FYS 3120/4120 Classical Mechanics and Electromagnetism

Note on Relativity and Covariant formalism

1 The four-dimensional space-time

The new idea of Einstein's Theory of Relativity is that there is no clear, physical distinction between space and time. They form together the continuum (*manifold*) of space-time, and the separation into space and time is not unique, but depends on the inertial frame to which an observer may refer.

In the same way as three-dimensional space is described in a natural way as a three-dimensional *vector space*, space-time is viewed as a four-dimensional vector space. The extension from three-dimensional space to four-dimensional space-time therefore leads to the extension of vectors $\mathbf{r}$ with components (x, y, z) to four-dimensional vectors with components (t, x, y, z), where t is the time coordinate. In order to have the same physical dimension for all four directions in space-time, we introduce, in the standard way, a time coordinate with dimension length, $x^0 = ct$, with c as the speed of light. Note the convention that the coordinates of space-time are written with lifted indices, so that,

$$x^0 = ct\ , \quad x^1 = x\ , \quad x^2 = y\ , \quad x^3 = z\ , \tag{1}$$

To distinguish the 4-vectors of space-time from the 3-vectors of space, we shall mainly use capital (boldface) letters for the 4-vectors. Thus, the position vector of a space time point is

$$\boldsymbol{R} = (ct, x, y, z) \tag{2}$$

and we shall also write it in the form,

$$\boldsymbol{R} = (x^0, \mathbf{r}) \tag{3}$$

where $\mathbf{r}$ is the space-component of the 4-vector $\boldsymbol{R}$.

Note that when we express the 4-vector $\mathbf{R}$ in terms of its time and space components as in (2) or (3), we refer to a specific (inertial) reference frame, since the separation in time and space is dependent of the choice of inertial frame. Although the representation in terms of components is often convenient, it is somewhat sloppy, since usually we consider the vector $\mathbf{R}$ to be independent of the choice of coordinate frame. (This is the same as for the 3-vector $\mathbf{r}$ which we consider as independent of choice of coordinate axes.) A more precise way is to express it in terms of a set of basis vectors,

$$\boldsymbol{R} = \sum_{\mu=0}^{3} x^\mu \mathbf{e}_\mu \tag{4}$$

where $\mathbf{e}_0$ is a unit vector in the time direction and $\mathbf{e}_k, k = 1, 2, 3$ is a set of orthogonal unit vectors in (three-dimensional) space. They can be identified with the standard unit vectors $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ of a Cartesian coordinate frame in 3-space. A change of reference frame will now correspond to a change of basis $\{\mathbf{e}_\mu\}$, and this is compensated for by a transformation of the coordinates x^μ so that the vector $\mathbf{R}$ is left unchanged.

Often it is convenient to consider the space-time components of the position vector $\mathbf{R}$ to form a 4×1 matrix

$$\mathbf{R} = \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} \tag{5}$$

Ideally we should use different symbols for the collection of coordinates (2) (and (3)), for the matrix (5) and the basis-independent vector (4). However, in order to avoid to many symbols we shall not do this, but rather trust that it is easy to see from the context what the symbols mean.

2 Minkowski diagrams

The position vectors $\mathbf{R}$ of space-time form a four dimensional vector space. Often it is useful to make a graphical representation of the space, but since we cannot make a good representation of *all* four dimensions we often make a restriction to the two-dimensional subspace spanned by the coordinates (x^0, x^1) or the three-dimensional subspace spanned by (x^0, x^1, x^2). Such a restricted representation may be sufficient when we consider motion in one or two (space) dimensions. The graphical representations of the subspaces are referred to as *Minkowski diagrams,* and the (flat) four-dimensional space-time of the special theory of relativity is often referred to as *Minkowski space*.

