Special Relativity

April 19, 2014

1 Galilean transformations

1.1 The invariance of Newton's second law

Newtonian second law,

$$\mathbf{F} = m\mathbf{a}$$

is a 3-vector equation and is therefore valid if we make any *rotation* of our frame of reference. Thus, if $O^i_{\ j}$ is a rotation matrix and we rotate the force and the acceleration vectors,

$$\begin{aligned} \tilde{F}^i &= O^i_{\ j} O^j \\ \tilde{a}^i &= O^i_{\ j} a^j \end{aligned}$$

then we have

$$\tilde{\mathbf{F}} = m\tilde{\mathbf{a}}$$

and Newton's second law is *invariant* under rotations. There are other invariances. Any change of the coordinates $\mathbf{x} \rightarrow \tilde{\mathbf{x}}$ that leaves the acceleration unchanged is also an invariance of Newton's law. Equating and integrating twice,

$$\begin{aligned} \tilde{\mathbf{a}} &= \mathbf{a} \\ \frac{d^2\tilde{\mathbf{x}}}{dt^2} &= \frac{d^2\mathbf{x}}{dt^2} \end{aligned}$$

gives

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{x}_0 - \mathbf{v}_0 t$$

The addition of a constant, $\mathbf{x}_0$, is called a *translation* and the change of velocity of the frame of reference is called a *boost*. Finally, integrating the equivalence $d\tilde{t} = dt$ shows that we may reset the zero of time (a *time translation*),

$$\tilde{t} = t + t_0$$

The complete set of transformations is

$$\begin{array}{cc} x'_i = \sum_j O_{ij} x_j & Rotations \\ \mathbf{x}' = \mathbf{x} + \mathbf{a} & Translations \\ t' = t + t_0 & Origin\ of\ time \\ \mathbf{x}' = \mathbf{x} + \mathbf{v}t & Boosts\ (change\ of\ velocity) \end{array}$$

There are three independent parameters describing rotations (for example, specify a direction in space by giving two angles (θ, φ) then specify a third angle, ψ, of rotation around that direction). Translations can be in the x, y or z directions, giving three more parameters. Three components for the velocity vector and one more to specify the origin of time gives a total of 10 parameters. These 10 transformations comprise the *Galilean group*. Newton's second law is invariant under the Galilean transformations.

Notice that all of the Galilean transformations are *linear*. This is crucial, because the position vectors $\mathbf{x}$ form a vector space, and only linear transformations preserve the linear combinations we require of vectors.

1.2 Failure of the Galilean group for electrodynamics

The same is not true of electrodynamics. For example, in the absence of sources, we have seen that Maxwell's equations lead to the wave equation,

$$-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}+\nabla^2\psi=0$$

for the each component of the fields and the potentials. But if we perform a boost of the coordinates, this equation is *not* invariant:

$$\begin{aligned}
0 &= -\frac{1}{c^2}\frac{\partial^2}{\partial\tilde{t}^2}\psi\left(\tilde{\mathbf{x}},\tilde{t}\right)+\tilde{\nabla}^2\psi\left(\tilde{\mathbf{x}},\tilde{t}\right) \\
&= -\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi\left(\tilde{\mathbf{x}},t\right)+\tilde{\nabla}^2\psi\left(\tilde{\mathbf{x}},t\right) \\
&= -\frac{1}{c^2}\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}\psi\left(\tilde{\mathbf{x}},t\right)-\frac{d\tilde{\mathbf{x}}}{dt}\cdot\tilde{\nabla}\psi\left(\tilde{\mathbf{x}},t\right)\right)+\nabla^2\psi\left(\mathbf{x},t\right) \\
&= -\frac{1}{c^2}\frac{\partial}{\partial t}\left(\frac{\partial}{\partial t}\psi\left(\tilde{\mathbf{x}},t\right)-\mathbf{v}_0\cdot\tilde{\nabla}\psi\left(\tilde{\mathbf{x}},t\right)\right)+\nabla^2\psi\left(\mathbf{x},t\right) \\
&= -\frac{1}{c^2}\left(\frac{\partial^2}{\partial t^2}\psi\left(\tilde{\mathbf{x}},t\right)-2\left(\mathbf{v}_0\cdot\tilde{\nabla}\right)\frac{\partial}{\partial t}\psi\left(\tilde{\mathbf{x}},t\right)+\left(\mathbf{v}_0\cdot\tilde{\nabla}\right)^2\psi\left(\tilde{\mathbf{x}},t\right)\right)+\nabla^2\psi\left(\mathbf{x},t\right) \\
&= \left[-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\psi\left(\tilde{\mathbf{x}},t\right)+\nabla^2\psi\left(\mathbf{x},t\right)\right]+\frac{2}{c^2}\left(\mathbf{v}_0\cdot\tilde{\nabla}\right)\frac{\partial}{\partial t}\psi\left(\tilde{\mathbf{x}},t\right)-\left(\frac{\mathbf{v}_0}{c}\cdot\tilde{\nabla}\right)^2\psi\left(\tilde{\mathbf{x}},t\right)
\end{aligned}$$

This means that there is an inherent conflict between the symmetry of Maxwell's equations and the symmetry of Newton's second law. They do not change in a consistent way if we change to a moving frame of reference. We must make a choice between modifying Maxwell's equations or modifying Newton's law. This is not as drastic as it sounds, since both of the troublesome terms above are of order $\left(\frac{v_0}{c}\right)^2 \ll 1$, but one must still be modified.

Since we know that Maxwell's equations actually *predict* the speed of light, it is not unreasonable to suppose that they are valid at large velocities. On the other hand, in 1900, the laws of Newtonian experiments had been tested only for $v \ll c$. We therefore begin by considering what set of boost transformations *does* leave the wave equation invariant.

1.3 Some definitions

Rewrite the wave equation,

$$-\frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2}+\nabla^2\psi=0$$

by introducing some systematic notation. Let

$$x^\alpha=(ct,x,y,z)\ \ for\,\alpha=0,1,2,3$$

and define the 4×4 object,

$$\eta^{\alpha\beta}\equiv\begin{pmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

We write derivatives with respect to these four coordinates as

$$\frac{\partial}{\partial x^\alpha}$$

so using the Einstein summation convention: any repeated index, one up and one down, is automatically summed,

$$x^\alpha \frac{\partial}{\partial x^\alpha} \equiv \sum_{\alpha=0}^{3} x^\alpha \frac{\partial}{\partial x^\alpha}$$

This will save us from writing large numbers of $\sum s$. With these definitions, the wave equation is simply

$$\eta^{\alpha\beta} \frac{\partial^2 \psi}{\partial x^\alpha \partial x^\beta} = 0$$

The double sum involves 16 terms, but since all but four components of $\eta^{\alpha\beta}$ are zero, only the four we need survive. Further details of this notation is given below.

