

Chapter 14

Relativistic covariance: Lorentz transformation and 4-vectors

December 3, 2019

The concept of covariance is extended to transformations between different cartesian coordinates in spacetime, in particular those that are in relative motion. The concepts of 4-scalars, 4-vectors and 4-tensors, as well as 4-scalar, 4-vector and 4-tensor fields, are introduced.

Contents

1	Introduction	
2	Lorentz transformation	
2.1	Events and coordinates	1
2.2	Galilean transformation	1
2.3	Speed of light and MM Experiment .	1
2.4	Setting up the transformation	1
2.5	The invariant	1
2.6	Condition on the transformation matrix	1
2.7	Choice of units	1
2.8	Difference form	1
3	Scalars, vectors and tensors	
3.1	Scalars and vectors	1
3.2	Tensors	1
3.3	Transformation law for covariant index	1
3.4	Examples of tensors	1
3.5	Contraction theorem	1
4	Scalar, vector and tensor fields	
4.1	Scalar field	1
4.2	Vector fields	1
4.3	Tensor fields	1
5	Differentiation	10
6	Covariance under Lorentz transforma- tions	10

1 Introduction

The laws of physics must be the same according to observers (a) whose axes are oriented differently, because there is no preferred orientation in space, and also (b) who are in relative motion, because there is no preferred state of rest. This Chapter examines the consequences of the latter constraint, developing the concepts of Lorentz transformations and 4-vectors, in parallel with the concepts of rotations and 3-vectors in the last Chapter. Basic knowledge of special relativity is assumed.

This long Chapter can be summarized by one single statement:

The indices will automatically take care of all transformation and covariance properties.

2 Lorentz transformation

2.1 Events and coordinates

In relativity, the basic object is not a point P , but an *event* E , e.g., the emission of a photon from an atom. Each event occurs at a definite time t and a definite place \mathbf{x} , and is characterized by 4 coordinates:

$$E = (t, x^1, x^2, x^3)$$

How do these 4 coordinates transform between a frame S and another frame S' moving at a velocity \mathbf{V} relatively to S ? For simplicity, assume

$$V^1 = V \quad , \quad V^2 = V^3 = 0$$

and the origins coincide at $t = t' = 0$ (**Figure 1**).

2.2 Galilean transformation

Start with the familiar Galilean transformations. These will turn out to be incorrect. Assume that (a) clocks are not affected by motion: $t' = t$; and

(b) lengths are not affected by motion. Then from **Figure 2**,

$$x' = x - Vt \quad (1)$$

where x is the coordinate of a particle in S and x' is the coordinate in S' .

The velocities

$$v = \frac{dx}{dt}, \quad v' = \frac{dx'}{dt'}$$

are then related by

$$v' = v - V \quad (2)$$

2.3 Speed of light and MM Experiment

The Michelson–Morley (MM) experiment showed that the speed of light¹ c is the same according to all observers and in all directions — which in particular means that (2) is not correct. Refer to any standard textbook for an account of the MM experiment.

Although this is the usual and the historical account of the experimental fact, in line with our stance that relativity is *logically* prior to EM and light, you can replace “velocity of light” everywhere by “the highest possible velocity of any particle” and pretend the MM experiment dealt with the latter concept.

No absolute motion

If Galilean transformations were correct, the velocity V of a train can be determined without referring to the outside (**Figure 3**): measure the velocity of light coming from the front and from the back. They should be $c \pm V$; the difference would reveal absolute motion.

Actually, such is not the case. Even for a moving observer, the speed of light is the same in all directions (isotropic), and is always $c = 3.0 \times 10^8 \text{ m s}^{-1}$. Instead of trying to *explain* this fact, we use it as the starting point to derive the transformation between moving coordinates.

2.4 Setting up the transformation

The derivation follows closely the analog of rotations in the last Chapter.

Basic object

The basic object is an event E , with coordinates

$$\begin{aligned} E &= (t, x^1, x^2, x^3) \text{ in } S \\ E &= (t', x'^1, x'^2, x'^3) \text{ in } S' \end{aligned}$$

¹In these notes, we always mean speed of light in vacuum.

Linear assumption

Assume that they are related linearly:

$$\begin{aligned} t' &= ?t + ?x^1 + ?x^2 + ?x^3 \\ x'^1 &= ?t + ?x^1 + ?x^2 + ?x^3 \\ x'^2 &= ?t + ?x^1 + ?x^2 + ?x^3 \\ x'^3 &= ?t + ?x^1 + ?x^2 + ?x^3 \end{aligned}$$

The 16 coefficients would be expressed more conveniently using index notation or matrix notation, and with two conventions: (a) call time the “zeroth component”;² (b) convert this component to the same unit. Since the speed of light is a universal constant, define

$$x^0 = ct, \quad x'^0 = ct'$$

Index notation and summation convention

Vectors³ in 4D are denoted as $\vec{x}, \vec{y}, \vec{p}$ etc., e.g.,

$$\vec{x} = (x^0, \mathbf{x}) = (x^0, x^1, x^2, x^3)$$

The column vector is denoted as

$$[x] = \begin{bmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{bmatrix}$$

The components are collectively denoted as x^μ . Sometimes we shall say loosely “the vector x^μ ” meaning “the vector whose components are x^μ ”. An index such as μ, ν, \dots runs from 0 to 3 (whereas an index such as i runs from 1 to 3).

If an index such as μ appears twice in the same term, once up, once down,⁴ it is understood to be summed from 0 to 3:

$$a_\mu b^\mu \equiv \sum_{\mu=0}^3 a_\mu b^\mu$$

Thus it is necessary to distinguish between an upper index and a lower index. The coordinate vector is defined with upper indices. Vectors with lower indices will be constructed later.

General transformation

With this convention, the most general linear transformation is

$$\begin{aligned} x'^\mu &= L^\mu{}_\nu x^\nu \\ [x'] &= [L][x] \end{aligned} \quad (3)$$

²Some authors call it the “fourth component”; it does not matter.

