Lecture I: A review of special relativity

(Dated: August 21, 2019)

I. OVERVIEW

This lecture introduces special relativity and the transformation laws. We will work from the point of view of symmetries and coordinate transformations (there are several ways to introduce the subject, but I will start from the constancy of the speed of light as a postulate since this is historically how special relativity was built).

This material corresponds to approximately Chapter 1 of Schutz.

II. SOME CONVENTIONS FOR RELATIVITY

[This roughly corresponds to §§1.3–1.4 of the book.]

Before we continue, we will set up a few conventions.

A. Units

In order to simplify the algebra, in this class we will work in the relativistic system of units where $c = 1$. Since c has units of velocity (299792458 m/s in SI), that means we can treat time and length as having equivalent units: e.g., 1 year = 1 light-year, or 1 second = 1 light-second. Velocity is dimensionless in this system.

If you want to write a relativistic formula in SI, then you will have to find the right places to put factors of c back into the equations, e.g., when we write $\gamma = (1 - v^2)^{-1/2}$ (where γ is dimensionless) then you should think $\gamma = (1 - v^2/c^2)^{-1/2}$.

B. Coordinates

As in the book, we will write spacetime coordinates in the order (t, x, y, z). A *reference frame* is a system of coordinates (t, x, y, z) that assigns a time (t) and spatial positions (x, y, z) to every event.

We will sometimes have to give the coordinates numbers (as appropriate when they are entries in the matrix). We will start the numbering of coordinates at 0 (as you might do if you program in C or Python):

$$x^0 = t, \quad x^1 = x, \quad x^2 = y, \quad \text{and} \quad x^3 = z. \tag{1}$$

Note that we write coordinates with a superscript ("up" index), which may seem unusual to you coming from undergraduate physics. The reason is that in relativity we are going to have a separate use for subscripts ("down" indices) later. The superscript on a coordinate is not an exponent; if we mean an exponent, we will write this with parentheses, e.g., $(\Delta x^1)^2$ for the square of the change in x^1.

We will sometimes have to write a generic index, x^α where α could be any of 0, 1, 2, or 3. We will use Greek letters to mean any of 0, 1, 2, or 3; Latin indices (e.g., x^i) will be used to mean any of the spatial components: $i = 1$, 2, or 3.

C. Events

A point in spacetime (defined by both "when" and "where": all 4 coordinates specified) is called an *event*. We denote events with script letters $\mathcal{P}$, $\mathcal{Q}$, etc. (we do not use $\mathcal{O}$, which will indicate a reference frame).

III. REMINDER OF REFERENCE FRAMES IN NEWTONIAN PHYSICS

We begin the content of our class with a review of reference frames in Newtonian physics. In Newtonian physics, once an origin is specified, the time t is universal and has the same numerical value in all reference frames. We will consider only inertial frames in this lecture (we will consider accelerated frames in special relativity later).

A. Change of velocity

In Newtonian physics, you learned that the laws of physics could be equally well described by observers in different reference frames. In particular, if an observer $\bar{\mathcal{O}}$ moves at constant velocity $\boldsymbol{V}$ with respect to an inertial observer $\mathcal{O}$, then in $\bar{\mathcal{O}}$'s frame, the time and spatial coordinates are related to $\mathcal{O}$'s coordinates by

$$\bar{t} = t \quad \text{and} \quad \bar{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{V}t. \tag{2}$$

The velocity of an object is defined by

$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}, \tag{3}$$

and transforms according to

$$\bar{\boldsymbol{v}} = \frac{\mathrm{d}\bar{\boldsymbol{x}}}{\mathrm{d}\bar{t}} = \frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{x} - \boldsymbol{V}t) = \boldsymbol{v} - \boldsymbol{V}. \tag{4}$$

The acceleration $\boldsymbol{a} = \mathrm{d}\boldsymbol{v}/\mathrm{d}t$ is invariant: $\bar{\boldsymbol{a}} = \boldsymbol{a}$, a fact which underlies most of your applications of Newton's 2nd law.

A transformation of reference frame of the form Eq. (2) is called a *boost*.