1.4 Invariance of the d'Alembertian wave equation

As with the Galilean transformation, we require our transformations to be linear,

$$\tilde{x}^\alpha = \Lambda^\alpha_{\ \beta} x^\beta$$

(Notice what this expression means. The symbol $\Lambda^\alpha_{\ \beta}$ represents a 4×4 constant matrix which multiplies the components of vector x^β to give the new components. We write one index up and one down because we want to sum over the index β, while the raised position of the free index α must be the same in all terms. There is a rigorous meaning to the two index positions which would take us too far afield – for now, I will simply use the correct index positions. We can always sum one raised index and one lowered one, and the objects you are used to calling "vectors" have a raised index: the components of a 3-vector $\mathbf{v}$ are written as v^i.

To transform the wave equation, we use the chain rule to write the derivative with respect to the new coordinates,

$$\frac{\partial}{\partial \tilde{x}^\alpha} = \frac{\partial x^\beta}{\partial \tilde{x}^\alpha} \frac{\partial}{\partial x^\beta}$$

Letting $\bar{\Lambda}^\alpha_{\ \beta}$ be the inverse of $\Lambda^\alpha_{\ \beta}$, so that $\Lambda^\alpha_{\ \mu} \bar{\Lambda}^\mu_{\ \beta} = \delta^\alpha_\beta$, we have

$$\begin{aligned}
x^\alpha &= \bar{\Lambda}^\alpha_{\ \mu} \tilde{x}^\mu \\
\frac{\partial x^\alpha}{\partial \tilde{x}^\beta} &= \frac{\partial}{\partial \tilde{x}^\beta} \left(\bar{\Lambda}^\alpha_{\ \mu} \tilde{x}^\mu \right) \\
&= \bar{\Lambda}^\alpha_{\ \mu} \frac{\partial}{\partial \tilde{x}^\beta} \tilde{x}^\mu \\
&= \bar{\Lambda}^\alpha_{\ \mu} \delta^\mu_\beta \\
&= \bar{\Lambda}^\alpha_{\ \beta}
\end{aligned}$$

The transformation of the derivative operator is therefore

$$\frac{\partial}{\partial \tilde{x}^\alpha} = \bar{\Lambda}^\beta_{\ \alpha} \frac{\partial}{\partial x^\beta}$$

and the d'Alembertian of ψ becomes

$$\begin{aligned}
\eta^{\alpha\beta} \frac{\partial^2 \psi}{\partial \tilde{x}^\alpha \partial \tilde{x}^\beta} &= \eta^{\alpha\beta} \bar{\Lambda}^\mu_{\ \alpha} \frac{\partial}{\partial x^\mu} \left(\bar{\Lambda}^\nu_{\ \beta} \frac{\partial \psi}{\partial x^\nu} \right) \\
&= \eta^{\alpha\beta} \bar{\Lambda}^\mu_{\ \alpha} \bar{\Lambda}^\nu_{\ \beta} \frac{\partial^2 \psi}{\partial x^\mu \partial x^\nu}
\end{aligned}$$

We want this to hold regardless of ψ, so the matrix of partial derivatives is arbitrary. Therefore, in order for the wave equation to be invariant, we must have

$$\eta^{\alpha\beta} \bar{\Lambda}^\mu_{\ \alpha} \bar{\Lambda}^\nu_{\ \beta} = \eta^{\mu\nu}$$

Exercise: Prove that this is equivalent to

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

We consider the special case of motion in the x-direction, so that

$$\begin{pmatrix} a & b & & \\ c & d & & \\ & & 1 & \\ & & & 1 \end{pmatrix}\begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}\begin{pmatrix} a & c & & \\ b & d & & \\ & & 1 & \\ & & & 1 \end{pmatrix} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

so we only need to solve

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} -1 & \\ & 1 \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix} &= \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \\
\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} -a & -c \\ b & d \end{pmatrix} &= \begin{pmatrix} -1 & \\ & 1 \end{pmatrix} \\
\begin{pmatrix} b^2 - a^2 & -ac + bd \\ -ac + bd & d^2 - c^2 \end{pmatrix} &= \begin{pmatrix} -1 & \\ & 1 \end{pmatrix}
\end{aligned}$$

This gives us three equations,

$$\begin{aligned}
b^2 - a^2 &= -1 \\
-ac + bd &= 0 \\
d^2 - c^2 &= 1
\end{aligned}$$

Solving the center equation, $d = \frac{ac}{b}$, so the third equation becomes

$$\begin{aligned}
\frac{a^2}{b^2}c^2 - c^2 &= 1 \\
\left(a^2 - b^2\right)c^2 &= b^2 \\
c^2 &= b^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
c &= \pm b \\
d &= \pm a
\end{aligned}$$

To solve the first equation, we let $b = \sinh\zeta$, and immediately find $a = \cosh\zeta$.

Therefore,

$$\bar{\Lambda}^{\mu}{}_{\alpha} = \begin{pmatrix} \cosh\zeta & \sinh\zeta & & \\ \sinh\zeta & \cosh\zeta & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

where the choice of the $+$ sign preserves the direction of x and t. Inverting,

$$\Lambda^{\mu}{}_{\alpha} = \begin{pmatrix} \cosh\zeta & -\sinh\zeta & & \\ -\sinh\zeta & \cosh\zeta & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

Changing the parameterization puts this in a more familar form. Let

$$\tanh\zeta = \frac{v}{c}$$

Then

$$\begin{aligned}
\cosh\zeta &= \frac{1}{\sqrt{1-\tanh^2\zeta}} \\
&= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \\
\sinh\zeta &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}\frac{v}{c}
\end{aligned}$$

Defining

$$\gamma \equiv \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}$$

we find the transformation of coordinates $\tilde{x}^\alpha = \Lambda^\alpha_{\ \beta} x^\beta$ with

$$\Lambda^\mu_{\ \alpha} = \begin{pmatrix} \gamma & -\frac{\gamma v}{c} & & \\ -\frac{\gamma v}{c} & \gamma & & \\ & & 1 & \\ & & & 1 \end{pmatrix}$$

This gives

$$\begin{aligned}
c\tilde{t} &= \gamma\left(ct - \frac{vx}{c}\right) \\
\tilde{x} &= \gamma\left(x - vt\right) \\
\tilde{y} &= y \\
\tilde{z} &= z
\end{aligned}$$

which is the typical form for a relativistic boost. In the limit as $c \gg v$, we have $\gamma \approx 1$ and this transformation reduces to

$$\begin{aligned}
\tilde{t} &= t \\
\tilde{x} &= x - vt \\
\tilde{y} &= y \\
\tilde{z} &= z
\end{aligned}$$

so we recover the Galilean boost and identify the parameter v with the relative velocity of the frames.