³For the moment, the word “vector” just means such a collection of 4 numbers. Further conditions on transformation properties will be imposed later.

⁴If it appears twice, but both up or both down, then almost certainly you made a mistake.

written first in component form and then in matrix form. The elements of $[L]$ are just the 16 “?” shown earlier. The matrix form looks a little “dangerous” because it does not distinguish between upper and lower indices. However, this does not matter because the indices will all work out correctly in the end according to the following rules.

- A free index such as μ in (3) must appear once in every term, either always up or always down.

$a^\mu = b^\mu$	Yes
$a^\mu = b_\mu$	No

- A dummy index must appear twice in one term, once up, once down.

$a^\mu b_\mu$	Yes
$a^\mu b^\mu$	No

2.5 The invariant

We now claim that the MM experiment tells us that the following is an invariant

$$\begin{aligned}\sigma^2 &= -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2 \\ &= -(ct)^2 + |\mathbf{x}|^2\end{aligned}$$

In other words, we claim that it is equal to

$$\begin{aligned}\sigma'^2 &= -(x'^0)^2 + (x'^1)^2 + (x'^2)^2 + (x'^3)^2 \\ &= -(ct')^2 + |\mathbf{x}'|^2\end{aligned}$$

Although this quantity is written as σ^2 , it can in fact be negative. (We shall never take the square root of this quantity in general.) Also, unlike the case of rotations, \mathbf{x} is different in the two frames, and in general $|\mathbf{x}|^2 \neq |\mathbf{x}'|^2$. This is true even for the Galilean transformation (1).

We now show that $\sigma^2 = \sigma'^2$, in three steps.

(1) Proportional

Suppose a short pulse of light is emitted from the origin at the time when the two origins coincide, i.e., at $(ct, x^1, x^2, x^3) = (ct', x'^1, x'^2, x'^3) = (0, 0, 0, 0)$. Let the event E be the receiving of the pulse by an observer. Because the velocity of light is exactly c in S :

$$\begin{aligned}(ct)^2 &= (x^1)^2 + (x^2)^2 + (x^3)^2 \\ \sigma^2 &= 0\end{aligned}$$

Similarly, because the velocity of light is also exactly c in S' :

$$\begin{aligned}(ct')^2 &= (x'^1)^2 + (x'^2)^2 + (x'^3)^2 \\ \sigma'^2 &= 0\end{aligned}$$

In other words, $\sigma^2 = 0$ if and only if $\sigma'^2 = 0$. Thus these two quantities are proportional for any event:⁵

$$\sigma^2 = A(\mathbf{V}) \sigma'^2$$

The proportionality constant A may depend on the relative velocity.

(2) The constant depends only on magnitude of relative velocity

The quantities σ^2 and σ'^2 are scalars under spatial rotation. So they (and hence their ratio) cannot depend on the direction of any vector. Therefore $A(\mathbf{V})$ depends only on $|\mathbf{V}|$. In particular

$$A(\mathbf{V}) = A(-\mathbf{V})$$

(3) Consider reverse transformation

Consider the reverse transformation. Interchange the role of σ^2 and σ'^2 . Because the relative velocity is now opposite, $A(\mathbf{V})$ is changed to $A(-\mathbf{V})$.

$$\sigma'^2 = A(-\mathbf{V}) \sigma^2$$

Combining these equations,

$$\begin{aligned}\sigma'^2 &= A(-\mathbf{V}) \{A(\mathbf{V}) \sigma'^2\} \\ &= \{A(-\mathbf{V}) A(\mathbf{V})\} \sigma'^2 \\ &= A(\mathbf{V})^2 \sigma'^2\end{aligned}$$

Hence

$$A(\mathbf{V}) = 1$$

The reason for discarding the negative root is as follows. If V is small, obviously $x'^\mu \approx x^\mu$ and $A(\mathbf{V}) \approx 1$; so we take the + sign. The sign cannot change suddenly as \mathbf{V} is changed, so we must always take the + sign.

This proves that

$$\boxed{\sigma^2 = -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2} \quad (4)$$

is an invariant. It is called the *invariant interval*. We use a *spacelike metric*, meaning that the sign of σ^2 follows that of the spatial components. Some texts use the opposite convention.

The invariant is a mathematical way of stating the result of the MM experiment. It is the analog of Pythagoras' theorem in 3D. Since there are some minus signs, the space is said to be *pseudo-Euclidean*. More specifically a space where the invariant is given by (4), with three plus signs and one minus sign, is called *Minkowski space* in physics.

⁵This argument is not strictly rigorous, and it is left as an exercise to identify and fill in the gap.

Metric

We can write σ^2 as

$$\boxed{\sigma^2 = \eta_{\mu\nu} x^\mu x^\nu} \quad (5)$$

where

$$\eta_{\mu\nu} = \begin{cases} -1 & \mu = \nu = 0 \\ +1 & \mu = \nu = 1 \text{ or } 2 \text{ or } 3 \\ 0 & \mu \neq \nu \end{cases} \quad (6)$$

The matrix $\eta_{\mu\nu}$ takes the place of δ_{ij} in usual Euclidean space. In fact, if we change $\eta \mapsto \delta$, we would obtain the familiar results of the last Chapter.

Lower indices

For any vector x^μ , define the corresponding vector x_μ as

$$x_\mu \equiv \eta_{\mu\nu} x^\nu \quad (7)$$

The reverse process to raise an index goes as

$$x^\mu = \eta^{\mu\nu} x_\nu \quad (8)$$

where $\eta^{\mu\nu}$ is obviously the inverse of $\eta_{\mu\nu}$. In this case

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

(This is a very exceptional case in which the upper and lower indices do not balance.)