In Figure 1 a two-dimensional and a three-dimensional Minkowski diagram is shown. In the first diagram both the y-direction and the z-directions are suppressed, in the second diagram only the z-direction. The directions of the coordinate axes are defined relative to a given inertial frame. In both diagrams the *forward light cone* and the *backward light cone* are shown. These cones are defined relative to a space-time point (here chosen as the origin O), but they are independent of the choice of inertial frame. The inside of the forward light cone defines the *absolute future* relative to O. A space-time point in the forward light cone is in the diagrams represented by the *time-like* 4-vector $\mathbf{R}_A$. The inside of the backward light cone defines the *absolute past.* The outside of the light cones define the points of *relative simultaneity*. The space-like vector $\mathbf{R}_B$ represents such a point. Whether a point in this region of space-time is in the future, in the past or is simultaneous with $O,$ depends on the choice of reference frame.

3 Summation conventions

When using 4-vector notations some conventions are commonly used, and we shall make use of them also here. For example when a vector index is running over all the four vector indices

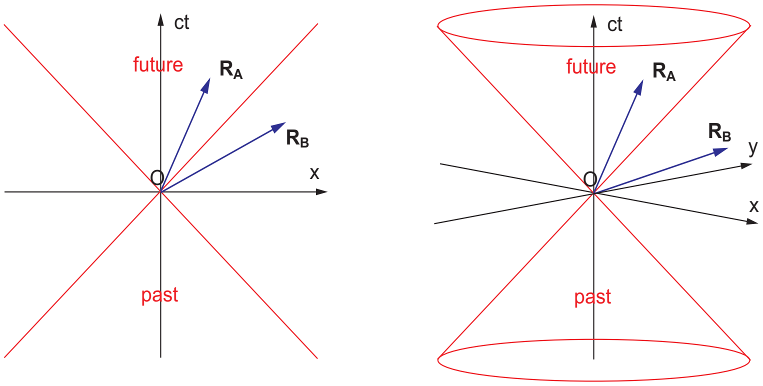

Figure 1: Two-dimensional and three-dimensional Minkowski diagrams. In both diagrams the the location of the light cones relative to the point O are shown. They indicate which space time points are causally connected or causally disconnected with O. The first kind of points is represented by the time-like vector $\mathbf{R}_A$ and the second type by the space-like vector $\mathbf{R}_B$.

of space-time, we label the index with greek letters, while a latin letter is normally indicate a restriction to the three space components. For example when x^μ is written, μ is allowed to take values from 0 to 3. If however we write x_i the index runs instead from 1 to 3.

Another convention we shall follow is Einstein's summing convention. Thus a repeated index in a product normally means that we sum over the index. Thus,

$$x_\mu x^\mu \equiv \sum_{\mu=0}^{3} x_\mu x^\mu \tag{6}$$

and

$$x_i x_i \equiv \sum_{\mu=1}^{3} x_i x_i \tag{7}$$

Again we note that the position of the index, up or down, is important in the 4-vector notation, but for the 3-vectors such a distinction is not important. We shall soon have a look at this distinction. The use of 4-vectors (and tensors) we refer to as covariant notation, and we note as a particular rule that in the *covariant notation* we only sum over pairs of indices, where one is an upper index and the other lower index. This is sometimes referred to as a *contraction*.

4 Metric tensor

In three-dimensional space the *distance* between two neighbouring points

$$ds^2 = d\mathbf{r} \cdot d\mathbf{r} = dx^2 + dy^2 + dy^2 \tag{8}$$

is an *invariant* quantity. It is unchanged when we redefine the Cartesian frame by rotating the coordinate axes. This quantity can be generalized to an invariant “distance” between points in four-dimensional space-time. This *invariant line element* is given by

$$ds^2 = d\mathbf{R} \cdot d\mathbf{R} = dx^2 + dy^2 + dy^2 - c^2 dt^2 \tag{9}$$

As opposed to the three dimensional case, the form of the line element ds^2 in four dimensions does not define a Euclidean geometry , since it is not positive for all vectors $d\mathbf{R}$. For space-time points with *time-like separation* ds^2 is negative (and hence the notation is somewhat misleading) while for *space-like separation* it is positive. For two points that can be connected by a light signal (*light-like separation*) ds^2 vanishes. Invariance of ds^2 means that it is independent of choice of inertial frame, and therefore that it is left unchanged by general Lorentz transformations. This invariance is directly related to the basic postulate of the theory of relativity, that the speed of light is the same in all reference frames.

