1 Ordinary vectors and rotation

We introduced 4-vectors as 4-component objects which transform in a certain way *under Lorentz transformations*. This may seem unrelated to how you first learned about (ordinary) vectors, which was probably without reference to any transformation. In fact, ordinary vectors (which we will also refer to as 3-vectors or Euclidean vectors) can be identified as 3-component objects which transform in a certain way *under rotations*. In this section, we will examine this way of thinking about 3-vectors and ordinary scalars, so that we can generalize properly to the relativistic (4-dimensional) case.

If a displacement vector $\vec{r}$ points at angle ϕ with respect to some axis, it has components $r\cos\phi$ and $r\sin\phi$ parallel and perpendicular to that axis. This geometric or trigonometric property of $\vec{r}$ is essential to figuring out the transformation matrix taking $\vec{r} = (x, y, z)$ to $\vec{r'} = (x', y', z')$ when the coordinate system is rotated.

Now any vector $\vec{A}$ (be it a velocity, force,...) has components $A\cos\phi$ and $A\sin\phi$ in directions parallel and perpendicular to an axis with respect to which the direction of $\vec{A}$ is at angle ϕ. This property is shared by all 3-vectors with the displacement vector $\vec{r}$. Hence the components of $\vec{A}$ should transform exactly the same way as the components of $\vec{r}$.

For example, if the x'-y' coordinate system is obtained by rotating the x-y system by angle θ (around the common z, z' axis), and if

$$\vec{A} \; = \; A_x\hat{\imath} + A_y\hat{\jmath} + A_z\hat{k} \; = \; A'_x\hat{\imath}' + A'_y\hat{\jmath}' + A'_z\hat{k}'$$

then

$$\begin{pmatrix} A'_x \\ A'_y \\ A'_z \end{pmatrix} \; = \; \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}$$

More generally, if under a rotation of coordinate axes, the position coordinates transform as

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \; = \; \mathcal{R} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

where $\mathcal{R}$ is an orthogonal 3×3 matrix, then the components of *any* 3-vector $\vec{A}$ will transform under that rotation as

$$\begin{pmatrix} A'_x \\ A'_y \\ A'_z \end{pmatrix} \; = \; \mathcal{R} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix} \; .$$

In fact, this may be thought of as the *definition* of a 3-vector:

Non-relativistic vectors

A 3-vector (or non-relativistic vector or 'ordinary' vector) is a three-component object whose components transform the same way under rotations as the components of displacement vectors.

This definition is consistent with the idea that a vector is a quantity that has both magnitude and direction. (Maybe this is how you were introduced to vectors?) A quantity having both magnitude and direction should be decompose-able into components according to the usual trigonometric rules, and therefore these components should transform under rotations the same way as the components of a displacement vector.

1.1 Non-examples

A vector is not just a collection of three numbers, i.e., not every 3-tuple is a vector. Examples of non-vectors:

1. Imagine an object with mass m and charge q. Of course, the velocity $\vec{v} = (v_x, v_y, v_z)$ of the object is a vector.

   What about the combination (q, m, v_z)? It has three components, each of which are physical properties related to the object. This is NOT a vector because these components do not transform correctly under rotation. The mass m and charge q do not change at all under rotation.

2. If $d\vec{r} = (dx, dy, dz)$ is a displacement vector, then the combination $(dx, 2dy, dz)$ is NOT a 3-vector.

   Why? Under a rotation around the z axis, we know that $dx' = dx\cos\theta - dy\sin\theta$ and $dy' = dx\sin\theta + dy\cos\theta$ because $d\vec{r}$ is a displacement. This is utterly incompatible with

   Totally incorrect!

   $$dx' = dx\cos\theta - (2dy)\sin\theta, \quad 2dy' = dx\sin\theta + (2dy)\cos\theta$$

   which is what one would need, if $(dx, 2dy, dz)$ were to be a valid 3-vector.

1.2 What is a scalar?

In non-relativistic physics, you might have learned that a *scalar* is an object which, unlike a vector, has no direction. Thus, you might have thought of a scalar as a single component object, i.e., something that can be described by a single number.

We will have to expand our understanding of scalars a bit. Not every single-component number is a scalar!

