Introduction to Relativity

Patrick Bryant
Carnegie Mellon University
patrickbryant@me.com
October 2019

I. Introduction

In 1861 James Maxwell published an early form of what we know today as Maxwell's equations. These four equations describe the intertwined dynamics of two seemingly distinct natural phenomena, electricity and magnetism. Thirty years earlier, Michael Faraday had observed a key physical manifestation of this connection: He was able to induce an electric current in a loop of wire by either moving a magnet through it, or by changing the current through an adjacent loop of wire. Clearly the two forces were related, but it was not until Maxwell that an underlying reason for this interplay began to emerge.

His equations were motivated by the study of physical objects - magnets, wires, batteries, etc. - but once they were written down it became possible to study their solutions in the absence of such experimental complexities. Maxwell noticed that his equations allowed for the electric and magnetic forces to propagate together through the vacuum. These wave solutions all traveled at the same speed, a speed which could be calculated by measuring two quantities in the laboratory:

(1) The *permeability* of free space μ_0 is a measure of how well the vacuum supports a magnetic field. The force between two magnets in vacuum is proportional to μ_0.

(2) The *permittivity* of free space ϵ_0 is a measure of how poorly (instead of 'how well' for unfortunate historic reasons) the vacuum can support an electric field. The force between two electric charges in a vacuum is inversely proportional to ϵ_0.

To the shock of the scientific community of the time, Maxwell's wave solutions propagated at a speed which was consistent with the recently measured speed of light. The agreement was so good, it lead Maxwell to make the following extraordinary claim:

> *We can scarcely avoid the inference that light consists in the transverse undulations of the same medium which is the cause of electric and magnetic phenomena.*

Perhaps the most surprising aspect of this revelation was that Maxwell's wave solutions had no notion of a reference frame: Light it seems travels at exactly the same speed, regardless of how fast you are moving when you turn on your flashlight. Standing still it leaves your flashlight at $299,792,458\,\mathrm{m/s}$. Traveling down the highway at $75\,\mathrm{mph}$ it leaves your flashlight at $299,792,458\,\mathrm{m/s}$. Send a signal from the Parker Solar Probe orbiting the sun at $190\,\mathrm{km/s}$, and it flies back to earth at *exactly* $299,792,458\,\mathrm{m/s}$.

II. Velocity Addition

We will come to terms with this intuitively absurd fact about the behaviour of light by deriving how relative velocities can be consistently added. It will turn out that there is only one possible formula for velocity addition and it has one free parameter.

Suppose we have three 2D coordinate systems which differ only by their relative velocities v_{AB} and v_{BC}:

$$(a_x, a_t) \in A \underset{v_{AB}}{\longrightarrow} (b_x, b_t) \in B \underset{v_{BC}}{\longrightarrow} (c_x, c_t) \in C$$

Imagine being at rest in coordinate system A and coordinate system B is moving past you at velocity v_{AB}. You could catch up to B by running v_{AB}. Then you will be at rest in coordinate system B. We call this a **boost** between reference systems by an amount v_{AB}. To get from A to C you could boost by v_{AB} and then again by v_{BC}. Equivalently you could do a single boost by v_{AC}:

$$A \underset{v_{AB}}{\longrightarrow} B \underset{v_{BC}}{\longrightarrow} C \equiv A \underset{v_{AC}}{\longrightarrow} C$$

This relationship is what we mean by velocity addition. Given two velocities v_{AB} and v_{AC} we should be able to uniquely determine v_{AC}.

We can write the coordinates of one system in terms of another with four coefficients which could be functions of their relative velocity. For example:

$$b_x = \Lambda_x{}^x(v_{AB})a_x + \Lambda_x{}^t(v_{AB})a_t$$

$$b_t = \Lambda_t{}^x(v_{AB})a_x + \Lambda_t{}^t(v_{AB})a_t$$

We will use simple logical relationships to pin down the allowed form of the $\Lambda_i{}^j(v)$. First consider the case where $v_{AB} = 0$ and by definition we have $A \equiv B$.

