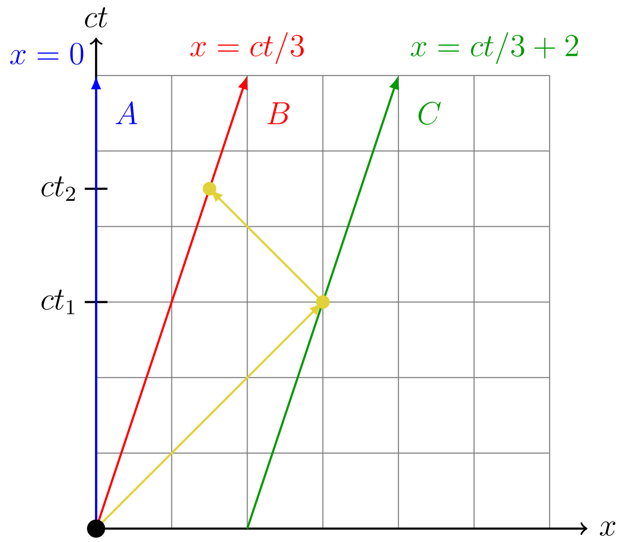

Figure 10: Worldlines of A, B, C, and two light beams in A's IRF $\mathcal{S}$.

From Figure 10, we see that C receives B's light beam at $ct_1 = 3$ and that B receives C's reflected light beam at $ct_2 = 4.5$. In other words, the time it takes for the light beam sent by B to reach C and that for the light beam to be reflected from C to B do not appear to A as being equal. This is because in $\mathcal{S}$, C appears to be moving away from the light beam sent by B and hence appears to reflect the light beam at a later time. It also appears to A that B receives the reflected light beam at an earlier time because B appears to move towards the light beam.

Let us now consider the hypothetical scenario in B's IRF. In Galilean relativity, we would shear the spacetime diagram in $\mathcal{S}$ to the right in the x direction. This is shown in Figure 11. In reality, however, this is incorrect. One of the reasons is that this method of transforming between two IRFs contradicts Einstein's second postulate, which states that the speed of light is invariant to any inertial observer.

The other slightly more subtle reason requires a bit of thinking. In our hypothetical scenario, it is mentioned that B and C travel at the same velocity to the right at 1/3 the speed of light; their worldlines only differ by their starting positions (and also by the distance between them at each moment in time, which is constant). Therefore, B and C appear motionless relative to each other. This means that in both B's and C's RFs, both the distance that the light beam sent by B travels and the time taken to do so is the same as those of the light beam reflected by C. As we can see, this directly contradicts A's notion of simultaneity since A would see the distance that the light beam travels from B to C and back to B as well as the time taken to do so as being unequal. Therefore, we can conclude that simultaneity is relative and not absolute.

Let ct_1' denote the time at which the light beam sent by B reaches C from their perspectives and ct_2' denote the time at which C reflects the light beam to B from their perspectives.

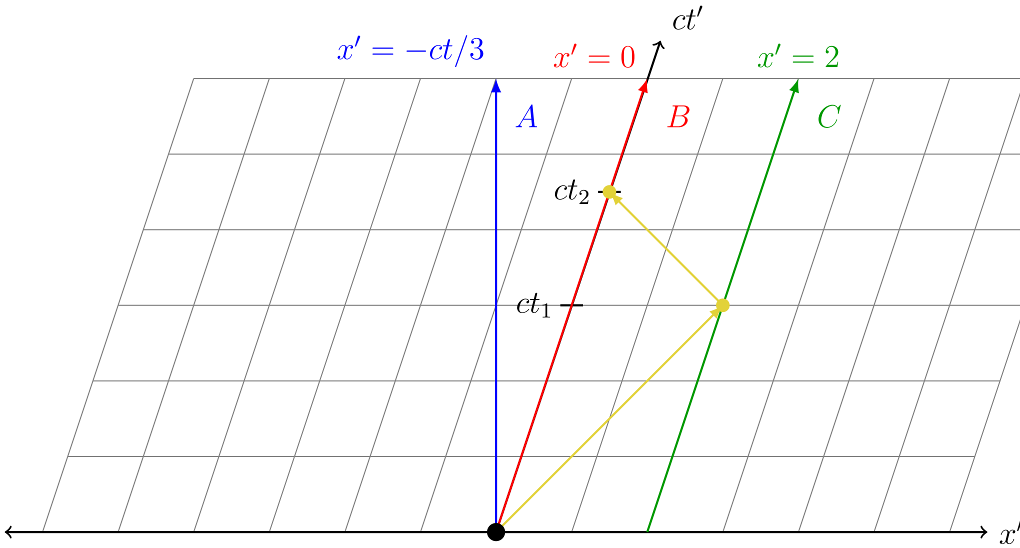

Figure 11: Worldlines of A, B, C, and two light beams in B's IRF $\mathcal{S}'$ under a Galilean transformation.

Since we have already established B and C would see that the time it takes for both of these to happen is equal, we would expect that

$$ct'_1 = \frac{ct'_2}{2}.$$

Again, this contradicts what A would see since in his IRF, $ct_1 = 3$ and $ct_2 = 4.5$, from which we can clearly see that

$$ct_1 \neq \frac{ct_2}{2}.$$

Using the information above, we can construct a *plane of simultaneity*[11] which passes through ct'_1 (which is the halfway point between the origin and ct'_2) and the event where C reflects the light beam to B. This is represented by a teal horizontal dotted line on the spacetime diagram in Figure 12, denoted by D. Simply put, a plane of simultaneity according to an observer is defined as the set of all events that appear to happen at the same time in that observer's RF. It therefore follows that multiple *planes of simultaneity* form the horizontal grid lines of a spacetime diagram in that observer's RF.

We can see from Figure 12 that IRFs transforming under Lorentz transformations have both their x- and ct-axes sheared in such a way that the angles between each of them and the worldline of a light beam are equal. In the following section, we will derive the actual mathematical equations that describe both a Lorentz transformation from an unprimed IRF to a primed IRF and one from a primed IRF to an unprimed IRF.

[11] Although on 2D spacetime diagrams it would be better to call them *lines of simultaneity*, we will stick with *planes of simultaneity* instead for generality.

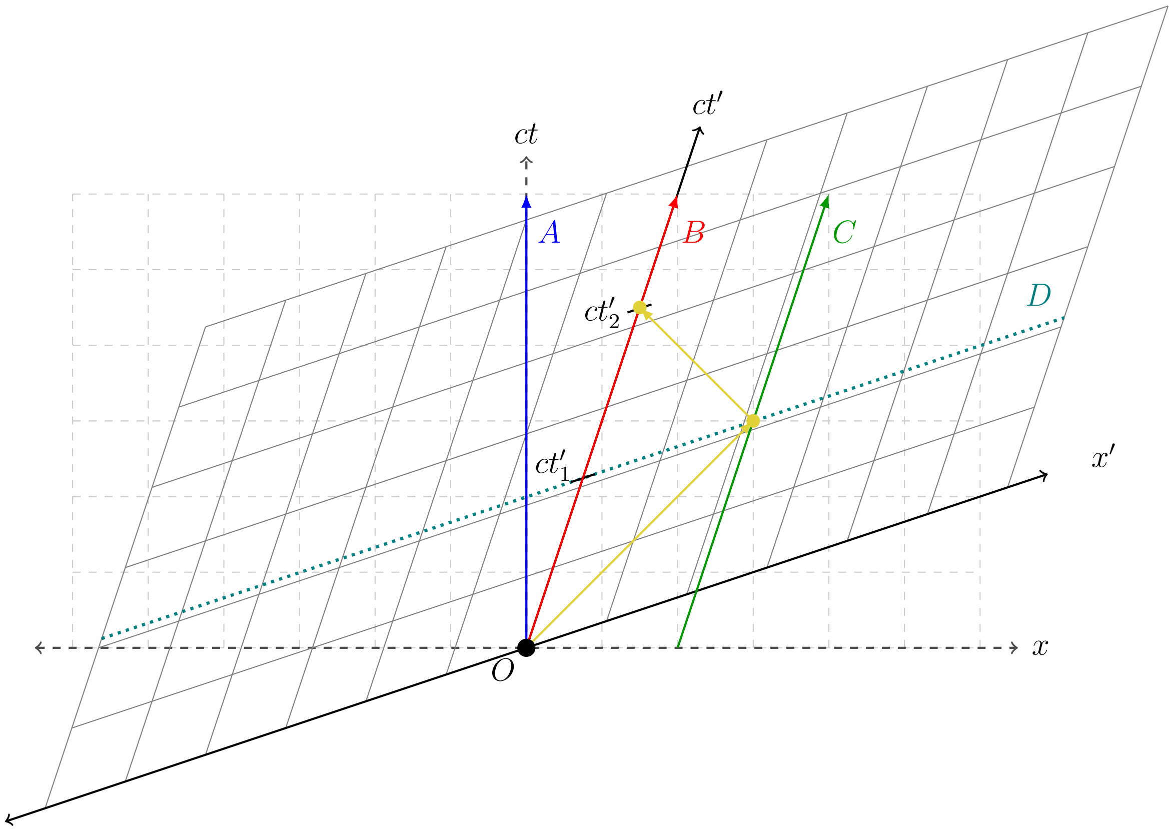

Figure 12: Worldlines of A, B, C, and two light beams in A's IRF $\mathcal{S}$ (lighter) B's IRF $\mathcal{S}'$ (darker) under a Lorentz transformation.

2.3 Derivation of the Lorentz Transformation Equations

We know from Equation 2.4 that the slope of the worldline of a rightward-travelling light beam is 1. From the hypothetical scenario mentioned in the section above, we also know that the worldline of B (and hence the ct'-axis) is described using the equation

$$x = \frac{1}{3}ct. \tag{2.5}$$

Equation 2.5 tells us that B travels at 1/3 the speed of light to the right relative to A. In units of m/s, we see that B travels at $v = 1 \times 10^8$ m/s. However, simply writing

$$x = vct$$

or

$$x = vt$$

would be incorrect since in the first equation, the dimensions (or units) are inconsistent, and in the second equation, we relate x with t instead of with ct. Hence, we must divide v by the speed of light c and obtain

$$x = \frac{v}{c}ct. \tag{2.6}$$

Here, we will define our new quantity β as

$$\beta \equiv \frac{v}{c},$$

where v is the velocity of any object or observer under question. This is called the *beta coefficient*, which represents the fraction of the speed of light at which the object travels. We may hence rewrite Equation 2.6 as

$$x = \beta ct. \tag{2.7}$$

We can also see that the slope of the ct'-axis is β, and since the angle between the worldline of a rightward-travelling light beam and the ct'-axis is equal to that between the light beam and the x'-axis, the x'-axis will have a reciprocal slope of $1/\beta$. Thus, the x'-axis is described by the equation

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

which, when we rearrange for ct, becomes

$$ct = \beta x. \tag{2.8}$$

Recall from Galilean relativity that a Galilean transformation from an unprimed IRF to a primed IRF is described by the equations

$$t' = t$$

and

$$x' = x - vt.$$

To obtain the Lorentz transformation analogues, we replace vt with βct for the x' coordinate and subtract βx from ct to obtain the ct' coordinate, as in

$$ct' = ct - \beta x \tag{2.9}$$

and

$$x' = x - \beta ct. \tag{2.10}$$

Note that βct represents the gap between the ct- and ct'-axes and βx represents the gap between the x- and x'-axes. We should also note that we do not know whether the spacing between the grid lines is unchanged after a Lorentz transformation. Thus, we will multiply an additional *gamma coefficient* γ to the expressions on the right-hand side of Equations 2.9 and 2.10, giving us

$$ct' = \gamma(ct - \beta x) \tag{2.11}$$

and

$$x' = \gamma(x - \beta ct). \tag{2.12}$$

To solve for γ, we can use the inverse Lorentz transformation equations, which are simply Equations 2.11 and 2.12 but with the signs of βx and βct flipped, as in

$$ct = \gamma(ct' + \beta x')$$

and

$$x = \gamma(x' + \beta ct').$$

Multiplying x' by x, we get

$$x'x = \gamma(x - \beta ct)\gamma(x' + \beta ct'). \tag{2.13}$$

We know that the speed of light is invariant in any IRF, so for a light beam, the Lorentz transformation must guarantee that $x = ct$ and $x' = ct'$. In other words, $c = x/t = x'/t'$. Hence, substituting these in for x' and x in Equation 2.13, we get

$$\begin{aligned} ct'ct &= \gamma(ct - \beta ct)\gamma(ct' + \beta ct') \\ &= \gamma ct(1 - \beta)\gamma ct'(1 + \beta) \\ &= \gamma^2 ct'ct\left(1 - \beta^2\right). \end{aligned}$$

Dividing both sides by $ct'ct$, we get

$$1 = \gamma^2\left(1 - \beta^2\right).$$

Finally, dividing both sides by $(1 - \beta^2)$ and taking the square root, we arrive at the following expression for γ:

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}}. \tag{2.14}$$

We have hence derived the Lorentz transformation equations and their inverse as well as γ, which we will call the *Lorentz factor* from here onward. Note that on a spacetime diagram, we will mainly be concerned with the ct and x coordinates, but like the Galilean transformation equations, the Lorentz transformation equations leave the y and z coordinate unchanged, that is, their unprimed and primed counterparts are equal. Therefore, we have the following equations describing a Lorentz transformation:

$$ct' = \gamma(ct - \beta x); \tag{2.15}$$

$$x' = \gamma(x - \beta ct); \tag{2.16}$$

$$y' = y; \tag{2.17}$$

$$z' = z. \tag{2.18}$$

We also have the following equations describing an inverse Lorentz transformation:

$$ct = \gamma(ct' + \beta x'); \tag{2.19}$$

$$x = \gamma(x' + \beta ct'); \tag{2.20}$$

$$y = y'; \tag{2.21}$$

$$z = z'. \tag{2.22}$$

Note that if we wish to isolate t by itself, we can alternatively write Equations 2.15 and 2.19 as

$$t' = \gamma\left(t - \frac{\beta x}{c}\right) = \gamma\left(t - \frac{vx}{c^2}\right) \tag{2.23}$$

and

$$t = \gamma\left(t' + \frac{\beta x'}{c}\right) = \gamma\left(t' + \frac{vx}{c^2}\right) \tag{2.24}$$

by dividing both sides of each equation by c. If we wish to work out only the difference between the time components of the unprimed and primed time coordinates, we can take out the expression containing x from Equations 2.23 and 2.24 to get

$$t' = \gamma t \tag{2.25}$$

and

$$t = \gamma t'; \tag{2.26}$$

or with time intervals, we get

$$\Delta t' = \gamma \Delta t \tag{2.27}$$

and

$$\Delta t = \gamma \Delta t'. \tag{2.28}$$

2.4 Time Dilation

Time dilation is a relativistic effect that is a consequence of the invariance of the speed of light and the Lorentz transformations. Time dilation is the slowing of time for an observer in a moving IRF relative to another observer in a stationary IRF. That is to say, the faster an object travels relative to an observer, it will appear to the observer that the object experiences the flow of time more slowly. We have already derived the associated equation that describes time dilation (see Equation 2.25), but we can visualise it using a spacetime diagram and explore using a thought experiment.

Let us consider another hypothetical scenario similar to that in Section 2.2. Suppose we have a person A standing stationary relative to the ground. Suppose we have another person B travelling away from A to the right at 1/2 the speed of light. Figure 13 shows the resultant spacetime diagram.

From Figure 13, we can clearly see that the ct-axis of $\mathcal{S}$ is not aligned with that of $\mathcal{S}'$. Take $ct = 3$, which is what A would call "3 seconds" in his stationary IRF. However, $ct' = 3$ is also what B would call "3 seconds" in his moving IRF. Notice from A's perspective, the point $ct' = 3$ (which is what B calls "3 seconds") appears to be at $ct \approx 3.46$. In other words, the notion of time for A and B are different in different IRFs — we say that time appears to be *dilated* in B's moving IRF relative to A. This is shown in Figure 13.

