Introduction to Tensor Calculus

Kees Dullemond & Kasper Peeters

This booklet contains an explanation about tensor calculus for students of physics and engineering with a basic knowledge of linear algebra. The focus lies mainly on acquiring an understanding of the principles and ideas underlying the concept of 'tensor'. We have not pursued mathematical strictness and pureness, but instead emphasise practical use (for a more mathematically pure resumé, please see the bibliography). Although tensors are applied in a very broad range of physics and mathematics, this booklet focuses on the application in special and general relativity.

We are indebted to all people who read earlier versions of this manuscript and gave useful comments, in particular G. Bäuerle (University of Amsterdam) and C. Dullemond Sr. (University of Nijmegen).

The original version of this booklet, in Dutch, appeared on October 28th, 1991. A major update followed on September 26th, 1995. This version is a re-typeset English translation made in 2008/2010.

1
The index notation

Before we start with the main topic of this booklet, tensors, we will first introduce a new notation for vectors and matrices, and their algebraic manipulations: the index notation. It will prove to be much more powerful than the standard vector notation. To clarify this we will translate all well-know vector and matrix manipulations (addition, multiplication and so on) to index notation.

Let us take a manifold (=space) with dimension n. We will denote the components of a vector $\vec{v}$ with the numbers $v_1, \ldots, v_n$. If one modifies the vector basis, in which the components $v_1, \ldots, v_n$ of vector $\vec{v}$ are expressed, then these components will change, too. Such a transformation can be written using a matrix A, of which the columns can be regarded as the old basis vectors $\vec{e}_1, \ldots, \vec{e}_n$ expressed in the new basis $\vec{e}_1{}', \ldots, \vec{e}_n{}'$,

$$\begin{pmatrix} v_1' \\ \vdots \\ v_n' \end{pmatrix} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \tag{1.1}$$

Note that the first index of A denotes the row and the second index the column. In the next chapter we will say more about the transformation of vectors.

According to the rules of matrix multiplication the above equation means:

$$\begin{array}{ccccccccc} v_1' & = & A_{11} \cdot v_1 & + & A_{12} \cdot v_2 & + & \cdots & + & A_{1n} \cdot v_n \,, \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ v_n' & = & A_{n1} \cdot v_1 & + & A_{n2} \cdot v_2 & + & \cdots & + & A_{nn} \cdot v_n \,, \end{array} \tag{1.2}$$

or equivalently,

$$\begin{aligned} v_1' &= \sum_{\nu=1}^{n} A_{1\nu} v_\nu \,, \\ &\vdots \quad \vdots \\ v_n' &= \sum_{\nu=1}^{n} A_{n\nu} v_\nu \,, \end{aligned} \tag{1.3}$$

or even shorter,

$$v_\mu' = \sum_{\nu=1}^{n} A_{\mu\nu}\, v_\nu \qquad (\forall \mu \in \mathbb{N} \quad | \quad 1 \leq \mu \leq n) \,. \tag{1.4}$$

In this formula we have put the essence of matrix multiplication. The index ν is a *dummy index* and μ is a *running index*. The names of these indices, in this case μ and

ν, are chosen arbitrarily. The could equally well have been called α and β:

$$v'_\alpha = \sum_{\beta=1}^{n} A_{\alpha\beta}\, v_\beta \qquad (\forall \alpha \in \mathbb{N} \quad | \quad 1 \leq \alpha \leq n)\,. \tag{1.5}$$

Usually the conditions for μ (in Eq. 1.4) or α (in Eq. 1.5) are not explicitly stated because they are obvious from the context.

The following statements are therefore equivalent:

$$\begin{aligned} \vec{v} \;&=\; \vec{y} \quad \Leftrightarrow \quad v_\mu = y_\mu \quad \Leftrightarrow \quad v_\alpha = y_\alpha\,, \\ \vec{v} \;&=\; A\,\vec{y} \quad \Leftrightarrow \quad v_\mu = \sum_{\nu=1}^{n} A_{\mu\nu} y_\nu \quad \Leftrightarrow \quad v_\nu = \sum_{\mu=1}^{n} A_{\nu\mu} y_\mu\,. \end{aligned} \tag{1.6}$$

This *index notation* is also applicable to other manipulations, for instance the inner product. Take two vectors $\vec{v}$ and $\vec{w}$, then we define the inner product as

$$\vec{v}\cdot\vec{w} := v_1 w_1 + \cdots + v_n w_n = \sum_{\mu=1}^{n} v_\mu w_\mu\,. \tag{1.7}$$

(We will return extensively to the inner product. Here it is just as an example of the power of the index notation). In addition to this type of manipulations, one can also just take the sum of matrices and of vectors:

$$\begin{aligned} C = A + B \quad &\Leftrightarrow \quad C_{\mu\nu} = A_{\mu\nu} + B_{\mu\nu} \\ \vec{z} = \vec{v} + \vec{w} \quad &\Leftrightarrow \quad z_\alpha = v_\alpha + w_\alpha \end{aligned} \tag{1.8}$$

or their difference,

$$\begin{aligned} C = A - B \quad &\Leftrightarrow \quad C_{\mu\nu} = A_{\mu\nu} - B_{\mu\nu} \\ \vec{z} = \vec{v} - \vec{w} \quad &\Leftrightarrow \quad z_\alpha = v_\alpha - w_\alpha \end{aligned} \tag{1.9}$$

▸ ***Exercises 1 to 6 of Section C.1.***

From the exercises it should have become clear that the summation symbols $\sum$ can always be put at the start of the formula and that their order is irrelevant. We can therefore in principle omit these summation symbols, if we make clear in advance over which indices we perform a summation, for instance by putting them after the formula,

$$\begin{aligned} \sum_{\nu=1}^{n} A_{\mu\nu} v_\nu \qquad &\rightarrow \qquad A_{\mu\nu} v_\nu \qquad\quad \{\nu\} \\ \sum_{\beta=1}^{n}\sum_{\gamma=1}^{n} A_{\alpha\beta} B_{\beta\gamma} C_{\gamma\delta} \quad &\rightarrow \quad A_{\alpha\beta} B_{\beta\gamma} C_{\gamma\delta} \qquad \{\beta,\gamma\} \end{aligned} \tag{1.10}$$

From the exercises one can already suspect that

- almost never is a summation performed over an index if that index only appears once in a product,
- almost always a summation is performed over an index that appears twice in a product,
- an index appears almost never more than twice in a product.

When one uses index notation in every day routine, then it will soon become irritating to denote explicitly over which indices the summation is performed. From experience (see above three points) one *knows* over which indices the summations are performed, so one will soon have the idea to introduce the convention that, unless explicitly stated otherwise:

- a summation is assumed over all indices that appear twice in a product, and
- *no* summation is assumed over indices that appear only once.

From now on we will write all our formulae in index notation with this particular convention, which is called the *Einstein summation convection*. For a more detailed look at index notation with the summation convention we refer to [4]. We will thus from now on rewrite

$$\begin{aligned} \sum_{\nu=1}^{n} A_{\mu\nu} v_\nu \qquad &\rightarrow \qquad A_{\mu\nu} v_\nu \,, \\ \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} A_{\alpha\beta} B_{\beta\gamma} C_{\gamma\delta} \quad &\rightarrow \quad A_{\alpha\beta} B_{\beta\gamma} C_{\gamma\delta} \,. \end{aligned} \tag{1.11}$$

► ***Exercises 7 to 10 of Section C.1.***

2

Bases, co- and contravariant vectors

In this chapter we introduce a new kind of vector ('covector'), one that will be essential for the rest of this booklet. To get used to this new concept we will first show in an intuitive way how one can imagine this new kind of vector. After that we will follow a more mathematical approach.

2.1. Intuitive approach

We can map the space around us using a coordinate system. Let us assume that we use a linear coordinate system, so that we can use linear algebra to describe it. Physical objects (represented, for example, with an arrow-vector) can then be described in terms of the basis-vectors belonging to the coordinate system (there are some hidden difficulties here, but we will ignore these for the moment). In this section we will see what happens when we choose another set of basis vectors, i.e. what happens upon a basis transformation.

In a description with coordinates we must be fully aware that the coordinates (i.e. the numbers) themselves have no meaning. Only with the corresponding basis vectors (which span up the coordinate system) do these numbers acquire meaning. It is important to realize that the object one describes is independent of the coordinate system (i.e. set of basis vectors) one chooses. Or in other words: an arrow does not change meaning when described an another coordinate system.

Let us write down such a basis transformation,

$$\begin{aligned} \vec{e}_1{}' &= a_{11}\,\vec{e}_1 + a_{12}\,\vec{e}_2\,, \\ \vec{e}_2{}' &= a_{21}\,\vec{e}_1 + a_{22}\,\vec{e}_2\,. \end{aligned} \tag{2.1}$$

This could be regarded as a kind of multiplication of a 'vector' with a matrix, as long as we take for the components of this 'vector' the basis vectors. If we describe the matrix elements with words, one would get something like:

$$\begin{pmatrix} \vec{e}_1{}' \\ \vec{e}_2{}' \end{pmatrix} = \begin{pmatrix} \text{projection of } \vec{e}_1{}' \text{ onto } \vec{e}_1 & \text{projection of } \vec{e}_1{}' \text{ onto } \vec{e}_2 \\ \text{projection of } \vec{e}_2{}' \text{ onto } \vec{e}_1 & \text{projection of } \vec{e}_2{}' \text{ onto } \vec{e}_2 \end{pmatrix} \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \end{pmatrix}. \tag{2.2}$$

Note that the basis vector-colums $\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$ are *not* vectors, but just a very useful way to write things down.

We can also look at what happens with the components of a *vector* if we use a different set of basis vectors. From linear algebra we know that the transformation

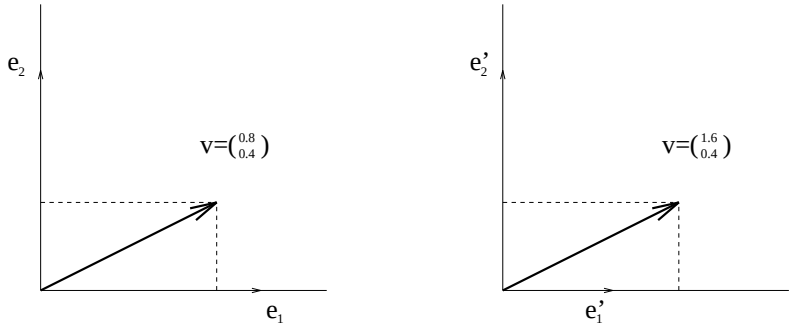

Figure 2.1: The behaviour of the transformation of the components of a vector under the transformation of a basis vector $\vec{e}_1{}' = \frac{1}{2}\vec{e}_1 \rightarrow v_1{}' = 2v_1$.

matrix can be constructed by putting the old basis vectors *expressed in the new basis* in the columns of the matrix. In words,

$$\begin{pmatrix} v_1{}' \\ v_2{}' \end{pmatrix} = \begin{pmatrix} \text{projection of } \vec{e}_1 \text{ onto } \vec{e}_1{}' & \text{projection of } \vec{e}_2 \text{ onto } \vec{e}_1{}' \\ \text{projection of } \vec{e}_1 \text{ onto } \vec{e}_2{}' & \text{projection of } \vec{e}_2 \text{ onto } \vec{e}_2{}' \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} . \tag{2.3}$$

It is clear that the matrices of Eq. (2.2) and Eq. (2.3) are *not* the same.

We now want to compare the basis-transformation matrix of Eq. (2.2) with the coordinate-transformation matrix of Eq. (2.3). To do this we replace all the primed elements in the matrix of Eq. (2.3) by non-primed elements and vice-versa. Comparison with the matrix in Eq. (2.2) shows that we also have to transpose the matrix. So if we call the matrix of Eq. (2.2) Λ, then Eq. (2.3) is equivalent to:

$$\vec{v}' = (\Lambda^{-1})^T \vec{v} . \tag{2.4}$$

The normal vectors are called 'contravariant vectors', because they transform contrary to the basis vector columns. That there must be a different behavior is also intuitively clear: if we described an 'arrow' by coordinates, and we then modify the basis vectors, then the coordinates must clearly change in the opposite way to make sure that the coordinates times the basis vectors produce the same physical 'arrow' (see Fig. 2.1).

In view of these two opposite transformation properties, we could now attempt to construct objects that, contrary to normal vectors, transform the same as the basis vector columns. In the simple case in which, for example, the basis vector $\vec{e}_1{}'$ transforms into $\frac{1}{2} \times \vec{e}_1$, the coordinate of this object must then also $\frac{1}{2}$ times as large. This is precisely what happens to the coordinates of a gradient of a scalar function! The reason is that such a gradient is the difference of the function per unit distance in the direction of the basis vector. When this 'unit' suddenly shrinks (i.e. if the basis vector shrinks) this means that the gradient must shrink too (see Fig. 2.1 for a one-dimensional example). A 'gradient', which we so far have always regarded as a true vector, will from now on be called a 'covariant vector' or 'covector': it transforms in the same way as the basis vector columns.