We may write the generalized distance in the form

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \tag{10}$$

where $g_{\mu\nu}$ is referred to as the *metric tensor* and can be thought of as (the matrix elements of) a 4×4 symmetric matrix. (Note that in (10) Einstein’s summation convention has been used.) This matrix is (in Cartesian coordinates) a diagonal matrix of the form

$$g = (g_{\mu\nu}) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{11}$$

5 Upper and lower indices

We have already stressed the convention that the coordinates of a 4-vector $\mathbf{R}$ are written with upper indices, as x^μ. We have also indicated that coordinates with lowered indices may be defined. The precise definition is,

$$x_\mu = g_{\mu\nu} x^\nu \tag{12}$$

Thus a four-vector can be associated with two sets of coordinates, those with upper indices which are the standard ones (referred to as *contravariant* components) and those with lower indices (referred to as *covariant* components). The metric tensor acts as a lowering operator on the coordinates. This gives a simple relation

$$x_0 = -x^0, x_1 = x^1, x_2 = x^2, x_3 = x^3 \tag{13}$$

Thus, the only change is that the sign of the 0’th component is reversed.

Initially it may seem cumbersome to operate with two sets of coordinates for a 4-vector, which are even so closely related. However, if one is careful to place the indices correctly the relativistic equations can be simplified, and if they are consistently used on both sides of the equations one will gain a guarantee that it keeps the form correctly when changing from one reference frame to another.

We note, as a special case, that the invariant line element can now be written without the metric tensor as

$$ds^2 = dx_\mu dx^\mu \tag{14}$$

More generally, summation over a pair of 4-indices, one lower and one upper will produce an invariant quantity.

The metric tensor acts as a lowering operator on the vector indices. Clearly there must be an inverse to this which acts as a raising operator. We write it as

$$x^\mu = g^{\mu\nu} x_\nu \tag{15}$$

Since it is the inverse to $g_{\mu\nu}$ we have the relation

$$g^{\mu\rho} g_{\rho\nu} = \delta^\mu_\nu \tag{16}$$

Note that the relativistic form of the Kronecker delta is written with one upper and one lower index. This is to have the indices of the two sides of the equation consistently placed.

We note from the matrix form of $g_{\mu\nu}$ that the square of the matrix is identical to the identity matrix. This means that the matrix is its own inverse and therefore $g_{\mu\nu}$ and $g^{\mu\nu}$ represent the same 4×4 matrix. Nevertheless, we insist on writing this matrix with lower indices when it is used as a lowering operator of vector indices in an equation and with upper indices when it is used as a raising operator. This is to be able to place consistently all vector indices in the relativistic equations.

6 Lorentz-transformations in matrix form

We assume that an inertial frame S' is moving with velocity v relative to another inertial frame S. The transition between the two reference frames corresponds to a special Lorentz transformation (a ”boost”) of the coordinates associated with a space time point. With the velocity directed in the x-direction (x^1-direction), the transformation is given by

$$\begin{aligned} x'^0 &= \gamma(x^0 - \beta x^1) \\ x'^1 &= \gamma(x^1 - \beta x^0) \\ x'^2 &= x^2 \\ x'^3 &= x^3 \end{aligned} \tag{17}$$

with $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. This transformation can be written in the form

$$x'^\mu = L^\mu{}_\nu \, x^\nu \tag{18}$$

or as a matrix equation

$$\mathbf{R}' = L\,\mathbf{R} \tag{19}$$

where L is the 4×4 matrix

$$L = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{20}$$

(Note how the indices of L are place, one raised and one lowered, in order to have a consistent expression for the indices of the transformation equation (18).)