2 Special relativity from Einstein's postulates

It is also possible to derive the reletivistic transformations from postulates.

2.1 The postulates

Special relativity is a combination of two fundamental ideas: the equivalence of inertial frames, and the invariance of the speed of light. Inertial frames are the same in relativistic mechanics as they are in Newtonian mechanics, i.e., frames of reference (sets of orthonormal basis vectors) in which Newton's second law holds. Newton's second law

$$\mathbf{F} = m\mathbf{a}$$

and as we have seen this is unchanged by the 10 distinct Galilean transformations. If we know one frame of reference in which Newton's second law holds, then these transformations give us a 10 parameter family of equivalent frames of reference. Einstein's first postulate is that these inertial frames of reference are indistinguishable. This means that there is no such thing as absolute rest. We can say that two frames move with constant relative velocity, but it is incorrect to say that one is at rest and the other moves.

The constancy of the speed of light, or perhaps better, the existence of a limiting velocity, is demonstrated by the Michaelson-Morely experiment. By measuring the speed of light in two different directions, at different times of the year so that the motion relative to the "fixed stars" is different for each, they found no effect of the motion on the travel time of the light. Difficulties explaining this, and especially difficulties when models were based on the disturbance travelling in a medium, lend support to the idea that light always travels in empty space with the limiting velocity, c. Notice that this is in dramatic conflict with our normal idea of addition of velocities. In Euclidean 3-space, if observer A moves with velocity $\mathbf{v}$ with respect to observer B, and A throws a ball with velocity $\mathbf{u}$, then the velocity of the ball with respect to B is $\mathbf{u}+\mathbf{v}$. But according to this postulate of special relativity, if A shines a light beam it travels with speed c relative to both A and B. We must combine $c\hat{\mathbf{n}}$ with $\mathbf{v}$ to get $c\hat{\mathbf{n}}'$, regardless of the directions of the unit vectors $\hat{\mathbf{n}}, \hat{\mathbf{n}}'$.

With some basic assumptions about the nature of space (specifically, spacetime is a vector space and inertial observers move in straight lines), the two postulates are:

1. The laws of physics are the same in all inertial frames of reference
2. There is a limiting velocity to all physical phenomena, c, found experimentally to be the speed at which light travels in vacuum (and theorized to be the speed at which gravitational waves travel in nearly flat spacetime). This velocity is independent of inertial frame, so that if in one inertial frame an object moves with speed c, then it moves with speed c in all inertial frames.

There are some other basic ideas we will use. Since there is strong evidence for the conservation of momentum, with momenta additively conserved, we still need the physical arena to be a vector space, with linear combinations of vectors giving other vectors. Also, we need a notions of straight lines and distance. We will assume that un-forced particles travel in straight lines, along their initial direction. What constitutes the length turns out to be the central difference between classical and relativistic models.

Another approach to these supplementary assumptions is given in problem 11.1. The assumption that spacetime is homogeneous and isotropic places strong constraints on the allowed transformations, since the transformation cannot depend on location or time. This approach also rules out position or time dependence of a scale factor, Λ, (see below). However, just as the 2-dimensional surface of a sphere is isotropic and homogeneous, there exist constant curvature 4-dimensional spaces which are homogeneous and isotropic, so some further assumption is required.

2.2 Lorentz transformations

In order for a position, $\mathbf{x}$, and time, t, to describe a vector in every frame of reference, we need to restrict possible transformations to linear transformations. Only linear transformations preserve the additivity properties of vectors. This means that the position and time in any two frames of reference must be related by a matrix,

$$\begin{pmatrix} c\tilde{t} \\ \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

or, from above, $\tilde{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta$. It is convenient to write this as a matrix equation, and simplifies the notation if we define

$$\begin{aligned} x_0 &= ct \\ x_1 &= x \end{aligned}$$

$$
\begin{aligned}
x_2 &= y \\
x_3 &= z
\end{aligned}
$$

where c is the postulated universal physical constant with units of velocity. Then we may write the transformation as

$$x'_a = \sum_{b=0}^{3} M_{ab} x_b$$

Now consider two inertial frames, with origins coinciding at time $t = t' = 0$, in which a pulse of light is emitted at time $t = 0$. Picture an expanding spherical wave with radius $ct = ct'$. Then we must have

$$
\begin{aligned}
ct &= \sqrt{x^2 + y^2 + z^2} \\
ct' &= \sqrt{x'^2 + y'^2 + z'^2}
\end{aligned}
$$

Write these relations as

$$
\begin{aligned}
x^2 + y^2 + z^2 - c^2t^2 &= 0 \\
x'^2 + y'^2 + z'^2 - c^2t'^2 &= 0
\end{aligned}
$$

Each of these must hold if the other does, and since the primed and unprimed coordinates are linearly related to one another, they must be proportional,

$$x^2 + y^2 + z^2 - c^2t^2 = \Lambda(\mathbf{v}) \left(x'^2 + y'^2 + z'^2 - c^2t'^2 \right)$$

To restrict Λ, suppose we relate x'_a to a third frame, x''_a. If the relative velocity is $\mathbf{u}$, then we must have

$$
\begin{aligned}
x^2 + y^2 + z^2 - c^2t^2 &= \Lambda(\mathbf{v}) \left(x'^2 + y'^2 + z'^2 - c^2t'^2 \right) \\
&= \Lambda(\mathbf{v}) \Lambda(\mathbf{u}) \left(x''^2 + y''^2 + z''^2 - c^2t''^2 \right)
\end{aligned}
$$

Now choose $\mathbf{u} = -\mathbf{v}$, so that we are back to the original frame, $x''_a = x_a$. Then we require

$$\Lambda(\mathbf{v}) \Lambda(-\mathbf{v})$$

and therefore

$$\Lambda(\mathbf{v}) = e^{f(\mathbf{v})}$$

where $f(-\mathbf{v}) = -f(\mathbf{v})$. Conventionally, we take $f(\mathbf{v}) = 0$, but there exist generalizations of relativity involving nontrivial factors. Setting $\Lambda = 1$ is equivalent to assuming that clocks maintain the same rate as they move from place to place. However, as long as the clock rates in different places are related by a single multiplicative function, there is no measurable effect comparing magnitudes of times that could demonstrate it. From here on, we will take $f(\mathbf{v}) = 0$ and $\Lambda = 1$.