The fact that gradients have usually been treated as ordinary vectors is that if the coordinate transformation transforms one cartesian coordinate system into the other (or in other words: one orthonormal basis into the other), then the matrices Λ en $(\Lambda^{-1})^T$ are the same.

▸ ***Exercises 1, 2 of Section C.2.***

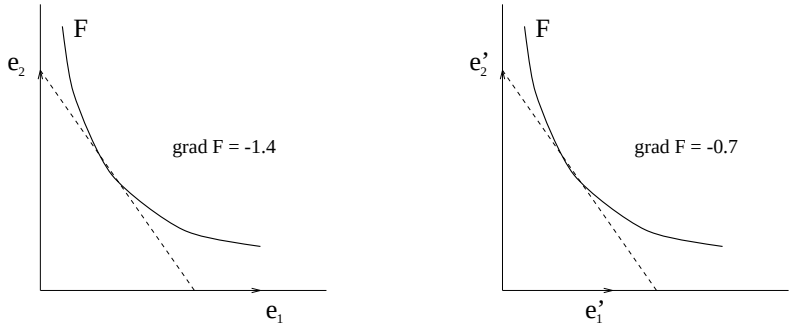

Figure 2.2: Basis vector $\vec{e}_1{}' = \frac{1}{2}\vec{e}_1 \rightarrow \vec{\nabla} f' = \frac{1}{2}\vec{\nabla} f$

As long as one transforms only between orthonormal basis, there is no difference between contravariant vectors and covariant vectors. However, it is not always possible in practice to restrict oneself to such bases. When doing vector mathematics in curved coordinate systems (like polar coordinates for example), one is often forced to use non-orthonormal bases. And in special relativity one is fundamentally forced to distinguish between co- and contravariant vectors.

2.2. Mathematical approach

Now that we have a notion for the difference between the transformation of a vector and the transformation of a gradient, let us have a more mathematical look at this.

Consider an n-dimensional manifold with coordinates $x_1, x_2, ..., x_n$. We define the gradient of a function $f(x_1, x_2, ..., x_n)$,

$$(\nabla f)_\mu := \frac{\partial f}{\partial x_\mu}\,. \tag{2.5}$$

The difference in transformation will now be demonstrated using the simplest of transformations: a homogeneous linear transformation (we did this in the previous section already, since we described all transformations with matrices). In general a coordinate transformation can also be non-homogeneous linear, (e.g. a translation), but we will not go into this here.

Suppose we have a vector field defined on this manifold $V : \vec{v} = \vec{v}(\vec{x})$. Let us perform a homogeneous linear transformation of the coordinates:

$$x'_\mu = A_{\mu\nu} x_\nu\,. \tag{2.6}$$

In this case not only the coordinates x_μ change (and therefore the dependence of $\vec{v}$ on the coordinates), but also the components of the vectors,

$$v'_\mu(\vec{x}) = A_{\mu\nu}\, v_\nu(\vec{x})\,, \tag{2.7}$$

where A is the same matrix as in Eq. (2.6) (this may look trivial, but it is useful to check it! Also note that we take as transformation matrix the matrix that describes the transformation of the vector components, whereas in the previous section we took for Λ the matrix that describes the transformation of the basis vectors. So A is equal to $(\Lambda^{-1})^T$).

Now take the function $f(x_1, x_2, ..., x_n)$ and the gradient w_α at a point P in the following way,

$$w_\alpha = \frac{\partial f}{\partial x_\alpha}\,, \tag{2.8}$$

and in the new coordinate system as

$$w'_\alpha = \frac{\partial f}{\partial x'_\alpha}\,. \tag{2.9}$$

It now follows (using the chain rule) that

$$\frac{\partial f}{\partial x'_1} = \frac{\partial f}{\partial x_1}\frac{\partial x_1}{\partial x'_1} + \frac{\partial f}{\partial x_2}\frac{\partial x_2}{\partial x'_1} + \ldots + \frac{\partial f}{\partial x_n}\frac{\partial x_n}{\partial x'_1}\,. \tag{2.10}$$

That is,

$$\frac{\partial f}{\partial x'_\mu} = \left(\frac{\partial x_\nu}{\partial x'_\mu}\right)\frac{\partial f}{\partial x_\nu}\,, \tag{2.11}$$

$$w'_\mu = \left(\frac{\partial x_\nu}{\partial x'_\mu}\right) w_\nu\,. \tag{2.12}$$

This describes how a gradient transforms. One can regard the partial derivative $\frac{\partial x_\nu}{\partial x'_\mu}$ as the matrix $(A^{-1})^T$ where A is defined as in Eq. (2.6). To see this we first take the inverse of Eq. (2.6):

$$x_\mu = (A^{-1})_{\mu\nu} x'_\nu\,. \tag{2.13}$$

Now take the derivative,

$$\frac{\partial x_\mu}{\partial x'_\alpha} = \frac{\partial((A^{-1})_{\mu\nu} x'_\nu)}{\partial x'_\alpha} = (A^{-1})_{\mu\nu}\frac{\partial x'_\nu}{\partial x'_\alpha} + \frac{\partial (A^{-1})_{\mu\nu}}{\partial x'_\alpha} x'_\nu\,. \tag{2.14}$$

Because in this case A does not depend on x'_α the last term on the right-hand side vanishes. Moreover, we have that

$$\frac{\partial x'_\nu}{\partial x'_\alpha} = \delta_{\nu\alpha}\,, \qquad \delta_{\nu\alpha} = \begin{cases} 1 & \text{when } \nu = \alpha, \\ 0 & \text{when } \nu \neq \alpha. \end{cases} \tag{2.15}$$

Therefore, what remains is

$$\frac{\partial x_\mu}{\partial x'_\alpha} = (A^{-1})_{\mu\nu}\delta_{\nu\alpha} = (A^{-1})_{\mu\alpha}\,. \tag{2.16}$$

With Eq. (2.12) this yields for the transformation of a gradient

$$w'_\mu = (A^{-1})^T_{\mu\nu} w_\nu\,. \tag{2.17}$$

The indices are now in the correct position to put this in matrix form,

$$w' = (A^{-1})^T w\,. \tag{2.18}$$

(We again note that the matrix A used here denotes the coordinate transformation from the coordinates x to the coordinates x').

We have shown here what is the difference in the transformation properties of normal vectors ('arrows') and gradients. Normal vectors we call from now on *contravariant vectors* (though we usually simply call them *vectors*) and gradients we call *covariant vectors* (or *covectors* or *one-forms*). It should be noted that not every covariant vector field can be constructed as a gradient of a scalar function. A gradient has the property that $\nabla \times (\nabla f) = 0$, while not all covector fields may have this property. To distinguish vectors from covectors we will denote vectors with an arrow ($\vec{v}$) and covectors with a tilde ($\tilde{w}$). To make further distinction between contravariant and covariant vectors we will put the contravariant indices (i.e. the indices of

contravariant vectors) as *superscript* and the covariant indices (i.e. the indices of covariant vectors) with *subscripts*,

y^α	: contravariant vector
w_α	: covariant vector, or covector

In practice it will turn out to be very useful to also introduce this convention for matrices. Without further argumentation (this will be given later) we note that the matrix A can be written as:

$$A: \quad A^\mu{}_\nu \,. \tag{2.19}$$

The transposed version of this matrix is:

$$A^T: \quad A_\nu{}^\mu \,. \tag{2.20}$$

With this convention the transformation rules for vectors resp. covectors become

$$v'^\mu = A^\mu{}_\nu v^\nu \,, \tag{2.21}$$

$$w'_\mu = (A^{-1})^T{}_\mu{}^\nu w_\nu = (A^{-1})^\nu{}_\mu w_\nu \,. \tag{2.22}$$

The delta δ of Eq. (2.15) also gets a matrix form,

$$\delta_{\mu\nu} \quad \rightarrow \quad \delta^\mu{}_\nu \,. \tag{2.23}$$

This is called the 'Kronecker delta'. It simply has the 'function' of 'renaming' an index:

$$\delta^\mu{}_\nu y^\nu = y^\mu \,. \tag{2.24}$$

3
Introduction to tensors

Tensor calculus is a technique that can be regarded as a follow-up on linear algebra. It is a generalisation of classical linear algebra. In classical linear algebra one deals with vectors and matrices. Tensors are generalisations of vectors and matrices, as we will see in this chapter.

In section 3.1 we will see in an example how a tensor can naturally arise. In section 3.2 we will re-analyse the essential step of section 3.1, to get a better understanding.

3.1. The new inner product and the first tensor

The inner product is very important in physics. Let us consider an example. In classical mechanics it is true that the 'work' that is done when an object is moved equals the inner product of the force acting on the object and the displacement vector $\vec{x}$,

$$W = \langle \vec{F}, \vec{x} \rangle \,. \tag{3.1}$$

The work W must of course be independent of the coordinate system in which the vectors $\vec{F}$ and $\vec{x}$ are expressed. The inner product as we know it,

$$s = \langle \vec{a}, \vec{b} \rangle = a^\mu b^\mu \quad \text{(old definition)} \tag{3.2}$$

does not have this property in general,

$$s' = \langle \vec{a}', \vec{b}' \rangle = A^\mu{}_\alpha a^\alpha A^\mu{}_\beta b^\beta = (A^T)^\mu{}_\alpha A^\mu{}_\beta a^\alpha b^\beta \,, \tag{3.3}$$

where A is the transformation matrix. Only if A^{-1} equals A^T (i.e. if we are dealing with *orthonormal* transformations) s will not change. The matrices will then together form the kronecker delta $\delta_{\beta\alpha}$. It appears as if the inner product only describes the physics correctly in a special kind of coordinate system: a system which according to our human perception is 'rectangular', and has physical units, i.e. a distance of 1 in coordinate x_1 means indeed 1 meter in x_1-direction. An orthonormal transformation produces again such a rectangular 'physical' coordinate system. If one has so far always employed such special coordinates anyway, this inner product has always worked properly.

However, as we already explained in the previous chapter, it is not always guaranteed that one can use such special coordinate systems (polar coordinates are an example in which the *local* orthonormal basis of vectors is not the coordinate basis).

The inner product between a vector x and a covector y, however, *is* invariant under all transformations,

$$s = x^\mu y_\mu \,, \tag{3.4}$$

because for all A one can write

$$s' = x'^\mu y'_\mu = A^\mu{}_\alpha x^\alpha (A^{-1})^\beta{}_\mu y_\beta = (A^{-1})^\beta{}_\mu A^\mu{}_\alpha x^\alpha y_\beta = \delta^\beta{}_\alpha x^\alpha y_\beta = s \tag{3.5}$$

With help of this inner produce we can introduce a new inner product between two contravariant vectors which also has this invariance property. To do this we introduce a covector w_μ and define the inner product between x^μ and y^ν *with respect to* this covector w_μ in the following way (we will introduce a better definition later):

$$s = w_\mu w_\nu x^\mu y^\nu \quad \text{(first attempt)} \tag{3.6}$$

(**Warning:** later it will become clear that this definition is not quite useful, but at least it will bring us on the right track toward finding an invariant inner product between two contravariant vectors). The inner product s will now obviously transform correctly, because it is made out of two invariant ones,

$$\begin{aligned} s' &= (A^{-1})^\mu{}_\alpha w_\mu (A^{-1})^\nu{}_\beta w_\nu A^\alpha{}_\rho x^\rho A^\beta{}_\sigma y^\sigma \\ &= (A^{-1})^\mu{}_\alpha A^\alpha{}_\rho (A^{-1})^\nu{}_\beta A^\beta{}_\sigma w_\mu w_\nu x^\rho y^\sigma \\ &= \delta^\mu{}_\rho \delta^\nu{}_\sigma w_\mu w_\nu x^\rho y^\sigma \\ &= w_\mu w_\nu x^\mu y^\nu \\ &= s \,. \end{aligned} \tag{3.7}$$

We have now produced an invariant 'inner product' for contravariant vectors by using a covariant vector w_μ as a measure of length. However, this covector appears twice in the formula. One can also rearrange these factors in the following way,

$$s = (w_\mu w_\nu) x^\mu y^\nu = \begin{pmatrix} x^1 & x^2 & x^3 \end{pmatrix} \begin{pmatrix} w_1 \cdot w_1 & w_1 \cdot w_2 & w_1 \cdot w_3 \\ w_2 \cdot w_1 & w_2 \cdot w_2 & w_2 \cdot w_3 \\ w_3 \cdot w_1 & w_3 \cdot w_2 & w_3 \cdot w_3 \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} . \tag{3.8}$$

In this way the two appearances of the covector w are combined into *one* object: some kind of product of w with itself. It is some kind of matrix, since it is a collection of numbers labeled with indices μ and ν. Let us call this object g,

$$g = \begin{pmatrix} w_1 \cdot w_1 & w_1 \cdot w_2 & w_1 \cdot w_3 \\ w_2 \cdot w_1 & w_2 \cdot w_2 & w_2 \cdot w_3 \\ w_3 \cdot w_1 & w_3 \cdot w_2 & w_3 \cdot w_3 \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} . \tag{3.9}$$

Instead of using a covector w_μ in relation to which we define the inner product, we can also directly define the object g: that is more direct. So, we define the inner product *with respect to* the object g as:

$$\boxed{s = g_{\mu\nu} x^\mu y^\nu} \quad \text{new definition} \tag{3.10}$$

Now we must make sure that the object g is chosen such that our new inner product reproduces the old one *if* we choose an orthonormal coordinate system. So, with Eq. (3.8) in an orthonormal system one should have

$$\begin{aligned} s &= g_{\mu\nu} x^\mu y^\nu \\ &= \begin{pmatrix} x^1 & x^2 & x^3 \end{pmatrix} \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \begin{pmatrix} y^1 \\ y^2 \\ y^3 \end{pmatrix} \\ &= x^1 y^1 + x^2 y^2 + x^3 y^3 \qquad \text{in an orthonormal system!} \end{aligned} \tag{3.11}$$

To achieve this, g must become, in an orthonormal system, something like a unit matrix:

$$g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{in an orthonormal system!} \tag{3.12}$$

One can see that one cannot produce this set of numbers according to Eq. (3.9). This means that the definition of the inner product according to Eq. (3.6) has to be rejected (hence the warning that was written there). Instead we have to start directly from Eq. (3.10). We do no longer regard g as built out of two covectors, but regard it as a matrix-like set of numbers on itself.

