The theory of special relativity

In physics we have a law which states that:

"Nothing moves faster than $c = 299792458$ m/s" (1

This is perhaps the most fundamental law in all of physics. All modern theories are based on it and the remarkable agreement obtained thus far between theory and experiment strongly suggest that this law is indeed a very solid one.

It turns out that this value $c \sim 3 \times 10^8$ m/s is also the value with which light moves in free space, which is a consequence that the photons have zero mass. However, this law has nothing to do with electromagnetism. It is a law on mechanics.

I cannot tell you why law (1) is true. It just is. Nature chose it this way. But we have to agree that it sort of makes sense. After all, when you restrict the speed of propagation of all particles, you also restrict the speed of propagation of interactions. If god wiggles the sun a little bit, the other stars in the solar system cannot feel the wiggling right away, that would be weird. Instead, this wiggling propagates through space with a finite velocity.

It doesn't take much to realize that law (1) is in complete disagreement with classical mechanics. For instance, if you apply a constant force F to a body, the velocity will evolve in time according to

$$v = v_0 + \frac{F}{m} t \tag{2}$$

wait long enough and v will become larger than c.

However, I want you to realize that the discrepancy is actually much deeper. Law (1) does not only disagree with our notions of dynamics and forces. It is inconsistent with our basic rudiments of <u>kinematics</u>.

All of classical mechanics is based on the Galileo transformation. Namely that if two inertial frames move relative to each other with a speed V, then the positions of a particle in the two frames, r and r', will be related by

$$r' = r + V t \tag{3}$$

Consequently,

$$v' = v + V \tag{4}$$

choose v and V conveniently and we can easily make $v' >$

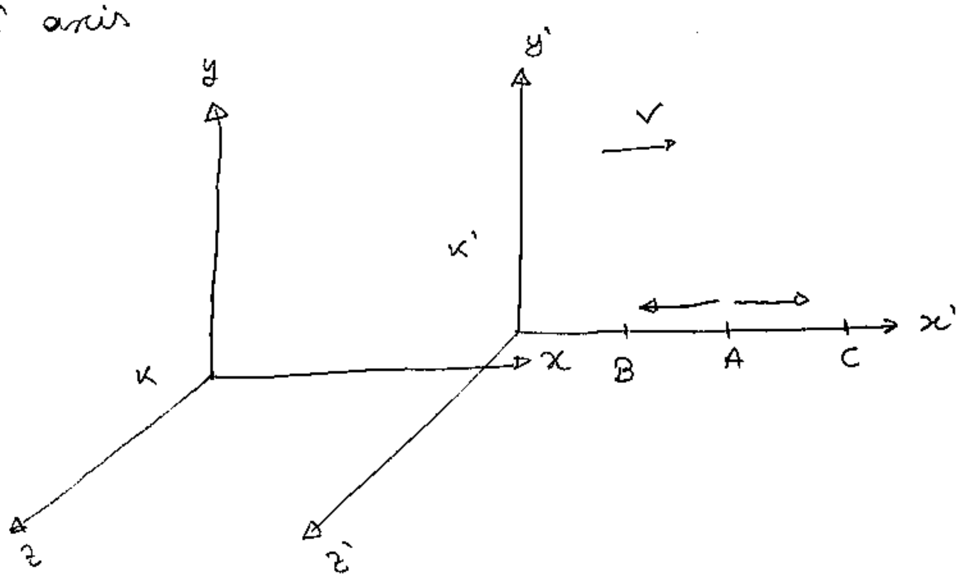
We therefore conclude that our very own notions of kinematics must be fundamentally wrong if we are to accept (1) as a fundamental physical law. In other words, Eq (3) must be modified. The final result will be a new transformation law called Lorentz transformation, which reduces to (3) when $V << c$.

Now suppose a particle is moving with $v = c$ in one frame. Then since it cannot move with a speed higher than c, its velocity in <u>any other</u> frame will also have to be c. This may sound weird at first, but it makes sense when you think about light: light propogates with speed c in the frame of reference attached at the earth. But it also propogates with c in any other inertial frame.

So we do not know how to modify Eq (3) just yet. But we know that, as a special case, if $v = c$ in one frame, then we must have $v' = c$ in any other frame. I can also state this as follows:

$$\boxed{\text{" The speed of light c is the same in any inertial frame"}} \quad (5)$$

This fact has deep consequences, which we may appreciate by considering the following example. Consider two frames, K and K', moving relative to each other with a speed V along the xx' axis



Suppose two light signals are emitted from point A directed toward the equally spaced points B and C. From the perspective of an observer in K' the signals will reach B and C simultaneously.

For an observer in K the signals will also move with speed c, but the points B and C are moving to the right. So for an observer in K the signal will reach B before it reaches C. This is a direct consequence of (5) (which in turn is a consequence of (1)).

Please stop and think about this for a moment. What this example is saying is that the law (1) affects our very notion of *simultaneity*: two events which are simultaneous in one frame, need not be simultaneous in another.

The law (1) therefore implies that *time cannot be absolute*. If we are to accept law (1), then we must completely abandon this very basic notion that time is absolute. We must rethink the very structure of *space* and *time*. So you see, the problem is not in the laws of dynamics. It is much deeper. To come up with a theory which satisfies (1), we must completely rethink our very notions of space and time.

In fact, we will learn that both must be treated together, as different parts of a more general structure called *spacetime*.

Intervals and the Lorentz transformation

We will now consider the motion of particles in different reference frames. For this we will frequently use the idea of an event. An event is specified by the position in space where it happened and the instant of time where it happened.