Assuming we are measuring time at a constant position of $x = 0$, using Equation 2.25, which is the equation for time dilation of the primed IRF in the unprimed IRF, we can see that indeed, what B would call "3 seconds" appears to be approximately 3.46 seconds (or $2\sqrt{3}$ seconds to be exact) in A's IRF:

$$\begin{aligned} t' &= \gamma t \\ &= \frac{3}{\sqrt{1 - \left(\frac{1}{2}\right)^2}} \\ &= \frac{3}{\sqrt{\frac{3}{4}}} \\ &\approx 3.46. \end{aligned}$$

However, from Equation 2.26, we can also see that an observer in a primed IRF would see the time coordinates of an unprimed IRF dilate by the same factor γ. So for our hypothetical scenario, while A sees B's time to be dilated, B will also see A's time as being dilated. This is because B is motionless relative to himself and will see A moving at 1/2 the speed of light

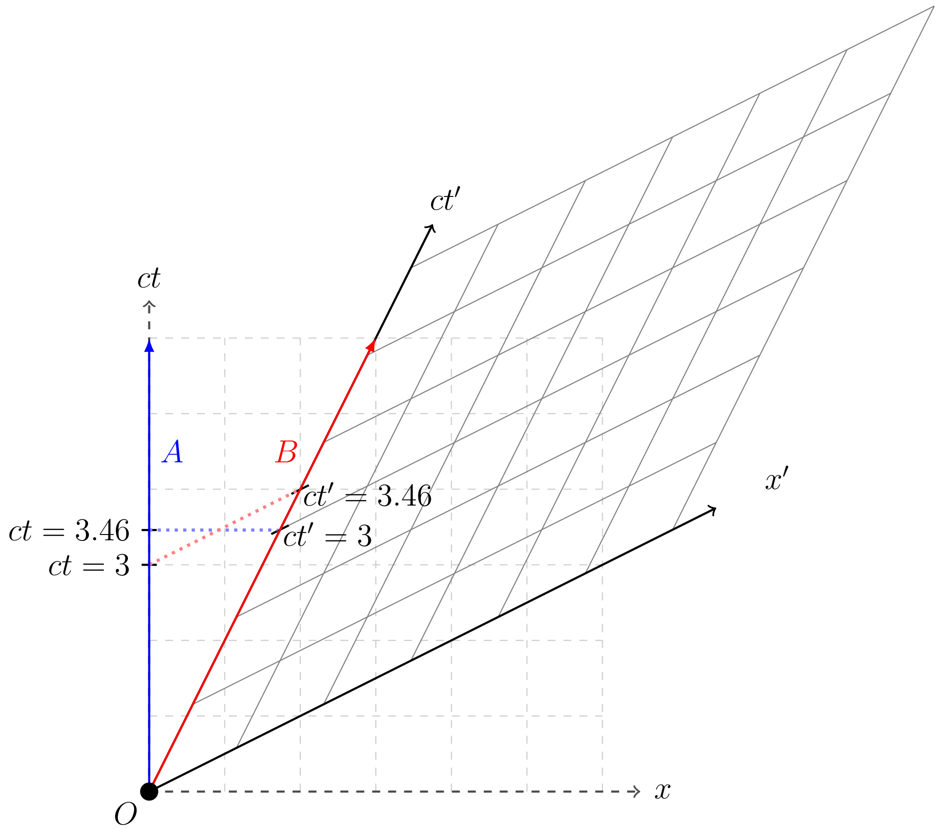

Figure 13: Worldlines of A and B in both A's IRF $\mathcal{S}$ (lighter) and B's IRF $\mathcal{S}'$ (darker) under a Lorentz transformation, demonstrating time dilation.

to the left with A's time being dilated by exactly the same factor γ. This is known as the *reciprocity of time dilation*, which is shown in Figure 13.

Time dilation can also be explained using another example with a train. Suppose an observer A stands motionless on a train relative to it. Inside the train, a light beam is emitted from a laser on the ground to a mirror on the train's ceiling and reflected back down again. Suppose we have another observer B standing on a platform, motionless relative to it inside a train station watching the train go past. The train travels to the right at some velocity v relative to B. We will denote the height of the train as h, the horizontal distance that B sees between the point at which the light beam is emitted and the point at which the light beam hits the ceiling as ℓ, and the distance the emitted light beam (and hence the reflected light beam) travels as α. This is shown in Figure 14. We will also denote A's measure of time in $\mathcal{S}$ as $\Delta\tau$ and that of B's in $\mathcal{S}'$ as $\Delta t'$.

Let us discuss what happens in $\mathcal{S}$ first. Since the train's height is h, we know that the light beam will travel a distance of $2h$ from when it is emitted to when it is reflected back down again. Since the light beam travels at a speed of c, the time taken for the light beam to finish one bounce[12] from A's perspective is

[12] We will take "bounce" to mean "emission of the light beam from the laser on the ground to the mirror on the ceiling and its reflection from the mirror back to the ground".

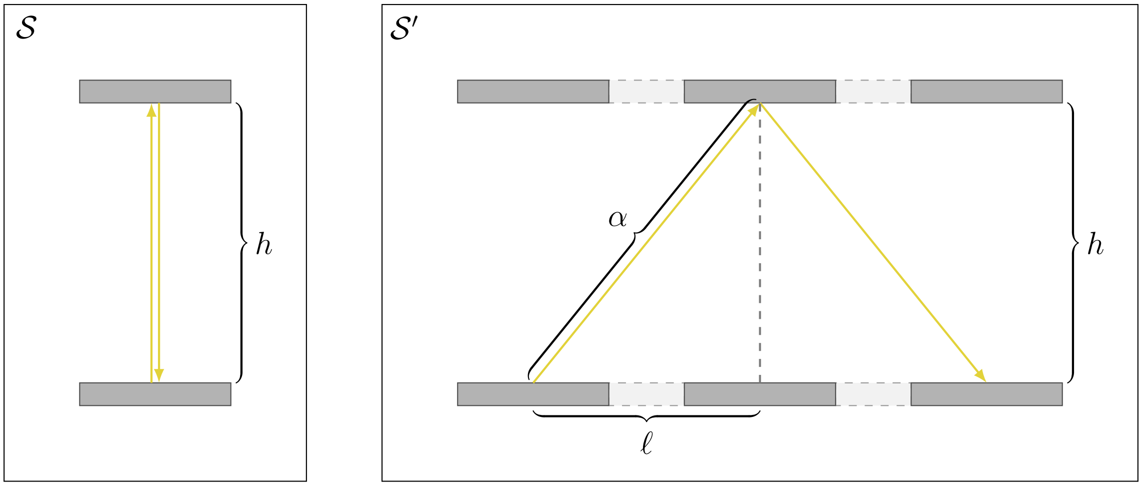

Figure 14: Bouncing light beam in both A's IRF $\mathcal{S}$ (left) and B's IRF $\mathcal{S}'$ (right), demonstrating time dilation.

$$\Delta\tau = \frac{2h}{c}. \tag{2.29}$$

Rearranging for h, we get

$$h = \frac{c\Delta\tau}{2}. \tag{2.30}$$

However, things do not appear the same in $\mathcal{S}'$. Since the light beam now appears to travel a distance of 2α, the time $\Delta t'$ taken from the light beam to finish one bounce from B's perspective is

$$\Delta t' = \frac{2\alpha}{c}. \tag{2.31}$$

Rearranging for α, we get

$$\alpha = \frac{c\Delta t'}{2}. \tag{2.32}$$

Using simple algebra, we also work out ℓ to be

$$\ell = \frac{v\Delta t'}{2}, \tag{2.33}$$

where v is the velocity of the train. By Pythagoras's theorem, we know that

$$\alpha^2 = \ell^2 + h^2. \tag{2.34}$$

Substituting expressions for α, ℓ, and h from Equations 2.32, 2.33, and 2.30, respectively, we have

$$\begin{aligned}\left(\frac{c\Delta t'}{2}\right)^2 &= \frac{c^2(\Delta t')^2}{4} \\ &= \left(\frac{v\Delta t'}{2}\right)^2 + \left(\frac{c\Delta t'}{2}\right)^2 \\ &= \frac{v^2(\Delta t')^2}{4} + \frac{c^2(\Delta \tau)^2}{4}.\end{aligned} \tag{2.35}$$

Multiplying both sides by 4 and rearranging so that only the $\Delta\tau$ term is on the right-hand side, we get

$$c^2(\Delta t')^2 - v^2(\Delta t')^2 = c^2(\Delta\tau)^2. \tag{2.36}$$

Dividing both sides by c^2 and factoring out $(\Delta t')^2$, we get

$$(\Delta t')^2\left(1 - \frac{v^2}{c^2}\right) = (\Delta\tau)^2. \tag{2.37}$$

Dividing by $(1 - v^2/c^2)$ and taking the square root, we obtain

$$\Delta t' = \frac{\Delta\tau}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{\Delta\tau}{\sqrt{1-\beta^2}} = \gamma\Delta\tau. \tag{2.38}$$

We have hence derived the equation for time dilation (Equation 2.38). $\Delta\tau$ is called the *proper time interval* or *proper time*, which is the time elapsed between two events measured by an observer in their own rest IRF. We will discuss more about this in Section 2.6.

2.5 Length Contraction

Another consequence of the invariance of the speed of light is *length contraction*. Length contraction is a phenomenon where the faster an object travels, the more its length appears to contract in the direction of motion from a stationary observer's perspective. Let us suppose we have a metal rod with a length of

$$\Delta\chi = \chi_2 - \chi_1$$

when measured by an observer A at rest relative to it in an IRF $\mathcal{S}$, where χ_1 and χ_2 are the endpoints of the rod. Let us suppose another observer B is moving away from the rod at some velocity v. B will attempt to measure the length of the rod in his moving IRF $\mathcal{S}'$. Figure 15 shows the resultant spacetime diagram. Note that we will denote the length of the rod measured by B in his moving IRF $\mathcal{S}'$ as

$$\Delta x' = x_2' - x_1'.$$

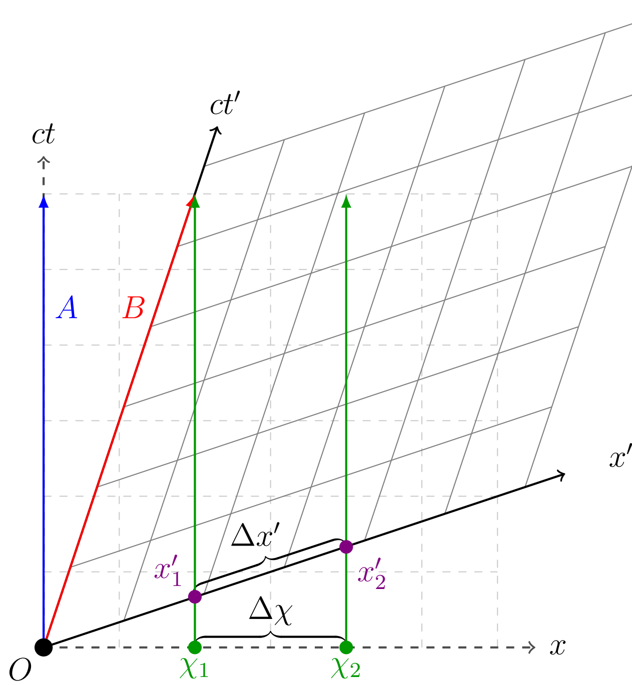

Figure 15: Worldlines of A, B, and the metal rod in both A's IRF $\mathcal{S}$ (lighter) and B's IRF $\mathcal{S}'$ (darker) under a Lorentz transformation, demonstrating length contraction.

Because B is a moving observer (relative to A and the rod), we will also assume that the times at which B measures each endpoint of the rod in $\mathcal{S}'$ are simultaneous according to his IRF. Hence, the duration $\Delta t'$ of the act of measuring by B in $\mathcal{S}'$ must be 0, that is, we have

$$\Delta t' = t'_2 - t'_1 = 0,$$

implying that $t'_1 = t'_2$. Note that we will denote the time interval measured by A in $\mathcal{S}$ as the proper time interval mentioned in Section 2.4, which is

$$\Delta\tau = \tau_2 - \tau_1.$$

Using Equations 2.16 and 2.23, we can write $\Delta x'$ and $\Delta t'$ as

$$\begin{aligned} \Delta x' &= x'_2 - x'_1 \\ &= \gamma(\chi_2 - v\tau_2) - \gamma(\chi_1 - v\tau_1) \\ &= \gamma(\chi_2 - \chi_1 - v\tau_2 + v\tau_1) \\ &= \gamma(\Delta\chi - v\Delta\tau). \end{aligned} \tag{2.39}$$

and

$$
\begin{aligned}
\Delta t' &= t'_2 - t'_1 \\
&= \gamma \left(\tau_2 - \frac{v\chi_2}{c^2} \right) - \gamma \left(\tau_1 - \frac{v\chi_1}{c^2} \right) \\
&= \gamma \left(\tau_2 - \tau_1 - \frac{v\chi_2}{c^2} + \frac{v\chi_1}{c^2} \right) \\
&= \gamma \left(\Delta\tau - \frac{v\Delta\chi}{c^2} \right).
\end{aligned} \tag{2.40}
$$

Since $\Delta t'$ is 0, we see that

$$
\Delta\tau = \frac{v\Delta\chi}{c^2}. \tag{2.41}
$$

Plugging Equation 2.41 in for $\Delta\tau$ in Equation 2.39, we obtain

$$
\begin{aligned}
\Delta x' &= \gamma \left(\Delta\chi - \frac{v^2\Delta\chi}{c^2} \right) \\
&= \gamma\Delta\chi(1 - \beta^2) \\
&= \Delta\chi \frac{1 - \beta^2}{\sqrt{1 - \beta^2}} \\
&= \frac{\Delta\chi}{\gamma}.
\end{aligned} \tag{2.42}
$$

We have hence derived the equation for length contraction (Equation 2.42). $\Delta\chi$ is called the *proper length*, which is the length of an object measured by an observer in their own IRF. We may also consider the same hypothetical scenario where the two observers A and B are trying to measure a metal rod, but this time, we will let the metal rod move away from A at some fixed velocity v such that the rod appears motionless relative to B. We will denote the length of the rod measured by A in $\mathcal{S}$ and that measured by B in $\mathcal{S}'$ as Δx and $\Delta\chi$, respectively. We will denote the duration of measurement by A and B in their own respective IRFs as Δt and $\Delta\tau$. Figure 16 shows the resultant spacetime diagram.

Because A is a moving observer (relative to B and the rod), the times at which A measures each endpoint of the rod in $\mathcal{S}$ must be simultaneous according to his IRF. Hence, the duration Δt of the act of measuring by B in $\mathcal{S}'$ must be 0, that is, we have

$$
\Delta t = t_2 - t_1 = 0,
$$

implying that $t_1 = t_2$. Note that we will denote the time interval measured by B in $\mathcal{S}'$ as the proper time interval mentioned in Section 2.4, which is

$$
\Delta\tau = \tau_2 - \tau_1.
$$

Using Equations 2.20 and 2.24, we can write Δx and Δt as

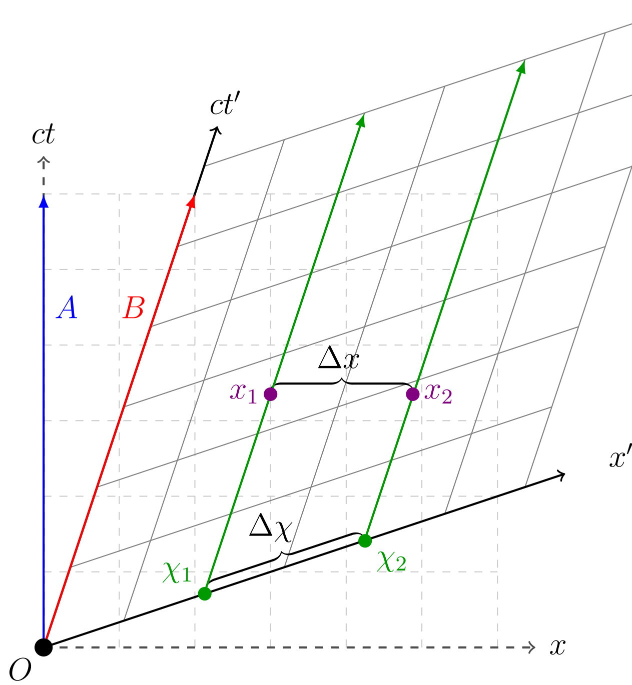

Figure 16: Worldlines of A, B, and the metal rod in both A's IRF $\mathcal{S}$ (lighter) and B's IRF $\mathcal{S}'$ (darker) under a Lorentz transformation, demonstrating length contraction.

$$\begin{aligned}\Delta x &= x_2 - x_1 \\ &= \gamma(\chi_2 + v\tau_2) - \gamma(\chi_1 + v\tau_1) \\ &= \gamma(\chi_2 - \chi_1 + v\tau_2 - v\tau_1) \\ &= \gamma(\Delta\chi + v\Delta\tau).\end{aligned} \tag{2.43}$$

and

$$\begin{aligned}\Delta t &= t_2 - t_1 \\ &= \gamma\left(\tau_2 + \frac{v\chi_2}{c^2}\right) - \gamma\left(\tau_1 + \frac{v\chi_1}{c^2}\right) \\ &= \gamma\left(\tau_2 - \tau_1 + \frac{v\chi_2}{c^2} - \frac{v\chi_1}{c^2}\right) \\ &= \gamma\left(\Delta\tau + \frac{v\Delta\chi}{c^2}\right).\end{aligned} \tag{2.44}$$

Since Δt is 0, we see that

$$\Delta\tau = -\frac{v\Delta\chi}{c^2}. \tag{2.45}$$

Plugging Equation 2.45 in for $\Delta\tau$ in Equation 2.43, we obtain

$$
\begin{aligned}
\Delta x &= \gamma \left(\Delta\chi - \frac{v^2 \Delta\chi}{c^2} \right) \\
&= \gamma \Delta\chi (1 - \beta^2) \\
&= \Delta\chi \frac{1 - \beta^2}{\sqrt{1 - \beta^2}} \\
&= \frac{\Delta\chi}{\gamma}.
\end{aligned} \tag{2.46}
$$

We have hence derived the other equation for length contraction (Equation 2.46).