A general expression for transformations between inertial frames has the form

$$x'^{\mu} = L^{\mu}{}_{\nu}\, x^{\nu} + a^{\mu} \tag{21}$$

where $L^{\mu}{}_{\nu}$ represents a general Lorentz transformation composed by a boost and a rotation and a^{μ} represents a shift of the origin. Such a general transformation is often referred to as a *Poincaré transformation*. If $L^{\mu}{}_{\nu}$ is to represent a (general) Lorentz transformation it has to satisfy a certain restriction, which follows from the requirement that the velocity of light is left unchanged by the transformation. As already discussed this can be expressed in terms of the invariant line element, so that

$$\begin{aligned} g_{\mu\nu} dx'^{\mu} dx'^{\nu} &= g_{\mu\nu} L^{\mu}{}_{\rho}\, L^{\nu}{}_{\sigma}\, dx^{\rho} dx^{\sigma} \\ &= g_{\mu\nu} dx^{\mu} dx^{\nu} \end{aligned} \tag{22}$$

For this equation to be valid for all dx^{μ}, the $L^{\mu}{}_{\nu}$ has to satisfy the restriction

$$g_{\mu\nu} L^{\mu}{}_{\rho} L^{\nu}{}_{\sigma} = g_{\rho\sigma} \tag{23}$$

In matrix form this can be written as

$$L^T g L = g \tag{24}$$

where L^T represents the transposed matrix. This equation, which determines whether the 4×4 matrix L represents a Lorentz transformation, corresponds to the following condition that rotation matrices $\mathcal{R}$ satisfy in three-dimensional space,

$$\mathcal{R}^T \mathcal{R} = \mathbf{1} \tag{25}$$

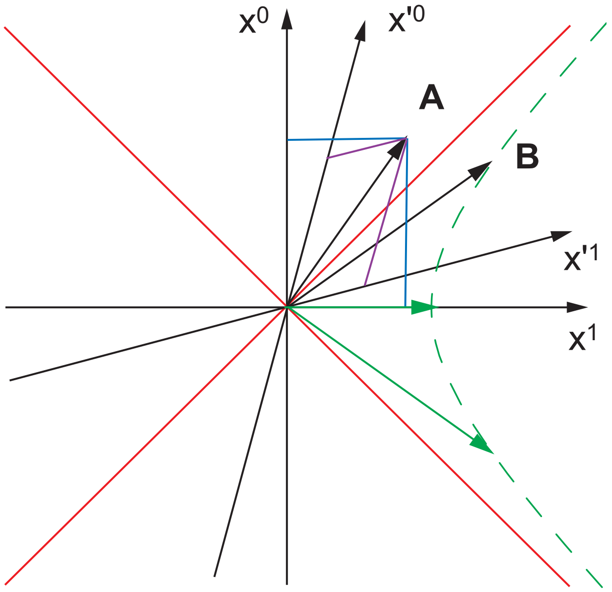

Figure 2: A two-dimensional Minkowski diagram with coordinate axes corresponding to two reference frames in relative motion. The coordinate axes of the un-primed system are perpendicular in the diagram, but not the primed axes. In reality both sets of axes define orthonormal directions in the sense of the relativistic scalar product. The decomposition of the time like vector **A** on both sets of coordinate axes are shown. The space-like vector **B** is orthogonal to **A** even if they are not perpendicular in the diagram. All the space-like vectors with tips at the hyperbolic curve have the same relativistic length, even if the Euclidean lengths in the diagram are quite different.

7 General 4-vectors

So far we have considered 4-vectors as being associated with points in four-dimensional space time. However, exactly as in three dimensions vectors can be more general objects, for example associated with velocity, acceleration, vector fields etc. A general 4-vector **A** is characterized by:

- it has four components $A^\mu, \mu = 0, 1, 2, 3$,
- it transforms as x^μ under Lorentz transformations, $A^\mu \rightarrow A'^\mu = L^\mu{}_\nu A^\nu$.