$$b_x = a_x = \Lambda_x{}^x(0)a_x + \Lambda_x{}^t(0)a_t$$

$$b_t = a_t = \Lambda_t{}^x(0)a_x + \Lambda_t{}^t(0)a_t$$

This implies the following constraint on the functional form $\Lambda_i{}^j(v)$:

$$\Lambda_x{}^x(0) = 1, \quad \Lambda_x{}^t(0) = 0$$

$$\Lambda_t{}^x(0) = 0, \quad \Lambda_t{}^t(0) = 1$$

We can use this fact to constrain the allowed form of velocity addition by computing the velocity of C in the coordinates of A. First write c_x in terms of A:

$$\begin{aligned} c_x &= \Lambda_x{}^x(v_{BC})b_x + \Lambda_x{}^t(v_{BC})b_t \\ &= \Lambda_x{}^x(v_{BC})\left[\Lambda_x{}^x(v_{AB})a_x + \Lambda_x{}^t(v_{AB})a_t\right] + \Lambda_x{}^t(v_{BC})\left[\Lambda_t{}^x(v_{AB})a_x + \Lambda_t{}^t(v_{AB})a_t\right] \end{aligned}$$

Then compute the rate of change of c_x with respect to a_t:

$$\frac{dc_x}{da_t} = \Lambda_x{}^x(v_{BC})\Lambda_x{}^t(v_{AB}) + \Lambda_x{}^t(v_{BC})\Lambda_t{}^t(v_{AB})$$

Now we need to switch from units of length in C to units of length in A. This conversion factor is the rate of change of a_x with respect to c_x which is just the reciprocal of

$$\frac{dc_x}{da_x} = \Lambda_x{}^x(v_{BC})\Lambda_x{}^x(v_{AB}) + \Lambda_x{}^t(v_{BC})\Lambda_t{}^x(v_{AB})$$

so the velocity of C in the coordinates of A must have the following form:

$$\begin{aligned} v_{AC} &= \frac{da_x}{dc_x}\frac{dc_x}{da_t} \\ &= \frac{\Lambda_x{}^x(v_{BC})\Lambda_x{}^t(v_{AB}) + \Lambda_x{}^t(v_{BC})\Lambda_t{}^t(v_{AB})}{\Lambda_x{}^x(v_{BC})\Lambda_x{}^x(v_{AB}) + \Lambda_x{}^t(v_{BC})\Lambda_t{}^x(v_{AB})} \end{aligned} \tag{1}$$

Consider the case where $v_{BC} = 0$ so $C \equiv B$:

$$\begin{aligned} v_{AC} = v_{AB} &= \frac{\cancelto{1}{\Lambda_x{}^x(0)}\Lambda_x{}^t(v_{AB}) + \cancelto{0}{\Lambda_x{}^t(0)}\Lambda_t{}^t(v_{AB})}{\cancelto{1}{\Lambda_x{}^x(0)}\Lambda_x{}^x(v_{AB}) + \cancelto{0}{\Lambda_x{}^t(0)}\Lambda_t{}^x(v_{AB})} \\ \Rightarrow v_{AB} &= \frac{\Lambda_x{}^t(v_{AB})}{\Lambda_x{}^x(v_{AB})} \end{aligned}$$

We have shown that $\Lambda_x{}^t(v)$ cannot be independent of $\Lambda_x{}^x(v)$, in particular we have

$$\Lambda_x{}^t(v) = v\Lambda_x{}^x(v) \tag{2}$$

for any boost v.

Plugging equation 2 into equation 1 we have:

$$\begin{aligned} v_{AC} &= \frac{\Lambda_x{}^x(v_{BC})v_{AB}\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_x{}^x(v_{BC})\Lambda_t{}^t(v_{AB})}{\Lambda_x{}^x(v_{BC})\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_x{}^x(v_{BC})\Lambda_t{}^x(v_{AB})} \\ &= \frac{v_{AB}\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_t{}^t(v_{AB})}{\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_t{}^x(v_{AB})} \end{aligned} \tag{3}$$

In the case where $v_{AC} = 0$ and $C \equiv A$ this implies

$$v_{AB}\Lambda_x{}^x(v_{AB}) = -v_{BC}\Lambda_t{}^t(v_{AB})$$

and if $v_{AC} = 0$ then $v_{BC} \equiv v_{BA} = -v_{AB}$. This gives the following additional constraint:

$$\Lambda_x{}^x(v) = \Lambda_t{}^t(v) \tag{4}$$

Plugging equation 4 into 3 gives

$$\begin{aligned} v_{AC} &= \frac{v_{AB}\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_x{}^x(v_{AB})}{\Lambda_x{}^x(v_{AB}) + v_{BC}\Lambda_t{}^x(v_{AB})} \\ &= \frac{v_{AB} + v_{BC}}{1 + v_{BC}\Lambda_t{}^x(v_{AB})/\Lambda_x{}^x(v_{AB})} \end{aligned} \tag{5}$$

Equation 5 must be true if we switch $v_{AB} \leftrightarrow v_{BC}$ since we can boost from A to C by v_{AB} and then v_{BC} or by v_{BC} and then v_{AB}. In other words, the choice of intermediate frame B on the way from A to C does not matter.