2.6 The Spacetime Interval and Light Cones

Special relativity is about determining what quantities are relative in different IRFs and what quantities remain invariant. We know that position, length, and simultaneity are relative to different inertial observers under Lorentz transformations. We also know that in Galilean relativity, the distance $\Delta\mathbf{r}$ between two points in 3D Euclidean space,

$$
\Delta\mathbf{r}^2 \equiv (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2,
$$

is invariant under a Galilean transformation. We can speculate that the quantity Δs^2 — which is the 4D analogue of Euclidean distance called the *spacetime interval*, which we will temporarily define as

$$
(\Delta s)^2 \equiv (c\Delta t)^2 + (\Delta\mathbf{r})^2 = (c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2, \tag{2.47}
$$

is invariant under a Lorentz transformation; that is, we are trying to see if $(\Delta s)^2 = (\Delta s')^2$, or in full, if the equation

$$
(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 = (c\Delta t')^2 + (\Delta x')^2 + (\Delta y')^2 + (\Delta z')^2
$$

is correct. Using Equations 2.15, 2.16, 2.17, and 2.18, we can compute $(\Delta s')^2$:

$$
\begin{aligned}
(\Delta s')^2 &= \gamma^2 (c\Delta t - \beta\Delta x)^2 + \gamma^2 (\Delta x - \beta c\Delta t)^2 + (\Delta y)^2 + (\Delta z)^2 \\
&= \gamma^2 (c\Delta t)^2 - 2\gamma^2 \beta (c\Delta t)(\Delta x) + \gamma^2 \beta^2 (\Delta x)^2 \\
&\quad + \gamma^2 (\Delta x)^2 - 2\gamma^2 \beta (c\Delta t)(\Delta x) + \gamma^2 \beta^2 (c\Delta t)^2 \\
&\quad + (\Delta y)^2 + (\Delta z)^2 \\
&= \gamma^2 (c\Delta t)^2 (1 + \beta^2) + \gamma^2 (\Delta x)^2 (1 + \beta^2) \\
&\quad - 4\gamma^2 \beta (c\Delta t)(\Delta x) + (\Delta y)^2 + (\Delta z)^2 \\
&= (c\Delta t)^2 \frac{1 + \beta^2}{1 - \beta^2} + (\Delta x)^2 \frac{1 + \beta^2}{1 - \beta^2} - (c\Delta t)(\Delta x) \frac{4\beta}{1 - \beta^2} \\
&\quad + (\Delta y)^2 + (\Delta z)^2 \\
&\neq (\Delta s)^2.
\end{aligned} \tag{2.48}
$$

From Equation 2.48, we see that the definition of $(\Delta s)^2$ given in Equation 2.47 is not correct. However, we could define $(\Delta s)^2$ as[13]

$$(\Delta s)^2 \equiv (c\Delta t)^2 - (\Delta \mathbf{r})^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \tag{2.49}$$

Using the definition given in Equation 2.49 and using Equations 2.15, 2.16, 2.17, and 2.18, we can compute $(\Delta s')^2$ once again:

$$\begin{aligned}
(\Delta s')^2 &= \gamma^2(c\Delta t - \beta\Delta x)^2 - \gamma^2(\Delta x - \beta c\Delta t)^2 - (\Delta y)^2 - (\Delta z)^2 \\
&= \gamma^2(c\Delta t)^2 - 2\gamma^2\beta(c\Delta t)(\Delta x) + \gamma^2\beta^2(\Delta x)^2 \\
&\quad - \gamma^2(\Delta x)^2 + 2\gamma^2\beta(c\Delta t)(\Delta x) - \gamma^2\beta^2(c\Delta t)^2 \\
&\quad - (\Delta y)^2 - (\Delta z)^2 \\
&= \gamma^2(c\Delta t)^2(1-\beta^2) - \gamma^2(\Delta x)^2(1-\beta^2) \\
&\quad - (\Delta y)^2 - (\Delta z)^2 \\
&= (c\Delta t)^2\frac{1-\beta^2}{1-\beta^2} - (\Delta x)^2\frac{1-\beta^2}{1-\beta^2} - (\Delta y)^2 - (\Delta z)^2. \\
&= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \\
&= (\Delta s)^2.
\end{aligned} \tag{2.50}$$

From Equation 2.50, we can see that the definition of the spacetime interval $(\Delta s)^2$ given in Equation 2.49 allows it to remain invariant under a Lorentz transformation. For this reason, $(\Delta s)^2$ is called a *Lorentz scalar*, which is a scalar quantity that remains invariant under a Lorentz transformation. We can also derive another equivalent expression for $(\Delta s)^2$ from Equation 2.36 by substituting $(\Delta \mathbf{r}')^2/(\Delta t')^2$ in for v^2:

$$\begin{aligned}
(c\Delta t')^2 - (v\Delta t')^2 &= (c\Delta t')^2 - \left(\frac{\Delta \mathbf{r}'}{\Delta t'}\Delta t'\right)^2 \\
&= (c\Delta t')^2 - (\Delta \mathbf{r}')^2 \\
&= (\Delta s)^2 \\
&= (c\Delta \tau)^2,
\end{aligned}$$

implying that

$$\Delta\tau = \sqrt{\frac{(\Delta s)^2}{c^2}}$$

for the *proper time interval* $\Delta\tau$, which is indeed what we obtain by using the formula for time dilation (Equation 2.38):

[13]Some authors choose to define the spacetime interval as having a negative timelike component $-(c\Delta t)^2$ and a positive spacelike component $(\Delta \mathbf{r})^2$. Both definitions are valid since they show that time and space are fundamentally different. Since the worldline of a light beam is given by $(c\Delta t)^2 = (\Delta \mathbf{r})^2$, we can either subtract $(c\Delta t)^2$ or $(\Delta \mathbf{r})^2$ from both sides and define the resultant quantity as the spacetime interval. Here, however, we will use the definition given by Equation 2.49.

$$
\begin{aligned}
\Delta\tau &= \frac{\Delta t}{\gamma} \\
&= \Delta t\sqrt{1-\frac{v^2}{c^2}} \\
&= \Delta t\sqrt{1-\frac{(\Delta\mathbf{r})^2}{c^2(\Delta t)^2}} \\
&= \sqrt{(\Delta t)^2-\left((\Delta t)^2\frac{(\Delta\mathbf{r})^2}{c^2(\Delta t)^2}\right)} \\
&= \sqrt{(\Delta t)^2-\left(\frac{(\Delta\mathbf{r})^2}{c^2}\right)} \\
&= \sqrt{\frac{(\Delta s)^2}{c^2}}.
\end{aligned}
$$

We will also define the *proper distance* $\Delta\sigma$ to be the square root of the negative of the spacetime interval, namely,

$$\Delta\sigma = \sqrt{-(\Delta s)^2}.$$

However, $(\Delta s)^2$ being positive makes $\Delta\sigma$ have an imaginary value. Hence, we must set some restrictions. Before setting these restrictions, let us discuss the concept of *light cones* in spacetime, which is the path that a flash of light emitted from a single event travelling in all directions would take through spacetime. This is visualised in Figure 17 on a 2D spacetime diagram.[14]

We see from Figure 17 that the spacetime diagram is separated into two types of regions. The blue region represents all events that are within the light cone, meaning that the events are *timelike* separated relative to the origin.[15] In other words, for any given event within that region, an object is able to reach the event from the origin by travelling slower than the speed of light. *If two events are timelike separated relative to each other, then there exists an IRF where both events occur at the same position but at different times. If two events are timelike separated relative to each other, then this will always be the case for any IRF.*

The red region represents all events that are outside the light cone, meaning that the events are *spacelike* separated relative to the origin. In other words, for any given event within that region, a light beam is unable to reach the event from the origin in any given amount of time; or, an object must travel faster than the speed of light to reach the event, which is impossible given our current knowledge of physics. *If two events are spacelike separated relative to each other, then there exists an IRF where both events occur at the same time but at different positions. Similarly, if two events are spacelike separated relative to each other, then this will always be the case for any IRF.*

[14] Instead of the usual x-axis, we will instead have the $\mathbf{r}$ axis representing all of x, y, and z.

[15] The upper blue region represents the *future light cone* and the lower blue region represents the *past light cone*; both regions are timelike separated relative to the origin.

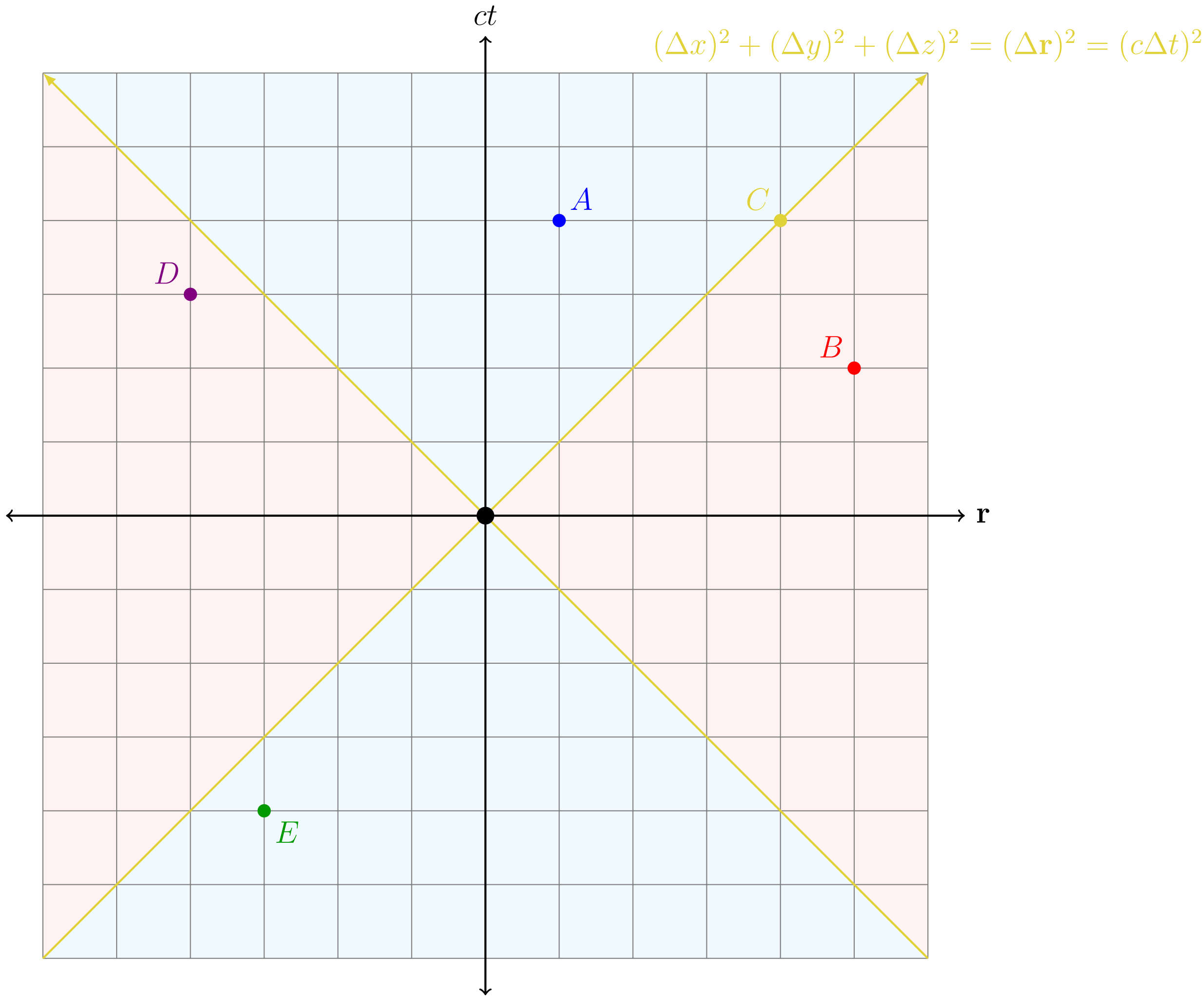

Figure 17: A 2D spacetime diagram showing the timelike region (blue), spacelike region (red), a light cone, and some events.

All events lying on the light cone itself are *lightlike* separated relative to the origin. In other words, for any given event on the light cone, it is exactly far enough away from the origin that a light beam travelling from the origin is able to reach it. *If two events are lightlike separated relative to each other, then this will always be the case for any IRF.*

Let us take event A as an example. We see that it is timelike separated relative to the origin since it is in the blue region. Also, notice that the separation in time is greater than that in space. Hence, we say that the separation between two events is timelike if

$$(c\Delta t)^2 > (\Delta \mathbf{r})^2,$$

implying that

$$(\Delta s)^2 > 0.$$

Let us take event E. We see that it is also timelike separated relative to the origin. However, we also see that $\Delta\mathbf{r} = -3$ and $c\Delta t = -4$, so we have that $c\Delta t < \Delta\mathbf{r}$. This is why we square each quantity (i.e. having $(c\Delta t)^2$ and $(\Delta\mathbf{r})^2$) instead of having

$$c\Delta t > \Delta\mathbf{r}$$

as the condition for a timelike separation; if we allow for negative values, then the definition of a timelike interval would not hold. Hence, we set a restriction saying that the proper time interval $\Delta\tau$ is only defined for values of $(\Delta s)^2 > 0$.

Event B is spacelike separated relative to the origin since it is in the red region. Notice that the separation in space is greater than that in time. Hence, we say that the separation between two events is spacelike if

$$(c\Delta t)^2 < (\Delta\mathbf{r})^2,$$

implying that

$$(\Delta s)^2 < 0.$$

Let us now take event D. We see that it is also spacelike separated relative to the origin. However, we see that $\Delta\mathbf{r} = -4$ and $c\Delta t = -3$, so we have that $c\Delta t > \Delta\mathbf{r}$. This is why, again, we take the square of each quantity instead of having

$$c\Delta t < \Delta\mathbf{r}$$

as the condition for a spacelike separation. Hence, we set another restriction saying that the proper distance $\Delta\sigma$ is only defined for values of $(\Delta s)^2 < 0$.

Finally, let us take event C. We see that it is lightlike separated relative to the origin since it lies on the light cone. Notice that the separation in time is equal to that in space. Hence we say that the separation between two events is lightlike if

$$(c\Delta t)^2 = (\Delta\mathbf{r})^2,$$

implying that

$$(\Delta s)^2 = 0.$$

To summarise, if two events are timelike separated (i.e. $(\Delta s)^2 > 0$), then we get the proper time interval between them, which is

$$\Delta\tau = \sqrt{\frac{(\Delta s)^2}{c^2}} = \sqrt{(\Delta t)^2 - \left(\frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{c^2}\right)}. \tag{2.51}$$

If two events are spacelike separated (i.e. $(\Delta s)^2 < 0$), then we get the proper distance between them, which is

$$\Delta\sigma = \sqrt{-(\Delta s)^2} = \sqrt{-(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}. \tag{2.52}$$

2.7 4-Vectors and the Minkowski Metric Tensor

A *vector* is a quantity with both a magnitude and a direction. On a 2D Cartesian coordinate system, a position vector is represented by an arrow spanning from the origin to some other point on the coordinate system, both of which are specified by coordinates (x, y). A 3D position vector is analogous to a 2D Euclidean vector, but instead, it is specified by three coordinates (x, y, z). Hence, we can see that a position vector represents the position of a point in space. Consider a position vector $\mathbf{u} = (4, 3)$ on a 2D Cartesian coordinate system. Figure 18 shows this.

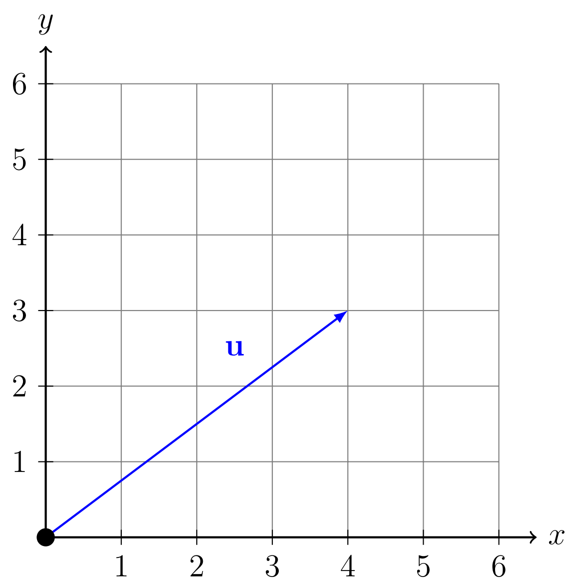

Figure 18: A 2D Cartesian coordinate system with a position vector $\mathbf{u}$.

Notice that $\mathbf{u}$ can be written in terms of an x-vector component and a y-vector component. To arrive at $\mathbf{u}$, we go across 4 grids in the x direction and go up 3 grids in the y direction. We will denote the x- and *y-vector components* of $\mathbf{u}$ as $\mathbf{u}_x$ and $\mathbf{u}_y$, respectively. They can each be further decomposed into a scalar (which we call the *scalar component* of each vector component) and a basis vector. A *scalar* is a quantity with only a magnitude and no direction. A *basis vector* is a unit vector used to represent the axes of a Cartesian coordinate system, and a *unit vector* is a vector of length 1. Basis vectors in the x and y direction are denoted as $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$, respectively. It follows that $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ have coordinates $(1, 0)$ and $(0, 1)$, respectively.

Using a combination of scalar components and basis vectors, we write

$$\mathbf{u}_x = 4\hat{\mathbf{e}}_x \quad \text{and} \quad \mathbf{u}_y = 3\hat{\mathbf{e}}_y$$

for the x- and y-vector components of $\mathbf{u}$, respectively, and

$$\mathbf{u} = \mathbf{u}_x + \mathbf{u}_y = 4\hat{\mathbf{e}}_x + 3\hat{\mathbf{e}}_y$$

for the position vector $\mathbf{u} = (4, 3)$. Figure 19 shows a fuller decomposition of $\mathbf{u}$.