In matrix notation

$$\begin{aligned} \sigma^2 &= [x^T]^\bullet [\eta]_{\bullet\bullet} [x]^\bullet \\ [x]_\bullet &= [\eta]_{\bullet\bullet} [x]^\bullet \\ [x]^\bullet &= [\eta]^{\bullet\bullet} [x]_\bullet \\ [\eta]^{\bullet\bullet} &= [\eta^{-1}]_{\bullet\bullet} \end{aligned}$$

The dots are placeholders to indicate whether the index is upper or lower; they are not needed once you become familiar with the notation.

The minus sign

Raising or lowering an index just changes the sign of the 0 component, e.g.,

$$\begin{aligned} x^\mu &= (1, 7, 4, 2) \\ x_\mu &= (-1, 7, 4, 2) \end{aligned}$$

The metric $[\eta]$ is just a way of taking care of this minus sign, which comes from (4) — i.e., the square of the 0 component enters the invariant with a minus sign.

Why do we need a matrix $[\eta]$ with 16 elements to deal with a mere minus sign? Answer: This notation provides a good stepping stone to general relativity, in which $\eta_{\mu\nu}$ is replaced by a more general (and position-dependent) matrix $g_{\mu\nu}$. Thus,

to go from Euclidean space to special relativity to general relativity simply involves

$$[\delta] \rightarrow [\eta] \rightarrow [g]$$

Another question is often asked: Can we define

$$x^0 = ict \quad ???$$

instead? If so, (4) would appear “nicely” with all plus signs. Some elementary books do this, but it is very bad practice, for two reasons.

- In Euclidean space, a subset such as $\sigma^2 = 1$ is bounded, because each component is bounded (in this case at most 1). But in Minkowski space, a subset such as $\sigma^2 = 1$ is unbounded — the components can be as large as you want. Thus there is a difference in topology, which is obscured if we “hide” the minus sign.
- In quantum mechanics, the factor i appears, with expressions such as $\Psi^* \Psi$; Ψ^* means changing $i \mapsto -i$. But this applies only to the “genuine” i ’s, not to the “fake” i ’s introduced in $x^0 = ict$. It would be a nightmare to keep track of the genuine i ’s versus the “fake” i ’s⁶ that appear when one considers relativistic quantum mechanics.

For these reasons, we shall continue with the matrix $[\eta]$.

2.6 Condition on the transformation matrix

The invariant places conditions on the transformation matrix $[L]$, to be next derived in three equivalent ways.

(1) Explicitly in components for (1+1)D

Notation: (1+1)D means 1 space + 1 time dimension.

Consider a case where the relative velocity \mathbf{V} is along the 1-direction. Then x^2 and x^3 are not transformed: $x'^2 = x^2$, $x'^3 = x^3$; we only need to transform x^0, x^1 . The matrix $[L]$ is reduced to 2×2 .

$$\begin{bmatrix} x'^0 \\ x'^1 \end{bmatrix} = [L] \begin{bmatrix} x^0 \\ x^1 \end{bmatrix}$$

Let

$$[L] = \begin{bmatrix} p & s \\ q & r \end{bmatrix}$$

⁶Also, in EM, we often write plane waves as, say $\mathbf{E} = E_0 \hat{\mathbf{e}}_x \exp[i(kz - \omega t)]$, and the i that appears in the phase is also a “genuine” one in this sense, to be reversed in sign when encountering expressions such as $\mathbf{E}^* \cdot \mathbf{E}$. But this objection is less severe, since the use of the “genuine” i in EM is only a matter of convenience rather than necessity; we always mean just the real part. That is of course not the case in QM.

so that explicitly

$$\begin{aligned} x'^0 &= px^0 + sx^1 \\ x'^1 &= qx^0 + rx^1 \end{aligned}$$

Let $x^2 = x^3 = 0$. Then

$$\begin{aligned} \sigma'^2 &= -(x'^0)^2 + (x'^1)^2 \\ &= (-p^2 + q^2)(x^0)^2 + 2(-ps + qr)x^0x^1 \\ &\quad + (-s^2 + r^2)(x^1)^2 \end{aligned}$$

But this must equal

$$\sigma^2 = -1 \cdot (x^0)^2 + 0 \cdot x^0x^1 + 1 \cdot (x^1)^2$$

as an identity. Hence we get three conditions:

$$\begin{aligned} p^2 - q^2 &= 1 \\ ps - qr &= 0 \\ s^2 - r^2 &= -1 \end{aligned} \tag{9}$$

Except for some minus signs, these conditions are the same as those for rotations in the last Chapter. The minus signs come about because of the minus sign in the invariant σ^2 .

(2) In general using index notation

$$\begin{aligned} x'^\mu &= L^\mu{}_\rho x^\rho \\ x'^\nu &= L^\nu{}_\sigma x^\sigma \\ \sigma'^2 &= \eta_{\mu\nu} x'^\mu x'^\nu = \eta_{\mu\nu} (L^\mu{}_\rho x^\rho) (L^\nu{}_\sigma x^\sigma) \\ &= (\eta_{\mu\nu} L^\mu{}_\rho L^\nu{}_\sigma) x^\rho x^\sigma \\ \sigma^2 &= \eta_{\rho\sigma} x^\rho x^\sigma \end{aligned}$$

Since σ^2 and σ'^2 must be equal as an identity,

$$\boxed{\eta_{\mu\nu} L^\mu{}_\rho L^\nu{}_\sigma = \eta_{\rho\sigma}} \tag{10}$$

Compare with the analogous equations in the last Chapter; in particular, note that if we replace η by δ , these equations would reduce exactly to the ones for rotation.

Problem 1

Check that (10) gives exactly the three conditions in (9). §

Problem 2

Consider the case of (3+1)D.