It is convenient to bundle this together into a four-dimensional vector with time as one of the coordinates. Something like

$$\text{event} = (t, x, y, z)$$

But this is a little bit awkward because t has different units from x, y and z. A good fix is to use as the first coordinate the quantity ct, which has units of length. We use c only for convenience since it is a universal constant. Thus, we will write events as

$$\text{event} = (ct, x, y, z) \tag{6}$$

When we consider the motion of point particles, an event is also called a world point. The motion of the particle thus forms a world line.

The 4D vector (6) is also called a four-vector and this fictitious 4D space is known as Minkowski space.

Now consider a reference frame K and suppose a light signal is emitted from (ct_1, x_1, y_1, z_2) and collected at (ct_2, x_2, y_2, z_2). Since we are talking about light here, the signal will move with speed c and therefore

$$c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0 \tag{7}$$

If we consider the same thing from another inertial frame K', the result will be the same because of (5): if a signal moves with speed c in K, it also does so in K'. Thus

$$c^2(t_2' - t_1')^2 - (x_2' - x_1')^2 - (y_2' - y_1')^2 - (z_2' - z_1')^2 = 0 \tag{8}$$

this particular combination of time and space in Eqs (7) and (8) plays a very important role in the theory of relativity. It is called an <u>interval</u>:

$$\boxed{s_{12} = \sqrt{c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2}} \tag{9}$$

what equations (7) and (8) say is that, for a particle moving with speed c,

$$s_{12} = s_{12}' = 0 \tag{10}$$

In other words, if an interval is zero in one inertial frame, it is also zero in any other.

If a signal or a particle move with some speed $v < c$, we continue to define its interval as in (9). But now, because $v < c$ we will have $s_{12} > 0$.

In the ficticious case where a particle moves faster than c, the interval would become imaginary.

For two events infinitesimally close to each other, we use for the corresponding interval the notation

$$\boxed{ds^2 = c^2dt^2 - dx^2 - dy^2 - dz^2} \qquad (11)$$

we can think of the interval as a "distance" in our 4D space. Note how this distance is different then the Euclidian distance due to the minus signs in the spatial components. this is the main difference between Euclidian space and Minkowski space.

we have already seen that

$$\boxed{\text{If } \quad s = 0 \text{ in one frame then } \quad s' = 0 \text{ in any other frame}} \qquad (12)$$

which is the mathematical statement of (5).

we now show a very important result, from which all else follows. Namely that

$$\boxed{s = s'} \qquad (13)$$

for any two inertial frames, irrespective of the velocity of the particles.

This means that the particular combination of time and space in Eq. (11) is an <u>invariant</u>; ie something which is the same in any inertial frame. In a world where so much is relative, invariants are treated like Kings.

To show (13) we consider two frames K_a and K_b moving with a relative velocity v_{ab}. Let ds_a^2 and ds_b^2 be the intervals of same event in the two frames. These are differentials of the same order and are such that, when one is zero, so is the other. Thus, they must be related linearly as

$$ds_a^2 = f(v_{ab})\, ds_b^2 \tag{14}$$

where f is some function which can depend only on the magnitude of v_{ab}: it cannot depend on the coordinates and times in the two frames because then different points in space and different times would not be equivalent, which would violate the <u>homogeneity</u> of <u>space</u> and <u>time</u>. Also, it cannot depend on the direction of v_{ab} because that would violate the <u>isotropy</u> of <u>space</u>.

Now suppose we have 3 frames moving relative to each other with velocities v_{ab}, v_{ac} and v_{bc}. Then we will have, besides (14),

$$ds_a^2 = f(v_{ac})\, ds_c^2 \tag{15}$$
$$ds_b^2 = f(v_{bc})\, ds_c^2$$

Comparing (14) and (15) we find:

$$ds_a^2 = f(V_{ab})\, ds_b^2 = f(V_{ab})\, f(V_{bc})\, ds_c^2$$

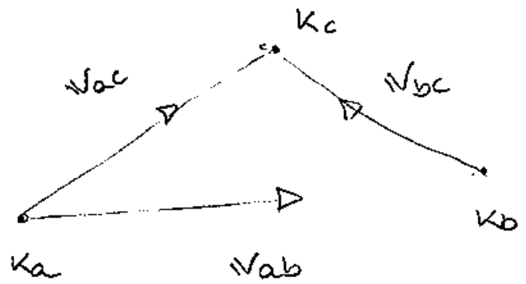
or

$$ds_a^2 = \frac{f(V_{ab})\, f(V_{bc})}{f(V_{ac})}\, ds_a^2$$

Thus, the function f must be such that

$$f(V_{ac}) = f(V_{ab})\, f(V_{bc}) \qquad (16)$$

This is merely a consistency requirement. Now let us look at this from the point of view of vector addition



we see that

$$\mathbf{V}_{ac} = \mathbf{V}_{ab} + \mathbf{V}_{bc}$$

Thus

$$V_{ac} = \sqrt{V_{ab}^2 + V_{bc}^2 + 2\mathbf{V}_{ab} \cdot \mathbf{V}_{bc}}$$

So from basic vector addition we see that V_{ac} must also depend on the angle between $\mathbf{V}_{ab}$ and $\mathbf{V}_{bc}$. But this angle does not appear on the right-hand side of (16).

The only possibility is therefore that f must be a constant, independent of V. Eq (16) then also determines that $f^2 = f$ so we must have $f = 1$.