We therefore define

$$s^2 \equiv x^2 + y^2 + z^2 - c^2t^2$$

and require $s'^2 = s^2$ between any two inertial frames. This equivalence defines the *Lorentz transformations*. Any linear transformation preserving the quantity s is a Lorentz transformation.

We check that for motion of O' along the positive x-axis of O, we have

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

where

$$\begin{aligned}\gamma &\equiv \frac{1}{\sqrt{1-\beta^2}}\\ \beta &\equiv \frac{v}{c}\end{aligned}$$

Substituting into s'^2 we have

$$\begin{aligned}s'^2 &= x'^2 + y'^2 + z'^2 - c^2t'^2\\ &= \left[\gamma\left(x-\beta ct\right)\right]^2 + y^2 + z^2 - \left[\gamma\left(ct-\beta x\right)\right]^2\\ &= \gamma^2\left(x^2 - 2\beta xct + \beta^2c^2t^2\right) + y^2 + z^2 - \gamma^2\left(c^2t^2 - 2\beta xct + \beta^2x^2\right)\\ &= \left(\gamma^2 - \gamma^2\beta^2\right)x^2 + \left(2\gamma^2\beta - 2\gamma^2\beta\right)xct - \left(\gamma^2-\gamma^2\beta^2\right)c^2t^2 + y^2 + z^2\end{aligned}$$

and since

$$\begin{aligned}\gamma^2 - \gamma^2\beta^2 &= \left(\frac{1}{\sqrt{1-\beta^2}}\right)^2\left(1-\beta^2\right)\\ &= \frac{1}{1-\beta^2}\left(1-\beta^2\right)\\ &= 1\end{aligned}$$

we have

$$\begin{aligned}s'^2 &= \left(\gamma^2 - \gamma^2\beta^2\right)x^2 + \left(2\gamma^2\beta - 2\gamma^2\beta\right)xct - \left(\gamma^2-\gamma^2\beta^2\right)c^2t^2 + y^2 + z^2\\ &= x^2 - c^2t^2 + y^2 + z^2\\ &= s^2\end{aligned}$$

proving that the transformation is a Lorentz transformation.

Notice that we can use a hyperbolic substitution to rewrite the Lorentz transformation. Define the *rapidity*, ζ, by

$$\beta \equiv \tanh\zeta$$

Then

$$\begin{aligned}\gamma &= \frac{1}{\sqrt{1-\beta^2}}\\ &= \frac{1}{\sqrt{1-\tanh^2\zeta}}\\ &= \frac{1}{\sqrt{1-\frac{\sinh^2\zeta}{\cosh^2\zeta}}}\\ &= \frac{\cosh\zeta}{\sqrt{\cosh^2\zeta - \sinh^2\zeta}}\\ &= \cosh\zeta\end{aligned}$$

and $\gamma\beta = \sinh\zeta$. Then, with $x_0 = ct$, we have

$$\begin{aligned}x'_0 &= x_0\cosh\zeta - x_1\sinh\zeta\\ x'_1 &= -x_0\sinh\zeta + x_1\cosh\zeta\\ x'_2 &= x_2\\ x'_3 &= x_3\end{aligned}$$

The similarity to a rotation,

$$\begin{aligned} x' &= x\cos\theta - y\sin\theta \\ y' &= x\sin\theta + y\cos\theta \end{aligned}$$

is not accidental, but will become clear when we find all Lorentz transformations.

Now suppose the velocity is in an arbitrary direction, $\boldsymbol{\beta}$. We can project the position coordinate $\mathbf{x}$ parallel and perpendicular to $\boldsymbol{\beta}$,

$$\begin{aligned} \mathbf{x}_{\parallel} &= \frac{1}{\beta^2}(\boldsymbol{\beta}\cdot\mathbf{x})\,\boldsymbol{\beta} \\ \mathbf{x}_{\perp} &= \mathbf{x} - \frac{1}{\beta^2}(\boldsymbol{\beta}\cdot\mathbf{x})\,\boldsymbol{\beta} \end{aligned}$$

with $\mathbf{x} = \mathbf{x}_{\perp} + \mathbf{x}_{\parallel}$. The component $\mathbf{x}_{\parallel}$ will behave just like the x-direction in the formula above, while the perpendicular directions $\mathbf{x}_{\perp}$ will be unchanged. The time transforms as before, so we have

$$\begin{aligned} ct' &= \gamma(ct - \boldsymbol{\beta}\cdot\mathbf{x}) \\ \mathbf{x}'_{\parallel} &= \gamma\left(\mathbf{x}_{\parallel} - \boldsymbol{\beta}ct\right) \\ \mathbf{x}'_{\perp} &= \mathbf{x}_{\perp} \end{aligned}$$

The last two may be combined as

$$\begin{aligned} \mathbf{x}' &= \mathbf{x}'_{\parallel} + \mathbf{x}'_{\perp} \\ &= \gamma\left(\mathbf{x}_{\parallel} - \boldsymbol{\beta}ct\right) + \mathbf{x}_{\perp} \\ &= \gamma\left(\frac{1}{\beta^2}(\boldsymbol{\beta}\cdot\mathbf{x})\,\boldsymbol{\beta} - \boldsymbol{\beta}ct\right) + \left(\mathbf{x} - \frac{1}{\beta^2}(\boldsymbol{\beta}\cdot\mathbf{x})\,\boldsymbol{\beta}\right) \\ &= \mathbf{x} + \frac{\gamma - 1}{\beta^2}(\boldsymbol{\beta}\cdot\mathbf{x})\,\boldsymbol{\beta} - \gamma\boldsymbol{\beta}ct \end{aligned}$$

3 Spacetime

We now consider properties the 4-dimensional physical arena called spacetime. The defining properties are that it is a 4-dimensional vector space in which the squared length of any vector (from $(x_1, y_1 z_1, ct_1)$ to $(x_2, y_2 z_2, ct_2)$ is given by

$$s^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 + (x_2 - x_1)^2 - c^2(t_2 - t_1)^2$$

We have seen that the value of s^2 is independent of the inertial frame of reference. If the time interval $(t_2 - t_1)^2$ is larger than the spatial separation, so that $s^2 < 0$, we use the equivalent length

$$c^2\tau^2 = c^2(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (x_2 - x_1)^2$$

To distinguish from 3-dimensional names, s is called the proper length and τ is called the proper time.