B. Rotations

Rotations are an ancient concept in geometry, pre-dating the concept of a coordinate system. I will discuss them at length here (even though less is said about them in the book) because the concepts will be useful for understanding other coordinate transformations. Both Newtonian physics and special relativity postulate that the laws of physics are the same in a rotated reference frame. We may describe such a rotation as:

$$\bar{t} = t \quad \text{and} \quad \bar{\boldsymbol{x}} = \mathbf{R}\boldsymbol{x}, \tag{5}$$

where $\mathbf{R}$ is a 3×3 *rotation matrix*. A proper rotation matrix is defined as a matrix with the properties that:

- It preserves dot products: $\bar{\boldsymbol{a}} \cdot \bar{\boldsymbol{b}} = \boldsymbol{a} \cdot \boldsymbol{b}$ for any two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^3$.
- It preserves orientation ($\det \mathbf{R} > 0$). [*Note* – if $\det \mathbf{R} < 0$, then the matrix flips left- and right-handed bases and is said to be improper. We don't need to consider improper rotations here.]

Let's pause here to think about what type of matrix $\mathbf{R}$ must be. Note that the first point means that

$$\bar{\boldsymbol{a}}^{\mathrm{T}}\bar{\boldsymbol{b}} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}, \tag{6}$$

or

$$\bar{\boldsymbol{a}}^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{R}\bar{\boldsymbol{b}} = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}. \tag{7}$$

for all $\boldsymbol{a}, \boldsymbol{b}$. This is true if and only if

$$\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbb{I}, \tag{8}$$

where $\mathbb{I}$ is the 3×3 identity. This looks like 9 equations for the 9 entries in $\mathbf{R}$, however since $\mathbf{R}^{\mathrm{T}}\mathbf{R}$ is symmetric regardless of the entries in $\mathbf{R}$ there are only 6 independent equations. Thus there are $9 - 6 = 3$ free parameters that we can specify to define a rotation. [In N dimensions, there would be $N(N-1)/2$ free parameters.] You can also see by taking the determinant of Eq. (8) that

$$1 = \det \mathbb{I} = \det \mathbf{R}^{\mathrm{T}} \, \det \mathbf{R} = (\det \mathbf{R})^2 \quad \rightarrow \quad \det \mathbf{R} = \pm 1. \tag{9}$$

Here the $+$ sign corresponds to a proper rotation, and the $-$ sign corresponds to an improper rotation.

A matrix $\mathbf{R}$ satisfying Eq. (8) is said to be *orthogonal*; if further $\det \mathbf{R} = 1$, it is *special orthogonal*.

If $\mathbf{R}_1$ and $\mathbf{R}_2$ are two proper rotation matrices, you can see that the composition $\mathbf{R}_1\mathbf{R}_2$ is a rotation matrix by noting:

$$(\mathbf{R}_1\mathbf{R}_2)^{\mathrm{T}}(\mathbf{R}_1\mathbf{R}_2) = \mathbf{R}_2^{\mathrm{T}}\mathbf{R}_1^{\mathrm{T}}\mathbf{R}_1\mathbf{R}_2 = \mathbf{R}_2^{\mathrm{T}}\mathbb{I}\mathbf{R}_2 = \mathbf{R}_2^{\mathrm{T}}\mathbf{R}_2 = \mathbb{I}. \tag{10}$$

Similarly, you can see that if $\mathbf{R}$ is a rotation matrix then so is $\mathbf{R}^{-1}$, and also that $\mathbb{I}$ is a rotation matrix. These properties, and the associativity of matrix multiplication, mean that the rotation matrices form a *group* in the mathematical sense: they have closure, inverses, an identity, and associativity. (We won't use or study group theory in this class.) Note that rotations do **not** commute.

You probably learned that a rotation by angle ϕ around the z-axis is described by

$$\begin{pmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \tag{11}$$

It can be shown that any proper rotation matrix $\mathbf{R}$ can be written as 3 such rotations; the angles of rotation are the 3 *Euler angles* ϕ, θ, and ψ:

$$\mathbf{R}(\phi,\theta,\psi) = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12}$$

C. Boost and rotation

A general change in reference frame can be written through a combination of a boost and a rotation:

$$\bar{t} = t \quad \text{and} \quad \bar{\boldsymbol{x}} = \mathbf{R}(\boldsymbol{x} - \boldsymbol{V}t). \tag{13}$$