Then

$$\begin{aligned} \frac{v_{AB} + v_{BC}}{1 + v_{BC}\Lambda_t{}^x(v_{AB})/\Lambda_x{}^x(v_{AB})} &= \frac{v_{BC} + v_{AB}}{1 + v_{AB}\Lambda_t{}^x(v_{BC})/\Lambda_x{}^x(v_{BC})} \\ \Rightarrow v_{BC}\frac{\Lambda_t{}^x(v_{AB})}{\Lambda_x{}^x(v_{AB})} &= v_{AB}\frac{\Lambda_t{}^x(v_{BC})}{\Lambda_x{}^x(v_{BC})} \end{aligned} \tag{6}$$

The left hand side of equation 6 is proportional to v_{BC} which implies that

$$\frac{\Lambda_t{}^x(v)}{\Lambda_x{}^x(v)} = kv \tag{7}$$

for some constant k. Plugging equation 7 into 5 gives us the only possible form for velocity addition:

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + k\, v_{AB} v_{BC}} \tag{8}$$

If we choose $k = 0$ we get the intuitive formula

$$v_{AC} = v_{AB} + v_{BC} \tag{9}$$

for velocity addition in **Galilean Relativity**. Unfortunately for your intuition, this cannot explain Maxwell's observation that light speed is the same in all reference frames.

Maxwell's discovery in 1861 of the relationship between electricity and magnetism predicted wave solutions which traveled at a velocity determined only by the *permeability* and *permittivity* of the vacuum. It had no dependence on the choice of rest frame of the wave. A, B and C would all observe these electromagnetic waves propagating at this velocity.

Let us return to the more general equation 11 to see if we can understand how Maxwell's prediction could possibly be consistent with our understanding of coordinate systems. Let c denote the velocity of Maxwell's wave solutions (not to be confused with coordinate system C). If we consider the case $v_{BC} = c$ then we need $v_{AC} = c$, ie. the speed of light should be the same in coordinate system A or B for an arbitrary choice of v_{AB}:

$$\begin{aligned} c &= \frac{v_{AB} + c}{1 + \mathrm{k} c v_{AB}} \\ \Rightarrow c + \mathrm{k} c^2 v_{AB} &= v_{AB} + c \\ \Rightarrow \mathrm{k} &= \frac{1}{c^2} \end{aligned} \tag{10}$$

Then velocity addition can be consistent with Maxwell's equations if and only if the following equation holds:

$$v_{AC} = \frac{v_{AB} + v_{BC}}{1 + v_{AB} v_{BC}/c^2} \tag{11}$$

This is a remarkable fact about space and time! The observation that the dynamics of electricity and magnetism allow for wave solutions which depend only on the *permeability* and *permittivity* of the vacuum *requires* a very non-intuitive formula for the composition of boosts.

Maxwell had discovered a physical law which served as a counter example to the assumption that the universe respected *Galilean Relativity*. The fact that light propagates at a velocity which is independent of inertial reference frame is proof that physical laws either respect this new **Special Relativity**, or no relativity at all.

This revelation was deeply disturbing to many physicists at the time. From our every day perspective the universe clearly respects *Galilean Relativity*. The reason our intuition turns out to be so wrong is that the speed of light is enormous relative to our everyday experience. This results in a separation of scales and it is worth studying the implications of *Special Relativity* in different regimes. Consider equation 11 in the regime most relevant to our daily lives - the limit $v_{AB}, c_{BC} \ll c$:

$$v_{AC} \approx (v_{AB} + v_{BC})(1 - v_{AC} v_{BC}/c^2 + ...)$$

For everyday velocities, we observe *Galilean Relativity* with a correction of order $v_{AB}v_{BC}/c^2$. Consider a baseball pitcher running at 10 mph ≈ 4 m/s and releasing a pitch at 90 mph ≈ 40 m/s:

$$-\frac{4m/s \times 40m/s}{(300,000,000m/s)^2} \approx -0.00005\%$$

In other words the pitcher thinks they threw a solid 90mph fastball with a 10mph boost from running, but *Special Relativity* gives the batter an advantage by making the ball cross the plate at a paltry 99.99995mph from their perspective.