3.1 Contravariant Vectors

We will discuss the reasons for this notation later, but from now on, the coordinate labels will be written raised. Thus, for Greek indices $\alpha, \beta, \ldots \in (0, 1, 2, 3)$, we write

$$x^\alpha = \left(x^0, x^1, x^2, x^3\right)$$

where $x^0 = ct$ and x^i for Latin indices $i = 1, 2, 3$ are the usual spatial (x, y, z). This means that use of the Greek or Latin alphabet tells us whether an object is four or three dimensional. We know how the coordinates x^α change when we change to a different frame of reference. We now define a (contravariant) vector, or 4-vector, to be any set of four quantities,

$$\begin{aligned} A^\alpha &= \left(A^0, A^1, A^2, A^3\right) \\ &= \left(A^0, A^i\right) \\ &= \left(A^0, \mathbf{A}\right) \end{aligned}$$

that transform in the same way, i.e.,

$$\begin{aligned} A'^0 &= \gamma\left(A^0 - \boldsymbol{\beta}\cdot\mathbf{A}\right) \\ A'_\parallel &= \gamma\left(A_\parallel - \beta A^0\right) \\ \mathbf{A}'_\perp &= \gamma\mathbf{A}_\perp \end{aligned}$$

It follows immediately that the quantity

$$\|A^\alpha\|^2 = -\left(A^0\right)^2 + \mathbf{A}\cdot\mathbf{A}$$

is the same in any inertial reference frame. This is the length of the 4-vector A^α. We will give alternative notation for this later.

Now let A^α and B^α be any two 4-vectors. Their scalar product or inner product is given by

$$-A^0B^0 + \mathbf{A}\cdot\mathbf{B}$$

This is seen to be invariant by noting that $A^\alpha + B^\alpha$ is also a vector, and writing it as

$$\begin{aligned} -A^0B^0 + \mathbf{A}\cdot\mathbf{B} &= \frac{1}{2}\left(\left(-\left(A^0+B^0\right)^2 + (\mathbf{A}+\mathbf{B})\cdot(\mathbf{A}+\mathbf{B})\right) - \left(-\left(A^0\right)^2 + \mathbf{A}\cdot\mathbf{A}\right) - \left(-\left(B^0\right)^2 + \mathbf{B}\cdot\mathbf{B}\right)\right) \\ &= \frac{1}{2}\left(\|A^\alpha + B^\alpha\|^2 - \|A^\alpha\|^2 - \|B^\alpha\|^2\right) \end{aligned}$$

Since each of the three terms on the right is invariant (that is, unchanged by change of reference frame), the sum is as well, so the inner product is unchanged as well.

More generally, if we write the general form of a Lorentz transformation as

$$x'^\alpha = \sum_{\beta=0}^{3} M^\alpha{}_\beta x^\beta$$

then a 4-vector is any set of four functions A^α which transform as

$$A'^\alpha = \sum_{\beta=0}^{3} M^\alpha{}_\beta A^\beta$$

3.2 Causality

In graphing spacetime, time is generally taken as the vertical axis. Points in spacetime are called *events* and denoted $P(t, \mathbf{x})$. The invariant separation between two events $P(t_1, \mathbf{x}_1), P(t_2, \mathbf{x}_2)$ is given by the invariant interval

$$s^2 = |\mathbf{x}_1 - \mathbf{x}_2|^2 - c^2\left(t_1 - t_2\right)^2$$

or

$$c^2\tau^2 = c^2\left(t_1 - t_2\right)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2$$

whichever is positive. When $s^2 > 0$ the separation is called *spacelike* and when $\tau^2 > 0$ the separation is called *timelike*. When $s^2 = c^2\tau^2 = 0$, the separation of the two events is called *lightlike* or *null*.

The minus sign between the time and space coordinates in the expression for the interval is responsibility for *causal relations* in spacetime. Consider the lightlike lines from any fixed spacetime event, P. This set of null lines is called the *light cone*, and its position in spacetime is agreed on by all inertial observers. As a result, the region above these lines is agreed by all inertial observers to occur at later times (larger values of t) and constitutes the *future* of P. Events lying below the lowest set of lightlike lines are agreed to have earlier values of t, and this region is therefore called the *past* of P. The remaining points of spacetime are called *elsewhere*.

Any object travelling from P at the speed of light must follow a null curve; objects travelling slower than the speed of light follow curves contained inside the future light cone. Moreover, the path lies of a particle travelling slower than the speed of light lies in the future of *every* event on its path. Such a path is called the *world line* of the particle and is said to be a timelike curve.

Suppose P_1 and P_2 are two events on the world line of a particle. Then there exists a frame of reference in which P_1 and P_2 occur at the same spatial location. In this frame of reference,

$$\begin{aligned} c^2\tau_{12}^2 &= c^2 (t_1 - t_2)^2 - |\mathbf{x}_1 - \mathbf{x}_2|^2 \\ &= c^2 (t_1 - t_2)^2 \end{aligned}$$

so that in this particular frame of reference, the proper time interval equals the difference in time coordinates, $\tau_{12} = t_1 - t_2$.