Hence we conclude, based on (14), that for any two inertial frames,

$$\boxed{ds = ds'} \tag{17}$$

Integrating the infinitesimal intervals finally leads to the invariance of finite intervals, Eq (13). Even though it is not obvious, this is all a consequence of law (1).

We now introduce the idea of proper time, as the time measured in the reference frame where the particle is at rest. From (17) we have

$$c^2dt^2 - dx^2 - dy^2 - dz^2 = c^2dt'^2 - dx'^2 - dy'^2 - dz'^2 \tag{18}$$

If the frame K' is moving together with the particle, then $dx' = dy' = dz' = 0$. We therefore conclude that

$$c^2dt'^2 = c^2dt^2 - dx^2 - dy^2 - dz^2$$

But

$$\frac{dx^2 + dy^2 + dz^2}{dt^2} = V^2$$

is the velocity of the particle (and of K'). Thus we may write

$$\boxed{dt' = \frac{ds}{c} = dt\sqrt{1 - V^2/c^2}} \tag{19}$$

This relates the proper time with the time in other reference frames. For finite times we obtain

$$\Delta t' = t_2' - t_1' = \int_{t_1}^{t_2} dt \sqrt{1 - v^2/c^2} \qquad (20)$$

For uniform motion, v is independent of time and this simplifies further to

$$\Delta t' = \Delta t \sqrt{1 - v^2/c^2} \qquad (21)$$

Since $v < c$ we have

$$\sqrt{1 - v^2/c^2} < 1 \qquad (22)$$

this means that

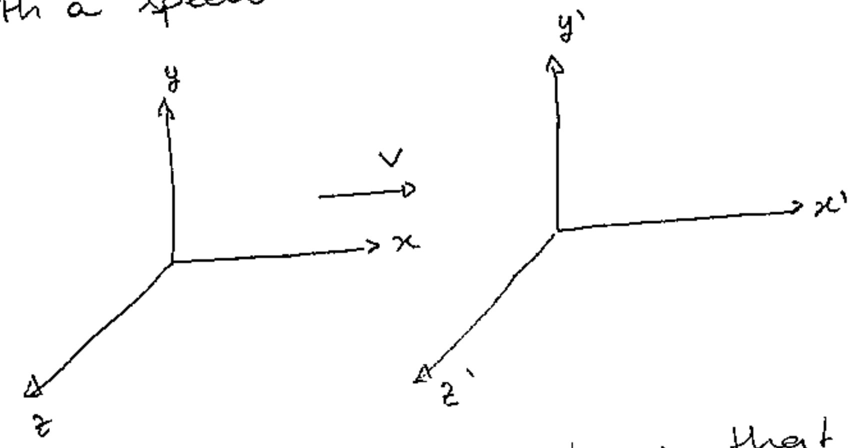
$$\Delta t' < \Delta t \qquad (23)$$

This is called <u>time dilation</u>. The classical experimental verification of this effect is based on the detection of muons produced in the atmosphere due to cosmic rays. If it were not for time dilation, we would never detect them at the surface of the earth because they would have all decayed. The decay time is measured in the proper time frame. Thus, the decay time for us here at earth will be larger by a factor $1/\sqrt{1-v^2/c^2}$. For these muons $v/c \simeq 0.994$ so

$$\frac{1}{\sqrt{1-v^2/c^2}} \approx 9.14$$

For an observer at the surface of the earth, the decay rate of muons is therefore around 9 times larger.

Let us now extend (19) and obtain the full transformation law between two inertial frames. Suppose K' moves relative to K with a speed v in the xx' direction



and suppose we adjust the clocks so that the origins coincide at $t = t' = 0$.

By the isotropy of space, the coordinates y and z should not change:

$$y = y' \qquad (24)$$
$$z = z'$$

Thus, the invariance of the interval implies that

$$c^2 t^2 - x^2 = c^2 t'^2 - x'^2 \qquad (25)$$

We have seen a similar thing when we discussed rotatio in 3D or 2D. We saw that rotations where those operations which left $x^2 + y^2 = x'^2 + y'^2$ and we saw that the transformation involved sines and cosines. Here the difference is in the minus sign. We can take care of it in a way similar to rotations by using hyperbolic functions.

The most general transformation satisfying (25) is

$$\begin{aligned} x &= x' \cosh\phi + ct' \sinh\phi \\ ct &= x' \sinh\phi + ct' \cosh\phi \end{aligned} \tag{26}$$

we can check this explicitly:

$$(ct)^2 - x^2 = x'^2 \underbrace{(\sinh^2\phi - \cosh^2\phi)}_{-1} + (ct')^2 \underbrace{(\cosh^2\phi - \sinh^2\phi)}_{1}$$

$$= (ct')^2 - x'^2.$$

All we need to do now is determine ϕ. To do this we note that if we consider the motion of the origin of K' then $x'=0$ and we obtain

$$x = ct' \sinh\phi$$
$$ct = ct' \cosh\phi$$

From which it follows that

$$\frac{x}{ct} = \tanh\phi$$

But we are looking at the origin of K' so $x/t = v$ is the velocity of K' with respect to K. Thus

$$\boxed{\tanh\phi = \frac{v}{c}} \tag{27}$$

To obtain $\cosh\phi$ and $\sinh\phi$ we use hyperbolic trigonometry:

$$\cosh^2\phi - \sinh^2\phi = 1$$

So (28)

$$\sinh\phi = \frac{V/c}{\sqrt{1-V^2/c^2}} \qquad \cosh\phi = \frac{1}{\sqrt{1-V^2/c^2}}$$

(29)