Non-relativistic scalars

In non-relativistic physics, a scalar is a single-component quantity that remains invariant under rotations.

Examples of scalars:

- The dot product of two 3-vectors is a scalar.

  If the rotation matrix is $\mathcal{R}$, then the dot product of $\vec{A}$ and $\vec{B}$ transforms under rotation to

  $$\vec{A}' \cdot \vec{B}' = A'^{\mathrm{T}} B' = (\mathcal{R}A)^{\mathrm{T}}(\mathcal{R}B) = A^{\mathrm{T}}(R^{\mathrm{T}}R)B = A^{\mathrm{T}}B = \vec{A} \cdot \vec{B}$$

  where we have used the convention that A, B are column vectors and A^{T}, B^{T} are row vectors.

Important: Inner products produce scalars!

- Thus, the work done on an object ($\vec{F} \cdot d\vec{r}$) is a non-relativistic scalar, and so is the kinetic energy which is the total work done in getting an object up to its current speed. Note that work (or energy) is not invariant under a Galilean transformation or a Lorentz transformation. Nevertheless, this is a non-relativistic scalar — the requirement is invariance under **rotations**, not any other type of transformation.

- In particular, the magnitudes of vectors are scalars, because they are obtained by taking the inner product of the vector with itself. If $\vec{v}$ is a velocity, the speed $v = |\vec{v}| = \sqrt{\vec{v} \cdot \vec{v}}$ is a non-relativistic scalar.

  The speed of a massive particle has no invariance under Galilean or Lorentz transformations, only under rotations.

Magnitudes of vectors are scalars!

- Physical examples of (non-relativistic) scalars: the mass or charge carried by an object. In non-relativistic physics they are not affected by the choice of coordinate system or axis orientation.

- We could define numerical values which are independent of reference frame. For example, the numbers 42 and $3 - i$ are scalars, as long as we understand them to have these values independent of the orientation of reference frame.

1.2.1 A non-scalar

A component of a 3-vector is NOT a scalar. For example, if $\vec{v}$ is a 3-velocity, then the component v_x is NOT a scalar. It gets transformed to something else under rotations, i.e., is not invariant. For example, under a rotation around the z axis, it gets transformed to

$$v'_x = v_x \cos\theta - v_y \sin\theta$$

which, except for very special cases, $\neq v_x$. Thus, even though v_x is a single number, it is not a scalar.

1.3 Summary

We have learned that a vector is not just a collection of three numbers, and a scalar is not just any single number. The transformation properties under rotation is what determines whether a three-component object is a vector or not, and whether a one-component object is a scalar or not.

2 Four-vectors and invariants

We are now ready to define 4-vectors, in analogy to 3-vectors, in terms of their transformation properties. For good measure, we will also define analogs of 3-scalars in the 4-dimensional world. We will discuss some properties of these objects, and define inner products between 4-vectors.

2.1 Deinfing 4-vectors

Definition: 4-vectors

The 4-tuplet $A = (A^0, A^1, A^2, A^3)$ is a 4-vector if the components A^μ transform under Lorentz transformations in the same way as a spacetime interval (cdt, dx, dy, dz) or as a spacetime coordinate (ct, x, y, z).

Comments:

1. Reminder: A Lorentz transformation is any 4×4 transformation matrix Λ that satisfies $\Lambda^{\mathrm{T}} g \Lambda = g$. Lorentz transformations consist of boosts, rotations, and any combination thereof.

2. Since the last three components of a spacetime coordinate are the coordinates of a displacement 3-vector, the last three components of any 4-vector must form an ordinary vector or Euclidean 3-vector. In other words, if $A = (A^0, A^1, A^2, A^3)$ is a 4-vector, then $(A^1, A^2, A^3) = \vec{A}$ is an ordinary vector.

3. If inertial frames Σ and $\tilde{\Sigma}$ are related by a standard boost, then the components of the 4-vector A as measured from the two frames will be related by

$$\tilde{A}^0 = \gamma_v \left(A^0 - (v/c) A^1 \right),$$
$$\tilde{A}^1 = \gamma_v \left(-(v/c) A^0 + A^1 \right),$$
$$\tilde{A}^2 = A^2, \quad \tilde{A}^3 = A^3,$$

   i.e., exactly the same way as spacetime coordinates transform under the standard boost.