This can be written in matrix form:

$$\begin{pmatrix} \bar{t} \\ \bar{x} \\ \bar{y} \\ \bar{z} \end{pmatrix} = \mathbf{\Lambda} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix}, \tag{14}$$

where the 4×4 matrix $\mathbf{\Lambda}$ can be written in blocks as:

$$\mathbf{\Lambda} = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \mathbf{R} \end{array}\right) \left(\begin{array}{c|c} 1 & 0 \\ \hline -\boldsymbol{V} & \mathbb{I} \end{array}\right) = \left(\begin{array}{c|c} 1 & 0 \\ \hline -\mathbf{R}\boldsymbol{V} & \mathbf{R} \end{array}\right) \tag{15}$$

(the lower-right block is 3×3). It takes 6 parameters (3 components of $\boldsymbol{V}$ and 3 Euler angles) to define the matrix $\mathbf{\Lambda}$: if you really want, you can write $\mathbf{\Lambda}(V_x, V_y, V_z, \phi, \theta, \psi)$. It can be shown that the matrices $\mathbf{\Lambda}$ form a group.

Despite the fact that we have written space and time as part of the same vector here, in Newtonian physics the two are not unified. Time is an absolute invariant: since $\bar{t} = t$, any observers $\mathcal{O}$ and $\bar{\mathcal{O}}$ always agree on the time. In contrast, the three spatial coordinates transform in a complicated way between reference frames. Moreover, since velocities transform as

$$\bar{\boldsymbol{v}} = \frac{\mathrm{d}\bar{\boldsymbol{x}}}{\mathrm{d}\bar{t}} = \frac{\mathrm{d}}{\mathrm{d}t}[\mathbf{R}(\boldsymbol{x} - \boldsymbol{V}t)] = \mathbf{R}(\boldsymbol{v} - \boldsymbol{V}), \tag{16}$$

one can see that light, traveling at speed $c = 1$ in one observer's frame, need not be traveling at c in another observer's frame. This directly contradicts your derivation of the speed of light in E&M, which tells you the speed of light in terms of fundamental constants; thus it is the same in every reference frame.

IV. THE LORENTZ TRANSFORMATIONS

We are now ready to construct an alternative set of transformations of spacetime that do preserve the speed of light. I will do this algebraically, as opposed to the book (which does this graphically). We will assume the transformations to be linear, which is appropriate when changing between rectilinear coordinate systems on a flat surface.

A. Derivation

We will leave the rotations alone – they work the same way in special relativity as in Newtonian physics or Euclidean geometry. Instead, we will think about the boosts. Let's imagine a transformation $\mathbf{\Lambda}$ corresponding to changing to the reference frame of an observer moving at velocity V in the x-direction. Symmetry considerations tell us that many entries in $\mathbf{\Lambda}$ must be 0:

$$\mathbf{\Lambda} = \left(\begin{array}{c|ccc} \mu & \nu & 0 & 0 \\ \hline \sigma & \gamma & 0 & 0 \\ 0 & 0 & \rho & 0 \\ 0 & 0 & 0 & \rho \end{array}\right) \qquad \leftrightarrow \qquad \begin{cases} \bar{t} = \mu t + \nu x \\ \bar{x} = \sigma t + \gamma x \\ \bar{y} = \rho y \\ \bar{z} = \rho z \end{cases}. \tag{17}$$

This uses the fact that, e.g., there is no difference between y and $-y$, so $\bar{t}$ cannot have a term proportional to y; and similarly $\bar{y}$ cannot have a term proportional to t. Similarly, the $\bar{y}y$ and $\bar{z}z$ entries must be the same since there is nothing to distinguish the y from the z-axis.

We first recall that observer $\bar{\mathcal{O}}$ is moving along the trajectory $x = Vt$, $y = 0$, $z = 0$. The barred position of $\bar{O}$ is:

$$\bar{x} = \sigma t + \gamma x = \sigma t + \gamma V t = (\sigma + \gamma V)t. \tag{18}$$

Since $\bar{\mathcal{O}}$ is at the spatial origin in their own rest frame, $\bar{x} = 0$, and so we must have

$$\sigma = -\gamma V. \tag{19}$$

Let's now test the constancy of the speed of light. We first imagine light going through some displacement $(\cos\theta, \sin\theta, 0)$ as seen by observer $\mathcal{O}$. The spatial distance traveled is 1, so $\mathcal{O}$ must see this process taking a time 1. The spacetime displacement of the light is $\Delta x^\alpha = (1, \cos\theta, \sin\theta, 0)$. In the barred frame, $\bar{\mathcal{O}}$ (who is moving with respect to $\mathcal{O}$) measures a displacement:

$$\Delta\bar{t} = \mu + \nu\cos\theta, \quad \Delta\bar{x} = \sigma + \gamma\cos\theta, \quad \Delta\bar{y} = \rho\sin\theta, \quad \Delta\bar{z} = 0. \tag{20}$$