- (a) How many conditions are there in (10)? Hint: note the symmetry under $\rho \leftrightarrow \sigma$.
- (b) Hence determine how many free parameters there are in $L^\mu{}_\rho$. Half of these correspond to rotation and half of these correspond to relative motion. Explain physically why there are this number of free parameters. §

(3) Using matrix notation

$$\begin{aligned} [x']^\bullet &= [L]^\bullet \bullet [x]^\bullet \\ [x'^T]^\bullet &= [x^T]^\bullet [L^T]_\bullet^\bullet \end{aligned}$$

In the second equation, the two dots in $[L]$ have been transposed.

$$\begin{aligned} \sigma'^2 &= [x'^T]^\bullet [\eta]_{\bullet\bullet} [x']^\bullet \\ &= \left\{ [x^T]^\bullet [L^T]_\bullet^\bullet \right\} [\eta]_{\bullet\bullet} \{ [L]^\bullet \bullet [x]^\bullet \} \\ &= [x^T]^\bullet \{ [L^T]_\bullet^\bullet [\eta]_{\bullet\bullet} [L]^\bullet \bullet \} [x]^\bullet \\ \sigma^2 &= [x^T]^\bullet [\eta]_{\bullet\bullet} [x]^\bullet \end{aligned}$$

Since these are equal as an identity,

$$\boxed{[L^T]_\bullet^\bullet [\eta]_{\bullet\bullet} [L]^\bullet \bullet = [\eta]_{\bullet\bullet}} \tag{11}$$

This is the best way to remember the result.

With experience, there is no need to write these dots. They can be constructed using two rules, which we explain with reference to this example.

- First choose the indices on the RHS, say both down. Then the leftmost and rightmost indices on the LHS must follow suit.
- The indices in the middle are all paired in the sense of matrix multiplication (e.g., the second and third dots on the LHS). The two dots of a pair must be one up, one down.

Problem 3

Derive (10) from (11). §

Determinant

Using (11), it is readily shown that

$$\det[L] = 1$$

in which the negative solution is discarded if $[L]$ is continuously linked to the identity, i.e., it is a *proper* Lorentz transformation.

Using minimum number of parameters

We do this only for (1+1)D. The situation is similar to the case of rotations and is left as a problem.

Problem 4

Of the 4 parameters p, q, r, s in $[L]$, regard s as the free parameter and define $s = -\sinh \alpha$. Find p, q, r in terms of α by using the three equations in (9). You will need to choose the sign of a square root. Explain the physical meaning of your choice. Your answer should be

$$[L] = \begin{bmatrix} \cosh \alpha & -\sinh \alpha \\ -\sinh \alpha & \cosh \alpha \end{bmatrix} \tag{12}$$

Note that the two off-diagonal entries have the same sign. §

Relate to relative velocity

Of course we can use α to specify the transformation. But it is more usual to use the relative speed V . We need to relate the two.

Let S' be moving along the +1-axis at a speed V relative to S . Thus the origin of S' ($x'^1 = 0$) is described in S as $x^1 = Vt$. Put this into the transformation:

$$\begin{aligned} x'^1 &= -\sinh \alpha x^0 + \cosh \alpha x^1 \\ 0 &= -\sinh \alpha (ct) + \cosh \alpha (Vt) \\ \tanh \alpha &= \frac{V}{c} \end{aligned} \quad (13)$$

It is conventional to define

$$\beta = V/c$$

as the dimensionless relative velocity. For “ordinary” speeds, $|\beta| \ll 1$; for relative motion near the speed of light, $\beta \approx 1$. It is easy to show

$$\begin{aligned} \cosh \alpha &= \gamma \\ \sinh \alpha &= \gamma \beta \end{aligned}$$

where

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

For small V , $\gamma \approx 1$. For $V \approx c$, $\gamma \rightarrow \infty$.

Putting these back into [L] gives the usual form of the *Lorentz transformation*:

$$\begin{aligned} x'^0 &= \gamma(x^0 - \beta x^1) \\ x'^1 &= \gamma(-\beta x^0 + x^1) \end{aligned} \quad (14)$$

Problem 5

Verify that $[L(\alpha_2)][L(\alpha_1)] = [L(\alpha_1 + \alpha_2)]$. §

Problem 6

Show that when two transformations are performed one after another, the relative velocities β “add” as

$$\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1 \beta_2}$$

Hint: Use the addition law for $\tanh \alpha$. §

Problem 7

Based on the above result, show that if $|\beta_1| < 1$, $|\beta_2| < 1$, then $|\beta| < 1$. Can you also give a more direct proof using the angle α ? §

Problem 8

Start from (14) and solve for x^0, x^1 in terms of x'^0, x'^1 . Show that the reverse transformation derived this way has the same form, but with $\beta \rightarrow -\beta$. §

Signs

The signs can be remembered as follows. (a) The diagonal terms (i.e., relating x'^0 to x^0 , or x'^1 to x^1) always have + signs. (b) The off-diagonal terms with β have the same sign. (c) Whether the off-diagonal sign is + or – is easily determined by considering $\beta \rightarrow 0$, $\gamma \rightarrow 1$, for which the Galilean transformation should be valid. For example, from the second equation in (14)

$$\begin{aligned} x'^1 &\approx (-\beta x^0 + x^1) = x^1 - \beta ct \\ &= x^1 - Vt \end{aligned}$$

The sign of the Vt term is easily checked with reference to **Figure 2**.

2.7 Choice of units

Analogy

A stupid surveyor measures distances along the east-west direction (x^1) in m and distance along the north-south direction (x^2) in km. His “Pythagoras theorem” would read

$$\sigma^2 = (x^1)^2 + k^2(x^2)^2 = \text{invariant}$$

where $k = 1000$ is a conversion factor to get his measurements into the same units. It would be much smarter to use the same units for both directions and get rid of k . Moreover, it would be silly to ask about the physical significance of the “fundamental constant” k — its value is not a property of nature, but a (stupid) human convention.

In the same way, the factors of c that appear, e.g., in

$$\sigma^2 = -(ct)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$$

or

$$(ct') = \gamma(ct - \beta x^1)$$

are there because we do not use the same unit for the zeroth component.

Choosing velocity of light to be unity

We can get rid of c by setting

$$c = 3.00 \times 10^8 \text{ m s}^{-1} = 1$$

This means, for example, measuring times in years and distances in light years. All formulas become simpler. When we get a final result that is dimensionally “wrong”, simply multiply by a suitable power of $c (= 1)$.