Let

$$\gamma = \frac{1}{\sqrt{1-V^2/c^2}} > 1$$

then Eq (26) becomes

$$x = \gamma\left(x' + \frac{V}{c}ct'\right)$$

$$ct = \gamma\left(x'\frac{V}{c} + ct'\right)$$

or (30)

$$\boxed{\begin{aligned} x &= \gamma(x' + Vt') \\ t &= \gamma\left(t' + \frac{Vx'}{c^2}\right) \end{aligned}}$$

These are called the <u>Lorentz transformations</u>. They represent the transformation rule between inertial frames such that law (1) is satisfied. Indeed, we see that if $x' = ct'$ (meaning the particle propagates with speed c in K') then

$$x = \gamma(c+V)t'$$

$$t = \gamma\left(1 + \frac{V}{c}\right)t'$$

so (31)

$$x' = ct' \implies x = ct$$

If a particle moves with c in K', it also moves with c in K.

when $V \ll c$ we get $\gamma \approx 1$ so (30) reduces to

$$x \simeq x' + Vt'$$
$$t \approx t' \qquad (32)$$

which are the Galileo transformations. Thus, indeed, at low speeds we recover the usual results of classical mechanics.

We can also obtain the inverse transformation easily using physical arguments: K' moves with speed V relative to K, so K moves with speed $-V$ relative to K'. Changing V to $-V$ therefore gives (t', x') as a function of t, x:

$$x' = \gamma(x - Vt)$$
$$t' = \gamma\left(t - \frac{Vx}{c^2}\right) \qquad (33)$$

Let Δx and $\Delta x'$ be the lengths of a rod as measured in the two frames. Considering (30) at the same instant t' we get

$$\Delta x = \gamma \Delta x' = \frac{\Delta x'}{\sqrt{1 - V^2/c^2}} \qquad (34)$$

The quantity $\Delta x'$ is the <u>proper</u> <u>length</u> of the rod; i.e. the length as measured from a reference frame at rest. We therefore see that Δx as measured from K is always smaller than $\Delta x'$. Bodies in motion look contracted.

From the Lorentz transformation (30) we may also obtain the law of transformation of velocities

$$v_x = \frac{dx}{dt} \qquad v_x' = \frac{dx'}{dt'} \tag{35}$$

The infinitesimal version of (30) is

$$\begin{aligned} dx &= \gamma(dx' + V dt') \\ dt &= \gamma\left(dt' + \frac{V}{c^2} dx'\right) \end{aligned} \tag{36}$$

so

$$v_x = \frac{dx}{dt} = \frac{dx' + V dt'}{dt' + \frac{V}{c^2} dx'} = \frac{v_x' + V}{1 + \frac{V v_x}{c^2}}$$

Thus

$$\boxed{v_x = \frac{v_x' + V}{1 + \frac{V v_x}{c^2}}} \tag{37}$$

For the other directions we have $y = y'$ so

$$v_y = \frac{dy}{dt} = \frac{dy'}{\gamma\left(dt' + \frac{V dx'}{c^2}\right)}$$

Hence

$$\boxed{v_y = \frac{v_y' \sqrt{1 - V^2/c^2}}{1 + \frac{V v_x}{c^2}}} \tag{38}$$

and similarly for v_z.

Four vectors

We have learned that to accept the laws of relativity we must work with space and time together, as forming a single entity called spacetime. A point in spacetime is described by a four vector (ct, x, y, z). We will from now on denote the components of this four vector by x^μ, $\mu = 0, 1, 2, 3$; ie

$$x^0 = ct \qquad x^1 = x \qquad x^2 = y \qquad x^3 = z \tag{39}$$

We will also use the notation

$$x^\mu = (ct, \mathbf{r}) \tag{40}$$

In which the first component is the time part and the second is the space part.

What we have learned is that given a four vector x^μ. The interval

$$s^2 = (ct)^2 - x^2 - y^2 - z^2 = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \tag{41}$$

is an invariant; ie. it is the same in any non-inertial reference frame. What is more striking about this result is that the time and space components do not enter in the same way.

To treat this we introduce another four vector

$$x_\mu = (ct, -\mathbf{r}) \tag{42}$$

Its time part is the same as x^μ and the space part gains a minus sign. A common way of writing this more compactly is as follows:

$$x_0 = x^0 \qquad x_i = -x^i, \quad i = 1, 2, 3 \tag{43}$$

This is a standard convention which we will always adopt: greek indices such as $\mu, \nu, \ldots$ always range from 0 to 3, whereas roman indices i, j, etc range from 1 to 3.

The vector x^μ is called contravariant, whereas x_μ is called covariant. A mnemonic to remember this is that the word "contravariant" is larger than "covariant", so its index goes on top.

The Lorentz transformation (33) is for a contravariant four vector:

$$
\begin{aligned}
x'^0 &= \gamma\left(x^0 - \frac{V}{c^2} x^1\right) \\
x'^1 &= \gamma\left(x^1 - V x^0\right) \\
x'^2 &= x^2 \\
x'^3 &= x^3
\end{aligned}
\qquad (44)
$$

If you want the covariant version you simply change signs in the spatial part

$$
\begin{aligned}
x'_0 &= \gamma\left(x_0 - \frac{V}{c^2}(-x_1)\right) \\
-x'_1 &= \gamma\left(-x_1 - V x_0\right) \\
-x'_2 &= -x_2 \\
-x'_3 &= -x_3
\end{aligned}
$$

or, organizing

$$
\begin{aligned}
x'_0 &= \gamma\left(x_0 + \frac{V}{c^2} x_1\right) \\
x'_1 &= \gamma\left(x_1 + V x_0\right) \\
x'_2 &= x_2 \\
x'_3 &= x_3
\end{aligned}
\qquad (45)
$$

With the two vectors x^μ and x_μ we may now construct the interval more succintly as

$$s^2 = \sum_{\mu=0}^{3} x_\mu x^\mu = x_0 x^0 + x_1 x^1 + x_2 x^2 + x_3 x^3$$
$$= (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \tag{46}$$

But there is one final twist to the story: it is called the <u>Einstein summation rule</u>: repeated indices should always be summed. That is, we will usually write (46) as

$$s^2 = x_\mu x^\mu \tag{47}$$

The summation is implicit because μ appears repeated. (by the way, $x_\mu x^\mu = x^\mu x_\mu$; the order is irrelevant).