We want to impose the condition that the speed of light is always 1 in the frame of $\bar{\mathcal{O}}$. To do this, let's first consider light going in the x-direction ($\theta = 0$). Here the speed of light in $\bar{\mathcal{O}}$'s frame is

$$1 = \frac{\Delta\bar{x}}{\Delta\bar{t}} = \frac{\sigma + \gamma}{\mu + \nu}. \tag{21}$$

Thus we see that we need

$$\mu + \nu = \sigma + \gamma = (1 - V)\gamma. \tag{22}$$

Similarly, for light going in the $-x$ direction ($\theta = \pi$), the speed of light in $\bar{\mathcal{O}}$'s frame is:

$$-1 = \frac{\Delta\bar{x}}{\Delta\bar{t}} = \frac{\sigma - \gamma}{\mu - \nu} \tag{23}$$

(I put in a -1 since the light moves to the left). Thus

$$\mu - \nu = \gamma - \sigma = (1 + V)\gamma. \tag{24}$$

This implies algebraically

$$\mu = \gamma \quad \text{and} \quad \nu = -V\gamma. \tag{25}$$

Finally, let's consider light moving in the y-direction in the $\mathcal{O}$ frame, thus $\theta = \frac{\pi}{2}$. The speed of light in the barred frame is

$$1 = \frac{\sqrt{\Delta\bar{x}^2 + \Delta\bar{y}^2}}{\Delta\bar{t}} = \frac{\sqrt{\sigma^2 + \rho^2}}{\mu} = \frac{\sqrt{V^2\gamma^2 + \rho^2}}{\gamma}, \tag{26}$$

so

$$\rho^2 = \gamma^2 - V^2\gamma^2 = (1 - V^2)\gamma^2 \quad \text{or} \quad \rho = \sqrt{1 - V^2}\,\gamma. \tag{27}$$

Thus all of our coefficients are determined except for an overall scaling γ.

You might be tempted to wonder if we can determine γ by considering the speed of light going at an arbitrary angle θ. But based on what we already know, the speed of light as seen by $\bar{\mathcal{O}}$ is:

$$\begin{aligned} 1 &= \frac{\sqrt{\Delta\bar{x}^2 + \Delta\bar{y}^2 + \Delta\bar{z}^2}}{\bar{t}} \\ &= \frac{\sqrt{(\sigma + \gamma\cos\theta)^2 + \rho^2\sin^2\theta}}{\mu + \nu\cos\theta} \\ &= \frac{\sqrt{(-V\gamma + \gamma\cos\theta)^2 + (1-V^2)\gamma^2\sin^2\theta}}{\gamma - V\gamma\cos\theta} \\ &= \frac{\sqrt{(-V + \cos\theta)^2 + (1-V^2)\sin^2\theta}}{1 - V\cos\theta} \\ &= \frac{\sqrt{V^2 - 2V\cos\theta + \cos^2\theta + \sin^2\theta - V^2\sin^2\theta}}{1 - V\cos\theta} \\ &= \frac{\sqrt{V^2\cos^2\theta - 2V\cos\theta + 1}}{1 - V\cos\theta} \\ &= \frac{1 - V\cos\theta}{1 - V\cos\theta} = 1. \end{aligned} \tag{28}$$

Therefore, the matrix we have constructed:

$$\mathbf{\Lambda} = \left(\begin{array}{c|ccc} \gamma & -V\gamma & 0 & 0 \\ \hline -V\gamma & \gamma & 0 & 0 \\ 0 & 0 & \sqrt{1-V^2}\,\gamma & 0 \\ 0 & 0 & 0 & \sqrt{1-V^2}\,\gamma \end{array}\right) \quad \leftrightarrow \quad \begin{cases} \bar{t} = \gamma(t - Vx) \\ \bar{x} = \gamma(-Vt + x) \\ \bar{y} = \sqrt{1-V^2}\,\gamma y \\ \bar{z} = \sqrt{1-V^2}\,\gamma z \end{cases}, \tag{29}$$

we have automatically satisfied the condition that the speed of light at any arbitrary angle is the same in all reference frames. This still does not determine γ.