However, it does not have the transformation properties of a classical matrix. Remember that the matrix A of the previous chapter had one index up and one index down: $A^{\mu}{}_{\nu}$, indicating that it has mixed contra- and co-variant transformation properties. The new object $g_{\mu\nu}$, however, has both indices down: it transforms in both indices in a covariant way, like the $w_\mu w_\nu$ which we initially used for our inner product. This curious object, which looks like a matrix, but does not transform as one, is an example of a *tensor*. A matrix is also a tensor, as are vectors and covectors. Matrices, vectors and covectors are special cases of the more general class of objects called 'tensors'. The object $g_{\mu\nu}$ is a kind of tensor that is neither a matrix nor a vector or covector. It is a new kind of object for which only tensor mathematics has a proper description.

The object g is called a *metric*, and we will study its properties later in more detail: in Chapter 5.

With this last step we now have a complete description of the new inner product between contravariant vectors that behaves properly, in that it remains invariant under any linear transformation, and that it reproduces the old inner product when we work in an orthonormal basis. So, returning to our original problem,

$$W = \langle \vec{F}, \vec{x} \rangle = g_{\mu\nu} F^\mu x^\nu \qquad \text{general formula}\,. \tag{3.13}$$

In this section we have in fact put forward two new concepts: the new inner product and the concept of a 'tensor'. We will cover both concepts more deeply: the tensors in Section 3.2 and Chapter 4, and the inner product in Chapter 5.

3.2. Creating tensors from vectors

In the previous section we have seen how one can produce a tensor out of two covectors. In this section we will revisit this procedure from a slightly different angle.

Let us look at products between matrices and vectors, like we did in the previous chapters. One starts with an object with two indices and therefore n^2 components (the matrix) and an object with one index and therefore n components (the vector). Together they have $n^2 + n$ componenten. After multiplication one is left with one object with one index and n components (a vector). One has therefore lost $(n^2 + n) - n = n^2$ numbers: they are 'summed away'. A similar story holds for the inner product between a covector and a vector. One starts with $2n$ components and ends up with one number,

$$s = x^\mu y_\mu = x^1 y_1 + x^2 y_2 + x^3 y_3\,. \tag{3.14}$$

Summation therefore reduces the number of components.

In standard multiplication procedures from classical linear algebra such a summation usually takes place: for matrix multiplications as well as for inner products. In index notation this is denoted with paired indices using the summation convention. However, the index notation also allows us to multiply vectors and covectors

without pairing up the indices, and therefore without summation. The object one thus obtains does not have fewer components, but *more*:

$$s^{\mu}{}_{\nu} := x^{\mu} y_{\nu} = \begin{pmatrix} x^1 \cdot y_1 & x^1 \cdot y_2 & x^1 \cdot y_3 \\ x^2 \cdot y_1 & x^2 \cdot y_2 & x^2 \cdot y_3 \\ x^3 \cdot y_1 & x^3 \cdot y_2 & x^3 \cdot y_3 \end{pmatrix} . \tag{3.15}$$

We now did not produce one number (as we would have if we replaced ν with μ in the above formula) but instead an ordered set of numbers labelled with the indices μ and ν. So if we take the example $\vec{x} = (1,3,5)$ and $\tilde{y} = (4,7,6)$, then the tensor-components of $s^{\mu}{}_{\nu}$ are, for example: $s^2{}_3 = x^2 \cdot y_3 = 3 \cdot 6 = 18$ and $s^1{}_1 = x^1 \cdot y_1 = 1 \cdot 4 = 4$, and so on.

This is the kind of 'tensor' object that this booklet is about. However, this object still looks very much like a matrix, since a matrix is also nothing more or less than a set of numbers labeled with two indices. To check if this is a true matrix or something else, we need to see how it transforms,

$$s'^{\alpha}{}_{\beta} = x'^{\alpha} y'_{\beta} = A^{\alpha}{}_{\mu} x^{\mu} (A^{-1})^{\nu}{}_{\beta} y_{\nu} = A^{\alpha}{}_{\mu} (A^{-1})^{\nu}{}_{\beta} (x^{\mu} y_{\nu}) = A^{\alpha}{}_{\mu} (A^{-1})^{\nu}{}_{\beta} s^{\mu}{}_{\nu} . \tag{3.16}$$

▸ ***Exercise 2 of Section C.3.***

If we compare the transformation in Eq.(3.16) with that of a true matrix of exercise 2 we see that the tensor we constructed is indeed an ordinary matrix. But if instead we use two covectors,

$$t_{\mu\nu} = x_{\mu} y_{\nu} = \begin{pmatrix} x_1 \cdot y_1 & x_1 \cdot y_2 & x_1 \cdot y_3 \\ x_2 \cdot y_1 & x_2 \cdot y_2 & x_2 \cdot y_3 \\ x_3 \cdot y_1 & x_3 \cdot y_2 & x_3 \cdot y_3 \end{pmatrix} , \tag{3.17}$$

then we get a tensor with different transformation properties,

$$\begin{aligned} t'_{\alpha\beta} = x'_{\alpha} y'_{\beta} &= (A^{-1})^{\mu}{}_{\alpha} x_{\mu} (A^{-1})^{\nu}{}_{\beta} y_{\nu} \\ &= (A^{-1})^{\mu}{}_{\alpha} (A^{-1})^{\nu}{}_{\beta} (x_{\mu} y_{\nu}) \\ &= (A^{-1})^{\mu}{}_{\alpha} (A^{-1})^{\nu}{}_{\beta} t_{\mu\nu} . \end{aligned} \tag{3.18}$$

The difference with $s^{\mu}{}_{\nu}$ lies in the first matrix of the transformation equation. For s it is the transformation matrix for contravariant vectors, while for t it is the transformation matrix for covariant vectors. The tensor t is clearly *not* a matrix, so we indeed created something new here. The g tensor of the previous section is of the same type as t.

The beauty of tensors is that they can have an arbitrary number of indices. One can also produce, for instance, a tensor with 3 indices,

$$A_{\alpha\beta\gamma} = x_{\alpha} y_{\beta} z_{\gamma} . \tag{3.19}$$

This is an ordered set of numbes labeled with three indices. It can be visualized as a kind of 'super-matrix' in 3 dimensions (see Fig. 3.2).

These are tensors of *rank 3*, as opposed to tensors of rank 0 (scalars), rank 1 (vectors and covectors) and rank 2 (matrices and the other kind of tensors we introduced so far). We can distinguish between the contravariant rank and covariant rank. Clearly $A_{\alpha\beta\gamma}$ is a tensor of covariant rank 3 and contravariant rank 0. Its total rank is 3. One can also produce tensors of, for instance, contravariant rank 2

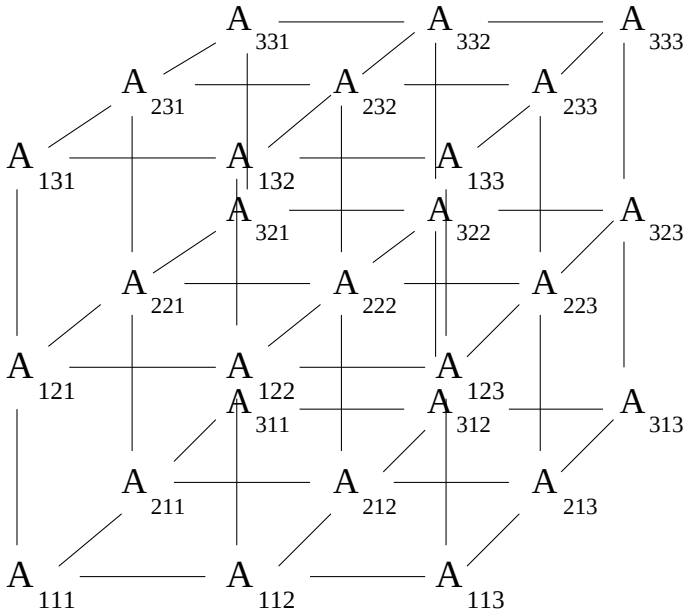

Figure 3.1: A tensor of rank 3.

and covariant rank 3 (i.e. total rank 5): $B^{\alpha\beta}{}_{\mu\nu\phi}$. A similar tensor, $C^{\alpha}{}_{\mu\nu\phi}{}^{\beta}$, is also of contravariant rank 2 and covariant rank 3. Typically, when tensor mathematics is applied, the meaning of each index has been defined beforehand: the first index means this, the second means that etc. As long as this is well-defined, then one can have co- and contra-variant indices in any order. However, since it usually *looks better* (though this is a matter of taste) to have the contravariant indices first and the covariant ones last, usually the meaning of the indices is chosen in such a way that this is accomodated. This is just a matter of how one chooses to assign meaning to each of the indices.

Again it must be clear that although a multiplication (without summation!) of m vectors and m covectors produces a tensor of rank $m + n$, *not every tensor of rank* $m + n$ *can be constructed as such a product*. Tensors are much more general than these simple products of vectors and covectors. It is therefore important to step away from this picture of combining vectors and covectors into a tensor, and consider this construction as nothing more than a simple example.

4

Tensors, definitions and properties

Now that we have a first idea of what tensors are, it is time for a more formal description of these objects.

We begin with a formal definition of a tensor in Section 4.1. Then, Sections 4.2 and 4.3 give some important mathematical properties of tensors. Section 4.4 gives an alternative, in the literature often used, notation of tensors. And finally, in Section 4.5, we take a somewhat different view, considering tensors as operators.

4.1. Definition of a tensor

'Definition' of a tensor: An (N, M)-tensor at a given point in space can be described by a set of numbers with $N + M$ indices which transforms, upon coordinate transformation given by the matrix A, in the following way:

$$\boxed{t'^{\alpha_1 \ldots \alpha_N}{}_{\beta_1 \ldots \beta_M} = A^{\alpha_1}{}_{\mu_1} \ldots A^{\alpha_N}{}_{\mu_N} (A^{-1})^{\nu_1}{}_{\beta_1} \ldots (A^{-1})^{\nu_M}{}_{\beta_M} t^{\mu_1 \ldots \mu_N}{}_{\nu_1 \ldots \nu_M}} \tag{4.1}$$

An (N, M)-tensor in a three-dimensional manifold therefore has $3^{(N+M)}$ components. It is contravariant in N components and covariant in M components. Tensors of the type

$$t_\alpha{}^\beta{}_\gamma \tag{4.2}$$

are of course not excluded, as they can be constructed from the above kind of tensors by rearrangement of indices (like the transposition of a matrix as in Eq. 2.20).

Matrices (2 indices), vectors and covectors (1 index) and scalars (0 indices) are therefore also tensors, where the latter transforms as $s' = s$.

4.2. Symmetry and antisymmetry

In practice it often happens that tensors display a certain amount of symmetry, like what we know from matrices. Such symmetries have a strong effect on the properties of these tensors. Often many of these properties or even tensor equations can be derived solely on the basis of these symmetries.