4. If inertial frames Σ and $\tilde{\Sigma}$ are related by a mere rotation (and no boost), then the components transform as

$$\begin{pmatrix} \tilde{A}^0 \\ \tilde{A}^1 \\ \tilde{A}^2 \\ \tilde{A}^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & \mathcal{R} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix}$$

where $\mathcal{R}$ is the rotation matrix. This is exactly how spacetime coordinates ($x^0 = ct, x^1 = x, x^2 = y, x^3 = z$) would transform under rotations. Remember that rotations are Lorentz transformations, as the transformation matrix appearing above satisfies $\Lambda^T g \Lambda = g$.

This equation above shows visually that the components (A^1, A^2, A^3) transform as an ordinary vector under rotations, and hence that these components form a 3-vector.

5. We have used superscripts for the coordinate indices. This could be potentially confusing as we are used to thinking of superscripted numbers as powers (exponents): does A^2 refer to A-squared or to the 2nd spatial component of A? Unfortunately, we have to rely on context to make sure we know which meaning is attached.

6. A 4-vector should transform like an interval between two spacetime events, $(c\Delta t, \Delta x, \Delta y, \Delta z)$ or (cdt, dx, dy, dz). Lorentz transformations are defined to be homogeneous, i.e., the zero event $(0, 0, 0, 0)$ is defined to be the same for every inertial frame. Therefore we can equally well talk about transformations of the coordinates of a single event (ct, x, y, z), because this can be thought of as the interval between this event and the zero event.

7. Since a four-vector $A = (A^0, A^1, A^2, A^3)$ transforms like a spacetime interval, this means that the combination

$$(A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2$$

is invariant under Lorentz transformations. We will later interpret this as the 'norm' of the 4-vector A.

Other than spacetime events, we have met another 4-vector: the 4-momentum of an object, $(E/c, p_x, p_y, p_z)$. Therefore the combination

$$\left(\frac{E}{c}\right)^2 - p_x^2 - p_y^2 - p_z^2 = \frac{E^2}{c^2} - \vec{p} \cdot \vec{p} = \frac{E^2}{c^2} - p^2$$

is invariant under LTs. This is consistent with the relativistic expressions we have learned for E and $\vec{p}$, because

$$\frac{E^2}{c^2} - p^2 = \frac{p^2c^2 + m^2c^4}{c^2} - p^2 = m^2c^2$$

where m is the rest mass, a property of the object, independent of the inertial frame.

8. Four-vectors are also known as Minkowski vectors. In more advanced physics literature based on special relativity (e.g., in a text on quantum field theory) they might just be called vectors. One should be able to

figure out from the context whether 'vector' refers to an ordinary vector or a 4-vector. Of course, in other branches of physics or math, a 'vector' may mean something else altogether, for example, wavevectors or 'ket' vectors in quantum mechanics.

9. In your introductory mechanics and your electromagnetism class, characters representng 3-vectors (usually a Latin character) were usually distinguished typographically from scalars, either by using an arrow on top of the character or using boldface. We have been following this tradition, using an arrow on top to indicate that a character represents a 3-vector.

   It would be nice if there was always a typographic signal to tell us what type of object a character represents. But this would be difficult — there are just too many types of objects. For example, when a letter represents a matrix, or a row vector, there is no common convention to show this typographically. You need to figure out from the context what type of object is represented by which character.

   Unfortunately, this is also true for 4-vectors. The most common convention in typed literature is to represent them with characters in italics without any special markings. We have done this above, by saying that A is a 4-vector. Clearly, you need to be extra careful and ask yourself continuously what type of object is being represented by which symbol.

   However, below in this chapter, we will use a thick arrow convention to represent 4-vectors, thus, $\overset{\Longrightarrow}{A}$ is a 4-vector. Hopefully, this will ease the transition into dealing with 4-vectors.

2.2 Defining Lorentz scalars

Now for scalars in relativity. These may be called 4-scalars, which sounds silly because they are single-component objects. "Lorentz scalar" is a more appropriate name. In a textbook on quantum field theory, they would be just referred to as a scalar.