So we need one more assumption in order to fix γ, and that is to presume the existence of such a thing as a fundamental standard of time (e.g., the frequency of the ^{133}Cs hyperfine transition, if you use SI). More precisely, two observers in the same frame should be able to agree on how many seconds have passed between two events in their laboratory. We set this up by doing a transformation from $\mathcal{O}$ to $\bar{\mathcal{O}}$, and then a back-transformation (boost by velocity $-V$) from $\bar{\mathcal{O}}$ to $\mathcal{O}$. The two transformations should have the same value of γ, since there is nothing to distinguish V from $-V$ (rotational invariance!). The second transformation is thus:

$$\mathbf{\Lambda}_{\rm back} = \left(\begin{array}{c|ccc} \gamma & V\gamma & 0 & 0 \\ \hline V\gamma & \gamma & 0 & 0 \\ 0 & 0 & \sqrt{1-V^2}\,\gamma & 0 \\ 0 & 0 & 0 & \sqrt{1-V^2}\,\gamma \end{array}\right), \tag{30}$$

and – by matrix multiplication –

$$\begin{aligned} \mathbf{\Lambda}_{\rm back}\mathbf{\Lambda} &= \left(\begin{array}{c|ccc} \gamma & V\gamma & 0 & 0 \\ \hline V\gamma & \gamma & 0 & 0 \\ 0 & 0 & \sqrt{1-V^2}\,\gamma & 0 \\ 0 & 0 & 0 & \sqrt{1-V^2}\,\gamma \end{array}\right) \left(\begin{array}{c|ccc} \gamma & -V\gamma & 0 & 0 \\ \hline -V\gamma & \gamma & 0 & 0 \\ 0 & 0 & \sqrt{1-V^2}\,\gamma & 0 \\ 0 & 0 & 0 & \sqrt{1-V^2}\,\gamma \end{array}\right) \\ &= \left(\begin{array}{c|ccc} (1-V^2)\gamma^2 & 0 & 0 & 0 \\ \hline 0 & (1-V^2)\gamma^2 & 0 & 0 \\ 0 & 0 & (1-V^2)\gamma^2 & 0 \\ 0 & 0 & 0 & (1-V^2)\gamma^2 \end{array}\right) = (1-V^2)\gamma^2\mathbb{I}. \end{aligned} \tag{31}$$

We can see that this "transformation back" only works and leaves $\mathcal{O}$ with a consistent measurement of time ($t = t$) if $(1-V^2)\gamma^2 = 1$. We thus see that

$$\gamma = \frac{1}{\sqrt{1-V^2}}, \tag{32}$$

and the Lorentz transformation is

$$\mathbf{\Lambda}_{\text{back}} = \left(\begin{array}{c|ccc} \gamma & V\gamma & 0 & 0 \\ \hline V\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \tag{33}$$

(we will often leave γ in the equations).

B. Simple implications

The most important – and surprising – result from the Lorentz transformation is that observers in different frames do not agree on the time of an event: we cannot have $\bar{t} = t$. This was the price to pay for forcing these observers to agree on the speed of light. Moreover, observers don't agree on the notion of simultaneity: $\bar{\mathcal{O}}$ sees two events as simultaneous if $\Delta\bar{t} = 0$ or if $\Delta t + V\Delta x = 0$. Thus two spatially separated events may be simultaneous for one observer but not another.

Another implication is that the speed of light is a universal speed limit: the above formulas only make sense for $|V| < 1$. We will see when we discuss accelerated observers how the "speed limit" is enforced.

V. MATHEMATICAL PROPERTIES OF THE TRANSFORMATIONS

We note several ways in which the Lorentz transformations are similar to rotation matrices. First, we see that $\det \mathbf{\Lambda} = 1$ – this means that the spacetime volume element $dt\,dx\,dy\,dz$ is conserved under change of reference frame.