A tensor t is called *symmetric* in the indices μ and ν if the components are equal upon exchange of the index-values. So, for a 2nd rank contravariant tensor,

$$t^{\mu\nu} = t^{\nu\mu} \qquad \text{symmetric (2,0)-tensor}\,. \tag{4.3}$$

A tensor t is called *anti-symmetric* in the indices μ and ν if the components are equal-but-opposite upon exchange of the index-values. So, for a 2nd rank contravariant tensor,

$$t^{\mu\nu} = -t^{\nu\mu} \qquad \text{anti-symmetric (2,0)-tensor}\,. \tag{4.4}$$

It is not useful to speak of symmetry or anti-symmetry in a pair of indices that are not of the same type (co- or contravariant). The properties of symmetry only remain invariant upon basis transformation if the indices are of the same type.

▸ ***Exercises 2 and 3 of Section C.4.***

4.3. Contraction of indices

With tensors of at least one covariant and at least one contravariant index we can define a kind of 'internal inner product'. In the simplest case this goes as,

$$t^{\alpha}{}_{\alpha}\,, \tag{4.5}$$

where, as usual, we sum over α. This is the trace of the matrix $t^{\alpha}{}_{\beta}$. Also this construction is invariant under basis transformations. In classical linear algebra this property of the trace of a matrix was also known.

▸ ***Exercise 3 of Section C.4.***

One can also perform such a *contraction of indices* with tensors of higher rank, but then some uncontracted indices remain, e.g.

$$t^{\alpha\beta}{}_{\alpha} = v^{\beta}\,, \tag{4.6}$$

In this way we can convert a tensor of type (N, M) into a tensor of type $(N-1, M-1)$. Of course, this contraction procedure causes information to get lost, since after the contraction we have fewer components.

Note that contraction can only take place between one contravariant index and one covariant index. A contraction like $t^{\alpha\alpha}$ is not invariant under coordinate transformation, and we should therefore reject such constructions. In fact, the summations of the summation convention *only* happen over a covariant and a contravariant index, be it a contraction of a tensor (like $t^{\mu\alpha}{}_{\alpha}$) or an inner product between two tensors (like $t^{\mu\alpha}y_{\alpha\nu}$).

4.4. Tensors as geometrical objects

Vectors can be seen as columns of numbers, but also as arrows. One can perform calculations with these arrows if one regards them as linear combinations of 'basis arrows',

$$\vec{v} = \sum_{\mu} v^{\mu}\vec{e}_{\mu}\,. \tag{4.7}$$

We would like to do something like this also with covectors and eventually of course with all tensors. For covectors it amounts to constructing a set of 'unit covectors' that serve as a basis for the covectors. We write

$$\tilde{w} = \sum_{\mu} w_{\mu}\tilde{e}^{\mu}\,. \tag{4.8}$$

Note: we denote the basis vectors and basis covectors with indices, but we do not mean the components of these basis (co-)vectors (after all: in which basis would that

be?). Instead we label the basis (co-)vectors themselves with these indices. This way of writing was already introduced in Section 2.1. In spite of the fact that these indices have therefore a slightly different meaning, we still use the usual conventions of summation for these indices. Note that we mark the co-variant basis vectors with an *upper* index and the contra-variant basis-vectors with a *lower* index. This may sound counter-intuitive ('did we not decide to use upper indices for contra-variant vectors?') but this is precisely what we mean with the 'different meaning of the indices' here: this time they *label* the vectors and do not denote their components.

The next step is to express the basis of covectors in the basis of vectors. To do this, let us remember that the inner product between a vector and a covector is always invariant under transformations, independent of the chosen basis:

$$\vec{v} \cdot \tilde{w} = v^{\alpha} w_{\alpha} \tag{4.9}$$

We now substitute Eq. (4.7) and Eq. (4.8) into the left hand side of the equation,

$$\vec{v} \cdot \tilde{w} = v^{\alpha} \vec{e}_{\alpha} \cdot w_{\beta} \tilde{e}^{\beta} \,. \tag{4.10}$$

If Eq. (4.9) and Eq. (4.10) have to remain consistent, it automatically follows that

$$\tilde{e}^{\beta} \cdot \vec{e}_{\alpha} = \delta^{\beta}{}_{\alpha} \,. \tag{4.11}$$

On the left hand side one has the inner product between a covector $\tilde{e}^{\beta}$ and a vector $\vec{e}_{\alpha}$. This time the inner product is written in abstract form (i.e. not written out in components like we did before). The indices α and β simply denote which of the basis covectors (β) is contracted with which of the basis vectors (α). A geometrical representation of vectors, covectors and the geometric meaning of their inner product is given in appendix B.

Eq. (4.11) is the (implicit) definition of the co-basis in terms of the basis. This equation does not give the co-basis explicitly, but it defines it nevertheless uniquely. This co-basis is called the 'dual basis'. By arranging the basis covectors in columns as in Section 2.1, one can show that they transform as

$$\tilde{e}'^{\alpha} = A^{\alpha}{}_{\beta} \tilde{e}^{\beta} \,, \tag{4.12}$$

when we choose A (that was equal to $(\Lambda^{-1})^{T}$) in such a way that

$$\vec{e}'_{\alpha} = ((A^{-1})^{T})_{\alpha}{}^{\beta} \vec{e}_{\beta} \,. \tag{4.13}$$

In other words: basis-covector-columns transform as vectors, in contrast to basis-vector-columns, which transform as covectors.

The description of vectors and covectors as geometric objects is now complete. We can now also express tensors in such an abstract way (again here we refer to the mathematical description; a geometric graphical depiction of tensors is given in appendix B). We then express these tensors in terms of 'basis tensors'. These can be constructed from the basis vectors and basis covectors we have constructed above,

$$t = t^{\mu\nu}{}_{\rho} \vec{e}_{\mu} \otimes \vec{e}_{\nu} \otimes \tilde{e}^{\rho} \,. \tag{4.14}$$

The operator $\otimes$ is the 'tensor outer product'. Combining tensors (in this case the basis (co-)vectors) with such an outer product means that the rank of the resulting tensor is the sum of the ranks of the combined tensors. It is the geometric (abstract) version of the 'outer product' we introduced in Section 3.2 to create tensors. The operator $\otimes$ is not commutative,

$$\vec{a} \otimes \vec{b} \neq \vec{b} \otimes \vec{a} \,. \tag{4.15}$$

4.5. Tensors as operators

Let us revisit the new inner product with v a vector and w a covector,

$$w_\alpha v^\alpha = s\,. \tag{4.16}$$

If we drop index notation and we use the usual symbols we have,

$$\tilde{w}\vec{v} = s\,, \tag{4.17}$$

This could also be written as

$$\tilde{w}(\vec{v}) = s\,, \tag{4.18}$$

where the brackets mean that the covector $\tilde{w}$ *acts* on the vector $\vec{v}$. In this way, $\tilde{w}$ is an operator (a function) which 'eats' a vector and produces a scalar. Therefore, written in the usual way to denote maps from one set to another,

$$\tilde{w}: \quad \mathbb{R}^n \to \mathbb{R}\,. \tag{4.19}$$

A covector is then called a linear 'function of direction': the result of the operation (i.e. the resulting scalar) is linearly dependent on the input (the vector), which is a directional object. We can also regard it the other way around,

$$\vec{v}(\tilde{w}) = s\,. \tag{4.20}$$

where the vector is now the operator and the covector the argument. To prevent confusion we write Eq. (4.19) from now on in a somewhat different way,

$$\vec{v}: \quad \mathbb{R}^{*3} \to \mathbb{R}\,. \tag{4.21}$$

An arbitrary tensor of rank (N, M) can also be regarded in this way. The tensor has as input a tensor of rank (M, N), or the product of more than one tensors of lower rank, such that the number of contravariant indices equals M and the number of covariant indices equals N. After contraction of all indices we obtain a scalar. Example (for a $(2,1)$-tensor),

$$t^{\alpha\beta}{}_\gamma a_\alpha b_\beta c^\gamma = s\,. \tag{4.22}$$

The tensor t can be regarded as a function of 2 covectors (a an b) and 1 vector (c), which produces a real number (scalar). The function is linearly dependent on its input, so that we can call this a 'multilinear function of direction'. This somewhat complex nomenclature for tensors is mainly found in older literature.

Of course the fact that tensors can be regarded to some degree as functions or operators does not mean that it is always useful to do so. In most applications tensors are considered as objects themselves, and not as functions.

5

The metric tensor and the new inner product

In this chapter we will go deeper into the topic of the new inner product. The new inner product, and the metric tensor g associated with it, is of great importance to nearly all applications of tensor mathematics in non-cartesian coordinate systems and/or curved manifolds. It allows us to produce mathematical and physical formulae that are invariant under coordinate transformations of any kind.

In Section 5.1 we will give an outline of the role that is played by the metric tensor g in geometry. Section 5.2 covers the properties of this metric tensor. Finally, Section 5.3 describes the procedure of raising or lowering an index using the metric tensor.

5.1. The metric as a measuring rod

If one wants to describe a physical system, then one typically wants to do this with numbers, because they are exact and one can put them on paper. To be able to do this one must first define a coordinate system in which the measurements can be done. One could in principle take the coordinate system in any way one likes, but often one prefers to use an orthonormal coordinate system because this is particularly practical for measuring distances using the law of Pythagoras.

However, often it is not quite practical to use such cartesian coordinates. For systems with (nearly) cylindrical symmetry or (nearly) spherical symmetry, for instance, it is much more practical to use cylindrical or polar coordinates, even though one *could* use cartesian ones. In some applications it is even impossible to use orthonormal coordinates, for instance in the case of mapping of the surface of the Earth. In a typical projection of the Earth with meridians vertically and parallels horizontally, the northern countries are stretched enormously and the north/south pole is a line instead of a point. Measuring distances (or the length of a vector) on such a map with Pythagoras would produce wrong results.

What one has to do to get an impression of 'true' distances, sizes and proportions is to draw on various places on the Earth-map a 'unit circle' that represents, say, 100 km in each direction (if we wish to measure distances in units of 100 km). At the equator these are presumably circles, but as one goes toward the pole they tend to become ellipses: they are circles that are stretched in horizonal (east-west) direction. At the north pole such a 'unit circle' will be stretched infinitely in horizontal direction. There is a strong relation between this unit circle and the metric g. In appendix B one can see that indeed a metric can be graphically represented with

a unit circle (in 2-D) or a unit sphere (in 3-D).

So how does one measure the distance between two points on such a curved manifold? In the example of the Earth's projection one sees that g is different at different lattitudes. So if points A and B have different lattitudes, there is no unique g that we can use to measure the distance. Indeed, there is no objective distance that one can give. What what can do is to define a path from point A to point B and measure the distance along this path in little steps. At each step the distance travelled is so small that one can take the g at the middle of the step and get a reasonably good estimate of the length ds of that step. This is given by

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \qquad \text{with} \qquad g_{\mu\nu} = g_{\mu\nu}(\vec{x})\,. \tag{5.1}$$

By integrating ds all the way from A to B one thus obtains the distance *along this path*. Perhaps the 'true' distance is then the distance along the path which yields the shortest distance between these points.

What we have done here is to chop the path into such small bits that the curvature of the Earth is negligible. Within each step g changes so little that one can use the linear expressions for length and distance using the local metric g.

Let us take a bit more concrete example and measure distances in 3-D using polar coordinates. From geometric considerations we know that the length $\mathrm{d}l$ of an infinitesimally small vector at position (r, θ, ϕ) in polar coordinates satisfies

$$\mathrm{d}l^2 = s = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\phi^2\,. \tag{5.2}$$

We could also write this as

$$\mathrm{d}l^2 = s = \begin{pmatrix}\mathrm{d}x^1 & \mathrm{d}x^2 & \mathrm{d}x^3\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & r^2 & 0\\ 0 & 0 & r^2\sin^2\theta\end{pmatrix}\begin{pmatrix}\mathrm{d}x^1\\ \mathrm{d}x^2\\ \mathrm{d}x^3\end{pmatrix} = g_{\mu\nu}\mathrm{d}x^\mu\mathrm{d}x^\nu\,. \tag{5.3}$$

All the information about the concept of 'length' or 'distance' is hidden in the metric tensor g. The metric tensor field $g(\vec{x})$ is usually called the 'metric of the manifold'.

5.2. Properties of the metric tensor

The metric tensor g has some important properties. First of all, it is symmetric. This can be seen in various ways. One of them is that one can always find a coordinate system in which, locally, the metric tensor has the shape of a unit 'matrix', which is symmetric. Symmetry of a tensor is conserved upon any transformation (see Section 4.2). Another way to see it is that the inner product of two vectors should not depend on the order of these vectors: $g_{\mu\nu}v^\mu w^\nu = g_{\mu\nu}v^\nu w^\mu \equiv g_{\nu\mu}v^\mu w^\nu$ (in the last step we renamed the running indices).