Similarly, suppose two events P_1 and P_2 have spacelike separation

$$s_{12}^2 = |\mathbf{x}_1 - \mathbf{x}_2|^2 - c^2 (t_1 - t_2)^2 > 0$$

Then there exists a frame of reference in which the two events occur at the same value of t, and the proper interval becomes equal to the spatial separation of the events:

$$s_{12}^2 = |\mathbf{x}_1 - \mathbf{x}_2|^2$$

Now consider the world line of a particle. We know (and will demonstrate later) that such a particle always moves with speed less than c. The proper time along its world line is the physical time for the particle. Consider two infinitesimally separated points on the world line. Choose a frame of reference (any will do!) and specify the position of the particle in that frame of reference by $\mathbf{x}(t)$, so that the infinitesimal change in proper time is

$$\begin{aligned} d\tau &= \sqrt{dt^2 - \frac{1}{c^2} d\mathbf{x}^2} \\ &= dt\sqrt{1 - \frac{1}{c^2}\left(\frac{d\mathbf{x}}{dt}\right)^2} \end{aligned}$$

We may not integrate along the world line between any two events A, B, to find the elapsed proper time for the particle,

$$\begin{aligned} \tau_{AB} &= \int_{t_A}^{t_B} dt\sqrt{1 - \frac{1}{c^2}\left(\frac{d\mathbf{x}}{dt}\right)^2} \\ &= \int_{t_A}^{t_B} dt\sqrt{1 - \frac{\mathbf{v}(\mathbf{t})^2}{c^2}} \end{aligned}$$

This shows that the elapsed time for physical processes depends on the motion.

3.3 General derivation of the Lorentz transformations (this section is optional)

We define Lorentz transformations to be those linear transformations of the spacetime coordinates $x^\alpha = (ct, x, y, z)$, $\alpha = 0, 1, 2, 3$ for which

$$\begin{aligned} ds^2 &= \eta_{\alpha\beta} x^\alpha x^\beta \\ &= -c^2t^2 + x^2 + y^2 + z^2 \end{aligned}$$

or equivalently, those linear transformations that preserve the wave equation. Either of these specifications leads to the necessary and sufficient condition

$$\eta_{\mu\nu} \Lambda^\mu_{\ \alpha} \Lambda^\nu_{\ \beta} = \eta_{\alpha\beta}$$

We find all such transformations by first considering infinitesmal ones,

$$\Lambda^\mu_{\ \alpha} = \delta^\mu_{\ \alpha} + \varepsilon^\mu_{\ \alpha}$$

3.3.1 Infinitesimal transformations

The defining condition becomes

$$\begin{aligned} \eta_{\mu\nu} \left(\delta^\mu_{\ \alpha} + \varepsilon^\mu_{\ \alpha}\right) \left(\delta^\nu_{\ \beta} + \varepsilon^\nu_{\ \beta}\right) &= \eta_{\alpha\beta} \\ \eta_{\mu\nu} \delta^\mu_{\ \alpha} \delta^\nu_{\ \beta} + \eta_{\mu\nu} \delta^\mu_{\ \alpha} \varepsilon^\nu_{\ \beta} + \eta_{\mu\nu} \varepsilon^\mu_{\ \alpha} \delta^\nu_{\ \beta} + \eta_{\mu\nu} \varepsilon^\mu_{\ \alpha} \varepsilon^\nu_{\ \beta} &= \eta_{\alpha\beta} \\ \eta_{\alpha\beta} + \eta_{\alpha\nu} \varepsilon^\nu_{\ \beta} + \eta_{\mu\beta} \varepsilon^\mu_{\ \alpha} + O\left(\varepsilon^2\right) &= \eta_{\alpha\beta} \end{aligned}$$

Dropping the higher order term, cancelling $\eta_{\alpha\beta}$, and defining

$$\varepsilon_{\alpha\beta} \equiv \eta_{\alpha\mu} \varepsilon^\mu_{\ \beta}$$

we have

$$\varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} = 0$$

so that $\varepsilon_{\alpha\beta}$ must be antisymmetric.

The most general antisymmetric 4×4 matrix may be written as

$$\begin{aligned} \begin{pmatrix} 0 & a_3 & a_2 & a_3 \\ -a_1 & 0 & b_3 & -b_2 \\ -a_2 & -b_3 & 0 & b_1 \\ -a_3 & b_2 & -b_1 & 0 \end{pmatrix} &= a_1 \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &\quad + \cdots + b_1 \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \\ &= a^i K_i + b^i J_i \end{aligned}$$

where

$$\begin{aligned} a^i &= (a_1, a_2, a_3) \\ b^i &= (b_1, b_2, b_3) \end{aligned}$$

and

$$[K_1]_{\alpha\beta} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$[K_2]_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$[K_3]_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

and finally,

$$[J_1]_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{J}_1 \end{pmatrix}$$

$$[J_2]_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{J}_2 \end{pmatrix}$$

$$[J_3]_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \tilde{J}_3 \end{pmatrix}$$

The 3×3 matrices $\tilde{J}_i$ have components given by the Levi-Civita tensor,

$$\left[\tilde{J}_i\right]_{jk} = \varepsilon_{ijk}$$

To compute the transformations, we need to raise the first index of K_i and J_i using the inverse metric. The three K_i change:

$$\begin{aligned}
[K_1]^{\alpha}_{\ \beta} &= \eta^{\alpha\mu} [K_1]_{\mu\beta} \\
&= \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
[K_2]_{\alpha\beta} &= \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
[K_3]_{\alpha\beta} &= \begin{pmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

while the J_i stay the same, that is, $[J_i]^{\alpha}_{\ \beta}$ has identical components to $[J_i]_{\alpha\beta}$. The transformations generated by the J_i with therefore be different from those generated by K_i.

3.4 Finite transformations

We build up finite transformations as the limit of an infinite number of infinitesmial transformations.

3.4.1 Rotations

Consider the transformations involving J_i first. A general, infinitesimal J_i-type transformation is given by

$$\Lambda = 1 + \mathbf{b} \cdot \mathbf{J}$$

To find a *finite* transformation we take the limit of many infinitesimal ones. Write the infinitesimal vector $\mathbf{b}$ as $\mathbf{b} = \varepsilon \mathbf{n}$ where $\mathbf{n}$ is a unit vector. Then define the finite transformation,

$$\Lambda(\mathbf{n}, \theta) = \lim_{n \to \infty} (1 + \varepsilon \mathbf{n} \cdot \mathbf{J})^n$$

where we take the limit in such a way that $\varepsilon n \to \theta$. To evaluate this we use the binomial theorem,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

to write

$$\begin{aligned} \Lambda(\mathbf{n}, \theta) &= \lim_{n \to \infty} \sum_{k=0}^{n} \binom{n}{k} 1^{n-k} (\varepsilon \mathbf{n} \cdot \mathbf{J})^k \\ &= \lim_{n \to \infty} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \varepsilon^k (\mathbf{n} \cdot \mathbf{J})^k \end{aligned}$$