Definition: Lorentz scalars

In relativistic physics, a quantity is called a scalar if it is invariant under Lorentz transformations.

Comments and examples:

1. An obvious Lorentz scalar is the speed of light c — this quantity is Lorentz invariant by definition.

2. Another set of Lorentz scalars: numbers which are defined independently of reference frame. If we define $\alpha = 2.0$ independently of reference frame, then α is a scalar.

3. The interval $c\Delta\tau = \sqrt{c^2(\Delta t)^2 - |\Delta\vec{r}|^2}$ between two events is a scalar.

   Note that the interval Δt between two events is not a Lorentz scalar, as it is not an invariant under LT, and is generally different when measured from another (boosted) frame. Of course, Δt is a scalar in non-relativistic physics as time is invariant under non-relativistic rotations.

   Similarly, $|\Delta\vec{r}|$ is the Euclidean norm of a 3-vector, and hence is a non-relativistic scalar, but it is not a Lorentz scalar.

4. In general, a non-relativistic scalar is not necessarily a Lorentz scalar, but a Lorentz scalar has to be a non-relativistic scalar as well because rotations are valid LTs.

5. Another example of a Lorentz scalar is the rest mass of an object. We have noted previously that the rest mass is related to the 4-momentum:

$$m = \sqrt{\frac{E^2}{c^4} - \frac{p^2}{c^2}}.$$

   In non-relativistic physics, both E and $p = \sqrt{\vec{p}\cdot\vec{p}}$ are scalars, as they are invariant under rotations. However, neither of these quantities are Lorentz scalars.

6. There was a long period (perhaps half of the 20th century) during which it was common to refer to the combination $\tilde{m} = \gamma_v m$ as 'mass'. (Some popular science books still use this terminology.) This was tempting because momentum would then have the familiar form of 'mass' $\times$ velocity. The Feynman lectures, dating from the 1960s, use this langauge as well.

   This terminlogy is no longer used in the professional physics literature. One argument is that m is Lorentz invariant (a scalar) while $\gamma_v m$ is not; hence m is more deserving of the fundamental-sounding name 'mass'. The combination $\gamma_v m$, if it is to be given a name at all, could be called the 'relativistic mass'.

2.3 Adding & multiplying — Inner products & norms

A 4-vector multiplied (or divided) by a Lorentz scalar gives another 4-vector. If $\overrightarrow{A} = (A^0, A^1, A^2, A^3)$ is a 4-vector and s is a Lorentz scalar, then $\overrightarrow{B} = s\overrightarrow{A} = (sA^0, sA^1, sA^2, sA^3)$ is a 4-vector.

Exercise: Show that the elements of $\overrightarrow{B}$ indeed transform as a 4-vector under a Lorentz transformation.

The sum of two 4-vectors is also a 4-vector: If $\overrightarrow{C} = (C^0, C^1, C^2, C^3)$ and $\overrightarrow{D} = (D^0, D^1, D^2, D^3)$ are 4-vectors, then so is $\overrightarrow{E} = \overrightarrow{C} + \overrightarrow{D} = (C^0 + D^0, C^1 + D^1, C^2 + D^2, C^3 + D^3)$. Exercise: Show!

Note that we have implicity defined the operations: of multiplication by a scalar and addition of two 4-vectors.

Next, we will define an **inner product** between two 4-vectors. Recall that the inner (dot) product of two 3-vectors produces a non-relativistic scalar. By analogy, we would like an inner product of two 4-vectors to produce a Lorentz scalar, i.e., a single-component object invariant under Lorentz transformations. This is achieved by the following definition.

Definition: Inner product of two 4-vectors

The inner product of 4-vectors (A^0, A^1, A^2, A^3) and (B^0, B^1, B^2, B^3) is the quantity $A^0B^0 - A^1B^1 - A^2B^2 - A^3B^3$.

If we want to continue to use our thick-arrow notation to distinguish 4-vectors, we might write this is as

$$\overrightarrow{A} \star \overrightarrow{B} = A^0B^0 - A^1B^1 - A^2B^2 - A^3B^3.$$

I've used a $\star$ to denote our new inner prouct. After you get used to 4-vectors and their inner products, from the next chapter, we will dispense with these notational crutches and just write this quantity as $A \cdot B$ or even just AB. In more advanced texts, this is done with no apology. See the panel for "index notation" and additional comments on notation.