The big conclusion about the invariance of the interval under Lorentz transformations may now be written as

$$x'_\mu x'^\mu = x_\nu x^\nu \tag{48}$$

Here I am being extra careful and writting different indices on the two sides, just to emphasize that both sides should be summed.

The operation $x_\mu x^\mu$, where you sum over μ, with one up and the other down, is called a <u>contraction</u>.

when we discussed rotations we established a rule to call something a "vector". We said "a vector is a collection of 3 numbers which correctly transforms under rotations".

Here we have the same thing for four vectors. A set of four quantities A^ν will have the honor of being called a four-vector if it correctly transforms under a Lorentz transformation as does x^ν. That is, A^ν will be a four vector if when x^ν transforms as in (44) then

$$
\begin{aligned}
A'^0 &= \gamma\left(A^0 - \frac{v}{c^2}A^1\right) \\
A'^1 &= \gamma\left(A^1 - vA^0\right) \\
A'^2 &= A^2 \\
A'^3 &= A^3
\end{aligned}
\tag{49}
$$

Given any two four-vectors A^ν and B^ν, we define their scalar product as

$$A^\nu B_\nu = A^0 B_0 + A^1 B_1 + A^2 B_2 + A^3 B_3 \tag{50}$$

Note that

$$A^\nu B_\nu = A_\nu B^\nu \tag{51}$$

because if we lower one index and raise the other, the minus signs cancel.

The quantity $A^\nu B_\nu$ is called a scalar. It is invariant under Lorentz transformations. This happens because we chose A^ν and B^ν to be legitimate four-vectors, which therefore transform similarly under Lorentz transformations.

Objects with more than one index are called tensors. The rank of the tensor is the number of indices it has. For instance

$$A^{\mu\nu} \quad A^{\mu}{}_{\nu} \quad A_{\mu}{}^{\nu} \quad A_{\mu\nu}$$

are all tensors of rank 2. The indices can be moved up or down using the same rule we had for vectors (the spatial part changes sign). Thus, we have for instance

$$A^{00} = A^{0}{}_{0} = A_{00} = A_{0}{}^{0}$$

but

$$A^{01} = -A^{0}{}_{1} = -A_{01} = A_{0}{}^{1}$$

and

$$A^{12} = -A^{1}{}_{2} = A_{12} = -A_{1}{}^{2}$$

etc. (always lower or raise one index at a time). Also, please be careful with the order. Don't write A^{μ}_{ν}, write $A^{\mu}{}_{\nu}$ or $A_{\nu}{}^{\mu}$. These are different things.

Just like we had with vectors, not any object $A^{\mu\nu}$ may be called a tensors. To be called a tensor, each index must transform in a Lorentz transformation as a four-vector. I will return to this point in a moment.

The most important tensor is called the <u>metric</u>. It has components

$$\eta^{00} = 1 \qquad \eta^{11} = \eta^{22} = \eta^{33} = -1 \tag{52}$$

In matrix notation

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \tag{53}$$

or we sometimes also write

$$\eta^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1) \tag{54}$$

to emphasize that η is diagonal

If you lower one index of the metric we get

$$\eta^{0}{}_{0} = 1 \qquad \eta^{1}{}_{1} = \eta^{2}{}_{2} = \eta^{3}{}_{3} = 1$$

So

$$\eta^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} = \begin{cases} 1 & \text{if } \mu = \nu \\ 0 & \text{if } \mu \neq \nu \end{cases} \tag{55}$$

written like $\eta^{\mu}{}_{\nu}$ or $\eta_{\mu}{}^{\nu}$ the metric becomes the <u>Kronecker delta</u>.

If we lower both indices of $\eta^{\mu\nu}$ nothing changes

$$\eta^{\mu\nu} = \eta_{\mu\nu} \tag{56}$$

The metric is usually used to lower or raise the indices of vectors and tensors. The typical procedure looks like this:

$$\boxed{x^\mu = \eta^{\mu\nu} x_\nu} \qquad (57)$$

Please stop and look at this formula for a second. It summarizes the notation that we just introduced. In a formula like this we will always follow two rules:

1) Repeated indices, one up, one down, are always summed (contracted). This is ν in (57)

2) Non-contracted indices, like μ in (57), should be on the same floor on both sides of the equation.

Always follow these rules. And use them as a consistency check. It will avoid you many errors.

For instance Eq (57) is actually four equations bundled together.

$$\mu = 0: \quad x^0 = \eta^{0\nu} x_\nu = \eta^{00} x_0 + \eta^{01} x_1 + \eta^{02} x_2 + \eta^{03} x_3$$

$$= (+1) x_0$$

$$\mu = 1: \quad x^1 = \eta^{1\nu} x_\nu = \eta^{10} x_0 + \eta^{11} x_1 + \eta^{12} x_2 + \eta^{13} x_3$$

$$= (-1) x_1$$

etc.