Now we need powers of $\mathbf{n} \cdot \mathbf{J}$. This is easiest if we focus on the nontrivial 3×3 part, since

$$(\mathbf{n} \cdot \mathbf{J})^n = \begin{pmatrix} 0 & 0 \\ 0 & \left(\mathbf{n} \cdot \tilde{\mathbf{J}}\right)^n \end{pmatrix}$$

Since every term in Λ lies in the lower 3×3 corner, this transformation only affects x, y, z and not t. Let $\tilde{\Lambda}(\mathbf{n}, \theta)$ be the 3-dim transformation,

$$\tilde{\Lambda}(\mathbf{n}, \theta) = \lim_{n \to \infty} \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \varepsilon^k \left(\mathbf{n} \cdot \tilde{\mathbf{J}}\right)^k$$

with $\tilde{\mathbf{J}}$ the 3×3 form of the generators. We find

$$\begin{aligned} \left(\mathbf{n} \cdot \tilde{\mathbf{J}}\right)^2 &= \begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix}^2 \\ &= \begin{pmatrix} -n_2^2 - n_3^2 & n_1 n_2 & n_1 n_3 \\ -n_1 n_2 & -n_1^2 - n_3^2 & n_2 n_3 \\ n_1 n_3 & n_2 n_3 & -n_1^2 - n_2^2 \end{pmatrix} \\ &= \begin{pmatrix} n_1 n_1 & n_1 n_2 & n_1 n_3 \\ n_1 n_2 & n_2 n_2 & n_2 n_3 \\ n_1 n_3 & n_2 n_3 & n_3 n_3 \end{pmatrix} - \left(n_1^2 + n_2^2 + n_3^2\right) \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \\ \left[\left(\mathbf{n} \cdot \tilde{\mathbf{J}}\right)^2\right]^i{}_j &= -\left(\delta^i{}_j - n^i n_j\right) \end{aligned}$$

Taking one more power,

$$\begin{aligned}
\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{3} &= \left[\begin{pmatrix} n_1n_1 & n_1n_2 & n_1n_3 \\ n_1n_2 & n_2n_2 & n_2n_3 \\ n_1n_3 & n_2n_3 & n_3n_3 \end{pmatrix} - \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}\right]\begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} n_1n_1 & n_1n_2 & n_1n_3 \\ n_1n_2 & n_2n_2 & n_2n_3 \\ n_1n_3 & n_2n_3 & n_3n_3 \end{pmatrix}\begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix} - \begin{pmatrix} 0 & n_3 & -n_2 \\ -n_3 & 0 & n_1 \\ n_2 & -n_1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right) \\
&= -\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)
\end{aligned}$$

so we have come back to the original matrix except for a sign. If we define $\tilde{\mathbf{M}} \equiv -\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{2}$ then we may write all powers as

$$\begin{aligned}
\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{2m} &= (-1)^{m}\,\tilde{\mathbf{M}} \\
\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{2m+1} &= (-1)^{m}\,\mathbf{n}\cdot\tilde{\mathbf{J}}
\end{aligned}$$

and the series becomes

$$\begin{aligned}
\tilde{\Lambda}(\mathbf{n},\theta) &= \lim_{n\to\infty}\sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}\varepsilon^{k}\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{k} \\
&= \lim_{n\to\infty}\left[\tilde{\mathbf{1}} + \sum_{m=1}^{n}\frac{n!}{(2m)!\,(n-2m)!}\varepsilon^{2m}\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{2m} + \sum_{m=0}^{n}\frac{n!}{(2m+1)!\,(n-2m-1)!}\varepsilon^{2m+1}\left(\mathbf{n}\cdot\tilde{\mathbf{J}}\right)^{2m+1}\right] \\
&= \tilde{\mathbf{1}} - \tilde{\mathbf{M}} + \tilde{\mathbf{M}}\lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n!}{(2m)!\,(n-2m)!}\varepsilon^{2m} + \mathbf{n}\cdot\tilde{\mathbf{J}}\lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n!}{(2m+1)!\,(n-2m-1)!}\varepsilon^{2m+1}
\end{aligned}$$

where we add and subtract $\tilde{\mathbf{M}}$ so that m starts at 0 in the first sum. Now look at the remaining sums. Multiplying and dividing the first by n^{2m},

$$\begin{aligned}
\lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n!}{(2m)!\,(n-2m)!}\varepsilon^{2m} &= \lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n!}{(2m)!n^{2m}\,(n-2m)!}(n\varepsilon)^{2m} \\
&= \lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n(n-1)(n-2)\cdots(n-2m+1)}{(2m)!n^{2m}}(n\varepsilon)^{2m} \\
&= \lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{2m-1}{n}\right)}{(2m)!}(n\varepsilon)^{2m} \\
&= \sum_{m=0}^{\infty}\frac{(-1)^{m}\,\theta^{2m}}{(2m)!} \\
&= \cos\theta
\end{aligned}$$

and similarly for the second

$$\begin{aligned}
\lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,n!}{(2m+1)!\,(n-2m-1)!}\varepsilon^{2m+1} &= \lim_{n\to\infty}\sum_{m=0}^{n}\frac{(-1)^{m}\,\eta^{2m+1}}{(2m+1)!} \\
&= \sin\theta
\end{aligned}$$

The full transformation is therefore

$$\tilde{\Lambda}(\mathbf{n},\theta) = \tilde{\mathbf{1}} - \tilde{\mathbf{M}} + \tilde{\mathbf{M}}\cos\theta + \mathbf{n}\cdot\tilde{\mathbf{J}}\sin\theta$$

where

$$\tilde{M}_{ij} = \delta_{ij} - n_i n_j$$

This is a rotation through an angle θ about the $\mathbf{n}$ direction. To see this, consider the effect on an arbitrary position vector, x^i. In components, remembering that $\left[\tilde{J}_i\right]_{jk} = \varepsilon_{ijk}$,

$$\begin{aligned}
\tilde{\Lambda}^i{}_j x^j &= \left[\delta^i_j - \left(\delta^i_j - n^i n_j\right)\right] x^j + \left(\delta^i_j - n^i n_j\right) x^j \cos\theta + n_k \varepsilon^{ki}{}_j \sin\theta \\
&= n^i n_j x^j + \left(x^i - n^i n_j x^j\right)\cos\theta + n_k \varepsilon^{ki}{}_j x^j \sin\theta \\
&= n^i n_j x^j + \left(x^i - n^i n_j x^j\right)\cos\theta - \varepsilon^i{}_{kj} n^k x^j \sin\theta \\
\tilde{\mathbf{\Lambda}}\mathbf{x} &= \mathbf{n}(\mathbf{n}\cdot\mathbf{x}) + (\mathbf{x} - \mathbf{n}(\mathbf{n}\cdot\mathbf{x}))\cos\theta - (\mathbf{n}\times\mathbf{x})\sin\theta
\end{aligned}$$