The Minkowski diagram is convenient to give a geometric representation of general 4-vectors. A reference frame corresponds to a choice of basis vectors $\{\mathbf{e}_\mu, \mu = 0, 1, 2, 3\}$ and a vector **A** can be decomposed on any set of basis vectors, corresponding to different inertial reference frames,

$$\mathbf{A} = A^\mu \mathbf{e}_\mu = A'^\mu \mathbf{e}'_\mu \tag{26}$$

Thus a Lorentz transformation simply corresponds to a change of basis in Minkowski space.

This is illustrated in Figure 3, where the basis vectors $\mathbf{e}_\mu$ are represented as orthogonal vectors, while $\mathbf{e}'_\mu$ are represented by non-orthogonal vectors. Note that there is no intrinsic difference between these two sets of basis vectors, but since the geometry of Minkowski space is non-Euclidean they cannot both be represented as orthogonal in the Minkowski diagram.

The geometry of Minkowski space is determined by the *scalar product*, defined by

$$\mathbf{A} \cdot \mathbf{B} = g_{\mu\nu} A^\mu B^\nu = A^\mu B_\mu \tag{27}$$

This scalar product is not positive definite and separates the vectors in three different classes

$$\begin{aligned} \mathbf{A}^2 &> 0 \quad space - like \\ \mathbf{A}^2 &= 0 \quad light - like \\ \mathbf{A}^2 &< 0 \quad time - like \end{aligned} \tag{28}$$

In the two-dimensional Minkowski diagram these three classes are represented by vectors lying outside the light cone, on the light cone or inside respectively.

Note that the basis vector $\mathbf{e}_0$ is a time-like vector and cannot therefore be normalized to unity. Instead one chooses this vector to be normalized to -1. Thus, the basis vectors associated with an inertial frame satisfy the orthonormality condition

$$\mathbf{e}_\mu \cdot \mathbf{e}_\nu = g_{\mu\nu} \tag{29}$$

They are orthogonal in the sense that the scalar product of two different basis vectors vanishes, but the normalization is different for the time like and the space like basis vectors.

As already noticed, orthogonality in the sense that the scalar product of two 4-vectors vanishes does not mean that they *appear* as orthogonal in the Minkowski diagram. In the diagram orthogonality ($\mathbf{A} \cdot \mathbf{B} = 0$) means that they have directions symmetrically about the light cone, and in particular a light like vector will be orthogonal to itself.

Thus, even if graphical representations in terms of the two- (or three-) dimensional Minkowski diagram is often useful, one has to remember that the geometry is distorted relative to the standard Euclidean geometry. As already discussed, angles may not be faithfully represented, and also distances will generally not coincide with the Euclidean distances. This can be seen by introducing units on the coordinates axes of the reference frames S and S'. We also note that a path that is short, in the sense that the integral of the invariant line element $ds = \sqrt{dx^\mu dx_\mu}$ is small, may appear as a long path in the diagram due to cancellations of contributions from the space and time parts of the vector $d\mathbf{R}$.

8 Lorentz transformation of vector components with lower index

The index of a general 4-vector can be lowered by application of the metric tensor, in the same way as for the position vector x^μ,

$$A_\mu = g_{\mu\nu} A^\nu \tag{30}$$

This relation leads to different transformation properties for vector components with upper indices (often called *contravariant* components) and lower indices (called *covariant* components). We find the following expression for the transformed covariant components

$$\begin{aligned} A'_\mu &= g_{\mu\nu} A'^\nu \\ &= g_{\mu\nu} L^\nu{}_\rho A^\rho \\ &= g_{\mu\nu} L^\nu{}_\rho \, g^{\rho\sigma} \, A_\sigma \\ &\equiv L_\mu{}^\sigma \, A_\sigma \end{aligned} \tag{31}$$