To give an example of a consistency check, we could also have written (57) as

$$x_\mu = \eta_\mu{}^\nu x_\nu$$

The indices are all in the right position. It is a consistent equation. Something like $x^\mu = \eta^\mu{}_\nu x_\nu$ is crap: ν is being summed and it is not in the up-down configuration. But

$$x^\mu = \eta^\mu{}_\nu x^\nu$$

is ok. In fact, since $\eta^\mu{}_\nu$ is the Kronecker delta, we can clearly see that this transformation is correct.

We may use the metric to write the scalar product as

$$A^\mu B_\mu = \eta_{\mu\nu} A^\mu B^\nu$$

Note how everything is consistent. Two indices up, two down.

The same works with tensors:

$$A^\mu{}_\nu = \eta_{\nu\alpha} A^{\mu\alpha}$$

or

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

To practice, try to check if the following formulas are internally consistent (it isn't)

$$A^{\alpha\beta\gamma} = x^\alpha \eta_{\mu\nu} C^{\mu\beta} \Gamma^\nu{}_\omega \Psi_\omega{}^\gamma$$

Lorentz transformations in general

Before, we demonstrated a very general result. Due to the finiteness and universality of the speed of ligth, the interval of a four vector is an invariant; ie $x_\mu x^\mu$ is the same in any inertial reference frame

We then went on to demonstrate that for the case where the two frames are moving with respect to each other with a velocity V along the x axis, then

$$
\begin{aligned}
x'^0 &= x^0 \cosh\phi - x^1 \sinh\phi \\
x'^1 &= -x^0 \sinh\phi + x^1 \cosh\phi \\
x'^2 &= x^2 \\
x'^3 &= x^3
\end{aligned}
\tag{58}
$$

where $\tanh\phi = V/c$ [I will from now on use the inverse of (30) and express x'^μ in terms of x^μ]. This type of transformation is called a boost.

These Lorentz transformations are linear and therefore may be written as

$$\boxed{x'^\mu = \Lambda^\mu{}_\nu \, x^\nu} \tag{59}$$

where

$$
\begin{aligned}
\Lambda^0{}_0 &= \Lambda^1{}_1 = \cosh\phi \\
\Lambda^0{}_1 &= \Lambda^1{}_0 = -\sinh\phi \\
\Lambda^2{}_2 &= \Lambda^3{}_3 = 1
\end{aligned}
$$

Or, in matrix notation

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \cosh\phi & -\sinh\phi & 0 & 0 \\ -\sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \tag{60}$$

Sometimes people write $\Lambda = (\cdots)$, but you need to be a bit careful in clarifying that Λ is the matrix with elements $\Lambda^{\mu}{}_{\nu}$ not, for example, $\Lambda^{\mu\nu}$.

Let us now consider the most general linear transformation and ask what is the condition for it to leave the interval invariant. The most general transformation may be written as

$$\boxed{x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu}} \tag{61}$$

for some transformation matrix $\Lambda^{\mu}{}_{\nu}$ and some four vector a^{μ}.

The vector a^{μ} describes <u>spacetime</u> <u>translations</u>, whereas $\Lambda^{\mu}{}_{\nu}$ may encompass the usual <u>rotations</u> <u>in</u> $\underline{\mathbb{R}}^3$ and <u>boosts</u> (like (60)), which are hyperbolic rotations.

The a^μ part will not leave $x_\mu x^\mu$ invariant, but that is ok. What is really invariant is the interval, which involves the distance between two points. In this case a^μ cancels out.

So for now let us set $a^\mu = 0$. We then have

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \tag{62}$$

and we want to determine the conditions on Λ such that $x^\mu x_\mu = x'^\alpha x'_\alpha$. Let us first write this as

$$x^\mu x_\mu = \eta_{\mu\nu} x^\mu x^\nu$$
$$x'^\alpha x'_\alpha = \eta_{\alpha\beta} x'^\alpha x'^\beta$$

Now we use (62):

$$x'^\alpha x'_\alpha = \eta_{\alpha\beta} x'^\alpha x'^\beta = \eta_{\alpha\beta} (\Lambda^\alpha{}_\mu x^\mu)(\Lambda^\beta{}_\nu x^\nu)$$
$$= (\eta_{\alpha\beta} \Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu) x^\mu x^\nu$$

We want this to be equal to $x^\mu x_\mu = \eta_{\mu\nu} x^\mu x^\nu$. Thus we get

$$\boxed{\eta_{\alpha\beta} \Lambda^\alpha{}_\mu \Lambda^\beta{}_\nu = \eta_{\mu\nu}} \tag{63}$$

This is the condition a matrix Λ must satisfy if it is to leave an interval invariant.

If we denote by η the matrix $\eta^{\mu\nu}$ and by Λ the matrix $\Lambda^\mu{}_\nu$ then we may also write (63) as

$$\boxed{\Lambda^T \eta \Lambda = \eta} \tag{64}$$

Any Lorentz transformation may be decomposed into at most 6 transformations, which correspond to rotations around

$$tx,\ ty,\ tz,\ xy,\ xz,\ yz$$

The $tx, \dots,$ rotations are hyperbolic (boosts) and the others are regular rotations.

For completeness I will write down the general formula for rotations and boosts. A general rotation of an angle θ around a unit vector $\hat{n}$ in $\mathbb{R}^3$ has the transformation matrix

$$\Lambda^0{}_0 = \Lambda^0{}_i = \Lambda^i{}_0 = 0 \tag{65}$$

$$\Lambda^i{}_j = \cos\theta \, \delta_{ij} + (1-\cos\theta)\, n^i n^j - \sin\theta \, \epsilon_{ijk} n^k$$

[Note that, for aesthetics, I am breaking our rules here!].