III. The Geometry of Spacetime

In the previous section we used simple consistency conditions in the relationships between coordinate systems differing only by relative velocity to derive the only possible formula for velocity addition. Such coordinate systems are called **inertial frames**. The formula had one free parameter which we showed is fixed by the observational fact that the speed of light is independent of your choice of inertial frame. Along the way we also found several constraints on the equations relating one inertial frame to another. In particular we found equations 2, 4 and 7 where $k = 1/c^2$.

$$\begin{aligned}
\Lambda_x^{\ x}(v) &= \Lambda_x^{\ x}(v) \\
\Lambda_x^{\ t}(v) &= v\Lambda_x^{\ x}(v) \\
\frac{\Lambda_t^{\ x}(v)}{\Lambda_x^{\ x}(v)} &= \frac{v}{c^2}
\end{aligned}$$

We can use these to write all of the functions $\Lambda_\mu^{\ \nu}(v)$ in terms of v, c and $\Lambda_x^{\ x}(v) \equiv \gamma(v)$:

$$\begin{aligned}
\Lambda_x^{\ x}(v) &= \gamma(v) & \Lambda_x^{\ t}(v) &= v\gamma(v) \\
\Lambda_t^{\ x}(v) &= \frac{v}{c^2}\gamma(v) & \Lambda_t^{\ t}(v) &= \gamma(v)
\end{aligned}$$

All that remains to convert between inertial frames is to determine the function $\gamma(v)$. Since every component of the coordinate transformation depends linearly on $\gamma(v)$, it must be true that $\gamma(v) = \gamma(-v)$. This is also a reflection of the fact that the speed of light doesn't depend on the direction of the light.

We know that if we boost by v to get from A to B, and then boost back by $-v$ we should recover A:

$$\begin{aligned}
a_x &= \gamma(v)\left[b_x - vb_t\right] \\
&= \gamma(v)\left[\gamma(v)(a_x + va_t) - v\gamma(v)(\frac{v}{c^2}a_x + a_t)\right] \\
&= \gamma(v)^2\left[a_x + va_t - \frac{v^2}{c^2}a_x - va_t\right] \\
&= \gamma(v)^2\left[a_x - \frac{v^2}{c^2}a_x\right] \\
&= a_x\gamma(v)^2\left[1 - \frac{v^2}{c^2}\right] \\
\Rightarrow \gamma(v) &= \frac{1}{\sqrt{1 - v^2/c^2}}
\end{aligned}$$

To make our lives simpler, we can use units of space and time where $c = 1$. As an example, if we use units of length given by the distance light travels in one second or a *light-second*, then

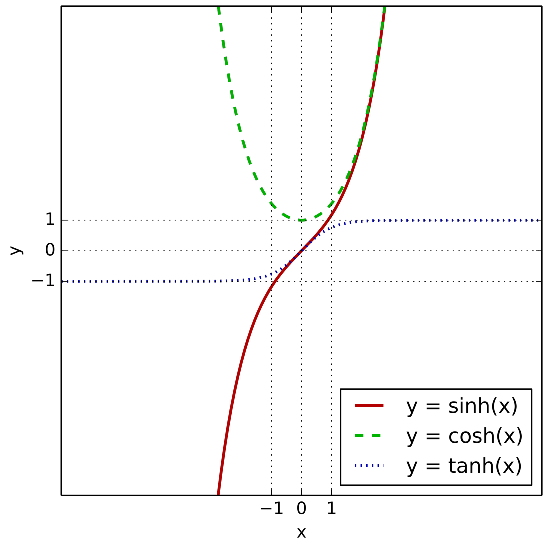

Figure 1: *By Fylwind at English Wikipedia - Own work, Public Domain*

$c = 1$ light-second per second. In these units γ is well defined for $-1 < v < 1$ (from now on we will drop the functional notation and γ will implicitly be a function of v). Then there exists some real number η such that

$$v = \tanh \eta$$

which you can see in figure 1. Conveniently the hyperbolic identity

$$\cosh^2 \eta = \frac{1}{1 - \tanh^2 \eta}$$

shows that $\gamma = \cosh \eta$. Then boosts between coordinate frames can be represented as