Now divide $\mathbf{x}$ into parts parallel and perpendicular to $\mathbf{n}$,

$$\begin{aligned}
\mathbf{x}_\parallel &= \mathbf{n}(\mathbf{n}\cdot\mathbf{x}) \\
\mathbf{x}_\perp &= \mathbf{x} - \mathbf{n}(\mathbf{n}\cdot\mathbf{x})
\end{aligned}$$

and notice that $\mathbf{n}\times\mathbf{x}$ is perpendicular to both.

$$\tilde{\mathbf{\Lambda}}\mathbf{x} = \mathbf{x}_\parallel + \mathbf{x}_\perp(\cos\theta - 1) - (\mathbf{n}\times\mathbf{x}_\perp)\sin\theta$$

The part of $\mathbf{x}$ parallel to $\mathbf{n}$ is unaffected by the transformation, while the perpendicular part undergoes a rotation by θ in the plane perpendicular to $\mathbf{n}$

3.4.2 Boosts

Now consider the transformations generated by K_i. The basic approach is identical, with only the generators differing. Identical steps, taking the limit of many infinitesimal transformations, lead to

$$\Lambda(\mathbf{n},\theta) = \lim_{n\to\infty}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\varepsilon^k(\mathbf{n}\cdot\mathbf{K})^k$$

where the limit is taken with $n\varepsilon \to \zeta$ where ζ is finite. The powers of $\mathbf{n}\cdot\mathbf{K}$ again split into even and odd. Starting with

$$\begin{aligned}
\mathbf{n}\cdot\mathbf{K} &= \begin{pmatrix} 0 & -n_1 & -n_2 & -n_3 \\ -n_1 & 0 & 0 & 0 \\ -n_2 & 0 & 0 & 0 \\ -n_3 & 0 & 0 & 0 \end{pmatrix} \\
(\mathbf{n}\cdot\mathbf{K})^2 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & n_1 n_1 & n_2 n_1 & n_3 n_1 \\ 0 & n_1 n_2 & n_2 n_2 & n_3 n_2 \\ 0 & n_1 n_3 & n_2 n_3 & n_3 n_3 \end{pmatrix} \\
(\mathbf{n}\cdot\mathbf{K})^3 &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & n_1 n_1 & n_2 n_1 & n_3 n_1 \\ 0 & n_1 n_2 & n_2 n_2 & n_3 n_2 \\ 0 & n_1 n_3 & n_2 n_3 & n_3 n_3 \end{pmatrix} \begin{pmatrix} 0 & -n_1 & -n_2 & -n_3 \\ -n_1 & 0 & 0 & 0 \\ -n_2 & 0 & 0 & 0 \\ -n_3 & 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & -n_1 & -n_2 & -n_3 \\ -n_1 & 0 & 0 & 0 \\ -n_2 & 0 & 0 & 0 \\ -n_3 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

so in general,

$$\begin{aligned}
(\mathbf{n}\cdot\mathbf{K})^{2m+1} &= \mathbf{n}\cdot\mathbf{K} \\
(\mathbf{n}\cdot\mathbf{K})^{2m} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & n_1n_1 & n_2n_1 & n_3n_1 \\ 0 & n_1n_2 & n_2n_2 & n_3n_2 \\ 0 & n_1n_3 & n_2n_3 & n_3n_3 \end{pmatrix} \equiv \mathbf{N}
\end{aligned}$$

Notice that there is no alternating sign now. The power series rearranges as before to give

$$\begin{aligned}
\Lambda(\mathbf{n},\theta) &= \lim_{n\to\infty}\sum_{k=0}^{n}\frac{n!}{k!\,(n-k)!}\varepsilon^k(\mathbf{n}\cdot\mathbf{K})^k \\
&= \mathbf{1}-\mathbf{N}+\mathbf{N}\sum_{m=0}^{\infty}\frac{\zeta^{2m}}{(2m)!}+\mathbf{n}\cdot\mathbf{K}\sum_{m=0}^{\infty}\frac{\zeta^{2m+1}}{(2m+1)!} \\
&= \mathbf{1}-\mathbf{N}+\mathbf{N}\cosh\zeta+\mathbf{n}\cdot\mathbf{K}\sinh\zeta
\end{aligned}$$

If we define $n^\alpha = \left(0, n^i\right)$ and $m^\alpha = (1, \mathbf{0})$ then we may write

$$\begin{aligned}
[\mathbf{N}]^\alpha_{\ \beta} &= n^\alpha n_\beta + m^\alpha m_\beta \\
[\mathbf{n}\cdot\mathbf{K}]^\alpha_{\ \beta} &= -m^\alpha n_\beta - n^\alpha m_\beta
\end{aligned}$$

and the Lorentz boost becomes

$$\Lambda^\alpha_{\ \beta} = \delta^\alpha_\beta + (n^\alpha n_\beta + m^\alpha m_\beta)(\cosh\zeta - 1) - (m^\alpha n_\beta + n^\alpha m_\beta)\sinh\zeta$$

To see that this is a boost, let $\mathbf{n}$ lie in the x-direction. Then

$$\begin{aligned}
\mathbf{n}\cdot\mathbf{K} &= \begin{pmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
\mathbf{N} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}$$

and therefore

$$\begin{aligned}
\Lambda(\mathbf{n},\theta) &= \mathbf{1}-\mathbf{N}+\mathbf{N}\cosh\zeta+\mathbf{n}\cdot\mathbf{K}\sinh\zeta \\
&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \cosh\zeta & 0 & 0 & 0 \\ 0 & \cosh\zeta & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \cosh\zeta & -\sinh\zeta & 0 & 0 \\ -\sinh\zeta & \cosh\zeta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\end{aligned}$$

just as we found previously as on of the transformations preserving the wave equation.