We shall adopt a “half-way” convention: we shall keep the factors of c , but typically would write (ct) as one combination.

Problem 9

- (a) Convert a time of 3.0 m to conventional units.
- (b) Convert an energy of 10^{-10} kg to conventional units. §

2.8 Difference form

Instead of the coordinates x^μ of one event, we can consider the displacement Δx^μ between two events. Because the transformation of coordinates is linear, we get immediately

$$\Delta x'^\mu = L^\mu_\nu \Delta x^\nu$$

From this, we can also write

$$\boxed{\frac{\partial x'^\mu}{\partial x^\nu} = L^\mu_\nu}$$

The difference (and differential) form has another advantage. In general relativity, spacetime is curved. But if we look at a *small* portion of a curved surface, it is nearly flat. So locally, i.e., in terms of *small* displacements Δx^μ (strictly speaking infinitesimal displacements dx^μ) the formulas in general relativity look very similar to the above difference and differential forms.

3 Scalars, vectors and tensors

This Section introduces scalars, vectors and tensors in spacetime, i.e., 4-scalars, 4-vectors and 4-tensors. The prefix “4” will be omitted if there is no danger of confusion.

3.1 Scalars and vectors

Four-scalars

A 4-scalar is any quantity that remains unchanged under both a rotation of axes and a transformation to another uniformly moving frame. Example: the mass (also called the rest mass). Note that time is not a 4-scalar. We shall see in the next Chapter that charge is a 4-scalar.

Basic 4-vector

The basic 4-vector $\Delta \vec{x}$ is a line joining two neighboring points in spacetime, in other words the spacetime displacement. Its cartesian coordinates are

$$\begin{aligned}\Delta x^\mu &= (\Delta x^0, \Delta x^1, \Delta x^2, \Delta x^3) \\ &= (c\Delta t, \Delta \mathbf{x})\end{aligned}$$

Note that 4-vectors are denoted as \vec{x} for example, whereas 3-vectors are denoted as \mathbf{x} for example.

Transformation laws

Under a transformation, the components change according to (3), in which the matrix $[L]$ is independent of the vector.

Rotations can be regarded as a special case of Lorentz transformations, with

$$[L] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & [R] \\ 0 & 0 & 0 \end{bmatrix}$$

Other 4-vectors

A collection of any four quantities

$$v^\mu = (v^0, v^1, v^2, v^3)$$

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

$$\begin{aligned}v'^\mu &= L^\mu_\nu v^\nu \\ [v'] &= [L][v]\end{aligned}$$

is a 4-vector by definition. Obviously 4-vectors may be obtained by adding or subtracting 4-vectors, or multiplying or dividing a 4-vector by a 4-scalar (or invariant).

“Length” of a vector

The length of a displacement vector Δx is defined through

$$(\Delta s)^2 = \Delta \vec{x} \cdot \Delta \vec{x} = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu$$

similar to (4). The lengths of other 4-vectors are defined in the same way. Note that $(\Delta s)^2$ may be $+$, $-$ or 0 .

Metric and lower indices

The matrix $\eta_{\mu\nu}$ is defined by (5), and serves to lower and raise indices through (7) and (8). The definition of lower indices may be applied to any vector, and in fact also to tensors. We adopt the following names.

$$\begin{aligned}v^\mu &= (1, 0) \text{ vector or} \\ &\quad \text{a contravariant vector} \\ v_\mu &= (0, 1) \text{ vector or} \\ &\quad \text{a covariant vector}\end{aligned}$$

The first notations, though not conventional, is more systematic. It may be even more suggestive to write $(1, 0)$ and $(0, 1)$ as columns, but that would be more cumbersome on the page.

Condition on the transformation matrix

Since the length has to be an invariant (i.e., a 4-scalar), a condition is imposed on $[L]$, namely (10) or (11).

Dot product

Consider a 4-vector $\vec{z} = \vec{x} + \vec{y}$. Now the following is a 4-scalar

$$\vec{z} \cdot \vec{z} = \vec{x} \cdot \vec{x} + 2\vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{y}$$

It follows that

$$\vec{x} \cdot \vec{y} = \eta_{\mu\nu} x^\mu y^\nu$$

is also a 4-scalar, called the *dot product*.

3.2 Tensors

Basic rank-2 tensor

Let \vec{x}, \vec{y} be two 4-vectors and define the 16 quantities as the basic rank 2 tensor:

$$t^{\mu\nu} = x^\mu y^\nu$$

To be specific, this is a $(2, 0)$ tensor, and we can likewise define a $(0, 2)$ tensor

$$t_{\mu\nu} = x_\mu y_\nu$$

and a $(1, 1)$ tensor

$$t^\mu{}_\nu = x^\mu y_\nu$$

These are related to each other through the raising or lowering of indices by $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$.

The transformation law for $t^{\mu\nu}$ is

$$\begin{aligned} t'^{\mu\nu} &= x'^\mu y'^\nu = (L^\mu{}_\rho x^\rho)(L^\nu{}_\sigma y^\sigma) \\ &= L^\mu{}_\rho L^\nu{}_\sigma x^\rho y^\sigma \\ t'^{\mu\nu} &= L^\mu{}_\rho L^\nu{}_\sigma t^{\rho\sigma} \end{aligned}$$

General rank-2 tensor

Any 16 quantities $t^{\mu\nu}$ which transform in this way is called a rank-2 tensor.

Higher rank tensors

A rank-3 tensor is 4^3 quantities which transform as

$$t'^{\mu\nu\rho} = L^\mu{}_\alpha L^\nu{}_\beta L^\rho{}_\gamma t^{\alpha\beta\gamma}$$

The definitions for higher ranks, and for lower indices, are similar.

By the way, the term “tensor” is the general one; a $(0, 0)$ tensor is also called a scalar, and a $(1, 0)$ or $(0, 1)$ tensor is also called a vector.