Note in the last line we have introduced a modified symbol for the transformation matrix

$$L_\mu{}^\sigma = g_{\mu\nu} L^\nu{}_\rho \, g^{\rho\sigma} \tag{32}$$

where we have followed the general rule that $g_{\mu\nu}$ acts as a lowering operator and $g^{\mu\nu}$ as a raising operator. With $L^\mu{}_\nu$ as the matrix elements of the 4×4 matrix L, $L_\mu{}^\nu$ then are the matrix elements of the matrix

$$\begin{aligned} \tilde{L} &= gLg \\ &= (L^T)^{-1} \end{aligned} \tag{33}$$

The last expression is derived from the identity (24), which is satisfied by all Lorentz transformation matrices L. We note that the covariant and contravariant components in a sense transform in inverse ways, which is consistent with the fact that the scalar product of two vectors, which can be written as the product upper index components with lower index components, is invariant under Lorentz transformations.

9 Tensors

The 4-vector notation is useful in order to express the relativistic equations in a compact form which applies to all inertial reference frames. However, all physical quantities cannot be written as vectors, but they can be expressed in terms of *tensors*, which are multicomponent objects that transform in ways closely related to that of vectors. A general tensor is chracterized by *a set of space-time indices*. The number of indices is called the *rank* of the tensor.

The 4-vector is a special case, it is a rank 1 tensor. A rank 2 tensor is written as

$$T^{\mu\nu} \,, \quad \mu = 0, 1, 2, 3; \nu = 0, 1, 2, 3. \tag{34}$$

It has all together 16 components. The important property of a tensor is the way it transforms under a change of reference frame. The transformation is determined by the number and position (up or down) of its space time indices. Thus there is one Lorentz transformation matrix for each index, so that the rank 2 tensor transforms as

$$T^{\mu\nu} \to T'^{\mu\nu} = L^\mu{}_\rho L^\nu{}_\sigma T^{\rho\sigma} \tag{35}$$

As an example, we may from two vectors $\mathbf{A}$ and $\mathbf{B}$ easily form a rank 2 tensor

$$C^{\mu\nu} = A^{\mu}B^{\nu} \tag{36}$$

This composition is called the *tensor product* of the two vectors. Another rank 2 tensor that we will meet later in the course is the *electromagnetic field tensor* $F^{\mu\nu}$. This tensor is antisymmetric in μ and ν and is composed by the electric and magnetic field strengths so that $F^{0k}, k = 1, 2, 3$ are the electric components and $F^{kl}, k, l = 1, 2, 3$ are the magnetic components.

Tensors may, like vectors, be written with upper indeces or lower indices. These are related by the action of the metric tensor. For rank 2, we then have three related tensors

$$T^{\mu\nu}, \quad T^{\mu}_{\ \nu} = g_{\nu\rho}T^{\mu\rho}, \quad T_{\mu\nu} = g_{\mu\rho}g_{\nu\sigma}T^{\rho\sigma} \tag{37}$$

With the introduction of tensors we have a series of different, but related relativistic objects at our disposal:

A	rank 0 (scalar)	no vector index	(1 component)
B^{μ}	rank 1 (vector)	one vector index	(4 components)
$C^{\mu\nu}$	rank 2	two vector indices	(16 components)
$D^{\mu\nu\rho}$	rank 3	three vector indices	(64 components)
etc.			

We note that a *contraction*, *i.e.*, a summation of one upper and one lower index will transform a tensor into a new tensor, with rank reduced by 2. For example $A = A^{\mu}{}_{\mu}$ is a scalar, $B^{\mu} = B^{\mu\nu}{}_{\nu}$ is a vector etc.

When the relativistic equations expressed in terms of tensors, they are said to be in *covariant form*. When the equations are written in covariant form they are expressed in terms of variables with simple, standardized transformation properties. One can then easily check that the two sides of the equation transforms in the same way, so that the equation is valid in any reference system. To check that a covariant equation has the correct form we note that

- free indices (that are not summed over) have the same positions (up or down) on both sides of the equation,

- repeated indices that are summed over appear with one in the upper position and one in the lower position.