A. Hyperbolic function description

You might wonder if the Lorentz transformations have a deeper relation to the rotations. In fact, they do. To see this, let's consider the Lorentz transformation with velocity

$$V = \tanh\alpha, \tag{34}$$

where $\alpha \in \mathbb{R}$ is the *boost parameter*. For small $\alpha \ll 1$, we have $V \approx \alpha$, but for very large boost parameter α, $V \to 1$ (we consider boosts to a frame very near the speed of light). Then using the hyperbolic identities:

$$\gamma = \frac{1}{\sqrt{1-V^2}} = \frac{1}{\sqrt{1-\tanh^2\alpha}} = \cosh\alpha. \tag{35}$$

Similarly,

$$V\gamma = \tanh\alpha\cosh\alpha = \sinh\alpha. \tag{36}$$

The Lorentz transformation matrix is thus:

$$\mathbf{\Lambda} = \left(\begin{array}{c|ccc} \cosh\alpha & -\sinh\alpha & 0 & 0 \\ \hline -\sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right). \tag{37}$$

From this point of view, the Lorentz transformation looks very much like a rotation. The difference is that it uses the hyperbolic rather than trigonometric functions. This is a minor difference from an analytic point of view – after all, every "rule" for trigonometric functions has a hyperbolic counterpart with a few sign flips (this is basically because $\cos\alpha = \cosh \mathrm{i}\alpha$ and $\mathrm{i}\sin\alpha = \sinh \mathrm{i}\alpha$). But from a global point of view it is hugely important, because trigonometric functions are periodic (rotation by 2π brings one back to the original frame), whereas if one keeps on boosting a reference frame there is no such periodicity.

B. Invariant products

In our study of rotations, we learned that there was an invariant dot product, $\boldsymbol{a}\cdot\boldsymbol{b}$, between two vectors, which in matrix notation was simply $\boldsymbol{a}^{\rm T}\boldsymbol{b}$. One might wonder if there is such a thing for the Lorentz transformations, but with 4-component vectors. It turns out that there is, but because of the sign flips associated with the hyperbolic functions we have to change the invariant. We write

$$\mathbf{M} = \left(\begin{array}{c|ccc} -1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right), \tag{38}$$

and then consider the quantity

$$\boldsymbol{a}^{\rm T}\mathbf{M}\boldsymbol{b} = -a^0b^0 + a^1b^1 + a^2b^2 + a^3b^3. \tag{39}$$

Now it is obvious that under rotations, $\boldsymbol{a}^{\rm T}\mathbf{M}\boldsymbol{b}$ will be invariant, since a^0 and b^0 are not affected by rotations, and $a^1b^1 + a^2b^2 + a^3b^3$ is the 3D dot product (which we already know to be invariant). Under the Lorentz boost, we have

$$\bar{\boldsymbol{a}}^{\rm T}\mathbf{M}\bar{\boldsymbol{b}} = \boldsymbol{a}^{\rm T}\boldsymbol{\Lambda}^{\rm T}\mathbf{M}\boldsymbol{\Lambda}\boldsymbol{b}. \tag{40}$$

By direct multiplication, we see that

$$\boldsymbol{\Lambda}^{\rm T}\mathbf{M}\boldsymbol{\Lambda} = \left(\begin{array}{c|ccc} \cosh\alpha & -\sinh\alpha & 0 & 0 \\ \hline -\sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c|ccc} -1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \left(\begin{array}{c|ccc} \cosh\alpha & -\sinh\alpha & 0 & 0 \\ \hline -\sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) = \left(\begin{array}{c|ccc} -1 & 0 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) = \mathbf{M}, \tag{41}$$

so

$$\bar{\boldsymbol{a}}^{\rm T}\mathbf{M}\bar{\boldsymbol{b}} = \boldsymbol{a}^{\rm T}\mathbf{M}\boldsymbol{b}. \tag{42}$$

Thus $\boldsymbol{a}^{\rm T}\mathbf{M}\boldsymbol{b}$ is an invariant. In fact, one normally defines the *Lorentz group* to be the group of 4×4 matrices $\boldsymbol{\Lambda}$ that satisfy $\boldsymbol{\Lambda}^{\rm T}\mathbf{M}\boldsymbol{\Lambda} = \mathbf{M}$. It really is analytically just like defining rotations to satisfy $\mathbf{R}^{\rm T}\mathbf{R} = \mathbb{I}$.

Note – This definition of the Lorentz group allows transformations that flip the direction of time (*non-orthochronous*) and/or switch left- and right-handed coordinate systems. Again, we won't usually consider such transformations, as there is no continuous way to realize them.