3.3 Transformation law for covariant index

How does a covariant (i.e., lower) index transform?

$$\begin{aligned} \Delta x'_\mu &= \eta_{\mu\nu} \Delta x^\nu \\ &= \eta_{\mu\nu} L^\nu{}_\rho \Delta x^\rho \\ &= \eta_{\mu\nu} L^\nu{}_\rho \eta^{\rho\sigma} \Delta x_\sigma \end{aligned} \quad (15)$$

The first way to view this is that the two $[\eta]$ factors simply raise and lower the indices on $[L]$, i.e.,

$$\eta_{\mu\nu} L^\nu{}_\rho \eta^{\rho\sigma} = L_\mu{}^\sigma$$

Hence

$$\Delta x'_\mu = L_\mu{}^\sigma \Delta x_\sigma$$

In other words, transform by $[L]$; keep indices in usual way.

Problem 10

Show that the transformation of a lower index can also be written as

$$\Delta x'_\mu = \Delta x_\sigma [L^{-1}]^\sigma{}_\mu$$

Thus: transform by $[L^{-1}]$ rather than $[L]$, but on the right. §

The above result is particularly convenient when we want to show that contracting an upper index with a lower index leads to an invariant.

$$\begin{aligned} x'_\mu y'^\mu &= \{x_\sigma [L^{-1}]^\sigma{}_\mu\} \{[L]^\mu{}_\nu y^\nu\} \\ &= x_\sigma \{[L^{-1}][L]\}^\sigma{}_\nu y^\nu \\ &= x_\sigma [I]^\sigma{}_\nu y^\nu = x_\sigma y^\sigma \end{aligned} \quad (16)$$

3.4 Examples of tensors

The metric is a tensor

If $\eta_{\mu\nu}$ is given by (5) in every frame is to be a $(0, 2)$ tensor, we must have

$$\eta_{\mu\nu} = \eta'_{\mu\nu} \stackrel{?}{=} L_\mu{}^\rho L_\nu{}^\sigma \eta_{\rho\sigma} \quad (17)$$

Problem 11

Show that (17) follows from (11). Also show that $\eta^{\mu\nu}$ is a $(2, 0)$ tensor. §

Metric with one upper and one lower index

What happens if we raise *one* index in $\eta_{\mu\nu}$? In general

$$t^\mu{}_\nu = \eta^{\mu\rho} t_{\rho\nu}$$

So putting $t \mapsto \eta$,

$$\eta^\mu{}_\nu = \eta^{\mu\rho} \eta_{\rho\nu}$$

$$\begin{aligned} \text{\color{red}\$} \eta^\mu{}_\nu &= \text{\color{red}\$} \eta^{\mu\rho} t_{\rho\nu} \\ &= \text{\color{red}\$} L^\mu{}_\alpha L^\nu{}_\beta \eta_{\alpha\beta} \\ &= \text{\color{red}\$} L^\mu{}_\alpha L^\nu{}_\beta \delta^\alpha{}_\beta \\ &= \text{\color{red}\$} L^\mu{}_\alpha L^\nu{}_\beta \epsilon_{\alpha\beta\gamma\delta} \epsilon^{\gamma\delta} \\ &= \text{\color{red}\$} L^\mu{}_\alpha L^\nu{}_\beta \epsilon_{\alpha\beta\gamma\delta} \epsilon^{\gamma\delta} \end{aligned}$$

Problem 12

Show that $\eta^\mu{}_\nu$ as defined above is identical with $\delta^\mu{}_\nu$. Hence $\delta^\mu{}_\nu$ is a $(1, 1)$ tensor. §

The antisymmetric symbol

In 4D, the totally antisymmetric symbol $\epsilon_{\mu\nu\alpha\beta}$ is defined as (a) + 1 for the indices being 0123 or an even permutation; (b) - 1 for the indices being an odd permutation of 0123; and (c) 0 if any two indices are equal. In fact, for any 4×4 matrix (with rows and columns labelled 0 to 3),

$$\det A = \epsilon_{\mu\nu\alpha\beta} A^{0\mu} A^{1\nu} A^{2\alpha} A^{3\beta}$$

We now claim that $\epsilon_{\mu\nu\alpha\beta}$ is a $(0, 4)$ tensor. Since $\epsilon_{\mu\nu\alpha\beta}$ is defined to have the same value in any frame, we need to check

$$\begin{aligned}\epsilon_{\mu\nu\alpha\beta} &= \epsilon'_{\mu\nu\alpha\beta} \\ &\stackrel{?}{=} L_\mu^\rho L_\nu^\sigma L_\alpha^\gamma L_\beta^\delta \epsilon_{\rho\sigma\gamma\delta}\end{aligned}$$

Let us check the case of $(\mu\nu\alpha\beta) = (0123)$. Then this amounts to checking

$$L_0^\rho L_1^\sigma L_2^\gamma L_3^\delta \epsilon_{\rho\sigma\gamma\delta} \stackrel{?}{=} 1$$

But the LHS is just $\det L$, which we have shown is unity. The other components are similar.

Note that there is no analogous rank 3 tensor $\epsilon_{\mu\nu\alpha}$ in 4D. Hence there is no cross product between two 4-vectors.

3.5 Contraction theorem

Again we consider an example for the forward theorem. Let x^μ be a $(1, 0)$ vector and y_μ be a $(0, 1)$ vector, then

$$\varphi = x^\mu y_\mu \quad (18)$$

is a $(0, 0)$ tensor, i.e., a scalar. The proof is similar to the case of 3-vectors and is left as an exercise.

Problem 13

Prove the reverse theorem: If in (18), it is known that φ is a $(0, 0)$ tensor (i.e., scalar) for every $(1, 0)$ vector x^μ , then y_μ is a $(0, 1)$ vector. §

Summary

Although the mathematics of 4-vectors looks slightly complicated, the index notation automatically keeps track of everything. This is really all that needs to be remembered.