The most general boost with velocity $v = c\tanh\phi$ around a direction $\hat{n} \in \mathbb{R}^3$ has the formula

$$\Lambda^0{}_0 = \cosh\phi \qquad \Lambda^0{}_i = \Lambda^i{}_0 = -n^i \sinh\phi \tag{66}$$

$$\Lambda^i{}_j = \delta_{ij} + (\cosh\phi - 1)\, n^i n^j$$

when $\hat{n} = (0,0,1)$ we recover (60)

Four - velocity

As seen in (37) and (38), the velocity of a particle transform under Lorentz transformations in a strange non-linear way. That is because

$$v = \frac{d\mathbf{r}}{dt}$$

and both $\mathbf{r}$ and t transform under Lorentz transformations

In order to construct an object which resembles a velocity but which correctly transforms under Lorentz transformations, we consider the construct

$$\boxed{u^\mu = \frac{dx^\mu}{ds}} \qquad (67)$$

this is certainly a four-vector because ds is a scalar and dx^μ is a four-vector. We call this quantity the "four-velocity". It is not really a velocity since it is infact dimensionless. But we will now see that it does look like one.

From (19)

$$ds = c\,dt\sqrt{1 - v^2/c^2} \qquad (68)$$

Thus

$$u^\mu = \left(\frac{1}{\sqrt{1 - v^2/c^2}} , \frac{v/c}{\sqrt{1 - v^2/c^2}} \right) \qquad (69)$$

The four-velocity satisfies

$$u_\nu u^\nu = \gamma^2 (1 - v^2/c^2) = 1 \qquad (70)$$

which therefore shows that it is indeed a four-vector (because $u_\nu u^\nu$ is an invariant)

The four velocity may be interpreted as the four-vector tangent to the world line of the particle.

Relativistic mechanics

What we have done so far is kinematics. Now we will determine the laws describing the motion of a relativistic particle. It is a remarkable feat that we can do this using only symmetry arguments.

The action and the Lagrangian of a particle are related by

$$S = \int_{t_a}^{t_b} \mathcal{L}\, dt \qquad (71)$$

the action governs the physics of a system and the physics must be invariant under Lorentz transformations since the laws of physics should be the same in any inertial frame. Hence, S must be Lorentz invariant and so does $\mathcal{L}dt$.

Let us use (19),

$$ds = c\,dt\sqrt{1-v^2/c^2}$$

to write

$$S = \int_{t_1}^{t_2} \frac{\mathcal{L}}{c\sqrt{1-v^2/c^2}}\, ds$$

Since ds (an interval) is Lorentz invariant, we conclude that the same must be true for the combination

$$F = \frac{\mathcal{L}}{c\sqrt{1-v^2/c^2}}$$

This means that F may depend only on scalars, such as

$$x_\nu x^\nu, \quad x_\nu u^\nu, \quad u_\nu u^\nu.$$

Any other scalar would lead to higher order derivatives and therefore to weird terms like $\dddot{x}$ in the equations of motion. Thus, it suffices to consider these 3. Please note that there really are <u>no</u> <u>other</u> invariants.

Since we are talking of a free particle, however, F cannot depend explicitly on x^ν, because this would violate the homogeneity of spacetime. It would imply that one corner of spacetime is more important than another. Thus, F cannot depend on $x_\nu x^\nu$ or $x_\nu u^\nu$. Moreover, $u_\nu u^\nu = 1$ so the only possibility is for F to be a constant.

We therefore conclude, based only on Lorentz invariance and the homogeneity of spacetime, that the Lagrangian of a free particle must be

$$\mathcal{L} = cF\sqrt{1 - v^2/c^2}$$

where F is a constant.

To determine the value of F we look at the particular case $v \ll c$. We then get

$$\mathcal{L} \simeq cF\left(1 - \frac{v^2}{2c^2}\right) = cF - \frac{F v^2}{2c}$$

The first term is just a constant and the second should match the classical value $\frac{1}{2}mv^2$. Thus we find that

$$F = -mc$$

We then obtain for the Lagrangian of a free particle

$$\boxed{\mathcal{L} = -mc^2\sqrt{1 - v^2/c^2}} \qquad (72)$$

and for the action

$$S = -mc\int_a^b ds = -mc^2\int_{t_a}^{t_b}\sqrt{1-v^2/c^2}\,dt \qquad (73)$$

Note now that we are not allowed to add to $\mathcal{L}$ an arbitrary constant. This would violate Lorentz invariance.

Now that we have $\mathcal{L}$ we do what we know to do best: we compute stuff! the generalized momentum is

$$\boxed{p^i = \frac{\partial \mathcal{L}}{\partial v^i} = \frac{mv^i}{\sqrt{1-v^2/c^2}}} \qquad (74)$$

and the energy is

$$E = \sum_i v^i p^i - \mathcal{L}$$

or

$$\boxed{E = \frac{mc^2}{\sqrt{1-v^2/c^2}}} \qquad (75)$$

If the particle is at rest we find

$$E = mc^2 \tag{76}$$

Recalling the four-velocity u^μ in (69), note that

$$p^i = mc\,u^i \tag{77a}$$

and

$$E = mc^2 u^0 \tag{77b}$$

This hints us at defining a new four-vector

$$\boxed{p^\nu = (E/c, \mathbf{p})} \tag{78}$$

such that

$$p^\nu = mc\,u^\nu \tag{79}$$

Since u^ν is a four-vector, so will p^ν. We therefore see that energy and momentum combine exactly like time and space, which makes sense given what we know about energy and momentum.