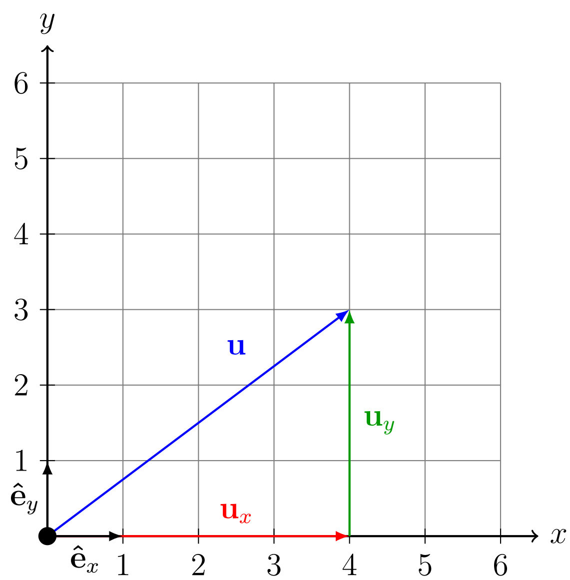

Figure 19: A 2D Cartesian coordinate system with a position vector $\mathbf{u}$.

Notice what happens if we make the basis vectors $\hat{\mathbf{e}}_x$ and $\hat{\mathbf{e}}_y$ each have a length of 2 units. We will leave the position vector $\mathbf{u}$ unchanged. Figure 20 shows the resultant graph.

After we double the length of each basis vectors, we find that the scalar components halve in length, that is, we have that

$$\mathbf{u}_x = 2\hat{\mathbf{e}}_x \quad \text{and} \quad \mathbf{u}_y = 1.5\hat{\mathbf{e}}_y.$$

By logic, if we halve the length of each basis vector, then the scalar components will double in length. We see that the scalar components and basis vectors vary in opposite ways. Hence, we say that the vector components $\mathbf{u}_x$ and $\mathbf{u}_y$ are *contravariant components*. We will also replace any indices involving x and y with 1 and 2. Therefore, for any 2D position vector $\mathbf{r}$, we can express it as

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 = r^1\hat{\mathbf{e}}_1 + r^2\hat{\mathbf{e}}_2.$$

We may also express $\mathbf{r}$ using *covariant components* — meaning that the scalar components and basis vectors vary in the same way, like so:[16]

[16] For our purposes, we do not need to understand how covariant components actually work. We only need to think of the covariant representation as another way of representing vectors. Here is a YouTube video that perhaps will give you some insight into contravariant and covariant components: `https://youtu.be/rG2q77qunSw`.

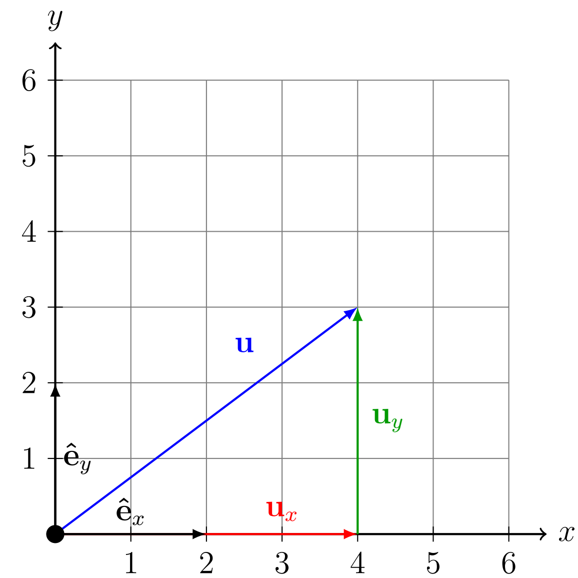

Figure 20: A 2D Cartesian coordinate system with a position vector $\mathbf{u}$ and basis vectors with doubled lengths.

$$\mathbf{r} = \mathbf{r}_1 + \mathbf{r}_2 = r_1\hat{\mathbf{e}}^1 + r_2\hat{\mathbf{e}}^2.$$

Here, we can see that upper indices represent contravariant components and lower indices represent covariant components. To generalise, we can express any vector $\mathbf{r}$ as a sum of the products of a contravariant or covariant component with a basis vector. For contravariant components, we write

$$\mathbf{r} = \sum_{i=1}^{n} r^i\hat{\mathbf{e}}_i = r^1\hat{\mathbf{e}}_1 + r^2\hat{\mathbf{e}}_2 + \cdots + r^n\hat{\mathbf{e}}_n. \tag{2.53}$$

For covariant components, we write

$$\mathbf{r} = \sum_{i=1}^{n} r_i\hat{\mathbf{e}}^i = r_1\hat{\mathbf{e}}^1 + r_2\hat{\mathbf{e}}^2 + \cdots + r_n\hat{\mathbf{e}}^n. \tag{2.54}$$

The covariant components r_i of a vector $\mathbf{r}$ are given by the dot product of $\mathbf{r}$ with the corresponding basis vector $\hat{\mathbf{e}}_i$,[17] as in

$$r_i = \mathbf{r} \cdot \hat{\mathbf{e}}_i.$$

In special relativity, we frequently deal with spacetime, which is four-dimensional. Hence, we use *4-vectors* to represent vectors in 4D. A 4-vector is a vector with a time component

[17] See Appendix B for vector operations.

and three space components. The 4D analogue of the position vector in Euclidean space is the *4-position* **X**, which we denote using contravariant components as

$$\begin{aligned}
X^\mu &= \sum_\mu X^\mu \hat{\mathbf{e}}_\mu \\
&= X^0\hat{\mathbf{e}}_0 + X^1\hat{\mathbf{e}}_1 + X^2\hat{\mathbf{e}}_2 + X^3\hat{\mathbf{e}}_3 \\
&= \left(X^0, X^1, X^2, X^3\right) \\
&= (ct, x, y, z).
\end{aligned} \tag{2.55}$$

In Equation 2.55, we have an index variable μ that appears twice within a summation, implying that we are summing $X^\mu \hat{\mathbf{e}}_\mu$ over all possible values of μ. Hence, assuming the index μ ranges over the values 0, 1, 2, and 3, we write X^μ as[18]

$$X^\mu = X^\mu \hat{\mathbf{e}}_\mu. \tag{2.56}$$

This is known as the *Einstein summation convention* or *Einstein notation*. Similarly, denoting **X** using covariant components, we have[19]

$$\begin{aligned}
X_\mu &= \sum_\mu X_\mu \hat{\mathbf{e}}^\mu \\
&= X_0\hat{\mathbf{e}}^0 + X_1\hat{\mathbf{e}}^1 + X_2\hat{\mathbf{e}}^2 + X_3\hat{\mathbf{e}}^3 \\
&= (X_0, X_1, X_2, X_3) \\
&= (ct, -x, -y, -z),
\end{aligned} \tag{2.57}$$

which, when expressed using Einstein notation, becomes[20]

$$X_\mu = X_\mu \hat{\mathbf{e}}^\mu. \tag{2.58}$$

We can hence see that

$$X^\mu = \begin{pmatrix} X^0 \\ X^1 \\ X^2 \\ X^3 \end{pmatrix} = \begin{pmatrix} X_0 \\ -X_1 \\ -X_2 \\ -X_3 \end{pmatrix} \tag{2.59}$$

and that

[18]Note that when X^μ is written on its own, it refers to the vector components of **X** expressed using contravariant components.

[19]Note that when we mention contravariance or covariance, we only talk about a vector's components being contravariant or covariant. A vector remains unchanged no matter how we represent it, but it can be represented by contravariant or covariant components. The same applies to tensors, as we will see later on in this section.

[20]Similarly, when X_μ is written on its own, it refers to the vector components of **X** expressed using covariant components.

$$X_\mu = \begin{pmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} X^0 \\ -X^1 \\ -X^2 \\ -X^3 \end{pmatrix}. \tag{2.60}$$

In special relativity, however, it is common to denote the 4-position $\mathbf{X}$ using contravariant components because it is more familiar to our original understanding of how vectors work or are represented. Though, we will quickly see the purpose of introducing the covariant representation of $\mathbf{X}$.

Recall that the spacetime interval $(\Delta s)^2$ is written as

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2.$$

We can also write $(\Delta s)^2$ as the dot product or scalar product of the displacements of the contravariant and covariant 4-positions, which we will call the *4-displacement*, $\Delta\mathbf{X}$:

$$\begin{aligned} (\Delta s)^2 &= \Delta X_\mu \Delta X^\mu \\ &= \begin{pmatrix} \Delta X_0 & \Delta X_1 & \Delta X_2 & \Delta X_3 \end{pmatrix} \begin{pmatrix} \Delta X^0 \\ \Delta X^1 \\ \Delta X^2 \\ \Delta X^3 \end{pmatrix} \\ &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \end{aligned} \tag{2.61}$$

Note that the expression $\Delta X_\mu \Delta X^\mu$ in Equation 2.61 uses Einstein notation, so to write the summation out explicitly, we have

$$\begin{aligned} \Delta X_\mu \Delta X^\mu &= \sum_\mu \Delta X_\mu \Delta X^\mu \\ &= \Delta X_0 \Delta X^0 + \Delta X_1 \Delta X^1 + \Delta X_2 \Delta X^2 + \Delta X_3 \Delta X^3 \\ &= (c\Delta t)(c\Delta t) + (-\Delta x)(\Delta x) + (-\Delta y)(\Delta y) + (-\Delta z)(\Delta z) \\ &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \end{aligned}$$

Given Equations 2.59 and 2.60, we can see that X^μ can be written in terms of the matrix product of a specific 4×4 matrix with X_μ and that X_μ can also be written in terms of the matrix product of the same 4×4 matrix with X^μ. That 4×4 matrix is called the *Minkowski metric tensor*, which we will denote as $\boldsymbol{\eta}$. The contravariant and covariant representations of $\boldsymbol{\eta}$ are equal, so we write[21]

[21] We have used a metric signature of $(+ - - -)$ to define the Minkowski metric tensor $\boldsymbol{\eta}$. Some authors would define $\boldsymbol{\eta}$ as having a metric signature of $(- + + +)$. Again, this is all down to convention and both are valid, but we would have to then be consistent with the definitions of other quantities (such as the spacetime interval $(\Delta s)^2$).

$$\boldsymbol{\eta} = \eta^{\mu\nu} = \eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{2.62}$$

Writing the matrices of $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$ in index form, we have[22]

$$\eta^{\mu\nu} = \begin{pmatrix} \eta^{00} & \eta^{01} & \eta^{02} & \eta^{03} \\ \eta^{10} & \eta^{11} & \eta^{12} & \eta^{13} \\ \eta^{20} & \eta^{21} & \eta^{22} & \eta^{23} \\ \eta^{30} & \eta^{31} & \eta^{32} & \eta^{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and

$$\eta_{\mu\nu} = \begin{pmatrix} \eta_{00} & \eta_{01} & \eta_{02} & \eta_{03} \\ \eta_{10} & \eta_{11} & \eta_{12} & \eta_{13} \\ \eta_{20} & \eta_{21} & \eta_{22} & \eta_{23} \\ \eta_{30} & \eta_{31} & \eta_{32} & \eta_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

The Minkowski metric tensor is an example of a *metric tensor*; a metric tensor is a mathematical object that defines properties including distances, angles, and in the case of special relativity, time and types of separation (timelike, spacelike, and lightlike) between events in a geometric space or on a manifold.[23] In the case of special relativity, it also captures the causal structure of spacetime. In this paper and special relativity, the spacetime we deal with, also called *Minkowski spacetime*, is *flat*, meaning the spatial components (x, y, z) obey the Pythagorean theorem. In non-Euclidean geometry, the Pythagorean theorem for distances no longer holds. Hence, the Minkowski metric tensor describes a space with flat geometry.

Recall that the distance $\Delta\mathbf{r}$ between two points in 3D Euclidean space is calculated using the equation

$$\Delta\mathbf{r}^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.$$

Since 3D Euclidean distance consists of three components (x, y, z) each with a plus sign, we know that the *Euclidean metric tensor* must have all its diagonal elements be $+1$. This is similar to the 4D Minkowski metric tensor, which has all its diagonal elements being ± 1. Hence, setting (x, y, z) as equivalent to (X^1, X^2, X^3) and expressing $(\Delta\mathbf{r})^2$ in terms of the Euclidean metric tensor (which we will denote as g_{ij}) in matrix notation, we have[24]

[22]For all 2D matrices, the left index denotes the row of the matrix and the right index denotes the column of the matrix.

[23]A *manifold* is a geometric or topological space (or object) that locally resembles Euclidean space at each point.

[24]See Appendix C for tensor operations.

$$
\begin{aligned}
(\Delta\mathbf{r})^2 &= \Delta X^i g_{ij} \Delta X^j \\
&= \begin{pmatrix} \Delta X^1 & \Delta X^2 & \Delta X^3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta X^1 \\ \Delta X^2 \\ \Delta X^3 \end{pmatrix} \\
&= (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2.
\end{aligned}
$$

Here, the indices i and j range over the values 1, 2, and 3. We see that g_{ij} is the 3×3 identity matrix $\mathbf{I}_3$, which we may alternatively denote as a function called the *Kronecker delta* δ_{ij}, which is defined as

$$
\delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.63}
$$

Hence, another way of determining whether a space is flat or curved is by seeing if we can apply a transformation to the metric tensor (describing that space) so that its entries become the Kronecker delta δ_{ij} (or $\delta_{\mu\nu}$ for a 4×4 matrix).[25] For the 3D case, the entries of the Euclidean metric tensor already equal δ_{ij}, hence the space is flat. For the 4D case, we can apply the *inverse Minkowski metric tensor* $\boldsymbol{\eta}^{-1}$ to $\boldsymbol{\eta}$, which is by definition equal to the Minkowski metric tensor itself, to obtain the Kronecker delta, like so:[26]

$$
\begin{aligned}
\boldsymbol{\eta}^{-1}\boldsymbol{\eta} = \eta^{\mu\sigma}\eta_{\sigma\nu} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \delta^{\mu}{}_{\nu}.
\end{aligned}
$$

Hence, we know that Minkowski spacetime is flat.

We now know that g_{ij} and $\eta_{\mu\nu}$ are metric tensors for 3D Euclidean space and 4D Minkowski spacetime, respectively. They each determine the concept of distance and the *standard basis* of the space they are describing. We have briefly discussed the concept of basis vectors earlier in this section — in fact, the standard basis of a geometric space is defined by basis vectors. More specifically, the standard basis of a geometric space is the set of all *orthonormal* vectors (that is, vectors that each have a length of 1 and are mutually orthogonal) — these are the basis vectors. For instance, in 3D Euclidean space, the standard basis is made up of three orthonormal bases $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$ (or three orthonormal dual bases

[25] Note that the *entries* of the identity matrix are equal to the Kronecker delta; the Kronecker delta is not itself a matrix.

[26] Note that $\delta_{\mu\nu} = \delta^{\mu\nu} = \delta_{\mu}{}^{\nu} = \delta^{\mu}{}_{\nu}$.

$\hat{\mathbf{e}}^1$, $\hat{\mathbf{e}}^2$, and $\hat{\mathbf{e}}^3$),[27] each of which has a length (or magnitude) of 1. They are defined as[28]

$$\hat{\mathbf{e}}_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \; \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \; \hat{\mathbf{e}}_3 = \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

and

$$\hat{\mathbf{e}}^1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \; \hat{\mathbf{e}}^2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; \hat{\mathbf{e}}^3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Note that if two bases or dual bases are mutually orthogonal, their dot products are 0.[29] That is,

$$\begin{aligned} \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_3 = \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_3 = 0, \\ \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^2 = \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^3 = \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^3 = 0. \end{aligned} \tag{2.64}$$

However, the dot product of a basis vector or dual basis vector with themselves gives 1, as in

$$\begin{aligned} \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_3 = 1, \\ \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^1 = \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^2 = \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^3 = 1. \end{aligned} \tag{2.65}$$

Hence, we can see that the Euclidean metric tensor g_{ij} and its inverse g^{ij} can be expressed in terms of the dot product of basis vectors and dual basis vectors, respectively. Using Equations 2.64 and 2.65, we can write

$$g_{ij} = \hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \begin{pmatrix} \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$g^{ij} = \hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}^j = \begin{pmatrix} \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^3 \\ \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^3 \\ \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The same applies to the standard basis and metric tensor of Minkowski spacetime. The standard basis of Minkowski spacetime consists of four bases $\hat{\mathbf{e}}_0$, $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$ (or four dual bases $\hat{\mathbf{e}}^0$, $\hat{\mathbf{e}}^1$, $\hat{\mathbf{e}}^2$, and $\hat{\mathbf{e}}^3$), each defined as

$$\hat{\mathbf{e}}_0 = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}, \; \hat{\mathbf{e}}_1 = \begin{pmatrix} 0 & i & 0 & 0 \end{pmatrix}, \; \hat{\mathbf{e}}_2 = \begin{pmatrix} 0 & 0 & i & 0 \end{pmatrix}, \; \hat{\mathbf{e}}_3 = \begin{pmatrix} 0 & 0 & 0 & i \end{pmatrix} \tag{2.66}$$

[27] *Dual bases* are bases with covariant scalar components.