$$\begin{bmatrix} b_x \\ b_t \end{bmatrix} = \begin{bmatrix} \gamma & v\gamma \\ v\gamma & \gamma \end{bmatrix} \begin{bmatrix} a_x \\ a_t \end{bmatrix} \tag{12}$$

$$= \begin{bmatrix} \cosh \eta & \tanh \eta \cosh \eta \\ \tanh \eta \cosh \eta & \cosh \eta \end{bmatrix} \begin{bmatrix} a_x \\ a_t \end{bmatrix}$$

$$= \begin{bmatrix} \cosh \eta & \frac{\sinh \eta}{\cosh \eta} \cosh \eta \\ \frac{\sinh \eta}{\cosh \eta} \cosh \eta & \cosh \eta \end{bmatrix} \begin{bmatrix} a_x \\ a_t \end{bmatrix}$$

$$= \begin{bmatrix} \cosh \eta & \sinh \eta \\ \sinh \eta & \cosh \eta \end{bmatrix} \begin{bmatrix} a_x \\ a_t \end{bmatrix} \tag{13}$$

This might all look like complicated mathematical tricks, but what we are actually doing is illustrating the geometric structure of spacetime. Remember that rotations in the plane can be represented by an angle θ and the usual sine and cosine functions:

$$\begin{bmatrix} b_x \\ b_y \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} a_x \\ a_y \end{bmatrix} \tag{14}$$

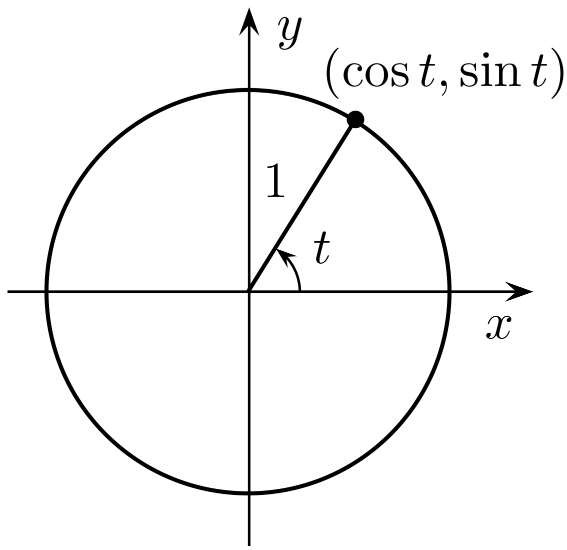

Figure 2: *By Gustavb - Own work, CC BY-SA 3.0*

This works because sine and cosine are *defined* by the geometry of the circle as illustrated in figure 2.

The hyperbolic sine and cosine are analogously defined by the *hyperbola* $x^2 - y^2 = 1$ as shown in figure 3:

In the same way that rotations mix together units of length, width and height, boosts mix together units of space and time! Rather than the usual geometry of rotations, boosts are hyperbolic rotations! The spacetime in which we live thus has a circular geometry relating the different spatial directions and a hyperbolic geometry relating space and time. This is called **Lorentz** geometry. Combined with the space and time translations, we call the geometric structure of spacetime the **Poincaré** group.

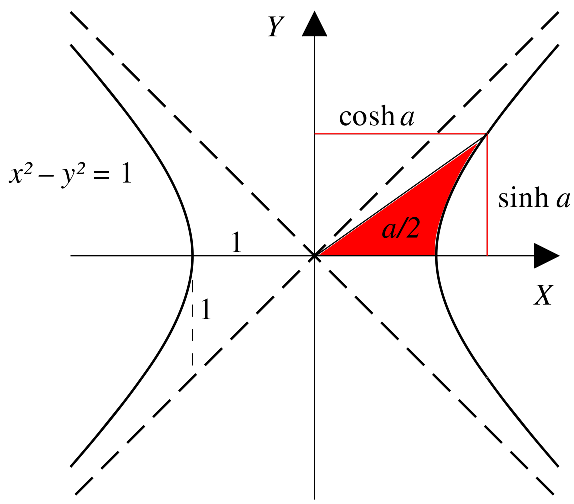

Figure 3: *The original uploader was Marco Polo at English Wikipedia: Derivative work by Jeandavid54 Public Domain*