4 Scalar, vector and tensor fields

A *field* is a quantity that depends on the position in spacetime, i.e., a function of x^μ ($\mu = 0, 1, 2, 3$) in 4D.

4.1 Scalar field

Suppose there is a quantity φ with the following properties.

- *Field*: To every spacetime point P (i.e., event), there is a number $\varphi(P)$.
- *Scalar*: $\varphi(P)$ has the same value in all frames.

$$\varphi(P) = \varphi'(P)$$

In this case, different frames are related by Lorentz transformations, with rotations as special cases.

There is no known example of *classical* scalar fields. (a) Only long-range fields can be detected classically, i.e., at macroscopic distances. (b) Such fields must correspond to particles with exactly zero mass; see argument about Yukawa potentials in Chapter 1. (c) A mass can be exactly zero only if there is a principle enforcing it, namely a gauge principle — which applies only to some vector and tensor fields, but not to scalar fields.⁷

4.2 Vector fields

Suppose at each point P in spacetime (i.e., each event), there are 4 quantities

$$A^\mu(P) = (A^0(P), A^1(P), A^2(P), A^3(P))$$

such that under a coordinate transformation

$$A'^\mu(P) = L^\mu_\nu A^\nu(P) \quad (19)$$

where $[L]$ is the transformation matrix for coordinate displacements. Then A^μ is said to be a 4-vector field.

The most important example is the 4-vector potential in EM:

$$\vec{A} = (\Phi/c, \mathbf{A}) \quad (20)$$

where Φ is the scalar potential and \mathbf{A} is the vector potential, i.e.,

$$\begin{aligned}\mathbf{E} &= -\nabla\Phi - \frac{\partial\mathbf{A}}{\partial t} \\ \mathbf{B} &= \nabla \times \mathbf{A}\end{aligned} \quad (21)$$

Note that the factor of c ensures that all components of (20) have the same units.

We can either try to check that $(\Phi/c, \mathbf{A})$ as defined by (21) transforms like (19), that is, if we know how \mathbf{E} and \mathbf{B} transform; or alternatively, we can *postulate* that there is a 4-vector potential A^μ and show that (21) follows. The latter approach will be adopted later in this course.

4.3 Tensor fields

Suppose at each point P in spacetime (i.e., each event), there are 4^2 quantities $F^{\mu\nu}(P)$, such that under a coordinate transformation

$$F'^{\mu\nu}(P) = L^\mu_\rho L^\nu_\sigma F^{\rho\sigma}(P)$$

Then $F^{\mu\nu}$ is said to be a $(2, 0)$ tensor field. Tensor fields of higher rank are defined in a similar manner.

⁷Students are not expected to completely understand these statements at this time.

The most important example is the EM field tensor (which is antisymmetric) to be introduced in the next Chapter. Other examples include the gravitational field in relativity. Weak gravitational fields can be regarded as a tensor $h^{\mu\nu}(x)$ (which is symmetric) on flat spacetime.

5 Differentiation

Consider functions defined for points $x = (x^0, x^1, x^2, x^3)$ in spacetime. (To allow for curved spacetime, we shall not put an arrow on x .) The case of 3-space is easily recovered by setting $x^0 = 0$. Moreover, in this section all displacements Δx are understood to be infinitesimal.

Differentiation of a scalar

Start with a scalar field $\varphi(x)$ and compare its value at two neighbouring points. Then

$$\begin{aligned}\Delta\varphi &= \varphi(x + \Delta x) - \varphi(x) \\ &= \frac{\partial\varphi}{\partial x^0} \Delta x^0 + \frac{\partial\varphi}{\partial x^1} \Delta x^1 + \dots \\ &= \frac{\partial\varphi}{\partial x^\mu} \Delta x^\mu\end{aligned}$$

Now $\Delta\varphi$ is a scalar, and Δx^μ is a $(1, 0)$ vector. Hence, by the inverse contraction theorem, $\partial\varphi/\partial x^\mu$ transforms like a $(0, 1)$ vector. Note that an upper index in the denominators behaves like a lower index.

The gradient operator

In other words,

$$\partial_\mu = \frac{\partial}{\partial x^\mu}$$

transforms like a $(0, 1)$ vector operator. The corresponding $(1, 0)$ vector operator is ∂^μ .

$$\begin{aligned}\partial_\mu &= \left(\frac{1}{c} \frac{\partial}{\partial t}, \boldsymbol{\nabla} \right) \\ \partial^\mu &= \left(-\frac{1}{c} \frac{\partial}{\partial t}, \boldsymbol{\nabla} \right)\end{aligned}$$

The d'Alembertian

By the contraction theorem, the following is a 4-scalar:

$$\partial^\mu \partial_\mu = -\frac{1}{c^2} \frac{\partial}{\partial t^2} + \boldsymbol{\nabla}^2$$

which is the natural wave operator for a field that travels at the speed c .

6 Covariance under Lorentz transformations

So with all this apparatus developed, let us come back and ask: How do we know that electrodynamics is relativistically covariant? In other words, if S and S' are two observers in relative motion, would they be able to apply the same laws of physics? The problem can be approached at four levels.

(1) One phenomenon at a time

Take one phenomenon or experiment at a time, and examine it in the two frames. An example was given at the beginning of this course:

- According to S , a point charge moves across a magnetic field, and there is a magnetic force $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$.
- According to S' , a magnetic field moves across a stationary charge. The changing \mathbf{B} field causes an electric field \mathbf{E} (Faraday's Law), and \mathbf{E} acts on the point charge. (The point charge, being stationary, cannot feel any magnetic force.)

Such analysis is possible, and indeed necessary in elementary courses to build up physical intuition. But the task would never be finished, since there are infinitely many phenomena to be examined.

(2) Examining the laws of physics

All phenomena are governed by a finite number of laws; in EM they are just the Lorentz force law and Maxwell's equations. We can check that all these — as conventionally written — remain the same under a Lorentz transformation. This is so messy we would not even bother.