Every symmetric second rank covariant tensor can be transformed into diagonal form in which the diagonal elements are either 1,0 or -1. For the metric tensor they will then all be 1: in an orthonormal coordinate system. In fact, an orthonormal coordinate system is *defined* as the coordinate system in which $g_{\mu\nu} = \mathrm{diag}(1,1,1)$. A normal metric can always be put into this form. However, mathematically one could also conceive a manifold that has a metric that can be brought into the following form: $g_{\mu\nu} = \mathrm{diag}(1,1,-1)$. This produces funny results for some vectors. For instance, the vector $(0,0,1)$ would have an imaginary length, and the vector $(0,1,1)$ has length zero. These are rather pathological properties and appear to be only interesting to mathematicians. A real metric, after all, should always produce positive definite lengths for vectors unequal to the zero-vector. However, as one can see in appendix A, in special relativity a metric will be introduced that can be brought to

the form $\mathrm{diag}(-1,1,1,1)$. For now, however, we will assume that the metric is positive definite (i.e. can be brought to the form $g_{\mu\nu} = \mathrm{diag}(1,1,1)$ through a coordinate transformation).

It should be noted that a metric of signature $g_{\mu\nu} = \mathrm{diag}(1,1,-1)$ can *never* be brought into the form $g_{\mu\nu} = \mathrm{diag}(1,1,1)$ by coordinate transformation, and vice versa. This signature of the metric is therefore an invariant. Normal metrics have signature $\mathrm{diag}(1,1,1)$. This does not mean that they *are* $g_{\mu\nu} = \mathrm{diag}(1,1,1)$, but they can be brought into this form with a suitable coordinate transformation.

▸ ***Exercise 1 of Section C.5.***

So what about the inner product between two *co*-vectors? Just as with the vectors we use a kind of metric tensor, but this time of *contravariant* nature: $g^{\mu\nu}$. The inner product then becomes

$$\boxed{s = g^{\mu\nu} x_\mu y_\nu}\,. \tag{5.4}$$

The distinction between these two metric tensors is simply made by upper resp. lower indices. The properties we have derived for $g_{\mu\nu}$ of course also hold for $g^{\mu\nu}$. And both metrics are intimitely related: once $g_{\mu\nu}$ is given, $g^{\mu\nu}$ can be constructed and vice versa. In a sense they are two realisations of the same metric object g. The easiest way to find $g^{\mu\nu}$ from $g_{\mu\nu}$ is to find a coordinate system in which $g_{\mu\nu} = \mathrm{diag}(1,1,1)$: then $g^{\mu\nu}$ is also equal to $\mathrm{diag}(1,1,1)$. But in the next section we will derive a more elegant relation between the two forms of the metric tensor.

5.3. Co versus contra

Before we got introduced to co-vectors we presumably always used orthonormal coordinate systems. In such coordinate systems we never encountered the difference between co- and contra-variant vectors. A gradient was simply a vector, like any kind of vector. An example is the electric field. One can regard an electric field as a gradient,

$$\tilde{E} = -\nabla V \quad \Leftrightarrow \quad E_\mu = -\frac{\partial V}{\partial x^\mu}\,. \tag{5.5}$$

On the other hand it is clearly also an arrow-like object, related to a force,

$$\vec{E} = \frac{1}{q}\vec{F} = \frac{1}{q}m\vec{a} \quad \Leftrightarrow \quad E^\mu = \frac{1}{q}F^\mu = \frac{1}{q}ma^\mu\,. \tag{5.6}$$

In an orthonormal system one can interchange $\vec{R}$ and $\tilde{R}$ (or equivalently E^μ and E_μ) without punishment. But as soon as we go to a non-orthonormal system, we will encounter problems. Then one is suddenly forced to distinguish between the two.

So the question becomes: if one has a potential field V, how does one obtain the *contravariant* $\vec{E}$? Perhaps the most obvious solution is: first produce the covariant tensor $\tilde{E}$, then transform to an orthonormal system, then switch from co-vector form to contravariant vector form (in this system their components are identical, at least for a positive definite metric) and then transform back to the original coordinate system. This is possible, but very time-consuming and unwieldy. A better method is to use the metric tensor:

$$E^\mu = g^{\mu\nu} E_\nu\,. \tag{5.7}$$

This is a proper tensor product, as the contraction takes place over an upper and a lower index. The object E^μ, which was created from E_ν is now contravariant. One can see that this vector must indeed be the contravariant version of E_ν by taking an orthonormal coordinate system: here $g_{\mu\nu} = \mathrm{diag}(1,1,1)$, i.e. $g_{\mu\nu}$ is some kind of unit

'matrix'. This is the same as saying that in the orthonormal basis $E^\mu = E_\nu$, which is indeed true (again, only for a metric that can be diagonalised to $\mathrm{diag}(1,1,1)$, which for classical physics is always the case). However, in contrast to the formula $E^\mu = E_\nu$, the formula Eq. (5.7) is now also valid after transformation to any coordinate system,

$$\begin{aligned}(A^{-1})^\mu{}_\alpha E_\mu &= (A^{-1})^\mu{}_\alpha (A^{-1})^\nu{}_\beta g_{\mu\nu} A^\beta{}_\sigma E^\sigma \\ &= (A^{-1})^\mu{}_\alpha \delta^\nu{}_\sigma g_{\mu\nu} E^\sigma \\ &= (A^{-1})^\mu{}_\alpha g_{\mu\nu} E^\nu\,,\end{aligned} \tag{5.8}$$

(here we introduced σ to avoid having four ν symbols, which would conflict with the summation convention). If we multiply on both sides with $A^\alpha{}_\rho$ we obtain exactly (the index ρ can be of course replaced by μ) Eq. (5.7).

One can of course also do the reverse,

$$E_\mu = g_{\mu\nu} E^\nu\,. \tag{5.9}$$

If we first lower an index with $g_{\nu\rho}$, and then raise it again with $g^{\mu\nu}$, then we must arrive back with the original vector. This results in a relation between $g_{\mu\nu}$ en $g^{\mu\nu}$,

$$E^\mu = g^{\mu\nu} E_\nu = g^{\mu\nu} g_{\nu\rho} E^\rho = \delta^\mu{}_\rho E^\rho = E^\mu \quad \Rightarrow \quad g^{\alpha\nu} g_{\nu\beta} = \delta^\alpha{}_\beta\,. \tag{5.10}$$

With this relation we have now defined the contravariant version of the metric tensor in terms of its covariant version.

We call the conversion of a contravariant index into a covariant one "lowering an index", while the reverse is "raising an index". The vector and its covariant version are each others "dual".

6
Tensor calculus

Now that we are familiar with the concept of 'tensor' we need to go a bit deeper in the calculus that one can do with these objects. This is not entirely trivial. The index notation makes it possible to write all kinds of manipulations in an easy way, but makes it also all too easy to write down expressions that have no meaning or have bad properties. We will need to investigate where these pitfalls may arise.

We will also start thinking about differentiation of tensors. Also here the rules of index notation are rather logical, but also here there are problems looming. These problems have to do with the fact that ordinary derivatives of tensors do no longer transform as tensors. This is a problem that has far-reaching consequences and we will deal with them in Chapter 7. For now we will merely make a mental note of this.

At the end of this chapter we will introduce another special kind of tensor, the Levi-Civita tensor, which allows us to write down cross-products in index notation. Without this tensor the cross-product cannot be expressed in index notation.

6.1. The 'covariance' of equations

In chapter 3 we saw that the old inner product was not a good definition. It said that the work equals the (old) inner product between $\vec{F}$ en $\vec{x}$. However, while the work is a physical notion, and thereby invariant under coordinate transformation, the inner product was not invariant. In other words: the equation was only valid in a prefered coordinate system. After the introduction of the new inner product the equation suddenly kept its validity in any coordinate system. We call this 'universal validity' of the equation: 'covariance. Be careful not to confuse this with 'covariant vector': it is a rather unpractical fate that the word 'covariant' has two meanings in tensor calculus: that of the type of index (or vector) on the one hand and the 'universal validity' of expressions on the other hand.

The following expressions and equations are therefore covariant,

$$x_\mu = y_\mu \,, \tag{6.1}$$

$$x^\mu = y^\mu \,, \tag{6.2}$$

because both sides transform in the same way. The following expression is *not* covariant:

$$x_\mu = y^\mu \,, \tag{6.3}$$

because the left-hand-side transforms as a covector (i.e. with $(A^{-1})^T$) and the right-hand-side as a contravariant vector (i.e. with A). If this equation holds *by chance* in

one coordinate system, then it will likely *not* hold in another coordinate system. Since we are only interested in vectors and covectors are *geometric objects* and not in their components in some arbitrary coordinate system, we must conclude that covectors cannot be set equal to contravariant vectors.

The same as above we can say about the following equation,

$$t_{\mu\nu} = s^{\mu}{}_{\nu}\,. \tag{6.4}$$

This expression is also not covariant. The tensors on both sides of the equal sign are of different type.

If we choose to stay in orthonormal coordinates then we can drop the distinction between co- and contra-variant vectors and we do not need to check if our expressions or equations are covariant. In that case we put all indices as subscript, to avoid giving the impression that we distinguish between co- and contravariant vectors.

6.2. Addition of tensors

Two tensors can only be added if they have the same rank: one cannot add a vector to a matrix. Addition of two equal-rank tensors is only possible if they are of the same type. So

$$v^{\mu} + w_{\mu} \tag{6.5}$$

is not a tensor. (The fact that the index μ occurs double here does not mean that they need to be summed over, since according to the summation convention one only sums over equal indices occuring in a product, not in a sum. See chapter 1). If one adds two equal type tensors but with unequal indices one gets a meaningless expression,

$$v_{\mu} + w_{\nu}\,. \tag{6.6}$$

If tensors are of the same type, one can add them. One should then *choose* the indices to be the same,

$$x^{\mu\nu} + y^{\mu\nu}\,, \tag{6.7}$$

or

$$x^{\alpha\beta} + y^{\alpha\beta}\,. \tag{6.8}$$

Both the above expressions are *exactly* the same. We just chose μ and ν as the indices in the first expression and α and β in the second. This is merely a matter of choice.

Now what about the following expression:

$$t^{\alpha}{}_{\beta}{}^{\gamma} + r^{\alpha\gamma}{}_{\beta}\,. \tag{6.9}$$

Both tensors are of same rank, have the same number of upper and lower indices, but their order is different. This is a rather peculiar contruction, since typically (as mentioned before) one assigns meaning to each index before one starts working with tensors. Evidently the meaning assigned to index 2 is different in the two cases, since in the case of $t^{\alpha}{}_{\beta}{}^{\gamma}$ it is a covariant index while in the case of $r^{\alpha\gamma}{}_{\beta}$ it is a contravariant index. But this expression, though highly confusing (and therefore not recommendable) is not formally wrong. One can see this if one assumes that $r^{\alpha\gamma}{}_{\beta} = u^{\alpha}v^{\gamma}w_{\beta}$. It is formally true that $u^{\alpha}v^{\gamma}w_{\beta} = u^{\alpha}w_{\beta}v^{\gamma}$, because multiplication is commutative. The problem is, that a tensor has only meaning for the user if one can, beforehand, assign meaning to each of the indices (i.e. "first index means this, second index means that..."). This will be confused if one mixes the indices up. Fortunately, in most applications of tensors the meaning of the indices will be clear from the context. Often symmetries of the indices will help in avoiding confusion.

6.3. Tensor products

The most important property of the product between two tensors is:

> *The result of a product between tensors is again a tensor if in each summation the summation takes place over one upper index and one lower index.*

According to this rule, the following objects are therefore not tensors, and we will forbid them from now on:

$$x^\mu T^{\mu\nu}\,, \qquad h^\alpha{}_\beta{}^\gamma t^{\alpha\beta}{}_\gamma\,. \tag{6.10}$$

The following products *are* tensors,

$$x_\mu T^{\mu\nu} \qquad h^\alpha{}_\beta{}^\gamma t_\alpha{}^\beta{}_\gamma\,. \tag{6.11}$$

and the following expression is a covariant expression:

$$t^{\mu\nu} = a^{\mu\rho}{}_\sigma b^{\nu\sigma}{}_\rho\,. \tag{6.12}$$

If we nevertheless want to sum over two covariant indices or two contravariant indices, then this is possible with the use of the metric tensor. For instance: an inner product between $t^{\mu\nu}$ en v^ν could look like:

$$w^\mu = g_{\alpha\beta} t^{\mu\alpha} v^\beta\,. \tag{6.13}$$

The metric tensor is then used as a kind of 'glue' between two indices over which one could normally not sum because they are of the same kind.

6.4. First order derivatives: non-covariant version

Up to now we have mainly concerned ourselves with the properties of individual tensors. In differential geometry one usually uses *tensor fields*, where the tensor depends on the location given by $\vec{x}$. One can now take the derivative of such tensor fields and these derivatives can also be denoted with index notation. We will see in Chapter 7 that in curved coordinates or in curved manifolds, these derivatives will not behave as tensors and are not very physically meaningful. But for non-curved coordinates on a flat manifold this problem does not arise. In this section we will therefore, temporarily, avoid this problem by assuming non-curved coordinates on a flat manifold.