[28] In some cases, we will represent dual vectors or dual bases using row vectors. We may for our purposes regard it as convention.

[29] See Appendix B for vector operations.

and

$$\hat{\mathbf{e}}^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \ \hat{\mathbf{e}}^1 = \begin{pmatrix} 0 \\ -i \\ 0 \\ 0 \end{pmatrix}, \ \hat{\mathbf{e}}^2 = \begin{pmatrix} 0 \\ 0 \\ -i \\ 0 \end{pmatrix}, \ \hat{\mathbf{e}}^3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ -i \end{pmatrix}, \tag{2.67}$$

where $i^2 = -1$. The reason why the spatial dual basis vectors each have a component containing a $-i$ is due to the following definition:

$$\hat{\mathbf{e}}^\mu \cdot \hat{\mathbf{e}}_\nu = \delta^\mu{}_\nu.$$

That is, when we take the dot product of a basis vector $\hat{\mathbf{e}}_\nu$ and a dual basis vector $\hat{\mathbf{e}}^\mu$ with the same index (i.e. where $\mu = \nu$), we get 1.[30] Otherwise, we get 0.

Note that the spatial basis vectors and basis dual vectors each have a length of 1 since $|i| = |-i| = 1$[31] but in a complex geometric space. We can see that the Minkowski metric tensor $\eta_{\mu\nu}$ and its inverse $\eta^{\mu\nu}$ can be expressed in terms of the dot product of basis vectors and dual basis vectors, respectively. Using Equations 2.66 and 2.67, we can write

$$\eta_{\mu\nu} = \hat{\mathbf{e}}_\mu \cdot \hat{\mathbf{e}}_\nu = \begin{pmatrix} \hat{\mathbf{e}}_0 \cdot \hat{\mathbf{e}}_0 & \hat{\mathbf{e}}_0 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_0 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_0 \cdot \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_0 & \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_1 \cdot \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_0 & \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_2 \cdot \hat{\mathbf{e}}_3 \\ \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_0 & \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \cdot \hat{\mathbf{e}}_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

and

$$\eta^{\mu\nu} = \hat{\mathbf{e}}^\mu \cdot \hat{\mathbf{e}}^\nu = \begin{pmatrix} \hat{\mathbf{e}}^0 \cdot \hat{\mathbf{e}}^0 & \hat{\mathbf{e}}^0 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^0 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^0 \cdot \hat{\mathbf{e}}^3 \\ \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^0 & \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^1 \cdot \hat{\mathbf{e}}^3 \\ \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^0 & \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^2 \cdot \hat{\mathbf{e}}^3 \\ \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^0 & \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^1 & \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^2 & \hat{\mathbf{e}}^3 \cdot \hat{\mathbf{e}}^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Let us now return to discussing 4-vectors. Using the Minkowski metric tensor $\boldsymbol{\eta}$, we can convert between the contravariant 4-displacement ΔX^μ and the covariant 4-displacement ΔX_μ, as in

[30] We can verify this: $(-i)(i) = -(i^2) = -(-1) = 1$.

[31] The vertical bars $||$ represent the *absolute value* or *magnitude* of a number; this is analogous to the double vertical bars $\|\|$, which represent the *length* of a vector.

$$
\begin{aligned}
\Delta X^{\mu} &= \eta^{\mu\nu}\Delta X_{\nu} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \Delta X_0 \\ \Delta X_1 \\ \Delta X_2 \\ \Delta X_3 \end{pmatrix} \\
&= \begin{pmatrix} \Delta X_0 \\ -\Delta X_1 \\ -\Delta X_2 \\ -\Delta X_3 \end{pmatrix} \\
&= \begin{pmatrix} \Delta ct \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix}
\end{aligned} \tag{2.68}
$$

and

$$
\begin{aligned}
\Delta X_{\mu} &= \eta_{\mu\nu}\Delta X^{\nu} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \Delta X^0 \\ \Delta X^1 \\ \Delta X^2 \\ \Delta X^3 \end{pmatrix} \\
&= \begin{pmatrix} \Delta X^0 \\ -\Delta X^1 \\ -\Delta X^2 \\ -\Delta X^3 \end{pmatrix} \\
&= \begin{pmatrix} \Delta ct \\ -\Delta x \\ -\Delta y \\ -\Delta z \end{pmatrix}.
\end{aligned} \tag{2.69}
$$

The spacetime interval is thus given by

$$
\begin{aligned}
(\Delta s)^2 &= \Delta X^{\mu}\eta_{\mu\nu}\Delta X^{\nu} \\
&= \begin{pmatrix} \Delta X^0 & \Delta X^1 & \Delta X^2 & \Delta X^3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \Delta X^0 \\ \Delta X^1 \\ \Delta X^2 \\ \Delta X^3 \end{pmatrix} \\
&= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2.
\end{aligned} \tag{2.70}
$$

Equivalently, the spacetime interval can also be obtained by

$$
\begin{aligned}
(\Delta s)^2 &= \Delta X_\mu \eta^{\mu\nu} \Delta X_\nu \\
&= \begin{pmatrix} \Delta X_0 & \Delta X_1 & \Delta X_2 & \Delta X_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \Delta X_0 \\ \Delta X_1 \\ \Delta X_2 \\ \Delta X_3 \end{pmatrix} \\
&= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2.
\end{aligned} \tag{2.71}
$$

2.8 The Lorentz Transformation Matrix

We know that a Lorentz transformation from $\mathbf{X}$ to $\mathbf{X}'$ involves the following four equations:

$$
\begin{aligned}
ct' &= \gamma(ct - \beta x) \\
x' &= \gamma(x - \beta ct) \\
y' &= y \\
z' &= z.
\end{aligned}
$$

We can hence define the *Lorentz transformation matrix* $\mathbf{\Lambda}$ to be the matrix such that

$$
\mathbf{X}' = \mathbf{\Lambda X}. \tag{2.72}
$$

In index and matrix notation, the Lorentz transformation from $\mathbf{X}$ to $\mathbf{X}'$ is written as

$$
X'^\mu = \Lambda^\mu{}_\nu X^\nu = \begin{pmatrix} \Lambda^0{}_0 & \Lambda^0{}_1 & \Lambda^0{}_2 & \Lambda^0{}_3 \\ \Lambda^1{}_0 & \Lambda^1{}_1 & \Lambda^1{}_2 & \Lambda^1{}_3 \\ \Lambda^2{}_0 & \Lambda^2{}_1 & \Lambda^2{}_2 & \Lambda^2{}_3 \\ \Lambda^3{}_0 & \Lambda^3{}_1 & \Lambda^3{}_2 & \Lambda^3{}_3 \end{pmatrix} \begin{pmatrix} X^0 \\ X^1 \\ X^2 \\ X^3 \end{pmatrix}, \tag{2.73}
$$

which, when more explicitly written, is also expressed as

$$
\begin{aligned}
\begin{pmatrix} X'^0 \\ X'^1 \\ X'^2 \\ X'^3 \end{pmatrix} &= \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \\
&= \begin{pmatrix} \gamma(ct - \beta x) \\ \gamma(x - \beta ct) \\ y \\ z \end{pmatrix} \\
&= \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix}.
\end{aligned} \tag{2.74}
$$

Since we have the Lorentz transformation matrix $\mathbf{\Lambda}$ that transforms 4-vectors from an unprimed IRF to a primed IRF, we may define the *inverse Lorentz transformation matrix*

$\mathbf{\Lambda}^{-1}$ to be the matrix that transforms 4-vectors from a primed IRF to an unprimed IRF. Since $\mathbf{\Lambda}^{-1}$ is the inverse of $\mathbf{\Lambda}$, it is the matrix such that

$$\mathbf{\Lambda}^{-1}\mathbf{\Lambda} = \mathbf{\Lambda}\mathbf{\Lambda}^{-1} = \mathbf{I}_4, \tag{2.75}$$

where $\mathbf{I}_4$ is the 4×4 *identity matrix*, written as

$$\mathbf{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The inverse Lorentz matrix is denoted as $(\Lambda^{-1})^{\mu}{}_{\nu}$; however, we will alternatively denote it as $\tilde{\Lambda}^{\mu}{}_{\nu}$. Hence, the transformation from $\mathbf{X}'$ to $\mathbf{X}$ under an inverse Lorentz transformation is given by

$$X^{\mu} = \tilde{\Lambda}^{\mu}{}_{\nu} X'^{\nu} = \begin{pmatrix} \tilde{\Lambda}^{0}{}_{0} & \tilde{\Lambda}^{0}{}_{1} & \tilde{\Lambda}^{0}{}_{2} & \tilde{\Lambda}^{0}{}_{3} \\ \tilde{\Lambda}^{1}{}_{0} & \tilde{\Lambda}^{1}{}_{1} & \tilde{\Lambda}^{1}{}_{2} & \tilde{\Lambda}^{1}{}_{3} \\ \tilde{\Lambda}^{2}{}_{0} & \tilde{\Lambda}^{2}{}_{1} & \tilde{\Lambda}^{2}{}_{2} & \tilde{\Lambda}^{2}{}_{3} \\ \tilde{\Lambda}^{3}{}_{0} & \tilde{\Lambda}^{3}{}_{1} & \tilde{\Lambda}^{3}{}_{2} & \tilde{\Lambda}^{3}{}_{3} \end{pmatrix} \begin{pmatrix} X'^{0} \\ X'^{1} \\ X'^{2} \\ X'^{3} \end{pmatrix}, \tag{2.76}$$

which, when explicitly written, is also expressed as

$$\begin{aligned} \begin{pmatrix} X^0 \\ X^1 \\ X^2 \\ X^3 \end{pmatrix} &= \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} \\ &= \begin{pmatrix} \gamma(ct' + \beta x') \\ \gamma(x' + \beta ct') \\ y' \\ z' \end{pmatrix} \\ &= \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}, \end{aligned} \tag{2.77}$$

corresponding accurately to the inverse Lorentz transformation equations below:

$$\begin{aligned} ct &= \gamma(ct + \beta x) \\ x &= \gamma(x + \beta ct) \\ y &= y' \\ z &= z'. \end{aligned}$$

We can multiply and check that Equation 2.75 does indeed hold:

$$
\begin{aligned}
\mathbf{\Lambda}^{-1}\mathbf{\Lambda} &= \begin{pmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \gamma^2 - \gamma^2\beta^2 & -\gamma^2\beta + \gamma^2\beta & 0 & 0 \\ \gamma^2\beta - \gamma^2\beta & \gamma^2 - \gamma^2\beta^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \gamma^2(1-\beta^2) & 0 & 0 & 0 \\ 0 & \gamma^2(1-\beta^2) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \mathbf{I}_4.
\end{aligned}
$$

3 Electrodynamics

3.1 Maxwell's Equations

Maxwell's equations are four partial differential equations that govern the behaviour of electric and magnetic fields, involving the divergence and curl operators.[32] The first equation is Gauss's law for electricity, written as

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \tag{3.1}$$

Equation 3.1 states that the divergence of the electric field is proportional to the charge density. For instance, a proton has a positive charge density, implying that the electric field diverges outwards from the proton (i.e. positive divergence). The constant ϵ_0 is called the *vacuum electric permittivity*, defined as $8.854\,187\,812\,8(13) \times 10^{-12}$ F/m (farads per metre). However, in the absence of a charge, as is the case for electric and magnetic fields in free space, Gauss's law for electricity becomes

$$\nabla \cdot \mathbf{E} = 0. \tag{3.2}$$

The second equation is Gauss's law for magnetism, written as

$$\nabla \cdot \mathbf{B} = 0. \tag{3.3}$$

[32]See Appendix D for vector calculus operators.

Equation 3.3 states that there is no divergence of the magnetic field since the presence of one necessitates the existence of magnetic monopoles, which is impossible given our current knowledge of physics.

The third equation is Faraday's law of induction, written as

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{3.4}$$

Equation 3.4 states that the curl of the electric field is equal to the negative of the rate of change of the magnetic field. One example is the production of a magnetic field (similar to that of a bar magnet) as current flows through a solenoid.

The fourth and final law is Ampère's law, written as

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right). \tag{3.5}$$

Equation 3.5 states that the curl of the magnetic field is proportional to the sum of the current density of a wire and the rate of change of the electric field with respect to time. One example is the production of a rotational magnetic field around a straight wire in which a current is flowing through. The constant μ_0 is called the *vacuum magnetic permeability*, defined as $1.256\,637\,062\,12(19) \times 10^{-6}\,\mathrm{N/A^2}$ (newtons per ampere). However, in the absence of an electric current, as is the case for electric and magnetic fields in free space, Ampere's law becomes

$$\nabla \times \mathbf{B} = \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \tag{3.6}$$

3.2 Derivation of the Constancy of the Speed of Light

In this section, we will go through the mathematical derivation of the constancy of the speed of light c, making use of Maxwell's equations. Note that we will assume an absence of charge and current throughout the derivation (meaning we are deriving the speed of light in free space). First, we take the curl of Faraday's law of induction (Equation 3.4) to get

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left(-\frac{\partial \mathbf{B}}{\partial t} \right) = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}). \tag{3.7}$$

Using the curl of curl identity in vector calculus, which is

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E},$$

we can plug the rightmost expression from Equation 3.7 in for $\nabla \times (\nabla \times \mathbf{E})$ to get

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\frac{\partial}{\partial t}(\nabla \times \mathbf{B}). \tag{3.8}$$

Assuming an absence of current, using Equation 3.6, we see that

$$-\frac{\partial}{\partial t}(\nabla \times \mathbf{B}) = -\mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{3.9}$$

Plugging Equation 3.9 into Equation 3.8, we obtain

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = -\mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

Because we are assuming an absence of charge (i.e. in free space), the divergence of the electric field is 0 (see Equation 3.2), implying that its gradient is also 0. Hence, we get

$$\nabla^2\mathbf{E} = \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{3.10}$$

Note that the negative signs are cancelled out in Equation 3.10. The equation above is the 3D wave equation, however, we only need to focus on the propagation of a wave in the x direction, meaning we take ∇^2 to be the second derivative with respect to x. Then, from equation 3.10, we have

$$\begin{aligned} \mu_0\epsilon_0\frac{\partial^2 \mathbf{E}}{\partial t^2} &= \frac{\partial^2 \mathbf{E}}{\partial x^2} \\ &= \frac{\partial}{\partial x}\left(\frac{\partial \mathbf{E}}{\partial x}\right) \\ &= \frac{\partial}{\partial x}\left(\frac{\partial \mathbf{E}}{\partial t}\frac{\partial t}{\partial x}\right). \end{aligned} \tag{3.11}$$

Since the velocity of the propagating wave in the x direction is

$$v = \frac{\partial x}{\partial t},$$

we can rewrite Equation 3.11 as

$$\begin{aligned} \frac{\partial^2 \mathbf{E}}{\partial x^2} &= \frac{\partial}{\partial x}\left(\frac{\partial \mathbf{E}}{\partial t}\frac{1}{v}\right) \\ &= \frac{\partial}{\partial t}\frac{\partial t}{\partial x}\left(\frac{\partial \mathbf{E}}{\partial t}\frac{1}{v}\right) \\ &= \frac{\partial}{\partial t}\frac{1}{v}\left(\frac{\partial \mathbf{E}}{\partial t}\frac{1}{v}\right). \end{aligned}$$

Hence, we are left with the equation

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 \mathbf{E}}{\partial t^2}. \tag{3.12}$$

From Equations 3.10 and 3.12, we can see that

$$\frac{1}{v^2} = \mu_0 \epsilon_0,$$

and so

$$v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = c. \tag{3.13}$$

Since μ_0 and ϵ_0 are constants, $1/\sqrt{\mu_0 \epsilon_0}$ must also be a constant. And since light is an electromagnetic wave, Equation 3.13 is valid in describing the speed of light following Maxwell's equations. We have hence derived the constancy of the speed of light c.

3.3 Lorentz Force and the Electromagnetic Tensor

In this section, we will discuss the mathematical formulation of the force exerted on charged particles by electric and magnetic fields. We will also introduce (later on) a new type of tensor related to the electric and magnetic fields and some new 4-vectors that will help us with the mathematical formulation.

The force exerted on a charged particle by an electric field, called the *electric force*, is given by

$$\mathbf{F_E} = q\mathbf{E},$$

where q denotes the charge of the particle and $\mathbf{E}$ denotes the electric field (which is a vector field). The force exerted on a charged particle by a magnetic field, called the *magnetic force*, is given by

$$\mathbf{F_B} = q(\mathbf{v} \times \mathbf{B}),$$

where $\mathbf{B}$ denotes the magnetic field (which is a vector field) and $\times$ denotes the cross product.[33] The total force $\mathbf{F}$ exerted on the charged particle is thus given by

$$\mathbf{F} = \mathbf{F_E} + \mathbf{F_B} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{3.14}$$

Equation 3.14 is called the *Lorentz force law*. From Equation 3.14, we can see that for a motionless particle (relative to some observer), the magnetic force on the particle vanishes. It is also the case that the faster the particle, the larger the magnetic force on the particle.

We can write Equation 3.14 in full in terms of its components (F_x, F_y, F_z), as in

$$\begin{aligned} F_x &= q(E_x + v_y B_z - v_z B_y), \\ F_y &= q(E_y + v_z B_x - v_x B_z), \\ F_z &= q(E_z + v_x B_y - v_y B_x). \end{aligned} \tag{3.15}$$

[33]See Appendix B for vector operations.