The inner product of two 4-vectors is invariant under Lorentz transformations: $\overrightarrow{A'} \star \overrightarrow{B'} = \overrightarrow{A} \star \overrightarrow{B}$. To show this, let's use matrix notation and the metric tensor to rewrite the inner product as

$$\overrightarrow{A} \star \overrightarrow{B} = A^{\mathrm{T}} g B$$

where B is a column matrix or column vector — a 4×1 object. Clearly A^{T} is a row vector. Under a Lorentz transformation Λ, the inner product

A common notation used for the Minkowski inner product of 4-vectors A and B is: $A_\mu B^\mu$, or $A^\mu B_\mu$. This "index notation" will be introduced in a later chapter. For now, it will occassionally be useful to write the inner product as $A_\mu B^\mu$ or $A^\mu B_\mu$. Similarly, the norm of a 4-vector A is $A_\mu A^\mu$ or $A^\mu A_\mu$.
This is more or less universally understood notation for inner products and norms of 4-vectors. However, since we have not explained index notation yet, we will mingle this with the $\star$ and thick-vector notation for now. The $\star$ and thick-vector notation is NOT standard, and is used here as a temporary instrument until you get comfortable with 4-vectors.

becomes

$$\begin{aligned}
\overset{\Longrightarrow}{A'} \star \overset{\Longrightarrow}{B'} &= (A')^{\mathrm{T}} gB' = (\Lambda A)^{\mathrm{T}} g(\Lambda B) = \left(A^{\mathrm{T}}\Lambda^{\mathrm{T}}\right) g\Lambda B \\
&= A^{\mathrm{T}}\left(\Lambda^{\mathrm{T}} g\Lambda\right) B \qquad \begin{cases}\text{using associativity of} \\ \text{matrix multiplication}\end{cases} \\
&= A^{\mathrm{T}} gB \qquad \begin{cases}\text{using the defintion of} \\ \text{Lorentz transformations}\end{cases} \\
&= \overset{\Longrightarrow}{A} \star \overset{\Longrightarrow}{B} .
\end{aligned}$$

It's almost as if the defintion of the inner product was designed to exploit the definition of the Lorentz transformation so that inner products could be invariant under LT's. (In fact, it was so designed.)

For 3-vectors, the magnitude (or **norm**) of a vector is naturally expressed in terms of the inner product: $\left|\vec{A}\right| = \sqrt{\vec{A}\cdot\vec{A}}$. We define the norm of a 4-vector by analogy:

Definition: Norm of a 4-vector

The norm of a 4-vector (A^0, A^1, A^2, A^3) is its inner product with itself: $(A^0)^2 - (A^1)^2 - (A^2)^2 - (A^3)^2$. This is a Lorentz scalar.

We could have tried to define the norm with a square root, as is common for Euclidean vectors. But for 4-vectors, the inner product of a vector with itself can be negative. Taking square roots would therefore be clumsy and is best avoided.

Since the norm of a 4-vector is a Lorentz invariant (scalar), whenever we meet a 4-vector, we can infer a corresponding invariant. In the next section, as we list or construct a number of physical 4-vector quantities, we will construct an invariant for each of them.

2.3.1 The really bad news — Dependence on metric convention

I have some bad news.

Our definitions of the inner product and the metric have assumed the negative-trace metric: the diagonal elements of g are $(1, -1, -1, -1)$. With the other metric convention (positive-trace metric), the signs would be reversed: the inner product would be $-A^0B^0 + A^1B^1 + A^2B^2 + A^3B^3$ and the norm would be $-(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2$.

The inner product has the form $A^{\mathrm{T}}gB$. When the sign of g is flipped, the inner product acquires an overall negative sign. So, whether an inner product (or a norm) is positive or negative depends on the choice of metric.

The really bad news is that this notational ambiguity will continue to torment you as long as you study relativity or any physics based on it. Most but not all texts on quantum field theory use the negative-trace metric; most but not all texts on general relativity use the positive-trace metric. Due to this lack of consensus on convention, the experience of comparing a definition or formula in two different texts can range from mildly annoying to utterly nightmarish. I am so sorry.