We start with the gradient of a scalar field, which we already know,

$$(\mathrm{grad}\, f)_\mu = \frac{\partial f}{\partial x^\mu}\,. \tag{6.14}$$

In a similar fashion we can now define the 'gradient' of a vector field:

$$(\mathrm{grad}\, \vec{v})^\mu{}_\nu = \frac{\partial v^\mu}{\partial x^\nu}\,, \tag{6.15}$$

or of a tensor-field,

$$(\mathrm{grad}\, t)_{\mu\nu\alpha} = \frac{\partial t_{\mu\nu}}{\partial x^\alpha}\,. \tag{6.16}$$

As long as we employ only linear transformations, the above objects are all tensors.

Now let us introduce an often used notation,

$$v^\mu{}_{,\nu} := \partial_\nu v^\mu := \frac{\partial v^\mu}{\partial x^\nu}\,. \tag{6.17}$$

For completeness we also introduce

$$v^{\mu,\nu} := \partial^\nu v^\mu := g^{\nu\rho}\frac{\partial v^\mu}{\partial x^\rho}\,. \tag{6.18}$$

With index notation we can also define the divergence of a vector field (again all under the assumption of non-curved coordinates),

$$\nabla\cdot\vec{v} = \frac{\partial v^\rho}{\partial x^\rho} = \partial_\rho v^\rho = v^\rho{}_{,\rho}\,. \tag{6.19}$$

Note, just to avoid confusion: the symbol ∇ will be used in the next chapter for a more sophisticated kind of derivative. We can also define the divergence of a tensor,

$$\frac{\partial T^{\alpha\beta}}{\partial x^\beta} = \partial_\beta T^{\alpha\beta} = T^{\alpha\beta}{}_{,\beta}\,. \tag{6.20}$$

6.5. Rot, cross-products and the permutation symbol

A useful set of numbers, used for instance in electrodynamics, is the permutation symbol ϵ. It is a set of numbers with three indices, if we work in a 3-dimensional space, or n indices if we work in n-dimensional space.

The object ϵ does not transform entirely like a tensor (though almost), and it is therefore not considered a true tensor. In fact it is a *tensor density* (or *pseudo tensor*), which is a near cousin of the tensor family. Tensor densities transform as tensors, but are additionally multiplied by a power of the Jacobian determinant of the transformation matrix. However, to avoid these complexities, we will in this section assume an orthonormal basis, and forget for the moment about covariance.

The ϵ symbol is completely anti-symmetric,

$$\epsilon_{ijk} = -\epsilon_{jik} = -\epsilon_{ikj} = -\epsilon_{kji}\,. \tag{6.21}$$

If two of the three indices have then same value, then the above equation clearly yields zero. The only elements of ϵ that are non-zero are those for which none of the indices is equal to another. We usually define

$$\epsilon_{123} = 1\,, \tag{6.22}$$

and with Eq. (6.21) all the other non-zero elements follow. Any permutation of indices yields a -1. As a summary,

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } ijk \text{ is an even permutation of 123,} \\ -1 & \text{if } ijk \text{ is an odd permutation of 123,} \\ 0 & \text{if two indices are equal.} \end{cases} \tag{6.23}$$

This pseudo-tensor is often called the Levi-Civita pseudo-tensor.

The contraction between two epsilons yields a useful identity,

$$\epsilon_{\alpha\mu\nu}\epsilon_{\alpha\rho\sigma} = \delta_{\mu\rho}\delta_{\nu\sigma} - \delta_{\mu\sigma}\delta_{\nu\rho}\,. \tag{6.24}$$

From this it follows that

$$\epsilon_{\alpha\beta\nu}\epsilon_{\alpha\beta\sigma} = 2\delta_{\nu\sigma}\,, \tag{6.25}$$

and also

$$\epsilon_{\alpha\beta\gamma}\epsilon_{\alpha\beta\gamma} = 6\,. \tag{6.26}$$

With the epsilon object we can express the cross-product between two vectors $\vec{a}$ and $\vec{b}$,

$$\vec{c} = \vec{a} \times \vec{b} \quad \rightarrow \quad c_i = \epsilon_{ijk} a_j b_k \,. \tag{6.27}$$

For the first component of c one therefore has

$$c_1 = \epsilon_{1jk} a_j b_k = \epsilon_{123} a_2 b_3 + \epsilon_{132} a_3 b_2 = a_2 b_3 - b_2 a_3 \,, \tag{6.28}$$

which is indeed the familiar first component of the cross-product between a and b.

This notation has advantages if we, for instance, want to compute the divergence of a rotation. When we translate to index notation this becomes

$$c = \nabla \cdot (\nabla \times \vec{a}) = \nabla_i (\epsilon_{ijk} \nabla_j a_k) = \epsilon_{ijk} \partial_i \partial_j a_k \,, \tag{6.29}$$

because

$$\nabla = \begin{pmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \frac{\partial}{\partial x_3} \end{pmatrix} \quad \text{and therefore} \quad \nabla_i = \frac{\partial}{\partial x_i} := \partial_i \,. \tag{6.30}$$

On the right-hand-side of Eq. (6.29) one sees the complete contraction between a set that is symmetric in the indices i and j ($\partial_i \partial_j$) and a set that is antisymmetric in the indices i and j (ϵ_{ijk}). From the exercise below we will see that such a contraction always yields zero, so that we have proven that a divergence of a rotation is zero.

▸ ***Exercise 4 of Section C.4.***

One can also define another kind of rotation: the *generalised rotation* of a covector,

$$(\mathrm{rot}\, \vec{w})_{\alpha\beta} = \partial_\alpha w_\beta - \partial_\beta w_\alpha \,. \tag{6.31}$$

This transforms as a tensor (in non-curved coordinates/space) and is anti-symmetric.

7
Covariant derivatives

One of the (many) reasons why tensor calculus is so useful is that it allows a proper description of physically meaningful derivatives of vectors. We have not mentioned this before, but taking a derivative of a vector (or of any tensor for that matter) is a non-trivial thing if one wants to formulate it in a covariant way. As usual, in a rectangular orthonormal coordinate system this is usually not an issue. But when the coordinates are curved (like in polar coordinates, for example) then this becomes a big issue. It is even more of an issue when the manifold is curved (like in general relativity), but this is beyond the scope of this booklet – even though the techniques we will discuss in this chapter will be equally well applicable for curved manifolds.

We want to construct a derivative of a vector in such a way that it makes physical sense even in the case of curved coordinates. One may ask: why is there a problem? We will show this by a simple example in Section **??**. In this section we will also show how one defines vectors in curved coordinates by using a local set of basis vectors derived from the coordinate system. In Section **??** we will introduce without proof the mathematical formula for a covariant derivative, and introduce thereby the so-called Christoffel symbol. In Section **??** we will return to the polar coordinates to demonstrate how this machinery works in practice.

This chapter will, however, take quite large steps. The reason is that the concepts of covariant derivatives are usually described in detail in introductory literature on general relativity and Riemannian geometry. In principle this is well beyond the scope of this booklet. However, since tensors find much of their application in problems in which differential operators are used, we wish to at least briefly address the concept of covariant derivative. For further details we refer to ********** **LITERATURE LIST** **********

7.1. Vectors in curved coordinates

Defining vectors in curvilinear coordinates is not trivial. Since the coordinates are curved (maybe even the manifold itself is curved, like the surface of the Earth), there is no meaningful global set of basis vectors in which one can express the vectors. If a vector or tensor field is given, then at each location this vector or tensor could be decomposed into the *local* basis (co-)vectors. This basis may be different from location to location, but we always assume that the changes are smooth.

In principle there are two useful set of basis vectors one can choose:

1. *A local orthonormal set of basis vectors*: If the coordinates are such that they are locally perpendicular with respect to each other, then it is possible to choose

the basis vectors such that they point along these coordinates (though not necessarily are normalized the same as the coordinate units).

2. *A local coordinate basis*: The covector basis can be seen as unit steps in each of the directions: $\tilde{e}^\mu = dx^\mu$. The contravariant basis vectors are then $\vec{e}_\mu = \partial/\partial x^\mu$. Typically this coordinate basis is not orthonormal.

So far in the entire booklet we have implicitly assumed a local coordinate basis (remember that we used the words coordinate transformation on equal footing as basis transformation). For tensor mathematics this basis is usually the easiest to use. But in realistic applications (like in Chapter **??**) a local orthonormal basis has more physical meaning. In the application of Chapter **??** we will first convert everything to the coordinate basis in order to do the tensor math, and then, once the final answer is there, we will convert back to the local orthonormal basis, which is physically more intuitive. But since the tensor mathematics works best for the local coordinate basis, we will assume this basis for the rest of this chapter.

Now let us take the example of circular coordinates in two dimensions. We can express the usual x- and y-coordinates in terms of r and ϕ as: $x = r\cos\phi$ and $y = r\sin\phi$. In r-direction the coordinate basis vector $\partial/\partial r$ can act as a normal basis vector. In the ϕ-direction, however, the coordinate basis vector $\partial/\partial\phi$ has a different length at different radii. However, $\frac{1}{r}\partial/\partial\phi$ is again a normalized (unit) vector. The set of basis vectors $(\partial/\partial r, \partial/\partial\phi)$ is the local coordinate basis, while the set of basis vectors $(\partial/\partial r, \frac{1}{r}\partial/\partial\phi)$ is a local orthonormal basis, which happens to be parallel to the coordinates.

Now let us use the coordinate basis. Let us take a vector field that, at every location, when expressed in this local basis, takes the form $(1, 0)$. This is shown in Fig. **XXXX**. Expressed in cartesian coordinates (for which a global basis is possible) this would be a vector field $(x/\sqrt{x^2+y^2}, y/\sqrt{x^2+y^2})$. Clearly this vector field is not constant. In cartesian coordinates one would write, for instance,

$$\frac{\partial}{\partial x}\begin{pmatrix} x/\sqrt{x^2+y^2} \\ y/\sqrt{x^2+y^2}\end{pmatrix} = \begin{pmatrix} y^2/(x^2+y^2)^{3/2} \\ -xy/(x^2+y^2)^{3/2}\end{pmatrix}, \tag{7.1}$$

which is clearly non-zero. However, if we switch to circular coordinates, and take the derivative with respect to, for instance, ϕ, then we obtain zero,

$$\frac{\partial}{\partial\phi}\begin{pmatrix} 1 \\ 0\end{pmatrix} = \begin{pmatrix} 0 \\ 0\end{pmatrix}. \tag{7.2}$$

In the circular coordinate system it *looks* as if the vector does not change in space, but in reality it does. This is the problem one encounters with derivatives of tensors in curved coordinates. Going to the local orthonormal basis does not help. One has to define a new, 'covariant' form of the derivative.

7.2. The covariant derivative of a vector/tensor field

In this section we will introduce the covariant derivative of a vector field. We will do so *without proof*. Again, for a more detailed discussion we refer to the literature. We define ∂_μ to be the ordinary coordinate-derivative $\partial/\partial x^\mu$, and ∇_μ to be the new *covariant derivative*. The covariant derivative of a vector field v^μ is defined as

$$\nabla_\mu v^\alpha = \partial_\mu v^\alpha + \Gamma^\alpha_{\mu\nu} v^\nu, \tag{7.3}$$

where $\Gamma^\alpha_{\mu\nu}$ is an object called the *Christoffel symbol*. The Christoffel symbol is *not* a tensor because it contains all the information about the curvature of the coordinate system and can therefore be transformed entirely to zero if the coordinates are

straightened. Nevertheless we treat it as any ordinary tensor in terms of the index notation.

The Christoffel symbol can be computed from the metric $g_{\mu\nu}$ (and its companion $g^{\alpha\beta}$) in the following way,

$$\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2} g^{\alpha\beta} \left(\frac{\partial g_{\beta\nu}}{\partial x^{\mu}} + \frac{\partial g_{\beta\mu}}{\partial x^{\nu}} - \frac{\partial g_{\mu\nu}}{\partial x^{\beta}} \right) \equiv \frac{1}{2} g^{\alpha\beta} (g_{\beta\nu,\mu} + g_{\beta\mu,\nu} - g_{\mu\nu,\beta}) \,. \tag{7.4}$$

The covariant derivative of a covector can also be defined with this symbol,

$$\nabla_{\mu} w_{\alpha} = \partial_{\mu} w_{\alpha} - \Gamma^{\nu}_{\mu\alpha} w_{\nu} \,. \tag{7.5}$$

The covariant derivative of a tensor $t^{\alpha\beta}$ is then

$$\nabla_{\mu} t^{\alpha\beta} = \partial_{\mu} t^{\alpha\beta} + \Gamma^{\alpha}_{\mu\sigma} t^{\sigma\beta} + \Gamma^{\beta}_{\mu\sigma} t^{\alpha\sigma} \,, \tag{7.6}$$

and of a tensor $t^{\alpha}{}_{\beta}$,

$$\nabla_{\mu} t^{\alpha}{}_{\beta} = \partial_{\mu} t^{\alpha}{}_{\beta} + \Gamma^{\alpha}_{\mu\sigma} t^{\sigma}{}_{\beta} - \Gamma^{\sigma}_{\mu\beta} t^{\alpha}{}_{\sigma} \,. \tag{7.7}$$

▸ ***NOW MAKE EXERCISES TO SHOW WHAT HAPPENS WITH THE ABOVE EXAMPLE***

From the exercise we can see that the above recipe indeed gives the correct answer even in curved coordinates. We can also prove that the covariant derivative of the metric itself is always zero.