Before we continue with the Lorentz force law, it would be convenient to introduce two new quantities here. The first quantity we will introduce is the *electric potential* ϕ, which is a scalar field. It is defined as the electric potential energy U divided by the charge of a particle q, as in

$$\phi = \frac{U}{q}. \tag{3.16}$$

The other quantity we will introduce is the *magnetic vector potential* $\mathbf{A}$, which is a vector field. Magnetic vector potential is defined as a quantity such that its curl is the magnetic field $\mathbf{B}$. Hence, we have our first important equation linking magnetic vector potential and the magnetic field:

$$\mathbf{B} = \nabla \times \mathbf{A}. \tag{3.17}$$

We know that Equation 3.17 is valid because if we look at Gauss's law for magnetism (Equation 3.3), substituting the equivalent expression $\nabla \times \mathbf{A}$ for $\mathbf{B}$ into Equation 3.3, we get

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0.$$

This follows from a vector identity in vector calculus, stating that the divergence of the curl of any vector field is 0. Another important equation that we will need later on is one relating the electric field $\mathbf{E}$, magnetic vector potential $\mathbf{A}$, and electric potential ϕ. The equation is

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \tag{3.18}$$

By rearranging Equation 3.18 and taking the curl on both sides, we recover Faraday's law of induction:

$$\begin{aligned} \nabla \times \mathbf{E} + \nabla \times (\nabla\phi) &= \nabla \times \mathbf{E} \\ &= -\frac{\partial}{\partial t}(\nabla \times \mathbf{A}) \\ &= -\frac{\partial \mathbf{B}}{\partial t}. \end{aligned}$$

The reason why $\nabla \times \mathbf{E} + \nabla \times (\nabla\phi) = \nabla \times \mathbf{E}$ is that the curl of the gradient of ϕ evaluates to 0. This follows from another vector calculus identity stating that the curl of the gradient of any scalar field is 0.

We will now introduce two new 4-vectors. The first one is the *4-potential*, denoted as $\boldsymbol{\mathsf{A}}$. The 4-potential comprises a scalar quantity ϕ/c as the time component and components of a 3-vector[34] $\mathbf{A}$ as the spatial components. The contravariant form of $\boldsymbol{\mathsf{A}}$ is given by

[34]A *3-vector* is a vector in 3D.

$$
\begin{aligned}
A^\mu &= \eta^{\mu\nu} A_\nu \\
&= \left(A^0, A^1, A^2, A^3\right) \\
&= \left(\frac{\phi}{c}, A_x, A_y, A_z\right).
\end{aligned} \tag{3.19}
$$

The covariant form of $\mathbf{A}$ is given by

$$
\begin{aligned}
A_\mu &= \eta_{\mu\nu} A^\nu \\
&= (A_0, A_1, A_2, A_3) \\
&= \left(\frac{\phi}{c}, -A_x, -A_y, -A_z\right).
\end{aligned} \tag{3.20}
$$

The other 4-vector we will introduce is the *4-gradient*, denoted as $\boldsymbol{\partial}$. The 4-gradient consists of a partial derivative with respect to time (multiplied by some scaling factor) and the ordinary del operator (that is, derivatives with respect to x, y, and z).[35]

The contravariant form of $\boldsymbol{\partial}$ is given by

$$
\begin{aligned}
\partial^\mu &= \eta^{\mu\nu} \partial_\nu \\
&= \left(\partial^0, \partial^1, \partial^2, \partial^3\right) \\
&= \left(\frac{\partial}{\partial X_0}, \frac{\partial}{\partial X_1}, \frac{\partial}{\partial X_2}, \frac{\partial}{\partial X_3}\right) \\
&= \left(\frac{1}{c}\frac{\partial}{\partial t}, -\nabla\right) \\
&= \left(\frac{1}{c}\frac{\partial}{\partial t}, -\frac{\partial}{\partial x}, -\frac{\partial}{\partial y}, -\frac{\partial}{\partial z}\right).
\end{aligned} \tag{3.21}
$$

The covariant form of $\boldsymbol{\partial}$ is given by

$$
\begin{aligned}
\partial_\mu &= \eta_{\mu\nu} \partial^\nu \\
&= (\partial_0, \partial_1, \partial_2, \partial_3) \\
&= \left(\frac{\partial}{\partial X^0}, \frac{\partial}{\partial X^1}, \frac{\partial}{\partial X^2}, \frac{\partial}{\partial X^3}\right) \\
&= \left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right) \\
&= \left(\frac{1}{c}\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right).
\end{aligned} \tag{3.22}
$$

Let us now expand Equations 3.17 and 3.18 and see what happens. We can write $\mathbf{B}$ and $\mathbf{E}$ in terms of their x, y, and z components. Writing out Equation 3.17 in full, we get

[35] See Appendix D fro vector calculus operators.

$$
\begin{aligned}
B_x &= \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z}, \\
B_y &= \frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z}, \\
B_z &= \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y}.
\end{aligned} \tag{3.23}
$$

We can write Equation 3.23 in terms of ∂^μ and A^μ, as in

$$
\begin{aligned}
B_x &= \left(-\partial^2 A^3 + \partial^3 A^2\right), \\
B_y &= \left(-\partial^1 A^3 + \partial^3 A^1\right), \\
B_z &= \left(-\partial^1 A^2 + \partial^2 A^1\right).
\end{aligned} \tag{3.24}
$$

Let us now turn to Equation 3.18. Writing it out fully, we get

$$
\begin{aligned}
E_x &= -\frac{\partial \phi}{\partial x} - \frac{\partial A_x}{\partial t}, \\
E_y &= -\frac{\partial \phi}{\partial y} - \frac{\partial A_y}{\partial t}, \\
E_z &= -\frac{\partial \phi}{\partial z} - \frac{\partial A_z}{\partial t}.
\end{aligned} \tag{3.25}
$$

We can write Equation 3.25 in terms of ∂^μ and A^μ, as in

$$
\begin{aligned}
E_x &= c\left(\partial^1 A^0 - \partial^0 A^1\right), \\
E_y &= c\left(\partial^2 A^0 - \partial^0 A^2\right), \\
E_z &= c\left(\partial^3 A^0 - \partial^0 A^3\right).
\end{aligned} \tag{3.26}
$$

Here, we shall introduce a new 4×4 tensor called the *electromagnetic tensor*, denoted as $\mathbf{F}$. We will now abbreviate the term "electromagnetic tensor" as "EM tensor". In matrix form, the contravariant form of $\mathbf{F}$ is expressed as

$$
F^{\mu\nu} = \eta^{\mu\sigma} F_{\sigma\rho} \eta^{\rho\nu} = \begin{pmatrix} F^{00} & F^{01} & F^{02} & F^{03} \\ F^{10} & F^{11} & F^{12} & F^{13} \\ F^{20} & F^{21} & F^{22} & F^{23} \\ F^{30} & F^{31} & F^{32} & F^{33} \end{pmatrix}. \tag{3.27}
$$

We will also use the following definition for $F^{\mu\nu}$:

$$
F^{\mu\nu} \equiv \partial^\mu A^\nu - \partial^\nu A^\mu. \tag{3.28}
$$

Finally, from Equations 3.24, 3.26, 3.27, and 3.28 we see that the contravariant form of the EM tensor is given by

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{3.29}$$

The covariant form of the EM tensor is expressed in matrix form as

$$F_{\mu\nu} = \eta_{\mu\sigma} F^{\sigma\rho} \eta_{\rho\nu} = \begin{pmatrix} F_{00} & F_{01} & F_{02} & F_{03} \\ F_{10} & F_{11} & F_{12} & F_{13} \\ F_{20} & F_{21} & F_{22} & F_{23} \\ F_{30} & F_{31} & F_{32} & F_{33} \end{pmatrix}. \tag{3.30}$$

Similar to the definition of $F^{\mu\nu}$ given in Equation 3.28, we define $F_{\mu\nu}$ as

$$F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{3.31}$$

Using the same approach as we did for deriving $F^{\mu\nu}$ or by using the covariant Minkowski metric tensor $\boldsymbol{\eta}$ twice to lower the indices of $F^{\mu\nu}$ (see Equation 3.30),[36] we see that $F_{\mu\nu}$ is given by

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}. \tag{3.32}$$

For the sake of completeness, we will write the corresponding expressions for the mixed EM tensors ${F^\mu}_\nu$ and ${F_\mu}^\nu$:

$$\begin{aligned} {F^\mu}_\nu &= F^{\mu\sigma} \eta_{\sigma\nu} \\ &= \eta^{\mu\sigma} F_{\sigma\nu} \\ &= \begin{pmatrix} {F^0}_0 & {F^0}_1 & {F^0}_2 & {F^0}_3 \\ {F^1}_0 & {F^1}_1 & {F^1}_2 & {F^1}_3 \\ {F^2}_0 & {F^2}_1 & {F^2}_2 & {F^2}_3 \\ {F^3}_0 & {F^3}_1 & {F^3}_2 & {F^3}_3 \end{pmatrix} \\ &= \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}; \end{aligned} \tag{3.33}$$

[36] See Appendix C for tensor operations.

$$
\begin{aligned}
F_\mu{}^\nu &= F_{\mu\sigma}\eta^{\sigma\nu} \\
&= \eta_{\mu\sigma}F^{\sigma\nu} \\
&= \begin{pmatrix} F_0{}^0 & F_0{}^1 & F_0{}^2 & F_0{}^3 \\ F_1{}^0 & F_1{}^1 & F_1{}^2 & F_1{}^3 \\ F_2{}^0 & F_2{}^1 & F_2{}^2 & F_2{}^3 \\ F_3{}^0 & F_3{}^1 & F_3{}^2 & F_3{}^3 \end{pmatrix} \\
&= \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}.
\end{aligned} \tag{3.34}
$$

Now that we are more familiar with the EM tensor $\mathbf{F}$, let us move onto other 4-vectors and the representation of the Lorentz force law using the EM tensor and 4-vectors. We know that the infinitesimal proper time interval $d\tau$ is written as

$$
d\tau = \frac{dt}{\gamma}
$$

as per the definition of the proper time interval given by Equation 2.51. Rearranging for γ, we find that

$$
\gamma = \frac{dt}{d\tau}. \tag{3.35}
$$

Velocity in 4D is represented by the *4-velocity* vector, denoted as $\mathbf{U}$, which is given by

$$
\mathbf{U} = \frac{d\mathbf{X}}{d\tau}. \tag{3.36}
$$

The contravariant form of $\mathbf{U}$ is given by

$$
\begin{aligned}
U^\mu &= \eta^{\mu\nu}U_\nu \\
&= \frac{dX^\mu}{d\tau} \\
&= \left(U^0, U^1, U^2, U^3\right) \\
&= \left(\frac{dX^0}{d\tau}, \frac{dX^1}{d\tau}, \frac{dX^2}{d\tau}, \frac{dX^3}{d\tau}\right) \\
&= \left(\frac{cdt}{d\tau}, \frac{dx}{d\tau}, \frac{dy}{d\tau}, \frac{dz}{d\tau}\right).
\end{aligned} \tag{3.37}
$$

Using Equation 3.35, we can also write

$$
\begin{aligned}
U^{\mu} &= \gamma \frac{dX^{\mu}}{dt} \\
&= \left(\gamma \frac{dX^0}{dt}, \gamma \frac{dX^1}{dt}, \gamma \frac{dX^2}{dt}, \gamma \frac{dX^3}{dt} \right) \\
&= (\gamma c, \gamma v_x, \gamma v_y, \gamma v_z) \,.
\end{aligned} \tag{3.38}
$$

The covariant form of $\mathbf{U}$ is given by

$$
\begin{aligned}
U_{\mu} &= \eta_{\mu\nu} U^{\nu} \\
&= \frac{dX_{\mu}}{d\tau} \\
&= (U_0, U_1, U_2, U_3) \\
&= \left(\frac{dX_0}{d\tau}, \frac{dX_1}{d\tau}, \frac{dX_2}{d\tau}, \frac{dX_3}{d\tau} \right) \\
&= \left(\frac{cdt}{d\tau}, -\frac{dx}{d\tau}, -\frac{dy}{d\tau}, -\frac{dz}{d\tau} \right).
\end{aligned} \tag{3.39}
$$

Using Equation 3.35, we can also write

$$
\begin{aligned}
U_{\mu} &= \gamma \frac{dX_{\mu}}{dt} \\
&= \left(\gamma \frac{dX_0}{dt}, \gamma \frac{dX_1}{dt}, \gamma \frac{dX_2}{dt}, \gamma \frac{dX_3}{dt} \right) \\
&= (\gamma c, -\gamma v_x, -\gamma v_y, -\gamma v_z) \,.
\end{aligned} \tag{3.40}
$$

Similar to that of velocity in 3D, momentum in 4D is represented by the *4-momentum* vector, denoted as $\mathbf{P}$, which is given by

$$
\mathbf{P} = m\mathbf{U}, \tag{3.41}
$$

where m is an object's *rest mass*, which is the mass of an object measured in its own IRF. Rest mass m is defined as

$$
m = \frac{m_{\text{rel}}}{\gamma},
$$

where m_{rel} is relativistic mass, which is the mass of an object measured by an observer moving relative to it at some velocity v. The contravariant form of $\mathbf{P}$ is given by

$$
\begin{aligned}
P^{\mu} &= \eta^{\mu\nu} P_{\nu} \\
&= mU^{\mu} \\
&= \left(P^0, P^1, P^2, P^3 \right) \\
&= (\gamma mc, \gamma m v_x, \gamma m v_y, \gamma m v_z).
\end{aligned} \tag{3.42}
$$

The covariant form of $\mathbf{P}$ is given by

$$
\begin{aligned}
P_\mu &= \eta_{\mu\nu} P^\nu \\
&= m U_\mu \\
&= (P_0, P_1, P_2, P_3) \\
&= (\gamma mc, -\gamma m v_x, -\gamma m v_y, -\gamma m v_z).
\end{aligned} \tag{3.43}
$$

We know that in classical mechanics, force $\mathbf{F}$ is expressed as

$$
\mathbf{F} = \frac{d\mathbf{p}}{dt}, \tag{3.44}
$$

where the non-relativistic (or classical) momentum $\mathbf{p}$ is the product of the rest mass m of the object under question with its velocity $\mathbf{v}$, as in

$$
\mathbf{p} = m\mathbf{v} = (mv_x, mv_y, mv_z).
$$

Force in 4D is represented by the *4-force* vector, denoted as f. In our case, we will let f represent the Lorentz force, given by

$$
\mathsf{f} = \frac{d\mathsf{P}}{d\tau} = m\frac{d\mathsf{U}}{d\tau} = m\frac{d^2\mathsf{X}}{d\tau^2}. \tag{3.45}
$$

The contravariant form of f is given by

$$
\begin{aligned}
f^\mu &= \frac{dP^\mu}{d\tau} \\
&= \left(\frac{dP^0}{d\tau}, \frac{dP^1}{d\tau}, \frac{dP^2}{d\tau}, \frac{dP^3}{d\tau}\right) \\
&= \left(m\frac{dU^0}{d\tau}, m\frac{dU^1}{d\tau}, m\frac{dU^2}{d\tau}, m\frac{dU^3}{d\tau}\right) \\
&= \left(\gamma mc\frac{d}{d\tau}, \gamma m\frac{dv_x}{d\tau}, \gamma m\frac{dv_y}{d\tau}, \gamma m\frac{dv_z}{d\tau}\right).
\end{aligned} \tag{3.46}
$$

The covariant form of f is given by

$$
\begin{aligned}
f_\mu &= \frac{dP_\mu}{d\tau} \\
&= \left(\frac{dP_0}{d\tau}, \frac{dP_1}{d\tau}, \frac{dP_2}{d\tau}, \frac{dP_3}{d\tau}\right) \\
&= \left(m\frac{dU_0}{d\tau}, m\frac{dU_1}{d\tau}, m\frac{dU_2}{d\tau}, m\frac{dU_3}{d\tau}\right) \\
&= \left(\gamma mc\frac{d}{d\tau}, -\gamma m\frac{dv_x}{d\tau}, -\gamma m\frac{dv_y}{d\tau}, -\gamma m\frac{dv_z}{d\tau}\right).
\end{aligned} \tag{3.47}
$$

Working with the contravariant form of f first, we can now relate the x, y, and z components of f^μ with Equation 3.14, as in

$$
\begin{aligned}
f^1 &= \frac{dP^1}{d\tau} = \gamma m \frac{dv_x}{d\tau} = \gamma q(E_x + v_y B_z - v_z B_y), \\
f^2 &= \frac{dP^2}{d\tau} = \gamma m \frac{dv_y}{d\tau} = \gamma q(E_y + v_z B_x - v_x B_z), \\
f^3 &= \frac{dP^3}{d\tau} = \gamma m \frac{dv_z}{d\tau} = \gamma q(E_z + v_x B_y - v_y B_x).
\end{aligned} \tag{3.48}
$$