(3) Explicitly covariant formalism

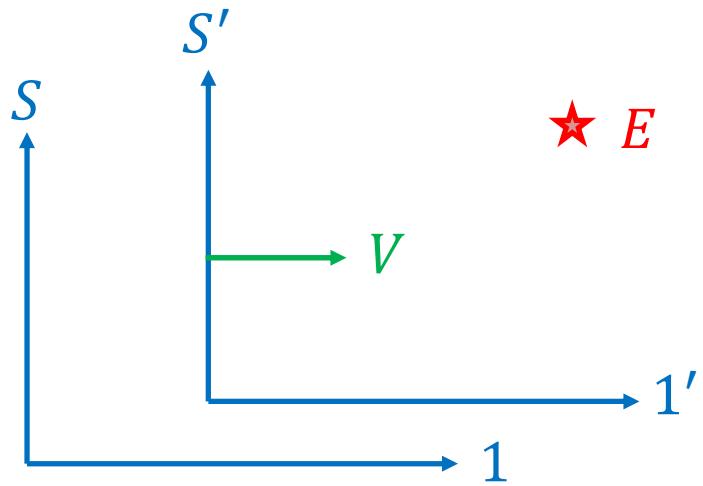
All this is made easier by writing the laws in a form that is explicitly covariant under Lorentz transformations, using 4-vectors. This will be the focus of the next Chapter.

(4) Invariant formalism

Even better, if we can formulate all the laws of EM in terms of an action formalism, i.e., to have everything emerge from a single statement

$$\delta\mathcal{S} = 0$$

and if we can guarantee that \mathcal{S} is a 4-scalar — the same thing according to S and S' — then of course the end results must be the same in the two frames. This we shall also develop later.



$$(t, x^1, x^2, x^3) \xrightarrow{?} (t', x'^1, x'^2, x'^3)$$

Figure 1 The coordinates of an event E in two reference frames

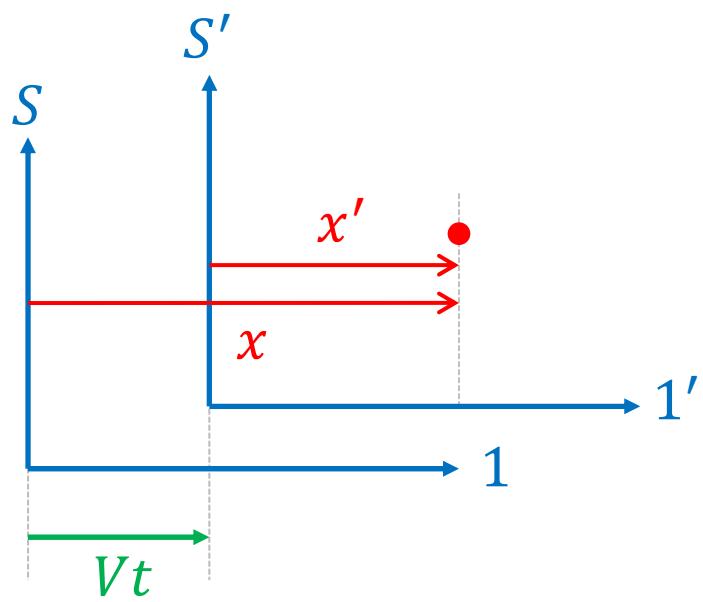


Figure 2 The relationship between the coordinates x and x' for two frames in relative motion

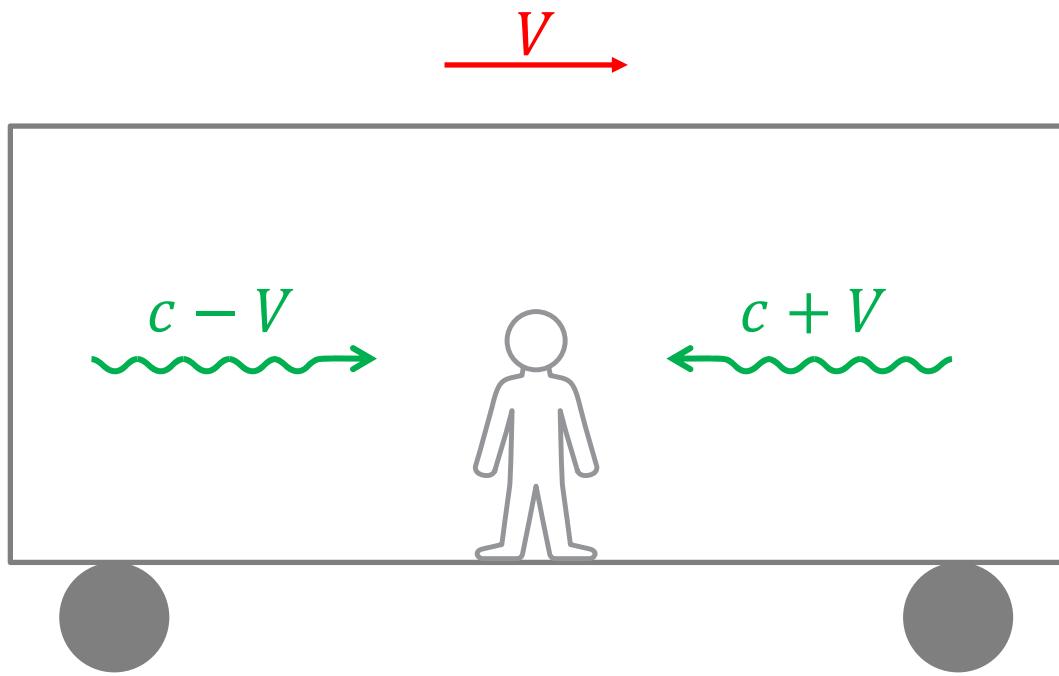


Figure 3 Determine the absolute motion by measuring the difference in velocity of the light coming from the front and from the back