▸ ***ANOTHER EXERCISE***

The covariant derivative ∇_{μ} produces, as its name says, covariant expressions. Therefore $\nabla_{\alpha} t^{\mu\nu}{}_{\gamma}$ is a perfectly valid tensor. We can also contract indices: $\nabla_{\alpha} t^{\alpha\nu}{}_{\gamma}$, or with help of the metric: $g^{\alpha\gamma} \nabla_{\alpha} t^{\mu\nu}{}_{\gamma}$. Since $\nabla_{\mu} g_{\alpha\beta} = 0$ (as we saw in the above exercise) we can always bring the $g_{\alpha\beta}$ and/or $g^{\alpha\beta}$ inside or outside the ∇_{μ} operator. We can therefore write

$$g^{\alpha\gamma} \nabla_{\alpha} t^{\mu\nu}{}_{\gamma} = \nabla_{\alpha} (t^{\mu\nu}{}_{\gamma} g^{\alpha\gamma}) = \nabla_{\alpha} t^{\mu\nu\alpha} \tag{7.8}$$

We can also define:

$$\nabla^{\alpha} = g^{\alpha\beta} \nabla_{\beta} \tag{7.9}$$

A

Tensors in special relativity

Although tensors are mostly useful in general relativity, they are also quite convenient in special relativity. Moreover, when we express the formulas using tensors, they expose the structure of the theory much better than the formulas without tensors do. We will start here at the point when four-vectors are introduced into special relativity.

From the simple thought experiments with moving trains, meant to give a visual understanding of the Lorentz transformations, it has become clear that 3-dimensional space and 1-dimensional time are not separate entities, but rather should be seen as one. In matrix form one found for a pure Lorentz transformation

$$\begin{pmatrix} x^{0'} \\ x^{1'} \\ x^{2'} \\ x^{3'} \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix} , \tag{1.1}$$

with $\gamma = 1/\sqrt{1 - v^2/c^2}$ and $\beta = \frac{v}{c}$. Moreover, we found that the inner product now satisfies the somewhat strange formula

$$\langle \vec{a}, \vec{b} \rangle = -a^0 b^0 + a^1 b^1 + a^2 b^2 + a^3 b^3 . \tag{1.2}$$

With the knowledge which we now have about the 'metric tensor', this formula can be understood much better. Namely, we know that the inner product of two vectors does not really exist (at least not in a form which is independent of the coordinate system). Rather, we should be talking about the inner product of a vector with a covector. We can obtain one from the other by raising or lowering an index, making use of the metric tensor,

$$a_\alpha = g_{\alpha\beta} a^\beta . \tag{1.3}$$

Eq. (1.2) should thus really be viewed as

$$\langle \vec{a}, \vec{b} \rangle_g = a^\mu b_\mu = g_{\mu\nu} a^\mu b^\nu . \tag{1.4}$$

where $\tilde{b}$ is a covector. From the fact that inner products such as (1.2) are invariant under Lorentz transformations we now find the components of the metric tensor,

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{1.5}$$

Minkowski space, which is the space in which the rules of special relativity hold, is thus a space which has a metric which is different from the metric in an ordinary Euclidean space. In order to do computations in this space, we do not have to introduce any strange inner products. Rather, we simply apply our new knowledge about inner products, which automatically leads us to the modified metric tensor.

Let us end with a small related remark. We have seen in Chapter (5) that the metric can also be used to describe curved spaces. The Minkowski metric which we have encountered in the present section can be used to describe curved *space-time*. This concept is at the basis of the general theory of relativity. In that theory, gravity is explained as a consequence of the curvature of space-time. A good introduction to this theory is given in [5].

B
Geometrical representation

The tensors as we have seen them in the main text may appear rather abstract. However, it is possible to visualise their meaning quite explicitly, in terms of geometrical objects. A simple example of this is the arrow-vector. An arrow-vector is effectively just a set of numbers that transforms in a particular way under a basis transformation. However, it is useful to think about it in terms of an arrow, so as to visualise its meaning.

We can make a similar geometrical visualisation of a covector, as well as for all other tensors. The present appendix contains a list of these geometrical representations. As we go along we will also describe how various manipulations with tensors (such as additions and inner products) can be translated to this geometrical picture. None of this is necessary in order to be able to compute with tensors, but it may help in understanding such computations.

- **arrow-vector**

  An arrow-vector is represented by (how else) an arrow in an n-dimensional space. The components of the vector (i.e. the numbers v^μ) are obtained by determining the projections of the arrow onto the axes. See figure B.1.

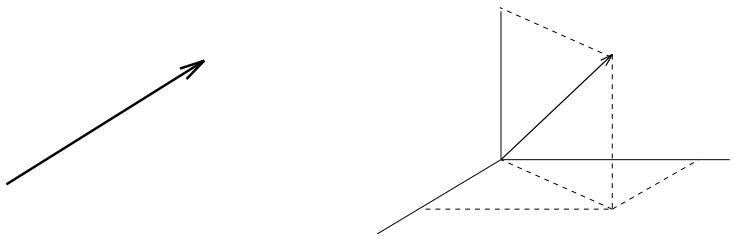

Figure B.1: Geometrical representation of a vector, and the determination of its components by projection onto the axes.

- **co-vector**

  A co-vectors is closely related to the gradient of a function $f(\vec{x})$, so it makes sense to look for a geometrical representation which is somehow connected to gradients. A useful related concept is that of contour lines (curves of equal height, in two dimensions) or contour surfaces (in three dimensions). Now consider two consecutive contour lines. The intersection they make with a neighbouring $x^2 =$ constant lijne is equal to the increase of x^1 necessary to

increase $f(\vec{x})$ by 1, if the other coordinate is kept constant. So in two dimensions,

$$\text{intersection at } x^1\text{-axis} = \left(\frac{\partial x^1}{\partial f}\right)_{x^2 \text{ constant}} = \left(\frac{\partial f}{\partial x^1}\right)^{-1}_{x^2 \text{ constant}} \tag{2.1}$$

Therefore, if we take the inverse of this intersection, we again obtain the first component of the gradient.

Given the above, it makes sense to use, for the generic representation of a co-vector, two consecutive lines (or, in three dimensions, planes), such that the intersections with the axes are equal to the components of the co-vector. Note that one should also keep track of the order of the lines or planes. In the representation sketched in figure B.2 we have drawn a flexible type of arrow. This is not an arrow-vector as above, but only meant to indicate the order of the surfaces. We could equally well have labelled the surfaces with different numbers or colours.

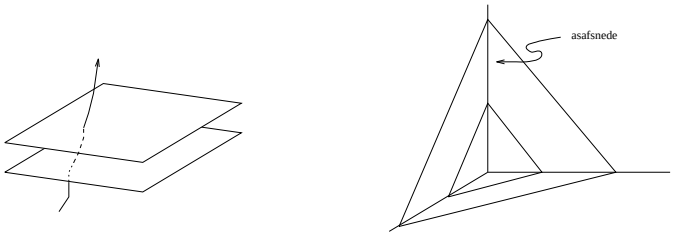

Figure B.2: Geometrical representation of a co-vector, and the determination of its components by computing the inverse of the intersection with the axes.

- **anti-symmetric contravariant second-order tensor** $t^{\mu\nu}$

  An anti-symmetric contravariant second order tensor can be made from two vectors according to

  $$t^{\mu\nu} = v^\mu w^\nu - w^\mu v^\nu \,. \tag{2.2}$$

  This is a bit like an outer product. The order of the two vectors is clearly of relevance. We can visualise this object as a surface with a given area and orientation. The components of the tensor can be obtained by measuring the projection of the surface onto the various coordinate surfaces. A negative component is obtained when the orientation of the surface assocatiated to the tensor is negative. Let us illustrate this with an example: if the surface of the projection on the x^1, x^2 surface is equal to 2, then the t^{12} component equals 2 and the t^{21} component equals -2. See figures B.3 and B.4.

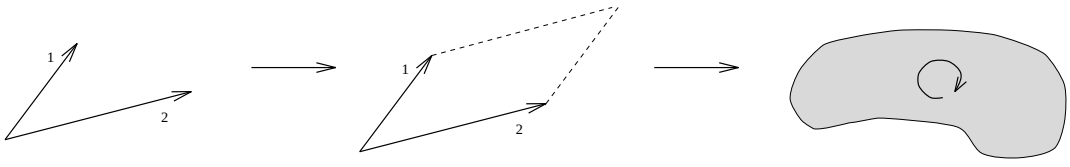

Figure B.3: Representation of an anti-symmetric contravariant 2nd-order tensor.

- **anti-symmetric covariant second order tensor** $t_{\mu\nu}$

  This object can again be seen as a kind of outer product of two vectors, or rather two co-vectors,

  $$t_{\mu\nu} = v_\mu w_\nu - w_\mu v_\nu \,. \tag{2.3}$$

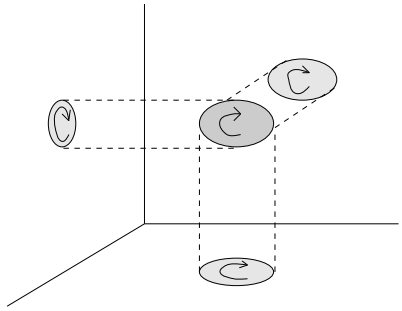

Figure B.4: Determination of the components of an anti-symmetric contravariant second order tensor.

We can visualise this object as a kind of tube with an orientation. The components can be found by computing the intersections of this tube with all coordinate surfaces, and inverting these numbers. Again one has to keep track of the orientation of the intersections. See figure B.5.

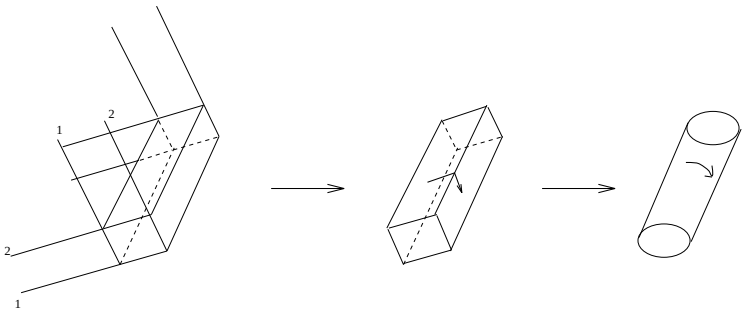

Figure B.5: Geometrical representation of an anti-symmetric covariant second order tensor.

- **addition of vectors**

  Addition of vectors is of course a well-known procedure. Simply translate one arrow along the other one until the top. This translated arrow now points to the sum-arrow. See figure B.6.

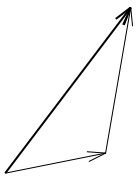

Figure B.6: Geometrical representation of the addition of vectors.

- **addition of co-vectors**

  The addition of covectors is somewhat more complicated. Instead of trying to describe this in words, it is easier to simply give the visual representation of this addition. See figure for a two-dimensional representation.

- **inner product between a vector and a co-vector**

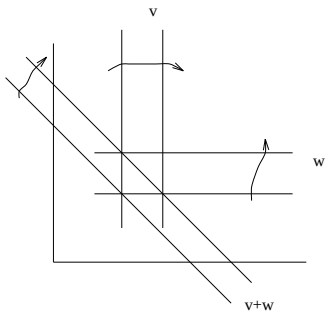

Figure B.7: Geometrical representation of the addition of co-vectors.

The inner product of a vector and a co-vector is the ratio of the length of the intersection which the co-vector has with the vector and the length of the vector. See figure B.8 for a three-dimensional representation.

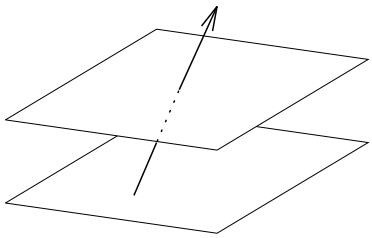

Figure B.8: Representation of the inner prdouct between a vector and a co-vector.