C. Invariant intervals

The most basic use of the dot product in Euclidean geometry is to define the square norm of a vector, $|\boldsymbol{a}|^2 = \boldsymbol{a}\cdot\boldsymbol{a}$. We can do the same thing with a 4D displacement vector $\Delta\boldsymbol{x} = (\Delta t, \Delta x, \Delta y, \Delta z)$ in spacetime. Its square norm is

$$\Delta s^2 = \Delta\boldsymbol{x}^{\rm T}\mathbf{M}\Delta\boldsymbol{x} = -\Delta t^2 + \Delta x^2 + \Delta y^2 + \Delta z^2. \tag{43}$$

This is known as the *invariant interval.* It looks like Euclidean geometry except for the $-$ sign. However, the $-$ sign is extremely important: in Euclidean geometry, all non-zero vectors have positive square norm. In the geometry of special relativity, Δs^2 may be positive, negative, or zero depending on whether the spatial displacement $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ is greater, equal to, or less than the time displacement $|\Delta t|$. Such displacements are called *spacelike*, *lightlike* (or *null*), or *timelike*, respectively. Because Δs^2 is an invariant, the classification as spacelike, lightlike, or timelike does not depend on the reference frame. Such a classification of directions of vectors has no analogue in Euclidean geometry.

For the timelike or null cases, one may also consider a displacement to be *future-directed* ($\Delta t > 0$) or *past-directed* ($\Delta t < 0$); again, as long as we restrict to orthochronous transformations, all observers will agree on this distinction (a timelike vector cannot be continuously transformed from future- to past-directed by Lorentz boosts, since they are topologically disconnected).

So far as we know, a physical object may only have a displacement that is timelike or null, and future-directed. (The mathematics of special relativity allow for other paths – after all, we are only talking about the path of an object through spacetime – but allowing spacelike trajectories causes problems with causality and the initial value formulation of physics since in someone's reference frame these trajectories go backward in time.)

VI. PROPER TIME

For the timelike displacements, with $\Delta s^2 < 0$, we may consider an observer $\bar{\mathcal{O}}$ who actually traverses that displacement, i.e., where the two points have zero spatial separation $\Delta\bar{x} = \Delta\bar{y} = \Delta\bar{z} = 0$. In this case, we must have $\Delta s^2 = -\Delta\bar{t}^2$. We thus define the *proper time interval* to be

$$\Delta\tau = \sqrt{-\Delta s^2}. \tag{44}$$

This is the time experienced on a path by an observer who actually traverses that path. Note that

$$\frac{\Delta\tau}{\Delta t} = \frac{\sqrt{-\Delta s^2}}{\Delta t} = \frac{\sqrt{\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2}}{\Delta t} = \sqrt{1 - \left(\frac{\Delta x}{\Delta t}\right)^2 - \left(\frac{\Delta y}{\Delta t}\right)^2 - \left(\frac{\Delta z}{\Delta t}\right)^2} = \sqrt{1 - v^2}, \tag{45}$$

where v is the velocity of the path. In particular, the proper time experienced on a spaceship (observer $\bar{\mathcal{O}}$) moving at constant velocity as it passes from event $\mathcal{P}$ to event $\mathcal{Q}$ is always less than the time span $\Delta t = t(\mathcal{Q}) - t(\mathcal{P})$ as measured in the lab frame (observer $\mathcal{O}$).

You might be worried at this point about the following paradox. In the aforementioned problem, I told you that the time measured on the spaceship (which is $\Delta\tau = \Delta\bar{t}$) is less than the time measured in the lab frame (Δt). That is really true: a fast-moving muon lives longer as measured in the lab frame than a muon at rest. It is often said that "time is slowed down" for a fast-moving observer. This seems weird, since in relativity we could always swap the two observers: by the same logic, the observer on the spaceship thinks time is slowed down for the observer in the lab. The resolution to this paradox is that we measured the times $\Delta\bar{t}$ and Δt between two specific events, $\mathcal{P}$ and $\mathcal{Q}$, which have definite spatial positions as well. (In the above example: $\mathcal{P}$ may be the creation of the muon, and $\mathcal{Q}$ may be its decay.) One observer (the observer on a spaceship moving along with the muon) sees these events happening at the same spatial point, and the other observer (in the lab) does not. Thus, even though relativity treats all inertial frames on an equal footing, the specific problem we are solving makes a distinction between the lab and the spaceship.