In accordance with the EM tensor, assuming the index i ranges over the values 1, 2, and 3, we may also write the following for f^i:

$$
\begin{aligned}
f^i &= qF^{i\nu}U_\nu \\
&= q\left(F^{i0}U_0 + F^{i1}U_1 + F^{i2}U_2 + F^{i3}U_3\right) \\
&= \gamma q\left(F^{i0}c - F^{i1}v_x - F^{i2}v_y - F^{i3}v_z\right) \\
&= \gamma q(E_x + v_y B_z - v_z B_y) + \gamma q(E_y + v_z B_x - v_x B_z) + \gamma q(E_z + v_x B_y - v_y B_x) \\
&= \gamma q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).
\end{aligned} \tag{3.49}
$$

Notice that dividing f^i by γ gives the classical or Newtonian force vector whilst also giving us Equation 3.14. Hence, using Equation 3.49, we can define the contravariant form of $\mathbf{f}$ as

$$
\begin{aligned}
f^\mu &= qF^{\mu\nu}U_\nu \\
&= q \begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ E_x/c & 0 & -B_z & B_y \\ E_y/c & B_z & 0 & -B_x \\ E_z/c & -B_y & B_x & 0 \end{pmatrix} \begin{pmatrix} \gamma c \\ -\gamma v_x \\ -\gamma v_y \\ -\gamma v_z \end{pmatrix} \\
&= \gamma q \begin{pmatrix} \left(E_x v_x + E_y v_y + E_z v_z\right)/c \\ (E_x + v_y B_z - v_z B_y) \\ (E_y + v_z B_x - v_x B_z) \\ (E_z + v_x B_y - v_y B_x) \end{pmatrix}.
\end{aligned} \tag{3.50}
$$

Using Equation 3.50, we can therefore see that

$$
\begin{aligned}
f^0 &= \frac{dP^0}{d\tau} \\
&= m\frac{d}{d\tau}\gamma c \\
&= qF^{0\nu}U_\nu \\
&= q\left(F^{00}U_0 + F^{01}U_1 + F^{02}U_2 + F^{03}U_3\right) \\
&= \frac{\gamma q}{c}\left(E_x v_x + E_y v_y + E_z v_z\right) \\
&= \frac{\gamma q}{c}(\mathbf{E} \cdot \mathbf{v}).
\end{aligned} \tag{3.51}
$$

By mass-energy equivalence, energy $\mathcal{E}$ is given by

$$\mathcal{E} = \gamma m c^2 = m_{\mathrm{rel}} c^2. \tag{3.52}$$

Therefore, from Equations 3.51 and 3.52, we see that

$$f^0 = \frac{1}{c}\frac{d\mathcal{E}}{d\tau} = m\frac{d}{d\tau}\gamma c,$$

and thus

$$cf^0 = \frac{d\mathcal{E}}{d\tau} = m\frac{d}{d\tau}\gamma c^2 = \gamma q(\mathbf{E}\cdot\mathbf{v}). \tag{3.53}$$

We can also define the contravariant form of **f** as

$$\begin{aligned} f^\mu &= qF^\mu{}_\nu U^\nu \\ &= q\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ E_x/c & 0 & B_z & -B_y \\ E_y/c & -B_z & 0 & B_x \\ E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} \\ &= \gamma q\begin{pmatrix} (E_x v_x + E_y v_y + E_z v_z)/c \\ (E_x + v_y B_z - v_z B_y) \\ (E_y + v_z B_x - v_x B_z) \\ (E_z + v_x B_y - v_y B_x) \end{pmatrix}. \end{aligned} \tag{3.54}$$

Applying the same logic, the covariant form of **f** is defined as

$$\begin{aligned} f_\mu &= qF_{\mu\nu} U^\nu \\ &= q\begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}\begin{pmatrix} \gamma c \\ \gamma v_x \\ \gamma v_y \\ \gamma v_z \end{pmatrix} \\ &= \gamma q\begin{pmatrix} (E_x v_x + E_y v_y + E_z v_z)/c \\ -(E_x + v_y B_z - v_z B_y) \\ -(E_y + v_z B_x - v_x B_z) \\ -(E_z + v_x B_y - v_y B_x) \end{pmatrix} \end{aligned} \tag{3.55}$$

and

$$
\begin{aligned}
f_\mu &= qF_\mu{}^\nu U_\nu \\
&= q\begin{pmatrix} 0 & -E_x/c & -E_y/c & -E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix}\begin{pmatrix} \gamma c \\ -\gamma v_x \\ -\gamma v_y \\ -\gamma v_z \end{pmatrix} \\
&= \gamma q\begin{pmatrix} (E_x v_x + E_y v_y + E_z v_z)/c \\ -(E_x + v_y B_z - v_z B_y) \\ -(E_y + v_z B_x - v_x B_z) \\ -(E_z + v_x B_y - v_y B_x) \end{pmatrix}.
\end{aligned} \tag{3.56}
$$

A Matrix Multiplication

Suppose we have a 2×2 matrix $\mathbf{A}$ and a 2×1 matrix $\mathbf{B}$, which we can write as

$$
\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} b_{11} \\ b_{21} \end{pmatrix}.
$$

The *matrix product* $\mathbf{C}$ of $\mathbf{A}$ and $\mathbf{B}$ is given by

$$
\mathbf{C} = \mathbf{AB},
$$

which can also be written as

$$
\mathbf{C} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} b_{11} \\ b_{21} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} \\ a_{21}b_{11} + a_{22}b_{21} \end{pmatrix} = \begin{pmatrix} c_{11} \\ c_{21} \end{pmatrix}.
$$

Let us consider another example where we have a 4×4 matrix $\mathbf{D}$ and a 4×1 matrix $\mathbf{E}$, written as

$$
\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{pmatrix} \quad \text{and} \quad \mathbf{E} = \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \end{pmatrix}.
$$

The matrix product $\mathbf{F}$ of $\mathbf{D}$ and $\mathbf{E}$ is given by

$$
\begin{aligned}
\mathbf{F} &= \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{pmatrix} \begin{pmatrix} e_{11} \\ e_{21} \\ e_{31} \\ e_{41} \end{pmatrix} \\
&= \begin{pmatrix} d_{11}e_{11} + d_{12}e_{21} + d_{13}e_{31} + d_{14}e_{41} \\ d_{21}e_{11} + d_{22}e_{21} + d_{23}e_{31} + d_{24}e_{41} \\ d_{31}e_{11} + d_{32}e_{21} + d_{33}e_{31} + d_{34}e_{41} \\ d_{41}e_{11} + d_{42}e_{21} + d_{43}e_{31} + d_{44}e_{41} \end{pmatrix} \\
&= \begin{pmatrix} f_{11} \\ f_{21} \\ f_{31} \\ f_{41} \end{pmatrix}.
\end{aligned}
$$

Notice that when we multiply a 2×2 matrix by a 2×1 matrix, we get a 2×1 matrix, and when we multiply a 4×4 matrix by a 4×1 matrix, we get a 4×1 matrix. We may hence generalise this can say that when we multiply an $m \times n$ matrix by an $n \times p$ matrix, we get a $m \times p$ matrix.[37] Note that the matrix product of two matrices is only defined if the number of columns in the first matrix equals the number of rows in the second matrix, both denoted by n.

Generalising even further, we can express any matrix multiplication between two arbitrary matrices $\mathbf{X}$ and $\mathbf{Y}$ as

$$
\begin{aligned}
\mathbf{XY} &= \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{pmatrix} \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{pmatrix} \\
&= \begin{pmatrix} x_{11}y_{11} + \cdots + x_{1n}y_{n1} & x_{11}y_{12} + \cdots + x_{1n}y_{n2} & \cdots & x_{11}y_{1p} + \cdots + x_{1n}y_{np} \\ x_{21}y_{11} + \cdots + x_{2n}y_{n1} & x_{21}y_{12} + \cdots + x_{2n}y_{n2} & \cdots & x_{21}y_{1p} + \cdots + x_{2n}y_{np} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}y_{11} + \cdots + x_{mn}y_{n1} & x_{m1}y_{12} + \cdots + x_{mn}y_{n2} & \cdots & x_{m1}y_{1p} + \cdots + x_{mn}y_{np} \end{pmatrix} \\
&= \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1p} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mp} \end{pmatrix} \\
&= \mathbf{Z}.
\end{aligned}
$$

One thing to note here is that generally, for two arbitrary matrices $\mathbf{X}$ and $\mathbf{Y}$, the matrix product $\mathbf{XY}$ does not equal $\mathbf{YX}$. That is to say, matrix multiplication is not *commutative*. Another thing to note is that matrix multiplication is *associative*. To illustrate, suppose we

[37] Here, m denotes the number of rows in the first matrix, n denotes the number of columns in the first matrix and the number of rows in the second matrix, and p denotes the number of columns in the second matrix.

have three arbitrary matrices **A**, **B**, and **C**. Also, suppose we have the following matrix product:

$$\mathbf{ABC}.$$

To compute **ABC**, we can either compute **BC** first and then multiply it by **A**, as in

$$\mathbf{A}(\mathbf{BC}),$$

or we can compute **AB** first and then multiply it by **C**, as in

$$(\mathbf{AB})\mathbf{C}.$$

Both ways of computing **ABC** are equal.

B Vector Operations

B.1 The Dot Product

In Cartesian coordinates, the *dot product* (or the *scalar product*) between two vectors $\mathbf{v} = (v_1, v_2, \cdots, v_n)$ and $\mathbf{w} = (w_1, w_2, \cdots, w_n)$[38] is expressed as

$$\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

The geometric definition of the dot product of two vectors **v** and **w** is given by

$$\mathbf{v} \cdot \mathbf{w} = \|\mathbf{v}\| \|\mathbf{w}\| \cos\theta,$$

where θ is the angle between **v** and **w**. In matrix form, the dot product of two vectors **v** and **w** is given by the matrix product of $\mathbf{v}^\mathsf{T}$ and **w**, written as

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^\mathsf{T}\mathbf{w} = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n,$$

where T denotes the *transpose* of a matrix. Transposing a matrix flips it over its diagonal, meaning its row and column indices are switched. That is to say, for an arbitrary $m \times n$ matrix **M**, its transpose $\mathbf{M}^\mathsf{T}$ is an $n \times m$ matrix.

[38] Note that the subscripts here do not specifically denote covariance; the lower indices are simply for labelling.

B.2 The Cross Product

The *cross product* of two vectors is only defined in 3D. Hence, in Cartesian coordinates, the cross product of two vectors $\mathbf{v} = (v_1, v_2, v_3)$ and $\mathbf{w} = (w_1, w_2, w_3)$ is expressed as

$$\begin{aligned} \mathbf{v} \times \mathbf{w} &= \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \\ &= \hat{\mathbf{e}}_x(v_2w_3 - v_3w_2) - \hat{\mathbf{e}}_y(v_1w_3 - v_3w_1) + \hat{\mathbf{e}}_z(v_1w_2 - v_2w_1) \\ &= \hat{\mathbf{e}}_x(v_2w_3 - v_3w_2) + \hat{\mathbf{e}}_y(v_3w_1 - v_1w_3) + \hat{\mathbf{e}}_z(v_1w_2 - v_2w_1), \end{aligned}$$

where the vertical bars enclosing the 3×3 matrix represents the determinant of the matrix. The geometric definition of the cross product of two vectors $\mathbf{v}$ and $\mathbf{w}$ is given by

$$\mathbf{v} \times \mathbf{w} = \|\mathbf{v}\|\|\mathbf{w}\| \sin(\theta)\, \mathbf{n},$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$ and $\mathbf{n}$ is the unit vector perpendicular to $\mathbf{v}$ and $\mathbf{w}$.

C Tensor Operations

C.1 The Tensor-Vector Product

Suppose we have a $(0, 2)$-tensor[39] $A_{\mu\nu}$ with covariant components and a $(1, 0)$-tensor[40] B^ν with contravariant components. We can express $T_{\mu\nu}$ and X^ν as

$$A_{\mu\nu} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \quad \text{and} \quad B^\nu = \begin{pmatrix} B^1 \\ B^2 \\ B^3 \\ B^4 \end{pmatrix}.$$

The *tensor-vector product* C_μ of $A_{\mu\nu}$ and B^ν is written as

$$C_\mu = A_{\mu\nu}B^\nu,$$

which can also be expressed as[41]

[39] A tensor of type (n, m) means a tensor with n contravariant indices and m covariant indices.

[40] A tensor of type $(1, 0)$ is simply a vector.

[41] The tensor-vector product is essentially the same as the matrix product but for tensors and vectors, which can be expressed in matrix form.

$$
\begin{aligned}
C_\mu &= \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \begin{pmatrix} B^1 \\ B^2 \\ B^3 \\ B^4 \end{pmatrix} \\
&= \begin{pmatrix} A_{11}B^1 + A_{12}B^2 + A_{13}B^3 + A_{14}B^4 \\ A_{21}B^1 + A_{22}B^2 + A_{23}B^3 + A_{24}B^4 \\ A_{31}B^1 + A_{32}B^2 + A_{33}B^3 + A_{34}B^4 \\ A_{41}B^1 + A_{42}B^2 + A_{43}B^3 + A_{44}B^4 \end{pmatrix} \\
&= \begin{pmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{pmatrix}.
\end{aligned}
$$

Now suppose we have a $(2,0)$-tensor $D^{\mu\nu}$ with contravariant components and a $(0,1)$-tensor[42] E_ν with covariant components. We can express $D^{\mu\nu}$ and E_ν as

$$
D^{\mu\nu} = \begin{pmatrix} D^{11} & D^{12} & D^{13} & D^{14} \\ D^{21} & D^{22} & D^{23} & D^{24} \\ D^{31} & D^{32} & D^{33} & D^{34} \\ D^{41} & D^{42} & D^{43} & D^{44} \end{pmatrix} \quad \text{and} \quad E_\nu = \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix}.
$$

The tensor-vector product F^μ of $D^{\mu\nu}$ and E_ν is given by

$$
\begin{aligned}
F^\mu &= D^{\mu\nu} E_\nu \\
&= \begin{pmatrix} D^{11} & D^{12} & D^{13} & D^{14} \\ D^{21} & D^{22} & D^{23} & D^{24} \\ D^{31} & D^{32} & D^{33} & D^{34} \\ D^{41} & D^{42} & D^{43} & D^{44} \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ E_3 \\ E_4 \end{pmatrix} \\
&= \begin{pmatrix} D^{11}E_1 + D^{12}E_2 + D^{13}E_3 + D^{14}E_4 \\ D^{21}E_1 + D^{22}E_2 + D^{23}E_3 + D^{24}E_4 \\ D^{31}E_1 + D^{32}E_2 + D^{33}E_3 + D^{34}E_4 \\ D^{41}E_1 + D^{42}E_2 + D^{43}E_3 + D^{44}E_4 \end{pmatrix} \\
&= \begin{pmatrix} F^1 \\ F^2 \\ F^3 \\ F^4 \end{pmatrix}.
\end{aligned}
$$

Suppose we have a $(1,1)$-tensor $G^\mu{}_\nu$ with mixed components and the $(1,0)$-tensor B^ν with contravariant components (which we have introduced before). We can express $G^\mu{}_\nu$ and B^ν as

[42] A tensor of type $(0,1)$ is a *covector*.

$$G^\mu{}_\nu = \begin{pmatrix} G^0{}_0 & G^0{}_1 & G^0{}_2 & G^0{}_3 \\ G^1{}_0 & G^1{}_1 & G^1{}_2 & G^1{}_3 \\ G^2{}_0 & G^2{}_1 & G^2{}_2 & G^2{}_3 \\ G^3{}_0 & G^3{}_1 & G^3{}_2 & G^3{}_3 \end{pmatrix} \quad \text{and} \quad B^\nu = \begin{pmatrix} B^0 \\ B^1 \\ B^2 \\ B^3 \end{pmatrix}.$$

The tensor-vector product H^μ of $G^\mu{}_\nu$ and B^ν is given by

$$\begin{aligned} H^\mu &= G^\mu{}_\nu B^\nu \\ &= \begin{pmatrix} G^0{}_0 & G^0{}_1 & G^0{}_2 & G^0{}_3 \\ G^1{}_0 & G^1{}_1 & G^1{}_2 & G^1{}_3 \\ G^2{}_0 & G^2{}_1 & G^2{}_2 & G^2{}_3 \\ G^3{}_0 & G^3{}_1 & G^3{}_2 & G^3{}_3 \end{pmatrix} \begin{pmatrix} B^0 \\ B^1 \\ B^2 \\ B^3 \end{pmatrix} \\ &= \begin{pmatrix} G^1{}_1 B^1 + G^1{}_2 B^2 + G^1{}_3 B^3 + G^1{}_4 B^4 \\ G^2{}_1 B^1 + G^2{}_2 B^2 + G^2{}_3 B^3 + G^2{}_4 B^4 \\ G^3{}_1 B^1 + G^3{}_2 B^2 + G^3{}_3 B^3 + G^3{}_4 B^4 \\ G^4{}_1 B^1 + G^4{}_2 B^2 + G^4{}_3 B^3 + G^4{}_4 B^4 \end{pmatrix} \\ &= \begin{pmatrix} H^1 \\ H^2 \\ H^3 \\ H^4 \end{pmatrix}. \end{aligned}$$

Suppose we have a $(1,1)$-tensor $J_\mu{}^\nu$ with mixed components and the $(0,1)$-tensor E_ν with covariant components (which we have introduced before). We can express $J_\mu{}^\nu$ and E_ν as

$$J_\mu{}^\nu = \begin{pmatrix} J_0{}^0 & J_0{}^1 & J_0{}^2 & J_0{}^3 \\ J_1{}^0 & J_1{}^1 & J_1{}^2 & J_1{}^3 \\ J_2{}^0 & J_2{}^1 & J_2{}^2 & J_2{}^3 \\ J_3{}^0 & J_3{}^1 & J_3{}^2 & J_3{}^3 \end{pmatrix} \quad \text{and} \quad E_\nu = \begin{pmatrix} E_0 \\ E_1 \\ E_2 \\ E_3 \end{pmatrix}.$$

The tensor-vector product K^μ of $J_\mu{}^\nu$ and E_ν is given by

$$
\begin{aligned}
K_\mu &= J_\mu{}^\nu E_\nu \\
&= \begin{pmatrix} J_0{}^0 & J_0{}^1 & J_0{}^2 & J_0{}^3 \\ J_1{}^0 & J_1{}^1 & J_1{}^2 & J_1{}^3 \\ J_2{}^0 & J_2{}^1 & J_2{}^2 & J_2{}^3 \\ J_3{}^0 & J_3{}^1 & J_3{}^2 & J_3{}^3 \end{pmatrix} \begin{pmatrix} E_0 \\ E_1 \\ E_2 \\ E_3 \end{pmatrix} \\
&= \begin{pmatrix} J_1{}^1 E_1 + J_1{}^2 E_2 + J_1{}^3 E_3 + J_1{}^4 E_4 \\ J_2{}^1 E_1 + J_2{}^2 E_2 + J_2{}^3 E_3 + J_2{}^4 E_4 \\ J_3{}^1 E_1 + J_3{}^2 E_2 + J_3{}^3 E_3 + J_3{}^4 E_4 \\ J_4{}^1 E_1 + J_4{}^2 E_2 + J_4{}^3 E_3 + J_4{}^4 E_4 \end{pmatrix} \\
&= \begin{pmatrix} K_1 \\ K_2 \\ K_3 \\ K_4 \end{pmatrix}.
\end{aligned}
$$

C.2 Tensor Contraction and Raising and Lowering Indices

Suppose we have a metric tensor $g_{\mu\nu}$ with covariant components and a vector v^ν with contravariant components. If we take their tensor-vector product

$$g_{\mu\nu} v^\nu,$$

then we produce the vector with covariant components (also called a covector)

$$v_\mu.$$

Notice that the index μ in v_μ is lowered, that is, we have gone from v^ν to v_μ. The process of taking the tensor-vector product of a covariant metric tensor with a vector and obtaining a covector with a lower index is called *lowering the index*. It also seems that the index ν has "disappeared"; this is because we know from Einstein's summation convention that the summation over the index ν is implied. This process is called *tensor contraction*, where indices belonging to different tensors with summation implied are contracted to produce a new tensor.