- **second order symmetric co-variant tensor** $g_{\mu\nu}$

  The metric tensor belongs to this class. It therefore makes sense to look for a representation which is related to the inner product. A suitable one is to consider the set of all points which have a unit distance to a given point $\vec{x}_0$,

  $$\{\vec{x} \in \mathbb{R}^3 \quad | \quad g_{\mu\nu} x^\mu x^\nu = 1\}\,. \tag{2.4}$$

  If the metric is position-dependent, $g = g(\vec{x})$, then the point $\vec{x}_0$ is equal to the point at which the corresponding metric $g_{\mu\nu}$ is evaluated.

  Now it is also clear what is the connection to the ellipses of chapter 5. See figure B.9 for a three-dimensional representation: an ellipsoid.

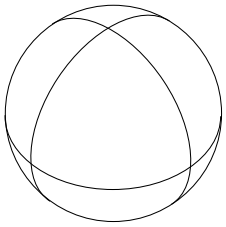

Figure B.9: Geometrical representation of a second order symmetric covariant tensor.

It is considerably more complicated to make a geometrical representation of a second order symmetric contravariant tensor, so we will omit that case.

- **turning an arrow-vector into a co-vector**

  The procedure of converting a vector to a co-vector can also be illustrated geometrically, see figure B.10 for a two-dimensional representation. The two long slanted lines are tangent to the circle, and the bottom horizontal line goes through the center of the circle. This representation fails when the arrow-vector is shorter than the radius of the circle.

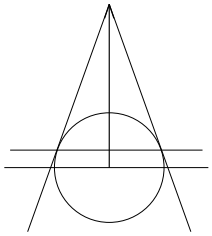

Figure B.10: Geometrical representation of turning an arrow-vector into a co-vector.

C
Exercises

C.1. Index notation

1. A, B and C are matrices. Assume that

$$A = BC$$

Write out this matrix multiplication using index notation.

2. A and B are matrices and x is a position vector. Show that

$$\sum_{\nu=1}^{n} A_{\mu\nu}(\sum_{\alpha=1}^{n} B_{\nu\alpha}x_\alpha) = \sum_{\nu=1}^{n}\sum_{\alpha=1}^{n}(A_{\mu\nu}B_{\nu\alpha}x_\alpha)$$

$$= \sum_{\alpha=1}^{n}\sum_{\nu=1}^{n}(A_{\mu\nu}B_{\nu\alpha}x_\alpha)$$

$$= \sum_{\alpha=1}^{n}(\sum_{\nu=1}^{n}(A_{\mu\nu}B_{\nu\alpha})x_\alpha)$$

3. Which of the following statements are true?

   (a) The summation signs in an expression can always be moved to the far left, without changing the meaning of the expression.

   (b) If all summation signs are on the far left of an expression, you can exchange their order without changing the meaning of the expression.

   (c) If all summation signs are on the far left of an expression, you cannot just change the order of the variables in the expression, because this changes the order in which matrices are multiplied, and generically $AB \neq BA$.

   (d) $A_{\mu\nu} = (A^T)_{\nu\mu}$

   (e) $A_{\mu\nu} = (A^T)_{\mu\nu}$

4. A, B, C, D and E are matrices. Write out the following matrix multiplications using index notation (with all summation signs grouped together).

   (a)

$$A = B(C + D)$$

   (b)

$$A = BCD$$

(c)

$$A = BCDE$$

5. Assume you have three vectors $\vec{x}, \vec{y}$ and $\vec{z}$ which satisfy the following relations,

$$\vec{y} = B\vec{x}$$

$$\vec{z} = A\vec{y}$$

Write these relations using index notation. Now write down the relation between $\vec{z}$ and $\vec{x}$ using index notation.

6. Write in matrix form

$$D_{\beta\nu} = \sum_{\mu 1}^{n} \sum_{\alpha=1}^{n} A_{\mu\nu} B_{\alpha\mu} C_{\alpha\beta} \,.$$

7. Try a couple of the previous exercises by making use of the summation convention.

8. Write as a matrix multiplication:

(a)

$$D_{\alpha\beta} = A_{\alpha\mu} B_{\mu\nu} C_{\beta\nu}$$

(b)

$$D_{\alpha\beta} = A_{\alpha\mu} B_{\beta\gamma} C_{\mu\gamma}$$

(c)

$$D_{\alpha\beta} = A_{\alpha\gamma}(B_{\gamma\beta} + C_{\gamma\beta})$$

9. Consider a vector field in an n-dimensional space,

$$\vec{v}(\vec{x}) \,.$$

We perform a coordinate transformation,

$$\vec{x}' = A\vec{x} \qquad A \quad \text{is a transformation matrix.}$$

Show that

$$\vec{v}' = A\vec{v} \,,$$

where the matrix A is the same matrix as in the first equation.

10. For a transformation we have

$$\vec{x}' = A\vec{x} \,.$$

This corresponds to

$$x'_{\mu} = \sum_{\nu=1}^{n} A_{\mu\nu} x_{\nu} \,.$$

Can you understand the expression

$$\sum_{\nu=1}^{n} x_{\nu} A_{\mu\nu} = ?$$

And how can you construct the matrix multiplication equivalent of

$$\sum_{\mu=1}^{n} x_{\mu} A_{\mu\nu} = ?$$

(note the position of the indices).

C.2. Co-vectors

1. The matrix for a rotation over an angle ϕ in the x-y plane is given by

$$\Lambda = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} .$$

Compute the inverse of this matrix (either by replacing ϕ by $-\phi$ or by matrix inversion). Show that $(\Lambda^{-1})^T$ is equal to Λ.

2. In Cartesian coordinate systems the basis vectors are orthogonal to each other,

$$\vec{e}_1 \cdot \vec{e}_2 = 0 .$$

(This is a somewhat misleading notation. You might be tempted to interpret this as the inner product of two basis vectors, in which the components of the basis vectors are expressed in the basis itself. That would of course *always* yield a trivial zero).

If we transform from one such Cartesian coordinate system to another one,

$$\vec{e}_1{}' = \Lambda\vec{e}_1 \quad \& \quad \vec{e}_2{}' = \Lambda\vec{e}_2 ,$$

(where Λ is the transformation matrix) then this relation should of course continue to hold,

$$\vec{e}_1{}' \cdot \vec{e}_2{}' = 0 .$$

Insert the transformation rule given above and show that, for transformations between two Cartesian systems, the tranformation matrix Λ is equal to $(\Lambda^{-1})^T$. These are thus orthogonal transformations.

C.3. Introduction to tensors

1. With the help of various mathematical operators you can make new tensors out of old ones. For instance

$$w_\mu = t_{\mu\nu} v^\nu ,$$

in which $t_{\mu\nu}$ and v^ν are tensors.

   (a) Show that in this case w_μ is also a tensor (that is, show that it transforms correctly under a basis transformation).
   (b) Show that w^μ is *not* a tensor if we make it according to

$$w^\mu = t^{\mu\nu} v^\nu .$$

2. A matrix can be viewed as an object which transforms one vector into another one. We have seen that this can be a normal transformation or a basis transformation. The question is now: does a normal transformation matrix transform too upon a change of basis. And how? Write this out in normal matrix notation, with S the matrix which transforms from the original system to the primed system, and A the matrix which is to be transformed.

Hint:

$$\vec{y} = A\vec{x} ,$$

and after a basis (or coordinate) transformation

$$A\vec{y} = S(A\vec{x}) ,$$

but there is also a matrix A' which satisfies

$$S\vec{y} = A'(S\vec{x}) .$$

3. Show that, if you start with a number of tensors and construct a new object out of them, and use the summation convention (that is, only sum over one upper and one lower index, never over two indices at the same position), the new object is again a tensor.

C.4. Tensoren, algemeen

1. You have seen that there are four types of 2nd-order tensors. How many 3rd order tensors are there?
2. Take an anti-symmetric tensor $t^{\mu\nu}$. Show that the property of anti-symmetry is preserved under a basis transformation. Do the same thing for a symmetric tensor. This shows that symmetry and anti-symmetry are fundamental properties of tensors.
3. Show, by writing out the transformation of a $(1,1)$-tensor, op te schrijven, dat het geen zin heeft om te spreken van symmetrie of anti-symmetrie van een tensor in indices die niet van hetzelfde soort zijn. (Laat dus zien dat die eigenschap na transformatie meestal verloren is).
4. Gegeven een tweede orde contravariante symmetrische tensor t en een tweede orde covariante antisymmetrische tensor r. Toon aan dat de dubbele contractie

   $$t^{\alpha\beta} r_{\alpha\beta}$$

   altijd gelijk is aan nul.
5. De vogende tensor is gedefinieerd uit twee vectoren:

   $$t^{\mu\nu} = v^\mu w^\nu - w^\nu v^\mu$$

   Laat zien dat deze antisymmetrisch is.
6. Hoe transformeert een kronecker-delta? De kronecker-delta is symmetrisch; waarom blijft de eigenschap van symmetrie hier wel behouden, terwijl toch is bewezen dat de symmetrie in een boven en onderindex niet behouden blijft?
7. We hebben in opgave 2 in paragraaf C.3 gezien hoe een matrix A transformeert. Laat aan de hand daarvan zien hoe de matrix A^T transformeert. Toon aan de hand van de twee gevonden vergelijkingen aan dat het object g uit de tekst niet transformeert als een matrix, maar ook niet als de getransponeerde van een matrix.
8. Neem twee tensoren: $s^{\mu\nu}$ en $t_{\mu\nu}$. Maak een product van die twee dat
   (a) geen vrije indices heeft.
   (b) twee vrije indices heeft.
   (c) vier vrije indices heeft.
9. Hoe transformeert een 3^e orde geheel covariante tensor? Kan dit in matrixvorm geschreven worden?
10. Je kunt een niet-lineaire coördinatentransformatie geven door de nieuwe coördinaten uit te drukken als een functie van de oude coördinaten:

    $$\begin{aligned} x'^1 &= f'^1(x^1, x^2 \ldots, x^n) \\ &\vdots \\ x'^n &= f'^n(x^1, x^2 \ldots, x^n) \end{aligned}$$

Deze functies zijn nu naar hun n coördinaten te Taylor-ontwikkelen om het punt $\vec{0}$. Maak zo'n Taylor-ontwikkeling.

Aanwijzing:

Bij het afleiden van de Taylorreeks van een gewone functie ging men uit van een eenvoudige machtreeks:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \dots$$

Door deze vergelijking steeds te differentiëren kon men de coëfficienten $a_{0\dots n}$ bepalen. Doe iets dergelijks ook in dit geval.

C.5. Metrische tensor

1. Laat zien dat het onmogelijk is om een metrische tensor die zowel 1'en als -1'en op de hoofddiagonaal heeft, te transformeren naar een vorm waarin er nog uitsluitend 1'en aanwezig zijn.

2. Het nieuwe inproduct, met de metrische tensor, is invariant onder coördinatentransformaties:
$$g'_{\mu\nu} x'^{\mu} x'^{\nu} = g_{\mu\nu} x^{\mu} x^{\nu}$$
Bij de behandeling van de Speciale Relativiteitstheorie komen we echter ook vaak de volgende vergelijking tegen:
$$g_{\mu\nu} x'^{\mu} x'^{\nu} = g_{\mu\nu} x^{\mu} x^{\nu}$$
(dus zonder accent bij g) Leg uit waarom we deze vergelijking mogen gebruiken in alle gevallen dat g de Minkowskimetriek is en we uitsluitend in lorentzframes werken.

C.6. Tensor calculus

1. Bewijs dat de volgende vergelijkingen covariant zijn:
$$z^{\alpha\beta} = x^{\beta\alpha}$$
$$z^{\alpha\beta} = x^{\alpha\beta} + y^{\beta\alpha}$$

2. Stel, je hebt 2 berekeningen uitgevoerd:
$$x^{\alpha} = \dots$$
en
$$y^{\beta} = \dots$$
Hoe kan je nu, terwijl de indices van x en y verschillend zijn, toch deze twee vectoren optellen?

3. Waarom is het bij
$$(\mathrm{grad}\vec{v})^{\mu}{}_{\nu} = \frac{\partial v^{\mu}}{\partial x^{\nu}}$$
niet nodig om de indices (aan de linkerkant van de vergelijking) op volgorde te zetten?

C
Bibliography

[1] Bäuerle G. G. A. (1981) *Syllabus 'Algemene Relativiteitstheorie I'* (UvA)

[2] Dullemond C. (1986) *Syllabus 'Algemene Relativiteitstheorie'* (KUN)

[3] Kay D. C. (1988) *'Tensor Calculus (Schaum's outline series)'* (McGraw-Hill)

[4] Schouten J. A. (1951) *'Tensor Analysis for Physicists'* (Oxford)

[5] Schutz B. F. (1985) *'A first course in general relativity'* (Cambridge)

[6] Takens R. J. (1985) *Syllabus 'Hemelmechanica'* (UvA)