VII. TRANSFORMATION OF VELOCITIES

The final problem we will solve is the transformation of velocities. Suppose that a particle is moving in frame $\mathcal{O}$ in the x-direction at velocity w; its trajectory is $x = vw$ (and $y = z = 0$). In the frame of an observer $\bar{\mathcal{O}}$ moving at velocity V in the x-direction, we have

$$\bar{w} = \frac{\Delta\bar{x}}{\Delta\bar{t}} = \frac{\gamma(\Delta x - V\Delta t)}{\gamma(\Delta t - V\Delta x)} = \frac{\gamma(w\Delta t - V\Delta t)}{\gamma(\Delta t - Vw\Delta t)} = \frac{w - V}{1 - Vw}. \tag{46}$$

(I am doing the parallel case for simplicity here.) This is to be compared with the formula from Newtonian physics, $\bar{w} = w - V$. In particular, we can see that in the limit of $w \to 1$, we have $\bar{w} \to 1$ (both the numerator and denominator go to $1 - V$), and if $w \to -1$ then $\bar{w} \to -1$.

A cool way to understand the transformation of Eq. (46) is to recall the difference-angle formula for tangents,

$$\tan(A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}. \tag{47}$$

There is an equivalent formula for hyperbolic tangents with one sign change,

$$\begin{aligned}
\tanh(A - B) &= \frac{\mathrm{e}^{A-B} - \mathrm{e}^{-(A-B)}}{\mathrm{e}^{A-B} + \mathrm{e}^{-(A-B)}} \\
&= \frac{2(\mathrm{e}^{A-B} - \mathrm{e}^{B-A})}{2(\mathrm{e}^{A-B} + \mathrm{e}^{B-A})} \\
&= \frac{(\mathrm{e}^{A+B} + \mathrm{e}^{A-B} - \mathrm{e}^{B-A} - \mathrm{e}^{-A-B}) - (\mathrm{e}^{A+B} + \mathrm{e}^{B-A} - \mathrm{e}^{A-B} - \mathrm{e}^{-A-B})}{(\mathrm{e}^{A+B} + \mathrm{e}^{A-B} + \mathrm{e}^{B-A} + \mathrm{e}^{-A-B}) - (\mathrm{e}^{A+B} - \mathrm{e}^{A-B} - \mathrm{e}^{B-A} + \mathrm{e}^{-A-B})} \\
&= \frac{(\mathrm{e}^{A} - \mathrm{e}^{-A})(\mathrm{e}^{B} + \mathrm{e}^{-B}) - (\mathrm{e}^{B} - \mathrm{e}^{-B})(\mathrm{e}^{A} + \mathrm{e}^{-A})}{(\mathrm{e}^{A} + \mathrm{e}^{-A})(\mathrm{e}^{B} + \mathrm{e}^{-B}) - (\mathrm{e}^{A} - \mathrm{e}^{-A})(\mathrm{e}^{B} - \mathrm{e}^{-B})} \\
&= \frac{\frac{\mathrm{e}^{A} - \mathrm{e}^{-A}}{\mathrm{e}^{A} + \mathrm{e}^{-A}} - \frac{\mathrm{e}^{B} - \mathrm{e}^{-B}}{\mathrm{e}^{B} + \mathrm{e}^{-B}}}{1 - \frac{\mathrm{e}^{A} - \mathrm{e}^{-A}}{\mathrm{e}^{A} + \mathrm{e}^{-A}}\frac{\mathrm{e}^{B} - \mathrm{e}^{-B}}{\mathrm{e}^{B} + \mathrm{e}^{-B}}} \\
&= \frac{\tanh A - \tanh B}{1 - \tanh A \tanh B}.
\end{aligned} \tag{48}$$

Thus in Eq. (46), the velocities subtract in the same way as hyperbolic tangents, and we can write

$$w = \tanh\psi, \quad \bar{w} = \tanh\bar{\psi}, \quad V = \tanh\alpha, \quad \text{and} \quad \bar{\psi} = \psi - \alpha. \tag{49}$$

The "angles" subtract (or add, if you prefer to write $\psi = \alpha + \bar{\psi}$). This should be familiar if you think about rotations in Euclidean geometry: the angle a curve makes relative to the coordinate axes changes in this way when you rotate the coordinate system. So the boost parameter α really does have some similarities to a rotation angle.