Now suppose we have the inverse metric tensor $g^{\mu\nu}$ with contravariant components and a covector v_ν with covariant components. If we take their tensor-vector product

$$g^{\mu\nu} v_\nu,$$

then we produce a vector with contravariant components, written as v^μ. Notice that the index μ in v^μ is raised; the process of taking the tensor-vector product of a contravariant metric tensor with a covector and obtaining a vector with a lower index is called *raising the index*. We also see that the index ν is contracted due to the implied summation.

Following this logic, we can formulate the subsequent expressions for tensor contractions of two tensors V and W:

$$T^{\mu}{}_{\nu} = V^{\mu\sigma}W_{\sigma\nu} = V^{\mu}{}_{\sigma}W^{\sigma}{}_{\nu},$$
$$T_{\mu}{}^{\nu} = V_{\mu\sigma}W^{\sigma\nu} = V_{\mu}{}^{\sigma}W_{\sigma}{}^{\nu},$$

as well as

$$T^{\mu\nu} = V^{\mu\sigma}W_{\sigma}{}^{\nu} = V^{\mu}{}_{\sigma}W^{\sigma\nu},$$
$$T_{\mu\nu} = V_{\mu\sigma}W^{\sigma}{}_{\nu} = V_{\mu}{}^{\sigma}W_{\sigma\nu}.$$

D Vector Calculus Operators

D.1 The Del Operator

The *del operator* is a vector operator which can act on scalar or vector fields (more precisely, scalar- or vector-valued functions). In three dimensions, its components consist of partial derivatives with respect to x, y, and z. The del operator is written as

$$\begin{aligned}\nabla &\equiv \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)\\ &= \hat{\mathbf{e}}_x\frac{\partial}{\partial x} + \hat{\mathbf{e}}_y\frac{\partial}{\partial y} + \hat{\mathbf{e}}_z\frac{\partial}{\partial z}.\end{aligned}$$

D.2 The Gradient Operator

The *gradient operator* is essentially a generalisation of the ordinary derivative. It applies to a scalar field and outputs a vector field. The gradient is a measure of the direction and rate of greatest change at each point on a scalar field. The gradient of a scalar-valued function S of three variables x, y, and z is written as

$$\begin{aligned}\nabla S(x,y,z) &= \left(\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}, \frac{\partial S}{\partial z}\right)\\ &= \hat{\mathbf{e}}_x\frac{\partial S}{\partial x} + \hat{\mathbf{e}}_y\frac{\partial S}{\partial y} + \hat{\mathbf{e}}_z\frac{\partial S}{\partial z}.\end{aligned}$$

D.3 The Divergence Operator

The *divergence operator* applies to a vector field and outputs a scalar field. Divergence is a measure of how much a vector field flows outwards or inwards from a point. The divergence of a vector-valued function $\mathbf{F}$ of three components F_x, F_y, and F_z is written as

$$\begin{aligned}\nabla \cdot \mathbf{F}(F_x, F_y, F_z) &= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) \cdot (F_x, F_y, F_z)\\ &= \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.\end{aligned}$$

D.4 The Curl Operator

The *curl operator* applies to a vector field and outputs a vector field. Curl is a measure of how much a vector field rotates or circulates about a point and the general direction of the rotation. The curl of a vector-valued function $\mathbf{F}$ of three components F_x, F_y, and F_z is written as

$$\begin{aligned}
\nabla \times \mathbf{F}(F_x, F_y, F_z) &= \begin{vmatrix} \hat{\mathbf{e}}_x & \hat{\mathbf{e}}_y & \hat{\mathbf{e}}_z \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{vmatrix} \\
&= \hat{\mathbf{e}}_x \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) - \hat{\mathbf{e}}_y \left(\frac{\partial F_z}{\partial x} - \frac{\partial F_x}{\partial z} \right) + \hat{\mathbf{e}}_z \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \\
&= \hat{\mathbf{e}}_x \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z} \right) + \hat{\mathbf{e}}_y \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x} \right) + \hat{\mathbf{e}}_z \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right).
\end{aligned}$$

D.5 The Laplace Operator

The *Laplace operator* (also known as the *Laplacian*) applies to a scalar field and outputs a scalar field. The Laplacian is equal to the divergence of the gradient (of a scalar field). The Laplacian of a scalar-valued function S of three variables x, y, and z is written as

$$\begin{aligned}
\nabla^2 S(x, y, z) &= \nabla \cdot \nabla S(x, y, z) \\
&= \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \cdot \left(\frac{\partial S}{\partial x}, \frac{\partial S}{\partial y}, \frac{\partial S}{\partial z} \right) \\
&= \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2}.
\end{aligned}$$

Bibliography

[1] Susskind, Leonard & Freidman, Art. *Special Relativity and Classical Field Theory – The Theoretical Minimum.* Allen Lane, 2018.

[2] Susskind, Leonard & Cabannes, Anfré. *General Relativity – The Theoretical Minimum.* Allen Lane, 2018.

[3] Myers, Alan L. *SPECIAL RELATIVITY.* University of Pennsylvania, 2016, `https://www.seas.upenn.edu/~amyers/SpecRel.pdf`.

[4] Myers, Alan L. *GENERAL RELATIVITY IRREDUCIBLE MINIMUM.* University of Pennsylvania, 2016, `https://www.seas.upenn.edu/~amyers/DualBasis.pdf`.

[5] Nadia, Zakamska L. *Theory of Special Relativity.* arXiv, 2018, `https://arxiv.org/abs/1511.02121`.

[6] Cresser, J.D. *THE SPECIAL THEORY OF RELATIVITY.* Macquarie University, 2003, `https://physics.mq.edu.au/~jcresser/Phys378/LectureNotes/SpecialRelativityNotes.pdf`.

[7] Tong, David. *7. Special Relativity.* University of Cambridge, 2013, `http://www.damtp.cam.ac.uk/user/tong/relativity/seven.pdf`.

[8] Tong, David. *5. Electromagnetism and Relativity.* University of Cambridge, 2015, `https://www.damtp.cam.ac.uk/user/tong/em/el4.pdf`.

[9] Tong, David. *Electromagnetism.* University of Cambridge, 2015, `https://www.damtp.cam.ac.uk/user/tong/em/em.pdf`.

[10] Tong, David. *Vector Calculus.* University of Cambridge, 2021, `https://www.damtp.cam.ac.uk/user/tong/vc/vc.pdf`.

[11] Hunt, Robert. *Chapter 3 – Cartesian Tensors.* University of Cambridge, 2002, `https://www.damtp.cam.ac.uk/user/reh10/lectures/nst-mmii-chapter3.pdf`.

[12] Rowe, Glenn. *RELATIVISTIC UNITS.* Physics Pages, 2021, `http://physicspages.com/pdf/Relativity/Relativistic%20units.pdf`.

[13] Rowe, Glenn. *THE LIGHT CLOCK AND TIME DILATION.* Physics Pages, 2023, `http://physicspages.com/pdf/Astrophysics/The%20light%20clock%20and%20time%20dilation.pdf`.

[14] Rowe, Glenn. *TIME DILATION AND PROPER TIME.* Physics Pages, 2021, `http://physicspages.com/pdf/Relativity/Time%20dilation%20and%20proper%20time.pdf`.

[15] Rowe, Glenn. *FOUR-VELOCITY.* Physics Pages, 2021, `http://physicspages.com/pdf/Relativity/Four-velocity.pdf`.

[16] Rowe, Glenn. *METRIC TENSOR – INVERSE AND RAISING AND LOWERING INDICES*. Physics Pages, 2021, https://physicspages.com/pdf/Relativity/Metric%20tensor%20-%20inverse%20and%20raising%20and%20lowering%20indices.pdf.

[17] Rowe, Glenn. *FORCE IN RELATIVITY*. Physics Pages, 2023, http://physicspages.com/pdf/Relativity/Force%20in%20relativity.pdf.

[18] Herr, Werner. *(Special) Relativity*. CERN Accelerator School, 2016, https://cas.web.cern.ch/sites/default/files/lectures/hamburg-2016/herrrelativity.pdf.

[19] Feynman, Richard. *The Feynman Lectures on Physics (Volume I): Chapter 15. The Special Theory of Relativity*. California Institute of Technology, 2013, https://www.feynmanlectures.caltech.edu/I_15.html.

[20] Feynman, Richard. *The Feynman Lectures on Physics (Volume I): Chapter 16. Relativistic Energy and Momentum*. California Institute of Technology, 2013, https://www.feynmanlectures.caltech.edu/I_16.html.

[21] Feynman, Richard. *The Feynman Lectures on Physics (Volume I): Chapter 17. Space-Time*. California Institute of Technology, 2013, https://www.feynmanlectures.caltech.edu/I_17.html.

[22] Feynman, Richard. *The Feynman Lectures on Physics (Volume II): Chapter 25. Electrodynamics in Relativistic Notation*. California Institute of Technology, 2013, https://www.feynmanlectures.caltech.edu/II_25.html.

[23] Feynman, Richard. *The Feynman Lectures on Physics (Volume II): Chapter 26. Lorentz Transformations of the Fields*. California Institute of Technology, 2013, https://www.feynmanlectures.caltech.edu/II_26.html.

[24] Kumar, Alok. *Covariant and Contravariant Vectors*. arXiv, 2010, https://arxiv.org/pdf/1002.3217.pdf.

[25] Sochi, Taha. *Tensor Calculus*. arXiv, 2016, https://arxiv.org/abs/1610.04347.

[26] Morin, David. *Chapter 12 – 4-vectors*. Kent State University, 2004, http://personal.kent.edu/~fwilliam/Chapter%2012%204-Vectors.pdf.

[27] Devanathan, V. *The Concept of Four-Vectors*. University of Madras, 2018, https://www.tnasc.com/wp-content/uploads/2018/04/4-vector-slide.pdf.

[28] Baird, Christopher S. *Lecture 13 Notes, Electromagnetic Theory II*. University of Massachusetts Lowell, n.d., https://www.wtamu.edu/~cbaird/BLecture13.pdf.

[29] Gubler, Martin. *Special Relativity with Four-Vectors*. PhysAstroMath, 2021, https://www.physastromath.ch/uploads/myPdfs/Relativ/STR%20with%20Four-Vectors.pdf.

[30] Komissarov, S.S. *SPECIAL RELATIVITY. MATH2410.* University of Leeds, 2018, https://www1.maths.leeds.ac.uk/~serguei/teaching/sr.pdf.

[31] Hirata, Christopher. *Lecture III: Vector algebra in special relativity.* The Ohio State University, 2019, https://hirata10.github.io/ph6820/lec03_vectors.pdf.

[32] Hughes, Scott A. *Lecture 4 – Spacetime, simultaneity, and the consequences of Lorentz.* Massachusetts Institute of Technology, 2021, https://web.mit.edu/sahughes/www/8.033/lec04.pdf.

[33] Hughes, Scott A. *Lecture 10 – Magnetic force; Magnetic fields; Ampere's law.* Massachusetts Institute of Technology, 2004, https://web.mit.edu/sahughes/www/8.022/lec10.pdf.

[34] Nardoni, Emily. *Lorentz tensor redux.* University of California, San Diego, 2015, https://keni.ucsd.edu/w15/Lorentz%20Transformations.pdf.

[35] Littlejohn, Robert G. *Lorentz Transformations in Special Relativity.* University of California, Berkeley 2019, https://bohr.physics.berkeley.edu/classes/221/1112/notes/lorentz.pdf.

[36] Landau, Rubin H. *FOUR-VECTORS AND LORENTZ TRANSFORMATIONS.* Wiley-VCH, 2004, https://onlinelibrary.wiley.com/doi/pdf/10.1002/9783527617432.app3.

[37] Challinor, Anthony. *Electrodynmaics – Lecture notes.* University of Cambridge, 2016, https://www.damtp.cam.ac.uk/user/examples/D21Lb.pdf.

[38] Porat, Boaz. *A Gentle Introduction to Tensors.* Israel Institute of Technology, 2014, https://www.ese.wustl.edu/~nehorai/Porat_A_Gentle_Introduction_to_Tensors_2014.pdf.

[39] Yamamoto, Hitoshi. *Chapter 1 – Lorentz Group and Lorentz Invariance.* The University of Hawaii, n.d., http://epx.phys.tohoku.ac.jp/~yhitoshi/particleweb/ptestup.pdf.

[40] Ramachers, Y.A. *Lecture Notes PX421: Relativity and Electrodynamics.* University of Warwick, 2014, https://warwick.ac.uk/fac/sci/physics/staff/academic/ramachers/px421_main.pdf.

[41] Haber, Howard E. *The velocity and momentum four-vectors.* University of California, Santa Cruz, 2015, http://scipp.ucsc.edu/~haber/ph171/uvector15.pdf.

[42] Playfer, Steve. *Electromagnetism – Lecture 18: Relativity & Electromagnetism.* The University of Edinburgh, n.d., https://www2.ph.ed.ac.uk/~playfer/EMlect18.pdf.

[43] Chang, Tsun-Hsu. *Chapter 10: Potentials and Fields.* National Tsing Hua University, n.d., http://www.phys.nthu.edu.tw/~thschang/notes/EM10.pdf.

[44] Tsymbal, Evgeny Y. *Section 6: Potentials and Fields.* University of Nebraska-Lincoln, n.d., https://unlcms.unl.edu/cas/physics/tsymbal/teaching/EM-914/section6-Potentials%20%26%20Fields.pdf.

[45] Christ. *Derivation of Lorentz Force Law.* Columbia University in the City of New York, 2021, https://www.columbia.edu/~nhc1/UN2802_s21/LorentzForce.pdf.

[46] Palmer, Christopher W.P. *S8: Covariant Electromagnetism.* University of Oxford, 2010, https://users.physics.ox.ac.uk/~palmerc/S8files/lectures.pdf.

[47] Baring, Matthew. *5. THE ELECTROMAGNETIC FIELD TENSOR.* Rice University, 2023, https://www.ruf.rice.edu/~baring/phys532/phys532_2023_lec_021723.pdf.

[48] Burgiel, Heidi. *V15.1 Del Operator.* Massachusetts Insitute of Technology, 2010, https://www.mit.edu/~hlb/1802/notes-apm-jmo/pdf/MIT18_02SC_MNotes_v15.1.pdf.

[49] Reany, P. *Vector Calculus Identities.* Advanced Mathematics, 2021, https://advancedmath.org/Math/Cliff-GA/VectorCalculus/VectCalc_Identities3.pdf.

[50] Renteln, Paul. *Lectures on Vector Calculus.* California State University, 2011, https://physics.csusb.edu/~prenteln/notes/vc_notes.pdf.

[51] Bower, Allan. *A Brief Introduction to Tensors and their properties.* Brown University, 2016, https://www.brown.edu/Departments/Engineering/Courses/En221/Notes/Tensors/Tensors.htm.