In particular note that since $u_\nu u^\nu = 1$, we have

$$p_\nu p^\nu = \frac{E^2}{c^2} - p^2 = m^2c^2 \tag{79'}$$

or

$$\boxed{E^2 = m^2c^4 + p^2c^2} \tag{80}$$

This relates the energy with the momentum. So we may now write the Hamiltonian of a relativistic particle

$$\boxed{H = \sqrt{m^2c^4 + p^2c^2}} \tag{81}$$

The Dirac Equation

A big difficulty in the beggining of quantum mechanics was to concile Schrödinger's quantization ideas with the theorie of special relativity.

In the non-relativistic case the Hamiltonian of a free particle is

$$H = \frac{p^2}{2m} \tag{82}$$

Schrödinger's recipe may be stated as

$$\begin{aligned} H &\rightarrow i\hbar \frac{\partial}{\partial t} \\ p &\rightarrow -i\hbar \frac{\partial}{\partial x} \end{aligned} \tag{83}$$

(where $\frac{\partial}{\partial x} \equiv \nabla$). Multiplying (82) on both sides by ψ and using this recipe we obtain Schrödinger's equation for a free particle

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \psi \tag{84}$$

This Eq has one great property: $|\psi|^2$ is a conserved quantity and may therefore be interpreted as the probability of finding a particle somewhere. Even though this is not obvious, this is a consequence of the fact that Eq (84) only involves the first derivative in time. In this sense, Schrödinger was very smart to use complex numbers: usually wave equations are second order in t. The only way to make a wave equation of first order is to use complex numbers.

To construct a relativistic wave equation we may try to use Schrödinger's recipe (83) in (81). Well, (81) is no good, because of the square root. So maybe we could use it in

$$H^2 = m^2c^4 + p^2c^2$$

we then obtain

$$-\hbar^2 \frac{\partial^2\psi}{\partial t^2} = m^2c^4\psi - \hbar^2c^2\nabla^2\psi \qquad (85)$$

This is known as the <u>Klein-Gordon equation</u>. In fact, Schrödinger derived this Eq before (84). But in (85) $|\psi|^2$ is not conserved. More than that actually, (85) has no conserved quantities which are always positive. So it does not serve to describe particles such as the electron since the number of electrons would not be conserved.

Now we know there is nothing wrong with (85). But it describes spin 0 particles and not electrons or protons.

Dirac tried to find a way out of this problem. So he had the following idea. Consider (79'):

$$p_\nu p^\nu - m^2c^2 = 0 \tag{86}$$

Is it possible to write this as

$$p_\nu p^\nu - m^2c^2 = (\gamma_\alpha p^\alpha - mc)(\gamma^\beta p_\beta + mc) \tag{87}$$

If this is true then we could apply Schrödinger's recipe (83) to the equations

$$\gamma_\alpha p^\alpha - mc = 0 \tag{88}$$

or

$$\gamma^\beta p_\beta + mc = 0$$

$$(\gamma_\alpha p^\alpha - mc)(\gamma^\beta p_\beta + mc) = \gamma_\alpha \gamma^\beta p^\alpha p_\beta - m^2c^2 + mc(\gamma_\alpha p^\alpha - \gamma^\beta p_\beta)$$

The linear terms are eliminated because

$$\gamma_\alpha p^\alpha = \gamma^\alpha p_\alpha$$

We must then have

$$\gamma_\alpha \gamma^\beta p^\alpha p_\beta = p^\nu p_\nu .$$

or, written slightly differently

$$p^\nu p_\nu = \gamma^\alpha \gamma^\beta p_\alpha p_\beta \tag{89}$$

So now we look for the values of γ^μ which satisfy this equation. Let us write it explicitly

$$(p^0)^2 - (p^1)^2 - (p^2)^2 - (p^3)^2 = (\gamma^0)^2(p^0)^2 + (\gamma^1)^2(p^1)^2 + (\gamma^2)^2(p^2)^2 + (\gamma^3)(p^3)^2$$

$$+ (\gamma^0\gamma^1 + \gamma^1\gamma^0)\, p_0 p_1 + (\gamma^0\gamma^2 + \gamma^2\gamma^0)\, p_0 p_2 + \dots$$

where ... means other similar combinations. If you look at this long enough you will see that it is impossible to find γ^μ to satisfy this Eq. Choosing $\gamma^0 = 1$ and $\gamma^i = i$ would satisfy the first part but does not eliminate the cross terms.

Now came Dirac's brilliant idea what if the γ^μ are matrices instead of numbers? then these matrices would have to satisfy

$$(\gamma^0)^2 = 1 \qquad (\gamma^i)^2 = -1$$

$$\gamma^\mu \gamma^\nu + \gamma^\nu \gamma^\mu = 0 \qquad \mu \neq \nu.$$

These relations can be written compactly as

$$\boxed{\{\gamma^\mu, \gamma^\nu\} = 2^{\mu\nu}} \qquad (90)$$

where $\{A, B\} = AB + BA$ is called the anti-commutator.

Dirac then looked for a set of 4 matrices which satisfied these relations. He found that these matrices had to be 4×4. So the quantization rule (88) would involve an object

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \end{pmatrix}$$

with four components! Two components would be ok because it was already known that the electron had spin 1/2 and therefore required two components. The fact that there were four components was surprising. Now we know it is related to the existence of antiparticles.